Intrinsic filtering on Lie groups with applications to attitude estimation

Axel Barrau, Silvère Bonnabel*

September 15, 2014

Abstract

This paper proposes a probabilistic approach to the problem of intrinsic filtering of a system on a matrix Lie group with invariance properties. The problem of an invariant continuous-time model with discrete-time measurements is cast into a rigorous stochastic and geometric framework. Building upon the theory of continuous-time invariant observers, we introduce a class of simple filters and study their properties (without addressing the optimal filtering problem). We show that, akin to the Kalman filter for linear systems, the error equation is a Markov chain that does not depend on the state estimate. Thus, when the filter's gains are held fixed, the noisy error's distribution is proved to converge to a stationary distribution, under some convergence properties of the filter with noise turned off. We also introduce two novel tools of engineering interest: the discrete-time Invariant Extended Kalman Filter, for which the trusted covariance matrix is shown to converge, and the Invariant Ensemble Kalman Filter. The methods are applied to attitude estimation, allowing to derive novel theoretical results in this field, and illustrated through simulations on synthetic data.

1 Introduction

The problem of probabilistic filtering aims at computing the conditional expectation $\hat{x}(t)$ of the state $x(t)$ of a system driven by process and observation noises conditioned

*A. Barrau and S. Bonnabel are with MINES ParisTech, PSL Research University, Centre for robotics, 60 Bd St Michel 75006 Paris, France. [axel.barrau,silvere.bonnabel]@mines-paristech.fr

on the past inputs and outputs. Filtering is a branch of stochastic process theory that has been strongly driven by applications in control and signal processing. In the particular case where the system is linear and corrupted by independent white Gaussian process and observation noises, the celebrated Kalman filter [20] solves the filtering problem. Yet, when the system is non-linear there is no general method to derive efficient filters, and their design has encountered important difficulties in practice. Indeed, general filtering equations describing the evolution of the state conditioned on past outputs can also be derived, see e.g., [10] and references therein. However, the equations are not closed, and in most cases the conditional densities can not be computed by solving a finite dimensional equation.

In engineering applications the most popular solution is the extended Kalman filter (EKF), which does not possess general theoretical properties, especially in a stochastic context, and is sometimes prone to divergence. The EKF amounts to linearize the system around the estimated trajectory, and build a Kalman filter for the linear model, which can in turn be implemented on the non-linear model. Despite the fact that it is subject to several problems both theoretical and practical, one merit of the EKF is to convey an estimation of the mean state $\hat{x}(t)$ together with an approximation of the covariance matrix $P(t)$ of the error, yielding a (trusted) ellipsoidal credible set that reflects the level of uncertainty in the estimation. In order to capture more closely the distribution, several rather computationally extensive methods have recently attracted growing interest such as particle filters and unscented Kalman filters (UKF) [19].

In this paper we consider a general filtering problem on Lie groups in a stochastic setting, and a class of simple

recursive finite dimensional filters, having a similar structure as the EKF, is introduced. Such recursively defined filters allow straightforward digital computer implementation. Besides, as the Kalman filter, the corresponding algorithms will be readily implementable with very little understanding of the theory leading to their development, as illustrated by the concrete examples of sections 4.2, 4.3 and 5.3, which can be read independently of the rest of the paper. As concerns their theoretical properties, they are motivated by the three following desirable properties a simple recursive filter should ideally have: 1- forgetting of the initial condition, 2- being able to capture the extent of uncertainty in the estimate, 3- being efficient, that is, minimizing the error's dispersion within a given class. In the filtering problem, the estimation error is initially distributed according to a prior distribution provided by the practitioner. The rationale for the first property is that, as the filter acquires information, the prior information should be given less (eventually no) importance by the filter. As the user generally possesses very little information on the system's state initially, the prior can be very far from the truth. Forgetting the initial condition allows the (subjective Bayesian) prior distribution to be as informative as noninformative without impacting the final estimates. Moreover, it corresponds to convergence properties of the filter, and thus deals with the ability to recover from catastrophic failures. In robotics, this leads to robustness to erroneous beliefs due for instance to perceptual ambiguity. The second property deals with the central motivation behind stochastic approaches. Indeed, all deterministic observers, no matter how strong properties they may have, have a common fundamental weakness: in the presence of sensor and model uncertainties, they provide no indication of the extent of uncertainty conveyed by the estimate. Being able to compute credible sets for the estimate may be needed in order to adapt control decisions to the situation, for instance to guarantee an absence of collision. The two first properties are general and apply to any class of sensible filters. The third property addresses the issue of gain tuning. When working with a class of (suboptimal) filters, the gains can be tuned over a training set in order to minimize the level of uncertainty conveyed by the estimate.

Several approaches to filtering for systems possessing a geometric structure have been developed in previous literature. For stochastic processes on Riemannian manifolds [18, 14] some results have been derived, see e.g., [29]. 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. [33, 13] and more recently [36]. For systems on Lie groups powerful tools to study the filtering equations (such as harmonic analysis [45, 32, 24, 43]) have been used, notably in the case of bilinear systems [44] and estimation of the initial condition of a Brownian motion in [15]. An extension of Gaussian distributions has also been developed and recently used for Bayesian filtering on Lie groups in [46] and also in [7] which devises a general adaptation of the EKF to Lie groups, without exploiting the geometrical properties of the specific invariant case such as in the present paper. 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 [9]. A certain class of least squares problems on the Euclidian group has been tackled in [17]. This approach has also been used to filter inertial data recently in [34], [26].

Although the paper is concerned with a general theory on (matrix) Lie groups, the several derived filters will systematically be applied to a benchmark and motivating filtering example on the special orthogonal group $SO(3)$. The problem of attitude estimation consists of estimating the orientation of a rigid body, which proves to be a central task in aeronautics, in order to stabilize or guide the system, or in mobile robotics, where orientation is a part of the robot's localization. When the system is equipped with gyroscopes, the delivered signals can be integrated over time to compute an orientation increment. However, the initial orientation may not be known, and even if it is, the gyroscopes are subject to noise and drift, so that pure integration may quickly result in a very erroneous attitude. Thus the estimations are regularly checked against some vectors' orientation, such as the earth magnetic field measured by magnetometers, or the earth gravity vector measured by accelerometers under quasi-stationary flight assumptions. The explosion over the last decade of low-cost noisy sensors integrated in unmanned aerial vehicles (UAV's), and the limited computational power on-board, have been a driving force for the development of simple filters on $SO(3)$. Work on the deterministic problem dates back to [37, 30] and has received much attention recently,

see [42, 27, 26, 4, 28, 38] to cite a few, and some related issues have been addressed as well, like angular velocity estimation [40] or attitude control [37, 39]. The beauty of the intrinsic problem formulation on $SO(3)$ and the properties stemming from it have been exploited in various papers in the deterministic case, see e.g. [26, 21, 5]. The problem has also been cast into a minimum energy filtering framework in [25]. In the stochastic framework, a general review of the classical techniques based on Gaussian assumptions is proposed in [27]. A first attempt to systematically exploit the invariance properties to design stochastic filters on $SO(3)$ appears in the preliminary work [1].

The contributions and the organization of the present paper are as follows. In Section 2, the problem of stochastic filtering on Lie groups with a continuous-time noisy model and discrete-time observations is rigorously posed. From a theoretical viewpoint, the choice of discrete-time measurements allows to circumvent some technical difficulties and to cast the problem into a perfectly rigorous framework. As this work is mainly guided by applications to aerospace, this choice is relevant as the evolution equations of the orientation generally rely on the integration of noisy inertial measurements that are delivered at a high rate, whereas the observations may be delivered at low(er) rate (GPS, vision). In order to derive proper filters, the problem is then transformed into a complete discrete-time model. The discretization is not straightforward because unlike the linear case, difficulties arise from non-commutativity.

Guided by the theory of symmetry-preserving observers that has been exclusively concerned with the continuous-time and deterministic case so far, we propose in Section 3 a class of natural filters. They appear as a mere transposition of linear observers (such as the linear Kalman filter) to the group, but where the addition has been replaced with the group multiplication law (due to non-commutativity, there is always some freedom in the transposition as left and right multiplications are different operations). We prove the proposed filters retain the invariances of the problem, and that the discrete-time error's evolution is independent of the system's trajectory, inheriting the properties of the deterministic continuous-time case [5] (or merely the linear case). Thus, the gains can be computed off-line, and the proposed filters require very few on-board computational resources. This is one

of their main advantages over current filters for the attitude estimation problems considered in simulations.

The proposed class is very broad and the tuning issue is far from trivial. Sections 4 and 5 explore two different routes, depending on if we are interested in the asymptotic or transitory phase. In Section 4, we propose to hold the gains fixed over time. As a result, the error equation becomes a *homogeneous* Markov chain. We first prove some new results about global convergence in a deterministic and discrete-time framework. Then, building upon the homogeneity property of the error, we prove that if the filter with noise turned off admits almost global convergence properties, the error with noise turned on converges to a stationary distribution. Mathematically this is a very strong and to some extent surprising result. From a practical viewpoint, the gains can be tuned numerically to minimize the asymptotic error's dispersion. This allows to "learn" sensible gains for very general types of noises. The theory is applied to two examples and gives convergence guarantees in each case. First an attitude estimation problem using two vector measurements and a gyroscope having isotropic noise (Section 4.2), then the construction of an artificial horizon with optimal gains, for a non-Gaussian noise model (Section 4.3). Each application is a contribution in itself and can be implemented without reading the whole paper.

In Section 5, we propose to optimize the convergence during the transitory phase using Gaussian approximations. We first introduce a method of engineering interest: the Invariant Extended Kalman Filter (IEKF) in discrete time. The IEKF was introduced in continuous time in [3] and developed and applied to localization problems in [6] and very recently in [2]. A discrete time version on $SO(3)$ was proposed in the preliminary paper [1]. The idea consists of linearizing the invariant error for the class of filters introduced in Section 3 around a fixed point (the identity element of the matrix group), build a Kalman filter for the linear model and implement it on the nonlinear model. As the linearizations always occur around the same point, the linearized model is time-invariant and thus the Kalman gains, as well as the Kalman covariance matrix, are proved to converge to fixed values. When onboard storage and computational resources are very limited, this advantageously allows to replace the gain with its asymptotic value. The IEKF is compared to the well-known MEKF [22, 12] and UKF [19, 11] on the attitude

estimation problem, and simulations illustrate some convergence properties that the latter lack. In the case where the error equation is totally autonomous, we introduce a new method based on off-line simulations, the IEnKF, which outperforms the other filters in case of large noises by capturing very accurately the error's dispersion.

2 Problem setting

2.1 Considered continuous-time model

Consider a state variable χ_t taking values in a matrix Lie group G with neutral element I_d, and the following continuous-time model with discrete measurements:

$$\frac{d}{dt}\chi_t = (\upsilon_t + w_t)\chi_t + \chi_t \omega_t \tag{1}$$

$$Y_n = h(\chi_{t_n}, V_n) \tag{2}$$

where υ_t and ω_t are inputs taking their values in the Lie algebra $\mathfrak{g}$ (i.e. the tangent space to the identity element of G), w_t is a continuous white Langevin noise with diffusion matrix Σ_t, whose precise definition will be discussed below in Subsection 2.3. $(Y_n)_{n\geqslant 0}$ are the discrete-time observations, belonging to some measurable space $\mathcal{Y}$ and V_n is a noise taking values in $\mathbb{R}^p$ for an integer $p > 0$. We further make the following additional assumption:

Assumption 1 (left-right equivariance) The output map h is left-right equivariant, i.e. there exists a left action of G on $\mathcal{Y}$ such that we have the equality in law:

$$\forall g, \chi \in G, h(\chi g, V_n) \overset{\mathscr{L}}{=} g^{-1}h(\chi, V_n) \tag{3}$$

$$\text{and } h(\chi g, 0) = g^{-1}h(\chi, 0) \tag{4}$$

The reader who is not familiar with stochastic calculus and Lie groups can view χ_t as a rotation matrix, and replace the latter property with the following more restrictive assumption:

Assumption 1 bis For any $n \geqslant 0$ there exists a vector b such that $h(\chi, V_n) = \chi^{-1}(b + V_n)$.

The assumptions could seem restrictive but are verified in practice in various cases as shown by the following examples, which will provide the reader with a more concrete picture than the formalism of Lie groups.

2.2 Examples

2.2.1 Attitude estimation on flat earth

Our motivating example for the model (1)-(2) is the attitude estimation of a rigid body assuming the earth is flat, and observing two vectors:

$$\frac{d}{dt}R_t = R_t(\omega_t + w_t^\omega)_\times,$$

$$Y_n = (R_{t_n}^T b_1 + V_n^1, R_{t_n}^T b_2 + V_n^2), \tag{5}$$

where as usual $R_t \in SO(3)$ represents the rotation that maps the body frame to the earth-fixed frame, and where $\omega_t \in \mathbb{R}^3$ is the instantaneous rotation vector, and $w_t^\omega \in \mathbb{R}^3$ is a continuous Gaussian white noise representing the gyroscopes' noise. We have let $(x)_\times \in \mathbb{R}^{3\times 3}$ denote the skew matrix associated with the cross product with a three dimensional vector, i.e., for $a, x \in \mathbb{R}^3$, we have $(x)_\times a = x \times a$. $(Y_n)_{n\geq 0}$ is a sequence of discrete noisy measurements of two vectors b_1, b_2 of the earth fixed frame verifying $b_1 \times b_2 \neq 0$, and V_n^1 and V_n^2 are sequences of independent isotropic Gaussian white noises. Note that the noise w_t^ω is defined, as the input ω_t, in the body frame (in other words it is multiplied on the left). Thus equations (5) do not match (1)-(2) which correspond to a noise defined in the earth-fixed frame. This can be remedied in the particular case where the gyroscope noise is isotropic, a restrictive yet relevant assumption in practice.

Definition 1. *A Langevin noise w_t of $\mathfrak{g}$ is said isotropic if $\tilde{w}_t := gw_t g^{-1} \overset{\mathscr{L}}{=} w_t$ for any $g \in G$.*

Note that for a noise taking values in $so(3)$ this definition corresponds to the physical intuition of an isotropic noise of $\mathbb{R}^3$. With this additional assumption the equation can be rewritten:

$$\frac{d}{dt}R_t = (\tilde{w}_t^\omega)_\times R_t + R_t(\omega_t)_\times,$$

$$Y_n = (R_{t_n}^T b_1 + V_n^1, R_{t_n}^T b_2 + V_n^2),$$

which corresponds to (1)-(2).

Remark 1. *If the gyrometer noise is not isotropic the new noise $\tilde{w}_t^\omega$ is related to R_t by $\tilde{w}_t^\omega = R_t w_t^\omega R_t^T$. Depending on the degree of anisotropy this can prevent the use of methods based on the autonomy of the error (see 4) but not of the discrete-time IEKF (see 5.1).*

2.2.2 Attitude estimation on a round rotating Earth

Another interesting case is attitude estimation on $SO(3)$ using an observation of the vertical direction (given by an accelerometer) and taking into account the rotation of the earth. Whereas the flat earth assumption perfectly suits low-cost gyroscopes, precise gyroscopes can measure the complete attitude by taking into account the earth's rotation. We define a geographic frame with axis North-west-up. The attitude R_t is the transition matrix from the body reference to the geographic reference. The earth instantaneous rotation vector $\upsilon \in \mathbb{R}^3$ and the gravity vector $g \in \mathbb{R}^3$ are expressed in the geographic reference. Both are constant in this frame. The gyroscope gives a continuous rotation speed ω_t, disturbed by an isotropic continuous white noise w_t^ω. The equation of the considered system reads:

$$\frac{d}{dt}R_t = (\upsilon)_\times R_t + R_t(\omega_t + w_t^\omega)_\times,$$
$$Y_n = R_{t_n}^T(g + V_n)$$

As w_t^ω is supposed isotropic the equation can be rewritten:

$$\frac{d}{dt}R_t = (\upsilon + \tilde{w}_t^\omega)_\times R_t + R_t(\omega_t)_\times,$$
$$Y_n = R_{t_n}^T(g + V_n),$$

which corresponds to (1)-(2).

2.2.3 Attitude and velocity estimation

We give here an example on a larger group. Consider the attitude $R_t \in SO(3)$ and speed $v_t \in \mathbb{R}^3$ of an aircraft evolving on a flat earth, equipped with a gyroscope and an accelerometer. The gyroscopes gives continuous increments $\psi_t \in \mathbb{R}^3$ with isotropic noise $w_t^\psi \in \mathbb{R}^3$, and the accelerometers gives continuous increments $a_t \in \mathbb{R}^3$ with isotropic noise $w_t^a \in \mathbb{R}^3$. The aircraft has noisy speed measurements in the earth reference frame $Y_n = v_{t_n} + V_n$, where V_n is a supposed to be a Gaussian isotropic white noise. The equations read:

$$\frac{d}{dt}R_t = R_t(\psi_t + w_t^\psi)_\times,$$
$$\frac{d}{dt}v_t = R_t(a_t + w_t^a) + g,$$

where R_t is the rotation mapping the body frame to the earth-fixed frame, v_t is the velocity expressed in the earth fixed frame, and g is the earth gravity vector, supposed to be (locally) constant. Using the matrix Lie group $SE(3)$ we introduce:

$$A_t = \begin{pmatrix} R_t^T & R_t^T v_t \\ 0 & 1 \end{pmatrix} \; ; \; \omega_t = \begin{pmatrix} 0 & g \\ 0 & 0 \end{pmatrix}$$

$$\upsilon_t = \begin{pmatrix} -(\psi_t)_\times & a_t \\ 0 & 0 \end{pmatrix} \; ; \; w_t = \begin{pmatrix} (w_t^\psi)_\times & w_t^a \\ 0 & 0 \end{pmatrix}$$

The problem can then be rewritten under the form (1)-(2):

$$\frac{d}{dt}A_t = (\upsilon_t + w_t)A_t + A_t\omega_t,$$
$$Y_{t_n} = A_{t_n}^{-1}\begin{pmatrix} V_n \\ 1 \end{pmatrix},$$

2.3 Interpretation of Langevin noises on Lie Groups

This present subsection provides some mathematical considerations about white noises on Lie groups. It can be skipped by the uninterested reader who is directly referred to the model (6).

Equation (1) is actually a Langevin equation that suffers from poly-interpretability because of its non-linearity, and its meaning must be clarified in the rigorous framework of stochastic calculus. Stratonovich stochastic differential equations on a Lie group can be intrinsically defined as in e.g. [23]. A somewhat simpler (but equivalent) approach consists of using the natural embedding of the matrix Lie group G in a matrix space and to understand the equation in the sense of Stratonovich, see e.g. [35]. The mathematical reasons stem from the fact that the resulting stochastic process is well-defined on the Lie group, whereas this is not the case when opting for an Ito interpretation as underlined by the following easily provable proposition.

Proposition 1. *If the stochastic differential equation* (1) *is taken in the sense of Itô, for $G = SO(3)$ embedded in $\mathbb{R}^{3\times3}$, an application of the Itô formula to $\chi_t^T \chi_t$ shows that the solution almost surely leaves the submanifold G for* any *time $t > 0$.*

Besides, the physical reasons for this stem from the fact that the sensors' noise are never completely white, and for colored noise the Stratonovich interpretation provides a better approximation to the true solution than Ito's, as advocated by the result of Wong and Zakai [47].

2.4 Exact discretization of the considered model

To treat rigorously the problem of integrating discrete measurements we need to discretize the continuous model with the same time step as the measurements'. Unlike the general case of non-linear estimation, the exact discrete-time dynamics corresponding to Equation (1) can be obtained, as proved by the following result:

Theorem 1. *Let $\chi_n = \chi_{t_n}$. Then the discrete system (χ_n, Y_n) satisfies the following equations:*

$$
\begin{aligned}
\chi_{n+1} &= \Upsilon_n W_n \chi_n \Omega_n, \\
Y_n &= h(\chi_n, V_n),
\end{aligned}
\tag{6}
$$

where W_n is a random variable with values in G and whose law depends on the values taken by υ_t for $t \in [t_n, t_{n+1}]$ and on the law of w_t for $t \in [t_n, t_{n+1}]$, and Υ_n and Ω_n are elements of G which only depend on the values taken respectively by υ_t and ω_t for $t \in [t_n, t_{n+1}]$.

Proof. For $n \in \mathbb{N}$ consider the value of the process χ_t is known until time t_n. Let Υ_t and Ω_t be the solutions of the following equations:

$$
\Upsilon_{t_n} = I_d, \ \frac{d}{dt}\Upsilon_t = \upsilon_t \Upsilon_t
$$

$$
\text{and } \Omega_{t_n} = I_d, \ \frac{d}{dt}\Omega_t = \Omega_t \omega_t
$$

Let W_t be the solution of the following (Stratonovich) stochastic differential equation:

$$
W_{t_n} = I_d, \qquad \frac{d}{dt}W_t = \Upsilon_t^{-1} w_t \Upsilon_t W_t
$$

Note that $\Upsilon_t^{-1} w_t \Upsilon_t$ being in $\mathfrak{g}$, W_t is ensured to stay in G. Define the process $\chi_{t|t_n} = \Upsilon_t W_t \chi_{t_n} \Omega_t$. We will show that for $t > t_n$ the processes χ_t and $\chi_{t|t_n}$ verify the same

stochastic differential equation. Indeed:

$$
\begin{aligned}
\frac{d}{dt}\chi_{t|t_n} &= (\frac{d}{dt}\Upsilon_t)W_t\chi_t\Omega_t + \Upsilon_t(\frac{d}{dt}W_t)\chi_t\Omega_t + \Upsilon_t W_t\chi_t(\frac{d}{dt}\Omega) \\
&= \upsilon_t\Upsilon_t W_t\chi_t\Omega_t + w_t\Upsilon_t W_t\chi_t\Omega_t + \Upsilon_t W_t\chi_t\Omega_t\omega_t \\
&= (\upsilon_t + w_t)\chi_{t|t_n} + \chi_{t|t_n}\omega_t
\end{aligned}
$$

Thus the two processes have the same law at time t_{n+1}, i.e. χ_{n+1} and $W_{t_{n+1}}\Upsilon_{t_{n+1}}\chi_{t_n}\Omega_{t_{n+1}}$ have the same law. Letting $\Upsilon_n = \Upsilon_{t_{n+1}}$, $\Omega_n = \Omega_{t_{n+1}}$ and $W_n = W_{t_{n+1}}$ we obtain the result. $\qquad\square$

Remark 2. *In many practical situations (for instance examples 2.2.1 and 2.2.2), the Langevin noise w_t is isotropic and we have thus $\Upsilon_t^{-1}w_t\Upsilon_t \overset{\mathscr{L}}{=} w_t$. Note that, the variable W_t depends also only on the law of w_t for $t \in [t_n, t_{n+1}]$.*

In the sequel, for mathematical reasons (the equations do not suffer from poly-interpretability), tutorial reasons (the framework of diffusion processes on Lie groups needs not be known), and practical reasons (any filter must be implemented in discrete time), we will systematically consider the discrete-time model (6). Moreover, the noise W_n will be a general random variable in G, not necessary a solution of the stochastic differential equation $\frac{d}{dt}W_t = \Upsilon_t^{-1}w_t\Upsilon_t W_t$.

3 A class of discrete-time intrinsic filters

3.1 Preliminary: Kalman filter for linear stationary systems with drift

Consider the particular case of a linear system in $\mathbb{R}^N$ (with arbitrary N) of the form $\frac{d}{dt}X_t = b_t + w_t$, w_t being a white noise and b_t a known deterministic input. Assume we have discrete measurements of the form $Y_n = HX_{t_n} + V_n \in \mathbb{R}^p$, V_n being a Gaussian noise in $\mathbb{R}^p$. A straightforward discretization yields:

$$
\begin{aligned}
X_{n+1} &= X_n + B_n + W_n, \\
Y_n &= HX_n + V_n,
\end{aligned}
\tag{7}
$$

With $B_n = \int_{t_n}^{t_{n+1}} b_s ds$ and $W_n = \int_{t_n}^{t_{n+1}} w_s ds$ is the value of a Brownian motion in $\mathbb{R}^N$. The optimal filter in this case is

the celebrated linear Kalman filter whose equations are:

$$\hat{X}'_{n+1} = \hat{X}_n + B_n,$$
$$\hat{X}_{n+1} = \hat{X}'_{n+1} + K_{n+1}(Y_{n+1} - H\hat{X}'_{n+1}) \tag{8}$$

The first equation is the so-called prediction step, and the second one the correction step. Let $e'_n = \hat{X}'_n - X_n$ and $e_n = \hat{X}_n - X_n$ denote the predicted and corrected state estimation errors, we get:

$$e'_{n+1} = e_n - W_n,$$
$$e_{n+1} = e'_{n+1} - K_{n+1}(He'_{n+1} - V_{n+1}) \tag{9}$$

The error equation is autonomous and can be optimized independently from the dynamical inputs B_n's. As a consequence, the optimal gains K_n can be computed off-line. It is a well-known result that they (as well as the error distribution) converge under observability and detectability assumptions, so all computations can be done off-line and only the first values need be stored, in which case only the equations (8) must be computed on-line, requiring extremely restricted computational power on board. Mimicking the form of equations (8) on the Lie group G where the addition is naturally replaced with the group multiplication law, leads to a similar result, as shown in the next section.

3.2 Proposed intrinsic filters

Inspired by the linear case and the theory of continuous-time symmetry-preserving observers on Lie groups [5] we propose to mimic the prediction and update steps (8) replacing the addition with the group multiplication yielding the class of filters defined by:

$$\hat{\chi}'_{n+1} = \Upsilon_n \hat{\chi}_n \Omega_n, \tag{10}$$
$$\hat{\chi}_{n+1} = K_{n+1}(\hat{\chi}'_{n+1} Y_{n+1})\hat{\chi}'_{n+1}, \tag{11}$$

where $K_{n+1}(\bullet)$ can be any function of $\mathscr{Y} \to G$, ensuring $K(h(I_d, 0)) = I_d$. There are two ways to understand the links between (8) and (10)-(11). First, by defining a an estimation error on the group G by $\chi_{n+1}\hat{\chi}_{n+1}^{-1}$ (see below), and by using the left-right equivariance hypothesis, which allows to interpret $\hat{\chi}'_{n+1} Y_{n+1} = h(\chi_{n+1}\hat{\chi}'^{-1}_{n+1}, V_{n+1})$ as a group equivalent to $H(X'_{n+1} - \hat{X}_{n+1}) + V_{n+1}$. Secondly, by viewing the linear case of Section 3.1 as a specific case of the proposed approach through the following proposition whose proof has been moved to the Appendix.

Proposition 2. *There exists an isomorphic representation of $\mathbb{R}^N$ as a matrix Lie group such that* (7) *takes the canonical form* (1)-(2)*, and* (8) *becomes the invariant filter* (10)-(11).

We now define the invariant output errors, which are a transposition of the linear error to our multiplicative group:

$$\eta_n = \chi_n \hat{\chi}_n^{-1}, \qquad \eta'_n = \chi_n \hat{\chi}'^{-1}_n \tag{12}$$

We have the following striking property, that is similar to the linear case:

Theorem 2. *The error variables η_n and η'_n are Markov processes, and are independent of the inputs Ω_n's.*

Proof. The equations followed by η'_n and η_n read:

$$\eta'_{n+1} = \chi_{n+1}\hat{\chi}'^{-1}_{n+1} = \Upsilon_n W_n \chi_n \Omega_n \Omega_n^{-1} \hat{\chi}_n^{-1} \Upsilon_n^{-1} = \Upsilon_n W_n \eta_n \Upsilon_n^{-1} \tag{13}$$

and:

$$\eta_{n+1} = \chi_{n+1}\hat{\chi}'^{-1}_{n+1} K_{n+1}(\hat{\chi}'_{n+1} Y_{n+1})^{-1}$$
$$= \eta'_{n+1} K_{n+1}(\hat{\chi}'_{n+1} h(\chi_{n+1}, V_{n+1}))^{-1} \tag{14}$$
$$\overset{\mathscr{L}}{=} \eta'_{n+1} K_{n+1}(h(\eta'_{n+1}, V_{n+1}))^{-1},$$

thanks to the equivariance property of the output. □

The most important consequence of this property is that if the inputs υ_t are known in advance, or are fixed, as it is the case in Examples 2.2.1 and 2.2.2, the gain functions K_n can be optimized off-line, independently of the trajectory followed by the system. In any case, numerous choices are possible to tune the gains K_n and the remaining sections are all devoted to various types of methods to tackle this problem.

4 Fixed gains filters

In certain cases, one can build an (almost) globally convergent observer for the associated deterministic system, i.e., with noise turned off, by using a family of constant gain function $K_n(\bullet) \equiv K(\bullet)$. If the filter with noise turned off has the desirable property of forgetting its initial condition, convergence to a single point is impossible to retain, because of the unpredictability of the noises that

"steer" the system, but convergence of the distribution can be expected, assuming as in the linear case that Υ_n is independent of n. Indeed in this section we prove that, when noise is turned on, the error forgets its initial distribution under mild conditions. The results are illustrated by an attitude estimation example for which we propose an intrinsic filter having strong convergence properties.

It should be noted that in practice, the convergence to an invariant distribution allows in turn to pick the most desirable gain $K(\bullet)$ among the family based on a performance criterion, such as convergence speed, or filter's precision (that is, ensuring low error's dispersion). This fact will be illustrated by an artificial horizon example of Subsection 4.3.

4.1 Convergence results

Here the left-hand inputs Υ_n are assumed fixed. This, together with constant gains K_n, makes the error sequence η_n a homogeneous Markov chain. Thus, under appropriate technical conditions, the chain has a unique stationary distribution and the sequence converges in distribution to this invariant distribution. Let d denote a right-invariant distance on the group G. We propose the following assumptions:

1. Confinement of the error: there exists a compact set C such that $\forall n \in \mathbb{N}, \eta_n \in C$ a.s. for any $\eta_0 \in C$

2. Diffusivity: the process noise has a continuous part with respect to Haar measure, with density positive and uniformly bounded from zero in a ball of radius $\alpha > 0$ around I_d.

3. Reasonable output noise: $\forall g \in G$, $\mathbb{P}[gK(h(g,V_n))^{-1} \in \mathscr{B}_o(gK(h(g,0))^{-1}, \frac{\alpha}{2})] > \varepsilon'$ for some $\varepsilon' > 0$.

The second assumption implies, and can in fact be replaced with, the more general technical assumption that there exists $\varepsilon > 0$ such that for any subset U of the ball $\mathscr{B}_o(I_d, \alpha)$ we have $P(W_n \subset U) > \varepsilon \mu(U)$ for all $n \geq 0$ where μ denotes the Haar measure. Those noise properties are relatively painless to verify, whereas the confinement property although stronger is automatically verified whenever G is compact, e.g. $G = SO(3)$. Intuitively, the last two assumptions guarantee the error process is well

approximated by its dynamics with noise turned off, followed by a small diffusion. In the theory of Harris chains, the latter diffusion step is a key element to allow probability laws to mix at each step and eventually forget their initial distribution.

Theorem 3. *For constant left-hand inputs* $\Upsilon_n \equiv \Upsilon$, *consider the filter:*

$$\hat{\chi}'_{n+1} = \Upsilon \hat{\chi}_n \Omega_n \tag{15}$$

$$\hat{\chi}_{n+1} = K(\hat{\chi}'_{n+1} Y_{n+1}) \hat{\chi}'_{n+1} \tag{16}$$

Suppose that Assumptions 1)-2)-3) are verified, where the compact set satisfies $C = cl(C^o)$, *cl denoting the closure and* ^{o} *the interior. When noises are turned off, the error equation* (13)-(14) *becomes:*

$$\gamma'_{n+1} = \Upsilon \gamma_n \Upsilon^{-1}, \\ \gamma_{n+1} = \gamma'_{n+1} K_{n+1} (h(\gamma'_{n+1}, 0))^{-1}, \tag{17}$$

Suppose that for any $\gamma_0 \in C$, *except on a set of null Haar measure,* γ_n *converges to* I_d. *Then there exists a unique stationary distribution* π *on* G *such that for any prior law* μ_0 *of the error* η_0 *supported by* C, *the law* $(\mu_n)_{n \geq 0}$ *of* $(\eta_n)_{n \geq 0}$ *satisfies the total variation (T.V.) norm convergence property:*

$$\lim_n \|\mu_n - \pi\|_{T.V.} \to 0$$

Corollary 1. *When the group* G *is compact, The convergence results of Theorem 3 hold globally, i.e. without the confinement assumption 1).*

Theorem 4. *Under the assumptions of Theorem 3, assuming only* $h(\gamma_n, 0) \to 0$ *instead of almost global convergence of* $(\gamma_n)_{n \geq 0}$, *that* G *is compact, the set* $K = \{g \in G, h(g,0) = h(I_d,0)\}$ *connected and* $h(\Upsilon, 0) = h(I_d, 0)$, *the results of Theorem 3 are still valid. Moreover, if* W_n *is isotropic, we have* $\pi(\tilde{\Upsilon}V) = \pi(V)$ *for any* $\tilde{\Upsilon} \in K$ *commuting with* Υ ($\tilde{\Upsilon}\Upsilon = \Upsilon\tilde{\Upsilon}$).

The proofs of the results above have all been moved to the Appendix.

4.2 Application to attitude estimation

Consider the attitude estimation example of Subsection 2.2.1. In a deterministic and continuous time setting, almost globally converging observers have been proposed

in several papers (see e.g. [37, 26]) and have since been analyzed and extended in a number of papers. In order to apply the previously developed theory to this example, the challenge is twofold. First, the deterministic observer must be adapted to the discrete time and proved to be almost globally convergent. Then, the corresponding filter must be proved to satisfy the assumptions of the theorems above in the presence of noise. In discrete time, the system equations read:

$$R_{n+1} = W_n R_n \Omega_n,$$
$$Y_n = (R_n^T b_1 + V_n^1, R_n^T b_2 + V_n^2)$$

with the notations introduced in 2.2.1. We propose the following filter on $SO(3)$:

$$\hat{R}'_{n+1} = \hat{R}_n \Omega_n,$$
$$\hat{R}_{n+1} = K(\hat{R}'_{n+1} Y_{n+1}) \hat{R}'_{n+1},$$
with $K(y_1, y_2) = \exp(k_1(y_1 \times b_1) + k_2(y_2 \times b_2)),$ (18)
$$k_1 > 0, \ k_2 > 0, \ k_1 + k_2 \leqslant 1$$

Proposition 3. *With noise turned off, the discrete invariant observer* (18) *is almost globally convergent, that is, the error converges to I_d for any initial condition except one.*

The proof has been moved to the Appendix, only the main idea is given here. For the continuous-time deterministic problem it is known, and easily seen, that $E : \gamma \to k_1||\gamma^T b_1 - b_1||^2 + k_2||\gamma^T b_2 - b_2||^2$ is a Lyapunov function, allowing to prove almost global convergence of the corresponding observer. In the discrete time deterministic case, the function above remains a Lyapunov function for the sequence $(\gamma_n)_{n \geq 0}$, allowing to derive Proposition 3. This is not trivial to prove, and stems from the more general following novel result:

Proposition 4. *Consider a Lie group G equipped with a left-invariant metric $\langle ., . \rangle$, and a left-invariant deterministic discrete equation on G of the form:*

$$\gamma_{n+1} = \gamma_n \exp(-k(\gamma_n))$$

Assume there exists a C^2 function $E : G \to \mathbb{R}_{\geq 0}$ with bounded sublevel sets, a global minimum at I_d, and a continuous and strictly positive function $u : G \to \mathbb{R}_{>0}$ such that: $\forall x \in G, k(x) = u(x)[x^{-1}.grad_E(x)]$. If the condition

$\forall x \in G, |\frac{\partial k}{\partial x}| \leqslant 1$ *(for the operator norm) is verified, for any initial value γ_0 such that I_d is the only critical point of E in the sublevel set $\{x \in G \mid E(x) \leq E(\gamma_0)\}$ we have:*

$$\gamma_n \underset{n \to \infty}{\to} I_d$$

The proof has been moved to the Appendix. Note that, the latter property is closely related to Lemma 2 of [41]. Using Theorem 3 and Proposition 3 we finally directly get:

Theorem 5. *The distribution of the error variable of the invariant filter* (18) *converges for the T.V. norm to an asymptotic distribution, which does not depend on its initial distribution.*

4.3 Learning robust gains: application to the design of an artificial horizon

As proved in 4.1, under appropriate conditions the error variable is a converging Markov chain whose asymptotic law depends on the gain function but not on the trajectory followed by the system (which is a major difference with most nonlinear filters, such as the EKF). Hence a fixed gain can be asymptotically optimized off-line, leading to a very low numerical cost of the on-line update.

A classical aerospace problem is the design of an artificial horizon using an inertial measurement unit (IMU). An estimation of the vertical is maintained using the observations of the accelerometer (which senses the body acceleration minus the gravity vector, expressed in the body frame) and the stationary flight assumption according to which the body's linear velocity is constant. The problem is that this approximation is not valid in dynamical phases (take-off, landing, atmospheric turbulence), which are precisely when the artificial horizon is most needed. The problem is generally stated as follows:

$$\frac{d}{dt} R_t = R_t(\omega_t + w_t)_\times,$$
$$Y_n = R_{t_n}^T g + V_n + N_n,$$

where R_t is the attitude of the aircraft (the rotation from body-frame coordinates to inertial coordinates), ω_t is the continuous-time gyroscope increment and Y_n is the observation of the accelerometer. The sensor noises w_t and V_n

can be considered as Gaussian, and N_n represents fluctuations due to accelerations of the aircraft that we propose to model as follows: N_n is null with high probability but when non-zero it can take large values. The N_n's are assumed to be independent as usually.

4.3.1 Convergence results

Consider the following class of filters:

$$\hat{R}'_{n+1} = \hat{R}_n \Omega_n,$$
$$\hat{R}_{n+1} = K(\hat{R}'_{n+1}Y_{n+1})\hat{R}'_{n+1}, \qquad (19)$$
$$\text{with } K(y) = \exp(f_{k,\lambda}(y))$$

where $f_{k,\lambda}(x) = k.\min(\text{angle}(x,g),\lambda)\frac{x \times g}{||x \times g||}$ if $x \times g \neq 0$ and $f_{k,\lambda}(x) = 0$ otherwise.

The rationale for the gain tuning is as follows: if the accelerometer measures a value y, we consider the smallest rotation giving to g the same direction as y. Conserving the same axis, the angle of this rotation is thresholded (hence the parameter λ) to give less weight to outliers (without purely rejecting them, otherwise the filter couldn't converge when initialized too far). Then we choose as a gain function a rotation by a fraction of the obtained angle. We begin with the following preliminary result:

Lemma 1. *For any $0 < k \leqslant 1$ and $0 < \lambda \leqslant \pi$ the output error $||Y_n - \hat{R}_n^T g||$ associated to the observer* (19) *with noise turned off converges to 0.*

Proof. Let us consider the error evolution when the noise is turned off. It writes $\gamma_{n+1} = \gamma_n \exp(-f_{k,\lambda}(\gamma_n^{-1}g))$, thus we have $\gamma_{n+1}^{-1}g = \exp(f_{k,\lambda}(\gamma_n^{-1}g))\gamma_n^{-1}g$. As for any $n \in \mathbb{N}$, $f_{k,\lambda}(\gamma_n^{-1}g)$ is orthogonal to g and $\gamma_n^{-1}g$, $\gamma_{n+1}^{-1}g$ stays in the plane spanned by g and $\gamma_0^{-1}g$, as well as the whole sequence $(\gamma_n^{-1}g)_{n \geqslant 0}$. Let $\phi_n = \text{angle}(\gamma_n^{-1}g, g)$. The dynamics of (ϕ_n) writes: $\phi_{n+1} = \phi_n - k.\min(\lambda, \phi_n)$. Thus ϕ_n goes to 0, i.e: $\gamma_n^{-1}g \underset{n \to \infty}{\to} g$, i.e. the observation error goes to 0. $\qquad \square$

The following result is a mere consequence of Lemma 1 and Theorem 4.

Proposition 5. *The error variable associated to the filter defined by* (19) *converges to a stationary distribution for the T.V. norm, which does not depend on its initial distribution.*

4.3.2 Numerical asymptotic gain optimization

To each couple (k, λ) we can associate an asymptotic error dispersion (computed in the Lie Algebra) associated to the corresponding stationary distribution, and try to minimize it. As all computations are to be done offline, the statistics of all distributions can be computed using particle methods. Table 1 gives the parameters of the model used in the following numerical experiment. Figure 1 displays the Root Mean Square Error $RMSE = \sqrt{\mathbb{E}(\eta_\infty g - g)}$, computed over a grid for the parameters (k, λ). The minimum is obtained for $k = 0.1202$ and $\lambda = 0.0029$. If we compare it to a MEKF, we observe a huge difference. For our asymptotic invariant filter we get $RMSE = 8.02 \times 10^{-4}$. For the MEKF, the observation noise matrix giving the best results leads to the value $RMSE = 4.3 \times 10^{-3}$. This result is not surprising due to the fact that the outliers significantly pollute the estimates of the Kalman filter (see Fig. 1). This illustrates the fact that when the noise is highly non-Gaussian, an asymptotic gain with some optimality properties can still be found.

5 Gaussian filters

The present section focuses more on the transitory phase, and the gain is computed at each step based on Gaussian noise approximations and linearizations. Beyond the local optimization underlying the gain computation, the two following filters have the merit to readily convey an approximation of the extent of uncertainty carried by the estimations. We first introduce the discrete-time invariant extended Kalman filter (IEKF), which is an EKF, but based on the linearization of an invariant state error. Then we introduce the IEnKF, which computes the empirical covariance matrix of the error using particles. As all computations can be done off-line, it is easily implementable, as long as the gains can be stored, and is shown to convey a better indication of the error's dispersion for large noises. Simulations illustrate the results.

5.1 The discrete-time Invariant EKF

In the present section we introduce the discrete-time IEKF. Indeed, as in the standard EKF theory, the idea is to linearize the error equation (13)-(14) assuming the noises
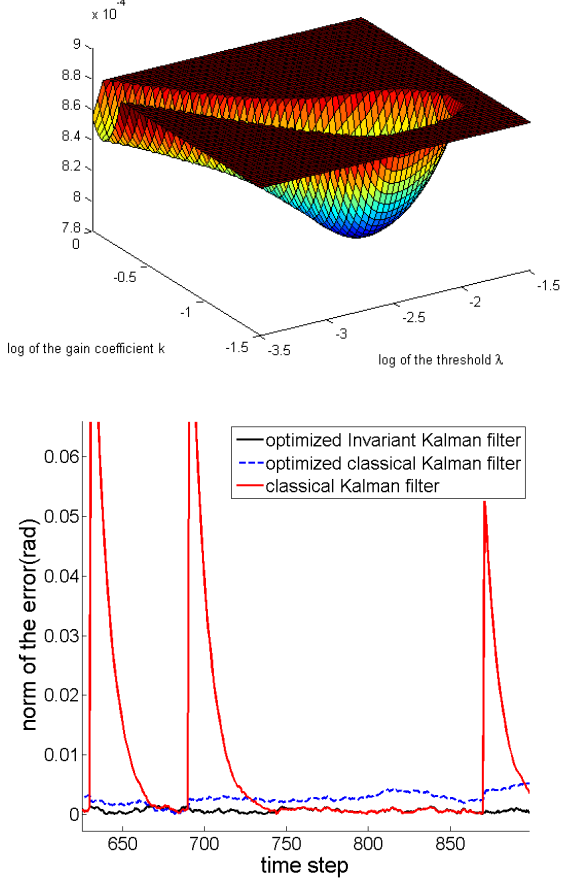
and the state error are small enough, use Kalman equations to tune the gains on this linear system, and implement the gain on the nonlinear model. As in the heuristic theory of the Invariant EKF in continuous time [3] [6] the gains and the error covariance matrix are proved to converge to a fixed value. In particular this allows (after some time) to advantageously replace the gain by its constant final value leading to a numerically very cheap asymptotic version of the IEKF.

Algorithm 1 Invariant Extended Kalman Filter

Returns at each time step $\hat{\chi}_n$, P_n such that $\chi_n \approx \exp(\xi_n)\hat{\chi}_n$, where ξ_n is Gaussian and $var(\xi_n) = P_n$. $\hat{\chi}_0$ and P_0 are inputs.
H_ξ, H_V defined by:

$$h(\exp(\xi'_n), V_n) = h(I_d, 0) + H_\xi \xi'_n + H_V V_n + O(\|\xi'_n\|^2) + O(\|V_n\|^2)$$

Set $n = 0$.
loop
    Compute the value $M_{t_{n+1}}$ solving the equation

$$M_{t_n} = 0, \qquad \frac{d}{dt}M_t = Var(w_t) + ad_{\upsilon_t} M_t ad_{\upsilon_t}^T$$

    The process noise covariance $Q_n^w = Var(Ad_{\Upsilon_n} w_n)$ is equal to $M_{t_{n+1}}$ (if w_n is isotropic or if $\upsilon_t \equiv 0$, merely set $Q_n^w = Var(w_n)(t_{n+1} - t_n)$),
    $Q_{n+1}^V = Var(V_{n+1})$,
    $P_{n+1|n} = Ad_{\Upsilon_n} P_n Ad_{\Upsilon_n}^T + Q_n^w$,
    $S_{n+1} = H_V Q_{n+1}^V H_V^T + H_\xi P_{n+1|n} H_\xi^T$,
    $L_{n+1} = P_{n+1|n} H_\xi^T S_{n+1}^{-1}$,
    $P_{n+1} = (I - L_{n+1} H_\xi) P_{n+1|n}$,
    $\hat{\chi}'_{n+1} = \Upsilon_n \hat{\chi}_n \Omega_n$,
    $\hat{\chi}_{n+1} = \exp(L_{n+1}[\hat{\chi}'_{n+1} Y_{n+1} - h(I_d, 0)])\hat{\chi}'_{n+1}$
    $n \leftarrow n + 1$
end loop

5.1.1 Linearization of the equations and IEKF formulas

We consider here the error equations (13)-(14). Assuming errors and noises are small, we introduce their projection

Figure 1: Artificial horizon example. Top plot: off-line optimization of the parameters of the asymptotic invariant filter (axis $x \to \lambda$; axis $y \to k$; axis $z \to$ Root Mean Square Error). The highest values are not displayed to improve visualization. The performance is optimal for $k = 0.1202, \lambda = 0.0029$. Bottom plot: error evolution of three artificial horizons : the invariant filter with optimized gain functions (black), a MEKF ignoring outliers (red), and a MEKF where the observation noise covariance matrix has been numerically adjusted to minimize the RMSE (dashed line). We see that the MEKF cannot filter the outliers and at the same time be efficient in the absence of outliers, contrarily to the proposed invariant filter.

in the tangent space at the identity I_d (the so-called Lie algebra $\mathfrak{g}$ of the group) using the matrix exponential map $\exp(.)$ (see e.g. [31]). As the matrix vector space $\mathfrak{g}$ can be identified with $\mathbb{R}^{\dim \mathfrak{g}}$ using the linear mapping $(.)_m$ (see Table 2) we can assume $\exp : \mathbb{R}^{\dim \mathfrak{g}} \to G$. This function is defined in Table 2 for usual Lie groups of mechanics. Its inverse function will be denoted by $\exp^{-1}$. We thus define the following quantities of $\mathbb{R}^{\dim \mathfrak{g}}$:

$$\xi_n = \exp^{-1}(\eta_n), \quad \xi'_n = \exp^{-1}(\eta'_n), \quad w_n = \exp^{-1}(W_n)$$

The simplest way to design the gain is to use a function which is linear in $\mathbb{R}^{\dim \mathfrak{g}}$ and then mapped to G through the exponential:

$$K_n : y \to \exp[L_n(y - h(I_d, 0))]$$

Equations (13) and (14) mapped to $\mathbb{R}^{\dim \mathfrak{g}}$ become:

$$\xi'_{n+1} = \exp^{-1}(\exp(Ad_{\Upsilon_n} w_n) \exp(Ad_{\Upsilon_n} \xi_n)),$$
$$\xi_{n+1} = \exp^{-1}(\exp(\xi'_{n+1}) \exp[-L_n(h(\exp(\xi'_{n+1}), V_{n+1}) - h(I_d, 0))]),$$

where we have introduced the adjoint matrix Ad:

Definition 2. *The following property of the adjoint matrix Ad can be used as a definition:*

$$\forall g \in G, \forall u \in \mathbb{R}^{\dim \mathfrak{g}}, \ \exp(Ad_g u) = g \exp(u) g^{-1}$$

The differential of $\xi \to Ad_{\exp(\xi)}$ at zero is denoted $\xi \to ad_\xi$.

To tune the gains L_n through the Kalman theory we need to evaluate at each time step, using $P_n = Var(\xi_n) \in \mathbb{R}^{\dim \mathfrak{g} \times \dim \mathfrak{g}}$, the following quantities:

$$P_{n+1|n} = Var(\xi'_{n+1}), \quad S_{n+1} = Var(Y_{n+1})$$

Our approach consists of using a coarse first-order development of the propagation and observation functions. To do so we use the Baker-Campbell-Hausdorff formula and keep only first order terms in ξ'_{n+1}, w_n and V_{n+1}, the crossed terms being also neglected:

$$\xi'_{n+1} = Ad_{\Upsilon_n} w_n + Ad_{\Upsilon_n} \xi_n$$
$$h(\exp(\xi'_{n+1}), V_{n+1}) = h(I_d, 0) + H_\xi \xi'_{n+1} + H_V V_{n+1}$$
$$\xi_{n+1} = \xi'_{n+1} - L_{n+1}(H_\xi \xi'_{n+1} + H_V V_{n+1})$$

The computation of the gains is then straightforward using the Kalman theory in $\mathbb{R}^{\dim \mathfrak{g}}$, which is a vector space. It is described in Algorithm 1.

5.1.2 Convergence of the gains

The main benefit of the filter is with respect to its convergence properties. Indeed, under very mild conditions, the covariance matrix P_n and the filter's gain K_n are proved to converge to fixed values [20]. The practical consequences are at least twofold: 1-the error covariance converges to a fixed value and is thus much easier to interpret by the user than a matrix whose entries keep on changing (see Fig. 2) 2-due to computational limitations on-board, the covariance may be approximated by its asymptotic value, leading to an asymptotic version of the IEKF being numerically very cheap.

Theorem 6. *If the noise matrices $Q_n^w = Var(w_n)$, $Q_n^V = Var(V_n)$ and the left inputs Υ_n are fixed, the pair (Ad_{Υ_n}, H_ξ) is observable and Q_n^w has full rank, then P_n and K_n converge to fixed values P_∞ and K_∞, as in the linear time-invariant case.*

5.2 The discrete-time Invariant EnKF

When the error equation is independent of the inputs, the exact density of the error variable can be sampled offline. This allows to radically improve the precision of the quantities involved in the Kalman gains computation of Section 5.1.1. We propose here the Invariant Ensemble Kalman filter (IEnKF) described in Algorithm 2. The idea is to compute recursively through Monte-Carlo simulations a sampling of the error density and to use it, instead of linearizations, to evaluate precisely the innovation and error covariance matrices used to compute the gains in the Kalman filter. The procedure is described in Algorithm 2.

5.3 Simulations

We will display here the results of Algorithms 1 and 2 for the attitude estimation problem described in Section 2.2.1:

$$\frac{d}{dt} R_t = R_t(\omega_t + w_t^\omega)_\times,$$
$$Y_n = (R_{t_n}^T b_1 + V_n^1, R_{t_n}^T b_2 + V_n^2),$$

where $R_t \in SO(3)$ represents the rotation that maps the body frame to the earth-fixed frame, and ω_t is the instantaneous angular rotation vector which is assumed to be measured by noisy gyroscopes. $(Y_n)_{n \geq 0}$ is a sequence of

Algorithm 2 IEnKF
***
Define H by $h(\exp(\xi),0) = h(I_d,0) + H\xi + O(\|\xi\|^2)$
Sample M particles $(\eta_0^i)_{1 \leqslant i \leqslant M}$ following the prior error density
for $n = 0$ **to** $N - 1$ **do**
    for $i = 1$ **to** M **do**
        $\eta'^i_{n+1} = W_n^i \Upsilon \eta_n^i \Upsilon^{-1}$,
        $y_{n+1}^i = h(\eta'^i_{n+1}, V_{n+1}^i)$,
    end for
    $P_{n+1|n} = \frac{1}{M} \sum_{i=1}^M \exp^{-1}(\eta'^i_{n+1}) \exp^{-1}(\eta'^i_{n+1})^T$
    $S_{n+1} = \frac{1}{M} \sum_{i=1}^M y_{n+1}^i y_{n+1}^i{}^T$,
    $L_{n+1} = P_{n+1|n} H^T S_{n+1}^{-1}$
    for $i = 1$ **to** M **do**
        $\eta_{n+1}^i = \eta'^i_{n+1} \exp(-L_{n+1}[y_{n+1}^i - h(I_d,0)])$
    end for
    Store the gain matrix L_{n+1}
end for
The prior estimation $\hat{\chi}_0$ is an input.
for $n = 0$ **to** N-1 **do**
    $\hat{\chi}'_{n+1} = \Upsilon \hat{\chi}_n \Omega_n$,
    $\hat{\chi}_{n+1} = \exp(L_{n+1}[\hat{\chi}'_{n+1} Y_{n+1} - h(I_d,0)]) \hat{\chi}'_{n+1}$
end for
***

discrete noisy measurements of the vectors b_1 and b_2 (for instance the gravity and the magnetic field), V_n^1 and V_n^2 are sequences of independent isotropic white noises. A simulation has been performed using the parameters given in Table 3. In this case the IEKF equations are described by Algorithm 3. As the state-of-the-art methods for this particular problem are the Multiplicative Extended Kalman Filter (MEKF, see e.g. [12]) and the USQUE filter [11] (a quaternion implementation of the Unscented Kalman Filter), the four methods are compared. The evolution of the gains is displayed on Fig. 2 and shows the interest of the invariant approach. The error variable is expressed in the Lie algebra and its projection to the first axis is given on Fig. 3 for each method (the projections on other axes being very similar, they are not displayed due to space limitations). All perform satisfactorily, but some differences are worthy of note. First we can see that the MEKF and the IEKF yield comparable performances but only the gains of the IEKF converge. Moreover, the linearizations lead the MEKF and the IEKF to fail capturing accurately the error dispersion through a 3σ-envelope. To this re-

spect, the USQUE performs better than the two latter filters but still does not succeed in capturing the uncertainty very accurately. On the other hand, the envelope provided by the EnKF is very satisfying. This result is not surprising: this envelope has been computed using a sampling of the true density of the error, there is thus no reason why it should not be a valid approximation as long as there are sufficiently many particles. This method is thus to be preferred to the other ones when the user is willing and able to perform extensive simulations, and has the capacity to store the gains over the whole trajectory.

***
Algorithm 3 Invariant Extended Kalman Filter on $SO(3)$
***
Returns at each time step $\hat{R}_n$, P_n such that $R_n \approx \exp(\xi_n)\hat{R}_n$, where ξ_n is a centered Gaussian and $var(\xi_n) = P_n$.
$\hat{R}_0$ and P_0 are inputs.
$H_\xi = \begin{pmatrix} (b_1)_\times \\ (b_2)_\times \end{pmatrix}$, $H_V = I_6$,
$Q_V = \begin{pmatrix} Var(V^1) & 0 \\ 0 & Var(V^2) \end{pmatrix}$, $\qquad Q_w = \int_{t_n}^{t_{n+1}} Var(w_s^\omega) ds$.
Set $n = 0$.
loop
    Let Ω_n be the solution at t_{n+1} of $T_{t_n} = I_3$, $\frac{d}{dt}T_t = (\omega_t)_\times$.
    $P_{n+1|n} = P_n + Q_w$,
    $S_{n+1} = H_V Q_V H_V^T + H_\xi P_{n+1|n} H_\xi^T$,
    $L_{n+1} = P_{n+1|n} H_\xi^T S_{n+1}^{-1}$,
    $P_{n+1} = (I - L_{n+1} H_\xi) P_{n+1|n}$,
    $\hat{R}'_{n+1} = \hat{R}_n \Omega_n$
    $\Delta_\chi = L_{n+1}[\begin{pmatrix} \hat{R}'_{n+1} Y_{n+1}^1 \\ \hat{R}'_{n+1} Y_{n+1}^2 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}]$,
    $\hat{R}_{n+1} = \exp(\Delta_\chi)\hat{R}'_{n+1}$ (see Table 2)
    $n \leftarrow n + 1$
end loop
***

6 Conclusion

This paper has introduced a proper stochastic framework for a class of filtering problems on Lie groups. A wide class of discrete-time filters with strong properties has
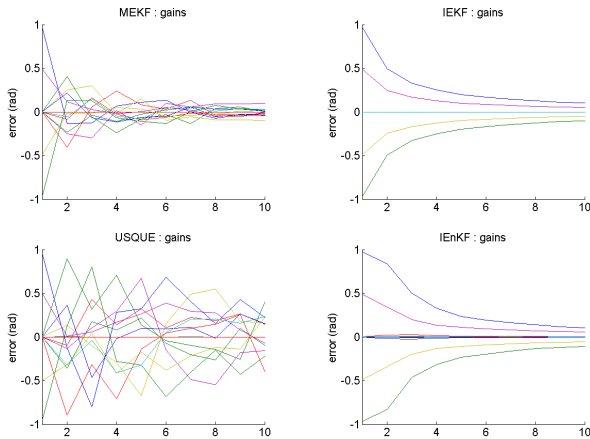
Figure 2: Evolution over time of the coefficients of the gain matrix for MEKF, IEKF, USQUE and IEnKF. The coefficients of the MEKF and USQUE have an erratic evolution, whereas those of the IEKF are convergent, allowing to save much computation power using asymptotic values. Moreover, in this case, only 4 of the 18 coefficients are not equal to zero for the IEKF, allowing sparse implementation.
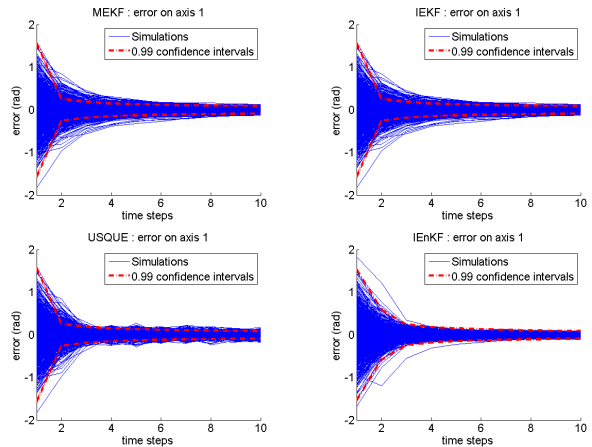
Figure 3: Evolution of the error variable, projected on the first axis, for 1000 simulations using MEKF, IEKF, USQUE and IEnKF. The MEKF and the IEKF give similar results. The USQUE is doing slightly better but is outperformed by the IEnKF. Only the latter captures the error dispersion satisfactorily, as the 3σ envelope contains 99 % of the simulated trajectories.

been proposed, along with several methods to tune the gains. When the gain is held fixed, the estimation error has the remarkable property to provably converge in distribution. Another route is to focus more on the transitory phase, and to seek to minimize the dispersion of the estimation error at each step based on linearizations and Gaussian noise approximations. This has lead to an invariant version of the celebrated EKF, having the property that the error covariance matrix converges asymptotically, as well as to an invariant version of the EnKF which has been shown to convey an accurate indication of the level of uncertainty carried by the estimate.

The results of the paper open the door to a rich variety of filters on Lie groups. Indeed, all the classical gain tuning approaches can still be considered, and a method based on the unscented transform for instance, completing the work [11] seems an interesting idea. In a situation where many observations are available at each time step, an information filter could also be derived mimicking the IEKF. Besides, we believe that the cases where the gains of the IEKF are sparse should be given special attention,

by studying when this property occurs and if significant computation power and memory can be saved. The possibilities offered by off-line simulations deserve also to be deepened. The computation of exact confidence sets can be explored, establishing a link with the approach of set-valued observers proposed in [8] and [38]. Concerning the asymptotic method introduced here, the optimization has been performed quite naively as efficient density simulation is beyond the scope of the paper, but there may certainly be some room for improvements using Markov Chain Monte-Carlo (MCMC) schemes to speed up convergence. In the future we will concentrate on proving the convergence in distribution of the IEKF, that can be conjectured at this stage, but does not seem easy.

A   Proof of Proposition 2

For any $i \in \mathbb{N}$, let $\mathcal{M}_i(\mathbb{R})$ denote the set of square matrices of size i. Consider the isomorphisms $\phi_X : X \in \mathbb{R}^N \to \begin{pmatrix} Id_N & 0_{N,1} \\ X^T & 1 \end{pmatrix} \in \mathcal{M}_{N+1}(\mathbb{R})$ and $\phi_Y : Y \in \mathbb{R}^p \to \begin{pmatrix} H^T \\ Y^T \end{pmatrix} \in$

$\mathscr{M}_{N+1,p}$. Letting $\chi_n = \phi_X(X_n)$, $Y_n^* = \phi_Y(Y_n)$, $V_n^* = \phi_Y(V_n)$, $W_n^* = \phi_X(W_n)$, $\Omega_n = \phi_X(B_n)$ and $\Upsilon_n = Id_{N+1}$ equations (7) become $\chi_{n+1} = \Upsilon_n W_n^* \chi_n \Omega_n$ and $Y_n^* = \chi_n^{-1} V_n^*$. Using the inverse isomorphisms ϕ_X^{-1} and ϕ_Y^{-1}, the invariant update $\hat{\chi}_{n+1} = K_n(\hat{\chi}'_{n+1} Y^*_{n+1}) \hat{\chi}_{n+1}$ corresponds in $\mathbb{R}^n$ to (7).

B Proofs of the results of Section 4.1

We first introduce here the basic notions about Harris Chains that we will need, to prove the stochastic convergence properties of the invariant filters.

Definition 3. *(First hitting time) Let S be a measurable space and $(X_n)_{n \geqslant 0}$ a Markov chain taking values in S. The first hitting time τ_C of a subset $C \subset S$ is defined by $\tau_C = \inf_{n \geqslant 0} \{X_n \in C\}$.*

Definition 4. *(Recurrent aperiodic Harris chain) Let S be a measurable space and $(X_n)_{n \geqslant 0}$ a Markov chain taking values in S. $(X_n)_{n \geqslant 0}$ is said to be a recurrent aperiodic Harris chain if there exist two sets $A, B \subset S$ satisfying the following properties:*

 i *For any initial state X_0 the first hitting time of A is a.s. finite.*

 ii *There exists a probability measure ρ on B, and $\varepsilon > 0$ such that if $x \in A$ and $D \subset B$ then $\mathbb{P}(X_1 \in D | X_0 = x) > \varepsilon \rho(D)$.*

 iii *There exists an integer $N \geqslant 0$ such that: $\forall x \in A, \forall n \geqslant N, \mathbb{P}(X_n \in A | X_0 = x) > 0$.*

The somewhat technical property ii means that any given area of B can be reached from each point of A with non-vanishing probability.

Theorem 7. *[Harris, 1956] A recurrent aperiodic Harris chain admits a unique stationary distribution ρ_∞ and the density of the state X_n converges to ρ_∞ in T.V. norm for any distribution of X_0.*

Proof. See [16], Theorems 6.5 and 6.8 □

In other words, if for any initialization we can ensure that the process will come back with probability 1 to a central area A where mixing occurs, we have a convergence property. The following technical result is a cornerstone to demonstrate the theorems of Subsection 4.1.

Lemma 2. *Let G be a locally compact Lie group provided with a right-invariant distance d, and $C \subset G$ be a measurable compact set endowed with the σ-algebra Σ. Consider a homogeneous Markov chain $(\eta_n)_{n \geqslant 0}$ defined by the relation*

$$\eta_0 \in C, \qquad \eta_{n+1} = q_z(\eta_n)$$

where $z \in \mathscr{Z}$ is some random variable belonging to a measurable space. Let 0 denote a specific point of $\mathscr{Z}$. Let $Q : (C, \Sigma) \to \mathbb{R}_+$, denote the transition kernel of the chain, that is $Q(x, V) = \mathbb{P}(q_z(\eta_n) \in V \mid \eta_n = x)$. We assume that

(a) *there exist real numbers $\alpha, \varepsilon > 0$ such that for any $x \in C$ and $V \subset \mathscr{B}_o(q_0(x), \alpha)$, the latter denoting the open ball center $q_0(x)$ and radius α, we have $Q(x, V) = \mathbb{P}(q_z(\eta_n) \in V \mid \eta_n = x) \geqslant \varepsilon \mu_G(V)$ where μ_G is the right-invariant Haar measure.*

(b) *q_0 admits a fixed point $x_0 \in C$, i.e. $q_0(x_0) = x_0$.*

Let $U_0 = \{x_0\}$. Define the sets $(U_n)_{n \geqslant 0}$ recursively by $U'_n = \{x \in G, d(x, U_n) < \frac{\alpha}{2}\} \cap C$ and $U_{n+1} = q_0^{-1}(U'_n)$. If there exists an integer $N \geqslant 0$ such that $U_N = C$, then the p.d.f. of η_n converges for the T.V. norm and its limit does not depend on the initialization.

More prosaically, z denotes the cumulated effect of the model and observation noise, and it is assumed that 1- the noiseless algorithm has some global convergence properties and 2- there are sufficiently many small enough noises so that there is an ε chance for η_n to jump from x to any neighborhood of the noiseless iterate $q_0(x)$ with a multiplicative factor corresponding to the neighborhood's size. It then defines a sequence of sets, by picking a fixed point of q_0, dilating it, and taking its pre-image. Dilatation and inversion are then reiterated until the whole set C is covered. In this case, forgetting of the initial distribution is ensured.

Proof. We will demonstrate the property trough three intermediate results:

1. There exists a sequence $\varepsilon_1, \cdots, \varepsilon_N$ such that: $\mathbb{P}(\eta_1 \in U_n | \eta_0 \in U_{n+1}) > \varepsilon_n > 0$.

2. The first hitting time of U_1 is a.s. finite.

3. $(\eta_n)_{n \geq 0}$ is an aperiodic recurrent Harris chain with $A = U_1$ and $B = U'_0$.

The conclusion will then immediately follow from Theorem 7. We give first a qualitative explanation of the approach. We have by assumption a sequence of sets $(U_n)_{n\geqslant 0}$. Intermediate result 1) states that starting from a set U_{n+1} the chain has a non-vanishing chance to jump to the "smaller" set U_n at the next time step. Intermediate result 2) states that starting from U_{n+1}, because of the non-vanishing property 1), the chain cannot avoid U_n forever and eventually reaches U_n, U_{n-1} and finally U_1. Intermediate result 3) brings out the fact that, $A = U_0'$ being totally included in the ball of radius α centered on any of its points, the chain can go from any given point of its pre-image $B = U_1$ to any area of A with controlled probability, which is the last property needed to obtain a recurrent aperiodic Harris chain.

1. Consider the closure $\overline{U_n'}$ of U_n'. For any $x \in \overline{U_n'}$ the set $\mathscr{B}_o(x,\alpha) \cap U_n$ is open and non-empty (as $d(x,U_n) \leqslant \frac{\alpha}{2}$ by definition of $\overline{U_n'}$) so $\mu_G(\mathscr{B}_o(x,\alpha) \cap U_n) > 0$. The function $f: x \to \mu_G(\mathscr{B}_o(x,\alpha) \cap U_n) > 0$ is continuous (since μ_G is a regular and positive measure) on the compact set $\overline{U_n'}$, so it admits a minimum $m_n > 0$. We get: $P(\eta_1 \in U_n | \eta_0 \in U_{n+1}) \geqslant P(\eta_1 \in \mathscr{B}_o(q_0(\eta_0),\alpha) \cap U_n | \eta_0 \in U_{n+1}) = P(\eta_1 \in \mathscr{B}_o(q_0(\eta_0),\alpha) \cap U_n | q_0(\eta_0) \in U_n') \geqslant \varepsilon m_n$ using the fact that, by assumption (a), $Q(x,V) \geqslant \varepsilon \mu_G(V)$ for $x \in C$ and $V \subset \mathscr{B}_o(q_0(x),\alpha)$. We set $\varepsilon_n = \varepsilon m_n$ to prove intermediate result 1).

2. We will prove by descending induction on n the property $\mathscr{P}_n : P(\tau_n < \infty) = 1$. $\mathscr{P}_N$ is obvious as $U_N = C$. Now assume $\mathscr{P}_{n+1}$ is true for $n \geqslant 1$. Due to homogeneity, we can construct a strictly increasing sequence of stopping times $(v_p)_{p\geqslant 0}$ such that: $\forall p \geqslant 0, \eta_{v_p} \in U_{n+1}$. But $\mathbb{P}(\eta_{v_p+1} \in U_n | \eta_{v_p} \in U_{n+1}) > \varepsilon_n$ by intermediate result 1). Thus letting T_k denote the event $\{\forall i \leqslant k, \eta_i \notin U_n\} = \{k < \tau_n\}$, we have for any $n \geqslant 0$: $\mathbb{P}(T_{v_p+1}) = \mathbb{P}(T_{v_p})\mathbb{P}(T_{v_p+1}|T_{v_p}) \leqslant \mathbb{P}(T_{v_p})(1 - \varepsilon_n)$. Thus $\mathbb{P}(T_{v_{p+1}}) \leqslant \mathbb{P}(T_{v_p+1}) \leqslant \mathbb{P}(T_{v_p})(1 - \varepsilon_n)$. A quick induction and a standard application of Borel-Cantelli Lemma prove there exists a.s. a rank p such that T_{v_p} is false. In other words, $\mathscr{P}_n$ is true. In particular, $\mathscr{P}_1$ is true: the first hitting time of U_1 is a.s. finite.

3. Define $A = U_1$ and $B = U_0' = \mathscr{B}_o(x_0,\frac{\alpha}{2}) \cap C$. For any $V \subset B$ and $x \subset A$ we have $V \subset \mathscr{B}_o(q_0(x),\alpha)$, thus

$Q(x,V) \geqslant \varepsilon \mu_G(V)$ by assumption (a). Thus $(\eta_n)_{n\geqslant 0}$ verifies property ii) of Definition 4. As $A = U_1$ the intermediate result 2) shows that $(\eta_n)_{n\geqslant 0}$ verifies the property i) of Definition 4. As there exists a.s. a rank k such that $\eta_k \in U_1$ by intermediate result 2), there exists an integer M such that $\mathbb{P}(\eta_M \in U_1) > 0$. As $U_1 \cap U_0' \subset \mathscr{B}_o(q(x),\alpha)$ for any $x \in U_1$ (by definition of U_0' and U_1) Assumption (a) gives immediately $\mathbb{P}(\eta_{M+1} \in U_1 \cap U_0') > \mathbb{P}(\eta_M \in U_1)\varepsilon \mu_G(U_1 \cap U_0')$. Using only the definition of U_0' and U_1, and Assumption (b), we see easily that both contain x_0. Thus the set $U_0' \cap U_1$ is non-empty and open. We have then $\mu_G(U_1 \cap U_0') > 0$ and $\mathbb{P}(\eta_{M+1} \in U_1 \cap U_0') > 0$. By an obvious induction on m, $\mathbb{P}(\eta_m \in U_1 \cap U_0') > 0$ for any $m > M$. As $U_1 \cap U_0' \subset A$ we get $\mathbb{P}(\eta_m \in A) > 0$ for any $m > M$ and the property iii) of Definition 4 is also verified. We finally obtain that $(\eta_n)_{n\geqslant 0}$ is an aperiodic recurrent Harris chain.

Using Theorem 7 we can conclude that the density of η_n converges in T.V. norm to its unique equilibrium distribution for any initialization. $\square$

Building upon the previous results, the proof of Theorem 3 is as follows. The Markov chain defined by equations (13) and (14) with fixed Υ can be written under the form $\eta_{n+1}' = q_z(\eta_n') = q_{1,z} \circ q_{2,z}(\eta_n')$ with $z = (W_n, V_n) \in G \times \mathbb{R}^p$, $q_{1,z}(x) = \Upsilon W_n x \Upsilon^{-1}$ and $q_{2,z}(x) = xK(h(x,V_n))^{-1}$. Let $0 = (I_d, 0)$. We will show that $(\eta_n')_{n\geqslant 0}$ has all the properties required in Lemma 2.

Let $\alpha, \varepsilon, \varepsilon'$ be defined as in 4.1. Let α' such as $\mathscr{B}_o(I_d,\alpha') \subset \Upsilon \mathscr{B}_o(I_d,\frac{\alpha}{2})\Upsilon^{-1}$ and $\varepsilon'' = \varepsilon'|Ad_{\Upsilon^{-1}}|\varepsilon$, where $|Ad_{\Upsilon^{-1}}|$ is the determinant of the adjoint operator on $\mathfrak{g}$. For any $x \in C$ and $V \subset \mathscr{B}_o(q_0(x),\alpha') = \mathscr{B}_o(I_d,\alpha')q_0(x)$ we have:

$$P(q_z(x) \in V) \geqslant P(q_z(x) \in V, q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x),\frac{\alpha}{2}))$$

$$\geqslant P(q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x),\frac{\alpha}{2})) \times$$

$$P(\Upsilon W_n q_{2,z}(x)\Upsilon^{-1} \in V | q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x),\frac{\alpha}{2}))$$

i.e. $P(q_z(x) \in V) \geqslant \varepsilon' P(W_n \in \Upsilon^{-1} V \Upsilon q_{2,z}(x)^{-1} | q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x),\frac{\alpha}{2}))$. Assuming $q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x),\frac{\alpha}{2})$ we have: $\Upsilon^{-1} V \Upsilon q_{2,z}(x)^{-1} \subset \Upsilon^{-1}\mathscr{B}_o(I_d,\alpha')q_0(x)\Upsilon q_{2,z}(x)^{-1} \subset \mathscr{B}_o(I_d,\frac{\alpha}{2})\Upsilon^{-1}q_0(x)\Upsilon q_{2,z}(x)^{-1} \subset$

$\mathscr{B}_o(I_d, \frac{\alpha}{2})q_{2,0}(x)q_{2,z}(x)^{-1} \subset \mathscr{B}_o(I_d, \frac{\alpha}{2} + \frac{\alpha}{2})$. As W_n is independent from the other variables we obtain: $P(W_n \in \Upsilon^{-1}V\Upsilon q_{2,z}(x)^{-1}|q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x), \frac{\alpha}{2})) \geqslant \varepsilon\mu_G(\Upsilon^{-1}V\Upsilon q_{2,z}(x)^{-1}) = \varepsilon\mu_G(\Upsilon^{-1}V\Upsilon) = \varepsilon\mu_G(V)|Ad_{\Upsilon^{-1}}|$
And finally:

$$P(q_z(x) \in V) \geqslant \varepsilon'|Ad_{\Upsilon^{-1}}|\varepsilon\mu(V) = \varepsilon''\mu(V)$$

As q_0 has I_d as a fix point we only have to verify that the sets U_n as defined in Lemma 2 eventually cover the whole set C. It suffices to consider for any $n \geqslant 0$ the set $D_n = \{x \in C, \forall k \geqslant n, q_0^k(x) \in U_0'\}$. As we have $q_0^n(x) \to I_d$ almost-everywhere on C we get: $\mu_G(\cup_{n\geqslant 0}D_n) = \mu_G(C)$. As the sequence of sets $(D_n)_{n\geqslant 0}$ increases we have: $\mu_G(D_n) \xrightarrow[n\to\infty]{} \mu_G(C)$. We introduce here the quantity $v_{\min} = \min_{x\in C}\mu_G(\mathscr{B}_o(x, \frac{\alpha'}{2})\cap C)$ (the property $C = cl(C^o)$ and the regularity of μ_G ensure that we have $v_{\min} > 0$). Let $N \in \mathbb{N}$ be such that $\mu_G(D_N) > \mu_G(C) - v_{\min}$. We have then: $\forall y \in C, d(y, D_N) < \frac{\alpha'}{2}$ (otherwise we would have $\mu_G(D_N) \leqslant \mu_G(C) - v_{\min}$). As we have $D_N \subset U_N$ we obtain $\forall y \in C, d(y, D_N) < \frac{\alpha'}{2}$ thus $U_N' = C$ and $U_{N+1}' = q_0^{-1}(U_N') = q_0^{-1}(C) = C$. So all the conditions of Lemma 2 hold and we can conclude convergence in T.V. norm to a stationary distribution which doesn't depend on the prior, provided that its support lies in C. Theorem 3 is proved.

Corollary 1 follows directly from the case $C = G$.

Theorem 4 can be proved as follows. Let $C = G$. Let $K = \{x \in G, h(x,0) = h(I_d, 0)\}$. Note that the left-right equivariance assumption ensures that K is a subgroup of G. First we show that there exists an integer N_1 such that $K \subset U_{N_1}$. As we have $\forall x \in K\ q_0(x) = \Upsilon x \Upsilon^{-1}$ the sequence of sets $Q_n = \Upsilon^{-n}U_n\Upsilon^n \cap K$ is growing and we have: $\{x \in K,\ d(x, Q_n) < \frac{\alpha}{2}\} \subset Q_{n+1}$. The set $Q_\infty = \cup_{n=1}^\infty Q_n$ is open in K as an union of open sets, but we have $\forall x \in Q_\infty, \mathscr{B}_o(x, \frac{\alpha}{2})\cap K \subset Q_\infty$ thus $\forall x \in K \setminus Q_\infty, \mathscr{B}_o(x, \frac{\alpha}{4})\cap K \subset K \setminus Q_\infty$. This implies that Q_∞ and $K \setminus Q_\infty$ are both open in K. As K is connected (see assumptions of Theorem 4) we obtain $Q_\infty = K$. As K is compact (as a closed subset of a compact) the open cover $\cup_{i=1}^\infty Q_n$ has a finite subcover and there exists an integer N_1 such that $Q_{N_1} = K$, i.e. $\Upsilon^{N_1}K\Upsilon^{-N_1} = U_{N_1}$. As K is a subgroup of G (see above) containing Υ we obtain $K \subset U_{N_1}$. Now we have to prove that the sets $(U_n)_{n\geqslant 0}$ eventually cover the whole set C. For any $x \in G$ we have $h(q_0^n(x), 0) \to h(I_d, 0)$. Thus

there exists a rank n such that $\forall k \geqslant n, q_0^k(x) \in U_N'$ (otherwise we could extract a subsequence from $(q_0^n(x))_{n\geqslant 0}$ which stays at a distance $> \frac{\alpha}{2}$ from K, as G is compact we could extract again a convergent subsequence and its limit $q_0^\infty(x)$ would be outside K, thus we would have $h(q_0^n(x)) \not\to h(I_d, 0)$). Here we can define as in the proof of Theorem 3 the sets $D_n = \{x \in C, \forall k \geqslant n, q_0^k(x) \in U_{N_1}'\}$ (note that U_1' has been replaced by U_{N_1}' in this definition of D_n). As for almost any $x \in G$ there exists a rank n such that $x \in D_n$ we have $\mu_G(D_n) \to \mu_G(C)$ and as in the proof of Theorem 3 the sets U_n eventually cover the set C.

The second part of the property (invariance to left multiplication) is easier. Consider the error process $\zeta_n = \tilde{\Upsilon}\eta_n$. The propagation and update steps read:

$$\zeta_{n+1}' = \tilde{\Upsilon}\eta_{n+1}' = \tilde{\Upsilon}\Upsilon W_n\eta_n\Upsilon^{-1} = \Upsilon\tilde{\Upsilon}W_n\eta_n\Upsilon^{-1} \overset{\mathscr{L}}{=} \Upsilon W_n\tilde{\Upsilon}\eta_n\Upsilon^{-1}$$
$$\overset{\mathscr{L}}{=} \Upsilon W_n\zeta\Upsilon^{-1}$$

$$\zeta_{n+1} = \tilde{\Upsilon}\eta_{n+1}' K_{n+1}(h(\eta_{n+1}', V_{n+1}))^{-1}$$
$$\overset{\mathscr{L}}{=} \tilde{\Upsilon}\eta_{n+1}' K_{n+1}(h(\tilde{\Upsilon}\eta_{n+1}', V_{n+1}))^{-1}$$
$$\overset{\mathscr{L}}{=} \zeta_{n+1}' K_{n+1}(h(\zeta', V_{n+1}))^{-1}$$

where we have used the property: $h(\eta_{n+1}', V_{n+1}) \overset{\mathscr{L}}{=} \eta_{n+1}'^{-1}h(I_d, V_{n+1}) \overset{\mathscr{L}}{=} \eta_{n+1}'^{-1}h(\tilde{\Upsilon}, V_{n+1}) \overset{\mathscr{L}}{=} h(\tilde{\Upsilon}\eta_{n+1}', V_{n+1})$. We see that the law of the error process is invariant under left multiplication by $\tilde{\Upsilon}$, thus the asymptotic distribution π inherits this property.

C Proofs of the results of Section 4.2

C.1 Proof of Proposition 4

We define the continuous process $\tilde{\gamma}_t : \mathbb{R} \to G$ as follows:

$$\forall n \in \mathbb{N}, \tilde{\gamma}_n = \gamma_n$$

$$\forall n \in \mathbb{N}, \forall t \in [n, n+1[, \frac{d}{dt}\tilde{\gamma}_t = -\tilde{\gamma}_t k(\tilde{\gamma}_n)$$

We obtain immediately that $\tilde{\gamma}_t$ is continuous. Besides, for any $n \in \mathbb{N}$ and $t \in]n, n+1[$ one has:

$$|\frac{d}{dt}\langle k(\tilde{\gamma}_t), k(\tilde{\gamma}_n)\rangle| = |\langle \frac{d}{dt}k(\tilde{\gamma}_t), k(\tilde{\gamma}_n)\rangle|$$

$$\leqslant |\frac{d}{dt}k(\tilde{\gamma}_t)||k(\tilde{\gamma}_n)|$$

$$\leqslant |\frac{d}{dt}\tilde{\gamma}_t||k(\tilde{\gamma}_n)| \text{ (due to } |\frac{\partial k}{\partial x}| \leqslant 1 \text{)}$$

$$\leqslant |k(\tilde{\gamma}_n)|^2 \text{ (as } \frac{d}{dt}\tilde{\gamma}_t = -\tilde{\gamma}_t k(\tilde{\gamma}_n) \text{)}$$

Thus:

$$\langle k(\tilde{\gamma}_t), k(\tilde{\gamma}_n)\rangle \geqslant \langle k(\tilde{\gamma}_n), k(\tilde{\gamma}_n)\rangle - (t-n)|k(\tilde{\gamma}_n)|^2 \geqslant 0$$

proving in the Lie algebra that $\langle k(\tilde{\gamma}_t), k(\tilde{\gamma}_n)\rangle \geqslant 0$. As u is a positive function we have thus $u(\tilde{\gamma}_t)^{-1}\langle \tilde{\gamma}_t k(\tilde{\gamma}_t), -\tilde{\gamma}_t k(\tilde{\gamma}_n)\rangle \leqslant 0$ immediately proving $\langle grad_E(\tilde{\gamma}_t), \frac{d}{dt}\tilde{\gamma}_t\rangle \leqslant 0$. The latter result being true for every $t \in \mathbb{R}\backslash\mathbb{N}$ and the function $E(\tilde{\gamma}_t)$ being continuous, it decreases and thus converges on $\mathbb{R}_{\geqslant 0}$. As we have supposed the sublevel sets of E are bounded (and closed as E is continuous), $\tilde{\gamma}_t$ is stuck in a compact set. Let γ_∞ an adherence value of $\tilde{\gamma}_n$. By continuity of E we have $E(\tilde{\gamma}_\infty) = E(\tilde{\gamma}_\infty \exp(-k(\tilde{\gamma}_\infty)))$. Thus the function $t \to E(\tilde{\gamma}_\infty \exp(-t k(\tilde{\gamma}_\infty)))$ is decreasing on $[0, 1]$ and has the same value at 0 and 1. Thus it is constant, its derivative at 0 is null proving $\gamma_\infty = I_d$. $(\tilde{\gamma}_n)_{n \geqslant 0}$ being confined in a compact set and having I_d as unique adherence value we finally get $\gamma_n = \tilde{\gamma}_n \to I_d$.

C.2 Proof of Proposition 3

Let $\gamma_n = R_n \hat{R}_n^T$ be the invariant error. Its equation reduces to:

$$\gamma_{n+1} = \gamma_n.\exp(-k_1(\gamma_n^T b_1) \times b_1 - k_2(\gamma_n^T b_2) \times b_2)$$

Let $k(\gamma) = k_1(\gamma^T b_1) \times b_1 + k_2(\gamma^T b_2) \times b_2$ and $E : \gamma \to k_1||\gamma^T b_1 - b_1||^2 + k_2||\gamma^T b_2 - b_2||^2$. To apply Proposition 4 we will first verify that 1) $\forall \gamma \in SO(3), \gamma.k(\gamma) = grad_E(\gamma)$ 2) $|\frac{\partial k}{\partial \gamma}| \leqslant 1$. To prove 1) consider the dynamics in $\in SO(3)$ defined by $\frac{d}{dt}\gamma_t = \gamma_t(\psi_t)_\times$ for some rotation vector path ψ_t. We have $\frac{d}{dt}E(\gamma_t) = k_1(\gamma_t^T b_1 - b_1)^T \frac{d}{dt}\gamma_t^T b_1 + k_2(\gamma_t^T b_2 - b_2)^T \frac{d}{dt}\gamma_t^T b_2$. Using triple product equalities this is equal to $\langle k_1(\gamma_t^T b_1)_\times b_1 +$

$k_2(\gamma_t^T b_2)_\times b_2, \psi_t\rangle = \langle k(\gamma_t), \psi_t\rangle = \langle \gamma_t.k(\gamma_t), \frac{d}{dt}\gamma_t\rangle$. Thus $\gamma.k(\gamma) = grad_E(\gamma)$. To prove 2) we analogously see that $\frac{d}{dt}k(\gamma_t) = -[k_1(b_1)_\times(\gamma_t^T b_1)_\times + k_2(b_2)_\times(\gamma_t^T b_2)_\times]\psi_t$. Thus $|\frac{d}{dt}k(\gamma_t)| \leqslant (k_1 + k_2)|\psi_t|$ and finally $|\frac{\partial k}{\partial \gamma}| \leqslant 1$. Now, except if initially γ_0 is the rotation of axis $b_1 \times b_2$ and angle π, the function E is strictly decreasing, and I_d is the only point in the sublevel set $\{\gamma \in SO(3) \mid E(\gamma) \leq E(\gamma_0)\}$ such that $grad_E(\gamma) = 0$. Applying Proposition 4 allows to prove Proposition 3.

References

[1] S. Bonnabel A. Barrau. Intrinsic filtering on SO(3) with discrete-time observations. In *IEEE Conference on Decision and Control*, 2013.

[2] M. Barczyk and A. F. Lynch. Invariant observer design for a helicopter uav aided inertial navigation system. *Control Systems Technology, IEEE Transactions on*, 21(3):791–806, 2013.

[3] S. Bonnabel. Left-invariant extended Kalman filter and attitude estimation. In *Decision and Control, 2007 46th IEEE Conference on*, pages 1027–1032. IEEE, 2007.

[4] S. Bonnabel, Ph. Martin, and P. Rouchon. Symmetry-preserving observers. *Automatic Control, IEEE Transactions on*, 53(11):2514–2526, 2008.

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

[6] 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.

[7] G. Bourmaud, R. Mégret, A. Giremus, Y. Berthoumieu, et al. Discrete extended Kalman filter on Lie groups. In *21st European Signal*

Processing Conference 2013 (EUSIPCO 2013), 2013.

[8] S. Brás, P. Rosa, C. Silvestre, and P. Oliveira. Global attitude and gyro bias estimation based on set-valued observers. *Systems & Control Letters*, 62(10):937–942, 2013.

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

[10] D.L. Burkholder, É. Pardoux, A.S. Sznitman, and P.L. Hennequin. *École d'été de probabilités de Saint-Flour XIX, 1989*. Springer-Verlag, 1991.

[11] J. L. Crassidis and F Landis Markley. Unscented filtering for spacecraft attitude estimation. *Journal of guidance, control, and dynamics*, 26(4):536–542, 2003.

[12] J. L. Crassidis, F. Landis Markley, and Yang Cheng. A survey of nonlinear attitude estimation methods. *Journal of Guidance, Control,and Dynamics*, 30(1):12–28, 2007.

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

[14] T.E. Duncan. Stochastic systems in Riemannian manifolds. *Journal of Optimization theory and applications*, 27(3):399–426, 1979.

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

[16] R. Durrett. *Probability: theory and examples*, volume 3. Cambridge university press, 2010.

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

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

[19] S.J. Julier and J.K. Uhlmann. A new extension of the Kalman Filter to Nonlinear Systems. *Proc. of AeroSense : The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls*, 1997.

[20] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Basic Eng., Trans. ASME, Ser. D,*, 83(3):95–108, 1961.

[21] Ch. Lageman, J. Trumpf, and R. Mahony. Gradient-like observers for invariant dynamics on a Lie group. *Automatic Control, IEEE Transactions on*, 55(2):367–377, 2010.

[22] E. J. Lefferts, F Landis Markley, and M. D. Shuster. Kalman filtering for spacecraft attitude estimation. *Journal of Guidance, Control, and Dynamics*, 5(5):417–429, 1982.

[23] M. Liao. *Lévy processes in Lie groups*. Cambridge University Press, 2004.

[24] JT-H Lo. Optimal estimation for the satellite attitude using star tracker measurements. *Automatica*, 22(4):477–482, 1986.

[25] J. Trumpf M. Zamani and R. Mahony. Near-optimal deterministic filtering on the rotation group. *IEEE Trans. on Automatic Control*, 56:6:1411 – 1414, 2011.

[26] R. Mahony, T. Hamel, and J-M Pflimlin. Nonlinear complementary filters on the special orthogonal group. *IEEE-Trans. on Automatic Control*, 53(5):1203–1218, 2008.

[27] F Landis Markley, J. L. Crassidis, and Y. Cheng. Nonlinear attitude filtering methods. In *AIAA Guidance, Navigation, and Control Conference*, 2005.

[28] Ph. Martin and E. Salaün. Invariant observers for attitude and heading estimation from low-cost inertial and magnetic sensors. In *46th IEEE Conf Decis Contr*, pages 1039–1045, 2007.

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

[30] H. Nijmeijer and T. I. Fossen. *New directions in nonlinear observer design*, volume 244. Springer, 1999.

[31] P. J. Olver. *Classical Invariant Theory*. Cambridge University Press, 1999.

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

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

[34] A. Saccon, J. Trumpf, R. Mahony, and A. P. Aguiar. Second-order-optimal filters on lie groups.

[35] S. Said. *Estimation and filtering of processes in matrix Lie groups*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2009.

[36] 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.

[37] S. Salcudean. A globally convergent angular velocity observer for rigid body motion. *Automatic Control, IEEE Transactions on*, 36(12):1493–1497, 1991.

[38] A. K. Sanyal, T. Lee, M. Leok, and N. H. McClamroch. Global optimal attitude estimation using uncertainty ellipsoids. *Systems & Control Letters*, 57(3):236–245, 2008.

[39] A. Tayebi, S. McGilvray, A. Roberts, and M. Moallem. Attitude estimation and stabilization of a rigid body using low-cost sensors. In *Decision and Control, 2007 46th IEEE Conference on*, pages 6424–6429. IEEE, 2007.

[40] J. Thienel and Robert M Sanner. A coupled nonlinear spacecraft attitude controller and observer with an unknown constant gyro bias and gyro noise. *Automatic Control, IEEE Transactions on*, 48(11):2011–2015, 2003.

[41] R. Tron, B. Afsari, and R. Vidal. Riemannian consensus for manifolds with bounded curvature. 2012.

[42] J.F. Vasconcelos, R. Cunha, C. Silvestre, and P. Oliveira. A nonlinear position and attitude observer on SE(3) using landmark measurements. *Systems Control Letters*, 59:155–166, 2010.

[43] 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.

[44] 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.

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

[46] K. C. Wolfe, M. Mashner, and G. S. Chirikjian. Bayesian fusion on Lie groups. *Journal of Algebraic Statistics*, 2(1):75–97, 2011.

[47] E. Wong and M. Zakai. Riemann-Stieltjes approximations of stochastic integrals. *Probability Theory and Related Fields*, 12(2):87–97, 1969.

Table 1: Artificial horizon: experiment parameters

Standard Deviation of the model noise	0.01 degree $= 1.75 \times 10^{-4}$ rad
Standard Deviation of the regular observations	0.1 degree $= 1.75 \times 10^{-3}$ rad
Standard Deviation of the outliers	30 degrees
Probability of the outliers to occur	0.01

Table 2: Useful formulas for some classical matrix Lie groups

G	$SO(3)$ ($\mathbb{R}^{\dim \mathfrak{g}} = \mathbb{R}^3$)	($\mathbb{R}^{\dim \mathfrak{g}} = \mathbb{R}^6$)
Embedding of G	$R \in \mathcal{M}_3(\mathbb{R}), R^T R = I_3$	$\begin{pmatrix} R & v \\ 0 & 1 \end{pmatrix}, R \in SO(3), u \in \mathbb{R}^3$
Embedding of $\mathfrak{g}$	$A_3 : \psi \in \mathcal{M}_3(\mathbb{R}), \psi^T = -\psi$	$\begin{pmatrix} \psi & u \\ 0 & 0 \end{pmatrix}, \psi \in A_3, u \in \mathbb{R}^3$
$(.)_m$	$\xi \to (\xi)_\times$	$\begin{pmatrix} \xi \\ u \end{pmatrix} \to \begin{pmatrix} (\xi)_\times & u \\ 0 & 0 \end{pmatrix}$
$x \to Ad_x$	$R \to R$	$\begin{pmatrix} R \\ T \end{pmatrix} \to \begin{pmatrix} R & 0 \\ (T)_\times R & R \end{pmatrix}$
$\xi \to ad_\xi$	$\xi \to (\xi)_\times$	$\begin{pmatrix} \xi \\ u \end{pmatrix} \to \begin{pmatrix} (\xi)_\times & 0 \\ (u)_\times & (\xi)_\times \end{pmatrix}$
exp	$\exp(x) = I_3 + \frac{sin(||x||)}{||x||}(x)_\times$ $+ \frac{1}{||x||^2}(1-cos|x|)(x)_m^2$	$\exp(x) = I_4 + (x)_m + \frac{[1-cos(||x||)]}{||x||^2}(x)_\times^2$ $+ \frac{1}{||x||^3}(||x|| - sin||x||)(x)_m^3$

Table 3: Observation of two vectors: parameters

Parameter	b_1	b_2	Q^{V_1}	Q^{V_2}
Value	$(1,0,0)$	$(0,1,0)$	$0,0873^2 I_3$	$0,0873^2 I_3$

Parameter	P_0	Q^w	N	Simulations
Value	$0.5236^2 I_3$	$0,01745^2 I_3$	50	1000