

# Iterated Invariant Extended Kalman Filter (IIEKF)

Sven Goffin\*, Axel Barrau†, Silvère Bonnabel‡, Olivier Brüls§, and Pierre Sacré¶

May 6, 2024

## Abstract

In this paper, we introduce the Iterated Invariant Extended Kalman Filter (IIEKF), which is an invariant extended Kalman filter (IEKF) where the updated state in the light of the latest measurement is defined as a maximum a posteriori (MAP) estimate. Under some compatibility requirements on the output map, we prove strong mathematical guarantees which echo those of the Kalman filter in the linear case. We apply the technique to two problems: solving a system of equations on a Lie group, and a problem of engineering interest, namely ego-localization of the hook of a crane. The latter serves as a benchmarking example, where the IIEKF favorably compares to other filters.

## 1 Introduction

The Kalman filter [1] solves in any respect the problem of estimating the state of a linear dynamical system from linear measurements corrupted by Gaussian noises. To accommodate nonlinearity of physical models and measurements, the Extended Kalman Filter (EKF), early developed in the Apollo program, and still very lively, relies on linearizations. However, linearization errors were soon recognized to degrade performance [2]. This motivated the development of iterative filtering algorithms such as the iterated EKF [3–5], iterated EKF on Lie groups [6, 7], and smoothers [8], where the discrepancy between the nonlinear output function and its first-order approximation is reduced by refining the operating point. It also inspired other iterative algorithms, such as the iterated extended Kalman particle filter [9].

On another note, the EKF, as an extension of the linear Kalman filter, inherently uses the vector space structure of the state space, and ignores all geometric properties the problem may possess. This latter concern has led to the development of geometric filtering methods, which today constitute a field of study in their own right. As concerns observer design, geometric methods may be traced back to attitude and pose estimation [10–12]. When turning to the more general problem of inertial navigation and localization, the invariant extended Kalman filter (IEKF) [13, 14] has become a key alternative to the standard EKF. Its theoretical properties include convergence guarantees [13], consistency properties for Simultaneous Localization and Mapping (SLAM) in robotics [14], see also related works [15–17], and have led to applications in various fields, [8, 14, 15, 18–21] and in the industry [14]. The field of observers has benefited from the introduction of new groups brought by the IEKF theory, see e.g., [22–24], namely the groups  $SE_2(3)$  and  $SE_k(d)$  introduced in [13, 25] and their relation to the navigation equations. A unifying perspective on the Lie groups that arise from the IEKF theory can be found in the recent paper [26]. The vibrant equivariant observer framework, see [27], is closely related.

In this paper, we introduce the iterated invariant extended Kalman filter (IIEKF). The rationale is to focus on the measurement update step of the invariant extended Kalman filter (IEKF) [14] and to improve it using the Gauss-Newton (GN) method [28]. Instead of attempting to explicitly account for the manifold structure of the state space to adapt the optimization algorithm of the iterated EKF at a general level as in [6], we make a limited number of modifications to the IEKF, by seeking the updated state as the maximum a posteriori (MAP) in terms of the linearized invariant error in the Lie algebra. We start by first explicating the sensible properties of the linear Kalman filter we seek to elicit in a

---

\*S. Goffin is a FRIA grantee of the Fonds de la Recherche Scientifique - FNRS.

†A. Barrau is with OFFROAD, France (axel@offroad.works).

‡S. Bonnabel is with MINES ParisTech, PSL Research University, France (silvere.bonnabel@mines-paristech.fr).

§O. Brüls is with the Department of Aerospace and Mechanical Engineering, University of Liège, Belgium (o.bruls@uliege.be).

¶S. Goffin and P. Sacré are with the Department of Electrical Engineering and Computer Science, University of Liège, Belgium (sven.goffin@uliege.be; p.sacre@uliege.be).

nonlinear context, and then we prove the IIEKF possesses them indeed. The differences between the proposed method and the general iterated EKF on Lie groups from [6] and [7] are detailed in Appendix D.

The fact the IIEKF recovers a number of provable properties and guarantees of the linear case, although in a nonlinear context, is made possible by the fact we consider specific state spaces (Lie groups), and specific output maps, namely the invariant measurements of [13]. To come up with mathematical properties, we consider the limit case where the measurement noise goes to zero. This limit case is particularly meaningful, as the improvements brought by iterative methods are sharper when measurements are accurate, since refining the linearization point then pays off greatly.

The method is applied to two different problems. The first problem is the analog to the mathematical problem of solving a system of linear equations, but when the unknown lives in a Lie group. The second problem, is a problem of engineering interest, that allows for comparisons between the proposed IIEKF, the IEKF, the EKF, and the iterated EKF.

Throughout the paper, the acronym IEKF is reserved for the invariant EKF, whereas we use IterEKF for the iterated EKF.

The paper is structured as follows. Section 2 recalls the equations of the Kalman filter, EKF, and iterated EKF. Section 3 analyzes the update step of the KF and EKF in the case of noise-free measurements. The desirable properties we expect from a probabilistic filter in the noise-free measurement case are then developed. Section 4 introduces the proposed IIEKF algorithm for both noisy and noise-free measurements. Section 5 presents the theoretical properties of the IIEKF. Section 6 applies the method to the problem of solving a system of equations on a Lie group. A simple example on  $SO(3)$  sheds light on the way the IIEKF operates. Finally, Section 7 compares the performance of the proposed algorithm and other state-of-the-art filters on the problem of estimating the extended pose (orientation, velocity, position) of the hook of a crane on which an inertial measurement unit (IMU) is mounted.

## 2 Kalman filter, EKF, iterated EKF

The goal of the present section is to fix notation and summarize a few known facts.

### 2.1 Kalman filter equations

Consider the linear system

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{w}_k, \quad (1a)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k, \quad (1b)$$

where  $\mathbf{x}_k \in \mathbb{R}^n$ ,  $\mathbf{y}_k \in \mathbb{R}^m$ ,  $\mathbf{u}_k \in \mathbb{R}^b$ ,  $\mathbf{F}_k \in \mathbb{R}^{n \times n}$ ,  $\mathbf{B}_k \in \mathbb{R}^{n \times b}$ , and  $\mathbf{H}_k \in \mathbb{R}^{m \times n}$ . The process and measurement noises are such that  $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$  and  $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{N}_k)$ .

The Kalman filter (KF) computes at all times the statistics of the state  $\mathbf{x}_k$  given the measurements gathered so far.

**Propagation** At propagation step, the KF computes

$$\hat{\mathbf{x}}_{k+1} = \mathbf{F}_k \hat{\mathbf{x}}_k + \mathbf{B}_k \mathbf{u}_k, \quad (2a)$$

$$\mathbf{P}_{k+1} = \mathbf{F}_k \mathbf{P}_k \mathbf{F}_k^T + \mathbf{Q}_k. \quad (2b)$$

Assuming  $p(\mathbf{x}_k | \mathbf{y}_0, \dots, \mathbf{y}_k) = \mathcal{N}(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ , we have that (2) ensures  $p(\mathbf{x}_{k+1} | \mathbf{y}_0, \dots, \mathbf{y}_k) = \mathcal{N}(\hat{\mathbf{x}}_{k+1}, \mathbf{P}_{k+1})$ .

**Update** Departing from  $p(\mathbf{x}_k | \mathbf{y}_0, \dots, \mathbf{y}_{k-1}) = \mathcal{N}(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ , and receiving measurement  $\mathbf{y}_k$  from (1b), the KF updates the estimates according to the equations

$$\mathbf{z}_k = \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k, \quad (3a)$$

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{N}_k, \quad (3b)$$

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^T \mathbf{S}_k^{-1}, \quad (3c)$$

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k + \mathbf{K}_k \mathbf{z}_k, \quad (3d)$$

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k, \quad (3e)$$

where  $\mathbf{z}_k$  is the innovation (i.e., a prediction error),  $\mathbf{S}_k$  is the innovation covariance, and  $\mathbf{K}_k$  is the Kalman gain. We have then  $p(\mathbf{x}_k | \mathbf{y}_0, \dots, \mathbf{y}_k) = \mathcal{N}(\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+)$ .

## 2.2 Extended Kalman filter equations

Consider now the nonlinear system

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k), \quad (4a)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}_k, \quad (4b)$$

where  $\mathbf{f}$  and  $\mathbf{h}_k$  are nonlinear functions. Owing to nonlinearity, the extended Kalman filter (EKF) maintains a Gaussian approximation, called ‘‘belief’’, of the (intractable) statistics of the state  $\mathbf{x}_k$ , given the measurements gathered so far.

**Propagation** The EKF propagates its belief through the state dynamics model as

$$\hat{\mathbf{x}}_{k+1} = \mathbf{f}(\hat{\mathbf{x}}_k, \mathbf{u}_k, \mathbf{0}), \quad (5a)$$

$$\mathbf{P}_{k+1} = \mathbf{F}_k \mathbf{P}_k \mathbf{F}_k^T + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T, \quad (5b)$$

where  $\mathbf{F}_k$  and  $\mathbf{G}_k$  are the Jacobians of  $\mathbf{f}$  with respect to (w.r.t.)  $\mathbf{x}_k$  and  $\mathbf{w}_k$  respectively, evaluated at  $(\hat{\mathbf{x}}_k, \mathbf{u}_k, \mathbf{0})$ .

**Update** The EKF linearizes the function  $\mathbf{h}_k$  around the current estimate  $\hat{\mathbf{x}}_k$ , yielding Jacobian  $\mathbf{H}_k$ . It uses the same update equations as the KF, except for the innovation that now writes  $\mathbf{z}_k = \mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_k)$ .

## 2.3 Updated state and the maximum a posteriori (MAP)

Consider again measurement (4b) given by

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}_k. \quad (6)$$

Let us start from a prior Gaussian belief  $p(\mathbf{x}_k \mid \mathbf{y}_0, \dots, \mathbf{y}_{k-1}) := \tilde{p}(\mathbf{x}_k)$  in the form of a Gaussian  $\mathcal{N}(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ . In the light of a present measurement  $\mathbf{y}_k$  the maximum a posteriori (MAP) estimate is obtained by minimizing the posterior negative log likelihood, i.e.,

$$\hat{\mathbf{x}}_k^* = \underset{\mathbf{x}}{\operatorname{argmin}} -\ln(\tilde{p}(\mathbf{x} \mid \mathbf{y}_k)), \quad (7a)$$

$$= \underset{\mathbf{x}}{\operatorname{argmin}} -\ln(p(\mathbf{y}_k \mid \mathbf{x})\tilde{p}(\mathbf{x})) + \text{constant}, \quad (7b)$$

$$= \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x} - \hat{\mathbf{x}}_k\|_{\mathbf{P}_k}^2 + \frac{1}{2} \|\mathbf{y}_k - \mathbf{h}_k(\mathbf{x})\|_{\mathbf{N}_k}^2, \quad (7c)$$

where we used the Bayes rule and  $\|\mathbf{x}\|_{\mathbf{P}}^2 = \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}$  denotes the squared Mahalanobis distance.

In the case where  $\mathbf{h}_k$  is linear, i.e.,  $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k$ , the estimate  $\hat{\mathbf{x}}_k^*$  provided by (3d) exactly solves the optimization problem (7c). However, when  $\mathbf{h}_k$  is a nonlinear function, the estimate provided by the EKF is the result of a linearization and is only an approximation to the solution of problem (7c).

## 2.4 The iterated EKF

The optimization problem of finding the MAP may be addressed through the Gauss-Newton method, leading to the iterated EKF, as shown by [3].

**Lemma 1** (from [3]). *Consider the optimization problem*

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x} - \bar{\mathbf{x}}\|_{\mathbf{A}}^2 + \frac{1}{2} \|\mathbf{b} - \bar{\mathbf{h}}(\mathbf{x})\|_{\mathbf{B}}^2, \quad (8)$$

with  $\mathbf{x}, \bar{\mathbf{x}} \in \mathbb{R}^n$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{B} \in \mathbb{R}^{m \times m}$  and  $\bar{\mathbf{h}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . This problem can be attacked with the Gauss-Newton (GN) method, yielding the sequence of estimates

$$\mathbf{x}^{i+1} = \bar{\mathbf{x}} + \bar{\mathbf{K}}^i (\mathbf{b} - \bar{\mathbf{h}}(\mathbf{x}^i) - \bar{\mathbf{H}}^i (\bar{\mathbf{x}} - \mathbf{x}^i)), \quad (9)$$

with

$$\bar{\mathbf{H}}^i = \bar{\mathbf{h}}'(\mathbf{x}^i), \quad (10a)$$

$$\bar{\mathbf{K}}^i = \mathbf{A} (\bar{\mathbf{H}}^i)^T (\bar{\mathbf{H}}^i \mathbf{A} (\bar{\mathbf{H}}^i)^T + \mathbf{B})^{-1}, \quad (10b)$$

where  $\bar{\mathbf{h}}'(\mathbf{x}^i)$  denotes the Jacobian of  $\bar{\mathbf{h}}$  evaluated at  $\mathbf{x}^i$ .

One may bring Lemma 1 to bear on the optimization problem (7c) of finding the most likely estimate in the light of latest measurement. It leads to the equations of the iterated EKF (IterEKF), see [3], that is,

$$\hat{\mathbf{x}}_k^{i+1} = \hat{\mathbf{x}}_k + \mathbf{K}_k^i (\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k^i) - \mathbf{H}_k^i (\hat{\mathbf{x}}_k - \mathbf{x}_k^i)), \quad (11a)$$

$$\mathbf{H}_k^i = \mathbf{h}'_k(\mathbf{x}_k^i), \quad (11b)$$

$$\mathbf{K}_k^i = \mathbf{P}_k (\mathbf{H}_k^i)^T (\mathbf{H}_k^i \mathbf{P}_k (\mathbf{H}_k^i)^T + \mathbf{N}_k)^{-1}. \quad (11c)$$

Since we initialize at  $\mathbf{x}_k^0 = \hat{\mathbf{x}}_k$ , we note the conventional EKF then amounts to one Gauss-Newton iteration.

### 3 Desirable properties to expect in the face of noise-free measurements

In this section, we inspect the case where measurement noise tends to zero. This allows us to set the stage for the problems we will solve.

#### 3.1 Noise-free measurements

As sensors get more accurate, the eigenvalues  $\lambda_j(\mathbf{N}_k)$  of covariance matrix  $\mathbf{N}_k$  get smaller and smaller. In the limit case of noise-free measurements, i.e.,  $\lambda_j(\mathbf{N}_k) = 0$  for all  $j$ , the unconstrained optimization problem (7c) becomes the following constrained optimization problem, see [29],

$$\hat{\mathbf{x}}_k^* = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{x} - \hat{\mathbf{x}}_k\|_{\mathbf{P}_k}^2 \mid \mathbf{h}_k(\mathbf{x}) = \mathbf{y}_k \right\}. \quad (12)$$

Let us focus on the linear case, and set  $\mathbf{N}_k = \mathbf{0}$ . The KF update then becomes

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k + \underbrace{\mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T)^{-1}}_{\mathbf{K}_k} \underbrace{(\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_k))}_{\mathbf{z}_k}. \quad (13)$$

Provided that  $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T$  is invertible, it returns the solution of the constrained optimization problem.

Noise-free measurements, which will play an important role in the sequel, completely remove uncertainty in the observed directions. This brings two issues.

**Issue 1** The covariance matrix  $\mathbf{P}_k$  can cease to be invertible, as noise-free implies null variance, in which case (12) is undefined.

**Issue 2** The innovation covariance  $\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T$  can cease to be invertible in such a way that the Kalman gain  $\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T)^{-1}$  in (13) is undefined. This occurs when the perfectly observed directions overlap between subsequent updates for example. This issue is solved by using a “noise-free limit gain” instead. It was developed in the preliminary conference paper [30] of the present article, and is given by

$$\mathbf{K}_k^{\text{nf}} = \lim_{\delta \rightarrow 0} \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \delta \mathbf{I})^{-1} = \mathbf{L}_k (\mathbf{H}_k \mathbf{L}_k)^{\dagger}, \quad (14)$$

where  $(\cdot)^{\dagger}$  is the Moore-Penrose pseudo-inverse,  $\mathbf{P}_k = \mathbf{L}_k \mathbf{L}_k^T$ , and nf stands for “noise-free”. This gain is always defined and is used for theoretical developments in the noise-free scenario. However, we have observed that it is not well suited to practical implementations, as it can be difficult to distinguish small singular values from actual zeros in the calculation of the pseudo-inverse, and haphazard thresholding can lead to the inversion of really small values, which can destabilize the filter. For this reason, the regularized gain

$$\mathbf{K}_k^{\text{rg}} = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \delta \mathbf{I})^{-1}, \quad \text{with } \delta \ll 1, \quad (15)$$

proves better suited to practical implementations.

## 3.2 Desirable filter properties in the degenerate case of noise-free measurements

Consider noise-free measurements in a nonlinear context

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k). \quad (16)$$

Such measurements convey sure information, namely that the state belongs to a subset of the state space.

**Definition 1.** *When facing noise-free measurement (16), we call observed set associated to  $\mathbf{y}_k$  the set*

$$\mathcal{S}_{\{\mathbf{h}_k(\mathbf{x})=\mathbf{y}_k\}} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{h}_k(\mathbf{x}) = \mathbf{y}_k\}. \quad (17)$$

### 3.2.1 EKF compatibility

One may wonder how this information should translate when using a Gaussian filter. To this aim, let us recall how uncertainty is represented. The EKF is based on a Gaussian belief about the state  $\mathbf{x}_k$  as follows

$$\mathbf{x}_k = \hat{\mathbf{x}}_k + \boldsymbol{\xi}_k, \quad \boldsymbol{\xi}_k \sim \mathcal{N}(0, \mathbf{P}_k), \quad (18)$$

and maintains it a Gaussian by systematically linearizing about the current estimate. The estimate  $\hat{\mathbf{x}}_k$  is considered as the “best” estimate, and  $\boldsymbol{\xi}_k$  encodes a small unknown dispersion around it. Following this rationale, suppose we know for sure the state is in the observed set (17). We have

$$\mathbf{x}_k \in \mathcal{S}_{\{\mathbf{h}_k(\mathbf{x})=\mathbf{y}_k\}} \Rightarrow \mathbf{h}_k(\mathbf{x}_k) = \mathbf{y}_k, \quad (19)$$

$$\Rightarrow \mathbf{h}_k(\hat{\mathbf{x}}_k + \boldsymbol{\xi}_k) = \mathbf{y}_k, \quad (20)$$

$$\Rightarrow \mathbf{h}_k(\hat{\mathbf{x}}_k) + \mathbf{H}_k \boldsymbol{\xi}_k = \mathbf{y}_k, \quad (21)$$

where  $\mathbf{H}_k$  denotes the Jacobian of  $\mathbf{h}_k$  at  $\hat{\mathbf{x}}_k$ , and where we neglect second-order terms in the error  $\boldsymbol{\xi}_k$ , that is, we linearize. As this should be true for all  $\boldsymbol{\xi}_k$ , and in particular for  $\boldsymbol{\xi}_k = \mathbf{0}$ , we necessarily get  $\mathbf{h}_k(\hat{\mathbf{x}}_k) = \mathbf{y}_k$  and  $\mathbf{H}_k \boldsymbol{\xi}_k = \mathbf{0}$ . As a consequence, the expectation  $E(\mathbf{H}_k \boldsymbol{\xi}_k (\mathbf{H}_k \boldsymbol{\xi}_k)^T) = \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}$ . This motivates the following definition, which encodes the certain information (16) in terms of the Gaussian parameters.

**Definition 2.** *The EKF belief (18) is said to be compatible with noise-free measurement (16) if we have both*

1.  $\hat{\mathbf{x}}_k \in \mathcal{S}_{\{\mathbf{h}_k(\mathbf{x})=\mathbf{y}_k\}},$

2.  $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0},$

where  $\mathbf{H}_k$  denotes the Jacobian of  $\mathbf{h}_k$  at  $\hat{\mathbf{x}}_k$ .

### 3.2.2 Resulting desirable properties

Given how a noise-free measurement ought to translate in terms of Gaussian parameters, we may deduce properties that should be satisfied when updating the state in the light of a noise-free measurement.

**Property 1: The update yields a compatible belief.** Starting from an arbitrary prior belief  $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ , the updated belief  $(\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+)$  in the light of measurement (16) should encode the information contained in (16) in the sense of Definition 2. Mathematically, it boils down to

$$\mathbf{h}_k(\hat{\mathbf{x}}_k^+) = \mathbf{y}_k, \quad (22)$$

$$\mathbf{H}_k^+ \mathbf{P}_k^+ (\mathbf{H}_k^+)^T = \mathbf{0}, \quad (23)$$

where  $\mathbf{H}_k^+$  denotes the Jacobian of  $\mathbf{h}_k$ , evaluated at  $\hat{\mathbf{x}}_k^+$ .

The second property is somewhat less obvious. It says that once a belief is compatible with a measurement, any immediate further—possibly unrelated—measurement should not “destroy” the certain information already encoded. Albeit meaningful, it is generally violated by the EKF.

**Property 2: Further updates preserve compatibility.** Suppose we start from a prior belief  $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$  already compatible with measurement (16),  $\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k)$ , i.e., ensuring the two conditions of Definition 2 are satisfied. Assume we *immediately* make a different (possibly noisy) arbitrary measurement  $\tilde{\mathbf{y}}_k = \tilde{\mathbf{h}}_k(\mathbf{x}_k) + \mathbf{n}_k$ . The updated belief  $(\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+)$  in the light of measurement  $\tilde{\mathbf{y}}_k$  should remain compatible with (16), that is,  $\mathbf{h}_k(\hat{\mathbf{x}}_k^+) = \mathbf{y}_k$  and  $\mathbf{H}_k^+ \mathbf{P}_k^+ (\mathbf{H}_k^+)^T = \mathbf{0}$ , with  $\mathbf{H}_k^+$  the Jacobian of  $\mathbf{h}_k$  evaluated at updated  $\hat{\mathbf{x}}_k^+$ , no matter  $\tilde{\mathbf{y}}_k$ . This means the deterministic information (16) has been appropriately encoded.

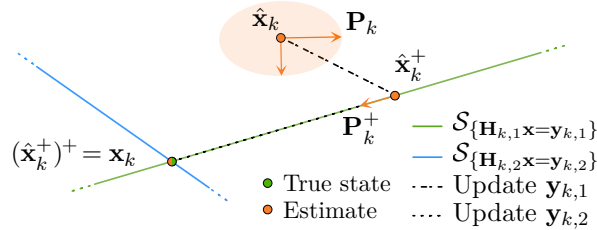


Figure 1: Graphical representation of two consecutive noise-free updates for a linear Kalman filter operating in 2D. The covariance matrix is represented by an ellipsoid where the true state lies with high probability, which becomes flat and aligned with the observed set after a perfect measurement is made.

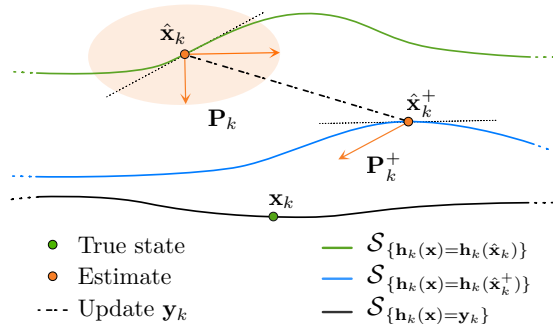


Figure 2: Graphical representation of EKF issues with Property 1. The level curves of  $h$  are represented, and the observed set is the bottom one. 1) we see the EKF update  $\hat{\mathbf{x}}_k^+$  fails to “land” on the observed set, so that (22) fails to hold. 2) in terms of updated uncertainty ellipsoid, it rightfully becomes flat, owing to sure information. Besides, tangential directions to level curves at  $\hat{\mathbf{x}}_k$  are  $\ker \mathbf{H}_k$ , with  $\mathbf{H}_k = \mathbf{h}'(\hat{\mathbf{x}}_k)$ . The tangent directions at updated state  $\hat{\mathbf{x}}_k^+$ , given by  $\ker \mathbf{H}_k^+$ , differ. As the EKF linearizes at point  $\hat{\mathbf{x}}_k$  to perform its update, we have  $\mathbf{H}_k \mathbf{P}_k^+ \mathbf{H}_k^T = 0$ , but no information about  $\hat{\mathbf{x}}_k^+$  is used, and  $\mathbf{H}_k^+ \mathbf{P}_k^+ (\mathbf{H}_k^+)^T \neq 0$ . Thus, (23) also fails to hold.

### 3.3 Noise-free measurements: linear vs nonlinear case

In the easier linear case, the Kalman filter is proved to possess the enunciated properties (see Appendix A). This stems from the fact that 1) we have  $\mathbf{H}_k^+ = \mathbf{H}_k$ , that is, there is no dependency on a “linearization point” that may vary with updates, and 2) the observed set becomes  $\mathcal{S}_{\{\mathbf{H}_k \mathbf{x} = \mathbf{y}_k\}}$ , that is, a linear subspace. Properties 1 and 2 are illustrated in Fig. 1, where updates can be viewed as projections.

When turning to the nonlinear case, the observed set  $\mathcal{S}_{\{\mathbf{h}_k(\mathbf{x}) = \mathbf{y}_k\}}$  is not a vector space. Moreover, the local information encoded around the current best estimate by linearizing  $\mathbf{h}_k$  does not contain any information about this function far away from that point (contrary to the linear case). This has several implications. First, there is no reason why the update should “land” on the desired set  $\mathcal{S}_{\{\mathbf{h}_k(\mathbf{x}) = \mathbf{y}_k\}}$ , i.e., satisfy (22). Then, the direction encoded by  $\mathbf{H}_k$  arbitrarily differs from that encoded by  $\mathbf{H}_k^+$ . Since  $\mathbf{P}_k$  is updated using  $\mathbf{H}_k$ , see (3e), there is no reason to have Property 1, see (23), as illustrated in Fig. 2. Finally, regarding Property 2, the issue stemming from linearizing in the face of nonlinear measurements is illustrated in Fig. 3. Thus, in general, there is no reason why any of the two desirable properties should hold when using an EKF.

## 4 The iterated invariant extended Kalman filter (IIEKF)

The invariant extended Kalman filter (IEKF) of [13] exactly handles the nonlinear dynamics of group affine systems when process noise tends to zero, owing to the so-called log-linearity property. Regarding updates, there is no counterpart of such an exact property: the IEKF only provides approximations when it updates its belief with a left- or right-invariant measurement, and this even if measurement noise tends to zero. The present paper aims at improving the IEKF to obtain such properties.

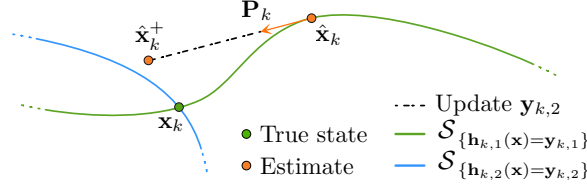


Figure 3: Graphical representation of EKF issues with Property 2. The EKF initial belief is assumed compatible with a prior noise-free measurement, i.e.,  $\mathbf{h}_{k,1}(\hat{\mathbf{x}}_k) = \mathbf{y}_1$  and  $\mathbf{H}_1 \mathbf{P}_k \mathbf{H}_1^T = \mathbf{0}$  with  $\mathbf{H}_1 = \mathbf{h}'_{k,1}(\hat{\mathbf{x}}_k)$ . Any immediately subsequent observation should maintain the estimate within the observed set  $\mathbf{h}_{k,1}(\mathbf{x}) = \mathbf{y}_{k,1}$ , as the information is sure. Contrary to the linear KF, though,  $\mathbf{P}_k$  only captures infinitesimal variability. Thus, when trying to incorporate a different measurement  $\mathbf{y}_{k,2}$ , although the EKF update points to a meaningful direction locally, the nonlinear nature of the observed set makes the updated estimate step out of it.

#### 4.1 The invariant EKF (IEKF)

Consider a state  $\chi_k$  evolving in a Lie group  $G$ . We assume  $G$  is a matrix Lie group ( $\chi_k \in \mathbb{R}^{N \times N}$ ) with the matrix multiplication as composition law. As this operation is not commutative, the IEKF can be designed following the left- or right-invariant formalism, depending on whether one is dealing with left-invariant or right-invariant measurements. In this paper, we focus on the left-IEKF, the right-IEKF “mirroring” the latter, see [14]. The state is defined through the following nonlinear system

$$\chi_{k+1} = \mathbf{f}(\chi_k, \mathbf{u}_k) \exp_G(\mathbf{w}_k), \quad (24a)$$

$$\mathbf{y}_k = \chi_k \mathbf{d}_k + \mathbf{n}_k, \quad (24b)$$

where  $\mathbf{u}_k \in \mathbb{R}^m$  is the system input,  $\mathbf{d}_k \in \mathbb{R}^N$ ,  $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$  is the process noise, and  $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{N}_k)$  is the measurement noise. The IEKF seeks to estimate  $\chi_k$ .

**Uncertainty representation** The left-IEKF defines the nonlinear error

$$\boldsymbol{\eta}_k = \hat{\chi}_k^{-1} \chi_k = \exp_G(\boldsymbol{\xi}_k), \quad (25)$$

where  $\exp_G$  is the exponential map of Lie group  $G$ , and where  $\boldsymbol{\xi}_k \in \mathbb{R}^n$  is called the linearized error<sup>1</sup>. The linearized error is assumed to follow a Gaussian distribution, so that the state follows a concentrated Gaussian distribution on  $G$ , centered on the current estimate  $\hat{\chi}_k$  [31–34], that is,

$$\chi_k = \hat{\chi}_k \exp_G(\boldsymbol{\xi}_k), \quad \text{with } \boldsymbol{\xi}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_k). \quad (26)$$

In this formula, which is a Lie group analog to (18),  $\hat{\chi}_k$  represents the noise-free “mean” of the Gaussian, and the dispersion arises by multiplying it with the random matrix  $\exp_G(\boldsymbol{\xi}_k)$ , see e.g., [14].

**Propagation** The belief is propagated through dynamics (24a) with no process noise. The linearized error thus evolves as  $\boldsymbol{\xi}_{k+1} = \mathbf{F}_k \boldsymbol{\xi}_k + \mathbf{G}_k \mathbf{w}_k$ , so that

$$\hat{\chi}_{k+1} = \mathbf{f}(\hat{\chi}_k, \mathbf{u}_k), \quad (27a)$$

$$\mathbf{P}_{k+1} = \mathbf{F}_k \mathbf{P}_k \mathbf{F}_k^T + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T, \quad (27b)$$

with  $\mathbf{F}_k$ ,  $\mathbf{G}_k$  the Jacobians w.r.t. the chosen linearized error and the process noise, see [14], evaluated at  $(\hat{\chi}_k, \mathbf{u}_k, \mathbf{0})$ . Note that the method developed in this paper does not require the function  $\mathbf{f}$  to be group affine, i.e., it does not need to satisfy (11) in [14].

**Update** Measurement (24b) brings partial information about the current error  $\boldsymbol{\xi}_k$ . In the (left) invariant filtering framework, the innovation (prediction error) is defined as

$$\mathbf{z}_k = \hat{\chi}_k^{-1} \mathbf{y}_k - \mathbf{d}_k, \quad (28a)$$

$$= \exp_G(\boldsymbol{\xi}_k) \mathbf{d}_k - \mathbf{d}_k + \hat{\chi}_k^{-1} \mathbf{n}_k, \quad (28b)$$

$$= \mathbf{H}_k \boldsymbol{\xi}_k + \chi_k^{-1} \mathbf{n}_k + \mathcal{O}(\|\boldsymbol{\xi}_k\|^2), \quad (28c)$$

<sup>1</sup>We identify the Lie algebra with  $\mathbb{R}^n$ , that is, letting  $\mathcal{L}_{\mathfrak{g}}(\cdot) : \mathbb{R}^n \rightarrow \mathfrak{g}$  be the bijective map between the vector space  $\mathbb{R}^n$  and the Lie algebra  $\mathfrak{g}$  of group  $G$ , with  $n = \dim \mathfrak{g}$ , see [13], we have  $\exp_G(\boldsymbol{\xi}) := \exp(\mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi}))$  with  $\exp$  the matrix exponential.

where  $\mathbf{H}_k$  is the invariant filtering output Jacobian

$$\exp_G(\boldsymbol{\xi})\mathbf{d}_k = \mathbf{d}_k + \mathbf{H}_k\boldsymbol{\xi} + \mathcal{O}(\|\boldsymbol{\xi}\|^2). \quad (29)$$

associated to map  $\mathbf{h}_k(\boldsymbol{\chi}) = \boldsymbol{\chi}\mathbf{d}_k$ , whose mathematical expression will be recalled at (46). A flagship property of invariant filtering is that, owing to the form of the output, the Jacobian  $\mathbf{H}_k$  does not depend on the current estimate  $\hat{\boldsymbol{\chi}}_k$  (only on  $\mathbf{d}_k$ ), see (28b). In the light of measurement  $\mathbf{y}_k$ , the (left) IEKF [13] updates its belief according to

$$\mathbf{K}_k = \mathbf{P}_k\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^T + \hat{\mathbf{N}}_k)^{-1}, \quad (30a)$$

$$\boldsymbol{\xi}_k^+ = \mathbf{K}_k\mathbf{z}_k, \quad (30b)$$

$$\hat{\boldsymbol{\chi}}_k^+ = \hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi}_k^+), \quad (30c)$$

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k, \quad (30d)$$

where  $\hat{\mathbf{N}}_k = \hat{\boldsymbol{\chi}}_k^{-1}\mathbf{N}_k(\hat{\boldsymbol{\chi}}_k^{-1})^T$ .

## 4.2 Derivation of the iterated invariant EKF (IEKF)

In this paper, we introduce the iterated IEKF, that attempts to solve the MAP problem by using the GN method.

**Proposition 1.** *Starting from the prior (26), supposed to encode the state distribution conditional on past information  $\mathbf{y}_0, \dots, \mathbf{y}_{k-1}$ , the MAP estimate in the light of latest measurement  $\mathbf{y}_k$  is given by  $\hat{\boldsymbol{\chi}}_k^* = \hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi}_k^*)$ , where  $\boldsymbol{\xi}_k^*$  is the solution to the optimization problem*

$$\boldsymbol{\xi}_k^* = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{\xi}\|_{\mathbf{P}_k}^2 + \frac{1}{2}\|\mathbf{z}_k - \exp_G(\boldsymbol{\xi})\mathbf{d}_k + \mathbf{d}_k\|_{\hat{\mathbf{N}}_k}^2. \quad (31)$$

*Proof.* We have

$$p(\boldsymbol{\xi}_k | \mathbf{y}_k) = p(\boldsymbol{\xi}_k | \mathbf{z}_k) \propto p(\mathbf{z}_k | \boldsymbol{\xi}_k)p(\boldsymbol{\xi}_k), \quad (32)$$

with all densities  $p$  implicitly conditional on past information  $\mathbf{y}_0, \dots, \mathbf{y}_{k-1}$ . Then, the density  $p(\boldsymbol{\xi}_k | \mathbf{y}_k)$  is proportional to  $\exp(-\frac{1}{2}\|\boldsymbol{\xi}_k\|_{\mathbf{P}_k}^2) \exp(-\frac{1}{2}\|\mathbf{z}_k - \exp_G(\boldsymbol{\xi}_k)\mathbf{d}_k + \mathbf{d}_k\|_{\hat{\mathbf{N}}_k}^2)$ .  $\square$

Akin to the conventional iterated EKF [3], we intend to solve (31) iteratively using the GN method. All we need to do is linearize the Lie exponential.

**Lemma 2.** *We have the following first-order expansion*

$$\exp_G(\bar{\boldsymbol{\xi}} + \boldsymbol{\delta})\mathbf{d} = \exp_G(\bar{\boldsymbol{\xi}})\mathbf{d} + \exp_G(\bar{\boldsymbol{\xi}})\mathbf{H}_k\mathcal{J}_r(\bar{\boldsymbol{\xi}})\boldsymbol{\delta} + \|\boldsymbol{\delta}\|^2, \quad (33)$$

with  $\mathbf{H}_k$  as in (29), and where the right Lie Jacobian [31] is defined by  $\exp_G(\boldsymbol{\xi} + \boldsymbol{\delta}) \approx \exp_G(\boldsymbol{\xi}) \exp_G(\mathcal{J}_r(\boldsymbol{\xi})\boldsymbol{\delta})$  neglecting terms of order  $\|\boldsymbol{\delta}\|^2$ . Its expression is given in Appendix B.

*Proof.* We have the first-order expansion in  $\boldsymbol{\delta}$

$$\exp_G(\bar{\boldsymbol{\xi}} + \boldsymbol{\delta})\mathbf{d} \approx \exp_G(\bar{\boldsymbol{\xi}})\exp_G(\mathcal{J}_r(\bar{\boldsymbol{\xi}})\boldsymbol{\delta})\mathbf{d}, \quad (34a)$$

$$\approx \exp_G(\bar{\boldsymbol{\xi}})(\mathbf{d} + \mathbf{H}_k\mathcal{J}_r(\bar{\boldsymbol{\xi}})\boldsymbol{\delta}). \quad (34b)$$

where we applied (29) with  $\boldsymbol{\xi} = \mathcal{J}_r(\bar{\boldsymbol{\xi}})\boldsymbol{\delta}$ .  $\square$

Now, letting  $\mathbf{b} := \mathbf{z}_k + \mathbf{d}_k$ ,  $\bar{\mathbf{h}}(\boldsymbol{\xi}) := \exp_G(\boldsymbol{\xi})\mathbf{d}_k$ , and  $\bar{\mathbf{x}} = \mathbf{0}$ , we see objectives (31) and (8) coincide, so that a direct application of Lemma 1 of Section 2.4 and Lemma 2 yields the following proposition.

**Proposition 2.** *The sequence of GN updates for the optimization problem (31) writes*

$$\boldsymbol{\xi}_k^{i+1} = \mathbf{K}_k^i(\mathbf{z}_k - \exp_G(\boldsymbol{\xi}_k^i)\mathbf{d}_k + \mathbf{d}_k + \mathbf{H}_k^i\boldsymbol{\xi}_k^i), \quad (35)$$

letting  $\mathbf{H}_k$  be the IEKF's standard Jacobian and

$$\mathbf{H}_k^i = \exp_G(\boldsymbol{\xi}_k^i)\mathbf{H}_k\mathcal{J}_r(\boldsymbol{\xi}_k^i), \quad (36)$$

$$\mathbf{K}_k^i = \mathbf{P}_k(\mathbf{H}_k^i)^T \left( \mathbf{H}_k^i\mathbf{P}_k(\mathbf{H}_k^i)^T + \hat{\mathbf{N}}_k \right)^{-1}. \quad (37)$$



Assuming the GN method converges after  $i^*$  iterations, the IIEKF eventually updates its estimate following the IEKF methodology (30c), i.e.,

$$\hat{\boldsymbol{\chi}}_k^+ = \hat{\boldsymbol{\chi}}_k \exp_G \left( \boldsymbol{\xi}_k^{i^*} \right). \quad (38)$$

Since Jacobian  $\mathbf{H}_k$  is independent from the current estimate, the Riccati update does not require any iteration, and the IIEKF updates its error covariance once and for all as follows

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k, \quad (39)$$

with  $\mathbf{H}_k = \mathbf{H}_k^0$  and  $\mathbf{K}_k = \mathbf{K}_k^0$ . The full algorithm is summarized in Algorithm 1, and we see we exactly recover the IEKF of [13] if we perform one iteration only.

---

**Algorithm 1** The iterated invariant extended Kalman filter (IIEKF)

---

```

1: Choose the initial state  $\hat{\boldsymbol{\chi}}_0 \in G$  and initial covariance  $\mathbf{P}_0 = \text{Cov}(\boldsymbol{\xi}_0) \in \mathbb{R}^{n \times n}$ .
2: loop
3:   Define dynamics Jacobian  $\mathbf{F}_k$  of IEKF framework.
4:   Define dynamics Jacobian  $\mathbf{G}_k$  w.r.t. process noise.
5:   Define output Jacobian  $\mathbf{H}_k$  of IEKF framework [13].
6:   Define process noise covariance  $\mathbf{Q}_k$ .
7:   Define measurement noise covariance  $\mathbf{N}_k$ .
8:
9:   {Measurement update}
10:   $i \leftarrow 0$ 
11:   $\boldsymbol{\xi}_k^i \leftarrow \mathbf{0}$ 
12:   $\Delta \boldsymbol{\xi}_k^i \leftarrow +\infty$ 
13:   $\mathbf{z}_k \leftarrow \hat{\boldsymbol{\chi}}_k^{-1} \mathbf{y}_k - \mathbf{d}_k$ 
14:   $\mathbf{z}_k^i \leftarrow \mathbf{z}_k$ 
15:   $\hat{\mathbf{N}}_k \leftarrow \hat{\boldsymbol{\chi}}_k^{-1} \mathbf{N}_k (\hat{\boldsymbol{\chi}}_k^{-1})^T$ 
16:  while  $\Delta \boldsymbol{\xi}_k^i > \text{tol}$  and  $i < N_{\max}$  do
17:     $\mathbf{H}_k^i \leftarrow \exp_G \left( \boldsymbol{\xi}_k^i \right) \mathbf{H}_k \mathcal{J}_r \left( \boldsymbol{\xi}_k^i \right)$ 
18:     $\mathbf{K}_k^i \leftarrow \mathbf{P}_k (\mathbf{H}_k^i)^T (\mathbf{H}_k^i \mathbf{P}_k (\mathbf{H}_k^i)^T + \hat{\mathbf{N}}_k)^{-1}$ 
19:     $\mathbf{z}_k^i \leftarrow \mathbf{z}_k - \exp_G \left( \boldsymbol{\xi}_k^i \right) \mathbf{d}_k + \mathbf{d}_k + \mathbf{H}_k^i \boldsymbol{\xi}_k^i$ 
20:     $\Delta \boldsymbol{\xi}_k^i \leftarrow \|\mathbf{K}_k^i \mathbf{z}_k^i - \boldsymbol{\xi}_k^i\|$ 
21:     $\boldsymbol{\xi}_k^i \leftarrow \mathbf{K}_k^i \mathbf{z}_k^i$ 
22:     $i \leftarrow i + 1$ 
23:  end while
24:   $\hat{\boldsymbol{\chi}}_k^+ \leftarrow \hat{\boldsymbol{\chi}}_k \exp_G \left( \boldsymbol{\xi}_k^i \right)$ 
25:   $\mathbf{K}_k \leftarrow \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \hat{\mathbf{N}}_k)^{-1}$ 
26:   $\mathbf{P}_k^+ \leftarrow (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k$ 
27:
28:  {Propagation}
29:   $\hat{\boldsymbol{\chi}}_{k+1} \leftarrow \mathbf{f}(\hat{\boldsymbol{\chi}}_k^+, \mathbf{u}_k)$ 
30:   $\mathbf{P}_{k+1} \leftarrow \mathbf{F}_k \mathbf{P}_k^+ \mathbf{F}_k^T + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T$ 
31: end loop

```

---

## 5 Theoretical properties of the IIEKF

Consider the left-invariant measurement (24b). Deriving properties in the presence of measurement noise seems out of reach, as is often the case in nonlinear filtering (e.g., the log-linear error property of [13] holds in the absence of process noise). However, in the degenerate case where the magnitude of the measurement noise tends to zero (whereas the state estimate and the propagation may be noisy), we will show the desirable properties of the linear case, outlined in Subsection 3.2, hold. This provides strong guarantees when using an IIEKF with noise-free or pseudo measurements, as in the application below. Besides, as there is a *continuum* between the noise-free and the noisy cases, the properties approximately hold in the case of small measurement noise.

## 5.1 The IIEKF in the face of noise-free measurements

To derive mathematical properties, we need to understand how the proposed IIEKF behaves when measurement noise vanishes. Let  $\tilde{\mathbf{N}}_k = \delta \mathbf{I}$ , with  $\delta \ll 1$ . In this context, Problem (31) becomes that of minimizing  $f_\delta(\boldsymbol{\xi}) = \frac{1}{2} \|\boldsymbol{\xi}\|_{\mathbf{P}_k}^2 + \frac{1}{2} \|\mathbf{z}_k - \exp_G(\boldsymbol{\xi}) \mathbf{d}_k + \mathbf{d}_k\|_{\delta \mathbf{I}}^2$ . The minimizer of  $f_\delta$  may be sought using the GN sequence of estimates of Proposition 2

$$\boldsymbol{\xi}_{k,\delta}^{i+1} = \mathbf{K}_{k,\delta}^i \cdot \underbrace{(\mathbf{z}_k - \exp_G(\boldsymbol{\xi}_{k,\delta}^i) \mathbf{d}_k + \mathbf{d}_k + \mathbf{H}_{k,\delta}^i \boldsymbol{\xi}_{k,\delta}^i)}_{\mathbf{z}_{k,\delta}^i}, \quad (40)$$

where  $\mathbf{H}_{k,\delta}^i = \exp_G(\boldsymbol{\xi}_{k,\delta}^i) \mathbf{H}_k \mathcal{J}_T(\boldsymbol{\xi}_{k,\delta}^i)$  and where we let  $\mathbf{K}_{k,\delta}^i = \mathbf{P}_k (\mathbf{H}_{k,\delta}^i)^T (\mathbf{H}_{k,\delta}^i \mathbf{P}_k (\mathbf{H}_{k,\delta}^i)^T + \delta \mathbf{I})^{-1}$ . Starting from  $\boldsymbol{\xi}_{k,\delta}^0 = \mathbf{0}$  and taking the limit for  $\delta \rightarrow 0$ , we get

$$\lim_{\delta \rightarrow 0} \boldsymbol{\xi}_{k,\delta}^{i+1} = \lim_{\delta \rightarrow 0} \mathbf{K}_{k,\delta}^i \cdot \lim_{\delta \rightarrow 0} \mathbf{z}_{k,\delta}^i, \quad (41)$$

so that the GN sequence approximating the solution to the noise-free optimization problem becomes

$$\boldsymbol{\xi}_k^{i+1} = \mathbf{L}_k (\mathbf{H}_k^i \mathbf{L}_k)^{\dagger} \mathbf{z}_k^i, \quad (42)$$

recalling (14). This provides the following limit IIEKF for noise-free measurements, described in Algorithm 2.

---

**Algorithm 2** The iterated invariant extended Kalman filter (IIEKF) with noise-free measurements

---

Same as Algorithm 1 with the following exceptions.

Line 18:  $\mathbf{K}_k^i \leftarrow \mathbf{L}_k (\mathbf{H}_k^i \mathbf{L}_k)^{\dagger}$

Line 25:  $\mathbf{K}_k \leftarrow \mathbf{L}_k (\mathbf{H}_k \mathbf{L}_k)^{\dagger}$

Line 26:  $\mathbf{P}_k^+ \leftarrow (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k$

---

This algorithm is not in our preliminary conference paper [30], where we derived an heuristic noise-free IEKF instead.

## 5.2 The IIEKF satisfies the compatibility property

Having derived the limit-case IIEKF, we may prove it satisfies the desirable Property 1 enunciated above.

### 5.2.1 First part (the estimate lands on the right subspace)

We start by discussing the first part of the property, namely desirable Equation (22). As measurement noise tends to  $\mathbf{0}$ , Problem (31) becomes that of finding the smallest  $\boldsymbol{\xi}$  in the sense of the metric induced by  $\mathbf{P}_k$  that satisfies the hard constraint  $\|\mathbf{z}_k - \exp_G(\boldsymbol{\xi}) \mathbf{d}_k + \mathbf{d}_k\|^2 = 0$ , which boils down to  $\hat{\boldsymbol{\chi}}_k^{-1} \mathbf{y}_k = \exp_G(\boldsymbol{\xi}) \mathbf{d}_k$ , see (28b). We thus have  $\hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi}_k^*) \mathbf{d}_k = \mathbf{y}_k$ , where  $\boldsymbol{\xi}_k^*$  is the argmin we search for. As long as IIEKF's iterate  $\boldsymbol{\xi}_k^i$  converges to it, the updated state  $\hat{\boldsymbol{\chi}}_k^+ = \hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi}_k^*)$  belongs to the observed set indeed, that is,  $\hat{\boldsymbol{\chi}}_k^+ \in \mathcal{S}_{\{\boldsymbol{\chi} \mathbf{d}_k = \mathbf{y}_k\}}$ .

Of course, this assumes the GN descent converges to a global minimizer. As the optimization problem is not convex, the GN sequence (42) converges towards the global minimizer provided that the true error  $\boldsymbol{\xi}_k$  is close enough to  $\mathbf{0}$ . It is important to note that a non-iterative update scheme, like the IEKF update, has no chance to satisfy  $\hat{\boldsymbol{\chi}}_k^+ \in \mathcal{S}_{\{\boldsymbol{\chi} \mathbf{d}_k = \mathbf{y}_k\}}$ .

### 5.2.2 Second part (the covariance correctly encodes the uncertainty drop)

The second part of Property 1, namely desirable Equation (23), is proved by the following result.

**Theorem 1.** *For noise-free measurements expressed in the left-invariant (or right-invariant) form, that is,  $\mathbf{y}_k = \boldsymbol{\chi}_k \mathbf{d}_k$  (or  $\mathbf{y}_k = \boldsymbol{\chi}_k^{-1} \mathbf{d}_k$ ), the invariant framework naturally ensures  $\mathbf{H}_k^+ \mathbf{P}_k^+ (\mathbf{H}_k^+)^T = \mathbf{0}$ , and also  $\mathbf{H}_k \mathbf{P}_k^+ (\mathbf{H}_k)^T = \mathbf{0}$ .*

*Proof.* From the theory of invariant filtering, left- or right-invariant measurements yield a Jacobian  $\mathbf{H}_k$  that is independent from the current estimate, see [14], meaning that  $\mathbf{H}_k = \mathbf{H}_k^+$ , which is the key property. This property is reminiscent of the linear Kalman filter, so that the same results hold. Either 1)  $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T$  is invertible, and thus  $\mathbf{H}_k \mathbf{P}_k^+ \mathbf{H}_k^T = \mathbf{H}_k ((\mathbf{I} - \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T)^{-1} \mathbf{H}_k) \mathbf{P}_k) \mathbf{H}_k^T = \mathbf{0}$  or 2) the same calculation (67) as in the linear case then holds.  $\square$

### 5.3 The IIEKF's belief is strongly compatible

Before we turn to desirable Property 2, we would like to underline that compatibility of the belief and the observed set is in fact very strong, as illustrated by the following result.

**Proposition 3.** *Under the compatibility assumptions of Definition 2, the entire probability distribution encoded by model (26) is in fact contained within the observed set. Mathematically, supposing the belief is compatible with noise-free measurement  $\mathbf{y}_k = \boldsymbol{\chi}\mathbf{d}_k$ , in the sense that*

$$\hat{\boldsymbol{\chi}}_k \in \mathcal{S}_{\{\boldsymbol{\chi}\mathbf{d}_k=\mathbf{y}_k\}} \quad \text{and} \quad \mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^T = \mathbf{0}, \quad (43)$$

with  $\mathbf{H}_k$  the Jacobian associated to  $\mathbf{d}_k$ , implies that for any  $\boldsymbol{\xi}$  drawn from  $\mathcal{N}(\mathbf{0}, \mathbf{P}_k)$ , we have  $\hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi}) \in \mathcal{S}_{\{\boldsymbol{\chi}\mathbf{d}_k=\mathbf{y}_k\}}$ .

The proof consists in noting that then  $\boldsymbol{\xi}$  is (almost surely) in the column space of  $\mathbf{P}_k$ , hence  $\mathbf{H}_k\boldsymbol{\xi} = \mathbf{0}$ , and in using a key result from the invariant filtering theory [35, 36].

**Lemma 3.** *Let  $\mathbf{H}_k$  denote the Jacobian from invariant filtering associated to map  $\mathbf{h}_k(\boldsymbol{\chi}) = \boldsymbol{\chi}\mathbf{d}_k$ . We have*

$$\begin{aligned} \hat{\boldsymbol{\chi}}_k \in \mathcal{S}_{\{\boldsymbol{\chi}\mathbf{d}_k=\mathbf{y}_k\}} \quad \text{and} \quad \mathbf{H}_k\boldsymbol{\xi} = \mathbf{0}, \quad \boldsymbol{\xi} \in \mathbb{R}^n \\ \Downarrow \\ \hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi}) \in \mathcal{S}_{\{\boldsymbol{\chi}\mathbf{d}_k=\mathbf{y}_k\}}. \end{aligned} \quad (44)$$

*Proof.* Let  $\mathcal{L}_{\mathfrak{g}}(\cdot) : \mathbb{R}^n \rightarrow \mathfrak{g}$  be the bijective map between the vector space  $\mathbb{R}^n$  and the Lie algebra  $\mathfrak{g}$  of group  $G$ , see [13]. As we are dealing with matrix Lie groups, the exponential coincides with the matrix exponential in the following sense

$$\exp_G(\boldsymbol{\xi}) = \mathbf{I} + \sum_{l=1}^{+\infty} \frac{1}{l!} \mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})^l. \quad (45)$$

Keeping only first-order terms proves, in passing, that Jacobian  $\mathbf{H}_k$  from (28b) writes

$$\mathbf{H}_k\boldsymbol{\xi} = \mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})\mathbf{d}_k. \quad (46)$$

To prove the result we seek, it then suffices to write

$$\hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi})\mathbf{d}_k = \hat{\boldsymbol{\chi}}_k \left( \mathbf{I} + \sum_{l=1}^{+\infty} \frac{1}{l!} \mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})^l \right) \mathbf{d}_k, \quad (47a)$$

$$= \hat{\boldsymbol{\chi}}_k\mathbf{d}_k + \hat{\boldsymbol{\chi}}_k \sum_{l=1}^{+\infty} \frac{1}{l!} \mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})^l \mathbf{d}_k, \quad (47b)$$

$$= \hat{\boldsymbol{\chi}}_k\mathbf{d}_k + \hat{\boldsymbol{\chi}}_k \sum_{l=1}^{+\infty} \frac{1}{l!} \mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})^{l-1} (\mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})\mathbf{d}_k). \quad (47c)$$

Since  $\mathbf{H}_k\boldsymbol{\xi} = \mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})\mathbf{d}_k = \mathbf{0}$  and  $\hat{\boldsymbol{\chi}}_k \in \mathcal{S}_{\{\boldsymbol{\chi}\mathbf{d}_k=\mathbf{y}_k\}}$ , the last line gives

$$\hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi})\mathbf{d}_k = \hat{\boldsymbol{\chi}}_k\mathbf{d}_k = \mathbf{y}_k, \quad (48)$$

which proves  $\hat{\boldsymbol{\chi}}_k \exp_G(\boldsymbol{\xi}) \in \mathcal{S}_{\{\boldsymbol{\chi}\mathbf{d}_k=\mathbf{y}_k\}}$ .  $\square$

The result of Proposition 3 shows that the local information encoded by the IIEKF is in fact global, exactly as in the linear case, where a Gaussian vector with rank-deficient covariance matrix has a distribution that is entirely contained within a linear subspace.

### 5.4 The IIEKF satisfies the “hard encoding” property

We now have what we need to actually show that the IIEKF possesses desirable Property 2: the filter inherently “hard codes” deterministic information.

**Theorem 2.** Consider a (left) IIEKF whose belief  $(\hat{\mathbf{X}}_k, \mathbf{P}_k)$  is compatible in the sense of Definition 2 with measurement (24b) with noise turned off, i.e.,  $\mathbf{y}_k = \mathbf{X}_k \mathbf{d}_k$ . This means  $\hat{\mathbf{X}}_k \in \mathcal{S}_{\{\mathcal{X}_{\mathbf{d}_k=\mathbf{y}_k}\}}$  and  $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}$  with  $\mathbf{H}_k$  the Jacobian associated with  $\mathbf{d}_k$ . Then, if we immediately make another measurement  $\tilde{\mathbf{y}}_k$ , no matter its expression, the updated belief  $(\hat{\mathbf{X}}_k^+, \mathbf{P}_k^+)$  corresponding to measurement  $\tilde{\mathbf{y}}_k$  remains compatible with the (first) noise-free measurement. Mathematically,

$$\begin{aligned} \hat{\mathbf{X}}_k \in \mathcal{S}_{\{\mathcal{X}_{\mathbf{d}_k=\mathbf{y}_k}\}} \quad & \text{and} \quad \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}, \\ & \Downarrow \\ \hat{\mathbf{X}}_k^+ \in \mathcal{S}_{\{\mathcal{X}_{\mathbf{d}_k=\mathbf{y}_k}\}} \quad & \text{and} \quad \mathbf{H}_k \mathbf{P}_k^+ \mathbf{H}_k^T = \mathbf{0}. \end{aligned} \quad (49)$$

*Proof.* Point 2) of Definition 2 is the easiest to prove. If we have  $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}$ , by symmetry it means  $\mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}$ . Any immediately subsequent update of the IIEKF, see line 26 of Algorithm 1, writes  $\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k \tilde{\mathbf{H}}_k) \mathbf{P}_k$ , hence  $\mathbf{P}_k^+$  verifies the same property whatever  $\tilde{\mathbf{H}}_k$ . And we have also  $\mathbf{H}_k^+ \mathbf{P}_k^+ \mathbf{H}_k^{+T} = \mathbf{0}$  owing to  $\mathbf{H}_k = \mathbf{H}_k^+$ .

Regarding part 1), we see at line 24 of Algorithm 1 that the update writes  $\hat{\mathbf{X}}_k^+ = \hat{\mathbf{X}}_k \exp_G(\boldsymbol{\xi}_k^i)$ . Using Lemma 3, we see that provided  $\boldsymbol{\xi}_k^i$  satisfies  $\mathbf{H}_k \boldsymbol{\xi}_k^i = \mathbf{0}$ , we have  $\hat{\mathbf{X}}_k^+ \in \mathcal{S}_{\{\mathcal{X}_{\mathbf{d}_k=\mathbf{y}_k}\}}$ , since we assumed  $\hat{\mathbf{X}}_k \in \mathcal{S}_{\{\mathcal{X}_{\mathbf{d}_k=\mathbf{y}_k}\}}$ . This follows from the fact that in the algorithm we have  $\boldsymbol{\xi}_k^i = \mathbf{K}_k^i \mathbf{z}_k^i = \mathbf{P}_k \dots \mathbf{z}_k^i$ , hence  $\mathbf{H}_k \boldsymbol{\xi}_k^i = \mathbf{0}$ , as by assumption we have  $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}$ , implying by symmetry  $\mathbf{H}_k \mathbf{P}_k = \mathbf{0}$ .  $\square$

## 6 Application: solving a system of equations on a matrix Lie group

In this section, we leverage the latter results to provide a novel algorithm to solve a problem of subset intersection in Lie groups, that possesses some mathematical guarantees. As a byproduct, we believe it is enlightening regarding the proposed IIEKF theory, and may help understanding its main features.

### 6.1 The linear case

To fix ideas, suppose we have a system of linear equations, or its equivalent form using row matrices

$$\begin{cases} 3x + 5y + z = 3, \\ 7x - 2y + 4z = 4, \\ -6x + 3y + 2z = 2, \end{cases} \quad \Leftrightarrow \quad \begin{cases} \mathbf{H}_1 \mathbf{x} = \mathbf{y}_1, \\ \mathbf{H}_2 \mathbf{x} = \mathbf{y}_2, \\ \mathbf{H}_3 \mathbf{x} = \mathbf{y}_3. \end{cases} \quad (50)$$

The standard Kalman filter (KF) may naturally be brought to bear on this system to find a solution, provided it exists.

**Proposition 4.** Suppose the row matrices  $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3$  are independent. Initialize a KF in  $\mathbb{R}^3$  with arbitrary belief  $(\mathbf{x}, \mathbf{P})$ , letting  $\mathbf{P}$  be full-rank. By consecutively feeding the KF with noise-free measurements  $\mathbf{y}_1, \mathbf{y}_2$  and  $\mathbf{y}_3$ , in any order, the final update lies in the intersection of the hyperplanes, that is, solves the system of linear equations.

*Proof.* Let us apply the Kalman update (3d), with null noise, i.e.,  $\mathbf{N}_k = \mathbf{0}$ . As  $\mathbf{P}$  is full-rank, we have  $\mathbf{H}_1 \mathbf{P} \mathbf{H}_1^T \neq 0$ . Hence we easily see the updated state in the light of  $\mathbf{y}_1$  satisfies  $\mathbf{H}_1 \mathbf{x}^+ = \mathbf{y}_1$ , no matter  $\mathbf{x}$ . Repeating the process, we find  $\mathbf{H}_2 (\mathbf{x}^+)^+ = \mathbf{y}_2$ , as  $\mathbf{H}_2 \mathbf{P}^+ \mathbf{H}_2^T \neq 0$  since after the first update the rank of  $\mathbf{P}$  has dropped by one dimension, along the column space of  $\mathbf{H}_1$ , which differs from that of  $\mathbf{H}_2$ . Similarly, we have  $\mathbf{H}_3 ((\mathbf{x}^+)^+)^+ = \mathbf{y}_3$ . A crucial point to conclude the proof, is to show the final update  $((\mathbf{x}^+)^+)^+$  satisfies the first two equations. This is the case, as the first update ensures  $\mathbf{H}_1 \mathbf{P}^+ \mathbf{H}_1^T = 0$ , and this cannot be undone by any subsequent updates, hence  $\mathbf{H}_1 \mathbf{P}$  remains null over the following measurements, hence so does  $\mathbf{H}_1 \mathbf{K}$ , ensuring further corrections preserve the observed set  $\mathbf{H}_1 \mathbf{x} = \mathbf{y}_1$ .  $\square$

The result is illustrated in 2D in Fig. 1. Note that, we used three equations for three unknowns. We could have very well considered an undetermined system, with less equations than unknowns. The method would then yield one solution, lying in the appropriate intersection of hyperplanes.

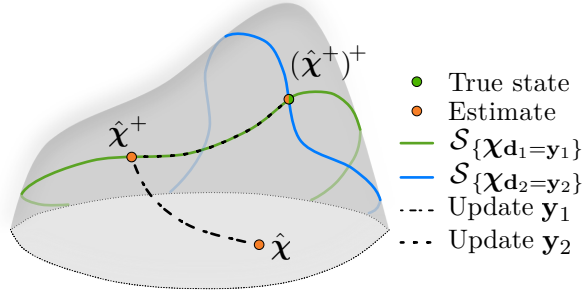


Figure 4: Graphical representation of two consecutive noise-free updates for an IIEKF. The filter is first updated with measurement  $\mathbf{y}_1$ , and then with measurement  $\mathbf{y}_2$ .

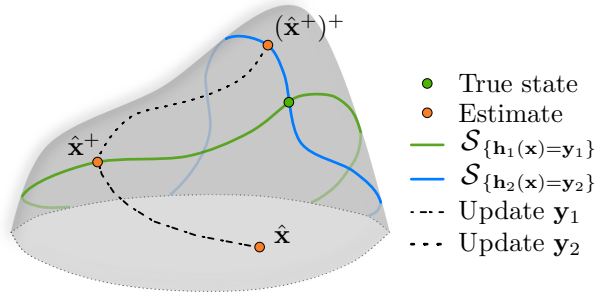


Figure 5: Graphical representation of two consecutive noise-free updates for a generic iterated EKF on Lie groups or manifolds. The filter is first updated with measurement  $\mathbf{y}_1$ , and then with measurement  $\mathbf{y}_2$ .

## 6.2 The Lie group case

Assume we have a system of equations on a Lie group

$$\begin{cases} \chi \mathbf{d}_1 = \mathbf{y}_1, \\ \chi \mathbf{d}_2 = \mathbf{y}_2, \end{cases} \quad (51)$$

where we limit ourselves to two equations for simplicity, but more may be added. Finding a solution amounts to finding a point  $\chi^*$  in the intersection of the associated observed sets, see Definition 1. The observed sets are of course nonlinear submanifolds (the equations could spuriously look linear, but  $\chi$  lives in a nonlinear subspace  $G$  of square matrices).

We see the IIEKF introduced in this paper provides a novel method to address Problem (51), and its extensions to more equations of the same form. Indeed, one can feed it with noise-free measurements, each corresponding to one of the equations above as we did in the linear case in the latter subsection. As long as the first optimization Gauss-Newton descent converges to the true minimum indeed, the first noise-free measurement is incorporated so that the estimate lands into the desired observed set; the rank of  $\mathbf{P}$  drops by one, and  $\mathbf{P}^+$  becomes compatible with the first measurement, see Section 5.2. Using Proposition 2, we see that all subsequent updates remain in this very observed set. As long as the intersection exists, this enables the IIEKF's final updated estimate to land within the intersection, and this whatever the order in which the measurements are given, provided each GN method succeeds in converging. This is schematically illustrated in Fig. 4.

By contrast, a generic iterated EKF (on Lie groups), would probably fail to converge to the intersection in one cycle over the measurements. Indeed, even if we assume each GN update converges to the true minimum, ensuring the first updated estimate lies in the first subspace, then, when the second measurement is fed to the EKF, its estimate will step out of the first observed set, making the second update satisfy the second equation, but not the first anymore. This is illustrated in Fig. 5.

**Remark 1.** When confronted with a system of equations in the right-invariant form, i.e., each of the form  $\chi^{-1} \mathbf{d} = \mathbf{y}$ , we may rewrite them as  $\mathbf{d} = \chi \mathbf{y}$ , and the same applies.

### 6.3 Illustration on $SO(3)$

Albeit a simple example, it is pedagogical to illustrate the results on a problem involving rotation matrices. Consider (51), where  $\chi \in SO(3)$ , and where  $\mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^3$  are noncolinear unitary vectors, and so are  $\mathbf{y}_1, \mathbf{y}_2$ . Assume we start with  $\hat{\chi} \in SO(3)$  close enough to the solution and  $\mathbf{P}$  invertible, and we feed the IIEKF with noise-free measurement  $\chi \mathbf{d}_1 = \mathbf{y}_1$ . The IIEKF will output an estimate  $\hat{\chi}^+$  which satisfies  $\hat{\chi}^+ \mathbf{d}_1 = \mathbf{y}_1$ . Then, if we feed the IIEKF with noise-free measurement  $\chi \mathbf{d}_2 = \mathbf{y}_2$ , all the iterates  $\chi^+ \exp_{SO(3)}(\xi^i)$  underlying the IIEKF will remain in the set defined by  $\chi \mathbf{d}_1 = \mathbf{y}_1$ , as guaranteed by Proposition 2. Provided the GN converges, they will converge to the solution of the system of equations  $(\hat{\chi}^+)^+$ . We may redo the calculations in this simple case, to gain some geometric insight.

**Proposition 5.** *Consider an IIEKF with  $\chi \in SO(3)$  and full-rank initial covariance matrix. Assume it is fed with measurement  $\chi \mathbf{d}_1 = \mathbf{y}_1$ , and that the GN manages to converge, so that  $\hat{\chi}^+ \mathbf{d}_1 = \mathbf{y}_1$ . Then, all immediate subsequent updates (or iterates) of the IIEKF consist of 2D rotations around the fixed axis  $\mathbf{d}_1$ , composed with the particular rotation  $\hat{\chi}^+$ .*

*Proof.* This is direct consequence of (49). Indeed, in the present case  $\exp_{SO(3)}(\delta) \mathbf{d}_1 \approx \mathbf{d}_1 + \mathbf{H}_1 \delta$  and furthermore from the definition of the exponential  $\exp_{SO(3)}(\delta) \mathbf{d}_1 \approx \mathbf{d}_1 + \delta \times \mathbf{d}_1$ , so that  $\mathbf{H}_1 \delta = \delta \times \mathbf{d}_1$ . The Riccati update ensures  $\mathbf{H}_1 \mathbf{P}^+ \mathbf{H}_1^T = \mathbf{0}$ , see Prop. 1, and the iterates are of the form  $\hat{\chi}^i = \hat{\chi}^+ \exp_{SO(3)}(\xi^i)$ , with  $\xi^i \in \text{Im } \mathbf{P}^+$ , see (35). Thus  $\mathbf{H}_1 \xi^i = \mathbf{0}$  and thus  $\xi^i \times \mathbf{d}_1 = \mathbf{0}$ . This immediately proves that  $\exp_{SO(3)}(\xi^i)$  is a 2D rotation around  $\mathbf{d}_1$ , as  $\exp_{SO(3)}(\xi^i) \mathbf{d}_1 = \sum_{k \geq 0} [(\xi^i)]_{\times}^k \mathbf{d}_1 = \mathbf{d}_1 + \sum_{k \geq 1} [(\xi^i)]_{\times}^{k-1} (\xi^i \times \mathbf{d}_1) = \mathbf{d}_1 + \mathbf{0}$ .  $\square$

This means that as soon as the first measurement is made, the IIEKF “focuses” on a problem of reduced dimension, namely seeking planar rotations in a 2D space. This makes it much more efficient as 1) it performs updates directly in a reduced subspace where the state is, and 2) it avoids destroying the information contained in the first measurement. The way noise-free measurements inherently remove degrees of freedom one by one in the estimation problem for the IIEKF certainly explains its fast convergence in the numerical experiments of the next section, although on a different problem.

## 7 Application of engineering interest

It is technically very feasible nowadays to mount an inertial measurement unit (IMU) on the hook of a crane that transmits its sensor readings [37], see illustration of Fig. 6. Estimating the position and velocity of the load from the IMU measurements may open the door to new automation capabilities since it allows for feedback, and may also increase safety [38]. However, the underlying dynamical model is unknown, as the hook does not necessarily follow a simple pendulum motion, because of external forces like friction and wind. The IMU dynamics, though, consist of kinematic relations which are satisfied at all times, whatever the motion. Moreover, when the cable is hanging, the distance between the hook and the cable hang-up point is the cable length, which is very accurately measured in modern cranes for safety reasons, see, e.g., [38]. We will use this as a very accurate–virtually perfect–measurement to gain information on the position of the hook.

We propose to compare in simulations the EKF, the iterated EKF (IterEKF) [3], the IEKF, and the proposed IIEKF on this estimation problem.

### 7.1 Dynamical system

Since they are rigidly attached, we assume the hook and the IMU share the same body frame. The state of the IMU, i.e., its “extended pose”, see [39], consists of: the rotation matrix  $\mathbf{R}_k \in SO(3)$  from the IMU to the world frame, the IMU velocity vector  $\mathbf{v}_k \in \mathbb{R}^3$  expressed in the world frame, and the IMU position vector  $\mathbf{p}_k \in \mathbb{R}^3$  expressed in the world frame. In this paper, the EKF and IterEKF estimate the state

$$\mathbf{x}_k = (\mathbf{R}_k, \mathbf{v}_k, \mathbf{p}_k) \in SO(3) \times \mathbb{R}^3 \times \mathbb{R}^3, \quad (52)$$

and use the estimation error defined as

$$\mathbf{e}_k = \begin{bmatrix} \log_{SO(3)}(\hat{\mathbf{R}}_k^{-1} \mathbf{R}_k) \\ \mathbf{v}_k - \hat{\mathbf{v}}_k \\ \mathbf{p}_k - \hat{\mathbf{p}}_k \end{bmatrix}, \quad (53)$$

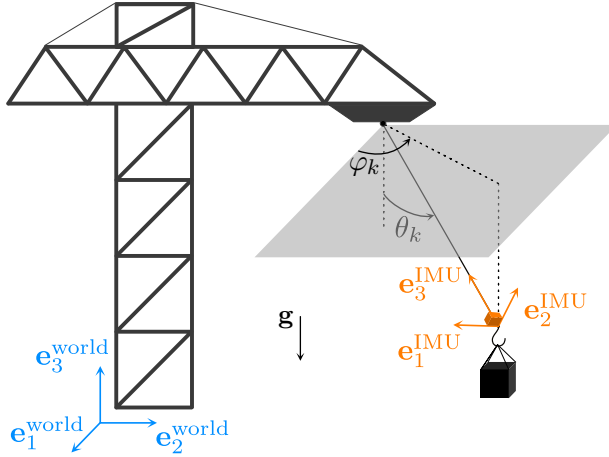


Figure 6: Crane with a wireless IMU mounted on its hook.

where  $\log_{SO(3)} : SO(3) \rightarrow \mathbb{R}^3$  is the logarithm map of group  $SO(3)$ . Similarly, we denote by  $\mathbf{e}_k^+$  the error achieved by the updated estimate  $\hat{\mathbf{x}}_k^+$ . The IEKF and IIEKF estimate the state

$$\boldsymbol{\chi}_k = \begin{bmatrix} \mathbf{R}_k & \mathbf{v}_k & \mathbf{p}_k \\ \mathbf{0} & 1 & 0 \\ \mathbf{0} & 0 & 1 \end{bmatrix} \in SE_2(3), \quad (54)$$

where  $SE_2(3)$  is the group of extended poses introduced in [13], see also [39]. Since measurements are expressed in the world frame, we use the left-invariant formalism, so that

$$\boldsymbol{\xi}_k := \log_{SE_2(3)}(\hat{\boldsymbol{\chi}}_k^{-1} \boldsymbol{\chi}_k), \quad (55)$$

where  $\log_{SE_2(3)} : SE_2(3) \rightarrow \mathbb{R}^9$  denotes the logarithm map of group  $SE_2(3)$ , see [14].

Without any prior knowledge about the motion, the dynamics of an IMU on flat Earth write

$$\mathbf{R}_{k+1} = \mathbf{R}_k \exp_{SO(3)}((\boldsymbol{\omega}_k + \mathbf{w}_k^\omega) dt), \quad (56a)$$

$$\mathbf{v}_{k+1} = \mathbf{v}_k + (\mathbf{R}_k(\mathbf{a}_k + \mathbf{w}_k^a) + \mathbf{g}) dt, \quad (56b)$$

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{v}_k dt, \quad (56c)$$

where the model inputs  $\boldsymbol{\omega}_k, \mathbf{a}_k \in \mathbb{R}^3$  are the angular velocity and the linear acceleration recorded by the IMU. They are both affected by normal noise  $(\mathbf{w}_k^\omega, \mathbf{w}_k^a) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$  and are assumed to be free of any bias. Vector  $\mathbf{g} = (0, 0, -9.81)$  is the gravity vector expressed in the world frame and  $dt$  is the integration time step. As shown in [13], the dynamics (56) are group affine, see [26] in discrete time. The Jacobians of the dynamics for each filter are available in Appendix C.3.

## 7.2 Measurements

The measurement  $\mathbf{y}_k$  is the position of the hang-up point of the crane cable expressed in the world frame. It can be modeled using the function

$$\mathbf{y}_k = \mathbf{R}_k \mathbf{r}_k + \mathbf{p}_k + \mathbf{n}_k, \quad (57)$$

where  $\mathbf{r}_k = (0, 0, l_k)$  and  $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{N})$ . This model assumes the cable length  $l_k$  to be known as well as the position  $\mathbf{y}_k$ . In modern cranes, they are both accurately measured via the encoders of the three motors of the crane [38]. The low noise inherent to those measurements is encapsulated within vector  $\mathbf{n}_k$ .

In the left-invariant framework, the model (57) writes

$$\begin{bmatrix} \mathbf{y}_k \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{\chi}_k \mathbf{d}_k + \begin{bmatrix} \mathbf{n}_k \\ 0 \\ 0 \end{bmatrix}, \quad (58)$$

where  $\mathbf{d}_k = (0, 0, l_k, 0, 1)$  and with  $\boldsymbol{\chi}_k$  as in (54). Note that the last two rows do not bring any useful information so that it is equivalent to (57). The Jacobians of the measurement function for each filter are available in Appendix C.3.

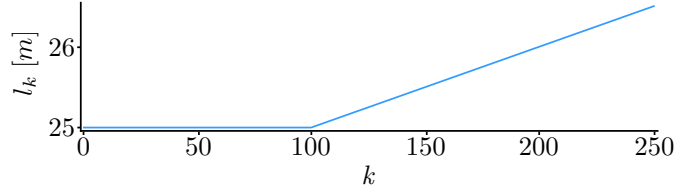


Figure 7: Evolution of the length of the crane cable  $l_k$  as a function of the time index, used in simulations.

### 7.3 Simulations

We consider three different scenarios. It is important to note that in all of them, process noise is present and non-negligible. For each of them, we run  $N_s = 200$  simulations in which the initial error is drawn randomly from the initial distribution. The performances are assessed by comparing the mean invariant error norm, that is,

$$E(\|\boldsymbol{\xi}_k^+\|) = \frac{1}{N_s} \sum_{n=1}^{N_s} \left\| \log_{SE_2(3)} \left( (\hat{\boldsymbol{\chi}}_{k,n}^+)^{-1} \boldsymbol{\chi}_k \right) \right\|, \quad (59)$$

where  $\hat{\boldsymbol{\chi}}_{k,n}^+$  is the updated estimate at time index  $k$  in the  $n^{\text{th}}$  simulation. The three different scenarios are as follows.

#### 7.3.1 Scenario 1

The ground-truth simulation corresponds to the trajectory followed by the hook with initial configuration:

$$\begin{aligned} \varphi_0 &= 0 \text{ rad}, & \theta_0 &= \frac{\pi}{4} \text{ rad}, \\ \dot{\varphi}_0 &= 1 \text{ rad/s}, & \dot{\theta}_0 &= 0 \text{ rad/s}, \end{aligned} \quad (60)$$

where  $\varphi_k$  and  $\theta_k$  are the azimuth and polar angles, see Fig. 6. No other external force than gravity is applied to the system. It is important to note that, in this configuration, the system is not observable, as shown in Appendix C.1. In addition to the average invariant error norm, we compute the average normalized estimation error squared (ANEES), denoted  $\bar{\epsilon}_k$ , in order to assess the consistency between the estimation error and the associated covariance matrix, see Appendix C.2 for more details. The trajectory of the hook is simulated for a total duration of 2.5 s. The length of the cable follows the profile shown in Fig. 7. The IMU and estimation filters operate at the same frequency of 100 Hz. The maximum number of iteration of the IterEKF and IIEKF is set to  $N_{\max} = 50$ . The gyroscope and accelerometer are assumed to be consumer-grade devices, respectively with a standard deviation of 0.017 rad/s and 0.1 m/s<sup>2</sup> along each axis, giving the covariance matrix

$$\mathbf{Q} = \begin{bmatrix} (0.017)^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & (0.1)^2 \mathbf{I} \end{bmatrix}. \quad (61)$$

The measurement noise is assumed to have a standard deviation of 1 m in every direction, giving  $\mathbf{N} = \mathbf{I}$ .

For a given simulation, all filters are initialized with the same initial estimation error. From one simulation to the next, the initial estimation error is randomly drawn from the initial distribution  $\mathcal{N}(\mathbf{0}, \mathbf{P}_0)$  with

$$\mathbf{P}_0 = \begin{bmatrix} (\frac{\pi}{6})^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (10)^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (10)^2 \mathbf{I} \end{bmatrix}. \quad (62)$$

The tolerance for the convergence of iterative algorithms is set to  $\text{tol} = 10^{-7}$ .

#### 7.3.2 Scenario 2 (restriction to 2D)

In this scenario, we assume the ground-truth trajectory is the trajectory followed by the hook having the following initial configuration:

$$\begin{aligned} \varphi_0 &= 0 \text{ rad}, & \theta_0 &= \frac{\pi}{4} \text{ rad}, \\ \dot{\varphi}_0 &= 0 \text{ rad/s}, & \dot{\theta}_0 &= 0 \text{ rad/s}. \end{aligned} \quad (63)$$



	Scenario 1	Scenario 2	Scenario 3
IterEKF	8.03	6.1	8.1
IIEKF	<b>6.2</b>	<b>4.75</b>	<b>3.15</b>

Table 1: Mean number of Gauss-Newton iterations needed by the IterEKF and IIEKF to converge in each scenario. The smallest number is highlighted in bold for each scenario.

We also assume that the gyroscope variance along the first and third axes is null, as well as the accelerometer variance along the second axis. In this way, the motion of the hook is contained within the plane  $\varphi_k = 0$ . In order to make the problem fully observable, we completely remove the uncertainty for rotations around the first and third axes, and for velocity and position around the second axis. All other parameters are kept identical to scenario 1.

### 7.3.3 Scenario 3 (restriction to 2D + noise-free measurements)

We now assume that the position of the cable hung-up point and the cable length are so accurately measured we may set  $\mathbf{N} = \mathbf{0}$ . All other parameters are kept identical to scenario 2. The system is therefore fully observable and measurements are noise-free. In accordance with the recommendations of Subsection 3.1, a regularized gain  $\mathbf{K}_k^{\text{rg}} = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \delta \mathbf{I})^{-1}$  is used for all filters, with  $\delta = 10^{-5}$ .

## 7.4 Results

Table 1 gathers the mean number of iterations needed for the GN method to converge in the update step of the IterEKF and IIEKF. In all three scenarios, the IIEKF requires less iterations. The biggest drop occurs in the third scenario, as could be anticipated from the insights of Section 6.3.

### 7.4.1 Scenario 1

The mean and standard deviation of the estimation error norm of each filter are displayed in graph (a) of Fig. 8. As expected for an unobservable system, the estimation error is lower bounded. The invariant methods outperform the EKF and IterEKF. A first factor contributing to this difference is the dependence on the state of the Jacobians used by the latter filters. When the estimate deviates significantly from the true trajectory, the Jacobians computed in such instances may differ completely from those calculated along the actual trajectory, leading to poor linearizations of the dynamics and output function.

A second factor is the problem of *false observability* inherent to the EKF: linearizing the output function (57) around the estimated trajectory may introduce spurious information that makes the system look observable. Graphs (b) and (c) of Fig. 8 show the evolution of the ANEES over time, which measures the ability of the filter to convey the actual extent of uncertainty associated to its estimates, see Appendix C.2 (the target value of the ANEES in the present example is 9). The EKF and IterEKF are overconfident about their estimation and their ANEES diverges. The IIEKF achieves an ANEES that almost entirely lies within the 95%-interval represented by the two horizontal dashed lines. This means it provides covariance matrices that are perfectly consistent with its estimation error, even in the transient. The IEKF initially under-estimates the estimation error before slowly recovering.

### 7.4.2 Scenario 2

The motion is now two-dimensional, so that the system is observable and the EKF and IterEKF do not suffer from false observability issues anymore. Fig. 9 shows that the IterEKF performs better than the EKF on average, with a smaller averaged error and a smaller variance across simulations. As in the previous scenario, the IEKF averaged error norm peaks after few time steps, before decreasing rapidly. The IIEKF exhibits the best performance: its estimation error is the smallest on average and presents the smallest variance across simulations. The IEKF and IIEKF have a much better convergence rate than the EKF and IterEKF. This is probably a direct consequence of the autonomy of their Jacobians w.r.t. the followed trajectory.

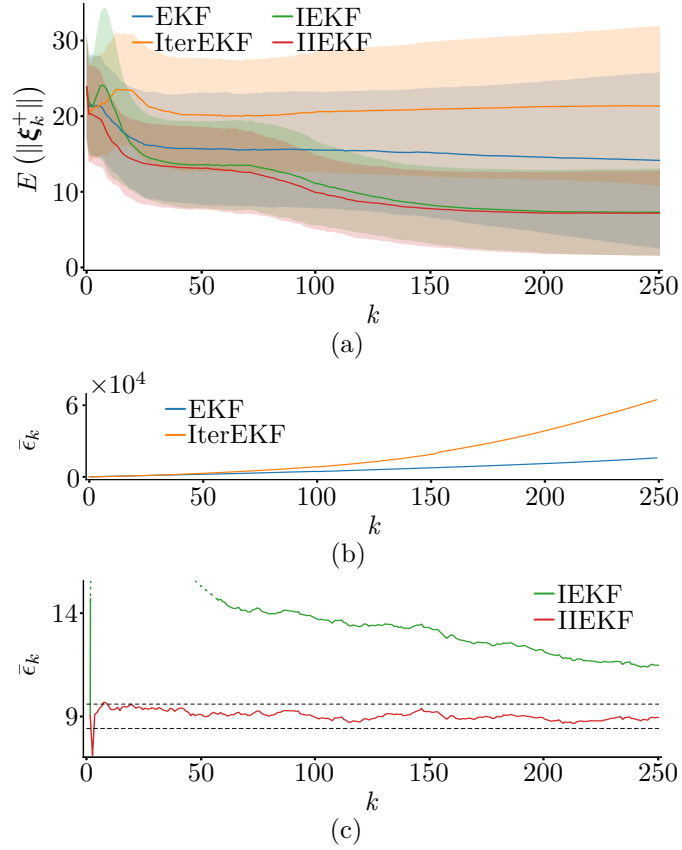


Figure 8: Results for scenario 1 (3D unobservable setting with process and measurement noises). (a) Mean and standard deviation (light color) of the estimation error norm as a function of the time index, computed over 200 simulations. (b) Evolution of the ANEES as a function of the time index over 200 simulations for the EKF and IterEKF. (c) Evolution of the ANEES as a function of the time index over 200 simulations for the IEKF and IIEKF. The theoretical value of the ANEES is  $N_{\text{dof}} = 9$ . The two dashed horizontal lines delimit a 95%-confidence interval for the ANEES.

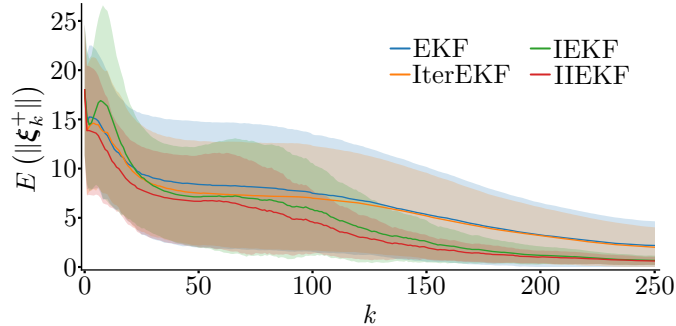


Figure 9: Results for scenario 2 (2D observable setting with noise). Mean and standard deviation (light color) of the estimation error norm as a function of the time index, computed over 200 simulations.

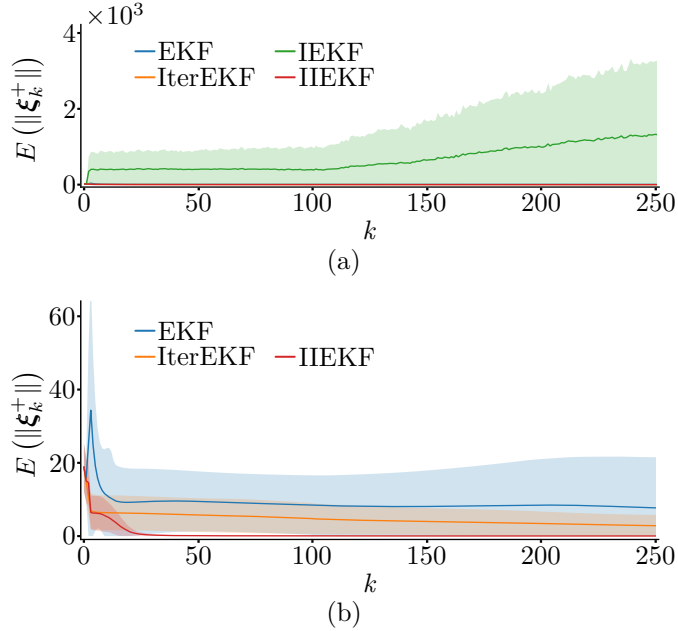


Figure 10: Results for scenario 3 (2D observable setting with process noise but noise-free measurements). (a) Mean and standard deviation (light color) of the estimation error norm as a function of the time index, computed over 200 simulations. (b) Same as (a) where the IEKF is removed.

### 7.4.3 Scenario 3

The measurements are now assumed to be noise-free and all four filters use the regularized gain (15) with  $\delta = 10^{-5}$ . The mean and standard deviation of the estimation error norm are displayed in Fig. 10. Surprisingly, the IEKF is unstable when the measurement noise is too low and diverges. As the IEKF, the EKF is not well-suited to handling noise-free measurements: its averaged error norm presents a high transient variation before decreasing with a low convergence rate. During the first time steps, both iterative filters achieve a significant drop in their averaged estimation error norm. Then, the estimation error of the IIEKF quickly drops to a value close to  $\mathbf{0}$  with almost no variance across simulations, while the error of the IterEKF slowly decreases towards zero with a relatively large variance across simulations. This scenario proves once more the superiority of the proposed IIEKF, and illustrates the theorems.

## 8 Conclusion

In this paper, we introduced the iterated invariant EKF. Its features and advantages, when applied to systems on Lie groups with left-invariant measurements, are as follows.

- In the absence of measurement noise, the IIEKF “hard codes” the measurement, inherently removing one degree of freedom in the estimation problem, making (immediate) subsequent updates more efficient.
- In the presence of low measurement noise, as there is a *continuum* between the noise-free and noisy case, this indicates the IIEKF incorporates the measurement by making the estimate land within a subspace where the state is very likely to be, and then performs updates that remain within (or close to) this subspace. This must explain its speed of convergence in the simulations.
- In the general case, whatever the level of noise, the IIEKF is anyway an improvement over the IEKF, which has been shown to be very well behaved in numerous applications. Indeed, the IIEKF leaves the IEKF’s Riccati update untouched, while refining its state estimate (the IEKF being exactly an IIEKF with one iteration).

In numerical experiments, the IIEKF showed unmatched performances in all presented scenarios while consistently estimating the error covariance. It was also shown to apply to the “novel” problem of solving a system of equations of the form  $\chi \mathbf{d} = \mathbf{y}$  on a matrix Lie group.

In terms of limitations, the IIEKF is reserved for systems with invariant measurements. However, note that the class of measurements possessing the required invariant property is large, and has recently been further extended in [26].

As a perspective, we would like to apply this iterative estimation technique to further problems where the IEKF proved successful, notably highly accurate navigation. We would also like to bring the IIEKF to bear on problems of state estimation with constraints, where constraints could be enforced by using them as pseudo-measurements, as was done in the subgraph intersection problem.

## A Consistency of the linear Kalman filter when facing noise-free measurements

**Theorem 3.** *Consider the linear noise-free measurement*

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k. \quad (64)$$

Let  $\mathbf{K}_k^{\text{nf}} = \mathbf{L}_k(\mathbf{H}_k \mathbf{L}_k)^\dagger$  be the Kalman gain in the limit case of noise-free measurement developed in our preliminary paper [30], with  $\mathbf{P}_k = \mathbf{L}_k \mathbf{L}_k^T$ . The Kalman update, defined as (3d)-(3e) but replacing the gain  $\mathbf{K}_k$  with  $\mathbf{K}_k^{\text{nf}}$ , satisfies Property 1, that is,  $\mathbf{H}_k \hat{\mathbf{x}}_k^+ = \mathbf{y}_k$  and  $\mathbf{H}_k \mathbf{P}_k^+ \mathbf{H}_k = \mathbf{0}$ .

*Proof.* Using (3d), it comes

$$\mathbf{H}_k \hat{\mathbf{x}}_k^+ = \mathbf{H}_k \hat{\mathbf{x}}_k + \mathbf{H}_k \mathbf{K}_k^{\text{nf}} \mathbf{z}_k, \quad (65a)$$

$$= \mathbf{y}_k - \mathbf{z}_k + \mathbf{H}_k \mathbf{K}_k^{\text{nf}} \mathbf{z}_k, \quad (65b)$$

$$= \mathbf{y}_k + (\mathbf{H}_k \mathbf{K}_k^{\text{nf}} - \mathbf{I}) \mathbf{z}_k, \quad (65c)$$

$$= \mathbf{y}_k + (\mathbf{H}_k \mathbf{K}_k^{\text{nf}} \mathbf{H}_k - \mathbf{H}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k). \quad (65d)$$

The estimation error  $(\mathbf{x}_k - \hat{\mathbf{x}}_k)$  follows the distribution  $\mathcal{N}(\mathbf{0}, \mathbf{L}_k \mathbf{L}_k^T)$ . Consequently, there always exists a vector  $\mathbf{v}$  such that  $(\mathbf{x}_k - \hat{\mathbf{x}}_k) = \mathbf{L}_k \mathbf{v}$ . Noting that  $\mathbf{K}_k^{\text{nf}} = \mathbf{L}_k(\mathbf{H}_k \mathbf{L}_k)^\dagger$  and using the property of the pseudo-inverse  $\mathbf{A} \mathbf{A}^\dagger \mathbf{A} = \mathbf{A}$ , yields

$$\mathbf{H}_k \hat{\mathbf{x}}_k^+ = \mathbf{y}_k + (\mathbf{H}_k \mathbf{L}_k (\mathbf{H}_k \mathbf{L}_k)^\dagger \mathbf{H}_k \mathbf{L}_k - \mathbf{H}_k \mathbf{L}_k) \mathbf{v}, \quad (66a)$$

$$= \mathbf{y}_k + (\mathbf{H}_k \mathbf{L}_k - \mathbf{H}_k \mathbf{L}_k) \mathbf{v} = \mathbf{y}_k. \quad (66b)$$

Then, using (3e) and  $\mathbf{K}_k^{\text{nf}} = \mathbf{L}_k(\mathbf{H}_k \mathbf{L}_k)^\dagger$ , it comes

$$\mathbf{H}_k \mathbf{P}_k^+ \mathbf{H}_k^T = (\mathbf{H}_k - \mathbf{H}_k \mathbf{L}_k (\mathbf{H}_k \mathbf{L}_k)^\dagger \mathbf{H}_k) \mathbf{P}_k \mathbf{H}_k^T, \quad (67a)$$

$$= (\mathbf{H}_k - \mathbf{H}_k \mathbf{L}_k (\mathbf{H}_k \mathbf{L}_k)^\dagger \mathbf{H}_k) \mathbf{L}_k \mathbf{L}_k^T \mathbf{H}_k^T, \quad (67b)$$

$$= (\mathbf{H}_k \mathbf{L}_k - \mathbf{H}_k \mathbf{L}_k (\mathbf{H}_k \mathbf{L}_k)^\dagger \mathbf{H}_k \mathbf{L}_k) \mathbf{L}_k^T \mathbf{H}_k^T, \quad (67c)$$

$$= (\mathbf{H}_k \mathbf{L}_k - \mathbf{H}_k \mathbf{L}_k) \mathbf{L}_k^T \mathbf{H}_k^T = \mathbf{0}. \quad (67d)$$

□

**Theorem 4.** *Let  $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$  be the belief of a Kalman filter that is compatible with the measurement  $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k$  in the sense of Definition 2. Consider a different (possibly noisy) measurement  $\tilde{\mathbf{y}}_k = \tilde{\mathbf{H}}_k \mathbf{x}_k + \mathbf{n}_k$ . The updated belief  $(\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+)$  in the light of measurement  $\tilde{\mathbf{y}}_k$  remains compatible with measurement  $\mathbf{y}_k$  in the sense of Definition 2, i.e., the Kalman filter satisfies Property 2.*

*Proof.* Let  $\tilde{\mathbf{K}}_k$  and  $\tilde{\mathbf{z}}_k$  be the Kalman gain and the innovation associated to measurement  $\tilde{\mathbf{y}}_k$ . Let  $\text{Im } \mathbf{A}$  and  $\text{Ker } \mathbf{A}$  respectively denote the image and kernel of matrix  $\mathbf{A}$ . The equality  $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}$  implies  $\text{Im } \mathbf{P}_k \subset \text{Ker } \mathbf{H}_k$ . By definition of the Kalman gain, we have

$$\text{Im } \tilde{\mathbf{K}}_k \subseteq \text{Im } \mathbf{P}_k \subset \text{Ker } \mathbf{H}_k, \quad (68)$$

in such a way that  $\mathbf{H}_k \tilde{\mathbf{K}}_k = \mathbf{0}$ . Consequently,  $\mathbf{H}_k \hat{\mathbf{x}}_k^+ = \mathbf{H}_k(\hat{\mathbf{x}}_k + \tilde{\mathbf{K}}_k \tilde{\mathbf{z}}_k) = \mathbf{H}_k \hat{\mathbf{x}}_k + \mathbf{H}_k \tilde{\mathbf{K}}_k \tilde{\mathbf{z}}_k = \mathbf{H}_k \hat{\mathbf{x}}_k = \mathbf{y}_k$ . Similarly, we get  $\mathbf{H}_k \mathbf{P}_k^+ \mathbf{H}_k^T = (\mathbf{H}_k - \mathbf{H}_k \tilde{\mathbf{K}}_k \tilde{\mathbf{H}}_k) \mathbf{P}_k \mathbf{H}_k^T = \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T = \mathbf{0}$ . In the case  $\tilde{\mathbf{y}}_k$  is a noise-free measurement and the noise-free gain  $\mathbf{K}_k^{\text{nf}}$  is used, the same result applies as  $\text{Im } \mathbf{K}_k^{\text{nf}} \subseteq \text{Im } \mathbf{P}_k$  too. □

## B Right Lie Jacobian

The matrix  $\mathcal{J}_r(\boldsymbol{\xi})$  denotes the right Jacobian of group  $G$  evaluated at  $\boldsymbol{\xi}$ , see [31]. It is defined through the relation

$$\mathcal{L}_{\mathfrak{g}}(\mathcal{J}_r(\mathbf{a})^{-1}\mathbf{b}) = \sum_{n=0}^{+\infty} \frac{B_n}{n!} \underbrace{[\mathcal{L}_{\mathfrak{g}}(\mathbf{a}), [\mathcal{L}_{\mathfrak{g}}(\mathbf{a}), \dots [\mathcal{L}_{\mathfrak{g}}(\mathbf{a}), \mathcal{L}_{\mathfrak{g}}(\mathbf{b})] \dots]]}_n, \quad (69)$$

where  $B_n$  are the Bernoulli numbers,  $[\mathbf{A}, \mathbf{B}]$  is the usual Lie bracket  $\mathbf{AB} - \mathbf{BA}$ , and  $\mathcal{L}_{\mathfrak{g}}(\cdot) : \mathbb{R}^n \rightarrow \mathfrak{g}$  is the bijective map between the Lie algebra  $\mathfrak{g}$  of  $G$  and the associated vector space, see  $\mathbb{R}^n$  [40], defined by  $\exp(\boldsymbol{\xi}) \approx \mathbf{I} + \mathcal{L}_{\mathfrak{g}}(\boldsymbol{\xi})$  neglecting second-order terms in  $\boldsymbol{\xi}$ . Closed-form expressions for the right Jacobians of the groups  $SO(3)$  and  $SE(3)$  are provided in [40]. The right Jacobian of the group of extended pose  $SE_2(3)$  is detailed in [39].

## C Crane application: implementation details

### C.1 Observability

Quite intuitively, the rotation around the gravity vector is not observable. Indeed, the sequence of IMU outputs produced by following a trajectory remains unchanged by any rotation of the initial state around the gravity vector. This makes it impossible to differentiate two trajectories starting from initial configurations that differ only in azimuth angle  $\varphi_0$ . This can be shown mathematically as followed. Considering no process and no measurement noise, the linearized error system from the invariant framework writes

$$\boldsymbol{\xi}_{k+1} = \mathbf{F}_k \boldsymbol{\xi}_k, \quad (70a)$$

$$\mathbf{z}_k = \exp_{SE_2(3)}(\boldsymbol{\xi}_k) \mathbf{d}_k - \mathbf{d}_k. \quad (70b)$$

Due to the log-linearity property of the IEKF, see [26], observability can be assessed using system (70). For this system to be observable, we must be able to retrieve  $\boldsymbol{\xi}_k$  from past inputs and past innovations. Since  $\mathbf{z}_k$  is a nonlinear function of  $\boldsymbol{\xi}_k$ , observability can only be assessed locally. Up to the first order, we get

$$\begin{bmatrix} \mathbf{z}_k \\ \mathbf{z}_{k-1} \\ \mathbf{z}_{k-2} \end{bmatrix} = \begin{bmatrix} \exp_{SE_2(3)}(\boldsymbol{\xi}_k) \mathbf{d}_k - \mathbf{d}_k \\ \exp_{SE_2(3)}(\mathbf{F}_{k-1}^{-1} \boldsymbol{\xi}_k) \mathbf{d}_{k-1} - \mathbf{d}_{k-1} \\ \exp_{SE_2(3)}(\mathbf{F}_{k-2}^{-1} \mathbf{F}_{k-1}^{-1} \boldsymbol{\xi}_k) \mathbf{d}_{k-2} - \mathbf{d}_{k-2} \end{bmatrix}, \quad (71a)$$

$$\approx \begin{bmatrix} \mathbf{H}_k \\ \mathbf{H}_{k-1} \mathbf{F}_{k-1}^{-1} \\ \mathbf{H}_{k-2} \mathbf{F}_{k-2}^{-1} \mathbf{F}_{k-1}^{-1} \end{bmatrix} \boldsymbol{\xi}_k \quad (71b)$$

$$\approx \underbrace{\begin{bmatrix} -(\mathbf{r}_k)_{\times} & \mathbf{0} & \mathbf{I} \\ \boldsymbol{\Psi} & -dt\boldsymbol{\Omega}_{k-1} & \boldsymbol{\Omega}_{k-1} \\ \boldsymbol{\Sigma} & -2dt\boldsymbol{\Phi} & \boldsymbol{\Phi} \end{bmatrix}}_{\mathcal{O}} \boldsymbol{\xi}_k, \quad (71c)$$

where  $\mathcal{O}$  is the observability matrix with

$$\boldsymbol{\Psi} = -(\mathbf{r}_{k-1} + dt^2 \mathbf{a}_{k-1})_{\times} \boldsymbol{\Omega}_{k-1}, \quad (72a)$$

$$\boldsymbol{\Phi} = \boldsymbol{\Omega}_{k-2} \boldsymbol{\Omega}_{k-1}, \quad (72b)$$

$$\boldsymbol{\Sigma} = -\boldsymbol{\Phi}(\boldsymbol{\Phi}^{-1}(\mathbf{r}_{k-2} + dt^2 \mathbf{a}_{k-2}) + 2dt^2 \boldsymbol{\Omega}_{k-1}^{-1} \mathbf{a}_{k-1})_{\times}, \quad (72c)$$

where  $\boldsymbol{\Omega}_k = \exp_{SO(3)}(\boldsymbol{\omega}_k dt)$  and  $(\cdot)_{\times} : \mathbb{R}^3 \rightarrow \mathfrak{so}(3)$  is the skew-symmetric operator associated to cross product on  $\mathbb{R}^3$ . The six first rows of matrix  $\mathcal{O}$  are linearly independent. Denoting by  $\mathbf{O}_i \in \mathbb{R}^{3 \times 9}$  the  $i^{\text{th}}$   $3 \times 9$  block-row of matrix  $\mathcal{O}$ , and performing the linear combination

$$\mathbf{O}_3 \leftarrow -\boldsymbol{\Phi}^{-1} \mathbf{O}_3 + 2\boldsymbol{\Omega}_{k-1}^{-1} \mathbf{O}_2 - \mathbf{O}_1, \quad (73)$$

it comes

$$\mathbf{O}_3 = [(\boldsymbol{\Phi}^{-1}(\mathbf{r}_{k-2} + dt^2 \mathbf{a}_{k-2}) - 2\boldsymbol{\Omega}_{k-1}^{-1} \mathbf{r}_{k-1} + \mathbf{r}_k)_{\times} \quad \mathbf{0}]. \quad (74)$$

The rank of  $\mathbf{O}_3$  is either 2 or 0 as the skew-symmetric matrix  $(\mathbf{b})_{\times}$  has a rank equal to 2 whenever  $\mathbf{b}$  is not  $\mathbf{0}$ . System (70) is therefore not observable as the observability matrix  $\mathcal{O}$  is not full rank.

## C.2 Average normalized estimation error squared (ANEES)

The ANEES is defined by

$$\bar{\epsilon}_k = \frac{1}{N_s} \sum_{n=1}^{N_s} \mathbf{e}_{k,n}^{+T} (\mathbf{P}_k^+)^{-1} \mathbf{e}_{k,n}^+, \quad (75)$$

for the EKF and IterEKF, and by

$$\bar{\epsilon}_k = \frac{1}{N_s} \sum_{n=1}^{N_s} \boldsymbol{\xi}_{k,n}^{+T} (\mathbf{P}_{k,n}^+)^{-1} \boldsymbol{\xi}_{k,n}^+, \quad (76)$$

for the IEKF and IIEKF. The subscript  $n$  is the index of the current simulation among the  $N_s$  simulations. Let  $N_{\text{dof}}$  be the number of freedom degrees of the considered motion. The ANEES converges towards  $N_{\text{dof}}$  and the random variable  $N_s \bar{\epsilon}_k$  follows a Chi-Square distribution with  $N_s N_{\text{dof}}$  degrees of freedom. A 95%-confidence interval  $[r_1, r_2]$  for  $\bar{\epsilon}_k$  can be computed, where

$$r_1 = \frac{1}{N_s} F_{\chi_{N_s N_{\text{dof}}}^2}^{-1} (0.025), \quad (77)$$

and

$$r_2 = \frac{1}{N_s} F_{\chi_{N_s N_{\text{dof}}}^2}^{-1} (0.975), \quad (78)$$

with  $F_{\chi_{N_s N_{\text{dof}}}^2}^{-1}$  the inverse cumulative distribution function of the Chi-Square distribution. An ANEES above  $r_2$  is the sign of an over-confident filter, while an ANEES below  $r_1$  characterizes an under-confident filter.

## C.3 Jacobians

The dynamics and output Jacobians of the EKF and IterEKF are given by

$$\mathbf{F}_k^{\text{EKF}} = \begin{bmatrix} \boldsymbol{\Omega}_k^{-1} & \mathbf{0} & \mathbf{0} \\ -\hat{\mathbf{R}}_k(\mathbf{a}_k)_{\times} dt & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} dt & \mathbf{I} \end{bmatrix}, \quad (79)$$

and

$$\mathbf{H}_k^{\text{EKF}} = [-\hat{\mathbf{R}}_k(\mathbf{r}_k)_{\times} \quad \mathbf{0} \quad \mathbf{I}], \quad (80)$$

where  $(\cdot)_{\times}$  is the skew-symmetric operator associated with cross product in  $\mathbb{R}^3$ , and where  $\boldsymbol{\Omega}_k = \exp_{SO(3)}(\boldsymbol{\omega}_k dt)$ . The Jacobian of the dynamics w.r.t. the process noise is given by

$$\mathbf{G}_k^{\text{EKF}} = \begin{bmatrix} \mathbf{J}_l(-\boldsymbol{\omega}_k dt) dt & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_k dt \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (81)$$

where  $\mathbf{J}_l(-\boldsymbol{\omega}_k dt)$  is the left Jacobian of  $SO(3)$  evaluated at  $-\boldsymbol{\omega}_k dt$ , see [40]. The recent two-frames theory [26] provides a clear and fast way to derive the dynamics and output Jacobians for the IEKF and IIEKF. For the problem at hand, the theory respectively yields, see Proposition 13 of [26],

$$\mathbf{F}_k^{\text{IEKF}} = \begin{bmatrix} \boldsymbol{\Omega}_k^{-1} & \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\Omega}_k^{-1}(\mathbf{a}_k)_{\times} dt & \boldsymbol{\Omega}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_k^{-1} dt & \boldsymbol{\Omega}_k^{-1} \end{bmatrix}, \quad (82)$$

and

$$\mathbf{H}_k^{\text{IEKF}} = [-(\mathbf{r}_k)_{\times} \quad \mathbf{0} \quad \mathbf{I}]. \quad (83)$$

The Jacobian of the dynamics w.r.t. the process noise is given by

$$\mathbf{G}_k^{\text{IEKF}} = \begin{bmatrix} \mathbf{J}_l(-\boldsymbol{\omega}_k dt) dt & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_k^{-1} dt \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (84)$$

In the following, the superscripts EKF and IEKF are dropped when the context is clear enough.

## D Difference with the iterated EKF on Lie groups

The IIEKF differs from the iterated EKF on Lie groups of [6] in various important respects, as [6] is concerned with adapting the iterated EKF to the manifold structure of the state space at a general level, whereas we pursue an approach tailored to a specific class of measurements. In particular the IIEKF boils down to the IEKF when using one iteration, which shows its specificity.

More in detail, the differences are: 1) We presented the IIEKF in the left-invariant formalism and assumed the state follows the concentrated Gaussian distribution (26), while [6] models uncertainty with the concentrated Gaussian defined through right multiplication, that is,

$$\chi_k = \exp_G(\xi_k)\hat{\chi}_k, \quad \text{with } \xi_k \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_k). \quad (85)$$

This is arbitrary anyway, as [6] makes no assumption regarding the output map, leading to a general algorithm, but without exploiting any relation between the uncertainty representation and the measurements, whereas in our work the relationship is very strong, as dramatically illustrated by Proposition 3. 2) We consider (left-)invariant measurements of the form (24b) and use the innovation defined by (28b). Our Jacobian  $\mathbf{H}_k$  is therefore independent from the current estimate, see [14]. This autonomy w.r.t. the current estimate makes our Riccati update (39) completely independent from the outcome of the GN iterative scheme, contrary to [6]. 3) The work about EKF on Lie groups [6, 41] focuses on standard groups such as  $SE(3)$  or direct products of the form  $SO(3) \times \mathbb{R}^k$  and does not seek to exploit the relation between the dynamics, or the measurement map, and the group structure, that allows in particular to define interesting innovation variables. 4) We presented the MAP optimization problem as an optimization over the variable  $\xi$ , see (31), so that the GN method is applied to a function whose domain is the vector space  $\mathbb{R}^n$ . By contrast, the LG-GN developed in [6], operates directly on the Lie group  $G$ . This requires the computation of additional Jacobians in the iterative scheme, see (53) in [6]. 5) We have derived strong theoretical guarantees for the IIEKF.

## References

- [1] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [2] S. F. Schmidt, “The kalman filter-its recognition and development for aerospace applications,” *Journal of Guidance and Control*, vol. 4, no. 1, pp. 4–7, 1981.
- [3] B. M. Bell and F. W. Cathey, “The iterated kalman filter update as a gauss-newton method,” *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, 1993.
- [4] J. Zhao, M. Netto, and L. Mili, “A robust iterated extended kalman filter for power system dynamic state estimation,” *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3205–3216, 2016.
- [5] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, “Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback,” *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.
- [6] G. Bourmaud, R. M egret, A. Giremus, and Y. Berthoumieu, “From intrinsic optimization to iterated extended kalman filtering on lie groups,” *Journal of Mathematical Imaging and Vision*, vol. 55, pp. 284–303, 2016.
- [7] B. Liu, H. Chen, and W. Zhang, “A general iterative extended kalman filter framework for state estimation on matrix lie groups,” in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 1177–1182.
- [8] N. van Der Laan, M. Cohen, J. Arsenault, and J. R. Forbes, “The invariant rauch-tung-striebel smoother,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5067–5074, 2020.
- [9] L. Liang-Qun, J. Hong-Bing, and L. Jun-Hui, “The iterated extended kalman particle filter,” in *IEEE International Symposium on Communications and Information Technology, 2005. ISCIT 2005.*, vol. 2. IEEE, 2005, pp. 1213–1216.
- [10] S. Bonnabel and P. Rouchon, “On invariant observers,” *Control and observer design for nonlinear finite and infinite dimensional systems*, pp. 53–65, 2005.

- [11] R. Mahony, T. Hamel, and J.-M. Pfimlin, “Nonlinear complementary filters on the special orthogonal group,” *IEEE Transactions on automatic control*, vol. 53, no. 5, pp. 1203–1218, 2008.
- [12] M. R. Cohen and J. R. Forbes, “Navigation and control of unconventional vtol uavs in forward-flight with explicit wind velocity estimation,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1151–1158, 2020.
- [13] A. Barrau and S. Bonnabel, “The invariant extended kalman filter as a stable observer,” *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1797–1812, 2016.
- [14] —, “Invariant kalman filtering,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 237–257, 2018.
- [15] R. Mahony and T. Hamel, “A geometric nonlinear observer for simultaneous localisation and mapping,” in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 2408–2415.
- [16] P. van Goor, R. Mahony, T. Hamel, and J. Trunpf, “A geometric observer design for visual localisation and mapping,” in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 2543–2549.
- [17] R. Mahony, T. Hamel, and J. Trunpf, “An homogeneous space geometry for simultaneous localisation and mapping,” *Annual Reviews in Control*, vol. 51, pp. 254–267, 2021.
- [18] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizzle, “Contact-aided invariant extended kalman filtering for robot state estimation,” *The International Journal of Robotics Research*, vol. 39, no. 4, pp. 402–430, 2020.
- [19] N. Pavlasek, A. Walsh, and J. R. Forbes, “Invariant extended kalman filtering using two position receivers for extended pose estimation,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5582–5588.
- [20] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, “An invariant-ekf vins algorithm for improving consistency,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 1578–1585.
- [21] S. Heo and C. G. Park, “Consistent ekf-based visual-inertial odometry on matrix lie group,” *IEEE Sensors Journal*, vol. 18, no. 9, pp. 3780–3788, May 2018.
- [22] M. Wang and A. Tayebi, “Hybrid nonlinear observers for inertial navigation using landmark measurements,” *IEEE Transactions on Automatic Control*, vol. 65, no. 12, pp. 5173–5188, 2020.
- [23] H. A. Hashim, “Gps-denied navigation: Attitude, position, linear velocity, and gravity estimation with nonlinear stochastic observer,” in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 1149–1154.
- [24] P. van Goor and R. Mahony, “Autonomous error and constructive observer design for group affine systems,” in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 4730–4737.
- [25] A. Barrau, “Non-linear state error based extended kalman filters with applications to navigation,” Ph.D. dissertation, Mines Paristech, 2015.
- [26] A. Barrau and S. Bonnabel, “The geometry of navigation problems,” *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 689–704, 2023.
- [27] R. Mahony and J. Trunpf, “Equivariant filter design for kinematic systems on lie groups,” *IFAC-PapersOnLine*, vol. 54, no. 9, pp. 253–260, 2021.
- [28] C. C. Paige and M. A. Saunders, “Lsq: An algorithm for sparse linear equations and sparse least squares,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 8, no. 1, pp. 43–71, 1982.
- [29] D. Simon and T. L. Chia, “Kalman filtering with state equality constraints,” *IEEE transactions on Aerospace and Electronic Systems*, vol. 38, no. 1, pp. 128–136, 2002.



- [30] S. Goffin, S. Bonnabel, O. Bröls, and P. Sacré, “Invariant kalman filtering with noise-free pseudo-measurements,” in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 8665–8671.
- [31] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 1: Classical Results and Geometric Methods*. Springer Science & Business Media, 2009.
- [32] Y. Wang and G. S. Chirikjian, “Error propagation on the euclidean group with applications to manipulator kinematics,” *IEEE Transactions on Robotics*, vol. 22, no. 4, pp. 591–602, 2006.
- [33] K. C. Wolfe, M. Mashner, and G. S. Chirikjian, “Bayesian fusion on lie groups,” *Journal of Algebraic Statistics*, vol. 2, no. 1, 2011.
- [34] G. Bourmaud, R. Mégret, M. Arnaudon, and A. Giremus, “Continuous-discrete extended kalman filter on matrix lie groups using concentrated gaussian distributions,” *Journal of Mathematical Imaging and Vision*, vol. 51, pp. 209–228, 2015.
- [35] P. Chauchat, A. Barrau, and S. Bonnabel, “Kalman filtering with a class of geometric state equality constraints,” in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 2581–2586.
- [36] A. Barrau and S. Bonnabel, “Extended kalman filtering with nonlinear equality constraints: a geometric approach,” *IEEE Transactions on Automatic Control*, vol. 65, no. 6, pp. 2325–2338, 2019.
- [37] F. Rauscher, S. Nann, and O. Sawodny, “Motion control of an overhead crane using a wireless hook mounted IMU,” in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 5677–5682.
- [38] S. Bonnabel and X. Claeys, “The industrial control of tower cranes: An operator-in-the-loop approach [applications in control],” *IEEE Control Systems Magazine*, vol. 40, no. 5, pp. 27–39, 2020.
- [39] M. Brossard, A. Barrau, P. Chauchat, and S. Bonnabel, “Associating uncertainty to extended poses for on lie group imu preintegration with rotating earth,” *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 998–1015, 2021.
- [40] T. D. Barfoot, *State estimation for robotics*. Cambridge University Press, 2017.
- [41] G. Bourmaud, R. Mégret, A. Giremus, and Y. Berthoumieu, “Discrete extended kalman filter on lie groups,” in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.