HAL
open science

Non-linear state error based extended Kalman filters with applications to navigation

Axel Barrau

To cite this version:

Axel Barrau. Non-linear state error based extended Kalman filters with applications to navigation. Signal and Image Processing. Ecole Nationale Supérieure des Mines de Paris, 2015. English. NNT : 2015ENMP0080 . tel-01344622

HAL Id: tel-01344622

https://pastel.hal.science/tel-01344622

Submitted on 12 Jul 2016

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

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

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n°432:
Sciences des Métiers de l'Ingénieur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité "Informatique temps-réel, robotique et automatique"

présentée et soutenue publiquement par

Axel BARRAU

le 15 septembre 2015

Non-linear state error based extended Kalman filters with applications to navigation

Filtres de Kalman étendus reposant sur une variable d'erreur non linéaire avec applications à la navigation

Directeur de thèse: **Silvère BONNABEL**
Co-encadrement de la thèse: **Xavier BISSUEL**

Jury

Brigitte d'ANDREA-NOVEL, Professeur, MINES ParisTech	Examinateur
Xavier BISSUEL, Expert en navigation inertielle, SAGEM	Examinateur
Silvère BONNABEL, Professeur, MINES Paristech	Examinateur
Jay FARRELL, Professeur, UC Riverside	Examinateur
Pascal MORIN, Professeur, Université Pierre et Marie Curie	Rapporteur
Christophe PRIEUR, Directeur de recherche, Gipsa-lab, CNRS	Rapporteur
Pierre ROUCHON, Professeur, MINES Paristech	Président du jury

T H È S E

MINES ParisTech
Centre de Robotique
60 bd Saint-Michel, 75006 Paris, France

Contents

1 Introduction **5**
1.1 Estimating the state of a dynamical system . . . 5
1.2 Probabilistic filtering . . . 5
1.3 Deterministic observers . . . 7
1.4 Autonomous error equations . . . 8
1.5 Invariant observers . . . 10
1.6 Contributions of the present thesis . . . 10
1.7 Organization of the document . . . 14

2 Mathematical preliminaries **15**
2.1 Group of direct planar isometries $SE(2)$ . . . 17
2.2 Group of rotation matrices $SO(3)$ . . . 17
2.3 Group of direct spatial isometries $SE(3)$ . . . 18
2.4 Group of double direct isometries $SE_2(3)$ . . . 18

I Invariant Kalman Filtering **20**

3 Autonomous error equations **21**
3.1 Introduction . . . 21
3.2 Autonomous innovation functions . . . 25
3.3 Autonomous error propagation . . . 26
3.4 A log-linearity property of errors having autonomous propagation . . . 27
3.5 A complete characterization of systems with autonomous errors . . . 30
3.6 Conclusion . . . 32

4 Invariant Extended Kalman Filtering **33**
4.1 Introduction . . . 33
4.2 Full noisy system . . . 34
4.3 Examples . . . 41
4.4 Conclusion . . . 55

5 Industrial applications **56**
5.1 Introduction . . . 56
5.2 Autonomous error equations in practice . . . 57

5.3 Industrial applications . . . 59
5.4 Conclusion . . . 65

II Probabilistic invariant filtering **68**

6 Stochastic properties of invariant observers **69**
6.1 Introduction . . . 69
6.2 Problem setting . . . 70
6.3 A class of discrete-time intrinsic filters . . . 75
6.4 Fixed gains filters . . . 75
6.5 Gaussian filters . . . 80
6.6 Conclusion . . . 86

7 Sampling-based methods **88**
7.1 Introduction . . . 88
7.2 Sampling biased dynamics . . . 88
7.3 The invariant Rao-Blackwellized filter . . . 89
7.4 Application to navigation with jumping GPS . . . 91
7.5 Conclusion . . . 96

III Nonlinear state error based EKFs **98**

8 Remedying EKF-SLAM inconsistency **99**
8.1 Introduction . . . 99
8.2 Kalman filter and observability . . . 100
8.3 SLAM inconsistency . . . 106
8.4 Solving false observability through the choice of the error variable . . . 111
8.5 A different approach to the same result . . . 115
8.6 Equivalence of the two approaches . . . 121
8.7 Conclusion . . . 122

9 EKF based on a non-linear error **124**
9.1 Introduction . . . 124
9.2 Building an EKF upon a non-linear error variable . . . 124
9.3 Non-linear error and false observability . . . 128
9.4 Nonlinear exponential update and singular information . . . 131
9.5 A geometric interpretation . . . 136
9.6 Link with Invariant Kalman filtering on Lie groups . . . 137
9.7 Conclusion . . . 140

10 Illustration by a tutorial example **141**
10.1 Introduction . . . 141
10.2 Perfect dynamics . . . 142
10.3 Classical EKF equations . . . 143

10.4 IEKF equations 144
10.5 Differences between the two filters 145
10.6 Non-linear EKF using implicit update 146
10.7 Experimental comparison 147
10.8 Explanation 148
10.9 Conclusion 151

11 Conclusion of the thesis **154**

A Proofs of the results of Chapter 4 **156**
A.1 Proof of theorem 7 156
A.2 Proof of proposition 8 160

B Proofs of the results of Chapter 6 **162**
B.1 Proofs of the results of Section 6.4.1 162
B.2 Proofs of the results of Section 6.4.2 166

Chapter 1

Introduction

1.1 Estimating the state of a dynamical system

Estimating unknown state of a dynamical system from indirect and imprecise measurements is essential in almost all technological fields. Navigation of aircrafts, weather forecast, autonomous vehicles control, underwater exploration, oil drilling and medical imaging to cite a few. From a mathematical point of view, if the state changes over time due to human intervention or physical laws, the problem is usually set down under the following general form:

$$\begin{aligned} \frac{d}{dt}X &= f_u(X,w), \\ Y &= h(X,V). \end{aligned} \tag{1.1}$$

X is the information of interest, that is, the state: orientation and position of a flying aircraft, air masses situation in meteorology, environment of a vehicle in robotics, etc. Its evolution over time, represented by the function f, depends on its current value X, some additional known parameters u, and some unknown, unpredictable, erratic elements denoted by w. If X is the state of a plane (its heading, velocity, position, etc.), f can represent the (known) laws of Mechanics, u the (known) velocity of the propeller and w all the uncertainties over movements: wind variations, imperfect operating of the piloting commands, approximate knowledge of the aerodynamics of the aircraft, etc.

Y denotes here all the measured variables. Through h, they are defined as a function of X, i.e. of the state of interest. Y can be for example a GPS signal received by a plane: it gives only the position, from which course and attitude have still to be deduced. Moreover, this observation also depends on a set of unsteady and unpredictable parameters denoted by V (GPS errors). The two main approaches to this kind of problems are probabilistic filtering and deterministic observers.

1.2 Probabilistic filtering

Probabilistic filtering is a branch of stochastic process theory that has been strongly driven by applications in control and signal processing. It aims at computing the expecta-

tion (or mean value) $\hat{X}$ of the state X defined above conditioned on the past information, which requires estimating its whole probability density function (p.d.f.). In the particular case where the system is linear and corrupted by independent white Gaussian process and observation noises, this function is Gaussian. This implies it can be fully described by its mean value and covariance matrix. The evolution of these two variables follows the equations of the celebrated Kalman filter [65]. Yet, when the system is non-linear there is no general method to derive efficient filters, and their design has encountered important difficulties in practice. Indeed, general filtering equations describing the evolution of the p.d.f. (see e.g., [27] and references therein) concern a function and cannot be reduced to a finite-dimensional computation.

For this reason, many approximation-based solutions have been proposed. In engineering applications the most popular of them is the extended Kalman filter (EKF). This method amounts to linearize the system around the estimated trajectory, and build a Kalman filter for the linear model, which can in turn be implemented on the non-linear model. One merit of the EKF is to convey an estimation of the whole density probability of the system through a mean state $\hat{X}_t$ and an approximation of the covariance matrix P_t of the error. These can be used to draw "uncertainty ellipsoids", that is, sets likely to contain the true value of X. They have to be seen as the confidence of the algorithm in its estimate $\hat{X}(t)$: the true state is not supposed to be exactly $\hat{X}(t)$ but at least to be in the area defined by the ellipsoid. They are hoped to be representative of the reality (i.e. to contain the true value of the state), but as they are obtained by ways of linearization, this has no reason to be true. A good illustration of this issue is given by Simultaneous Localization and Mapping (SLAM) problem. This problem consists for a robot in building a map of its environment while localizing itself in this map. A theoretical study performed in [61], as well as convincing simulations [8], showed that for this problem the EKF eventually becomes inconsistent, in the sense that the true state has excessive probability to be outside the trusted ellipsoid. This may be harmful as it may for instance result in a collision by misestimating the distance from the robot to an obstacle. In particular, the estimated uncertainty over heading decreases although the general orientation of the map and robot are not observable.

In order to capture more closely the distribution, several rather computationally extensive methods have recently attracted growing interest. Unscented Kalman Filter [62] proposes to compute the covariance matrices appearing in the EKF using a smart deterministic sampling. In most cases (see [107]) the results obtained are better than those of the EKF but [8] showed it does not fix the problems of EKF SLAM inconsistency. Particle filters [28,50] generate a random set of points (the "particles") following a procedure making the result (almost) equivalent to a direct sampling of the p.d.f of interest. Then, the expectation of any useful quantity (the state for instance) can be evaluated as its mean value on the obtained points. With a sufficient number of particles, the result can be made as accurate as wanted. The main drawback of such an approach is the growing size of the sample needed as the dimension of the system increases. In particular, too a sparse sampling compared to the dimensionality of the problem leads to a type of degeneracy: no particle matches the data, even approximately, and the weights vanish. Several solutions have been proposed to tackle this issue, the two main ideas being either to use the observations to smartly propagate the particles [100], or to sample only a reduced set

of relevant variables then use a classical filtering method for each particle to marginalize out the remaining variables. Of course, both can be used simultaneously. The latter is known as the Rao-Blackwellized Particle Filter [39]. Only some of the variables forming the state X are sampled, so that the evaluation of the remaining variables boils down to a linear problem. This residual estimation can be performed using classical Kalman filtering. This is the solution proposed in [84] to the non-linearity issues arising in SLAM: only the heading of the robot is sampled, and for each obtained value a Kalman filter builds the map and localizes the robot.

Filtering on manifolds If the system is naturally defined on a manifold (consider for instance 3D attitude estimation, where the state of the system is a rotation matrix), many works have sought to avoid parameterization and the implied distortions and singularities [41, 60, 88]. The specific situation where the process evolves in a vector space but the observations belong to a manifold has also been considered, see e.g. [40, 90] and more recently [95]. For systems on Lie groups powerful tools to study the filtering equations (such as harmonic analysis [73, 89, 109, 111]) have been used, notably in the case of bilinear systems [110] and estimation of the initial condition of a Brownian motion in [42].

Minimum energy filtering A somewhat different but related approach to filtering consists of finding the path that best fits the data in a deterministic setting, which is equivalent to compute the state maximizing the likelihood of the data. It is thus related to optimal control theory where geometric methods have long played an important role [25]. A certain class of least squares problems on the Euclidean group has been tackled in [53]. This approach has also been used to filter inertial data recently in [93], [76]. Note that for linear systems with Gaussian noises, this definition of the optimal estimate coincides with the previous one and is again given by the Kalman Filter.

1.3 Deterministic observers

The other route to estimate the value of an unknown dynamical system is the design of a deterministic observer. This general concept denominates a variable whose evolution has been chosen by the practitioner as a function of the available data, in a way ensuring the convergence to zero of the error $X(t) - \hat{X}(t)$ where $\hat{X}(t)$ denotes the estimate of the true state $X(t)$ of the dynamical system under consideration. The fundamental difference with the filtering approach is that it does not take into account the noise directly, that is we assume the noise is always turned off ($w = 0$ and $V = 0$ in the above equations). The very structure of the general form of observers (most of them are made of a copy of the dynamics plus a correction term) is built to make it perturbation- or noise-resistant, but the theoretical study focuses on its convergence only when noise is turned off. Although less intuitive, this simplification allows more powerful convergence results in some cases, and the rationale is that a globally convergent observer should be robust even to large noises as the error tends to return to zero no matter its size.

For linear systems, the problem of designing observers is now well understood and the celebrated Luenberger observer [74] allows to assign the velocity at which the estimation

error should exponentially return to zero. Its tuning is experimental, and depends on the level of noises, the idea being that for large noises the convergence velocity should not be too high, as the incoming data cannot be fully trusted. For non-linear systems, there is no general method to design observers, but, there has been very interesting solutions introduced for certain classes of systems. Systems known as linearizable by output injection [52] can be reduced to the linear case where usual methods apply. For system having observability normal form with triangular structure, high-gain observers are appropriate [2–5, 49, 85].

The Kalman filter as an observer An interesting property of the Kalman filter is to be also a convergent deterministic observer. Indeed, if the data are noiseless (i.e. $V = 0, w = 0$), and the equations are linear, for any choice of positive definite covariance matrices for the trusted noises (here they should be referred to as tuning matrices Q, R since we assume there is no noise) the estimate $\hat{X}$ converges to the true state X under natural assumptions [38]. In the non-linear case, the EKF does not possess any convergence guarantee, and its efficiency is aleatory. Indeed, its main flaw lies in its very nature: the Kalman gain is computed assuming the estimation error is sufficiently small to be propagated analytically through a first-order linearization of the dynamics about the *estimated* trajectory. When the estimate is actually far from the true state variable, the linearization is not valid, and results in an ill-adapted gain that may amplify the error. In turn, such positive feedback loop may lead to divergence of the filter. This is the reason why most of the papers dealing with the stability of the EKF (see [23, 24, 92, 102]) rely on the highly non-trivial assumption that the eigenvalues of the Kalman covariance matrix $P(t)$ computed about the *estimated* trajectory are lower and upper bounded by strictly positive scalars. Under this assumption, the error as long as it is sufficiently small can not be amplified in a way that makes the filter diverge. To the author's knowledge, only a few papers deal with the stability of the EKF without invoking this assumption [69]. It is then replaced by second-order properties whose verification can prove difficult in practical situations. This is also due to the fact that the filter *can* diverge indeed in a number of applications. Moreover, the lack of guarantees goes beyond the general theory: there are few engineering examples where the EKF is proved to be (locally) stable.

1.4 Autonomous error equations

Both Luenberger observer and Kalman filter allow to build estimation methods with convergence guarantees for deterministic linear problems defined by the general equations:

$$\frac{d}{dt}X_t = A_t X_t + B_t u_t,$$

$$Y_{t_n} = H_n X_{t_n},$$

where $X_t \in \mathbb{R}^n$ is a vector and A_t, B_t, H_n are matrices with appropriate dimensions. Note that we removed the noises V_n and w_t from the equations (see explanations above). These filters share the general form:

Propagation:

$$\frac{d}{dt}\hat{X}_t = A_t\hat{X}_t + B_t u_t, \qquad t_{n-1} < t < t_n,$$

Update:

$$\hat{X}_{t_n}^+ = \hat{X}_{t_n} + K_n\left(Y_n - H_n\hat{X}_{t_n}\right), \qquad t = t_n,$$

where $\hat{X}_t$ is the estimate, the superscript $+$ is used to denote the value of the estimate just after the update and Y_n is the observation at time t_n. Matrix K_n, called "gain", is defined differently depending on the chosen method. But in any case, the estimation error $e_t = X_t - \hat{X}_t$ has the following evolution:

$$\frac{d}{dt}e_t = A_t e_t,$$

$$e_{t_n}^+ = (Id - K_n H_n)e_{t_n}.$$

The estimate $\hat{X}_t$ and true state X_t do not appear in these equations. This property is the founding principle of all the convergence results of these methods. Indeed, it makes the analysis of the error e_t independent from a specific trajectory of the system. One of the contributions of the present work is to study non-linear systems offering the same advantage. In more mathematical terms, the properties we are interested in are the following:

Definition 1 (Autonomous error equation)**.** *Consider an observer providing an estimate $\hat{X}_t$ defined by the two usual steps:*

Propagation:

$$\frac{d}{dt}\hat{X}_t = f_{u_t}(\hat{X}_t), \qquad t_{n-1} < t < t_n,$$

Update:

$$\hat{X}_{t_n}^+ = \psi\left(\hat{X}_{t_n}, h(X_{t_n}), (u_s)_{s\leqslant t_n}\right), \qquad t = t_n,$$

where ψ is a function of the current estimate, the observation and the past inputs. The error variable $e_t = \varepsilon(X_t, \hat{X}_t)$ is told to follow an autonomous equation if there exist two functions g and ψ^e such that for any couple of solutions $X, \hat{X}$ of respectively the system and the observer (or filter) dynamics, the error $e_t = \varepsilon(X, \hat{X})$ verifies at each step:

Propagation:

$$\frac{d}{dt}e_t = g_{u_t}(e_t), \qquad t_{n-1} < t < t_n,$$

Update:

$$e_{t_n}^+ = \psi^e\left(e_{t_n}, (u_s)_{s\leqslant t_n}\right), \qquad t = t_n.$$

The terminology "fully autonomous error variable" will be used if the function ψ^e is in addition totally independent from the inputs.

Note that the terminology "autonomous" above is slightly abusive at it allows the error equation to depend on the time through the inputs, this is why we introduce the concept of "fully autonomous" which corresponds to the standard notion of autonomy in the ordinary differential equations literature.

A natural question is whether or not non-linear systems offering such an error variable exist. A first answer is given by the theory of invariant observers.

1.5 Invariant observers

Invariant observers are a class of deterministic observers meant to be appropriate to dynamical systems possessing symmetries. Roughly speaking, a dynamical system possesses symmetries, if it “looks” the same after having undergone a certain transformation. For instance, the motion of a car looks the same whether it is heading west or north. However, a badly tuned extended Kalman filter meant to estimate the motion of the car, might behave differently in both cases. Invariant observers remedy this issue: their general form is such that the estimation is identical for the transformed system. They were introduced to our best knowledge in [1] and well developed in [15, 19] (see also e.g. [10, 106]). Such observers do not have any guaranteed convergence properties. However, for a system living on a Lie group and having enough symmetries, namely a left-invariant dynamics with right-equivariant output (and vice versa), the estimation error between the true state and the state estimated by an invariant observer has an autonomous evolution [16]. This property has been discovered independently and leveraged by many authors in various theoretical and applicative contexts (see e.g. [51, 70, 77, 105, 106]). The idea of taking advantage of the properties of Lie groups to design convergent observers (see also [78–80]) is surprisingly new if compared to the wide literature dedicated to the subject of bringing the symmetries or Lie group structure to bear on the design of controllers, dating back to the 1970’s [64]. The popularity of the approach has persisted [26, 63] and has lead to very interesting and relatively recent methods [46, 86, 87].

1.6 Contributions of the present thesis

The present work has been made possible by previous research on invariant observers initiated in [16, 19]. We believe the breakthrough of this theory was to show that observers combining a non-linear error variable (error between true and estimated state) with a non-linear innovation (that is, correction) term could lead to an autonomous error equation. The same authors also suggested to linearize the introduced non-linear invariant error and to use Kalman equations to make it decrease, leading to the theory of the Invariant EKF (IEKF), see [14, 17]. By the time this PhD started, some questions were however left unanswered:

1. Invariant systems are generally not linear and linear system are generally not invariant, but they share the property of having an error variable with autonomous error equation. Is there a more general class of systems containing both but preserving their main characteristics ?

2. Beyond the IEKF, what properties should one expect from an EKF which is meant to make a linearized non-linear error variable decrease, that is a general function $\varepsilon(X,\hat{X})$ rather than the usual linear state error $\hat{X}-X$?

Industrial motivations The starting point of the present work is industrial applications to inertial navigation and aircraft geo-localization. Indeed, the present PhD was done in partnership with the company SAGEM, the No. 1 company in Europe and No. 3 worldwide for inertial navigation systems (INS). Half of the time was spent within the company. The initial goal was to apply the theory of symmetry-preserving (or invariant) observers for the problem of geo-localization. This came with two constraints:

1. Current state estimation in SAGEM's products essentially relies on extended Kalman filtering. Indeed, the EKF has proved to be reliable for a wide range of aerospace applications for a long time now (starting with the Apollo program). Industrialists are conscious of its potential shortcomings; but generally trust its performances and appreciate its easiness of implementation and low computational requirements (compared to sampling based state of the art particle filters). The novel algorithms developed should be essentially based on extended Kalman filtering as well to be industrially acceptable and lead to commercial products. Last but not least, the high rate of inertial measurements and the discrete nature of other measurements (GPS, altimeter, etc.) favored the use of continuous-time EKFs with discrete observations.

2. Even the simple system of an aircraft or Unmanned Aerial Vehicle (UAV) navigating on flat earth equipped with accelerometers and gyrometers does *not* define a left-invariant system on a Lie group, and thus does not fit into the framework of [16] that allows for an autonomous error equation, a key property being at the heart of our present work. Yet the true navigation equations as used by SAGEM engineers are even much more complicated than this. This raised the question of extending the approach to systems that were neither left nor right invariant systems on Lie groups.

For those reasons, the present work essentially focuses on the Invariant EKF (which is a form of EKF) on Lie groups for state estimation, with continuous time dynamics and discrete observations, and seeks to broaden the class of systems for which such a tool may prove useful.

The IEKF as a stable observer for a broad class of systems The main contribution of the present work is to characterize the class of systems defined on Lie groups such that in the absence of measurements, the (invariant) error between the true state and the estimated state (that is, any other solution to the dynamical system in question) follows an autonomous equation (to be more precise an equation that may depend on the system's inputs but which does not explicitly depend on the true nor estimated system's state). Those systems include left and right invariant dynamics, but also a combination of both, linear dynamics and some more complicated systems also. We further show that, for this class of systems the logarithm of the invariant error follows a *linear* differential equation in the Lie algebra, although the system is totally non-linear (the dynamics is

non-linear and the state space is not a vector space). The very unexpected latter property is leveraged to prove that for this class of systems on Lie groups, the IEKF viewed as a deterministic observer (that is turning all noises to zero) converges around *any* trajectory under standard observability conditions. This is a very remarkable feature as it is unusual to be able to design non-linear observers that converge around any trajectory for challenging non-linear systems of engineering interest, and here the linearized error system is time-dependent (through the system's inputs and the linearization around the estimated state) which makes it a challenge to stabilize. Many systems of engineering interest modeling the behavior of a body flying in space fit into the proposed framework, and the IEKF inherits the local convergence properties derived in the general case.

Stochastic convergence properties of invariant observers Then, the present thesis investigates the stochastic properties possessed by estimators with the autonomous error equation property. More precisely, we consider invariant observers (with constant non-linear gains) such that the estimation error follows a fully autonomous equation (that is does not even depend on the inputs). Even if the design is made by assuming noise is turned off, we turn the noise on and study how the observer behaves. In this case, the estimation error becomes a time-homogeneous Markov chain due to the autonomy of the error. We then show that under some assumptions, the error's distribution (as a random variable) converges asymptotically to some fixed probability distribution. This is an unusual result in the theory of filtering on Lie groups, and from a practical viewpoint, it allows to tune the gains so as to minimize the dispersion of the asymptotic distribution, that can be evaluated empirically through Monte-Carlo simulations. The gain tuning can thus possess some optimality properties although assessed *in advance*, i.e. independently from the actual measurements, a surprising and remarkable property.

Non-linear state errors for EKF-SLAM The last important contribution of the present work is to attempt to generalize the invariant approach, or in other terms attempt to understand at a more general level what makes it so appealing. Concretely, we raise the issue of the use of general non-linear state errors when devising EKFs, not necessarily on Lie groups. We wonder : what is the virtue of non-linear estimation errors rather than linear errors (beyond being able to prove some convergence properties) ? For example, given that one can devise two types of invariant errors on Lie groups (right-invariant and left-invariant, see [15]) which one should one prefer ? We propose two answers to those questions.

The first answer deals with the issue of *observability*. Indeed, if some variables (or some combination of variables) are not observable, then an observer, or an EKF should not try to estimate those variables. However, the EKF sometimes does not "see" those variables are not observable, and due to linearization it starts thinking it could estimate them. This leads to inconsistency of the EKF estimates (that is over-optimism with respect to its estimates) in the problem of [61] for example. A simple way to remedy this issue, is to devise a non-linear state error such that the non-observable directions are always the same for this error, independently of the state X. This is in general impossible to achieve, but becomes feasible when on can produce an error whose evolution is relatively

independent of the trajectory (ideally is autonomous). Note that, such a problem does not arise in linear systems as the non observable directions are the same whatever the state X. For this reason, we have managed to devise an EKF for the SLAM (simultaneous localization and mapping) problem that solves a part of the very well-known inconsistency issues in the EKF-SLAM [8, 31, 57–59, 61] . This is a novel result to our knowledge, and we believe it is a fundamental result in estimation theory and robotics.

The second approaches the problem from the other end: in many applications of the EKF (and especially in SLAM) some function of the state is observable with extreme precision while the remaining part is still uncertain. This is what happens for example if a robot travels for a long time with little heading information. It is able to build an accurate map of its close environment but has no idea of the absolute orientation of this map. The covariance matrix P of the EKF is then almost singular. When this kind of situation arises in a linear problem, the Kalman filter still works perfectly and its updates are contained in the "authorized" space, i.e. in the vector space spanned by the eigenvectors of P having non-zero eigenvalues. In our (non-linear) SLAM example, the corresponding property would be the local map to be only rotated by a heading update resulting from a late heading measurement (but without shape distortion of the map). We will show this is not verified at all by a classical EKF. But with an appropriate error variable and a modified update discussed in this document, an EKF natively ensuring this property can be obtained without any patchwork forcing the constraint.

Industrial contributions The theory developed in this document was successfully applied to some estimation problems arising in inertial navigation. This field aims at using inertial measurements coming from accelerometers and gyroscopes to compute the position, velocity and attitude of a vehicle. Merging the data provided by these sensors with other sources of information is usually performed using an EKF and gives a natural application to the class of filters we study here. As the current methods are already efficient, we focused on the situations where they reach their limits.

- The first one is the transitory phase of the convergence where the estimation error is possibly large. An EKF failing in this situation to achieve acceptable performance, some methods involving preliminary coarse estimation steps are currently in use. Their design is not trivial and has to be adapted to each new situation. Moreover, the coarse estimation phase extends the duration of the initialization, which can be critical in practice. The enhanced EKF we implemented preserves its performance even for very large initial errors and allows to skip the coarse estimation steps. The implementation and adaptation become straightforward, the time required for initialization is reduced and some constrained imposed to the pilot can be loosened. This method has been industrially approved and adopted by SAGEM to replace their state of the art EKF algorithm on a wide-spread commercial military device.

- The second one is the case of a quasi-static pose, with constant heading but polluted by a specific pattern of movement. If the EKF maintains its estimate using as an observation the fact that the position is approximately constant, a systematic error appears on the long run, due to the issue of false observability explained further down this document. This can be solved by some patch-up work designed

precisely to solve this problem, but raising also new issues. The method developed during the thesis for this filtering situation solved the encountered problem although not being specifically tuned for that.

1.7 Organization of the document

Chapter 2 is a brief tutorial giving the basic notions about Lie groups required to read this document. We divided the sequel into three parts. To some respect, the parts can be read independently. Note, this comes at the price of several repetitions throughout the manuscript.

Part I is dedicated to the IEKF. We first characterize in Chapter 3 a large class of systems on Lie groups where an error having autonomous propagation can be designed. This novel framework, containing left-invariant dynamics, right-invariant dynamics, combinations of both, linear dynamics, and some systems arising in inertial navigation, is shown to ensure novel and striking linearity properties. An IEKF for this extended class of systems is then presented in Chapter 4, along with new stability properties applied to a long list of examples. The benefits of the approach are illustrated in Chapter 5 by industrial applications and comparison with tried and tested commercial softwares.

Part II introduces a fully probabilistic framework and then considers the whole probability density function of the state of the system as the quantity of interest. The framework of stochastic processes on Lie groups is used in Chapter 6 to write the evolution of the error as a Markov chain (more precisely a Harris chain) and we show under proper conditions the almost global convergence property of the deterministic state error with noise turned off leads to a convergence in distribution of the error when noise is turned on. Still in a probabilistic framework, yet studying a quite different approach, a practical sampling method (that is particle filter) taking advantage of the autonomy property of the error, is proposed in Chapter 7, and applied to a non-Gaussian navigation problem.

Part III extends the use of non-linear error variables to the general design of an EKF and proves its dramatic implications. The exposure starts with the classical example of SLAM: Chapter 8 shows that the widely acknowledged flaw of the EKF for this problem is only due to the (classical yet arbitrary) linear error variable. A well chosen non-linear definition of the error is shown to solve the problem. The study is then generalized in Chapter 9 and novel properties are derived, which do not rely on first-order approximations (that is, some more global properties of EKFs based on alternative state errors are outlined). They are illustrated on another (non-SLAM) example which can be seen as a simplified version of the industrial applications described earlier, and provides physical insight into the EKFs based on non-linear state errors.

Chapter 2

Mathematical preliminaries

In this document we will have to design some error variables adapted to the different non-linear systems we will consider. A very convenient way to define an error is to start from an existing group operation. For this reason, the notion of matrix Lie group will be extensively used:

Definition 2. *A matrix group G is a set of square invertible matrices of size $n \times n$ on which matrix multiplication and inversion can be safely used without going outside the set. In other words, the following properties are verified:*

$$Id \in G, \qquad \forall g \in G, g^{-1} \in G \qquad \textit{and} \qquad \forall a, b \in G, ab \in G.$$

An example of matrix group is given by the group of 2D rotations $SO(2)$:

$$\left\{ \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \theta \in \mathbb{R} \right\}.$$

To obtain a Lie group, we also need a tangent space $\mathfrak{g}$ about identity Id, defined as the matrix subspace spanned by all possible first-order variations inside G around Id. For rotation matrices, Id corresponds to $\theta = 0$. For theta small, the first-order variation of the rotation matrix is:

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = Id + \theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \circ(\theta).$$

Thus, if $G = SO(2)$, $\mathfrak{g}$ is the one-dimensional matrix space spanned by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Note that G was parameterized by a single coordinate, and $\mathfrak{g}$ is one-dimensional. Of course this no coincidence, the dimension of $\mathfrak{g}$ is also the minimum number of real numbers needed to parameterize G, and is called the *dimension* of the Lie group G. Another way to see $\mathfrak{g}$ is to consider all processes $\gamma(t)$ taking values in G and starting from $\gamma(0) = Id$. $\mathfrak{g}$ is then the space of all possible values of $(\frac{d}{dt}\gamma)(0)$. This definition makes the derivation of $\mathfrak{g}$ more convenient in the following example:

Example 1. *$G = SO(3) = \{H \in \mathscr{M}_3(\mathbb{R}), H^T H = Id, det(H) > 0\}$. Consider a process $\gamma(t)$ on G and assume $\gamma(0) = Id$. We have $\forall t > 0, \gamma(t)^T \gamma(t) = Id$. Deriving this expression we get: $(\frac{d}{dt}\gamma)(t)^T \gamma(t) + \gamma(t)^T (\frac{d}{dt}\gamma)(t) = 0$. At $t = 0$ and denoting $(\frac{d}{dt}\gamma)(0)$ by ξ we obtain $\xi^T + \xi = 0$, i.e. ξ is skew-symmetric matrix. Reciprocally, any skew-symmetric matrix is the derivative at zero of a process taking values in G as the solution of the equation $M_0 = Id, \frac{d}{dt} M_t = M_t \omega_\times$ stays in $SO(3)$ for $t > O$ if $\omega_\times$ is a skew-symmetric matrix. Finally, $\mathfrak{g}$ is here the space of skew-symmetric matrices.*

If this tangent space about identity is defined, it can also be defined about any element g of G as the possible derivatives at zero of a process $\gamma(t)$ starting from $\gamma(0) = g$. Its characterization is obvious as the processes $g^{-1}\gamma(t)$ and $\gamma(t)g^{-1}$ start from Id. We obtain:

Proposition 1. *For $g \in G$, the tangent plan to G about g has the two equivalent characterizations $g \cdot \mathfrak{g}$ and $\mathfrak{g} \cdot g$.*

An interesting consequence of Proposition 1 is the property

$$g \cdot \mathfrak{g} \cdot g^{-1} = \mathfrak{g}.$$

In other words, for any $g \in G$ and $\xi \in \mathfrak{g}$ we have $g\xi g^{-1} \in \mathfrak{g}$. The latter property defines for any $g \in G$ a linear function from $\mathfrak{g}$ to $\mathfrak{g}$. It is called *adjoint operator* and denoted by $Ad_g : \mathfrak{g} \to \mathfrak{g}$. The differential of the operator Ad at identity is denoted ad, which means that if we consider a smooth process g_t taking values in G and verifying $g_0 = Id, \frac{d}{dt} g_t = \xi \in \mathfrak{g}$ then we have:

$$\frac{d}{dt} Ad_{g_t}|_{t=0} = ad_\xi.$$

An equivalent notation for $ad_{\xi_1}\xi_2$ is $[\xi_1, \xi_2]$, usually called the *Lie bracket* of ξ_1 and ξ_2. This multiplicative operation endows $\mathfrak{g}$ with the structure of an algebra, and is the reason why $\mathfrak{g}$ is often called the *Lie algebra* of G. As it proves useful to identify $\mathfrak{g}$ to $\mathbb{R}^{\dim \mathfrak{g}}$, a linear mapping $\mathscr{L}_\mathfrak{g} : \mathbb{R}^{\dim \mathfrak{g}} \to \mathfrak{g}$ will often be used. The mapping is defined below for the matrix Lie groups considered in this document, but it is a mere convention and any choice of a mapping $\mathscr{L}_\mathfrak{g}$ would be equivalent. An interesting property of the matrix exponential is the following:

Proposition 2. *Let $\exp_m$ denote the matrix exponential:*

$$\exp_m(A) = \sum_{i=1}^{+\infty} \frac{A^k}{k!}.$$

Then we have:

$$\exp_m(\mathfrak{g}) \subset G.$$

This gives a simple way to map $\mathfrak{g}$ to G (the rest of the document will give new arguments supporting the use of this function for estimation purposes). As well, $\mathbb{R}^{\dim \mathfrak{g}}$ can be mapped to G through a function $\exp$ defined by $\exp(\xi) = \exp_m(\mathscr{L}_\mathfrak{g}(\xi))$. For all matrix

Lie groups considered in this document, no matrix exponentiation is actually needed as there exist closed formulas, given thereafter. The following property of the exponential mapping will be useful:

Proposition 3. *The functions* Ad *and* $\exp$ *commute in the following sense:*

$$\forall g \in G, \forall \xi \in \mathfrak{g}, \exp(Ad_g \xi) = g \exp(\xi) g^{-1}.$$

We give now a short description of the most useful matrix Lie groups for our purpose.

2.1 Group of direct planar isometries $SE(2)$

We have here $G = SE(2)$ and $\mathfrak{g} = \mathfrak{se}(2)$, where :

$$SE(2) = \left\{ \begin{pmatrix} R(\theta) & x \\ 0_{1,2} & 1 \end{pmatrix}, \begin{pmatrix} \theta \\ x \end{pmatrix} \in \mathbb{R}^3 \right\},$$

$$\mathfrak{so}(2) = \left\{ \begin{pmatrix} 0 & -\theta & u_1 \\ \theta & 0 & u_2 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} \theta \\ u_1 \\ u_2 \end{pmatrix} \in \mathbb{R}^3 \right\},$$

$$\mathscr{L}_{\mathfrak{so}(2)} \begin{pmatrix} \theta \\ u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 & -\theta & u_1 \\ \theta & 0 & u_2 \\ 0 & 0 & 0 \end{pmatrix},$$

$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ denoting the rotation of angle θ, and the exponential mapping is:

$$\exp \begin{pmatrix} \theta \\ u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \theta \\ Re(z) \\ Im(z) \end{pmatrix}, \text{ where } z = (u_1 + iu_2) \frac{e^{i\theta} - 1}{i\theta}.$$

2.2 Group of rotation matrices $SO(3)$

G is here the group of rotation matrices $SO(3)$ and $\mathfrak{g}$ the space of skew-symmetric matrices $\mathfrak{so}(3)$:

$$SO(3) = \left\{ R \in \mathscr{M}_3(\mathbb{R}), RR^T = Id, \det(R) = 1 \right\},$$

$$\mathfrak{so}(3) = \left\{ A \in \mathscr{M}_3(\mathbb{R}), A = -A^T \right\},$$

$$\mathscr{L}_{\mathfrak{so}(3)} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}_\times = \begin{pmatrix} 0 & -\xi_3 & \xi_2 \\ \xi_3 & 0 & -\xi_1 \\ -\xi_2 & \xi_1 & 0 \end{pmatrix}.$$

The notation $(\cdot)_\times$ just defined will be used throughout this document. The exponential mapping is given by the formula:

$$\exp(\xi) = I_3 + \frac{\sin(||\xi||)}{||\xi||} (\xi)_\times + 2 \frac{\sin(||\xi||/2)^2}{||\xi||^2} (\xi)_\times^2.$$

2.3 Group of direct spatial isometries $SE(3)$

G is here the group of rigid motions $SE(3)$ and $\mathfrak{g} = \mathfrak{se}(3)$ its Lie algebra:

$$SE(3) = \left\{ \begin{pmatrix} R & x \\ 0_{1,3} & 1 \end{pmatrix}, R \in SO(3), x \in \mathbb{R}^3 \right\},$$

$$\mathfrak{se}(3) = \left\{ \begin{pmatrix} (\xi)_\times & x \\ 0_{1,3} & 1 \end{pmatrix}, \xi, x \in \mathbb{R}^3 \right\},$$

$$\mathscr{L}_{\mathfrak{se}(3)} \begin{pmatrix} \xi \\ x \end{pmatrix} = \begin{pmatrix} (\xi)_\times & x \\ 0_{1,3} & 0 \end{pmatrix}.$$

The exponential mapping is given by the formula:

$$\exp \begin{pmatrix} \xi \\ x \end{pmatrix} = I_4 + S + \frac{1-\cos(||\xi||)}{||\xi||^2})S^2 + \frac{||\xi|| - \sin(||\xi||)}{||\xi||^3} S^3,$$

where $S = \mathscr{L}_{\mathfrak{se}(3)} \begin{pmatrix} \xi \\ x \end{pmatrix}$.

2.4 Group of double direct isometries $SE_2(3)$

We introduce here a matrix Lie group proving very useful for inertial navigation and we call it "group of double direct isometries", with the notation $SE_2(3)$, for lack of existing conventions (to our best knowledge). Its Lie algebra will be denoted $\mathfrak{se}_2(3)$ and we have:

$$SE_2(3) = \left\{ \begin{pmatrix} R & v & x \\ 0_{1,3} & 1 & 0 \\ 0_{1,3} & 0 & 1 \end{pmatrix}, R \in SO(3), v, x \in \mathbb{R}^3 \right\},$$

$$\mathfrak{se}_2(3) = \left\{ \begin{pmatrix} (\xi)_\times & u & y \\ 0_{1,3} & 0 & 0 \\ 0_{1,3} & 0 & 0 \end{pmatrix}, \xi, u, y \in \mathbb{R}^3 \right\},$$

$$\mathscr{L}_{\mathfrak{se}_2(3)} \begin{pmatrix} \xi \\ u \\ y \end{pmatrix} = \begin{pmatrix} (\xi)_\times & u & y \\ 0_{1,3} & 0 & 0 \\ 0_{1,3} & 0 & 0 \end{pmatrix},$$

$$\exp \begin{pmatrix} \xi \\ u \\ y \end{pmatrix} = I_5 + S + \frac{1-\cos(||\xi||)}{||\xi||^2})S^2 + \frac{||\xi|| - \sin(||\xi||)}{||\xi||^3} S^3,$$

where $S = \mathscr{L}_{\mathfrak{se}_2(3)} \begin{pmatrix} \xi \\ u \\ y \end{pmatrix}$.

Table 2.1 – Formulas for most frequent matrix Lie groups in navigation

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

Part I

Invariant Kalman Filtering

Chapter 3

Autonomous error equations

Part of this chapter has been merged to Chapter 4 and submitted to a journal. The paper is currently under revision.

Chapter abstract In the present chapter we characterize dynamical systems on Lie groups that are such that the error (that is the difference in the sense of group multiplication) between two solutions follows an autonomous equation, a key property for the convergence results of the filters introduced in the next chapters. Previous publications noticed the error equation was autonomous for invariant systems [16, 19, 70] but did not try to characterize the systems having this property. We furthermore prove that for those systems the error surprisingly follows a linear differential equation, although the system at hand is non-linear. The two latter contributions are novel to our best knowledge.

3.1 Introduction

As stated in the introduction chapter, finding an error variable following an autonomous equation is very desirable in an estimation problem. The theory of invariant observers [16], starting from considerations regarding the symmetries of a system, has brought to bear the use of group operations on obtaining such a property. Consider indeed a system of the form:

$$\frac{d}{dt}\chi_t = \chi_t \omega_t,$$
$$Y_{t_n} = \chi_{t_n} \cdot b_n,$$

where χ_t belongs to a matrix Lie group G, the input ω_t to the associated Lie algebra $\mathfrak{g}$, $b_n \in \mathbb{R}^p$ is a known vector, and "$\cdot$" a group action. The reader feeling uncomfortable with this notion can assume "$\cdot$" is the classical product of a matrix and a vector. Define the error variable as $e_t = \chi_t^{-1}\hat{\chi}_t$ and the update as $\hat{\chi}_{t_n}^+ = \hat{\chi}_{t_n}\mathscr{K}_n(\hat{\chi}_t^{-1}Y_{t_n})$ where $\mathscr{K}_n$ can be any function of $\mathbb{R}^p \to G$. The evolution of e_t reads now:

$$\frac{d}{dt}e_t = -\omega_t e_t + e_t \omega_t,$$

$$e_{t_n}^+ = e_{t_n}\mathscr{K}_n(e_{t_n}^{-1} \cdot b_n).$$

We see the evolution of e_t is independent from the estimate. Thus, we are now able to characterize two families of systems on which an error variable with autonomous equation is defined: linear and invariant systems. Note that to obtain autonomous error equation, the innovation has been defined in a specific way: $z = \hat{\chi}_t^{-1} \cdot Y = e_{t_n}^{-1} \cdot b_n$, making it only a function of the error. Remark 1 below should convince the reader that a new class of estimation problems allowing autonomous error equation must exist, and contain both invariant and linear systems.

Remark 1. *Additive linear systems of the form $\frac{d}{dt}X_t = u_t$ are translation-invariant, but general linear systems are not a specific case of invariant systems, not more than invariant systems are a specific case of linear systems.*

Inspired by the latter example, we consider in this chapter systems defined on Lie groups and error variables of one of the two following forms:

$$e_t = \eta_t^L = \chi_t^{-1}\hat{\chi}_t \qquad \text{(left invariant)}, \tag{3.1}$$

$$e_t = \eta_t^R = \hat{\chi}_t\chi_t^{-1} \qquad \text{(right invariant)}. \tag{3.2}$$

The superscripts L and R refer to "left-invariant" and "right-invariant" as in [12]. We will characterize a broad class of systems (containing invariant *and* linear systems) where the innovation and error propagation depend only on the error variable. We look for them under the following general form:

$$\begin{aligned} \frac{d}{dt}\chi_t &= f_{u_t}(\chi_t), \\ Y_n &= h(\chi_{t_n}), \end{aligned} \tag{3.3}$$

where χ_t is the state of the system and Y the observation. Definition 1 has the drawback to refer to a specific observer. Yet, the property of autonomous error equation can surprisingly be stated only as a characteristic of the system and the error variable. This is what we show now, defining first the two following properties:

Definition 3 (Autonomous error propagation, Autonomous innovation)**.** *Let X_t be a state vector having an evolution of the form:*

$$\frac{d}{dt}X_t = f_{u_t}(X_t),$$

where u_t is a sequence of inputs, and let h be an observation function measured at instants $(t_n)_{n\geqslant 0}$:

$$Y_n = h(X_{t_n}).$$

An error variable e_t is defined as a function ε of the true state X_t and the estimate $\hat{X}_t$: $e_t = \varepsilon(X_t, \hat{X}_t)$, smooth and invertible with respect to each of its variables and equal to zero (or to the identity element on a Lie group) if its two arguments are identical. We define as follows two desirable properties of ε:

Autonomous error propagation: *ε is told to have autonomous propagation if there exist a function g such that for any couple of solutions $X, \hat{X}$ of the dynamics, $e_t = \varepsilon(X, \hat{X})$ verifies:*

$$\frac{d}{dt}e_t = g_{u_t}(e_t).$$

Autonomous innovation: *ε is told to ensure autonomous innovation if there exist two functions ψ_1 and ψ_2 such that for any couple of states $(X_t, \hat{X}_t)$ we have:*

$$\psi_1\left(\hat{X}_t, h(X_t)\right) = \psi_2\left(\varepsilon(X_t, \hat{X}_t)\right),$$

ψ_1 being invertible with respect to its second variable.

Remark 2. *The idea of autonomous error propagation is quite transparent: the evolution of the error e_t in the absence of observation depends on $\hat{X}_t$ and X_t only through the value of e_t itself. But the introduction of an "autonomous innovation" may deserve more explanation. Innovation measures the discrepancy between the expected observation $h(\hat{X}_t)$ and the true observation $h(X_t)$. We denote it by $\psi_1(\hat{X}_t, h(X_t))$ (instead of simply $\psi_1\left(h(\hat{X}_t), h(X_t)\right)$) to allow any additional dependency in the known variable $\hat{X}_t$. We define this innovation function as "autonomous" if the result depends on $\hat{X}_t$ and X_t only through the value of e_t. The point of introducing this notion here is to remove any reference to a specific observer. Instead, we defined the two basic ingredients allowing to build observers having autonomous error equation.*

In the context of multiplicative errors, definitions 3 and 1 are equivalent in the sense of Proposition 4:

Proposition 4. *If a system has autonomous error propagation and autonomous innovation as defined in Definition 3 then an observer having autonomous error equation can be build. If an observer having autonomous error equation can be build then the system has autonomous error propagation and autonomous innovation as defined in Definition 3.*

Proof. The error is assumed right-invariant ($\varepsilon(\chi_t, \hat{\chi}_t) = \hat{\chi}_t \chi_t^{-1}$), the proof being similar for a left-invariant error. Assume the system verifies the properties introduced in Definition 3. Then an observer with update step of the form $\hat{\chi}_{t_n}^+ = \mathscr{K}(\psi_1(\hat{\chi}_t, h(\chi_t)))\hat{\chi}_{t_n}$ has autonomous error equation. To show the converse, let ψ and ψ^e denote the update functions of definition 1(we removed the inputs, any sequence can be chosen here). Autonomy of the update step reads: $\psi(\hat{\chi}_t, h(\chi_t))\chi_t^{-1} = \psi^e(\hat{\chi}_t \chi_t^{-1})$. Thus function $\psi_1(\hat{\chi}_t, h(\chi_t)) = \psi(\hat{\chi}_t, h(\chi_t))\hat{\chi}_t^{-1}$ appears to be an autonomous innovation function:

$$\psi(\hat{\chi}_t, h(\chi_t))\hat{\chi}_t^{-1} = \psi(\hat{\chi}_t, h(\chi_t))\chi_t^{-1}\chi_t\hat{\chi}_t^{-1} = \psi^e(e)e^{-1}.$$

□

The characterization of dynamics f_{u_t} having autonomous error propagation on one hand, and the observation functions h ensuring autonomous innovation on the other hand, are independent and will be dealt with separately, in Section 3.2 for autonomous innovation, then Sections 3.3, 3.4 3.5 for autonomous propagation. Section 3.3 gives a first characterization, surprisingly simple although not allowing an exhaustive classification. Section 3.4 derives a fundamental log-linearity property of the systems having autonomous propagation. It will be advantageously leveraged in Chapter 4, but first used in Section 3.5 to link our study of the systems having autonomous propagation to the classification of Lie algebra homomorphisms.

Remark 3. *The reader could be wondering how restrictive is the framework we just adopted to study autonomous error equations. Although we don't claim perfect rigor here, we will try to give an argument showing Lie group structures and multiplicative errors arise quite naturally if error variables are at issue. Here the state χ belongs to any space $\mathscr{X}$ and we introduce an additional property of the error function: for any $\chi_1, \chi_2, \chi_3, \gamma_1, \gamma_2, \gamma_3 \in \mathscr{X}$ we assume*

$$\varepsilon(\chi_1, \chi_2) = \varepsilon(\gamma_1, \gamma_2) \text{ and } \varepsilon(\chi_2, \chi_3) = \varepsilon(\gamma_2, \gamma_3) \Rightarrow \varepsilon(\chi_1, \chi_3) = \varepsilon(\gamma_1, \gamma_3). \tag{3.4}$$

This property means that you can deduce the error between χ_1 and χ_3 from the error between χ_1 and χ_2 on one hand, χ_2 and χ_3 on the other hand, without knowing the value of χ_1, χ_2 or χ_3. In other words the errors can be composed defining a group law: for any errors e_1 and e_2 we choose any element χ_1 of $\mathscr{X}$, another one denoted by χ_2 such as $\varepsilon(\chi_1, \chi_2) = e_1$ (whose existence is ensured by the invertibility of ε w.r.t. its second variable) and a third one denoted by χ_3 such as $\varepsilon(\chi_2, \chi_3) = e_2$. The product $e_1 \cdot e_2$ is then defined by $\varepsilon(\chi_1, \chi_3)$. Hypothesis (3.4) *ensures the result is independent from the choice of χ_1 and χ_2. Error zero is the identity element, the inverse is defined by $\varepsilon(\chi_1, \chi_2)^{-1} = \varepsilon(\chi_2, \chi_1)$ and associativity is obtained considering a sequence of four elements $\chi_1, \chi_2, \chi_3, \chi_4$. This law is then transported on the state space $\mathscr{X}$: we can choose any point χ_0 as a reference to embed the state space into the error space using the bijection $\phi : \chi \to \varepsilon(\chi_0, \chi)$. The error becomes $\varepsilon(\chi_1, \chi_2) = \varepsilon(\chi_1, \chi_0) \cdot \varepsilon(\chi_0, \chi_2) = \varepsilon(\chi_0, \chi_1)^{-1} \cdot \varepsilon(\chi_0, \chi_2) = \chi_1^{-1} \cdot \chi_2$ and takes the general form introduced above.*

3.2 Autonomous innovation functions

Characterization of the observation functions ensuring autonomous innovation is the easier part of the problem as shown by theorem 1 below:

Theorem 1. *The observation functions $h(\chi)$ ensuring the existence of an autonomous innovation for the error $e = \chi^{-1}\hat{\chi}$ (resp.$e = \hat{\chi}\chi^{-1}$), i.e. two functions ψ_1 and ψ_2 such that*

$$\forall \chi, \hat{\chi} \in G, \psi_1(\hat{\chi}, h(\chi)) = \psi_2(\chi^{-1}\hat{\chi}), \tag{3.5}$$

are the left-actions (resp. right-actions) of G on a space $\mathscr{Y}$. Moreover, the associated autonomous innovation is then defined by $\psi(\hat{\chi}, Y) = \hat{\chi}^{-1}Y$ (resp. $\psi(\hat{\chi}, Y) = \hat{\chi}Y$).

To prove Theorem 1 we need the following lemma:

Lemma 1. *Let $\mathscr{Y}$ denote any set and $\psi_x : \mathscr{Y} \to \mathscr{Y}$ a function from $\mathscr{Y}$ to $\mathscr{Y}$ defined for any $x \in G$ (with $\psi_{Id} = Id$). If for an element $y_0 \in \mathscr{Y}$ the relation $\forall a, b \in G, \psi_a(\psi_b(y_0)) = \psi_{ba}(y_0)$ is verified then ψ is a right action of G over $\psi_G(y_0)$.*

Proof. The proof is straightforward: let $y \in \psi_G(y_0)$ and $c \in G$ such that $y = \psi_c(y_0)$. For any $a, b \in G$ we have $\psi_a(\psi_b(y)) = \psi_a(\psi_b(\psi_c(y_0))) = \psi_a(\psi_{cb}(y_0)) = \psi_{cba}(y_0) = \psi_{ba}(\psi_c(y_0)) = \psi_{ba}(y)$. □

Now we can prove Theorem 1:

Proof. Let $\mathscr{Y}$ denote the observation space and $h : G \to \mathscr{Y}$ the observation function. Assume h ensures the existence of an autonomous innovation. Then for any $\chi, \hat{\chi} \in G$ we have $\psi_1(\hat{\chi}, h(\chi)) = \psi_2(\chi^{-1}\hat{\chi})$. In particular, $\chi = Id$ gives

$$\forall \hat{\chi}, \psi_1(\hat{\chi}, h(Id)) = \psi_2(\hat{\chi}). \tag{3.6}$$

Thus $\psi_2(\chi)$ can be replaced by $\psi_1(\chi, h(Id))$ for any $\chi \in G$ and, denoting $\psi_1(\chi, y)$ by $\psi_\chi(y)$ in (3.5) then applying (3.6) the condition becomes:

$$\forall \chi, \hat{\chi}, \psi_{\hat{\chi}}(h(\chi)) = \psi_{\chi^{-1}\hat{\chi}}(h(Id)). \tag{3.7}$$

For $\hat{\chi} = Id$ we get $\psi_{Id}(h(\chi)) = \psi_{\chi^{-1}}(h(Id))$ thus $h(\chi) = \psi_{Id}^{-1} \circ \psi_{\chi^{-1}} \circ h(Id)$. Reintroducing this equality in (3.7) it becomes:

$$\forall \chi, \hat{\chi}, \psi_{\hat{\chi}} \circ \psi_{Id}^{-1} \circ \psi_{\chi^{-1}} \circ h(Id) = \psi_{\chi^{-1}\hat{\chi}} \circ h(Id).$$

Thus $\forall \chi, \hat{\chi}, (\psi_{Id}^{-1} \circ \psi_{\hat{\chi}}) \circ (\psi_{Id}^{-1} \circ \psi_{\chi^{-1}}) \circ h(Id) = (\psi_{Id}^{-1} \circ \psi_{\chi^{-1}\hat{\chi}}) \circ h(Id)$ which makes $\psi \to \psi_{Id}^{-1} \circ \psi_\chi$ a right-action of G over $Im(h)$ (see Lemma (3.7)), and h a left-action over $Im(h)$. The proof for the right-invariant error $e = \eta^R = \hat{\chi}\chi^{-1}$ is similar. □

3.3 Autonomous error propagation

To obtain autonomous error equations, the second property we seek is autonomy of the errors evolution, i.e. the existence of a function g_{u_t} (independent from χ or $\hat{\chi}$) such that the error evolution reads $\frac{d}{dt}e_t = g_{u_t}(e_t)$, with e_t defined as either $\chi_t^{-1}\hat{\chi}_t$ or $\hat{\chi}_t\chi_t^{-1}$. It is verified by a very specific class of systems, as shown by Theorem 2. We recall before the property we are seeking:

Definition 4. *The left-invariant and right-invariant errors are said to have autonomous propagation if they satisfy a differential equation of the form* $\frac{d}{dt}e_t = g_{u_t}(e_t)$.

Note that, in general the time derivative of e_t is a complicated function depending on u_t and both χ_t^{-1} and $\hat{\chi}_t$ in a way that does not boil down to a function of e_t. The following result allows to characterize the class of systems of the form (3.3) for which the property holds.

Theorem 2. *The three following conditions are equivalent for the dynamics* (3.3)*:*

1. *The left-invariant error has autonomous propagation*
2. *The right-invariant error has autonomous propagation*
3. *For all* $t > 0$ *and* $a, b \in G$ *we have (in the tangent space at* ab*):*

$$f_{u_t}(ab) = f_{u_t}(a)b + af_{u_t}(b) - af_{u_t}(I_d)b. \tag{3.8}$$

Moreover, if one of these conditions is satisfied we have

$$\frac{d}{dt}\eta_t^L = g_{u_t}^L(\eta_t^L) \quad \textit{where} \quad g_{u_t}^L(\eta) = f_{u_t}(\eta) - f_{u_t}(I_d)\eta, \tag{3.9}$$

$$\frac{d}{dt}\eta_t^R = g_{u_t}^R(\eta_t^R) \quad \textit{where} \quad g_{u_t}^R(\eta) = f_{u_t}(\eta_)) - \eta f_{u_t}(I_d). \tag{3.10}$$

Proof. Assume we have $\frac{d}{dt}\eta_t^L = g_{u_t}(\eta_t^L)$ for a certain function g_{u_t} and any process of the form $\eta_t^L = \chi_t^{-1}\hat{\chi}_t$, where χ_t and $\hat{\chi}_t$ are solutions of (3.3). We have:

$$\begin{aligned}
g_{u_t}(\chi_t^{-1}\hat{\chi}_t) &= \frac{d}{dt}(\chi_t^{-1}\hat{\chi}_t) \\
&= -\chi_t^{-1}\left[\frac{d}{dt}\chi_t\right]\chi_t^{-1}\hat{\chi}_t + \chi_t^{-1}\frac{d}{dt}\hat{\chi}_t \\
&= -\chi_t^{-1}f_{u_t}(\chi_t)\eta_t^L + \chi_t^{-1}f_{u_t}(\hat{\chi}_t) \\
g_{u_t}(\eta_t^L) &= -\chi_t^{-1}f_{u_t}(\chi_t)\eta_t^L + \chi_t^{-1}f_{u_t}(\chi_t\eta_t^L).
\end{aligned} \tag{3.11}$$

This has to hold for any χ_t and η_t. In the particular case where $\chi_t = I_d$ we obtain:

$$g_{u_t}(\eta_t^L) = f_{u_t}(\eta_t^L) - f_{u_t}(I_d)\eta_t^L. \tag{3.12}$$

Re-injecting (3.12) in (3.11) we obtain:

$$f_{u_t}(\chi_t\eta_t) = f_{u_t}(\chi_t)\eta_t + \chi_t f_{u_t}(\eta_t) - \chi_t f_{u_t}(I_d)\eta_t.$$

The converse is trivial and the proof is similar for right-invariant errors. □

The class of systems defined by Equation (3.8) has many similarities with linear systems, as will be shown in the sequel. It generalizes them in a way that preserves some interesting properties.

Remark 4. *The particular cases of left-invariant and right-invariant dynamics, or the combination of both as follows, verify the hypothesis of Theorem 2. Let* $f_{v_t,\omega_t}(\chi) = v_t\chi + \chi\omega_t$. *We have indeed:*

$$\begin{aligned} f_{v_t,\omega_t}(a)b + af_{v_t,\omega_t}(b) - af_{v_t,\omega_t}(I_d)b =& (v_t a + a\omega_t)b + a(v_t b + b\omega_t) - a(v_t + \omega_t)b \\ =& u_t ab + ab\omega_t \\ =& f_{v_t,\omega_t t}(ab). \end{aligned}$$

Remark 5. *In the particular case where* G *is a vector space with standard addition as the group composition law, the condition* (3.8) *boils down to* $f_{u_t}(a+b) = f_{u_t}(a) + f_{u_t}(b) - f_{u_t}(0)$ *and we recover the affine functions.*

3.4 A log-linearity property of errors having autonomous propagation

The general solution of a linear system of the form $\frac{d}{dt}X_t = A_t X_t + B_t$ is notoriously the sum of a particular solution $\bar{X}_t$ and a solution ξ_t of the homogeneous linear system $\frac{d}{dt}\xi_t = A_t\xi_t$. In the broader case defined by Theorem 2, a surprising generalization can be derived: the general solution of the system is the product of a specific solution by the exponential of the solution of a *linear* homogeneous system (Corollary 1).

In the sequel, we will systematically consider systems of the form (3.3) with the additional property (3.8), i.e. systems defined by

$$\begin{aligned} \frac{d}{dt}\chi_t &= f_{u_t}(\chi_t), \\ \text{where } \forall(u,a,b) \quad f_u(ab) &= af_u(b) + f_u(a)b - af_u(Id)b. \end{aligned} \tag{3.13}$$

For such systems, Theorem 2 proves that the left (resp. right) invariant error is a solution to the equation $\frac{d}{dt}\eta_t = g_{u_t}(\eta_t)$ where g_{u_t} is given by (3.9) (resp. (3.10)). For small errors η_t, that is η_t close to Id, the error system can be linearized as follows. As we have $g_{u_t}(Id) \equiv 0$, the Lie derivative of the vector field g_{u_t} along any other vector field only depends on the value of the vector field at Id. It can thus be identified to a linear operator (thus a matrix) in the tangent space at Id, and will be denoted by $Dg_{u_t}(Id)$. Furthermore, a small error η_t can be identified to an element of $\mathbb{R}^{\dim\mathfrak{g}}$ through the exponential map, that is $\eta_t = \exp(\xi_t)$ (see Chapter 2 for a definition of the exponential map). The linearized error equation thus reads $\frac{d}{dt}\xi_t = Dg_{u_t}(Id)\xi_t$. The next striking result proves this first order Taylor expansion of the non linear error system captures in fact the whole behavior of the error propagation.

Theorem 3. *Consider the left or right invariant error η_t^i as defined by* (3.1) *or* (3.2) *between two trajectories of* (3.13)*, the superscript i denoting indifferently L or R, and assume η_t^i is in the image of the exponential mapping (this image is the whole group G in all the examples considered in this document). Let $\xi_0^i \in \mathbb{R}^{dim\,\mathfrak{g}}$ be such that $\exp(\xi_0^i) = \eta_0^i$. If ξ_t^i is defined for $t > 0$ by the* linear *differential equation in $\mathbb{R}^{\dim \mathfrak{g}}$*

$$\frac{d}{dt}\xi_t^i = A_t^i \xi_t^i, \quad \textit{where} \quad A_t^i := Dg_{u_t}^i(Id),$$

then we have:

$$\forall t \geq 0 \quad \eta_t^i = \exp(\xi_t^i).$$

Recalling the definition of the invariant errors, the following corollary illustrated by Fig. 3.1 is easily derived.

Corollary 1. *Let $\hat{\chi}_t$ denote a particular solution of* (3.13)*. If the exponential mapping is surjective then the general solution of the system has the two following characterizations:*

- $\chi_t = \exp(F_t^R \xi_0)\hat{\chi}_t$*, where F_t^R is defined by $F_0^R = Id$ and $\frac{d}{dt}F_t^R = Dg_{u_t}^R(Id)F_t^R$.*
- $\chi_t = \hat{\chi}_t \exp(F_t^L \xi_0)$*, where F_t^L is defined by $F_0^L = Id$ and $\frac{d}{dt}F_t^L = Dg_{u_t}^L(Id)F_t^L$.*

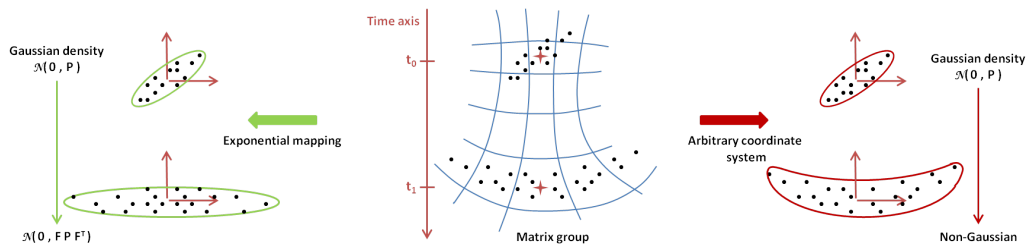


Figure 3.1 – Example illustration of the log-linear property of the error propagation. Central plot: a cloud of points centered on $\hat{\chi}_{t_0}$ having the property of being a Gaussian in the tangent space at $\hat{\chi}_{t_0}$ with covariance matrix P is transported by the flow (3.13) over the matrix group. Right plot: the cloud in the tangent space loses its Gaussianity due to the nonlinearity of the flow. Left plot: when using the exponential to map the tangent space to the manifold, we see at $t = t_1$ the cloud remains a Gaussian centered at $\hat{\chi}_{t_1}$, and the covariance matrix becomes $F_{t_1} P F_{t_1}^T$ where F_{t_1} is defined in Cor 1.

The last two properties are trivial in the linear case as the exponential associated to the additive structure is the identity function. But they show that a wide range of nonlinear problems (see examples of Chapter 4 Section 4.3) can lead to linear error equations provided the error variable is correctly chosen. The result implies that the propagation step of an EKF making use of the exponential mapping and the invariant errors is expected to reflect the true evolution of the error variable. This will be extensively used in Chapter 4 to prove stability properties of Invariant Extended Kalman Filters.

The proof of Theorem 3 is based upon the following lemmas.

Lemma 2. *Consider the system* (3.13) *and let* $\hat{\chi}_t$ *denote a particular solution. Consider the condition*

$$\forall a,b \in G, \quad g_u(ab) = ag_u(b) + g_u(a)b. \tag{3.14}$$

We have the following properties:

- *The function* $g_u^R(\eta) = f_{u_t}(\eta) - \eta f_{u_t}(Id)$ *verifies* (3.14) *and all the solutions of* (3.13) *have the form* $\chi_t = \eta_t^R \hat{\chi}_t$*, where* η_t^R *verifies* $\frac{d}{dt}\eta_t^R = g_{u_t}^R(\eta_t^R)$.
- *The function* $g_{u_t}^L(\eta) = f_{u_t}(\eta) - f_{u_t}(Id)\eta$ *verifies* (3.14) *and all the solutions of* (3.13) *have the form* $\chi_t = \hat{\chi}_t \eta_t^L$*, where* η_t^L *verifies* $\frac{d}{dt}\eta_t^L = g_{u_t}^L(\eta_t^L)$.

The verification of these two properties is trivial. The functions g_{u_t} governing the errors propagation turn out to possess an interesting property:

Lemma 3. *Let* Φ_t *be the flow associated to the system* $\frac{d}{dt}\eta_t = g_{u_t}(\eta_t)$*, where* g_{u_t} *verifies* (3.14)*. Then:*

$$\forall \eta_0, \eta_0' \in G, \forall p \in \mathbb{Z}, \Phi_t(\eta_0\eta_0') = \Phi_t(\eta_0)\Phi_t(\eta_0').$$

Proof. We simply have to see that $\Phi_t(\eta_0)\Phi_t(\eta_0')$ is solution of the system $\frac{d}{dt}\eta_t = g_{u_t}(\eta_t)$:

$$\frac{d}{dt}\left[\Phi_t(\eta_0)\Phi_t(\eta_0')\right] = g_{u_t}(\Phi_t(\eta_0))\Phi_t(\eta_0') + \Phi_t(\eta_0)g_{u_t}(\Phi_t(\eta_0')) = g_{u_t}(\Phi_t(\eta_0)\Phi_t(\eta_0')).$$

□

An immediate recursion gives then:

Lemma 4. *For any* $\eta_0 \in G$ *and* $p \in \mathbb{Z}$ *we have:*

$$\Phi_t(\eta_0^p) = \Phi_t(\eta_0)^p.$$

Lemmas 3 and 4 indicate the behavior of the flow infinitely close to Id dictates its behavior arbitrarily far from it, as the flow commutes with exponentiation (exponentiation in the sense of the iterated multiplication of an element of a group by itself). The use of the exponential (map) thus allows to derive an infinitesimal version of the Lemma 4, which is an equivalent formulation of Theorem 3 and its corollary.

Theorem 4. *Let* Φ_t *be the flow associated to the system* $\frac{d}{dt}\eta_t = g_{u_t}(\eta_t)$ *satisfying* (3.14)*. As we have* $g_{u_t}(Id) = 0$*, the Lie derivative of* g_{u_t} *along a vector field* v *verifying* $v(Id) = \xi_0$ *only depends on* ξ_0 *and will be denoted by* $Dg_{u_t}\xi_0$*. We have:*

$$\Phi_t(\exp(\xi_0)) = \exp(F_t\xi_0).$$

where F_t *is the solution of the matrix equation* $F_0 = Id, \frac{d}{dt}F_t = Dg_{u_t}F_t$.

Proof. Thanks to Lemma 4 we have

$$\Phi_t\left(e^{\xi_0}\right)=\Phi_t\left(\left[e^{\frac{1}{n}\xi_0}\right]^n\right)=\Phi_t\left(e^{\frac{1}{n}\xi_0}\right)^n=\left[e^{\frac{1}{n}D\Phi_t|_{Id}\xi_0+r_t\left(\frac{1}{n}\xi_0\right)}\right]^n,$$

where $r_t\left(\frac{1}{n}\xi_0\right)$ is a quadratic term, which ensures in turn $\left[e^{\frac{1}{n}D\Phi_t|_{Id}\xi_0+r_t\left(\frac{1}{n}\xi_0\right)}\right]^n=e^{D\Phi_t|_{Id}\xi_0+\frac{1}{n}r_t(\xi_0)}$. Letting $n\to\infty$ we get $\Phi_t\left(e^{\xi_0}\right)=e^{D\Phi_t|_{Id}\xi_0}$. We then introduce a parameterized curve $\psi_h(Id)$ on G such that $\psi_0(Id)=Id$ and $\frac{d}{dh}\psi_h(Id)=\xi_0$. We then depart from the definition of the flow Φ_t starting at $\psi_h(Id)$:

$$\left.\frac{d}{dt}\right|_{t=s}\left[\Phi_t\left(\psi_h(Id)\right)\right]=g_{u_s}\left(\Phi_s\left(\psi_h(Id)\right)\right).$$

Using $\frac{d}{dh}\Phi_t\left(\psi_h(Id)\right)=D\Phi_t|_{Id}\xi_0$ and $\frac{d}{dh}g_{u_s}\left(\Phi_s\left(\psi_h(Id)\right)\right)=Dg_{u_s}D\Phi_s|_{Id}\xi_0$ we get the final desired result

$$\frac{d}{dt}D\Phi_t|_{Id}=Dg_{u_t}D\Phi_t|_{Id},$$

letting F_t be equal to $D\Phi_t|_{Id}$. □

3.5 A complete characterization of systems with autonomous errors

Relation (3.8) can be easily checked and is thus a satisfactory alternative definition of systems having autonomous error propagation. But it is not a characterization as it does not give a complete list of the vector fields f verifying

$$\forall a,b, f(ab)=f(a)b+af(b)-af(Id)b. \tag{3.15}$$

This set will be denoted $\mathscr{F}$ and its characterization is the purpose of the present section.

Proposition 5. *Relation* (3.15) *being linear, the classical addition of vector fields makes $\mathscr{F}$ a vector space.*

To formulate the next result, we need to introduce some new objects and definitions.

Definition 5 (Lie algebra homomorphism)**.** *Given a Lie algebra $\mathfrak{a}$ with Lie bracket $[\cdot,\cdot]_{\mathfrak{a}}$, an homomorphism a of $\mathfrak{a}$ is a linear and invertible function from $\mathfrak{a}$ to $\mathfrak{a}$ ensuring:*

$$\forall\xi_1,\xi_2\in\mathfrak{a},[a\xi_1,a\xi_2]=a[\xi_1,\xi_2].$$

The homomorphisms of a Lie algebra $\mathfrak{a}$ constitute a Lie group A' whose operation is the operators composition.

The group of homomorphisms of $\mathfrak{g}$ will be denoted by G', and its Lie algebra by $\mathfrak{g}'$. These notations allow us to give a full characterization of the vector fields f verifying (3.15):

Theorem 5. *Assume G is connected. The functions $\mathfrak{F}_L$ and $\mathfrak{F}_R$ below are two linear injections of $\mathscr{F}$ into $\mathfrak{g}\times\mathfrak{g}'$:*

$$\mathfrak{F}_L : \mathscr{F} \to \mathfrak{g}\times\mathfrak{g}' \quad , \quad \mathfrak{F}_L : f \to (f(0), D|_{x=0}[f(x) - xf(0)]),$$
$$\mathfrak{F}_R : \mathscr{F} \to \mathfrak{g}\times\mathfrak{g}' \quad , \quad \mathfrak{F}_R : f \to (f(0), D|_{x=0}[f(x) - f(0)x]).$$

- *If G is simply connected then $\mathfrak{F}_L$ and $\mathfrak{F}_R$ are bijections.*
- *Otherwise, for any element $A \in (\mathfrak{g}\times\mathfrak{g}')$ there exists a vector field f defined on a neighborhood of Id such that $\mathfrak{F}_L(f) = A$ (resp. $\mathfrak{F}_R(f) = A$), and verifying* (3.15) *for a and b small enough.*

If defined, the result of the inverse operator $\mathfrak{F}_L^{-1}(\alpha,\check{\alpha})$ (resp. $\mathfrak{F}_R^{-1}(\alpha,\check{\alpha})$) is the vector field $x \to x\alpha + D\exp_{\log x}(\check{\alpha}\log x)$ (resp. $x \to \alpha x + D\exp_{\log x}(\check{\alpha}\log x)$).

If G is $\mathbb{R}^n$ ($n > 0$) with the vector sum as group operation, we have $\mathfrak{g} = \mathbb{R}$, $G' = Gl_n(\mathbb{R})$ (group of invertible matrices of size $n\times n$) and $\mathfrak{g}' = \mathscr{M}_n(\mathscr{R})$ (space of matrices of size $n\times n$). Theorem 5 boils down to the proposition stated in example 2:

Example 2. *The vector fields f on $\mathbb{R}^n$ verifying $\forall a,b \in \mathbb{R}^n, f(a+b) = f(a)+f(b)-f(0)$ are the functions of the form:*

$$f(X) = AX + B,$$

with $A \in \mathscr{M}_n(\mathbb{R})$ and $B \in \mathbb{R}^n$.

We see again that vector fields verifying (3.15) are very similar to affine functions, although defined on a general Lie group instead of $\mathbb{R}^n$.

The proof of theorem 5 is done here only for $\mathfrak{F}_L$, it would be similar for $\mathscr{F}_R$. It consists of two steps:

1. The operator $\mathfrak{F}_L(f) : f \to \big(f(0), g^L\big)$, where g^L is defined as in Lemma 2 by $\forall x \in G, g^L(x) = f(x) - xf(Id)$, is a bijection from $\mathscr{F}$ to $\big(\mathfrak{g}\times\mathscr{G}\big)$, where $\mathscr{G}$ is the space of vector fields verifying (3.14).
2. The derivative at Id is an injection of $\mathscr{G}$ into $\mathfrak{g}'$

1 is verified easily. To show 2 we first characterize the flows Φ_t induced by the vector fields verifying (3.14). Lemma 3 shows that Φ_t is a group homomorphism of G at any $t > 0$. As a consequence we have:

Proposition 6. *At any time $t > 0$, the differential $D\Phi_t|_{Id} = F_t$ is a Lie algebra homomorphism of $\mathfrak{g}$, i.e. verifies:*

$$\forall \xi_1,\xi_2 \in \mathfrak{g}, [F_t\xi_1, F_t\xi_2] = F_t[\xi_1,\xi_2].$$

Proof. Lemma 3 shows that Φ_t is a group homomorphism of G at any $t > 0$. As a consequence, F_t is a Lie Algebra homomorphism of $\mathfrak{g}$. This result can be either obtained

replacing Φ_t in Lemma 3 by its exponential form given in Theorem 4, then developing the Baker-Campbell-Hausdorff (BCH) formula up to the second-order, or found directly in [67] for instance. $\square$

We just showed that $D\Phi_t|_{Id}$ is in G'. As it is also solution of $\frac{d}{dt}D\Phi_t|_{Id} = Dg|_{Id}D\Phi_t|_{Id}$, the differential $Dg|_{Id}$ is its left-trivialized derivative and is thus in $\mathfrak{g}'$. To show the differential at Id is injective we consider a vector field $g \in \mathscr{G}$ verifying $Dg|_{Id} = 0$. According to Theorem 4 the flow of g is constant over time in an open neighborhood of Id. As this flow is a group homomorphism the property extends to the whole group out of connexity. Deriving the flow w.r.t. the time we get $\forall x \in G, g(x) = 0$ which proves the injectivity. To obtain the reciprocal we first consider an element α of $\mathfrak{g}'$ and define the (linear) flow A_t by $A_0 = Id, \frac{d}{dt}A_t = \check{\alpha}A_t$. A_t is an homomorphism of $\mathfrak{g}$. For $x \in G$ close to Id we consider now $\Phi_t(x) = \exp(A_t \log(x))$. If $x_1, A_t x_1, A_t x_2$ and x_2 are in the convergence radius of $\exp$ we have:

$$\begin{aligned}
\Phi_t(x_1)\Phi_t(x_2) &= \exp\big(A_t \log(x_1)\big) \exp\big(A_t \log(x_2)\big) \\
&= \exp\big(BCH\left(A_t \log(x_1), A_t \log(x_2)\right)\big) \\
&= \exp\big(A_t BCH(\log(x_1), \log(x_2))\big) \\
&= \exp\big(A_t \log(x_1 x_2)\big) \\
&= \Phi_t\big(x_1 x_2\big).
\end{aligned}$$

We used here the fact that A_t, as an homomorphism of $\mathfrak{g}$, commutes with all the terms of the BCH formula. The last step is simply to derive Φ_t w.r.t. time to obtain a function of $\mathscr{G}$ defined around Id. The result extends to the whole group G only if the exponential map associates a Lie group homomorphism of G to each homomorphism A_t of $\mathfrak{g}$. This is ensured if G is connected and simply connected (see Theorem 3.41 of [67] for instance).

3.6 Conclusion

This chapter characterized a class of problems for which the error variable can be designed to have autonomous propagation. For these systems, the error variable was showed to have even linear propagation if re-defined through a logarithmic map. Some concrete examples of such systems, mostly related to navigation, are described in Chapter 4. The advantages of such an approach are obvious, regardless of the chosen estimation method. If an observer relies on pre-computed gains, the dependence on the specific trajectory has to be made as weak as possible. If the error equation is linearized for an EKF-like on-line gain tuning, this operation becomes independent from the estimated state and deriving stability properties becomes possible (see Chapter 4). Moreover, the discrepancy between the system linearized over its true trajectory and the system linearized on the current estimate disappears, which Chapter 9 will prove to be of dramatic importance. An EKF for this class of systems will be studied in the next chapter.

Chapter 4

Invariant Extended Kalman Filtering

Chapter abstract The previous chapter proved a certain class of estimation problems to be of specific interest as they allow designing observers ensuring the evolution of the error variable to be fully independent from the estimate. In this chapter, we consider a *deterministic setting*, and we use a first-order expansion of the error to derive a generalization of the Invariant Extended Kalman Filter (IEKF) introduced in [14]. The terminology IEKF is retained although it is designed for a larger class of systems, and we allow for the tuning matrices (interpreted as noise covariance matrices) to depend on the filter's trajectory, as opposed to [14, 17] that presuppose fixed covariance matrices. The resulting IEKF is thus more suited to real implementation where the covariance of the noises may depend on the trajectory. We derive a novel local convergence result of the IEKF viewed as an observer, building upon the log-linearity property derived in Section 3.4 of Chapter 3. We provide a long list of challenging navigation examples and for each of them the IEKF is proved to converge around *any* trajectory under standard observablity conditions, a generally difficult to prove property in a non-linear setting.

4.1 Introduction

The multiplicative difference between two solutions χ_t and $\hat{\chi}_t$ of the systems introduced in Chapter 3, defined as $\chi_t^{-1}\hat{\chi}_t$ or $\hat{\chi}_t\chi_t^{-1}$, has been shown to have extremely convenient properties. In this Chapter, we put them to good use extending the IEKF (proposed in [14] for invariant systems) to this broader class, and improving it by the introduction of an exponential mapping in the update step (see Chapter 2), whose benefits will be discussed at length in Chapters 9 and 10. A novel stability result based on the property obtained in Section 3.4 of Chapter 3 is derived for this EKF-like method, whose assumptions are exactly those used in [38] for linear Kalman filtering. Of course, the computations involved in the proof proposed in this latter paper have been of great help here. The difference with the linear case is that global convergence is replaced by local stability. Yet, only first-order observability assumptions are necessary. This is a decisive difference with most of the previous work on convergence of the EKF, where hypotheses regarding higher orders of the system and/or global behavior of the *estimated* error covariance matrix are involved [23, 24, 92, 102]. Their verification is usually extremely difficult on concrete sys-

tems, to the opposite of the result derived here, easily applied to a long list of examples. We divided the exposure into two sections. Section 4.2 defines the framework, derives the equations of the IEKF then gives a novel stability property based on the results of Chapter 3. Section 4.3 details the computations in several specific cases corresponding to classical navigation problems, and applies each time the result of 4.2 to guarantee stability of the method.

4.2 Full noisy system

We first consider a noisy equation on a matrix Lie group $G \subset \mathbb{R}^{N\times N}$ of the form:

$$\frac{d}{dt}\chi_t = f_{u_t}(\chi_t) + \chi_t w_t, \tag{4.1}$$

where w_t is a continuous white noise belonging to $\mathfrak{g}$ (the Lie algebra of G) whose covariance matrix is denoted by Q_t (for a proper discussion on multiplicative noise for systems defined on Lie groups, see Chapter 6), u_t is an input and f_{u_t} verifies for any value of u_t the relation:

$$\forall a,b \in G, f_{u_t}(ab) = af_{u_t}(b) + f_u(a)b - af_{u_t}(Id)b.$$

Guided by the results of Chapter 3, we assume there exists an action $\cdot$ of G on a space $\mathscr{Y}$. The observations we are interested in are :

1. Left-invariant observations: $Y = \tilde{h}(\chi \cdot b_0, V)$ (with V a noise and $\tilde{h}(.,v)$ invertible for any v).
2. Right-invariant observations: $Y = \tilde{h}(\chi^{-1} \cdot b_0, V)$ (with V a noise and $\tilde{h}(.,v)$ invertible for any v).

In each case, $\tilde{h}$ is smooth and verifies $\tilde{h}(x,0) = x$. Before proceeding with the derivation of the filter we give some examples of such observation functions:

Example 3. *The attitude of a vehicle is usually represented by a rotation matrix $R \in G = SO(3)$. It is the rotation matrix mapping a vector expressed in the vehicle frame to a vector expressed in a fixed frame. The following observations have the general form just defined:*

- *$Y = R^T b + V$, where $V \in \mathbb{R}^3$ is a random (unknown) noise and b is a vector known in the fixed reference frame, for instance the gravity vector observed in the vehicle frame by an accelerometer. Here, the space $\mathscr{Y}$ is $\mathbb{R}^3$, the action of G on $\mathscr{Y}$ is simply the product of a matrix and a vector and the observation is of the right-invariant kind.*
- *$Y = \begin{pmatrix} R^T b_1 + V_1 \\ R^T b_2 + V_2 \end{pmatrix}$, where $V_1, V_2 \in \mathbb{R}^3$ are two random noises and b_1, b_2 are two vectors known in the fixed reference frame, for instance the gravity vector and earth magnetic field observed in the vehicle frame by an accelerometer. Here, the space $\mathscr{Y}$ is $\mathbb{R}^6$, the action of G on $\mathscr{Y}$ is defined by $R \cdot \begin{pmatrix} y_1 & y_2 \end{pmatrix} = \begin{pmatrix} Ry_1 & Ry_2 \end{pmatrix}$ and the observation is of the right-invariant kind.*

- *$Y = Ru + V$, where $V \in \mathbb{R}^3$ is a random noise and u is a vector known in the vehicle reference frame, for instance an axis attached to a quadrotor but measured by a fixed camera. Here, the space $\mathscr{Y}$ is $\mathbb{R}^3$, the action of G on $\mathscr{Y}$ is again the product of a matrix and a vector and the observation is of the left-invariant kind.*
- *$Y = RV$, where $V \in SO(3)$ is a random noise (note the values of V are rotation matrices). The observation is here the full state, resulting for instance from a reconstruction algorithm using a set of fixed cameras. Here, the space $\mathscr{Y}$ is $SO(3)$, the action of G on $\mathscr{Y}$ is the product of matrices and the observation is of the left-invariant kind. Moreover, the element b_0 is here the identity matrix.*

To make the description of the filters more concrete, we consider only the case where observations are vectors, and the group action is simply the multiplication of an element of G by a sequence of vectors. There is no specific difficulty in the extension of the approach to the general case. In this modified framework, the two kinds of observations are:

Left-invariant observations This family of outputs writes:

$$Y^1_{t_n} = \chi_{t_n}(d^1 + B^1_n) + V^1_n \qquad , \qquad \dots \qquad , \qquad Y^k_{t_n} = \chi_{t_n}(d^k + B^k_n) + V^k_n,$$

where $(d^i)_{i \leq k}$ are known vectors and the $(V^i_n)_{i \leq k}$, $(B^i_n)_{i \leq k}$ are noises with known characteristics. The Left-Invariant Extended Kalman Filter (LIEKF) is defined here through the following propagation and update steps:

$$\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t), \quad t_{n-1} \leq t \leq t_n \qquad \text{(propagation)}, \tag{4.2}$$

$$\hat{\chi}^+_{t_n} = \hat{\chi}_{t_n} \exp\left[L_n \begin{pmatrix} \hat{\chi}^{-1}_{t_n} Y^1_{t_n} - d^1 \\ \dots \\ \hat{\chi}^{-1}_{t_n} Y^k_{t_n} - d^k \end{pmatrix} \right] \quad \text{(update)}, \tag{4.3}$$

where the function $L_n : \mathbb{R}^{kN} \to \mathbb{R}^{\dim \mathfrak{g}}$ is to be defined in the sequel using error linearizations. A left-invariant error between true state χ_t and estimated state $\hat{\chi}_t$ can be associated to this filter:

$$\eta^L_t = \chi_t^{-1} \hat{\chi}_t.$$

During the Propagation step, as χ_t and $\hat{\chi}_t$ are two trajectories of the system (4.1) *with noise turned off* i.e. $w_t \equiv 0$, Theorem 2 states that the evolution of the error is independent from the estimate $\hat{\chi}_t$ (or alternatively from the true state χ_t). During the Update step, the evolution of the invariant error variable merely reads:

$$(\eta^L_{t_n})^+ = \chi_{t_n}^{-1} \hat{\chi}^+_{t_n} = \eta^L_{t_n} \exp\left[L_n \begin{pmatrix} (\eta^L_{t_n})^{-1} d^1 - d^1 + \hat{\chi}^{-1}_{t_n} V^1_n + B^1_n \\ \dots \\ (\eta^L_{t_n})^{-1} d^k - d^k + \hat{\chi}^{-1}_{t_n} V^k_n + B^k_n \end{pmatrix} \right]. \tag{4.4}$$

We see that, as long as output noise V^j_n is turned off, i.e. $\forall n, j \; V^j_n \equiv 0$, the nice geometrical structure of the LIEKF allows the updated error $(\eta^L_{t_n})^+$ to be only a function of the error just before update $\eta^L_{t_n}$, i.e. to be independent from the estimate $\hat{\chi}_{t_n}$.

Right-invariant observations This family of observations has the form:

$$Y^1_{t_n} = \chi_{t_n}^{-1}(d^1 + V^1_n) + B^1_n \qquad , \qquad \dots \qquad , \qquad Y^k_{t_n} = \chi_{t_n}^{-1}(d^k + V^k_n) + B^k_n,$$

with the same notations as in the previous section. The Right-Invariant EKF (RIEKF) is defined here as:

$$\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t), \qquad t_{n-1} < t < t_n \qquad \text{(propagation)}, \tag{4.5}$$

$$\hat{\chi}^+_{t_n} = \exp\left[L_n \begin{pmatrix} \hat{\chi}_{t_n} Y^1_{t_n} - d^1 \\ \dots \\ \hat{\chi}_{t_n} Y^k_{t_n} - d^k \end{pmatrix}\right] \hat{\chi}_{t_n} \qquad \text{(update)}. \tag{4.6}$$

A right-invariant error can be associated to this filter:

$$\eta^R_t = \hat{\chi}_t \chi_t^{-1}.$$

During the update state, the evolution of the invariant error variable reads:

$$(\eta^R_{t_n})^+ = \hat{\chi}^+_t \chi_t^{-1} = \exp\left[L_n \begin{pmatrix} \eta^R_{t_n} d^1 - d^1 + V^1_n + \hat{\chi}_{t_n} B^1_n \\ \dots \\ \eta^R_{t_n} d^k - d^k + V^k_n + \hat{\chi}_{t_n} B^k_n \end{pmatrix}\right] \eta^R_{t_n}.$$

Once again, due to Theorem 2, we see that when noise is turned off, the evolution of the error does not depend on the state of the system, for both Propagation and Update steps.

The case of full state observation An observation of the full state with multiplicative noise can be seen either as a left or as a right-invariant observation:

$$Y_{t_n} = V_n \chi_{t_n} B_n \qquad \text{or} \qquad Y_{t_n} = B_n^{-1} \chi_{t_n}^{-1} V_n^{-1}.$$

A wise choice is to keep the formulation which provides the more independent noisy error equation. We provide details for the first formulation ($Y_{t_n} = V_n \chi_{t_n} B_n$), left and right playing a symmetrical role here. The LIEKF is defined here as:

$$\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t), \qquad t_{n-1} < t < t_n \qquad \text{(propagation)}, \tag{4.7}$$

$$\hat{\chi}^+_{t_n} = \exp\left[L_n \exp^{-1}(\hat{\chi}_{t_n}^{-1} Y_{t_n})\right] \hat{\chi}_{t_n} \qquad \text{(update)}, \tag{4.8}$$

and here we have $L_n : \mathbb{R}^{\dim \mathfrak{g}} \mapsto \mathbb{R}^{\dim \mathfrak{g}}$. A left-invariant error can be associated to this filter:

$$\eta_t = \chi_t^{-1} \hat{\chi}_t.$$

During the update state, the evolution of the invariant error variable reads:

$$\eta^+_{t_n} = \chi_{t_n}^{-1} \hat{\chi}^+_{t_n} = \eta_{t_n} \exp\left[L_n \exp^{-1}\left((\hat{\chi}_{t_n}^{-1} V_n \hat{\chi}_{t_n}) \eta_{t_n}^{-1} B_n\right)\right]. \tag{4.9}$$

If there is no noise on the left side (i.e. $V_n = Id$) this evolution is independent from the estimate $\hat{\chi}_{t_n}$. In the case $B_n = Id$ the right-invariant filter is to be preferred.

4.2.1 Gain tuning and IEKF equations

As in a conventional EKF, the error equation has to be assumed to be small (here close to Id) so that the error system can be linearized to compute the gains L_n (the error is Id if $\hat{\chi}_t = \chi_t$). By definition, the Lie algebra $\mathfrak{g}$ represents the infinitesimal variations around Id of an element of G. Thus the natural way to define a vector error variable ξ_t in $\mathbb{R}^{\dim \mathfrak{g}}$ is (see Chapter 2):

$$\eta_t = \exp(\xi_t) = \exp_m\Big(\mathscr{L}_{\mathfrak{g}}(\xi_t)\Big). \tag{4.10}$$

During the Propagation step, that is for $t_{n-1} \leq t < t_n$, elementary computations based on the results of Theorem 2 show that for the noisy model (4.1) we have

$$\frac{d}{dt}\eta_t^L = g_{u_t}^L(\eta_t^L) - w_t\eta_t^L, \qquad \frac{d}{dt}\eta_t^R = g_{u_t}^R(\eta_t^R) - (\hat{\chi}_t w_t \hat{\chi}_t^{-1})\eta_t^R. \tag{4.11}$$

Defining $\hat{w}_t \in \mathbb{R}^{\dim \mathfrak{g}}$ by $\mathscr{L}_{\mathfrak{g}}(\hat{w}_t) = -w_t$ in the first case and $\mathscr{L}_{\mathfrak{g}}(\hat{w}_t) = -\hat{\chi}_t w_t \hat{\chi}_t^{-1}$ (i.e. $\hat{w}_t = -Ad_{\hat{\chi}_t}\mathscr{L}_{\mathfrak{g}}^{-1}(w_t)$) in the second case, and using the superscript i to denote indifferently L or R we end up with the linearized error equation in $\mathbb{R}^{\dim \mathfrak{g}}$:

$$\frac{d}{dt}\xi_t = A_{u_t}^i \xi_t + \hat{w}_t, \tag{4.12}$$

where $A_{u_t}^i$ is defined by $g_{u_t}^i(\exp(\xi)) = \mathscr{L}_{\mathfrak{g}}\left(A_{u_t}^i \xi\right) + \bigcirc\left(\|\xi_t\|^2\right)$ (This expression makes sense because the state is embedded into a matrix space, otherwise we should have to compose the left-hand side by $D\log|_{\exp(\xi)}$; see Section 3.3 for a more precise definition) and where we have neglected terms of order $\bigcirc(\|\xi_t\|^2)$ as well as terms of order $\bigcirc(\|\hat{w}_t\|\|\xi_t\|)$. Note that, Theorem 3 states that the first order Taylor expansion $D\log|_{\exp(\xi)} g_{u_t}^i(\exp(\xi)) = \mathscr{L}_{\mathfrak{g}}(A_{u_t}^i \xi)$ is in fact totally exact, i.e. all higher order terms are identically null. This implies the following result (illustrated by Figure 3.1):

Proposition 7. *With noise turned off, i.e. $w_t \equiv 0$, if ξ_t is defined as the solution to the* linearized *error system* (4.12) *and η_t is defined as the solution to the* nonlinear *error system* (4.11)*, then the equality* (4.10) *is verified at all times, even for arbitrarily large errors.*

To derive the equations we consider for instance the case of left-invariant observations, and define ξ_{t_n} through the exponential mapping (4.10), i.e. $\exp(\xi_{t_n}) = \eta_{t_n}^L$. Moreover, for $1 \leq i \leq k$ let $\hat{V}_n^i$ denote $\hat{\chi}_{t_n}^{-1} V_n^i$. To linearize the update equation (4.4) we proceed as follows. For $1 \leq i \leq k$ we have

$$\begin{aligned}
(\eta_{t_n})^{-1} d^i - d^i &= \exp_m(\mathscr{L}_{\mathfrak{g}}(\xi_{t_n}))^{-1} d^i - d^i + \hat{V}_n^i + B_n^i \\
&= \exp_m(-\mathscr{L}_{\mathfrak{g}}(\xi_{t_n})) d^i - d^i + \hat{V}_n^i + B_n^i \\
&= (I - \mathscr{L}_{\mathfrak{g}}(\xi)_{t_n}) d^i - d^i + \hat{V}_n^i + B_n^i + \bigcirc\left(\|\xi_{t_n}\|^2\right) \\
&= -\mathscr{L}_{\mathfrak{g}}(\xi)_{t_n} d^i + \hat{V}_n^i + B_n^i + \bigcirc\left(\|\xi_{t_n}\|^2\right),
\end{aligned}$$

using a simple Taylor expansion of the matrix exponential map. Expanding in the same way equation (4.4) yields:

$$I+\mathscr{L}_{\mathfrak{g}}(\xi_{t_n})^+ = I+\mathscr{L}_{\mathfrak{g}}(a)+\bigcirc\left(\|\xi_{t_n}\|^2\right), \qquad \text{with } a = L_n \begin{pmatrix} -\mathscr{L}_{\mathfrak{g}}(\xi_{t_n})d^1+\hat{V}_n^1+B_n^1 \\ \dots \\ -\mathscr{L}_{\mathfrak{g}}(\xi_{t_n})d^k+\hat{V}_n^k+B_n^k \end{pmatrix}. \quad (4.13)$$

Neglecting terms of order $\bigcirc\left(\|\xi_{t_n}\|^2\right)$ we finally get the following linearized error equation in $\mathbb{R}^{\dim\mathfrak{g}}$:

$$\xi_{t_n}^+ = \xi_{t_n}+L_n(H\xi_{t_n}+\hat{V}_n+B_n), \quad (4.14)$$

where $H \in \mathbb{R}^{kN\times\dim\mathfrak{g}}$, $\hat{V}_n \in \mathbb{R}^{kN}$ and $B_n \in \mathbb{R}^{kN}$ are defined by

$$H\xi = \begin{pmatrix} -\mathscr{L}_{\mathfrak{g}}(\xi)d^1 \\ \dots \\ -\mathscr{L}_{\mathfrak{g}}(\xi)d^k \end{pmatrix}, \qquad \hat{V}_n = \begin{pmatrix} \hat{V}_n^1 \\ \dots \\ \hat{V}_n^k \end{pmatrix}, \qquad B_n = \begin{pmatrix} B_n^1 \\ \dots \\ B_n^k \end{pmatrix}.$$

Note that, contrarily to the Propagation step (4.12), equation (4.14) is only a first-order approximation of the true error update (4.4). Let $\hat{Q}_t$ denote the covariance of the modified process noise $\hat{w}_t$, and $\hat{N}_n$ denote the covariance of the modified measurement noise $\hat{V}_n+B_n$. Note that, equations (4.12) and (4.14) mimic those of a Kalman filter designed for the following auxiliary linear system with discrete measurements: $\frac{d}{dt}x_t = A_{u_t}x_t+\hat{w}_t$, $y_n = Hx_{t_n}+\hat{V}_n+B_n$. The standard Kalman theory thus states that if L_n is computed through the following Riccati equation:

$$\begin{aligned} \frac{d}{dt}P_t = A_{u_t}P_t+P_tA_{u_t}^T+\hat{Q}_t, \qquad S_n = HP_{t_n}H^T+\hat{N}_n, \\ L_n = P_{t_n}H^TS^{-1}, \qquad P_{t_n}^+ = (I-L_nH)P_{t_n}, \end{aligned} \quad (4.15)$$

the dispersion of the linearized error ξ_t is minimized at each step, indicating that $\hat{\chi}_t$ should be a good estimate of χ_t despite the noise, as long as $\hat{\chi}_t$ and χ_t are sufficiently close. Moreover, Kalman theory states that the solution ξ_t to the linearized system (4.12)-(4.14) is a centered Gaussian with covariance matrix $P_t \in \mathbb{R}^{\dim\mathfrak{g}\times\dim\mathfrak{g}}$. Thus the true error $\eta_t = \chi_t^{-1}\hat{\chi}_t$ is approximately distributed as follows

$$\mathbb{P}(\eta_t \in \exp([\zeta,\zeta+d\zeta]) \approx \frac{1}{(2\pi)^{N/2}|P_t|^{1/2}}e^{-\frac{1}{2}\zeta^TP_t^{-1}\zeta}d\zeta, \quad (4.16)$$

where $[\zeta,\zeta+d\zeta]$ is a volume element.

4.2.2 Summary of IEFK equations

Gathering all the results above, the IEKF equations can be compactly written as follows

$$\begin{aligned}
\frac{d}{dt}\hat{\chi}_t &= f_{u_t}(\hat{\chi}_t), \quad t_{n-1} \leq t \leq t_n \\
\hat{\chi}_{t_n}^+ &= \hat{\chi}_{t_n} \exp\left[L_n \begin{pmatrix} \hat{\chi}_{t_n}^{-1} Y_{t_n}^1 - d^1 \\ \dots \\ \hat{\chi}_{t_n}^{-1} Y_{t_n}^k - d^k \end{pmatrix} \right] \text{ (LIEKF)} \\
&\underline{\text{or}} \\
\hat{\chi}_{t_n}^+ &= \exp\left[L_n \begin{pmatrix} \hat{\chi}_{t_n} Y_{t_n}^1 - d^1 \\ \dots \\ \hat{\chi}_{t_n} Y_{t_n}^k - d^k \end{pmatrix} \right] \hat{\chi}_{t_n} \text{ (RIEKF)}
\end{aligned} \tag{4.17}$$

where the LIEKF (resp. RIEKF) is to be used in the case of left (resp. right) invariant outputs. The gain L_n is obtained in each case through a Riccati equation:

$$\begin{aligned}
\frac{d}{dt}P_t &= A_{u_t}P_t + P_t A_{u_t}^T + \hat{Q}_t, \qquad t_{n-1} < t < t_n, \\
S_n = HP_{t_n}H^T + \hat{N}_n, &\qquad L_n = P_{t_n}H^T S^{-1}, \qquad P_{t_n}^+ = (I - L_n H)P_{t_n}.
\end{aligned} \tag{4.18}$$

As concerns the LIEKF, A_{u_t} is defined by $g_{u_t}^L(\exp(\xi)) = \mathscr{L}_{\mathfrak{g}}(A_{u_t}\xi) + \bigcirc(\|\xi\|^2)$, and $H \in \mathbb{R}^{kN \times \dim \mathfrak{g}}$ is defined by $H\xi = \left(-\mathscr{L}_{\mathfrak{g}}(\xi)d^1, \dots, -\mathscr{L}_{\mathfrak{g}}(\xi)d^k\right)^T$, $\hat{Q}_t \in \mathbb{R}^{\dim \mathfrak{g} \times \dim \mathfrak{g}}$ denotes the covariance of the modified process noise $\hat{w}_t = -\mathscr{L}_{\mathfrak{g}}^{-1}(w_t)$ and $\hat{N}_n$ the covariance matrix of the noise $\hat{V}_n + B_n$, $\hat{V}_n$ and B_n being defined as: $\hat{V}_n = (\hat{\chi}_{t_n}^{-1}V_n^1, \dots, \hat{\chi}_{t_n}^{-1}V_n^k)^T$, $B_n = (B_n^1, \dots, B_n^k)^T$.

As concerns the RIEKF, A_{u_t} is defined by $g_{u_t}^R(\exp(\xi)) = \mathscr{L}_{\mathfrak{g}}(A_{u_t}\xi) + \bigcirc(\|\xi\|^2)$, $H \in \mathbb{R}^{kN \times \dim \mathfrak{g}}$ is defined by $H\xi = \left(\mathscr{L}_{\mathfrak{g}}(\xi)d^1, \dots, \mathscr{L}_{\mathfrak{g}}(\xi)d^k\right)^T$, $\hat{Q}_t$ denotes the covariance of the modified process noise $\hat{w}_t = -Ad_{\hat{\chi}_t}\mathscr{L}_{\mathfrak{g}}^{-1}(w_t)$ and $\hat{N}_n$ the covariance matrix of the noise $V_n + \hat{B}_n$, V_n and $\hat{B}_n$ being defined as: $V_n = (V_n^1, \dots, V_n^k)^T$, $\hat{B}_n = (\hat{\chi}_{t_n}B_n^1, \dots, \hat{\chi}_{t_n}B_n^k)^T$.

4.2.3 Stability properties

The aim of the present section is to study the stability properties of the IEKF as a *deterministic* observer. Throughout the section, the noise is thus systematically turned off. Contrarily to the EKF, the IEKF has strong stability properties relying on the state trajectory independence of the deterministic part of the error equation. The stability of a filter is defined as its ability to recover from a perturbation or an erroneous initialization:

Definition 6. *Let $(x_0, t_0, t) \to X_{t_0}^t(x_0)$ denote a continuous flow on a space $\mathscr{X}$ endowed with a distance d. A flow $(z, t_0, t) \to \hat{X}_{t_0}^t(z)$ is an asymptotically stable observer of X about the trajectory $\left(X_{t_0}^t(x)\right)_{t \geq t_0}$ if there exists $\varepsilon > 0$ such that:*

$$d\left(x, \tilde{x}\right) < \varepsilon \Rightarrow d\left(X_{t_0}^t(x), \hat{X}_{t_0}^t(\tilde{x})\right) \to 0 \text{ when } t \to +\infty$$

Theorem 7 below is a consequence of Theorem 4 of Section 3.3. To maintain the independence of the different sections, and not to interrupt the flow of reading, its fairly technical proof has been removed to the Appendix. J. J. Deyst and C. F. Price have shown in [38] the following theorem, stating sufficient conditions for the Kalman filter to be a stable estimator for *linear* (time-varying) systems.

Theorem 6 (Deyst and Price, 1968)**.** *Consider the deterministic linear system* $\frac{d}{dt}x_t = A_t x_t$, $y_{t_n} = Hx_{t_n}$ *and let* $\Phi_{t_0}^t = Id + \int_{t_0}^t A_u du$. *If there exist* α_1, α_2, β_1, β_2, δ_1, δ_2, δ_3, $M > 0$ *such that:*

1. $(\Phi_{t_n}^{t_{n+1}})^T \Phi_{t_n}^{t_{n+1}} \succeq \delta_1 I \succeq 0$
2. $\exists q \in \mathbb{N}^*, \forall s > 0, \exists G_s \in \mathbb{R}^{q \times \dim \mathfrak{g}}, Q_s = G_s Q'$ *where* $Q' \succeq \delta_2 I_q \succeq 0$
3. $N_n \succeq \delta_3 I \succeq 0$
4. $\alpha_1 I \leq \int_{s=t_{n-M}}^{t_n} \Phi(t,s) Q_s \Phi(t,s)^T \leq \alpha_2 I$
5. $\beta_1 I \leq \sum_{i=n-M}^{n-1} \Phi(t_n, t_{i+1})^T H^T N_n^{-1} H \Phi(t_n, t_{i+1}) \leq \beta_2 I$

Then the linear Kalman filter tuned with covariance matrices Q *and* N *is asymptotically stable. More precisely there exist* $\gamma_{min}, \gamma_{max} > 0$ *such that* $\gamma_{min} I \preceq P_t I \preceq \gamma_{max} I$ *for all* t *and* $(\hat{x}_t - x_t)^T P_t^{-1} (\hat{x}_t - x_t)$ *has an exponential decay.*

The main theorem of this section is the extension of this linear result to the non-linear case when the Invariant Extended Kalman Filter is used for systems of Section 3.3.

Theorem 7. *Suppose the stability conditions of the linear Kalman filter given in Theorem 6 are verified about the* true *system's trajectory* χ_t *(i.e. are verified for the linear system obtained by linearizing the system* (3.13) *with left (resp. right) invariant output about* χ_t*). Then the Left (resp. Right) Invariant Extended Kalman Filter* $\hat{\chi}_t$ *is an asymptotically stable observer of* χ_t *in the sense of Definition 6. Moreover, the convergence radius* $\varepsilon > 0$ *is valid over the whole trajectory (i.e. is independent of the initialization time* t_0*).*

Proof. The full proof is technical and has been removed to Appendix A.1. The rationale is to compare the evolution of the logarithmic error ξ_t defined as $\eta_t = \exp(\xi_t)$, with its linearization. For the general EKF, the control of second-order terms in the error equation is difficult because: 1) they depend on the inputs u_t, 2) they depend on the linearization point $\hat{\chi}_t$, 3) the estimation error impacts the gain matrices. For the IEKF, the main difficulties vanish for the following reasons:

- Due to Theorem 3 the linear propagation $\frac{d}{dt}\xi_t = A_{u_t}\xi_t$ of the logarithmic error ξ_t is *exact*: all the non-linearity is contained in the update step.
- Due to the specific form of the IEKF update step, the non-linear rest r_n, defined (here for the left-invariant filter) as: $\exp\left[(I - L_n H)\xi_n - r_n(\xi)\right] = \exp\left[\xi\right] \exp\left[-L_n H(\exp[\xi]b - b)\right]$ is second-order in ξ, *uniformly* over n if L_n is bounded.

- Due to the error equation of the IEKF, the Riccati equation depends on the estimate only through the noise matrices $\hat{Q}_t$ and $\hat{N}_t$ which affect stability in a minor way, as shown by Theorem 6.

In the detailed proof of Appendix A.1 we introduce the flow $\Psi^t_{t_0}$ of the linear part of the equations governing ξ_t (that is, $\Psi^{t_0}_{t_0} = Id, \quad \frac{d}{dt}\Psi^t_{t_0} = A_{u_t}\Psi^t_{t_0}, \quad \Psi^{t+}_{t_0} = (I - L_n H)\Psi^t_{t_0}$) and decompose the solution ξ_t as:

$$\forall t \geq 0, \qquad \xi_t = \Psi^t_0 \xi_0 + \sum_{t_n < t} \Psi^t_{t_n} r_n(\xi_{t_n}). \tag{4.19}$$

All we have to verify is that the apparition of the second-order terms $r_n(\xi_{t_n})$ at each update is compensated by the exponential decay of $\Psi^t_{t_0}$ (Theorem 6).

□

The result displayed in Theorem 7 is in sharp contrast with the usual results available on the EKF which essentially assume the linearized system around the *estimated* trajectory is well-behaved. This is almost impossible to predict as when the estimate is (even slightly) away from the true state, the Kalman gain becomes erroneous, which can in turn amplify the discrepancy between estimate and true state. On the other hand, when considering an actual system undergoing a realistic physical motion (see the examples below), if sufficiently many sensors are available, the linearized system around the *true* trajectory is expected to posses all the desired properties. The following consequence proves useful in practice.

Theorem 8. *Assume the system linearized around the* true *trajectory has the following properties : the propagation matrix $A_t = A$ is constant, there exist matrices B, D such that $\hat{Q}_t = B\bar{Q}_t B^T$ and $\hat{N}_n = D\bar{N}_n D^T$ with $\bar{Q}_t$ and $\bar{N}_n$ upper- and lower-bounded over time, with (A, H, B, D) detectable and reachable. Then the conditions of Theorem 7 are satisfied and the IEKF is asymptotically stable.*

4.3 Examples

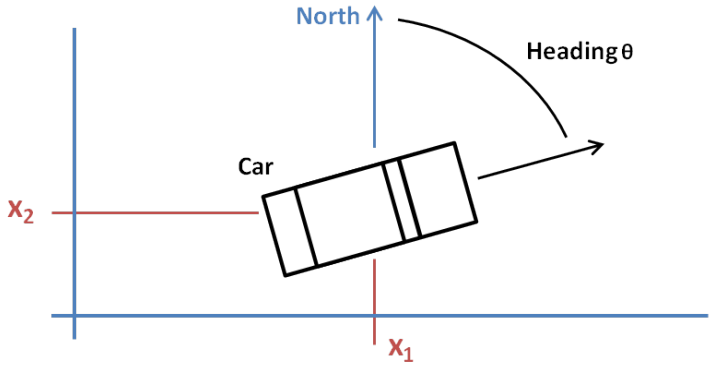
Each example can be read and implemented independently from the rest of the chapter, which is used only to prove the stability of the proposed algorithm. The computations require the basic notions about matrix Lie groups recalled in Chapter 2.

4.3.1 Simplified car or unicycle

Original model

Consider a (non-holonomic) car evolving on the 2D plan. Its heading is denoted by an angle $\theta_t \in [-\pi, \pi]$ and its position by a vector $x_t \in \mathbb{R}^2$. They follow the classical equations:

$$\frac{d}{dt}\theta_t = u_t v_t + w^\theta_t, \qquad \frac{d}{dt}x^1_t = \cos(\theta_t)(v_t + w^l_t) - \sin(\theta_t) w^{tr}_t,$$



$$\frac{d}{dt}x_t^2 = \sin(\theta_t)(v_t + w_t^l) + \cos(\theta_t)w_t^{tr},$$

where v_t is the velocity measured by an odometer, u_t (a function of) the steering angle, w_t^θ the differential odometry error, w_t^l the longitudinal odometry error and w_t^{tr} the transverse shift. Two kinds of observations are considered:

$$\bar{Y}_n = x_{t_n} + V_n, \tag{4.20}$$

or

$$\bar{Y}_n^k = R(\theta_{t_n})^T(x_{t_n} - p_k) + \bar{V}_n^k, \qquad k \in [1, K], \tag{4.21}$$

where $R(\theta)$ is a rotation of angle θ, V_n and $\bar{V}_n^k$ are random noises in $\mathbb{R}^2$. (4.20) represents a noisy position measure (GPS for instance) and has been considered in [43] where a classical EKF has been proposed, but without convergence guarantees. (4.21) represents a range-and-bearing observation of a sequence of known features located at $p_k \in \mathbb{R}^2$ for $k \in [1, K]$. This observation models the navigation of a robot in a known environment. Even in recent papers such as [47], the EKF is considered the best suited to a robot having low computation capacities. As opposed to previous work based on EKF, the IEKF will be shown to possess local convergence properties for the considered problem, and this under mild conditions and for a large class of noise covariances. Chapter 10 will be entirely dedicated to the discussion of this precise example, as it gives a simple and visual understanding of what shifting from EKF to IEKF concretely changes regarding the behavior of the filter.

Matrix form

This system can be embedded in the matrix Lie group $SE(2)$ [87] (see Chapter 2) using the following matrices:

$$\chi_t = \begin{pmatrix} \cos(\theta_t) & -\sin(\theta_t) & x_t^1 \\ \sin(\theta_t) & \cos(\theta_t) & x_t^2 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \nu_t = \begin{pmatrix} 0 & -u_t v_t & v_t \\ u_t v_t & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$w_t = \begin{pmatrix} 0 & -w_t^\theta & w_{t1}^l \\ w_t^\theta & 0 & w_t^{tr} \\ 0 & 0 & 0 \end{pmatrix}.$$

The equation of the system becomes:

$$\frac{d}{dt}\chi_t = \chi_t(\nu_t + w_t), \tag{4.22}$$

and the observations (4.20) and (4.21) respectively have the equivalent form:

$$Y_n = \begin{pmatrix} x_{t_n} + V_n \\ 1 \end{pmatrix} = \chi_{t_n} \begin{pmatrix} 0_{2\times 1} \\ 1 \end{pmatrix} + \begin{pmatrix} V_n \\ 0 \end{pmatrix}, \tag{4.23}$$

$$Y_n^k = \begin{pmatrix} R_{t_n}^T(x_{t_n} - p_k) \\ 1 \end{pmatrix} + \begin{pmatrix} \bar{V}_n^k \\ 0 \end{pmatrix} = -\chi_{t_n}^{-1} \begin{pmatrix} p_k \\ 1 \end{pmatrix} + \begin{pmatrix} V_n^k \\ 1 \end{pmatrix}. \tag{4.24}$$

The reader can verify relation (3.8).

IEKF equations and stability for the left-invariant output (4.20)

The LIEKF (4.17) equations for the system (4.22), (4.23) write:

$$\frac{d}{dt}\hat{\chi}_t = \hat{\chi}_t \nu_t \qquad , \qquad \hat{\chi}_{t_n}^+ = \hat{\chi}_{t_n} \exp\left(L_n \left[\hat{\chi}_{t_n}^{-1} Y_n - \begin{pmatrix} 0_{2\times 1} \\ 1 \end{pmatrix} \right] \right).$$

As the bottom element of $\left[\hat{\chi}_{t_n}^{-1} Y_n - \begin{pmatrix} 0_{2\times 1} \\ 1 \end{pmatrix} \right]$ is always zero we can use a reduced-dimension gain matrix $\tilde{L}_n$ defined by $L_n = \tilde{L}_n \tilde{p}$ with $\tilde{p} = (I_2, 0_{2,1})$. To compute the gains, we write the left-invariant error $\eta_t = \chi_t^{-1}\hat{\chi}_t$ whose evolution is:

$$\begin{aligned} \frac{d}{dt}\eta_t &= \eta_t \nu_t - \nu_t \eta_t - w_t \eta_t, \\ \eta_{t_n}^+ &= \eta_{t_n} \exp\left(\tilde{L}_n \tilde{p} \left[\eta_{t_n}^{-1} \begin{pmatrix} 0_{2\times 1} \\ 1 \end{pmatrix} - \begin{pmatrix} 0_{2\times 1} \\ 1 \end{pmatrix} + \hat{\chi}_{t_n}^{-1} \begin{pmatrix} V_n \\ 0 \end{pmatrix} \right] \right). \end{aligned} \tag{4.25}$$

To linearize of this equation we introduce the linearized error ξ_t defined as $\eta_t = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(\xi_t)$. Introducing $\eta_t = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(\xi_t)$, $\eta_t^+ = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(\xi_t^+)$, $\exp(u) = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(u)$ and $\eta_t^{-1} = I_3 - \mathscr{L}_{\mathfrak{so}(2)}(\xi_t)$ in (4.25) and removing the second-order terms in ξ_t, V_n and w_t we obtain:

$$\frac{d}{dt}\xi_t = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -u_t v_t \\ -v_t & u_t v_t & 0 \end{pmatrix} \xi_t - \begin{pmatrix} w_t^\theta \\ w_t^l \\ w_t^{tr} \end{pmatrix},$$

$$\xi_{t_n}^+ = \xi_{t_n} - \tilde{L}_n[(0_{2,1}, I_2)\xi_t - R(\hat{\theta}_{t_n})^T V_n],$$

The gains $\tilde{L}_n$ are thus computed using the Riccati equation (4.18) with:

$$A_t = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -u_t v_t \\ -v_t & u_t v_t & 0 \end{pmatrix}, \qquad H = (0_{2,1}, I_2), \qquad \hat{Q}_t = Cov \begin{pmatrix} w_t^\theta \\ w_t^l \\ w_t^{tr} \end{pmatrix},$$

$$\hat{N} = R(\hat{\theta}_{t_n}) Cov(V_n) R(\hat{\theta}_{t_n})^T.$$

Proposition 8. *If there exists $v_{\max}, v_{\min} > 0$ such that the displacement satisfies*

$$\|x_{t_{n+1}} - x_{t_n}\| \geq v_{\min} > 0,$$

and the input velocity satisfies $u_t \leq v_{\max}$ then the IEKF is an asymptotically stable observer.

The proof is a verification of the hypotheses of 7 and has been removed to Appendix A.2. Note that it seems very difficult to improve on the assumptions: if $\|x_{t_{n+1}} - x_{t_n}\| = 0$ the system becomes non-observable (the heading θ_t cannot be known), and if u_t is arbitrary high a large move between two observations makes the covariance of the propagation step infinite due to the uncertainty in the heading. In practice an arbitrary high velocity is anyway unfeasible.

IEKF equations and stability for the right-invariant output (4.21)

The RIEKF equations (4.17) for the system (4.22), (4.24) write:

$$\frac{d}{dt}\hat{\chi}_t = \hat{\chi}_t v_t \qquad , \qquad \hat{\chi}_{t_n}^+ = \exp\left(L_n\left[\hat{\chi}_{t_n} Y_n^1 + \begin{pmatrix} p_1 \\ 1 \end{pmatrix}; \ldots; \hat{\chi}_{t_n} Y_n^K + \begin{pmatrix} p_1 \\ 1 \end{pmatrix}\right]\right)\hat{\chi}_{t_n}.$$

As the bottom element of $\left[\hat{\chi}_{t_n}^{-1} Y_n^k + \begin{pmatrix} p^k \\ 1 \end{pmatrix}\right]$ is always zero we can use a reduced-dimension gain matrix $\tilde{L}_n$ defined by $L_n = \tilde{L}_n \tilde{p}$ with:

$$\tilde{p} = \begin{pmatrix} [I_2, 0_{2,1}] & & \\ & \ddots & \\ & & [I_2, 0_{2,1}] \end{pmatrix}.$$

To compute the gains $\tilde{L}_n$ we derive the evolution of the right-invariant error variable $\eta_t = \hat{\chi}_t \chi_t^{-1}$:

$$\begin{aligned} \frac{d}{dt}\eta_t &= -(\hat{\chi}_t w_t \hat{\chi}_t^{-1})\eta_t, \\ \eta_{t_n}^+ &= \exp(\tilde{L}_n \tilde{p}\left[-\eta_{t_n}\begin{pmatrix} p_1 \\ 1 \end{pmatrix} + \begin{pmatrix} p_1 \\ 1 \end{pmatrix} + \hat{\chi}_{t_n} V_n^1; \ldots; -\eta_{t_n}\begin{pmatrix} p_k \\ 1 \end{pmatrix} + \begin{pmatrix} p_k \\ 1 \end{pmatrix} + \hat{\chi}_{t_n} V_n^k\right])\eta_{t_n}. \end{aligned} \tag{4.26}$$

To linearize this equation we introduce the linearized error ξ_t defined as $\eta_t = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(\xi_t)$. Introducing $\eta_t = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(\xi_t)$, $\eta_t^+ = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(\xi_t^+)$, $\exp(u) = I_3 + \mathscr{L}_{\mathfrak{so}(2)}(u)$ and $\eta_t^{-1} =$

$I_3 - \mathscr{L}_{\mathfrak{so}(2)}(\xi_t)$ in (4.26) and removing the second-order terms in ξ_t, V_n and w_t we obtain:

$$\frac{d}{dt}\xi_t = -\begin{pmatrix} 1 & 0_{1,2} \\ (\hat{x}_t)_2 & R(\hat{\theta}_t) \\ -(\hat{x}_t)_1 & \end{pmatrix} \begin{pmatrix} w_t^\theta \\ w_t^l \\ w_t^{tr} \end{pmatrix},$$

$$\xi_{t_n}^+ = \xi_{t_n} - \tilde{L}_n \left[\begin{pmatrix} \begin{pmatrix} -p_2^1 & 1 & 0 \\ p_1^1 & 0 & 1 \end{pmatrix} \\ \dots \\ \begin{pmatrix} -p_2^k & 1 & 0 \\ p_1^k & 0 & 1 \end{pmatrix} \end{pmatrix} \xi_{t_n} - \begin{pmatrix} \hat{R}_{t_n} V_n^1 \\ \dots \\ \hat{R}_{t_n} V_n^k \end{pmatrix} \right].$$

The gains are thus computed using the Riccati equation (4.18) with $A_t, H, \hat{Q}$ and $\hat{N}$ defined as:

$$A_t = 0_{3,3}, \quad H = \begin{pmatrix} \begin{pmatrix} -p_2^1 & 1 & 0 \\ p_1^1 & 0 & 1 \end{pmatrix} \\ \dots \\ \begin{pmatrix} -p_2^k & 1 & 0 \\ p_1^k & 0 & 1 \end{pmatrix} \end{pmatrix},$$

$$\hat{Q} = \begin{pmatrix} 1 & 0_{1,2} \\ (\hat{x}_t)_2 & \hat{R}_t \\ -(\hat{x}_t)_1 & \end{pmatrix} Cov \begin{pmatrix} w_t^\theta \\ w_t^l \\ w_t^{tr} \end{pmatrix} \begin{pmatrix} 1 & 0_{1,2} \\ (\hat{x}_t)_2 & \hat{R}_t \\ -(\hat{x}_t)_1 & \end{pmatrix}^T,$$

$$\hat{N} = \begin{pmatrix} \hat{R}_{t_n} Cov(N^1) \hat{R}_{t_n} & & 0 \\ & \ddots & \\ 0 & & \hat{R}_{t_n} Cov(N^k) \hat{R}_{t_n} \end{pmatrix}.$$

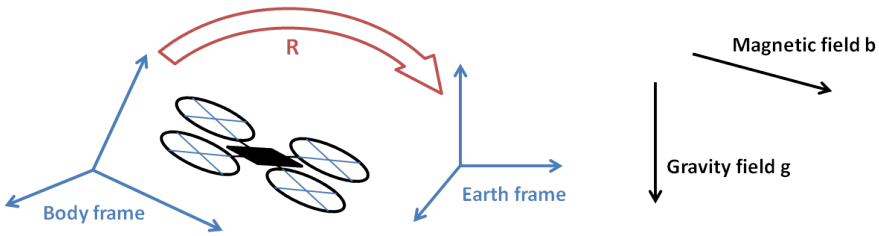
Proposition 9. *If at least two distinct points are observed then the IEKF is an asymptotically stable observer about any bounded trajectory.*

Proof. According to Theorem 8 it is sufficient to show that in this case the observation matrix H is full-rank, i.e. of rank 3. This is obvious as the position and the heading are easily computed from the observation of two vectors at known locations. □

4.3.2 Attitude estimation

Original model

The following example has been considered in numerous articles. An IEKF has already been built in [11], but a proof of stability is still lacking. There are well-known non-linear observers ensuring almost global convergence of the attitude estimate to the true state in the absence of noise. This property is retained by some variants ensuring simultaneous estimates of quantities of interest such as biases [79] or camera to IMU rotation parameters [97] to cite a few, which have flourished over the last decade. However, to our best knowledge those approaches rely on observers with time-invariant gain functions. They



cannot adapt easily to large prior uncertainty or statistics of the noise as do usual Gaussian approximation-based methods. The latter are detailed in [36] in the specific context of attitude estimation and include Multiplicative EKF, Additive EKF, UKF and backward smoothing. To the opposite of fixed-gains methods they have no proved convergence properties but can more easily take into account prior uncertainty or noise statistics. In the following we allow for any covariance matrix of the noises (note that even time-varying ones can be handled) which may result in gains depending on the trajectory (see Proposition 10). Contrarily to those previous works this ensures 1- first-order optimality around the estimated state, as a Kalman filter is used to handle the error system as in the standard EKF philosophy, 2- a tuning based on the trusted actual noises covariances, and 3- at the same time we still prove local convergence around the true trajectory, a feature which is in general very hard if not impossible to prove when using a standard EKF.

In [7], a complementary filter was tested, without convergence properties and in [79], gyro biases were taken into account and global convergence was achieved for constant gains. We would like to mention also that some near-optimality results using a different approach, yet based on a Riccati equation, have also been obtained in [75].

Consider the attitude of a vehicle, represented by the rotation matrix $R_t \in SO(3)$ mapping the coordinates of a vector expressed in the vehicle frame to its coordinates in the static frame. The vehicle is endowed with gyrometers giving an angular velocity ω_t. The equation of the dynamics reads:

$$\frac{d}{dt}R_t = R_t(\omega_t + w_t)_\times, \tag{4.27}$$

where $(b)_\times$ is the anti-symmetric matrix associated to vector b and w_t is a noise (see [11] for a discussion on the noise). Two kinds of observations are considered:

$$Y_n = R_{t_n} V_n, \tag{4.28}$$

or

$$Y_n = (R_t^T g + V_n^g; R_t^T b + V_n^b), \tag{4.29}$$

where V_n is a noise in $SO(3)$, g and b two vectors of $\mathbb{R}^3$ (the gravity and magnetic fields for instance) and V_n^g, V_n^b two noises in $\mathbb{R}^3$. (4.28) represents a full observation of the attitude returned for instance by an image processing algorithm. The noise V_n on the right side corresponds to an uncertainty known in (attached to) the camera frame. (4.28) represents the measurement in the vehicle frame of known vectors of the fixed frame (e.g. using an accelerometer under the quasi-static hypothesis, and a magnetometer), where g et b are assumed non-collinear. This process is already defined on a matrix Lie group (see Chapter 2). The reader can verify relation (3.8).

IEKF equations and stability for the left-invariant output (4.28)

The LIEKF (4.17) for the system (4.27), (4.28) is defined by:

$$\frac{d}{dt}\hat{R}_t = \hat{R}_t(\omega_t)_\times \qquad , \qquad \hat{R}_{t_n}^+ = \hat{R}_{t_n}\exp(L_n \exp^{-1}[\hat{R}_{t_n}^{-1}Y_n]).$$

To compute the gains L_n we define the left-invariant error as $\eta_t = R_t^T\hat{R}_t$ and its evolution reads:

$$\frac{d}{dt}\eta_t = -(\omega_t)_\times\eta_t + \eta_t(\omega_t)_\times - (w_t)_\times\eta_t \qquad , \qquad \eta_{t_n}^+ = \eta_{t_n}\exp(L_n\exp^{-1}(\eta_{t_n}^{-1}V_n)). \tag{4.30}$$

To linearize this equation we introduce the linearized error ξ_t defined as $\eta_t = I_3 + (\xi_t)_\times$. Introducing $\eta_t = I_3 + (\xi_t)_\times$, $\eta_t^+ = I_3 + (\xi_t^+)_\times$, $\exp(u) = I_3 + (u)_\times$, $\eta_t^{-1} = I_3 - (\xi_t)_\times$, $\exp^{-1}[I_3 + (u)_\times] = u$ and $\exp^{-1}(ab) = \exp^{-1}(a) + \exp^{-1}(b)$ in (4.30) and removing the second-order terms in ξ_t, V_n and w_t we obtain:

$$\frac{d}{dt}\xi_t = -(\omega_t)_\times\xi_t - w_t \qquad , \qquad \xi^+ = \xi - L_n(\xi - \tilde{V}_n). \tag{4.31}$$

The gains L_n are thus computed using the Riccati equation (4.18) with matrices $A_t, H, \hat{Q}$ and $\hat{N}$ defined as:

$$A_t = -(\omega_t)_\times, \qquad H = I_3, \qquad \hat{Q} = Cov(w_t), \qquad \hat{N} = Cov(V_n).$$

Proposition 10. *The conditions of Theorem 7 being obvious, the IEKF for the system* (4.27)*,* (4.28) *is an asymptotically stable observer.*

IEKF equations and stability for the right-invariant output (4.29)

The RIEKF (4.17) for the system (4.27), (4.29) is defined by:

$$\frac{d}{dt}\hat{R}_t = \hat{R}_t(\omega_t)_\times \qquad , \qquad \hat{R}_{t_n}^+ = \exp(L_n[\hat{R}_{t_n}Y_n - (g;b)])\hat{R}_{t_n},$$

with the notation $R(x_1;x_2) = (Rx_1;Rx_2)$ for $R \in SO(3)$ and $x_1, x_2 \in \mathbb{R}^3$. To compute the gains L_n we write the evolution of the right invariant error $\eta_t = \hat{R}_t R_t^T$:

$$\frac{d}{dt}\eta_t = -(\hat{R}_t w_t)_\times\eta_t \qquad , \qquad \eta_{t_n}^+ = \exp(L_n[\eta_t g - g + \hat{R}_t V_n^g; \eta_t b - b + \hat{R}_t V_n^b])\eta_{t_n}. \tag{4.32}$$

To linearize this equation we introduce the linearized error ξ_t defined as $\eta_t = I_3 + (\xi_t)_\times$. Introducing $\eta_t = I_3 + (\xi_t)_\times$, $\eta_t^+ = I_3 + (\xi_t^+)_\times$, $\exp(u) = I_3 + (u)_\times$, $\eta_t^{-1} = I_3 - (\xi_t)_\times$ and $\exp^{-1}[I_3 + (u)_\times] = u$ in (4.32) and removing the second-order terms in ξ_t, V_n^g, V_n^b and w_t we obtain:

$$\frac{d}{dt}\xi_t = -\hat{R}_t w_t \qquad , \qquad \xi_{t_n}^+ = \xi_{t_n} - L_n\left[\begin{pmatrix}(g)_\times \\ (b_\times)\end{pmatrix}\xi_{t_n} - \begin{pmatrix}\hat{R}_{t_n}V_n^g \\ \hat{R}_{t_n}V_n^b\end{pmatrix}\right].$$

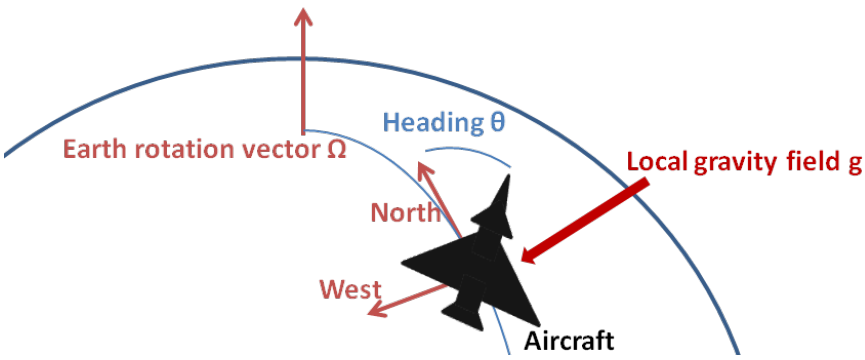
The gains L_n are thus computed using the Riccati equation (4.18) where matrices $A_t, H, \hat{Q}$ and $\hat{N}$ defined as:

$$A_t = 0_{3,3}, \qquad H = \begin{pmatrix}(g)_\times \\ (b_\times)\end{pmatrix}, \qquad \hat{Q} = \hat{R}_t Cov(w_t)\hat{R}_t^T, \qquad \hat{N} = \hat{R}_{t_n}Cov(V_n)\hat{R}_{t_n}^T.$$

Proposition 11. *The IEKF for the attitude estimation system* (4.27), (4.29), *with observation of two non-collinear vectors is an asymptotically stable observer, as a straightforward application of Theorem 8.*

4.3.3 Gyrocompass

Original model



Consider the attitude of a gyrocompass, represented by the rotation matrix $R_t \in SO(3)$ mapping the coordinates of a vector expressed in the gyrocompass frame to its coordinates in the static frame. The vehicle is endowed with gyrometers giving an angular velocity ω_t, and precise enough to measure the earth rotation. This situation did not receive much attention from academics due to the high cost of the IMU required, but is of decisive importance in inertial navigation and led to a great deal of patents. Up to our knowledge, none of them was interested in convergence properties although instability of the EKF can occur for this problem as shown on true data in Chapter 5. This is due to the false observability issues explained in Chapters 8 and 9. The equation of the dynamics reads:

$$\frac{d}{dt}R_t = (\Omega)_\times R_t + R_t(\omega_t + w_t)_\times, \tag{4.33}$$

where Ω is the earth rotation vector and w_t is a noise. An accelerometer gives moreover the gravity field in the reference frame of the gyrocompass (quasi-static hypothesis):

$$Y_n = R_t^T g + V_n^g. \tag{4.34}$$

This process is already defined on a matrix Lie group (see Chapter 2). The reader can verify the relation (3.8).

IEKF equations and stability for the right-invariant output (4.34)

The RIEKF (4.17) for the system (4.33), (4.34) is defined by:

$$\frac{d}{dt}\hat{R}_t = (\Omega)_\times \hat{R}_t + \hat{R}_t(\omega_t)_\times \qquad , \qquad \hat{R}_{t_n}^+ = \exp(L_n[\hat{R}_{t_n}Y_n - g])\hat{R}_{t_n}.$$

The invariant error is $\eta_t = \hat{R}_t R_t^T$ and its evolution reads:

$$\frac{d}{dt}\eta_t = (\Omega)_\times \eta_t - \eta_t (\Omega)_\times - (\hat{R}_t w_t)_\times \eta_t \qquad , \qquad \eta_{t_n}^+ = \exp\left(L_n\left[\eta_t g - g + \hat{R}_t V_n^g\right]\right)\eta_{t_n}. \quad (4.35)$$

To linearize this equation we introduce the linearized error ξ_t defined as $\eta_t = I_3 + (\xi_t)_\times$. Introducing $\eta_t = I_3 + (\xi_t)_\times$, $\eta_t^+ = I_3 + (\xi_t^+)_\times$, $\exp(u) = I_3 + (u)_\times$, $\eta_t^{-1} = I_3 - (\xi_t)_\times$ and $\exp^{-1}[I_3 + (u)_\times] = u$ in (4.35) and removing the second-order terms in ξ_t, V_n^g, V_n^b and w_t we obtain:

$$\frac{d}{dt}\xi_t = (\Omega)_\times \xi_t - \hat{R}_t w_t \qquad , \qquad \xi_{t_n}^+ = \xi_{t_n} - L_n\left[(g)_\times \xi_{t_n} - \hat{R}_{t_n} V_n^g\right].$$

The gains L_n are computed using the Riccati equation (4.18) and matrices $A_t, H, \hat{Q}$ and $\hat{N}$ defined as:

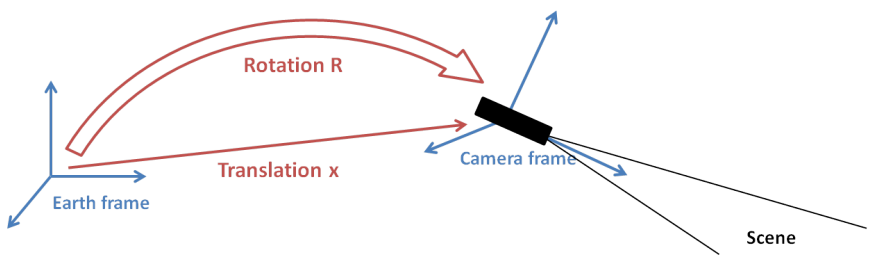
$$A_t = (\Omega)_\times, \qquad H = -(g)_\times, \qquad \hat{Q} = \hat{R}_t Cov(w_t)\hat{R}_t^T, \qquad \hat{N} = \hat{R}_{t_n} Cov(V_n)\hat{R}_{t_n}^T.$$

Proposition 12. *The IEKF for the gyrocompass problem* (4.33), (4.34) *is an asymptotically stable observer if g and Ω are non-collinear.Note that, if g and Ω are collinear (in other word the gyrocompass is on the north or south pole), the heading is anyway not observable.*

Proof. According to Theorem 8 we only have to show that the pair (A_t, H) is observable. The corresponding observability matrix for two successive observations reads $\begin{pmatrix} (g)_\times \\ (g)_\times R_\Omega \end{pmatrix}$, with $R_\Omega = \exp\left(\Omega_\times (t_{n+1} - t_n)\right)$. The rank is not increased if this matrix is multiplied on the left by $\begin{pmatrix} I_3 & 0_{3\times3} \\ 0_{3\times3} & R_\Omega^T \end{pmatrix}$. We obtain $\begin{pmatrix} (g)_\times \\ R_\Omega^T (g)_\times R_\Omega \end{pmatrix} = \begin{pmatrix} (g)_\times \\ (R_\Omega^T g)_\times \end{pmatrix}$. This is the same observation matrix as in the case where vectors g and $R_\Omega^T g$ are observed. This matrix is of rank 3 if g and $R_\Omega^T g$ are non-collinear, i.e. if g and Ω are non-collinear. □

4.3.4 Depth camera motion

Original model



The question of filtering poses computed from a depth camera has been investigated for navigation relying on scan matching [9, 54]. It is relevant also for image denoising

purposes, which can require to work on stabilized images. Some methods avoid reconstruction of camera pose, like [83] where the motion is estimated using a window of several frames then represented as a vector field on the image. Some other resort to filtering methods to maintain a pose estimate, such as particle filtering in [113]. Anyway, no stability proof has been proposed for now. Here, we adopt an approach close to [9] and [54] but use the previous results to study the stability of the method. We consider thus the problem of a hand-held depth-camera in a known environment, and an assumption of slow displacement. The attitude is denoted by the rotation matrix R_t and the position by the 3-dimensional vector x_t. The equations read:

$$\frac{d}{dt}R_t = (w_t^R)_\times R_t \qquad , \qquad \frac{d}{dt}x_t = (w_t^R)_\times x + w_t^x, \tag{4.36}$$

where w_t^R and w_t^x are noises. The process noise is assumed known in the static frame, as whatever the camera orientation is, the horizontal velocity is generally higher than the vertical one, resulting in a covariance linked to the static frame axes. We assume an algorithm returns the whole state (see e.g. [54]).

$$Y_n^R = R_{t_n}^T V_n^R \qquad , \qquad Y_n^x = R_{t_n}^T (x + V_n^x). \tag{4.37}$$

Matrix form

The system can be embedded in the matrix Lie group $SE(3)$ (see Chapter 2) :

$$\chi_t = \begin{pmatrix} R_t & x_t \\ 0_{1\times 3} & 1 \end{pmatrix} \qquad , \qquad w_t = \begin{pmatrix} (w_t^R)_\times & w_t^x \\ 0_{1\times 3} & 0 \end{pmatrix},$$

$$Y_n = \begin{pmatrix} Y_n^R & -Y_n^x \\ 0_{1,3} & 1 \end{pmatrix} \qquad , \qquad V_n = \begin{pmatrix} V_n^R & -V_n^x \\ 0_{1,3} & 1 \end{pmatrix}.$$

Equations (4.36), (4.37) become:

$$\frac{d}{dt}\chi_t = w_t \chi_t, \tag{4.38}$$

$$Y_n = \chi_{t_n}^{-1} V_n. \tag{4.39}$$

IEKF equations and stability for the right invariant output (4.39)

The RIEKF (4.17) for the system (4.38), (4.39) is defined by:

$$\frac{d}{dt}\hat{\chi}_t = 0_{4,4} \qquad , \qquad \hat{\chi}_{t_n}^+ = \exp(L_n[\exp^{-1}(\hat{\chi}_{t_n} Y_n)])\hat{\chi}_{t_n},$$

The right-invariant error is $\eta_t = \hat{\chi}_t \chi_t^{-1}$ and its evolution reads:

$$\frac{d}{dt}\eta_t = -\eta_t w_t \qquad , \qquad \eta_{t_n}^+ = \exp(L_n[\exp^{-1}(\eta_{t_n} V_n)])\eta_{t_n}. \tag{4.40}$$

To linearize this equation we introduce the linearized error ξ_t defined as $\eta_t = I_4 + \mathscr{L}_{\mathfrak{se}(3)}(\xi_t)$ and the linearized observation noise defined as $V_n = I_4 + \mathscr{L}_{\mathfrak{se}(3)}(v_n)$. Introducing $\eta_t =$

$I_4+\mathscr{L}_{\mathfrak{se}(3)}(\xi_t)$, $\eta_t^+=I_4+\mathscr{L}_{\mathfrak{se}(3)}(\xi_t^+)$, $\exp(u)=I_4+\mathscr{L}_{\mathfrak{se}(3)}(u)$, $\eta_t^{-1}=I_4-\mathscr{L}_{\mathfrak{se}(3)}(\xi_t)$, $\exp^{-1}[I_4+\mathscr{L}_{\mathfrak{se}(3)}(u)]=u$ and $\exp^{-1}(ab)\approx\exp^{-1}(a)+\exp^{-1}(b)$ in (4.40) and removing the second-order terms in ξ_t, $\exp^{-1}(V_n)$ and w_t we obtain:

$$\frac{d}{dt}\xi_t=-w_t \qquad , \qquad \xi_{t_n}^+=\xi_{t_n}-L_n(-\xi_{t_n}-v_n).$$

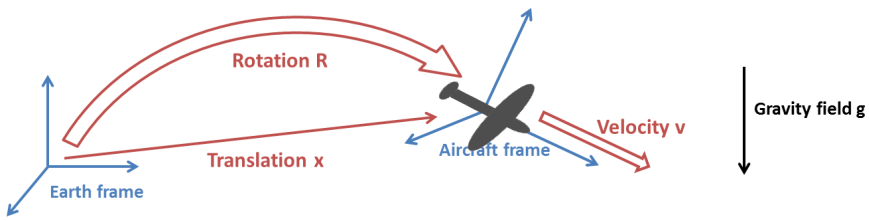
The gains L_n are computed using the Riccati equation (4.18) and matrices $A_t,H,\hat{Q}$ and $\hat{N}$ defined as:

$$A_t=0_{3,3}, \qquad H=-I_6, \qquad \hat{Q}=Cov(w_t), \qquad \hat{N}=Cov(\exp^{-1}(V_n)).$$

Proposition 13. *The IEKF for the movement estimation problem* (4.38), (4.39) *is an asymptotically stable observer, as a straightforward application of Theorem 8.*

4.3.5 3D pose and velocity estimation

Original Model



We consider here the more complicated model of a vehicle evolving in the 3D space and characterized by its attitude R_t, velocity v_t and position x_t. The vehicle is endowed with accelerometers and gyroscopes whose measures are denoted respectively by u_t and ω_t. The equation of the dynamics reads:

$$\frac{d}{dt}R_t=R_t(\omega_t+w_t^{\omega})_\times \qquad , \qquad \frac{d}{dt}v_t=g+R_t(u_t+w_t^u) \qquad , \qquad \frac{d}{dt}x_t=v_t. \tag{4.41}$$

Two kinds of observations are considered:

$$Y_n=x_n+V_n, \tag{4.42}$$

or

$$\left(Y_n^1,\dots,Y_n^K\right)=\left(R_{t_n}^T(p_1-x_1)+V_n^1,\dots,R_{t_n}^T(p_1-x_1)+V_n^K\right), \tag{4.43}$$

where V_n, $V_n^1,\dots,V_n^K$ are noises in $\mathbb{R}^3$. (4.42) represents a position measure given by a GPS. IMU-GPS data fusion is of decisive industrial interest and EKF is a well-established method for this purpose [29, 45, 91, 108]. Yet, no convergence property has been obtained. Invariant methods have been proposed for inertial navigation [10, 18], only in the

simplified situation where the system is reduced to attitude and velocity and assuming the GPS provides an absolute velocity measure. The resulting model is a combination of left- and right-invariant dynamics matching the usual theory. The extended framework proposed by the present document, combined with the introduction of a relevant group structure (see 2.4) makes straightforward the design of an IEKF for the whole attitude/velocity/position system although it does not verify the hypotheses required by invariant techniques. We don't have knowledge of an observer for this problem having convergence properties on uniformly observable trajectories, but the benefits of the IEKF below go beyond its nice theoretical guaranties. It is the core of a pool of patented methods developed for industrial purposes, whose results are presented in Chapter 5. They take into account many additional variables including bias and drifts and inherit largely of the global properties illustrated in Chapter 10.

(4.43) represents the measure of the relative position of known features (using for instance a depth camera): $(p_1, \dots, p_K)$ denote the position of the features in the static frame. The issue of merging IMU data with range and bearing measurements has been raised by the use of stereo cameras. Batch optimization has been used [68] but the most widespread approach is probably visual odometry followed by Kalman filtering (see, e.g., [30,68,98,106] for recent publications). In the last years, the same question received a renewed attention due to the evolution of depth cameras and LiDAR [56,66]. Theoretical stability of EKF-based methods used for this purpose have not been studied yet, but they can be obtained for the IEKF we propose as will be shown below.

Matrix form

As already noticed in the preliminary work [13], the system (4.41) can be embedded in the group of double homogeneous matrices (see Chapter 2) using the matrices χ_t, w_t and function $f_{\omega,u}$:

$$\chi_t = \begin{pmatrix} R_t & v_t & x_t \\ 0_{3,1} & 1 & 0 \\ 0_{3,1} & 0 & 1 \end{pmatrix}, \qquad w_t = \begin{pmatrix} (w_t^\omega)_\times & w_t^f & 0_{3,1} \\ 0_{1,3} & 0 & 0 \\ 0_{1,3} & 0 & 0 \end{pmatrix},$$

$$f_{\omega,u} : \begin{pmatrix} R & v & x \\ 0_{3,1} & 1 & 0 \\ 0_{3,1} & 0 & 1 \end{pmatrix} \to \begin{pmatrix} R(\omega)_\times & g + Ru & v \\ 0_{3,1} & 0 & 0 \\ 0_{3,1} & 0 & 0 \end{pmatrix}.$$

The equation of the dynamics becomes:

$$\frac{d}{dt}\chi_t = f_{\omega_t,u_t}(\chi_t) + \chi_t w_t, \tag{4.44}$$

and the observations (4.42) and (4.43) find respectively the equivalent forms:

$$Y_n = \chi_{t_n} \begin{pmatrix} 0_{4,1} \\ 1 \end{pmatrix} + \begin{pmatrix} V_n \\ 0_{2,1} \end{pmatrix}, \tag{4.45}$$

$$(Y_n^1, \dots, Y_n^K) = \left(\chi_{t_n}^{-1} \begin{pmatrix} p_1 \\ 0 \\ 1 \end{pmatrix} + V_n^1, \dots, \chi_{t_n}^{-1} \begin{pmatrix} p_K \\ 0 \\ 1 \end{pmatrix} + V_n^K \right). \tag{4.46}$$

Proposition 14. *The matrix function f_{ω_t,u_t} is neither left or right invariant, nor a combination of both, but verifies condition* (3.8)*. The demonstration is straightforward.*

IEKF equations for the left-invariant output (4.42)

The LIEKF (4.17) for the system (4.44), (4.45) reads:

$$\frac{d}{dt}\hat{\chi}_t = f_{\omega_t,u_t}(\hat{\chi}_t) \qquad , \qquad \hat{\chi}_t^+ = \hat{\chi}_t \exp(L_n(\hat{\chi}_t^{-1}Y_n)). \tag{4.47}$$

As the two last entries of matrix $\hat{\chi}_t^{-1}Y_n$ are always zero one we can use a reduced-dimension gain matrix $\tilde{L}_n$ defined by $L_n = \tilde{L}_n\tilde{p}$ with $\tilde{p} = (I_3, 0_{3,2})$. The left-invariant error is $\eta_t = \chi_t^{-1}\hat{\chi}_t$ and its evolution reads:

$$\frac{d}{dt}\eta_t = f_{\omega,u}(\eta_t) - f_{\omega,u}(I_5)\eta_t - w_t\eta_t, \tag{4.48}$$

$$\eta_{t_n}^+ = \eta_{t_n}\exp\left(\tilde{L}_n\tilde{p}\left[\eta_{t_n}^{-1}\begin{pmatrix}0_{4,1}\\1\end{pmatrix} + \hat{\chi}_{t_n}^{-1}\begin{pmatrix}V_n\\0_{2,1}\end{pmatrix}\right]\right). \tag{4.49}$$

To linearize this equation we introduce the linearized error ξ_t defined as $\eta_t = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t)$. Introducing $\eta_t = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t)$, $\eta_t^+ = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t^+)$, $\exp(u) = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(u)$ and $\eta_t^{-1} = I_3 - \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t)$ in (4.48) and removing the second-order terms in ξ_t, V_n and w_t we obtain:

$$\frac{d}{dt}\xi_t = \begin{pmatrix}-(\omega_t)_\times & 0_{3,3} & 0_{3,3}\\ -(u_t)_\times & -(\omega_t)_\times & 0_{3,3}\\ 0_{3,3} & I_3 & -(\omega_t)_\times\end{pmatrix}\xi_t - w_t,$$

$$\xi_{t_n}^+ = \xi_{t_n} - \tilde{L}_n\left[\begin{pmatrix}0_{3,3} & 0_{3,3} & I_3\end{pmatrix}\xi_{t_n} - \hat{R}_{t_n}^T V_n\right].$$

The gains $\tilde{L}_n$ are computed using the Riccati equation (4.18) and matrices $A_t, H, \hat{Q}$ and $\hat{N}$ defined as:

$$A_t = \begin{pmatrix}-(\omega_t)_\times & 0_{3,3} & 0_{3,3}\\ -(u_t)_\times & -(\omega_t)_\times & 0_{3,3}\\ 0_{3,3} & I_3 & -(\omega_t)_\times\end{pmatrix}, \qquad H = \begin{pmatrix}0_{3,3} & 0_{3,3} & I_3\end{pmatrix},$$

$$\hat{Q} = Cov(w_t), \qquad \hat{N} = Cov\left(\hat{R}_t^{-1}(V_n)\right).$$

Deriving a stability result of the IEKF for this system under the assumption the acceleration varies (as the state is not observable if the acceleration is constant) goes beyond the scope of the present work. However, due to the theoretical developments above, stability can be conjectured, and it is left for future work.

IEKF equations and stability for the right invariant output (4.43)

The RIEKF (4.17) for the system (4.44), (4.46) reads:

$$\frac{d}{dt}\hat{\chi}_t = f_{\omega_t,u_t}(\hat{\chi}_t) \qquad , \qquad \hat{\chi}_{t_n}^+ = \exp\left(L_n\begin{pmatrix}\hat{\chi}_{t_n}Y_n^1\\ \dots\\ \hat{\chi}_{t_n}Y_n^K\end{pmatrix}\right)\hat{\chi}_{t_n}.$$

As the two last entries of each matrix $\hat{\chi}_{t_n}^{-1} Y_n^k$ are always zero, one we can use a reduced-dimension gain matrix $\tilde{L}_n$ defined by $L_n = \tilde{L}_n \tilde{p}$ with $\tilde{p} = \begin{pmatrix} [I_3, 0_{3,2}] & & \\ & \ddots & \\ & & [I_3, 0_{3,2}] \end{pmatrix}$. The right-invariant error is $\eta_t = \hat{\chi}_t \chi_t^{-1}$ and its evolution reads:

$$\frac{d}{dt}\eta_t = f_{\omega,u}(\eta_t) - \eta_t f_{\omega,u}(I_5) - \left(\hat{\chi}_t w_t \hat{\chi}^{-1}\right)\eta_t, \tag{4.50}$$

$$\eta_{t_n}^+ = \exp\left(\tilde{L}_n \tilde{p} \begin{pmatrix} \eta_{t_n}\begin{pmatrix} p_1 \\ 0 \\ 1 \end{pmatrix} + \hat{\chi}_{t_n}\begin{pmatrix} V_n \\ 0_{2,1} \end{pmatrix} \\ \dots \\ \eta_{t_n}\begin{pmatrix} p_K \\ 0 \\ 1 \end{pmatrix} + \hat{\chi}_{t_n}\begin{pmatrix} V_n \\ 0_{2,1} \end{pmatrix} \end{pmatrix}\right)\eta_{t_n}. \tag{4.51}$$

To linearize this equation we introduce the linearized error ξ_t defined as $\eta_t = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t)$. Introducing $\eta_t = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t)$, $\eta_t^+ = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t^+)$, $\exp(u) = I_3 + \mathscr{L}_{\mathfrak{se}(3),2}(u)$ and $\eta_t^{-1} = I_3 - \mathscr{L}_{\mathfrak{se}(3),2}(\xi_t)$ in (4.50), (4.51) and removing the second-order terms in ξ_t, V_n and w_t we obtain:

$$\frac{d}{dt}\xi_t = \begin{pmatrix} 0_{3,3} & 0_{3,3} & 0_{3,3} \\ (g)_\times & 0_{3,3} & 0_{3,3} \\ 0_{3,3} & I_3 & 0_{3,3} \end{pmatrix}\xi_t - \begin{pmatrix} \hat{R}_t & 0_{3,3} & 0_{3,3} \\ (\hat{v}_t)_\times \hat{R}_t & \hat{R}_t & 0_{3,3} \\ (\hat{x}_t)_\times \hat{R}_t & 0_{3,3} & \hat{R}_t \end{pmatrix} w_t,$$

$$\xi_{t_n}^+ = \xi_{t_n} - \tilde{L}_n \left[\begin{pmatrix} (p_1)_\times & 0_{3,3} & -I_3 \\ & \dots & \\ (p_K)_\times & 0_{3,3} & -I_3 \end{pmatrix}\xi_{t_n} - \begin{pmatrix} \hat{R}_{t_n} V_n^1 \\ \dots \\ \hat{R}_{t_n} V_n^K \end{pmatrix} \right].$$

The gains $\tilde{L}_n$ are computed using the Riccati equation (4.18) and matrices $A_t, H, \hat{Q}$ and $\hat{N}$ defined as:

$$A_t = \begin{pmatrix} 0_{3,3} & 0_{3,3} & 0_{3,3} \\ (g)_\times & 0_{3,3} & 0_{3,3} \\ 0_{3,3} & I_3 & 0_{3,3} \end{pmatrix}, \qquad H = \begin{pmatrix} (p_1)_\times & 0_{3,3} & -I_3 \\ & \dots & \\ (p_K)_\times & 0_{3,3} & -I_3 \end{pmatrix},$$

$$\hat{Q} = \begin{pmatrix} \hat{R}_t & 0_{3,3} & 0_{3,3} \\ (\hat{v}_t)_\times \hat{R}_t & \hat{R}_t & 0_{3,3} \\ (\hat{x}_t)_\times \hat{R}_t & 0_{3,3} & \hat{R}_t \end{pmatrix} Cov(w_t) \begin{pmatrix} \hat{R}_t & 0_{3,3} & 0_{3,3} \\ (\hat{v}_t)_\times \hat{R}_t & \hat{R}_t & 0_{3,3} \\ (\hat{x}_t)_\times \hat{R}_t & 0_{3,3} & \hat{R}_t \end{pmatrix}^T,$$

$$\hat{N} = \begin{pmatrix} \hat{R}_{t_n} Cov(V_1) \hat{R}_{t_n}^T & & \\ & \ddots & \\ & & \hat{R}_{t_n} Cov(V_K) \hat{R}_{t_n}^T \end{pmatrix}.$$

Proposition 15. *If three non-collinear points are observed, then the IEKF is an asymptotically stable observer about any bounded trajectory.*

Proof. According to theorem 8 we only have to ensure the couple (A,H) is observable. Integrating the propagation on one step we obtain the discrete propagation matrix $\Phi = \begin{pmatrix} I_3 & 0_{3\times3} & 0_{3\times3} \\ t(g)_\times & I_3 & 0_{3\times3} \\ \frac{1}{2}t^2(g)_\times & tId_{3\times3} & I_3 \end{pmatrix}$. The observation matrix is denoted H. We will show that $[H;H\Phi]$ has rank 9. We can keep only the rows corresponding to the observation of three non-collinear features p_1, p_2, p_3. and denote the remaining matrix by $\mathscr{H}_1$. Matrices $\mathscr{H}_2$ and $\mathscr{H}_3$, obtained using elementary operations on the columns of $\mathscr{H}_1$, have a rank inferior or equal to the rank of $\mathscr{H}_1$:

$$\mathscr{H}_1 = \begin{pmatrix} (p_1)_\times & 0_{3\times3} & -I_3 \\ (p_2)_\times & 0_{3\times3} & -I_3 \\ (p_3)_\times & 0_{3\times3} & -I_3 \\ (p_1)_\times - \frac{1}{2}t^2(g)_\times & -tI_3 & -I_3 \\ (p_2)_\times - \frac{1}{2}t^2(g)_\times & -tI_3 & -I_3 \\ (p_3)_\times - \frac{1}{2}t^2(g)_\times & -tI_3 & -I_3 \end{pmatrix}, \quad \mathscr{H}_2 = \begin{pmatrix} (p_1)_\times & 0_{3\times3} & -I_3 \\ (p_2)_\times & 0_{3\times3} & -I_3 \\ (p_3)_\times & 0_{3\times3} & -I_3 \\ -\frac{1}{2}t^2(g)_\times & -tI_3 & 0_{3\times3} \\ -\frac{1}{2}t^2(g)_\times & -tI_3 & 0_{3\times3} \\ -\frac{1}{2}t^2(g)_\times & -tI_3 & 0_{3\times3} \end{pmatrix}$$

$$\mathscr{H}_3 = \begin{pmatrix} (p_1 - p_3)_\times & 0_{3\times3} & 0_{3\times3} \\ (p_2 - p_3)_\times & 0_{3\times3} & 0_{3\times3} \\ \frac{1}{2}t^2(g)_\times & tI_3 & 0_{3\times3} \\ (p_3)_\times & 0_{3\times3} & -I_3 \end{pmatrix}$$

The diagonal blocks $\begin{pmatrix} -(p_1 - p_3)_\times & 0_{3\times3} \\ -(p_2 - p_3)_\times & 0_{3\times3} \end{pmatrix}$, tI_3 and I_3 have rank 3 thus the full matrix has rank 9. □

4.4 Conclusion

The methodology of the IEKF has been applied to the systems defined in Chapter 3, whose unusual properties have been used to derive stability guaranties under the same assumptions as in the linear case. These hypotheses have been easily checked for a long list of simple navigation examples. More complicated situations arising in industrial applications are presented in the next Chapter, along with results obtained on real and simulated data.

Chapter 5

Industrial applications

The methods used in this chapter are adaptations of those described in Chapter 4. They led to the registration of two patents (n° 15-00654 and 14-01535) and the present results have been partially presented at the 2014 21st Saint Petersburg International Conference on Integrated Navigation Systems. Note that, throughout the chapter, the level of details is limited by confidentiality of the results.

Chapter abstract When dealing with high precision navigation problems, in order to accurately estimate the position of the body, slowly time-varying biases affecting the gyroscopes and accelerometers measurements need be estimated online. Although the theory developed in the previous chapters does not readily apply to this extended problem, we show how to handle in practice the estimation of such additional variables. Then, we present some industrial results obtained in partnership with SAGEM. The IEKF is shown to outperform the conventional EKF on real experimental data for three different navigation problems.

5.1 Introduction

The present PhD thesis has been partly funded by the aerospace and defense company SAGEM (part of the Safran group), to improve its inertial navigation techniques. We give here a brief description of the specific problems motivating the theory presented in this document, as well as a methodology adopted to approach some applications of a higher degree of complexity than those mentioned in other chapters. Even if the following chapter is of reduced length and the exposure is limited by confidentiality requirements, it has been the object of many experiments, simulations, and hard work, and the results have been considered convincing enough to use the proposed algorithms in a currently developed military device. As such, it constitutes a key - if not the main - output of this PhD thesis.

This chapter is divided into two sections. Section 5.2 discusses a generalized framework allowing to face most of the industrial problems encountered and Section 5.3 presents some inertial navigation issues and the striking improvements allowed by the algorithms implemented.

5.2 Autonomous error equations in practice

The perfect situation studied in chapter 4 is quite general, as illustrated by the various examples treated, but omits to simultaneously estimate biases in the inertial measurements. When it comes to high precision navigation, bias estimation is paramount. Of course, as there is no Lie group structure dictating the form of an EKF meant to simultaneously estimate biases, there is a great deal of freedom to devise such an EKF. What we propose is as follows.

5.2.1 A more general framework

The main idea allowing application of the EKFs described in this work to more complicated systems is to build an error following an autonomous equation for a simpler system with comparable behavior, then use it for the true system although most of the theoretical properties are lost. This approach is more justified than using a classical linear EKF which is, as already mentioned, an arbitrary and parameterization-dependent choice. To illustrate this approach we embed now the class of systems studied in the previous chapters into an environment impacting the dynamics and the observations:

$$\begin{aligned} \frac{d}{dt}\Theta &= l(\Theta), \\ \frac{d}{dt}\chi_t &= f_{(\Theta_t,u_t)}(\chi_t), \\ Y = h(\chi_t \cdot b, \Theta) \qquad &\text{or} \qquad h(\chi_t^{-1} b, \Theta). \end{aligned} \tag{5.1}$$

The core of the dynamics is still a function f verifying (3.13) but the inputs and observation function are both related to an additional vector variable $\Theta_t \in \mathbb{R}^p$. Moreover, the observation involves an additional function h possibly reducing the observability of the system. This framework can be used to describe most of the navigation problems.

5.2.2 "Imperfect" IEKF

The most straightforward way to extend the IEKF to the generalized situation (5.1) is to copy its error variable and add the errors regarding Θ as simple vector differences $\Theta_t - \hat{\Theta}_t$:

$$e_t = (\chi_t^{-1}\hat{\chi}_t, \Theta_t - \hat{\Theta}_t).$$

The matrices involved in the Riccati equation take the general form:

$$A_t = \begin{pmatrix} A_\chi & A_{\Theta,\chi} \\ 0 & A_\Theta \end{pmatrix} \qquad , \qquad H_n = Dh\left(H_\chi, H_\Theta\right),$$

where Dh denotes the differential of h at $\left(\hat{\chi}_{t_n} \cdot b, \hat{\Theta}_{t_n}\right)$. This method apparently sacrifices all the interesting properties of the IEKF, but its fundamental benefits are actually preserved. The theoretical properties of an EKF using this kind of non-linear error variable will be extensively discussed in Part III of this document.

5.2.3 Fast Riccati integration

In this paragraph we underline some tricks to lower the numerical burden of gain computation. A common approximation in the integration of the Riccati equations appearing in Kalman filtering is:

$$\int_{t_n}^{t_{n+1}} F_s^{t_{n+1}} Q_s (F_s^{t_{n+1}})^T \approx (t_{n+1} - t_n) Q_{t_{n+1}},$$

where $F_s^t = D\Phi_s^t|_{\hat{X}_s}$ is the differential of the flow associated to the dynamics. This leads to the simplified equation:

$$P_{t_{n+1}} = F_n P_{t_n} F_n^T + (t_{n+1} - t_n) Q_{t_{n+1}},$$

where $F_n = F_{t_n}^{t_{n+1}}$. This makes the complexity of the integration depend only on the dynamics of the system, not on the error variable chosen, as shown by the following remark:

Remark 6. *Integration of the Jacobian F_n is of equal difficulty for the left- and right-invariant error as their Jacobian are related through the relation:*

$$F_{t_n}^{t_{n+1}R} = Ad_{\tilde{\chi}_{t_{n+1}}} F_{t_n}^{t_{n+1}L} Ad_{\tilde{\chi}_{t_n}^{-1}},$$

where $\tilde{\chi}_t$ is any particular solution of the dynamics (note that the inversion of χ is trivial in all the examples given).

Proof. Let $\tilde{\chi}_t$ denote a particular solution of the dynamics and (Φ^L, Φ^R) the flows associated to the left- and right- invariant errors. For any other solution χ_t, with the usual notations $\eta_t^L = \chi_t^{-1}\tilde{\chi}_t$ and $\eta_t^R = \tilde{\chi}_t \chi_t^{-1}$ we have $\eta_{t_{n+1}}^R = \tilde{\chi}_{t_{n+1}} \eta_{t_{n+1}}^L \tilde{\chi}_{t_{n+1}}^{-1} = \tilde{\chi}_{t_{n+1}} \Phi_{t_n}^{t_{n+1}L}(\eta_{t_n}^L) \tilde{\chi}_{t_{n+1}}(\eta_t^R) = \tilde{\chi}_{t_{n+1}} \Phi_{t_n}^{t_{n+1}L}(\tilde{\chi}_{t_n}^{-1} \eta_{t_n}^R \tilde{\chi}_{t_n}) \tilde{\chi}_{t_{n+1}}$. A first-order expansion in η_t^R gives exactly:

$$F_{t_n}^{t_{n+1}R} = Ad_{\tilde{\chi}_{t_{n+1}}} F_{t_n}^{t_{n+1}L} Ad_{\tilde{\chi}_{t_n}^{-1}}.$$

□

Example 4. *Consider the Riccati equation obtained in example 4.3.1. The Jacobian of the left-invariant error follows the equation:*

$$\frac{d}{dt} F_{t_n}^{tL} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -u_t v_t \\ -v_t & u_t v_t & 0 \end{pmatrix} F_{t_n}^{tL}.$$

The estimate $\hat{\chi}_t$ is a particular solution over $[t_n, t_{n+1}]$, and the solution of the simplified Riccati equation associated to the right-invariant matrix is trivial ($F_{t_n}^{tR} = Id$). We can thus compute instantaneously the solution for the right-invariant error:

$$F_{t_n}^{t_{n+1}L} = Ad_{\hat{\chi}_{t_{n+1}}} Ad_{(\hat{\chi}_{t_n}^+)^{-1}}.$$

Going back to the general case (5.1), the Jacobian of the full system is the solution of:

$$F_0 = Id \qquad , \qquad \frac{d}{dt} F_t = \begin{pmatrix} A_\chi & A_{\Theta,\chi} \\ 0 & A_\Theta \end{pmatrix} F_t.$$

The solution takes the form $F_t = \begin{pmatrix} F_\chi & F_{\Theta,\chi} \\ 0 & F_\Theta \end{pmatrix}$. The method described above allows to optimize the computation of the top left block. Depending on the situation, the other blocks can become either trivial, or well approximated by a trivial solution. This can be illustrated by biased dynamics of the form $\frac{d}{dt}\chi_t = f(\chi_t) + \chi_t b$, for which the Jacobian describing the evolution of the left-invariant error is well approximated by:

$$F_t \approx \begin{pmatrix} F_\chi & \Delta T \times Id \\ 0 & Id \end{pmatrix}.$$

Remark 7. *In this section, we gave some directions of the way to take advantage of the properties of systems having an error variable with autonomous error equation, when these are only part of a more complicated environment as in most engineering applications. The main interest of the theory thus appears to provide the practitioner with relevant error variables for the use of the non-linear state error based EKFs as presented in Chapter 9. Even if they lose their autonomy property due to the increased complexity of the system, a linear error variable has no reason to be preferred to them a priori.*

To be more specific, if a first system has observability properties comparable to those of a second system for which an error variable with autonomous equation is available, then this error should be also used for the first system. This heuristic finds a rigorous justification in the case described by Theorem 20 of Chapter 9: recycling the non-linear error variable of a simpler system can be sufficient to get rid of a dangerous non-linear effect described in details in Chapter 8 and known as false observability.

Another route to take advantage of the particular structure of a part of the state is to use sampling-based methods. The sampling space can thus be reduced by a Rao - Blackwellization procedure benefiting from the specificities of the Riccati equation involved in the Invariant Extended Kalman Filter as shown in Part II of the present thesis. The strong convergence properties of this latter method can be expected to have interesting consequences on the accuracy of the obtained sampling. No result in this direction is given here but addressing this issue seems an interesting research avenue.

5.3 Industrial applications

5.3.1 Inertial navigation

The basic function of an inertial navigation system is to produce real-time information about the attitude, velocity and position of a vehicle (aircraft, ship, submarine, car, etc.). It relies on an Inertial Measurement Unit, consisting of a set of accelerometers and gyroscopes (generally 3 of each). An accelerometer measures the projection over one axis of the sum of vehicle acceleration and gravitation. This vector is fully measured by three accelerometers forming a basis of the space. A gyroscope measures the projection of the angular rate of the vehicle over one axis, three non-coplanar gyroscopes measure the full instantaneous rotation vector. These measures are polluted by a noise whose modeling is a difficult problem in itself. In the design of filtering methods for inertial navigation, it is generally divided into two parts: a systematic error called "bias" or "drift", constant over

time or having a well modeled behavior, and an additional perturbation having erratic behavior and assimilated to a white noise. The general equations in an inertial reference frame read:

$$\begin{aligned}
\frac{d}{dt}R_t &= R_t(\omega_t + d_t + w_t^{\omega})_\times, \\
\frac{d}{dt}v_t &= R_t(f_t + b_t + w_t^{f})_\times + G(x_t), \\
\frac{d}{dt}x_t &= v_t,
\end{aligned} \tag{5.2}$$

where $R_t \in SO(3)$ denotes the attitude in an inertial reference frame, $v_t \in \mathbb{R}^3$ denotes the velocity, $x_t \in \mathbb{R}^3$ the position, $\omega_t \in \mathbb{R}^3$ the measurements of the gyroscopes, $f_t \in \mathbb{R}^3$ the measurements of the accelerometers, $d_t \in \mathbb{R}^3$ and $b_t \in \mathbb{R}^3$ the biases of the gyroscopes and accelerometers respectively, w_t^{ω} and w_t^{f} the white additional noises and $G : \mathbb{R} \in \mathbb{R}^3 \to \mathbb{R} \in \mathbb{R}^3$ the gravity field.

Owing to its nature, inertial navigation faces two major drawbacks. First it allows to maintain an estimate of attitude, velocity and position only if their initial value is known. Second, the sensor errors accumulate and make the estimate drift on the long run. Note that due to roundness of the Earth, an acceleration error does not lead to a position error growing quadratically over time [99] as could be expected. Anyway, coupling inertial navigation with another information is the usual way to circumvent these flaws. Three different situations where an IEKF has been successfully tested on real data are described in this chapter.

5.3.2 Static alignment

As previously mentioned, an initial guess of the attitude, velocity and position of the vehicle is necessary to start an inertial navigation. Finding position and velocity is generally the easy part, as the starting point of a mission is known quite precisely and can be given manually to the IMU. Concerning the velocity, the vehicle is likely to be stationed when the IMU is turned on. Finding the attitude is more tricky as no sensor can measure it accurately (a magnetic compass is subject to large errors incompatible with the requirements of inertial navigation). Fortunately, a stationed vehicle knowing it is stationed need not more information to estimate its attitude if the accuracy of its gyros is sufficient to make out the Earth rotation, whose projection over the horizontal plan (given by the accelerometer) is the north direction. Obviously, the procedure cannot be that simple as the possible rotations of the vehicle during alignment have to be taken into account. Moreover, some sensors bias are observable, some can be or not depending on the attitude changes during the alignment, some correlate with the attitude, and this information has to be encoded to improve future navigation. It makes this problem a perfect candidate for application of an EKF: the dynamics is given by the equations of inertial navigation where sensor biases are modeled (5.2) and the observation can be either position or velocity:

$$Y_n = x_{t_n} + V_n^x \qquad \text{or} \qquad Y_n = v_{t_n} + V_n^v.$$

Note that no velocity nor position sensor is actually used: Y_n is the inertial position or velocity of the geographical point where the vehicle lies, V_x or V_v being the small variations

of position and velocity around these reference values. This kind of virtual measurement is usually called ZUPT (Zero velocity UPdaTe). The Kalman filtering framework proves extremely convenient here: the algorithm stays the same although the observability properties of the problem vary depending on the situation as mentioned above. This made EKF the classical approach to this problem. Unfortunately, application of this local method requires an accurate initialization. Thus, a procedure consisting of three steps is usually implemented:

Vertical direction search: over a few seconds, the accelerometer measures are corrected from rotations thanks to the gyro then averaged to obtain a first guess of the vertical.

Coarse alignment: an approximate but reliable estimate of the attitude is computed. Here different methods can be used, for instance building an approximate but linear modeling of the system and applying a classical Kalman filter.

Fine alignment: starting from a coarse estimate of the attitude (error within a few degrees), the EKF is applied on the full non-linear system to compute an estimate of all variables (including biases), as well as the covariance matrix of the errors.

Implementation of this procedure is heavy, as handling transitions between the different phases is not straightforward: the instant they are performed has to ensure the desired precision is achieved, at the risk of jeopardizing the final result. But waiting too much also means increasing the duration of the alignment. As the systems under consideration are embedded, having several phases also implies a more complicated validation process.

A way to circumvent these difficulties is IEKF. The problems caused here by large initial errors are similar to the one exposed in Chapter 10: state estimation error is mainly caused by one single variable (heading uncertainty) and the true state is thus (almost) constrained in a one-dimensional (non-linear) manifold. This constraint is natively preserved by the estimate of the IEKF, but not by the one of the EKF as shown in Chapter 10, for which the authorized manifold is left after the first update and never reached again (see Figure 10.6 of Chapter 10). Thus, the convergence radius of the IEKF is extremely large, which makes possible to use it without needing a first guess of the attitude. The alignment procedure then boils down to one step: applying an IEKF.

The method has been tested on real data from an IMU put in a car at stand, but with running engine producing small movements. The results are displayed on Figure 5.1. The blue line is the heading estimate of the IEKF, the green line the estimate of the classical method. The vertical line is a consequence of the three-steps procedure: although already computed during the coarse alignment phase, the heading estimate is displayed here only from the beginning of the fine alignment. The transition from a memorized value creates this artifact. As the experiment is performed in real conditions, no ground truth is available. Checking the validity of the method thus requires the assumption that the classical algorithm implemented in the software of the IMU is also correct. Looking at Figure 5.1, two conclusions can be drawn for the specific situation considered in the experiment:

1. The IEKF converges to the correct value of the heading, or at least to the same

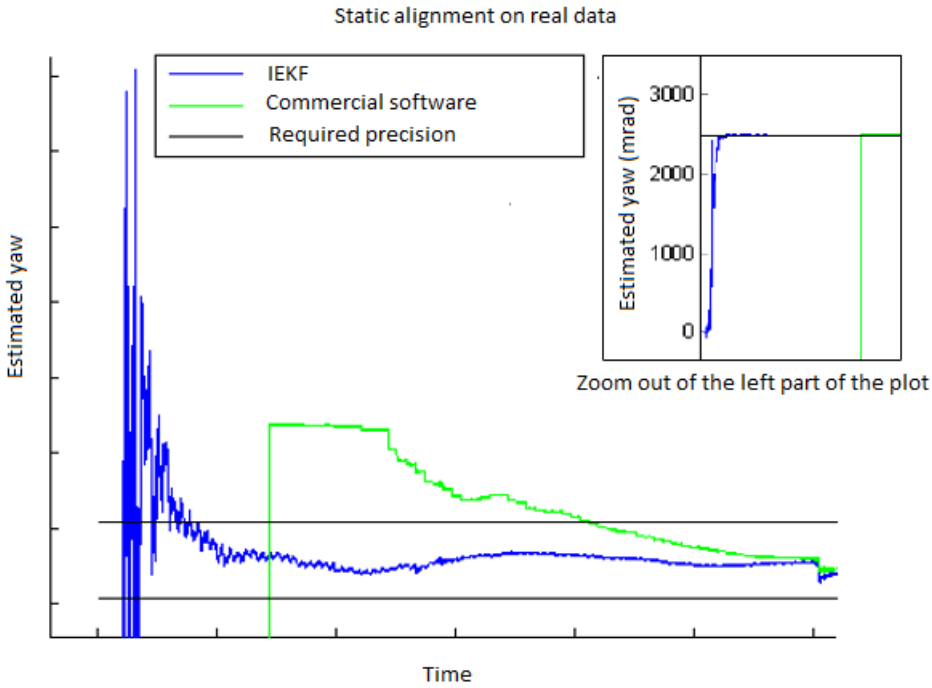


Figure 5.1 – Static alignment of an IMU planted in a car at stand with engine running. The blue line is the heading estimate of the IEKF, the green line the estimate of the commercial software. The vertical line is a consequence of the three-steps procedure: although already computed during the coarse alignment phase, the heading estimate is displayed here only from the beginning of the fine alignment. The transition from a memorized value creates this artifact. We see that the IEKF converges to the same value as the classical method although starting very far (120 degrees) from the true value.

value as the commercial software, although the initial heading error is extremely large (around 120 degrees here).

2. The coarse estimation steps being avoided, the convergence is much faster.

5.3.3 Prolonged static pose

Chapter 8 will explain in details the problem of “false observability”: due to updates of the linearization point the observability properties of the linearized system used by the EKF can differ from those of the true system. IEKF is an efficient response to this issue as shown by Remark 17 in chapter 9. A convincing illustration is given by the following problem encountered in operational situations. A vehicle keeps the same pose for several hours while undergoing small zero-mean translation movements, and maintains its attitude estimate using the ZUPT mentioned above. In this case, no information can ever be obtained concerning the projection of the gyro biases over the west direction. To

be more specific, distinguishing between a variation of this value and a variation of the heading is impossible. Yet, on the long run, the EKF creates artificial information as in the SLAM problem exposed in Chapter 9 and the estimated heading drifts as displayed on Figure 5.2 (purple plot). An EKF re-implemented from scratch and used on the same data encounters the exact same problem (black plot), which proves this is not due to a bug in the software.

A first solution to the problem is the Observability-Constrained EKF (OC-EKF) proposed in [57]. Once the false observability issue has been identified and the dimension of the non-observable subspace analytically computed, the linearization is tuned to ensure the non-observable subspace to keep the right dimension. The results of this algorithm are displayed on Figure 5.2 (green plot). We see that the problem is solved: the drift of the heading estimate disappeared. But the filter has been explicitly "told", by a user possessing side information on the experiment the IMU is undergoing, to maintain a given dimension of the non-observable subspace. This can be worrying regarding possible unexpected consequences, and brings an additional issue: if a true attitude change occurs, making the problematic direction observable, it has to be detected to allow transition to a classical EKF.

An IEKF has been implemented for the same problem, but no constraint regarding non-observability was added. The results are displayed on Figure 5.2 (red plot): this also solves the problem and leads to the same estimate as the OC-EKF, but without ah-hoc patch-up. The blue plot is a second IEKF, modeling only attitude, position and velocity which prevents of course false observability as all its variables are observable. We see that the IEKF modeling the full system (red plot) gives the same results and thus is not perturbed by non-observability. Moreover, a procedure detecting the problematic situation in order to trigger the OCEKF is not required anymore. Nevertheless, note that the OC-EKF has the advantage of being easily added to an existing system, which is not the case of the IEKF.

5.3.4 In-flight alignment

The last practical situation considered in this document is the initialization of the navigation performed while the vehicle is moving, using a GPS aid instead of the "small movement" information. Here again, this functionality is usually achieved through a gradual fine-tuning possibly demanding operational constraints. The results of the software in use in Sagem's IMUs are not going to be displayed here. Yet, the benefits of the IEKF can be easily illustrated first through a comparison with the standard EKF on real data, then through a statistical analysis of the results over a large sample of simulated trajectories.

Figure 5.3 displays the estimation error of an EKF and an IEKF, tuned with the exact same parameters, for a GPS-INS alignment using real inertial measurements. We see the estimate of the EKF deteriorates rapidly as the initial value of the heading error grows. Of course, this specific plot is not representative of the behavior of an operational system as the EKF is not used until a reliable first guess of the attitude has been obtained (see above). It simply illustrates why a much more complicated procedure is required, usually limiting the movements of the vehicle during the alignment. On the other hand, as in the static case, the convergence of the IEKF is not disturbed by large initial errors, which

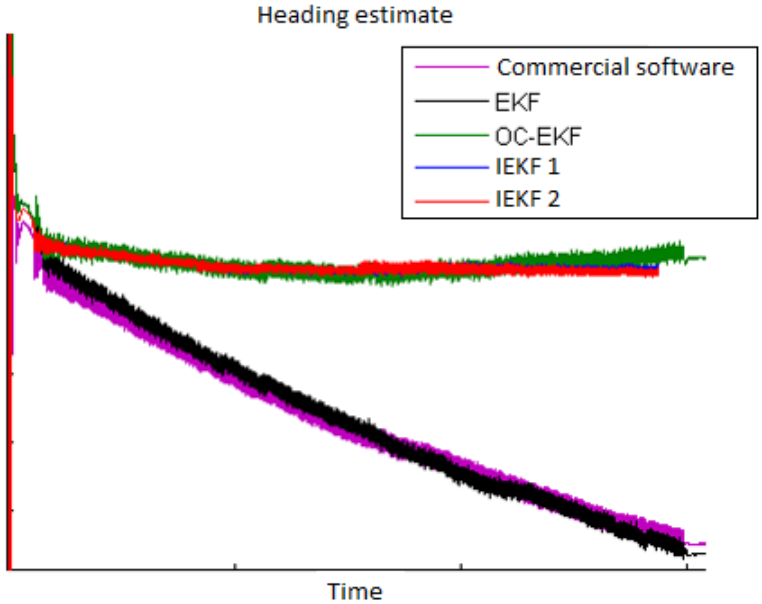


Figure 5.2 – Prolonged static pose during which a "small movement" measure is provided to the EKF. This information is sufficient to maintain the yaw estimate thanks to the earth rotation. The true yaw is constant, but a specific pattern of small translations of the IMU makes the linearization point of the EKF move slightly at each update. The resulting linearized system has a non-observable subspace of smaller dimension than the true system. This has dramatic consequences on the yaw estimate: it drifts over time (purple plot) although the true yaw is constant. To make sure this is not due to a bug in the software, an EKF has been re-implemented from scratch with the exact same results (black plot). An engineer proposed and successfully implemented the OC-EKF introduced in [58]: the dimension of the non-observable subspace is analytically derived and the linearization point of the EKF is modified to preserve it (green plot). A method based on the IEKF, designed without taking into account the possible false observability issue, has been used on the same data (red plot): it did not encounter the problem. Superposed with the red plot is the result of a similar method, but where only the observable part of the system is modeled. The results are almost equivalent. This proves that the first method, like linear filters, is not perturbed by the addition of non-observable states.

allows a great simplification of the general method: the successive fine-tuning steps can be skipped.

Results of Figure 5.4 were extracted from a simulations campaign aiming at the implementation of the IEKF on a commercial product. One of its goals was to verify the coherence of the estimation errors with the covariance matrix provided by the filter. Some results are displayed here but this time, the physical quantities estimated are not given, nor the units. 1000 sequences of increments are generated and the initializations are randomly distributed. We can see that the error over time exactly matches the $3-\sigma$ envelopes, as in the linear case. This is a good argument confirming that the performances of the method are close to optimality. After thousands of simulations including those presented here, the decision was made to shift from EKF to IEKF, at least for the considered product.

5.4 Conclusion

We gave in this chapter the most decisive applications of the theoretical results presented in the rest of the document, and the general form of the equations involved. The existence of commercial products based on the EKF allows a fair comparison with a method relying on the IEKF. In several crucial situations, the second one results in dramatic improvement, possibly allowing a great simplification of the algorithms in use, of their design and validation process, as well as better performance and more flexibility for the user.

This closes the first part of the thesis, devoted to the introduction and study of a class of estimation problems generalizing both invariant and linear systems.

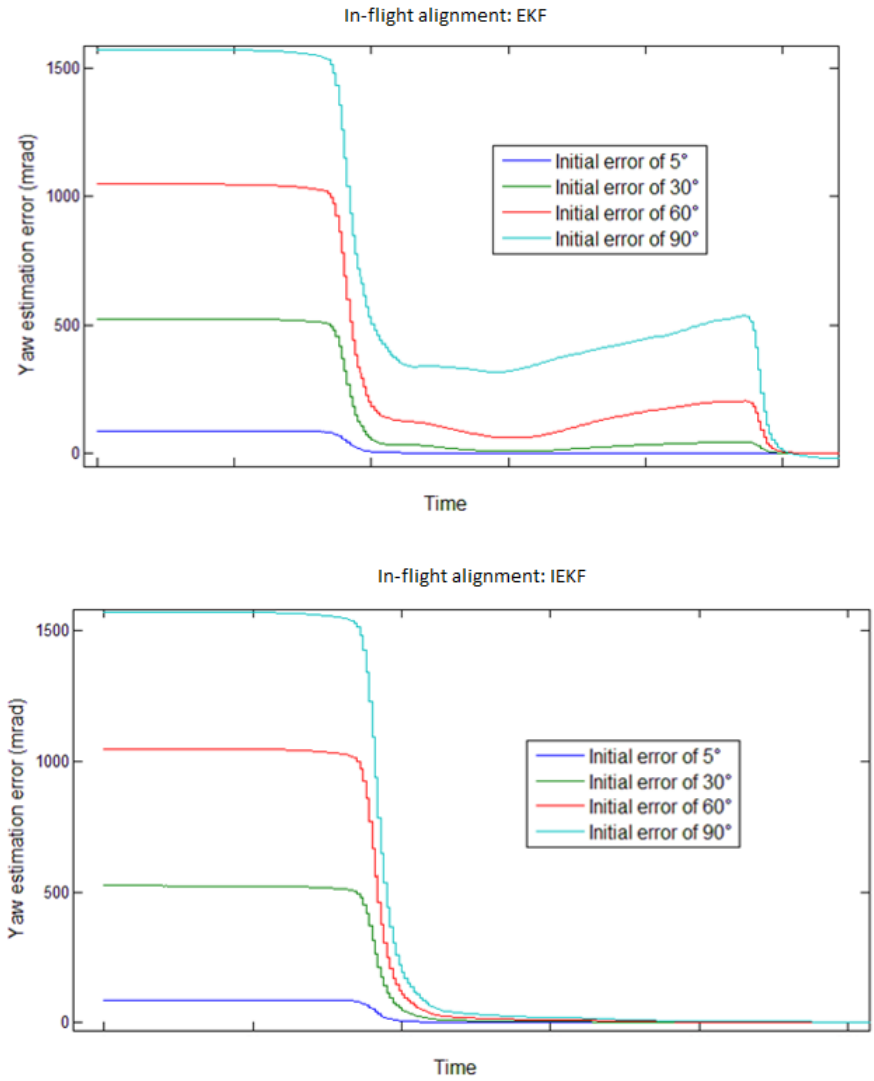


Figure 5.3 – In-flight alignment: comparison of an EKF and IEKF with same parameters. They are equivalent for small initial errors, but the convergence of the EKF deteriorates much faster with the increase of the initial error. It makes it useless during the transitory phase, to the contrary of the IEKF.

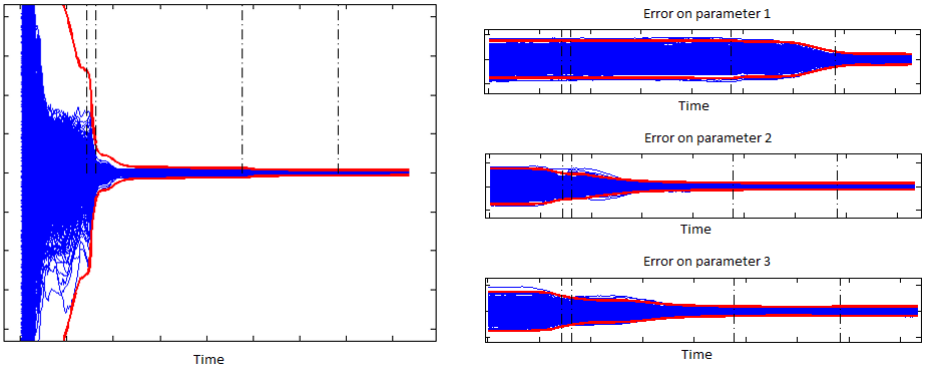


Figure 5.4 – Simultaneous alignment of an IMU and estimation of some physical parameters. A set of 1000 random values has been sampled and corresponding increments generated by a proven simulator. The $3-\sigma$ envelopes of the errors on the physical parameters are displayed over time, together with the true errors for the 1000 simulations. We see the filter gives consistent estimates although the initial attitude errors are possibly large.

Part II

Probabilistic invariant filtering

Chapter 6

Stochastic properties of invariant observers

The present chapter has been published in IEEE Transactions on Automatic Control.

Chapter abstract In the present chapter we address the problem of filtering in the context of an autonomous error equation: the state is now seen as a random process and the quantity of interest becomes its density over time. The interpretation of the stochastic equations involved (Ito or Stratoniovitch) is discussed. The error variable is shown to be a time-homogeneous Harris chain (a Markov chain taking values in a continuous space). We show that if the estimation error almost globally converges for noise turned off, then when noises are turned on the error converges in law to a distribution that is independent of the initial error density. This novel result is illustrated on concrete examples, and leveraged to optimize non-linear gains through the asymptotic error distribution they produce. The transitory phase is then studied and a gain tuning relying on particle sampling is proposed.

6.1 Introduction

In this chapter, the state of the system is studied as a stochastic process, which requires some clarifications regarding the sense of the equations involved (Itô or Stratonovitch). Because of this additional difficulty, the generalized framework built in Part I is temporarily dropped in favor of simplified equations. This makes all the computations much more readable as the dynamics appears under a concrete matrix form instead of involving a function f_{u_t} assumed to verify a given property. This is a handicap regarding the direct application of the results to some inertial navigation problems mentioned earlier, but the properties derived here would actually hold in the general framework under similar hypotheses. This simplification is used only in the present chapter.

Its contributions and organization are as follows. In Section 6.2, the problem of stochastic filtering on Lie groups with a continuous-time noisy model and discrete-time observations is rigorously posed. The problem is then transformed into a complete discrete-time model. The discretization is not straightforward because unlike the linear

case, difficulties arise from non-commutativity. We propose in Section 6.3 a class of natural filters appearing as a loosening of the IEKF presented in Chapter 4, where the gains are allowed to be tuned by any other method than a Riccati equation. The proposed class is very broad and the tuning issue is far from trivial. Sections 6.4 and 6.5 explore two different routes, depending on if we are interested in the asymptotic or transitory phase. In Section 6.4, we propose to hold the gains fixed over time. As a result, the error equation becomes a *homogeneous* Markov chain. We first prove some new results about global convergence in a deterministic and discrete-time framework. Then, building upon the homogeneity property of the error, we prove that if the filter with noise turned off admits almost global convergence properties, the error with noise turned on converges to a stationary distribution. Mathematically this is a very strong and to some extent surprising result. From a practical viewpoint, the gains can be tuned numerically to minimize the asymptotic error's dispersion. This allows to "learn" sensible gains for very general types of noises. The theory is applied to two examples and gives convergence guarantees in each case. First an attitude estimation problem using two vector measurements and a gyroscope having isotropic noise (Section 6.4.2), then the construction of an artificial horizon with optimal gains, for a non-Gaussian noise model (Section 6.4.3). Each application is a contribution in itself and can be implemented without reading the whole thesis. In Section 6.5, we propose to optimize the convergence during the transitory phase using Gaussian approximations. We first stick to the general purpose of this thesis manuscript and adapt the IEKF studied in previous chapters to this discrete situation. As the linearizations always occur around the same point, the linearized model is time-invariant and thus the Kalman gains, as well as the Kalman covariance matrix, are proved to converge to fixed values (note that when on-board storage and computational resources are very limited, this advantageously allows to replace the gain with its asymptotic value. The IEKF is compared to the well-known MEKF [34, 71] and UKF [33, 62] on the attitude estimation problem, and simulations illustrate some convergence properties that the latter lack. In the case where the error equation is fully autonomous, we introduce a new method based on off-line simulations, the IEnKF, which outperforms the other filters in case of large noises by capturing very accurately the error's dispersion.

6.2 Problem setting

6.2.1 Considered continuous-time model

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

$$\frac{d}{dt}\chi_t = (\upsilon_t + w_t)\chi_t + \chi_t \omega_t, \tag{6.1}$$

$$Y_n = h(\chi_{t_n}, V_n), \tag{6.2}$$

where υ_t and ω_t are inputs taking their values in the Lie algebra $\mathfrak{g}$ (i.e. the tangent space to the identity element of G), w_t is a continuous white Langevin noise with diffusion matrix Σ_t, whose precise definition will be discussed below in Subsection 6.2.3. $(Y_n)_{n\geqslant 0}$ are the discrete-time observations, belonging to some measurable space $\mathscr{Y}$ and V_n is a

noise taking values in $\mathbb{R}^p$ for an integer $p > 0$. We further make the following additional assumption:

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

$$\forall g, \chi \in G,\ h(\chi g, V_n) \overset{\mathscr{L}}{=} g^{-1} h(\chi, V_n), \tag{6.3}$$

$$\text{and } h(\chi g, 0) = g^{-1} h(\chi, 0). \tag{6.4}$$

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

Assumption 1 bis For any $n \geqslant 0$ there exists a vector b such that:

$$h(\chi, V_n) = \chi^{-1}(b + V_n).$$

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

6.2.2 Examples

Attitude estimation on flat earth

Our motivating example for the model (6.1)-(6.2) is the attitude estimation of a rigid body assuming the earth is flat, and observing two vectors:

$$\begin{aligned} \frac{d}{dt} R_t &= R_t(\omega_t + w_t^{\omega})_{\times}, \\ Y_n &= (R_{t_n}^T b_1 + V_n^1, R_{t_n}^T b_2 + V_n^2), \end{aligned} \tag{6.5}$$

where as usual $R_t \in SO(3)$ represents the rotation that maps the body frame to the earth-fixed frame, and where $\omega_t \in \mathbb{R}^3$ is the instantaneous rotation vector, and $w_t^{\omega} \in \mathbb{R}^3$ is a continuous Gaussian white noise representing the gyroscopes' noise. We have let $(x)_{\times} \in \mathbb{R}^{3\times 3}$ denote the skew matrix associated with the cross product with a three dimensional vector, i.e., for $a, x \in \mathbb{R}^3$, we have $(x)_{\times} a = x \times a$. $(Y_n)_{n\geq 0}$ is a sequence of discrete noisy measurements of two vectors b_1, b_2 of the earth fixed frame verifying $b_1 \times b_2 \neq 0$, and V_n^1 and V_n^2 are sequences of independent isotropic Gaussian white noises. Note that the noise w_t^{ω} is defined, as the input ω_t, in the body frame (in other words it is multiplied on the left). Thus equations (6.5) do not match (6.1)-(6.2) which correspond to a noise defined in the earth-fixed frame. This can be remedied in the particular case where the gyroscope noise is isotropic, a restrictive yet relevant assumption in practice.

Definition 7. *A Langevin noise w_t of $\mathfrak{g}$ is said isotropic if $\tilde{w}_t := g w_t g^{-1} \overset{\mathscr{L}}{=} w_t$ for any $g \in G$.*

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

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

which corresponds to (6.1)-(6.2).

Remark 8. *If the gyrometer noise is not isotropic the new noise $\tilde{w}_t^{\omega}$ is related to R_t by $\tilde{w}_t^{\omega} = R_t w_t^{\omega} R_t^T$. Depending on the degree of anisotropy this can prevent the use of methods based on the autonomy of the error (see 6.4) but not of the discrete-time IEKF (see 6.5.1).*

Attitude estimation on a round rotating Earth

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

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

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

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

which corresponds to (6.1)-(6.2).

Attitude and velocity estimation on a flat Earth

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

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

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

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

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

The problem can then be rewritten under the form (6.1)-(6.2):

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

6.2.3 Interpretation of Langevin noises on Lie Groups

This present subsection provides some mathematical considerations about white noises on Lie groups. It can be skipped by the uninterested reader who is directly referred to the model (6.6).

Equation (6.1) is actually a Langevin equation that suffers from poly-interpretability because of its non-linearity, and its meaning must be clarified in the rigorous framework of stochastic calculus. Stratonovich stochastic differential equations on a Lie group can be intrinsically defined as in e.g. [72]. A somewhat simpler (but equivalent) approach consists of using the natural embedding of the matrix Lie group G in a matrix space and to understand the equation in the sense of Stratonovich, see e.g. [94]. The mathematical reasons stem from the fact that the resulting stochastic process is well-defined on the Lie group, whereas this is not the case when opting for an Ito interpretation as underlined by the following easily provable proposition.

Proposition 16. *If the stochastic differential equation* (6.1) *is taken in the sense of Itô, for $G = SO(3)$ embedded in $\mathbb{R}^{3\times 3}$, an application of the Itô formula to $\chi_t^T\chi_t$ shows that the solution almost surely leaves the submanifold G for* any *time $t > 0$.*

Besides, the physical reasons for this stem from the fact that the sensors' noise are never completely white, and for colored noise the Stratonovich interpretation provides a better approximation to the true solution than Ito's, as advocated by the result of Wong and Zakai [112].

6.2.4 Exact discretization of the considered model

To treat rigorously the problem of integrating discrete measurements we need to discretize the continuous model with the same time step as the measurements'. Unlike the general case of non-linear estimation, the exact discrete-time dynamics corresponding to Equation (6.1) can be obtained, as proved by the following result:

Theorem 9. *Let $\chi_n = \chi_{t_n}$. Then the discrete system (χ_n, Y_n) satisfies the following equations:*

$$\begin{aligned} \chi_{n+1} &= \Upsilon_n W_n \chi_n \Omega_n, \\ Y_n &= h(\chi_n, V_n), \end{aligned} \tag{6.6}$$

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

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

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

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

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

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

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

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

Remark 9. *In many practical situations (for instance examples 6.2.2 and 6.2.2), the Langevin noise w_t is isotropic and we have thus $\Upsilon_t^{-1} w_t \Upsilon_t \overset{\mathscr{L}}{=} w_t$. Note that, the variable W_t depends also only on the law of w_t for $t \in [t_n, t_{n+1}]$.*

In the sequel, for mathematical reasons (the equations do not suffer from poly-interpretability), tutorial reasons (the framework of diffusion processes on Lie groups needs not be known), and practical reasons (any filter must be implemented in discrete time), we will systematically consider the discrete-time model (6.6). Moreover, the noise W_n will be a general random variable in G, not necessary a solution of the stochastic differential equation $\frac{d}{dt} W_t = \Upsilon_t^{-1} w_t \Upsilon_t W_t$.

6.3 A class of discrete-time intrinsic filters

6.3.1 Proposed intrinsic filters

Inspired by the theory of continuous-time symmetry-preserving observers on Lie groups [20] and the IEKF described in Chapter 4 we propose a class of multiplicative filters defined by:

$$\hat{\chi}_{n+1} = \Upsilon_n \hat{\chi}_n^+ \Omega_n, \tag{6.7}$$

$$\hat{\chi}_{n+1}^+ = K_{n+1}\left(\hat{\chi}_{n+1} Y_{n+1}\right) \hat{\chi}_{n+1}, \tag{6.8}$$

where the superscript $+$ denotes the value of the estimate after the update and $K_{n+1}(\bullet)$ can be any function of $\mathscr{Y} \to G$, ensuring $K(h(I_d, 0)) = I_d$. The multiplicative error does not change:

$$\eta_n = \chi_n \hat{\chi}_n^{-1}, \qquad \eta_n^+ = \chi_n \left(\hat{\chi}_n^+\right)^{-1}. \tag{6.9}$$

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

Theorem 10. *The error variables η_n and η_n^+ are Markov processes, and are independent of the inputs $(\Omega_n)_{n \geqslant 0}$.*

Proof. The equations followed by η_n and η_n^+ read:

$$\eta_{n+1} = \chi_{n+1}\hat{\chi}_{n+1}^{-1} = \Upsilon_n W_n \chi_n \Omega_n \Omega_n^{-1} \left(\hat{\chi}_n^+\right)^{-1} \Upsilon_n^{-1} = \Upsilon_n W_n \eta_n^+ \Upsilon_n^{-1}, \tag{6.10}$$

and:

$$\begin{aligned} \eta_{n+1}^+ &= \chi_{n+1}\hat{\chi}_{n+1}^{-1} K_{n+1}(\hat{\chi}_{n+1} Y_{n+1})^{-1} \\ &= \eta_{n+1} K_{n+1}(\hat{\chi}_{n+1} h(\chi_{n+1}, V_{n+1}))^{-1} \\ &\overset{\mathscr{L}}{=} \eta_{n+1} K_{n+1}(h(\eta_{n+1}, V_{n+1}))^{-1}, \end{aligned} \tag{6.11}$$

thanks to the equivariance property of the output. □

The most important consequence of this property is that if the inputs υ_t are known in advance, or are fixed, as it is the case in the two first examples of 6.2.2, the gain functions K_n can be optimized off-line, independently of the trajectory followed by the system. In any case, numerous choices are possible to tune the gains K_n and the remaining sections are all devoted to various types of methods to tackle this problem.

6.4 Fixed gains filters

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

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

It should be noted that in practice, the convergence to an invariant distribution allows in turn to pick the most desirable gain $K(\bullet)$ among the family based on a performance criterion, such as convergence speed, or filter's precision (that is, ensuring low error's dispersion). This fact will be illustrated by the artificial horizon example of Subsection 6.4.3.

6.4.1 Convergence results

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

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

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

3. Reasonable output noise:

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

The second assumption implies, and can in fact be replaced with, the more general technical assumption that there exists $\varepsilon > 0$ such that for any subset U of the ball $\mathscr{B}_o(I_d, \alpha)$ we have $P(W_n \subset U) > \varepsilon\mu(U)$ for all $n \geq 0$ where μ denotes the Haar measure. Those noise properties are relatively painless to verify, whereas the confinement property although stronger is automatically verified whenever G is compact, e.g. $G = SO(3)$. Intuitively, the last two assumptions guarantee the error process is well approximated by its dynamics with noise turned off, followed by a small diffusion. In the theory of Harris chains, the latter diffusion step is a key element to allow probability laws to mix at each step and eventually forget their initial distribution.

Theorem 11. *For constant left-hand inputs $\Upsilon_n \equiv \Upsilon$, consider the filter:*

$$\hat{\chi}_{n+1} = \Upsilon \hat{\chi}_n^+ \Omega_n, \tag{6.12}$$

$$\hat{\chi}_{n+1} = K(\hat{\chi}'_{n+1} Y_{n+1}) \hat{\chi}'_{n+1}. \tag{6.13}$$

Suppose that Assumptions 1)-2)-3) are verified, where the compact set satisfies $C = cl(C^o)$, cl denoting the closure and ^{o} the interior. When noises are turned off, the error equation (6.10)-(6.11) *becomes:*

$$\begin{aligned} \gamma_{n+1} &= \Upsilon \gamma_n^+ \Upsilon^{-1}, \\ \gamma_{n+1}^+ &= \gamma_{n+1} K_{n+1}(h(\gamma_{n+1}, 0))^{-1}. \end{aligned} \tag{6.14}$$

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

$$\|\mu_n - \pi\|_{T.V.} \underset{n\to\infty}{\longrightarrow} 0.$$

Corollary 2. *When the group G is compact, The convergence results of Theorem 11 hold globally, i.e. without the confinement assumption 1).*

Theorem 12. *Under the assumptions of Theorem 11, assuming only $h(\gamma_n, 0) \to 0$ instead of almost global convergence of $(\gamma_n)_{n\geq 0}$, that G is compact, the set $K = \{g \in G, h(g,0) = h(I_d, 0)\}$ connected and $h(\Upsilon, 0) = h(I_d, 0)$, the results of Theorem 11 are still valid. Moreover, if W_n is isotropic, we have $\pi(\tilde{\Upsilon} V) = \pi(V)$ for any $\tilde{\Upsilon} \in K$ commuting with Υ ($\tilde{\Upsilon}\Upsilon = \Upsilon\tilde{\Upsilon}$).*

The proofs of the results above have all been moved to Appendix B.1.

6.4.2 Application to attitude estimation

Consider the attitude estimation example of Subsection 6.2.2. In a deterministic and continuous time setting, almost globally converging observers have been proposed in several papers (see e.g. [76,96]) and have since been analyzed and extended in a number of papers. In order to apply the previously developed theory to this example, the challenge is twofold. First, the deterministic observer must be adapted to the discrete time and proved to be almost globally convergent. Then, the corresponding filter must be proved to satisfy the assumptions of the theorems above in the presence of noise. In discrete time, the system equations read:

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

with the notations introduced in 6.2.2. We propose the following filter on $SO(3)$:

$$\begin{aligned} \hat{R}_{n+1} &= \hat{R}_n^+ \Omega_n, \\ \hat{R}_{n+1}^+ &= K(\hat{R}_{n+1} Y_{n+1}) \hat{R}_{n+1}, \\ \text{with } K(y_1, y_2) &= \exp(k_1(y_1 \times b_1) + k_2(y_2 \times b_2)), \\ k_1 > 0,\ k_2 > 0,\ k_1 + k_2 &\leqslant 1. \end{aligned} \tag{6.15}$$

Proposition 17. *With noise turned off, the discrete invariant observer* (6.15) *is almost globally convergent, that is, the error converges to I_d for any initial condition except one.*

The proof has been moved to the Appendix, only the main idea is given here. For the continuous-time deterministic problem it is known, and easily seen, that $E : \gamma \to k_1 ||\gamma^T b_1 - b_1||^2 + k_2 ||\gamma^T b_2 - b_2||^2$ is a Lyapunov function, allowing to prove almost global convergence of the corresponding observer. In the discrete time deterministic case, the function above remains a Lyapunov function for the sequence $(\gamma_n)_{n \geq 0}$, allowing to derive Proposition 17. This is not trivial to prove, and stems from the more general following novel result:

Proposition 18. *Consider a Lie group G equipped with a left-invariant metric $\langle .,. \rangle$, and a left-invariant deterministic discrete equation on G of the form:*

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

Assume there exists a C^2 function $E : G \to \mathbb{R}_{\geq 0}$ with bounded sublevel sets, a global minimum at I_d, and a continuous and strictly positive function $u : G \to \mathbb{R}_{>0}$ such that: $\forall x \in G, k(x) = u(x)[x^{-1}.grad_E(x)]$. If the condition $\forall x \in G, |\frac{\partial k}{\partial x}| \leqslant 1$ (for the operator norm associated to the left-invariant metric) is verified, for any initial value γ_0 such that I_d is the only critical point of E in the sublevel set $\{x \in G \mid E(x) \leq E(\gamma_0)\}$ we have:

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

The proof has been moved to Appendix B.2.1. Note that, the latter property is closely related to Lemma 2 of [104]. Using Theorem 11 and Proposition 17 we finally directly get:

Theorem 13. *The distribution of the error variable of the invariant filter* (6.15) *converges for the T.V. norm to an asymptotic distribution, which does not depend on its initial distribution.*

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

As proved in 6.4.1, under appropriate conditions the error variable is a converging Markov chain whose asymptotic law depends on the gain function but not on the trajectory followed by the system (which is a major difference with most nonlinear filters, such as the

EKF). Hence a fixed gain can be asymptotically optimized off-line, leading to a very low numerical cost of the on-line update.

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

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

where R_t is the attitude of the aircraft (the rotation from body-frame coordinates to inertial coordinates), ω_t is the continuous-time gyroscope increment and Y_n is the observation of the accelerometer. The sensor noises w_t and V_n can be considered as Gaussian, and N_n represents fluctuations due to accelerations of the aircraft that we propose to model as follows: N_n is null with high probability but when non-zero it can take large values. The N_n's are assumed to be independent as usually.

Convergence results

Consider the following class of filters:

$$\begin{aligned}\hat{R}_{n+1} &= \hat{R}_n^+ \Omega_n,\\ \hat{R}_{n+1}^+ &= K(\hat{R}_{n+1} Y_{n+1})\hat{R}_{n+1},\\ \text{with } K(y) &= \exp(f_{k,\lambda}(y))\end{aligned} \tag{6.16}$$

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

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

Lemma 1. *For any $0 < k \leqslant 1$ and $0 < \lambda \leqslant \pi$ the output error $\|Y_n - \hat{R}_n^T g\|$ associated to the observer* (6.16) *with noise turned off converges to 0.*

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

The following result is a mere consequence of Lemma 1 and Theorem 12.

Proposition 19. *The error variable associated to the filter defined by* (6.16) *converges to a stationary distribution for the T.V. norm, which does not depend on its initial distribution.*

Numerical asymptotic gain optimization

To each couple (k, λ) we can associate an asymptotic error dispersion (computed in the Lie Algebra) associated to the corresponding stationary distribution, and try to minimize it. As all computations are to be done off-line, the statistics of all distributions can be computed using particle methods. Table 6.1 gives the parameters of the model used in the following numerical experiment. Figure 6.1 displays the Root Mean Square Error $RMSE = \sqrt{\mathbb{E}(\eta_\infty g - g)}$, computed over a grid for the parameters (k, λ). The minimum is obtained for $k = 0.1202$ and $\lambda = 0.0029$. If we compare it to a MEKF, we observe a huge difference. For our asymptotic invariant filter we get $RMSE = 8.02 \times 10^{-4}$. For the MEKF, the observation noise matrix giving the best results leads to the value $RMSE = 4.3 \times 10^{-3}$. This result is not surprising due to the fact that the outliers significantly pollute the estimates of the Kalman filter (see Fig. 6.1). This illustrates the fact that when the noise is highly non-Gaussian, an asymptotic gain with some optimality properties can still be found.

Table 6.1 – Artificial horizon: experiment parameters

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

6.5 Gaussian filters

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

6.5.1 The discrete-time Invariant EKF

In the present section we give the equations of the IEKF in the simplified situation of this chapter. As in Chapter 4 and standard EKF theory, the idea is to linearize the error

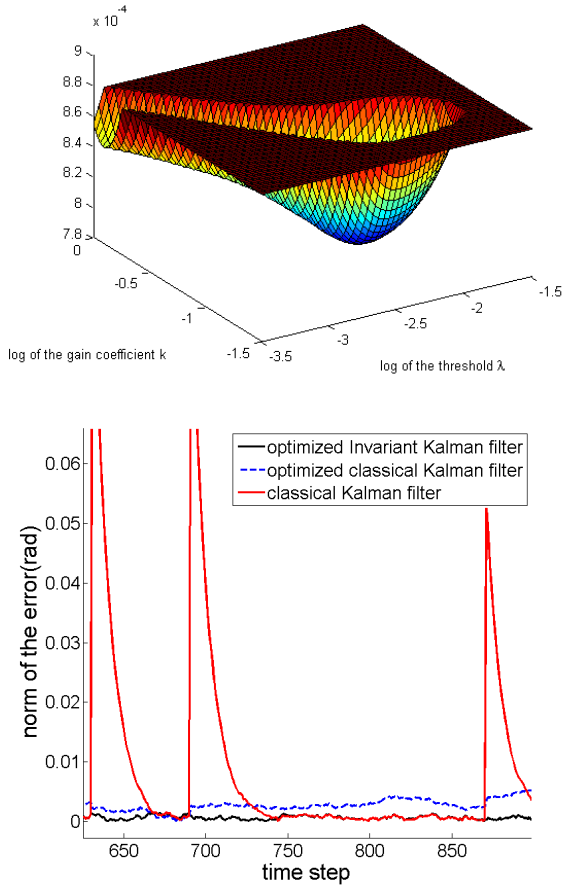


Figure 6.1 – Artificial horizon example. Top plot: off-line optimization of the parameters of the asymptotic invariant filter (axis $x \to \lambda$; axis $y \to k$; axis $z \to$ Root Mean Square Error). The highest values are not displayed to improve visualization. The performance is optimal for $k = 0.1202, \lambda = 0.0029$. Bottom plot: error evolution of three artificial horizons : the invariant filter with optimized gain functions (black), a MEKF ignoring outliers (red), and a MEKF where the observation noise covariance matrix has been numerically adjusted to minimize the RMSE (dashed line). We see that the MEKF cannot filter the outliers and at the same time be efficient in the absence of outliers, contrarily to the proposed invariant filter.

equation (6.10)-(6.11) assuming the noises and the state error are small enough, use Kalman equations to tune the gains on this linear system, and implement the gain on the nonlinear model. As in the heuristic theory of the IEKF in continuous time [14] [21] the gains and the error covariance matrix are proved to converge to a fixed value. In particular this allows (after some time) to advantageously replace the gain by its constant final value leading to a numerically very cheap asymptotic version of the IEKF.

Linearization of the equations and IEKF formulas

We consider here the error equations (6.10)-(6.11). Assuming errors and noises are small, we introduce their projection in the tangent space at the identity I_d (the so-called Lie algebra $\mathfrak{g}$ of the group) using the matrix exponential map $\exp(.)$ (see Chapter 2). As the matrix vector space $\mathfrak{g}$ can be identified with $\mathbb{R}^{\dim \mathfrak{g}}$ using the linear mapping $\mathscr{L}_{\mathfrak{g}}$ (see Table 2.1) we can assume $\exp : \mathbb{R}^{\dim \mathfrak{g}} \to G$. This function is defined in Table 2.1 for usual Lie groups of mechanics. Its inverse function will be denoted by $\exp^{-1}$. We thus define the following quantities of $\mathbb{R}^{\dim \mathfrak{g}}$:

$$\xi_n = \exp^{-1}(\eta_n), \quad \xi_n^+ = \exp^{-1}(\eta_n^+), \quad w_n = \exp^{-1}(W_n).$$

The gain is designed as previously using a function which is linear in $\mathbb{R}^{\dim \mathfrak{g}}$, then an exponential mapping:

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

Equations (6.10) and (6.11) mapped to $\mathbb{R}^{\dim \mathfrak{g}}$ become:

$$\begin{aligned} \xi_{n+1} &= \exp^{-1}\left(\exp\left(Ad_{\Upsilon_n} w_n\right) \exp\left(Ad_{\Upsilon_n} \xi_n^+\right)\right), \\ \xi_{n+1}^+ &= \exp^{-1}(\exp(\xi_{n+1}) \exp[-L_n(h(\exp(\xi_{n+1}, V_{n+1})) - h(I_d, 0))]). \end{aligned}$$

As in Chapter 4, the gains L_n are tuned on the linearized system through the classical Kalman theory. Within our current simplified framework, this first-order expansion reads:

$$\begin{aligned} \xi_{n+1} &= Ad_{\Upsilon_n} w_n + Ad_{\Upsilon_n} \xi_n^+, \\ h(\exp(\xi_{n+1}), V_{n+1}) &= h(I_d, 0) + H_\xi \xi_{n+1} + H_V V_{n+1}, \\ \xi_{n+1}^+ &= \xi_{n+1} - L_{n+1}(H_\xi \xi_{n+1} + H_V V_{n+1}). \end{aligned}$$

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

Convergence of the gains

The main benefit of the filter is with respect to its convergence properties. Indeed, under very mild conditions, the covariance matrix P_n and the filter's gain K_n are proved to converge to fixed values [65]. The practical consequences are at least twofold: 1-the error covariance converges to a fixed value and is thus much easier to interpret by the user than a matrix whose entries keep on changing (see Fig. 6.2) 2-due to computational limitations on-board, the covariance may be approximated by its asymptotic value, leading to an asymptotic version of the IEKF being numerically very cheap.

Theorem 14. *If the noise matrices $Q_n^w = Var(w_n)$, $Q_n^V = Var(V_n)$ and the left inputs Υ_n are fixed, the pair (Ad_{Υ_n}, H_ξ) is observable and Q_n^w has full rank, then P_n and K_n converge to fixed values P_∞ and K_∞, as in the linear time-invariant case.*

Algorithm 1 Invariant Extended Kalman Filter

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

$$h(\exp(\xi_n), V_n) = h(I_d, 0) + H_\xi \xi_n + H_V V_n + \bigcirc(\|\xi_n\|^2) + \bigcirc(\|V_n\|^2).$$

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

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

The process noise covariance $Q_n^w = Var(Ad_{\Upsilon_n} w_n)$ is equal to $M_{t_{n+1}}$ $\big($if w_n is isotropic or if $\upsilon_t \equiv 0$, merely set $Q_n^w = Var(w_n)(t_{n+1} - t_n)\big)$,
$Q_{n+1}^V = Var(V_{n+1})$,
$P_{n+1|n} = Ad_{\Upsilon_n} P_n Ad_{\Upsilon_n}^T + Q_n^w$,
$S_{n+1} = H_V Q_{n+1}^V H_V^T + H_\xi P_{n+1|n} H_\xi^T$,
$L_{n+1} = P_{n+1|n} H_\xi^T S_{n+1}^{-1}$,
$P_{n+1} = (I - L_{n+1} H_\xi) P_{n+1|n}$,
$\hat{\chi}_{n+1} = \Upsilon_n \hat{\chi}_n^+ \Omega_n$,
$\hat{\chi}_{n+1}^+ = \exp(L_{n+1}[\hat{\chi}_{n+1} Y_{n+1} - h(I_d, 0)])\hat{\chi}_{n+1}$,
$n \leftarrow n + 1$.
end loop

Algorithm 2 IEnKF

Define H by $h(\exp(\xi),0) = h(I_d,0) + H\xi + \bigcirc(\|\xi\|^2)$.
Sample M particles $(\eta_0^i)_{1\leqslant i\leqslant M}$ following the prior error density.
for $n = 0$ **to** $N-1$ **do**
  for $i = 1$ **to** M **do**
    $\eta_{n+1}^i = W_n^i \Upsilon \eta_n^{i+} \Upsilon^{-1}$,
    $y_{n+1}^i = h(\eta_{n+1}^i, V_{n+1}^i)$,
  end for
  $P_{n+1} = \frac{1}{M}\sum_{i=1}^M \exp^{-1}(\eta_{n+1}^i)\exp^{-1}(\eta_{n+1}^i)^T$,
  $S_{n+1} = \frac{1}{M}\sum_{i=1}^M y_{n+1}^i (y_{n+1}^i)^T$,
  $L_{n+1} = P_{n+1} H^T S_{n+1}^{-1}$,
  for $i = 1$ **to** M **do**
    $\eta_{n+1}^{i+} = \eta_{n+1}^i \exp(-L_{n+1}[y_{n+1}^i - h(I_d,0)])$,
  end for
  Store the gain matrix L_{n+1}.
end for
The prior estimation $\hat{\chi}_0$ is an input.
for $n = 0$ **to** $N-1$ **do**
  $\hat{\chi}_{n+1} = \Upsilon \hat{\chi}_n^+ \Omega_n$,
  $\hat{\chi}_{n+1}^+ = \exp(L_{n+1}[\hat{\chi}_{n+1} Y_{n+1} - h(I_d,0)])\hat{\chi}_{n+1}$.
end for

6.5.2 The discrete-time Invariant EnKF

When the error equation is independent of the inputs, the exact density of the error variable can be sampled off-line. This allows to radically improve the precision of the quantities involved in the Kalman gains computation of Section 6.5.1. We propose here the Invariant Ensemble Kalman filter (IEnKF) described in Algorithm 2. The idea is to compute recursively through Monte-Carlo simulations a sampling of the error density and to use it, instead of linearizations, to evaluate precisely the innovation and error covariance matrices used to compute the gains in the Kalman filter. The procedure is described in Algorithm 2.

6.5.3 Simulations

We will display here the results of Algorithms 1 and 2 for the attitude estimation problem described in Section 6.2.2:

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

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

the vectors b_1 and b_2 (for instance the gravity and the magnetic field), V_n^1 and V_n^2 are sequences of independent isotropic white noises. A simulation has been performed using the parameters given in Table 6.2. In this case the IEKF equations are described by Algorithm 3. As the state-of-the-art methods for this particular problem are the Multiplicative Extended Kalman Filter (MEKF, see e.g. [34]) and the USQUE filter [33] (a quaternion implementation of the Unscented Kalman Filter), the four methods are compared. The evolution of the gains is displayed on Fig. 6.2 and shows the interest of the invariant approach. The error variable is expressed in the Lie algebra and its projection to the first axis is given on Fig. 6.3 for each method (the projections on other axes being very similar, they are not displayed due to space limitations). All perform satisfactorily, but some differences are worthy of note. First we can see that the MEKF and the IEKF yield comparable performances but only the gains of the IEKF converge. Moreover, the linearizations lead the MEKF and the IEKF to fail capturing accurately the error dispersion through a 3σ-envelope. To this respect, the USQUE performs better than the two latter filters but still does not succeed in capturing the uncertainty very accurately. On the other hand, the envelope provided by the EnKF is very satisfying. This result is not surprising: this envelope has been computed using a sampling of the true density of the error, there is thus no reason why it should not be a valid approximation as long as there are sufficiently many particles. This method is thus to be preferred to the other ones when the user is willing and able to perform extensive simulations, and has the capacity to store the gains over the whole trajectory.

Algorithm 3 Invariant Extended Kalman Filter on $SO(3)$

Returns at each time step $\hat{R}_n$, P_n such that $R_n \approx \exp(\xi_n)\hat{R}_n$, where ξ_n is a centered Gaussian and $var(\xi_n) = P_n$.
$\hat{R}_0$ and P_0 are inputs.
$H_\xi = \begin{pmatrix} (b_1)_\times \\ (b_2)_\times \end{pmatrix}$, $H_V = I_6$,
$Q_V = \begin{pmatrix} Var(V^1) & 0 \\ 0 & Var(V^2) \end{pmatrix}$, $Q_w = \int_{t_n}^{t_{n+1}} Var(w_s^\omega) ds$.
Set $n = 0$.
loop
Let Ω_n be the solution at t_{n+1} of $T_{t_n} = I_3$, $\frac{d}{dt} T_t = (\omega_t)_\times$.
$P_{n+1} = P_n^+ + Q_w$,
$S_{n+1} = H_V Q_V H_V^T + H_\xi P_{n+1} H_\xi^T$,
$L_{n+1} = P_{n+1} H_\xi^T S_{n+1}^{-1}$,
$P_{n+1}^+ = (I - L_{n+1} H_\xi) P_{n+1}$,
$\hat{R}_{n+1} = \hat{R}_n^+ \Omega_n$,
$\hat{R}_{n+1}^+ = \exp\left(L_{n+1} \left[\begin{pmatrix} \hat{R}_{n+1} Y_{n+1}^1 \\ \hat{R}_{n+1} Y_{n+1}^2 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right] \right) \hat{R}_{n+1}$,
$n \leftarrow n + 1$.
end loop

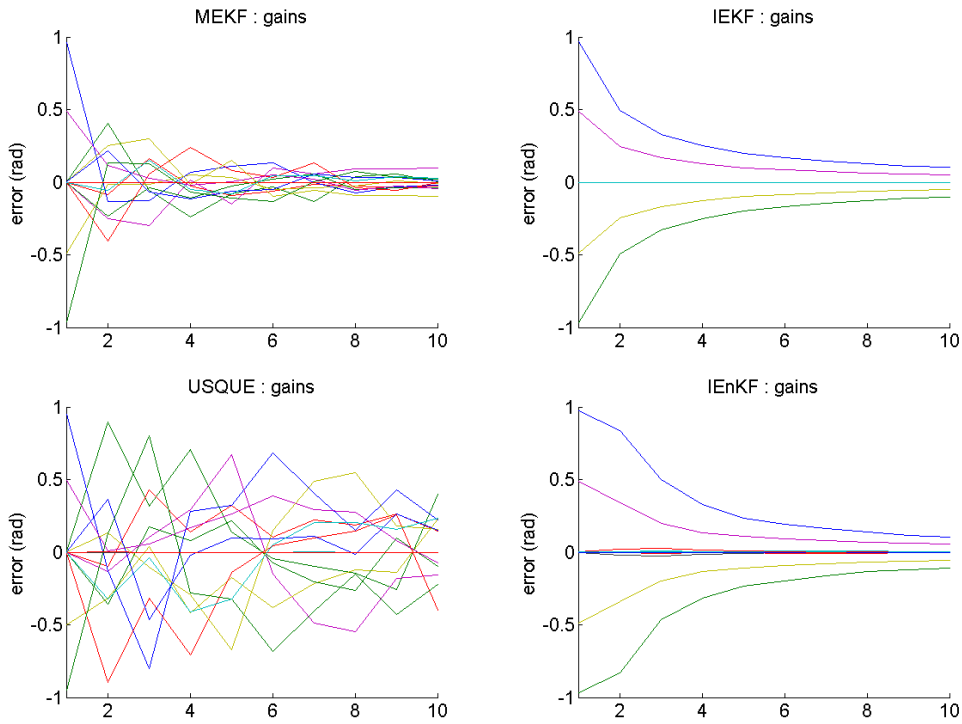


Figure 6.2 – Evolution over time of the coefficients of the gain matrix for MEKF, IEKF, USQUE and IEnKF. The coefficients of the MEKF and USQUE have an erratic evolution, whereas those of the IEKF are convergent, allowing to save much computation power using asymptotic values. Moreover, in this case, only 4 of the 18 coefficients are not equal to zero for the IEKF, allowing sparse implementation.

Table 6.2 – Observation of two vectors: parameters

Parameter	b_1	b_2	Q^{V_1}	Q^{V_2}
Value	$(1,0,0)$	$(0,1,0)$	$0,0873^2 I_3$	$0,0873^2 I_3$
Parameter	P_0	Q^w	N	Simulations
Value	$0.5236^2 I_3$	$0,01745^2 I_3$	50	1000

6.6 Conclusion

This chapter has introduced a proper stochastic framework for some filtering problems on Lie groups having autonomous error equation. This implies to consider not any more a classical dynamical system, but a full probability density function. Several methods to tune the gains have been proposed. When they are held fixed, the estimation error has the remarkable property to provably converge in distribution. Yet, the results obtained can look restricted compared to the difficulty of the corresponding proofs (see B.1). The deterministic properties derived for example in Chapter 4 where much more general and

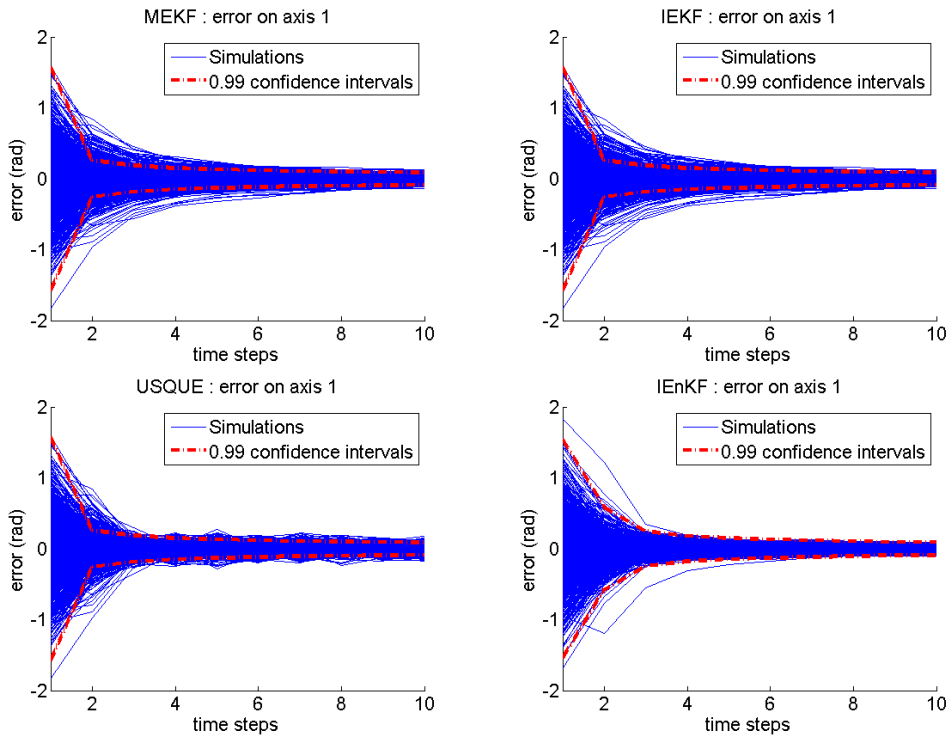


Figure 6.3 – Evolution of the error variable, projected on the first axis, for 1000 simulations using MEKF, IEKF, USQUE and IEnKF. The MEKF and the IEKF give similar results. The USQUE is doing slightly better but is outperformed by the IEnKF. Only the latter captures the error dispersion satisfactorily, as the 3σ envelope contains 99 $\%$ of the simulated trajectories.

have been easily applied to a large variety of examples. This shows the extreme complexity of the study of stochastic systems and explains the prevalence of the deterministic approach in the theory of estimation, as well as in the present work.

Chapter 7

Sampling-based methods

Part of the content of this chapter has been presented at the 2014 IEEE Conference on Decision and Control and led to the registration of a patent (n° 14-01535).

Chapter abstract This Chapter proposes to approach the discrepancy between actual systems and the ideal situation studied in Chapter 4 through sampling. We transpose the idea of Rao-blackwellized particle filtering [39], where only a subset of the variables is sampled and the rest is marginalized out using (if possible) linear Kalman filters running in parallel for each value of the sampled variables. Here, the remaining system is sought to have the form allowing its conditional density to be estimated using IEKFs running in parallel. Under proper conditions leading to autonomy of a partial error, the gains are identical for all particles, allowing an extremely cheap implementation of the particle filter.

7.1 Introduction

In this short chapter, we present two specific cases of 5.1 where sampling-based methods can be an interesting approach. We first discuss in Section 7.2 a property of biased dynamics. Then we develop in Section 7.3 the idea of introducing a sampling to reduce a complicated problem to the form studied in Part I. Of course, the procedure requires some hypotheses on the form of the equations. To make the understanding of the idea easier, we will start with two IEKFs running in parallel, without selection among them, and discuss the benefits of the situation over the general case of two EKFs. Then we introduce the full method, which resorts to a possibly large number of IEKF estimates. In Section, 7.4, we detail the example of a GPS-aided inertial navigation handling jumping biases in the position provided by the GPS.

7.2 Sampling biased dynamics

This section is a mere remark about dynamics of the form 5.1, where the additional variables gathered in Θ_t actually take the form of a bias b on the dynamics. The result

presented is very easy but has to be taken into account in any method requiring sampling-based propagation of a p.d.f. The equation considered takes the form:

$$\frac{d}{dt}\chi_t = f_{u_t}(\chi_t) + \chi_t b, \tag{7.1}$$

where f verifies relation (3.13) and b is an unknown element of the Lie algebra $\mathfrak{g}$. A nice separation principle is verified:

Proposition 20. *Let $\hat{b} \in \mathscr{B}$ denote a particular value of b and $\hat{\chi}_t$ a corresponding particular solution of Equation* (7.1)*. If (χ_t, b) is another couple of solutions, $\eta_t^L = \chi_t^{-1}\hat{\chi}_t$ the left-invariant error of the χ component and $\Delta b = b - \hat{b}$ the linear error of the b component we have:*

$$\eta_t^L = \eta_t^\chi \eta_t^b,$$

with:

$$\eta_0^\chi = \eta_0, \qquad \frac{d}{dt}\eta_t^\chi = g_{u_t}^L(\eta_t^\chi) + \eta_t^\chi \hat{b},$$

$$\eta_0^b = Id, \qquad \frac{d}{dt}\eta_t^b = g_{u_t}^L(\eta_t^b) + \eta_t^b \hat{b} + \eta_t^b(\Delta b),$$

where $g_{u_t}^L$ is defined by Equation (3.9)*.*

Verification is trivial. This result means the evolution of the error due to the initial perturbation η_0^L and the one due to the bias perturbation $\Delta b = b - \hat{b}$ can be fully decoupled and contained in separate variables (η_t^χ and η_t^b) , which is usually true only up to the first order for nonlinear systems. Once again, we managed to retrieve a property of the linear case. For a given value of $\hat{b}_t$ the dynamics $\frac{d}{dt}\chi_t = f(\chi_t) + \chi_t \hat{b}$ verifies (3.13) as the sum of two functions verifying (3.13)(see 5). Thus the first-order approximation of the evolution of η_t^χ is actually exact (see Theorem 4) and the only approximations in a first-order propagation of η_t^L come from second order terms in the integration of the bias. For instance, if sigma-points are sampled (see [62]), only those corresponding to a variation of the biases seem really necessary. Moreover, the remaining sigma-points are really required only if the second-order influence of biases is sufficiently important.

This is also true for particle filtering: a sampling of the full density requires to integrate χ_t for each sampled bias but then, all the values of χ_t for this bias and different initializations are contained in one particular solution and the first-order expansion of the error (Corollary 1). The cost of maintaining a representative sampling of the density thus goes down from $e^{\dim G + \dim b}$ to $e^{\dim b}(\dim G)^2$.

Next section gives a second way to leverage autonomous error properties for sampling-based methods.

7.3 The invariant Rao-Blackwellized filter

7.3.1 A conditional invariance hypothesis

Assume the considered system can be divided into a set of variables gathered in a vector Θ_t, and an element χ_t of a matrix Lie group G, such (Θ, χ_t) verifies an equation of the

form:

$$\frac{d}{dt}\Theta_t = \varphi(\Theta_t, \nu_t), \tag{7.2}$$

$$\frac{d}{dt}\chi_t = f_{u_t}(\Theta_t, \chi_t) + \chi_t w_t, \tag{7.3}$$

$$Y_{t_n} = \psi_{\Theta_{t_n}}(\chi_{t_n} \cdot d + V_n), \tag{7.4}$$

where w_t and ν_t are Langevin noises, $\psi_\Theta : \mathscr{Y} \to \mathscr{Y}$ is invertible for any Θ and $f_\cdot(\cdot,\cdot)$ verifies 3.8 as well as the additional hypothesis:

$$\forall u, \Theta, \eta, f_u(\Theta, \eta) - f_u(\Theta, I_d)\eta = f_u(0, \eta) - f_u(0, I_d)\eta. \tag{7.5}$$

The additional parameter Θ can affect the observation as in the example of Section 7.4 and/or the dynamics. An example of the latter case is given by the attitude χ_t (here a rotation matrix) of a body endowed with a gyroscope undergoing the combination of a bias Θ and an isotropic white noise, the stochastic evolution of the bias being possibly complicated.

7.3.2 Invariant Extended Kalman Filters running in parallel

We first illustrate the idea in a simplified situation where we consider two distinct trajectories $(\Theta_t^1)_{t\geq 0}$ and $(\Theta_t^2)_{t\geq 0}$ of (7.2) and let two Invariant Extended Kalman Filters run in parallel to filter the subsystem (7.3)-(7.4) where Θ_t is considered as an input.

Proposition 21. *Let $(\Theta_t^1)_{t\geq 0}$ and $(\Theta_t^2)_{t\geq 0}$ be two solutions of* (7.2)*. The equations* (7.3)-(7.4) *define then two different systems, one associated to Θ_t^1 and the other to Θ_t^2. If these systems are filtered by two IEKFs $(\hat{\chi}_t^i)_{i=1,2}$ of the form:*

$$\frac{d}{dt}\hat{\chi}_t^i = f_{u_t}\left(\Theta_t^i, \hat{\chi}_t^i\right), \tag{7.6}$$

$$\hat{\chi}_{t_n}^{i+} = \hat{\chi}_{t_n}^i \exp\left(K_n^i\left[(\hat{\chi}_t^i)^{-1}\psi_{\Theta_t^i}^{-1}(Y_n) - d\right]\right), \tag{7.7}$$

then the Kalman gains K_n^i are identical for both *systems.*

Proof. It suffices to write the error equation associated to (7.6)-(7.7) and use (7.5) to obtain:

$$\frac{d}{dt}\eta_t^i = f_{u_t}(0, \eta_t^i) - f_{u_t}(0, I_d)\eta_t^i - w_t\eta_t^i, \tag{7.8}$$

$$\eta_{t_n}^{i+} = \eta_{t_n}^i \exp\left[K_n\left(\left(\eta_{t_n}^i\right)^{-1} \cdot d - d + V_n\right)\right], \tag{7.9}$$

which is independent of Θ_t^i. The linearization can be performed as in Chapter 4. □

7.3.3 Invariant Rao-Blackwellized algorithm

Inspired by the Rao-Blackwellized particle Filter, the Invariant Rao - Blackwellization consists of fully exploiting the conditional invariance property in the following way. The state being partitioned into (Θ_t, χ_t), the variable Θ_t can be sampled using particles $(\Theta_t^j)_{1 \leq j \leq N_p}$, each particle having a weight w^j reflecting its likelihood assigned to it. There are numerous sampling methods that we will not detail here. Each time a new observation Y_n is available, the weight is updated and a re-sampling step may take place. The distribution of the state conditioned on the past outputs can then merely be factored using the chain rule as follows:

$$\mathbb{P}(\chi_t, \Theta_t | Y_1, \ldots, Y_n) = \mathbb{P}(\Theta_t | Y_1, \ldots, Y_n) \mathbb{P}(\chi_t | \Theta_t, Y_1, \ldots, Y_n).$$

This distribution can be approximated combining particle filters and IEKFs to approximate $\mathbb{P}(\Theta_t | Y_1, \ldots, Y_n)$ and $\mathbb{P}(\chi_t | \Theta_t, Y_1, \ldots, Y_n)$ in the following way:

- $\mathbb{P}(\Theta_t | Y_1, \ldots, Y_n)$ is computed using a particle filter, that is, is approximated by the distribution $\sum_{j=1}^{N_p} w^j \delta_{\Theta^j}(\Theta_j)$.
- $\mathbb{P}(\chi_t | \Theta_t^j, Y_1, \ldots, Y_n)$ is approximated by a normal law with mean and covariance output by an IEKF.

Using Proposition 21 we see that as soon as the system satisfies the conditional invariance property, the Kalman gains and covariances are identical for all particles, leading to a much decreased numerical complexity.

7.4 Application to navigation with jumping GPS

The method described above will now be used to integrate an advanced non-Gaussian and non-white GPS noise model into a localization algorithm. The problem of inertia/GPS fusion has been addressed using several kinds of Kalman filters [35, 55, 82], but most of them rely on Gaussian approximations. In [82] for instance, the GPS error is a Gaussian white noise, in [35] it is modeled as the sum of a bias following a small random walk and a white noise. Here, the GPS estimates are assumed to be polluted not only by white noise, but also by an offset that abruptly jumps at unknown times. Such a model reflects the realistic effects of highly correlated noises in GPS and the way we propose to handle it opens the door to the efficient integration of an even more sophisticated noise model.

7.4.1 Model

The inertial sensors delivering estimates at high rate, and the GPS at a much lower rate, we opt for the following continuous-time model with discrete observations, based on the

equations of a rigid body moving on a flat earth:

$$\frac{d}{dt}R_t = R_t(\omega_t + w_t^{\omega})_\times, \tag{7.10}$$

$$\frac{d}{dt}v_t = R_t(a_t + w_t^{a}) + g, \tag{7.11}$$

$$\frac{d}{dt}x_t = v_t, \tag{7.12}$$

$$Y_{t_n} = x_{t_n} + b_{t_n} + V_n, \tag{7.13}$$

where for $u \in \mathbb{R}^3$, $(u)_\times$ is the skew-symmetric matrix such that $\forall z \in \mathbb{R}^3, (u)_\times z = u \times z$, R_t encodes the orientation of the body, v_t its velocity, x_t its position, b_t is the (jumping) GPS bias, g the gravitation vector, Y_{t_n} the GPS measurement, $\omega_t \in \mathbb{R}^3$ the gyroscope measurement, $a_t \in \mathbb{R}^3$ the accelerometer measurement, w_t^{ω} the gyroscope noise, w_t^{a} the accelerometer noise (the covariance matrix of the concatenated noise (w_t^{ω}, w_t^{a}) will be denoted by Q), and V_n the GPS noise, assumed to be isotropic (see Remark 10) with variance $r^2 I_d$. The bias b_t is modeled as follows: there exist a sequence of times $(\tau_k)_{k\geqslant 0}$ independent of other variables and unknown to the user such that b_t is constant on $[\tau_k, \tau_{k+1}[$ and b_{τ_k} is a centered Gaussian variable with variance $\sigma_b^2 I_d$, independent of the past. The time elapsed between τ_k and τ_{k+1} follows an exponential law with known intensity, and we let p_J denote the probability of the bias to have jumped at leas once between two consecutive observations (p_J is constant here because we assume the GPS information arrives at a constant rate but all the computations hold if p_J is a function of $(t_n - t_{n-1})$).

Remark 10. *The isotropic GPS noise assumption is not really restrictive as the method still works if the noise V_n is isotropic in the horizontal directions x, y only, and larger along axis z. This hypothesis is classical in GPS models.*

7.4.2 IEKF for known biases

In this preliminary subsection we assume the bias sequence $\tilde{b}_n = b_{t_1}, \ldots, b_{t_n}$ to be known. In this case the biases can be removed from the sequence of GPS measurements $\tilde{Y}_n = Y_{t_1}, \ldots, Y_{t_n}$ and the obtained system is the one described in 4.3.5. Using the same matrix embedding $\hat{\chi}_t$ and letting the error be $\eta_t = \chi_t^{-1}\hat{\chi}_t$ we have

$$\frac{d}{dt}\eta_t = g_{u_t}(\eta_t, \omega_t, a_t) = f_{u_t}(\eta_t) - f_{u_t}(I_5)\eta_t,$$

and the matrices used in the gains computation are the same as in 4.3.5:

$$A_t = \begin{pmatrix} -(\omega_t)_\times & 0_{3\times 3} & 0_{3\times 3} \\ -(a_t)_\times & -(\omega_t)_\times & 0_{3\times 3} \\ 0_{3\times 3} & I_3 & -(\omega_t)_\times \end{pmatrix}, \quad H = \begin{pmatrix} 0_{3\times 3} \\ 0_{3\times 3} \\ I_3 \end{pmatrix}. \tag{7.14}$$

For χ_t close to $\hat{\chi}_t$ the Gaussian error approximation reads:

$$\mathbb{P}\left(\exp^{-1}(\chi_{t_n}^{-1}\hat{\chi}_{t_n}) = \xi \,\middle|\, \tilde{b}_n, (\tilde{Y}_n)\right) \approx \mathcal{N}(\xi, P_{t_n}, 0),$$

where $\mathcal{N}(x, \Sigma, \mu)$ is the value at x of the Gaussian density of mean μ and covariance matrix Σ.

Remark 11. *There are two remarkable features of this filter. First, the Kalman gains do not depend on the bias b_{t_n}. This is logical but it will play an important role in the sequel. Then, around a large and relevant class of trajectories defined by constant inputs ω_t, a_t, the Riccati equation followed by the error covariance matrix is time-independent, and the Kalman gains are thus expected to converge.*

7.4.3 Description of the proposed particle filter

The basis of the algorithm (and more generally of approximate Rao - Blackwellized Particle Filters) is the assumption that for a given sequence of bias values $\tilde{b}_n$ and an instant $t \in]t_n, t_{n+1}]$ the marginal density $\mathbb{P}(\chi_t|\tilde{b}_n, \tilde{Y}_n)$ is well approximated by a Gaussian with mean and covariance output by an IEKF. Thus, only the sequence of bias values has to be sampled, using N_p particles. As we are interested in the density of the current bias, each particle only stores the last bias b_t^j and conditional mean $\hat{\chi}_t^j$ (the conditional covariance matrix is computed using the invariant error equation). If the b_t^j are i.i.d sampled particles an empirical estimate of the posterior distribution $\mathbb{P}\left(b_t, \chi_t \middle| \tilde{Y}_n\right)$ of the whole current state (b_t, χ_t) then writes:

$$\mathbb{P}\left(b_t, \chi_t \middle| \tilde{Y}_n\right) = \frac{1}{N_p} \sum_{j=1}^{N_p} \delta_{b_t^j}(b_t) \mathbb{P}\left(\chi_t \middle| \tilde{b}_n^j\right).$$

As it is impossible to sample efficiently from the posterior distribution, we use importance sampling that weights the particles according to their likelihood. The following steps are thus repeated over time:

Propagation: the conditional estimates $\hat{\chi}_t^j$ evolve following the deterministic part of (7.10), (7.11), (7.12) and the Riccati equation defined by (7.14) is integrated.

Re-sampling: when a measure Y_n is available the bias of each particle is re-sampled following the law $\mathbb{P}(b_n|\tilde{b}_{n-1}^j, \tilde{Y}_n)$.

Re-weighting: the weight attached to each particle is multiplied by $\mathbb{P}(Y_n|\tilde{b}_{n-1}^j, \tilde{Y}_{n-1})$, then the weights are normalized. A re-sampling can be performed.

Marginal density update: the conditional mean $\hat{\chi}_{t_n}^j$ associated to each particle is updated using (4.47) and the sampled bias b_n^j. The (particle independent) invariant covariance matrix P_{t_n} is updated.

The detailed particle filter is described by Algorithm 4 and the various derivations leading to it are described in the following subsection.

7.4.4 Main formulas derivation

To implement the procedure we should be able to:

1. Compute $\mathbb{P}(\chi_{t_n}|\tilde{b}_n^j, \tilde{Y}_n)$ from $\mathbb{P}(\chi_{t_{n-1}}|\tilde{b}_{n-1}^j, \tilde{Y}_{n-1})$, b_n, Y_n.
2. Compute $\mathbb{P}(Y_n|\tilde{b}_{n-1}^j, \tilde{Y}_{n-1})$ from P_{t_n}, b_{n-1}^j and $\hat{\chi}_{t_n}^j$.
3. Compute $\mathbb{P}(b_n|\tilde{b}_{n-1}^j, \tilde{Y}_n)$ from P_{t_n}, b_{n-1}^j and $\hat{\chi}_{t_n}^j$.

Computation of $\mathbb{P}(\chi_{t_n}|\tilde{b}_n^j, \tilde{Y}_n)$

We obtained in Section 7.4.2 that $\mathbb{P}(\phi^{-1}(\chi_{t_n}^{-1}\hat{\chi}_{t_n}) = \xi|\tilde{b}_n, \tilde{Y}_n) \approx \mathcal{N}(\xi, P_{t_n}, 0)$. We can extract the relation $\mathbb{P}(x_{t_n} = x|\tilde{b}_n^j, \tilde{Y}_n) \approx \mathcal{N}(x, \hat{R}_{t_n}^j H P_{t_n} H^T \hat{R}_{t_n}^{jT}, \hat{x}_t^j)$ that will prove useful in the sequel.

Computation of $\mathbb{P}(Y_n|\tilde{b}_{n-1}^j, \tilde{Y}_{n-1})$

Let J_n be the event "the bias jumps between t_n and t_{n+1}". This event is independent of $\chi_{t_{n+1}}$. We let p_J denote its prior probability, and $\bar{J}$ its opposite. If J occurs, according to (7.13), Y_n is the sum of three independent Gaussian vectors: x_{t_n}, b_{t_n} and V_n. Thus

$$\mathbb{P}(Y_n|J, \tilde{b}_{n-1}^j, \tilde{Y}_{n-1}) = \mathcal{N}(Y_n,\ R_{t_n}^T H^T P_{t_n} H R_{t_n} + \sigma_b^2 I_3 + r^2 I_3, \hat{x}_{t_n}^j).$$

If $\bar{J}$ occurs, Y_n is the sum of two independent Gaussian, x_{t_n} and V_n, and a known vector b_{t_n}. Thus

$$\mathbb{P}(Y_n|\bar{J}, \tilde{b}_{n-1}^j, \tilde{Y}_{n-1}) = \mathcal{N}(Y_n,\ R_{t_n}^T H^T P_{t_n} H R_{t_n} + r^2 I_3, \hat{x}_{t_n}^j + b_{t_n}^j).$$

We can conclude introducing the quantities Π_1 and Π_2:

$$\Pi_1 = p_J \mathcal{N}(Y_n,\ R_{t_n}^T H^T P_{t_n} H R_{t_n} + \sigma_b^2 I_3 + r^2 I_3, \hat{x}_{t_n}^j),$$
$$\Pi_2 = (1 - p_J)\mathcal{N}(Y_n,\ R_{t_n}^T H^T P_{t_n} H R_{t_n} + r^2 I_3, \hat{x}_{t_n}^j + b_{t_n}^j),$$

yielding $\mathbb{P}(Y_n|\tilde{b}_{n-1}^j, \tilde{Y}_{n-1}) = \Pi_1 + \Pi_2$.

Computation of $\mathbb{P}(b_n|\tilde{b}_{n-1}^j, \tilde{Y}_n)$

This probability decomposes as follows:

$$\mathbb{P}(b_n|\tilde{b}_{n-1}^j, \tilde{Y}_n) = \mathbb{P}(J|\tilde{b}_{n-1}^j, \tilde{Y}_n)\mathbb{P}(b_n|J, \tilde{b}_{n-1}^j, \tilde{Y}_n) + \mathbb{P}(\bar{J}|\tilde{b}_{n-1}^j, \tilde{Y}_n)\mathbb{P}(b_n|\bar{J}, \tilde{b}_{n-1}^j, \tilde{Y}_n).$$

Conditioning on Y_n we have:

$$\mathbb{P}(J|\tilde{b}_{n-1}^j, \tilde{Y}_n) = \Pi_1/(\Pi_1 + \Pi_2),$$
$$\mathbb{P}(\bar{J}|\tilde{b}_{n-1}^j, \tilde{Y}_n) = \Pi_2/(\Pi_1 + \Pi_2).$$

If J occurs using the fact b_n is a Gaussian and $Y_n - \hat{x}_n^j$ a noisy measurement of it, the linear Kalman equations yield:

$$\mathbb{P}(b_n = b|J, \tilde{b}_{n-1}^j, \tilde{Y}_n) = \mathcal{N}(b, \hat{R}_{t_n}^{jT} P_n^b \hat{R}_{t_n}^j, \hat{R}_{t_n}^{jT} K_n^b \hat{R}_{t_n}^j (Y_n - \hat{x}_n^j)),$$

with $K_n^b = \sigma_b^2(H^T P_{t_n} H + \sigma_b^2 I_3 + r^2 I_3)^{-1}$ and $P_n^b = \sigma_b^2(I_3 - K_n^b)$. If $\bar{J}$ occurs the bias does not change : $\mathbb{P}(b_n = b|\bar{J}, \tilde{b}_{n-1}^j, \tilde{Y}_n) = \delta_{b_{n-1}^j}(b)$. Finally $\mathbb{P}(b_n = b|\tilde{b}_{n-1}^j, \tilde{Y}_n)$ writes:

$$\Pi_1 \mathcal{N}(b, \hat{R}_{t_n}^{jT} P_n^b \hat{R}_{t_n}^j, \hat{R}_{t_n}^{jT} K_n^b \hat{R}_{t_n}^j (Y_n - \hat{x}_n^j)) + \Pi_2 \delta_{b_{n-1}}(b).$$

Remark 12. *To sample from this law (see Algorithm 4) we first choose if the bias jumps (for instance comparing Π_1 to a uniform variable). Then, if it is the case, we use a Cholesky decomposition $P_n^b = (L_n^b)(L_n^b)^T$ and let $b_{n+1}^j = R_{t_n}^{jT} K_n^b R_{t_n}^j (Y_n - \hat{x}_{t_n}^j) + R_{t_n}^{jT} L_n^b X$ where X is sampled following a standard three dimensional normal law. Note that a direct Cholesky decomposition of the variance $R_{t_n}^{jT} L_n^b R_{t_n}$ would be a bad choice as the computation would have to be done for each particle.*

Algorithm 4 Invariant Rao - Blackwellization

The prior estimation $\hat{\chi}_0$ and variance P_0 are supposed known.
for $n = 1$ **to** T **do**
  Solve $R_0 = I_3,\ \frac{d}{dt}R_t = R_t\omega_{t_n+t}$ on $[0,\Delta t]$.
  Compute $v_{\Delta t} = \int_0^{\Delta t} R_t a_{t_n+t}$ and $x_{\Delta t} = \int_0^{\Delta t} v_t$.
  Solve $\frac{d}{dt}P_t = F_tP_t + P_tF_t^T + Q$.
  Compute $S_n = HP_{t_n}H^T + r^2I_3$ and $K_n = P_{t_n}H^TS_n^{-1}$.
  Compute $K_n^b = \sigma^2(H^TP_{t_n}H + \sigma^2I_3 + r^2I_3)^{-1}$, $P_n^b = \sigma^2(I_3 - K_n^b)$.
  Compute L_n^b such that $(L_n^b)(L_n^b)^T = P_n^{b+}$.
  for $j = 1$ **to** N_p **do**
    $R_{t_{n+1}}^j = R_{t_n}^{j+}R_{\Delta t},\ v_{t_{n+1}}^j = v_{t_n}^{j+} + R_{t_n}^{j+}v_{\Delta t} + (\Delta t)g$.
    $x_{t_{n+1}}^j = x_{t_n}^{j+} + (\Delta t)v_{t_n}^j + R_{t_n}^{j+}x_{\Delta t} + \frac{1}{2}(\Delta t)^2g$.
    Compute $\Pi_1 = \mathbb{P}(Y|J_n)$ and $\Pi_2 = \mathbb{P}(Y|\bar{J}_n)$.
    Sample $c \in [0,1]$ following an uniform law.
    Update the weight: $w^{j+} = (\Pi_1 + \Pi_2)w^j$.
    if $c < \Pi_1/(\Pi_1 + \Pi_2)$ **then**
      Sample X following a 3-dimensional normal law.
      $b^j = R_{t_n}^TK_b^nR_{t_n}Y_n + R_{t_n}^TL_n^bX$.
    end if
    Build $\chi_{t_{n+1}}^j$ using $R_{t_{n+1}}^j,\ v_{t_{n+1}}^j$ and $x_{t_{n+1}}^j$.
    $\chi_{t_{n+1}}^{j+} = \chi_{t_{n+1}}^j \exp(\mathscr{L}(K_nR_{t_{n+1}}^{jT}(Y_n - b^j - x_{t_{n+1}}^j)))$.
    Extract $R_{t_{n+1}}^{j+},\ v_{t_{n+1}}^{j+}$ and $x_{t_{n+1}}^{j+}$ from $\chi_{t_{n+1}}^{j+}$.
  end for
  Compute $P_{t_n}^+ = (I_9 - K_nH)P_{t_n}$.
  Normalize weights and possibly re-sample.
  Compute the mean position $\tilde{x}_t = \sum_{j=1}^{N_p} w^{j+}x^{j+}$.
end for

7.4.5 Some remarks about bias observability

Before the first GPS jump A constant bias is not observable. Thus, the best one can do before the bias has jumped is to efficiently filter the GPS and inertial sensors noise to reconstruct the state up to a position offset.

After one jump As the inertia gives an accurate short-term estimation of the trajectory, the bias jumps can be detected and the value updated using the difference between the estimated position and the GPS observation. But the residual centered noise V_n of the GPS (after bias correction) affects this first estimation and can be handled only using future GPS measurements. Hence the interest of sampling several values of the new bias then selecting those which best fit the observations to come.

After many jumps As the bias is assumed centered, it can be recovered asymptotically. Indeed in the proposed algorithm, the re-weighting will eventually eliminate the particles whose sequence of bias has a mean far from zero. This prevents the position to diverge from the GPS trajectory in the long run, this discrepancy being erroneously explained by a large non-centered estimated bias.

7.4.6 Results

Algorithm 4 has been implemented with the parameters given in table 7.1. The initial state is known. Then the estimate deviates from the true trajectory due to gyroscope, accelerometer and GPS noise. As a small set of parameters is sampled (the components of the GPS bias) using moreover an optimal sampling the filter gives reasonable results even with one particle. The accuracy of the localization is measured using the total Root Mean Square Error (RMSE) of the position over the whole trajectory. Its performance improves as the number of particles grows but stabilizes fast (see table 7.2) as the sampled variable is only 3-dimensional. The computation time is also displayed in table 7.2, for a naive Matlab implementation using "for" loops and it can be seen that the cost of 10 particles is only around twice the cost of one particle. Considering Algorithm 4 in detail the reader could even expect a much smaller difference between the cost of 1 and 10 particles but the result we obtain is affected by Matlab's poor ability to deal with loops over the particles. The end of the trajectory estimated by the Invariant Rao - Blackwellization is drawn on Figure 7.1 and compared to the true one. The apparent gap in the GPS observations is due to a bias change. We see that the filter is able to manage the non-Gaussian noise model of the GPS with an accuracy improved as the number of particles increases.

7.5 Conclusion

This chapter has introduced the idea of using autonomous errors variables to sample from a lower dimensional distribution in particle filtering. A class of system has been presented, for which the combination of invariant and particle filtering brings decisive prop-

Figure 7.1 – Zoom on the end of the trajectory estimated by the Invariant Rao - Blackwellization with 1 and 100 particles, illustrating the usefulness of sampling. The beginning is not informative as the IMU gives an accurate estimation on the short run in both cases. As an accurate near-optimal sampling is performed, a unique particle provides already a reasonable estimate. Sampling 100 particles improves the performance and, due to the properties of invariant filtering, is not much more expansive

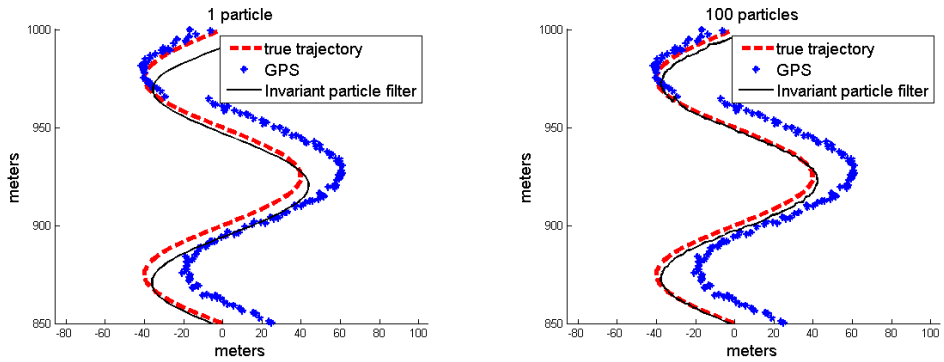


Table 7.1 – Parameters of the simulation

Duration	$1000s$
Variance of the gyroscope noise	$10^{-8} rad/s^2$
Variance of the accelerometer noise	$10^{-6} m^2/s^5$
Probability of a GPS bias jump	$10^{-3} s^{-1}$
Frequency of the GPS observations	$10Hz$
Variance of the regular GPS observations	$1m^2$
Variance of the GPS bias	$100m^2$

erties in terms of computational burden. The realism of the proposed model has been proved by the implementation of the method on a concrete non-linear localization problem with non-white noise. The strong convergence properties of the Invariant Extended Kalman Filter can be expected to have interesting consequences on the accuracy of the obtained sampling, but no result in this direction has been given here and the question is left for future work.

In Part III we come back to a deterministic setting, and step back a little to study, from a more general point of view, the benefits of an EKF based on a non-linear error variable.

Table 7.2 – Results of the simulation : RMSE of the error

Number of particles	1	10	30	100	1000
RMSE (m)	105,62	96,31	75,73	70,37	66,59
Computation time (s)	1,29	2,61	5,64	15,59	150,52

Part III

Nonlinear state error based EKFs

Chapter 8

Remedying EKF-SLAM inconsistency

Chapter abstract In this chapter, we first recall one of the main reasons why, although very well suited to loop closures and as such very appealing, the EKF SLAM has been essentially abandoned. It is known as “EKF-SLAM inconsistency” [8, 31, 59, 61], that is, the over-optimism of the EKF covariance matrix due to linearizations. Then, we show the problem can be fully solved by choosing to linearize a certain nonlinear state error variable inspired by Chapter 4. To our best knowledge, this result is entirely novel.

8.1 Introduction

The problem of a robot discovering a new environment, and trying to map it while estimating its own position on the map, is known as "Simultaneous Localization and Mapping". The probabilistic formulation of this problem introduced in [101], and its resolution using Extended Kalman filtering recommended by the authors of the paper, has been a great success. This is due to the impossibility to neglect the correlations between the different variables [32]. They have to be encoded one way or another and to this end, the Kalman filtering framework is difficult to avoid. Yet, further work [8, 31, 59, 61] showed the limits of the EKF SLAM: the non-linearity of the problem makes the filter inconsistent, in the sense that the error covariance matrix it provides is extremely optimistic, or in other words the estimation error is far beyond its (alleged) the 3-σ envelope. This has been related to a problem known as "false observability": the filter believes it has enough information to improve its estimate of some variables being actually non-observable. For instance, a robot using only relative measurements of unknown landmarks eventually considers its heading is known whereas an overall rotation of the system yields the entire observations sequence unchanged ans thus cannot be distinguished. To circumvent this drawback some work proposed to chose a linearization point forcing the non-observable subspace of the system to keep the right dimension [57, 58]. But in general, EKF was mostly abandoned in favor of two kinds of approaches: batch optimization and particle methods deriving from FastSlam. FastSLAM [84] is a sampling-based estimation method that carries a set of hypothetical trajectories along with weights. It performs data association jointly with

localization and mapping, and thus is very robust to erroneous feature identification. Yet, it comes at a price: the FastSLAM tends to become over confident in some trajectories and discard some other, less likely at a given time although proving more relevant once more information is obtained. In particular, when a loop closure occurs (that is, the same scene is viewed a long time after having been seen for the first time) the filter can fail to properly take it into account as the only thing it can do is increase the weight of the most likely hypothetical trajectories he has kept. On the other hand, the EKF-SLAM is very good at closing loops by maintaining full correlation between the features through the covariance matrix. Its weaknesses come from the linearization step which induces over confidence in the estimates and leads the filter to fail returning relevant estimates. Batch optimization takes as argument a large part of the trajectory [37]. These methods allows re-linearization of the whole sequence, which the EKF cannot make.

Here, we will carefully analyze the problem of false observability and show that it is strongly coordinate-dependent, in the sense that simply changing the definition of the error variable can make it vanish. This strange phenomenon is better grasped once the error variable has been understood as a local state-dependent basis. First, we introduce in the Section 8.2 the tools required to study observability issues arising in the context of the EKF. We use them in Section 8.3 to develop an analysis of EKF SLAM inconsistency mainly relying on existing ideas [31,61] yet presenting them, we believe, in a much simpler way. Then we show in Section 8.4 how to solve this problem through the choice of the error variable and we finally give in Section 8.5 and 8.6 an equivalent approach allowing a more physical understanding of the mechanisms at play.

8.2 Kalman filter and observability

8.2.1 Observability

Observability is a classical concept of the automatic control theory, whose definition is quite transparent. Consider a general dynamical system $X_t \in \mathbb{R}^n_t$, associated to a sequence of observations $(Y_n)_{n\geqslant 0}$.

$$\frac{d}{dt}X_t = f_{u_t}(X_t), \tag{8.1}$$

$$Y_n = h(X_{t_n}), \tag{8.2}$$

where f is the function describing the evolution of the system X_t, u_t an input and h the observation function. Between the instants t_1 and t_2 the differential equation moves any vector X to a to a new position denoted by $\Phi^{t_2}_{t_1}(X)$, where the function Φ is called the "flow" associated to f. This denomination is visual: if the system lets itself be driven by the equation $\frac{d}{dt}X_t = f_{u_t}(X_t)$ and is at X at time t_1, then it is at $\Phi^{t_2}_{t_1}(X)$ at t_2. Note that we can have $t_1 > t_2$, in which case the flow is traced back from t_2 to t_1. In more mathematical terms we have $\Phi^{t_2}_{t_1}(X) = \left[\Phi^{t_1}_{t_2}\right]^{-1}(X)$.

Definition 8 (non-observable transformation)**.** *We say a transformation $\phi : \mathbb{R}^n \to \mathbb{R}^n$ of the system is non-observable if for any $X_0 \in \mathbb{R}^n$ and $n \geqslant 0$ we have:*

$$h\left(\Phi_0^{t_n}\left[\phi(X_0)\right]\right) = h\left(\Phi_0^{t_n}(X_0)\right).$$

It concretely means that if the transformation is applied to the initial state then none of the observations Y_n are going to be affected. As a consequence, there is no way to guess this transformation has been applied. This notion has an infinitesimal counterpart strongly related to the notion of infinitesimal observability [48] that we define as follows:

Definition 9 (non-observable direction)**.** *Let $(X_t)_{t\geqslant 0}$ denote a solution of* (8.1)*. A vector $\delta X_0 \in \mathbb{R}^n$ is said to span a non-observable direction around X_0 if we have:*

$$\forall n \geqslant 0, H_n \delta X_{t_n} = 0,$$

where H_n is the linearization of h around X_{t_n} and $\delta X_{t_n} = D\Phi_0^{t_n}(X_0)\delta X_0$ is the first-order propagation of the perturbation δX_0 until t_n ($D\Phi_0^{t_n}(X_0)$ being the differential of $\Phi_0^{t_n}$ around X_0).

The denomination δX_t for a vector aims here at highlighting it has to be understood as an infinitesimal perturbation of the trajectory. This definition, in a more common language, could be reformulated as follows: if the initial state was $X_0 + \delta X_0$ instead of X_0, we would not see any difference on the sequence of observations up to the first order. The condition $H_n \delta X_{t_n} = 0$ means $h(X_{t_n} + \delta X_{t_n}) = h(X_{t_n})$, still up to the first order. If f and h are linear these notions are equivalent: a direction δX_0 is non-observable if and only if the translation of the system by the vector δX_0 is non-observable. An estimation method providing the uncertainty of its own result (as the EKF does) should be able to detect that a direction is not observable, as obtaining an estimate accurate along this direction is impossible. This is the question studied in the rest of this section.

8.2.2 Kalman filtering and Extended Kalman filtering

Consider the specific case of a process X_t following a linear dynamics, polluted by Gaussian white noise w_t:

$$\frac{d}{dt}X_t = A_t X_t + B_t + w_t,$$

where matrices A_t and B_t are known and w_t is a random Gaussian white noise with covariance matrix Q_t. The value of this process is estimated using a sequence of linear observations Y_n, also polluted by a white noise V_n having covariance matrix R_n:

$$Y_n = H_n X_{t_n} + V_n.$$

If the distribution of the initial state X_0 is Gaussian with mean $\hat{X}_0 = \mathbb{E}\left[X_0\right]$ and covariance matrix $P_0 = Cov(X_0) = \mathbb{E}\left[(X_0 - \hat{X}_0)(X_0 - \hat{X}_0)^T\right]$, then the exact expected value $\hat{X}_t$ of X_t, as well as the conditional covariance matrix P_t given the past observations, are given by the

equations of the Kalman filter. We recall them here, in a way allowing to establish the terminology we are going to use in the rest of this document.

Propagation (while no observation is available):

$$\frac{d}{dt}\hat{X}_t = A_t\hat{X}_t + B_t,$$

$$\frac{d}{dt}P_t = A_tP_t + P_tA_t^T + Q_t, \qquad t_{n-1} < t < t_n. \tag{8.3}$$

Update (and each time an observation Y_n appears):

- Computation of the innovation z_n and its covariance matrix S_n :

$$z_n = Y_n - H_n\hat{X}_{t_n},$$

$$S_n = H_nP_{t_n}H_n^T + R_n. \tag{8.4}$$

- Estimation of the error ΔX:

$$K_n = P_{t_n}H_n^TS_n^{-1},$$

$$\Delta X = K_nz_n.$$

- Update of the mean and covariance matrix:

$$\hat{X}_{t_n}^+ = \hat{X}_{t_n} + \Delta X,$$

$$P_{t_n}^+ = [I - K_nH_n]P_{t_n}.$$

This is the exact solution to the problem of filtering in the linear case: we have $\mathbb{P}(X_t|(Y_n)_{t_n<t}) = \mathcal{N}(X_t, \hat{X}_t, P_t)$ where $\mathcal{N}$ denotes the multivariate Gaussian density :

$$\mathcal{N}(x,\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}}\exp\left((x-\mu)^T\Sigma^{-1}(x-\mu)\right),$$ with k the dimension of x.

If we come back to the general non-linear case, the p.d.f. conditioned on the observations does not stay Gaussian and its exact description cannot be derived in a close form. The usual way to deal with this issue is simply to use a first-order linearization of the error $X_t - \hat{X}_t$ to maintain an approximation of the mean and covariance matrix of the system, denoted again by $\hat{X}_t$ and P_t. This approach, called extended Kalman filtering, consists of the following steps:

Propagation (while no observation is available):

$$\frac{d}{dt}\hat{X}_t = f_{u_t}(\hat{X}_t),$$

$$\frac{d}{dt}P_t = A_tP_t + P_tA_t^T + Q_t, \qquad t_{n-1} < t < t_n,$$

where A_t is the Jacobian matrix $\frac{\partial f}{\partial X}$ computed at $\hat{X}_t$, and $Q_t = \left(\frac{\partial f}{\partial w}|_{\hat{X}_{t_n}}\right)Cov(w_t)\left(\frac{\partial f}{\partial w}\hat{X}_{t_n}\right)^T$ is the first-order covariance matrix of the noise perturbing the evolution of X_t.

Update (each time an observation Y_n appears):

- Computation of the innovation z_n and its covariance matrix S_n :

$$z_n = Y_n - H_n \hat{X}_{t_n},$$

where H_n is the Jacobian matrix $\frac{\partial h}{\partial X}$ computed at $\hat{X}_{t_n}$.

$$S_n = H_n P_{t_n} H_n^T + R_n,$$

where $R_n = (\frac{\partial h}{\partial V}|_{\hat{X}_{t_n}}) Cov(V_n) (\frac{\partial h}{\partial V}|_{\hat{X}_{t_n}})^T$ is the observation noise covariance matrix computed at $\hat{X}_t$.

- Estimation of the error ΔX:

$$K_n = P_{t_n} H_n^T S_n^{-1},$$

$$\Delta X = K_n (Y_n - H_n \hat{X}_{t_n}).$$

- Update of the mean and covariance matrix:

$$\hat{X}_{t_n}^+ = \hat{X}_{t_n} + \Delta X,$$

$$P_{t_n}^+ = [I - K_n H_n] P_{t_n}.$$

The obtained estimate $\hat{X}_t$ and covariance matrix P_t stem now from approximations and their behavior is not obvious. In particular, if a direction is not observable, this should be encoded one way or another in P_t, meaning the filter "knows" it has no information along this direction. Failure to correctly handle this situation has been shown to be a major cause of failure of the EKF, especially in SLAM applications [8, 58, 61]. The property we seek is defined now more rigorously.

8.2.3 Error Covariance matrix and non-observability

The covariance σ_C^2 of any scalar scalar variable C defined as the projection of X_t over a known vector $u \in \mathbb{R}^n$ is given by $Cov(C) = u^T P_t u$. If u is the projection over coordinate i then we obtain $\sigma_C^2 = P_t^{i,i}$: the diagonal coefficients of P_t are the covariances of the components of X_t. The uncertainty of C is then usually represented through the standard deviation $\sigma_C = \sqrt{\sigma_C^2}$, as this value is homogeneous to C. If the system is linear, σ_C^2 is the exact covariance of the estimation error over the direction u. But if it is non-linear, then $\hat{X}_t$ and P_t have been obtained through approximations and their validity has to be discussed. This is where the standard deviation σ_C (derived from the computed P_t) proves very useful. Indeed, the difference between the estimate and the true state does not carry any useful information because 1) the notion of “large” difference has to be defined 2) an error due to non-observability does not mean the method is not working but only that the filter has not been given enough information. However, comparing the actual error with the computed standard deviation σ_C is meaningful as it is a test of the ability of the filter to estimate its own uncertainty. For instance, having frequently an error superior to $3\sigma_C$ is a good indicator of failure as the probability of having such an error should be less than 3.10^{-3}.

Note that this uncertainty information can also be very useful to the user or the robot, typically for safety issues: an accurate uncertainty estimation will prevent some risky decisions that could for instance lead to a collision.

Now we would like to translate the (deterministic) notion of non-observability in terms of covariance matrix. If a scalar component X_t^i, or more generally a scalar variable $C = u^T X_t$ is non-observable, its estimated covariance defined as $u^T P_t u$ should stay large in a certain sense. If not, we can conclude there is a problem with the estimation method. In this kind of analyses, remarks such as *"the estimated covariance of the variable C decreases although this direction is not observable"* are common but should not be understood literally. It would be erroneous to expect the covariance over a non-observable direction to be always decreasing as shown by the following linear counter-example:

Example 5. *Consider a linear system consisting of a state $(v, x_t^1, x_t^2)^T \in \mathbb{R}^3$ following the equation:*

$$\frac{d}{dt} v_t = 0, \qquad \frac{d}{dt} x_t^1 = v, \qquad \frac{d}{dt} x_t^2 = v,$$

and assume one observes $y = h(v, x_1, x_2) = x_2 + V$, where V is a white noise. A Kalman filter is run on this simple linear *problem with initial covariance matrix I_3 and observation covariance $Cov(V) = r = 1$. The results are displayed on the left plot of Figure 8.1. We see that the covariance of x_1 is not always increasing although x_1 is not observable. The explanation is simple: as the velocity v is unknown, the covariance of x_1 grows, then reduces as v is observed through x_2, but stays larger than its initial value. This shows that increasing covariance is not the property we seek. Note however that the covariance along this direction is lower bounded by its initial value which is logical.*

For the linear system of the previous example, we see that x_1 defines a non-observable direction as two solutions defined by the initial conditions $(v_0, x_0^1, x_0^2)^T$ and $(v_0, \tilde{x}_0^1, {x_0}^2)^T$ with $\tilde{x}_0^1 \neq x_0^2$ yield the same output trajectory (note that the direction $(0,1,0)$ spanned by their difference lies in $\ker H_x$, with $H_x : (v, x_1, x_2) \to x_2$). However, we see that the covariance matrix projected along this direction can decrease. As the system is linear, we know this not due to an approximation of in the computation of the covariance.

On the other hand, the inverse of the covariance matrix, the so-called *information* matrix, is actually much more suited to the notion of observability: its value is always decreasing along non-observable directions for any Kalman filter tuned with arbitrary noise covariance matrices as shown by the following result.

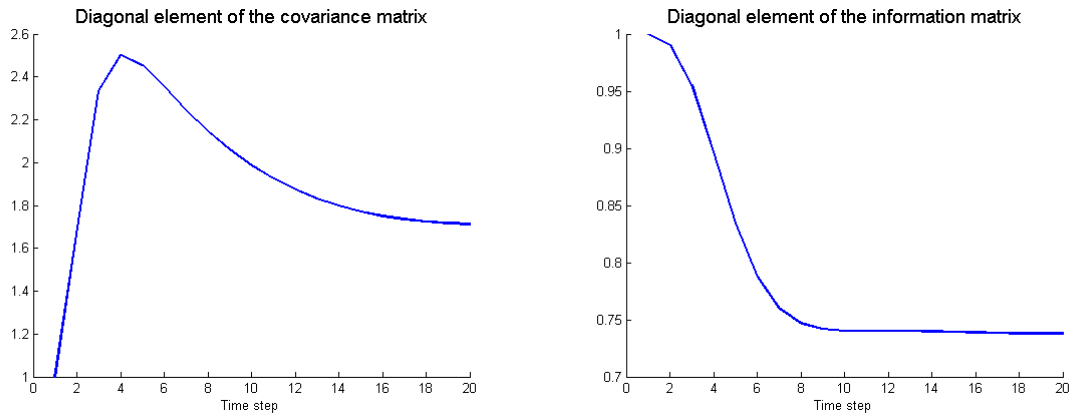


Figure 8.1 – Covariance and information over the non-observable direction x_1 for the simple linear system $\frac{d}{dt}v = 0$, $\frac{d}{dt}x_1 = v$, $\frac{d}{dt}x_2 = v$ with observation $y = x_2 + V$. The variable x_1 is not observable. Yet the variance of x_1 conditioned on the outputs (i.e. $P_t^{2,2}$ if the chosen order is v, x_1, x_2) can decrease as it is correlated to x_2 through the fact they have the same velocity (left plot). Note on the other hand the information over x_1 (defined as $(P_t^{-1})_{2,2}$) never increases, which is consistent with the fact it is not observable.

Proposition 22. *Consider the linear system*

$$\frac{d}{dt}X_t = A_t X_t + w_t,\ Y_{t_n} = H_n X_{t_n} + V_n,$$

and assume the direction δX_0 is non-observable, i.e., we have at all observation time t_n:

$$\delta X_{t_n} \in \ker H_n,$$

where δX_t verifies $\frac{d}{dt}\delta X_t = A_t \delta X_t$.

For arbitrary choices regarding the covariance matrices Q_t and R_n involved in the propagation and update steps of the Kalman filter (8.4)-(8.4)*, and keeping the notation P_t for the covariance matrix they compute, then the information $(\delta X)_t^T P_t^{-1} (\delta X)_t$ along the considered direction $(\delta X_t)_{t \geq 0}$ is monotonically decreasing.*

Proof. During the propagation step using the standard Kalman filter equations and the well-known fact that $\frac{d}{dt}P_t^{-1} = -P_t^{-1}(\frac{d}{dt}P_t)P_t^{-1}$ we have:

$$\begin{aligned}\frac{d}{dt}(\delta X)_t^T P_t^{-1}(\delta X)_t =& (A_t(\delta X)_t)^T P_t^{-1}(\delta X)_t + (\delta X)_t^T P_t^{-1} A_t (\delta X)_t \\ &- (\delta X)_t^T P_t^{-1}\big(A_t P_t + P_t A_t^T + Q_t\big)P_t^{-1}(\delta X)_t \\ =& -(\delta X)_t^T P_t^{-1} Q_t P_t^{-1}(\delta X)_t \leqslant 0.\end{aligned}$$

And at the update step (written here in information form as, e.g., in [103]) we have:

$$(\delta X)_{t_n}^T (P_{t_n}^+)^{-1}(\delta X)_{t_n} = (\delta X)_{t_n}^T \big[P_{t_n}^{-1} + H_n^T R_t^{-1} H_n\big](\delta X)_{t_n} = (\delta X)_{t_n}^T P_{t_n}^{-1}(\delta X)_{t_n},$$

as $H_n(\delta X)_{t_n} = 0$. Thus the information $(\delta X)_{t_n}^T (P_{t_n}^+)^{-1} (\delta X)_{t_n}$ is always decreasing. □

The remarks of this section allow us to give a mathematical definition of a good behavior of an EKF regarding non-observable directions:

Definition 10. *An EKF preserves the non-observable direction δX_0 if the quantity $(\delta X_t)^T P_t^{-1} (\delta X_t)$ is always decreasing, where δX_t follows the differential equation $\frac{d}{dt}\delta X_t = A_t \delta X_t$, A_t denoting the Jacobian matrix $\frac{\partial f}{\partial X}$ computed at $\hat{X}_t$.*

Remark 13. *An immediate consequence of Proposition 22 is that a non-observable direction δX_t of the non-linear system is preserved by an EKF if it is also a non-observable direction of the system linearized over the estimated trajectory $\hat{X}_t$. This not surprising: the EKF "sees" only the linearized system and computes its uncertainty consequently.*

The next section approaches a famous false observability problem and proposes a new solution involving no complicated patch-up job.

8.3 SLAM inconsistency

It is now well-known, thanks to the works of [8, 58, 61], that the EKF-SLAM algorithm with relative observation of unknown features tends to gain information, especially on the heading uncertainty, although this variable cannot ever be inferred (a rotation of the full system is non-observable). This is essentially due to the changes in the linearization point over the observation sequence, which is a typically nonlinear problem, and has been well analyzed [57, 61]. In this section, we advocate part of the EKF-SLAM inconsistency can be remedied by choosing appropriate estimation error variables when designing an EKF. This will first be shown on the standard "steered" bicycle model [43]. The state is defined as:

$$\chi_t = (\omega_t, X_t, p_t),$$

where $\theta_t \in \mathbb{R}$ denotes the heading, $X_t \in \mathbb{R}^2$ the 2D position of the vehicle, $p_t \in \mathbb{R}$ the position of an unknown feature and χ_t a synthetic notation for the state of the system. Its dynamics reads:

$$\begin{aligned} \frac{d}{dt}\theta_t &= \omega_t, \\ \frac{d}{dt}X_t &= R(\theta_t)\begin{pmatrix} v_t \\ 0 \end{pmatrix}, \\ \frac{d}{dt}p_t &= 0, \end{aligned} \tag{8.5}$$

where $\omega_t \in \mathbb{R}$ denotes the angular velocity of the vehicle, v_t the odometric velocity and $R(\theta)$ is the matrix encoding a rotation by an angle θ:

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Note that we simplified the notations: ω_t can be expressed as a function of the velocity, wheel base and steer angle but keeping the formulas as simple as possible looked preferable. The observation of feature can be any function defined in the reference frame of the vehicle:

$$Y_n = h\left[R(\theta_t)^T\left(p - X_{t_n}\right)\right], \tag{8.6}$$

where $Y_n \in \mathbb{R}^p$ is the observation of the feature at time t_n and h is a function of $\mathbb{R}^2 \to \mathbb{R}^p$.

Remark 14. *Note that the observation model* (8.6) *encompasses the usual range and bearing observations used in the SLAM problem as they read:*

$$Y_n = h_{r,b}\left(R(\theta_t)^T\left(p - X_{t_n}\right)\right),$$

$$\textit{with:} \qquad h_{r,b}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \sqrt{y_1^2 + y_2^2} \\ \arctan\left(\frac{y_2}{y_1}\right) \end{pmatrix}.$$

If we choose $h\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \arctan\left(\frac{y_2}{y_1}\right)$ *instead, we obtain monocular SLAM, another classical model. Note also we do not provide any form for the noise in the output: this is because the properties we are about to prove only depend on the deterministic part of the system so they are insensitive to the way the noise enters the system.*

All the measures being in the reference frame of the vehicle, an initial rotation and/or translation of both the vehicle and the map (i.e. the set of features) is non-observable. We will now examine the system linearized around an estimated trajectory $(\hat{\theta}_t, \hat{X}_t, \hat{p}_t)_{t>0}$ to see if an infinitesimal rotation of the vehicle and map is ensured to be a non-observable direction of this linear system, i.e., ensured to be considered non-observable by an extended Kalman filter.

Integration of the dynamics gives a discrete evolution between two observation times:

$$\theta_{n+1} = \theta_n + \bar{\omega}_n,$$
$$X_{n+1} = X_n + R(\theta_n)\bar{v}_n,$$

with $\bar{\omega}_n = \int_{t_n}^{t_{n+1}} \omega_t \in \mathbb{R}$ and $\bar{v}_n \in \mathbb{R}^2$ the solution at time t_{n+1} of the following equation starting at t_n: $\bar{v}_{t_n} = (0,0)^T$, $\frac{d}{dt}\bar{v}_t = R\left(\int_{t_n}^t \omega_t dt\right)\begin{pmatrix} v_t \\ 0 \end{pmatrix}$. We also used the short-hand notations $\theta_n = \theta_{t_n}$ and $X_n = X_{t_n}$. The same notations will be used for the estimates: $\hat{\theta}_n = \hat{\theta}_{t_n}$, $\hat{X}_n = \hat{X}_{t_n}$, $\hat{p}_n = \hat{p}_{t_n}$ and $\hat{\chi}_n = \hat{\chi}_{t_n}$. The propagation and observation functions can be linearized around an estimate $\hat{X}_n$:

$$\begin{pmatrix} \theta_{n+1} - \hat{\theta}_{n+1} \\ X_{n+1} - \hat{X}_{n+1} \\ p - \hat{p}_{n+1} \end{pmatrix} \approx F_n \begin{pmatrix} \theta_n - \hat{\theta}_n \\ X_n - \hat{X}_n \\ p - \hat{p}_n \end{pmatrix},$$

$$h\begin{pmatrix} \theta_n \\ X_n \\ p \end{pmatrix} - h\begin{pmatrix} \hat{\theta}_n \\ \hat{X}_n \\ \hat{p}_n \end{pmatrix} \approx H_n \begin{pmatrix} \theta_n - \hat{\theta}_n \\ X_n - \hat{X}_n \\ p - \hat{p}_n \end{pmatrix},$$

with:

$$F_n = \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ JR\left(\hat{\theta}_n\right)\bar{v}_n & I_2 & 0_{2,2} \\ 0_{2,1} & 0_{2,2} & I_2 \end{pmatrix}, \qquad H_n = \left(Dh_{\hat{\chi}_n}\right) \begin{pmatrix} (\hat{p}_n - \hat{X}_n)^T R(\hat{\theta}_n) J \\ -R(\hat{\theta}_n) \\ R(\hat{\theta}_n) \end{pmatrix}^T, \tag{8.7}$$

where $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $Dh_{\hat{\chi}_n}$ denotes the Jacobian of h computed at the point $R(\hat{\theta}_n)^T \left[\hat{p}_n - \hat{X}_n\right] \in \mathbb{R}^2$.

Now, let us define carefully the linear error system "seen" by the EKF for a sequence of estimates $(\hat{\chi}_n, \hat{\chi}_n^+)_{n \geqslant 0}$. Note that we distinguish intentionally the estimates before the update $(\hat{\chi}_n)$ and after the update $\hat{\chi}_n^+$. Indeed, in the EKF, the dynamics is linearized around the updated estimate $\hat{\chi}_n^+$ although the observation is linearized around the estimate before update $\hat{\chi}_n$ (as the updated value is not known yet). The linear system "seen" by the EKF is defined by the sequence of propagation and observation matrices:

$$F_n = \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ JR(\hat{\theta}_n^+)\bar{v}_n & I_2 & 0_2 \\ 0_{2,1} & 0_2 & I_2 \end{pmatrix},$$

$$H_n = \left(Dh_{\hat{\chi}_n}\right) \begin{pmatrix} (\hat{p}_n - \hat{X}_n)^T R(\hat{\theta}_n) J \\ -R(\hat{\theta}_n) \\ R(\hat{\theta}_n) \end{pmatrix}^T.$$

Thus, following definition 9, a direction is non-observable for the linearized system if we have for any $m \geqslant 0$:

$$H_{m+1}\left(\Pi_{n=0}^{m} F_n\right) \delta X_0 = 0. \tag{8.8}$$

The initial perturbations we are interested in here correspond to overall rotations of the system. They are defined in Proposition 23 below:

Proposition 23. *Let* $\hat{\chi} = \begin{pmatrix} \hat{\theta} \\ \hat{X} \\ \hat{p} \end{pmatrix}$ *be an estimate of the state. The first-order perturbation of the estimate corresponding to an infinitesimal rotation of the system through an angle* $\delta\alpha$ *around the origin reads* $\begin{pmatrix} 1 \\ J\hat{X} \\ J\hat{p} \end{pmatrix} \delta\alpha$, *with* $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Proof. The first-order perturbation is computed considering a rotation through an angle $\delta\alpha$ of the overall system then keeping only the first-order terms. The heading changes as follows:

$$\hat{\theta} \to \hat{\theta} + \delta\alpha.$$

The position of the vehicle changes as follows:

$$\hat{X} \to \begin{pmatrix} \cos(\delta\alpha) & -\sin(\delta\alpha) \\ \sin(\delta\alpha) & \cos(\delta\alpha) \end{pmatrix} \hat{X} \approx \delta\alpha \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \hat{X}.$$

The position of the feature changes as follows:

$$\hat{p} \to \begin{pmatrix} \cos(\delta\alpha) & -\sin(\delta\alpha) \\ \sin(\delta\alpha) & \cos(\delta\alpha) \end{pmatrix} \hat{p} \approx \delta\alpha \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \hat{p}.$$

Stacking these results we obtain the first-order variation of the full state vector:

$$\begin{pmatrix} \hat{\theta} \\ \hat{X} \\ \hat{p} \end{pmatrix} \to \begin{pmatrix} \hat{\theta} \\ \hat{X} \\ \hat{p} \end{pmatrix} + \begin{pmatrix} 1 \\ J\hat{X} \\ J\hat{p} \end{pmatrix} \delta\alpha.$$

□

Denoting by $\delta\chi_0^R = \begin{pmatrix} 1 \\ J\hat{X}_0 \\ J\hat{p}_0 \end{pmatrix} \delta\alpha$ such a perturbation of the initial state we come to the main result of this section:

Proposition 24. *Overall rotations of the system are non-observable. But once the equations have been linearized, the observability of $\delta\chi_0^R$ for the linearized system depends on the linearization points:*

1. *If the linearization point $\hat{\chi}_{t_n}$ is propagated over time but never updated ($\forall n, \hat{\chi}_n^+ = \hat{\chi}_n$) then $\delta\chi_0^R$ is a non-observable direction of the linearized system.*
2. *In general, $\delta\chi_0^R$ is not a non-observable direction of the system linearized on $(\hat{\chi}_n, \hat{\chi}_n^+)_{n\geqslant 0}$.*

Result 1 confirms [57]: linearizing the system on the prior trajectory prevents false observability. But result 2 is quite impractical as an EKF does not "see" anything else than the system linearized on $(\hat{\chi}_n, \hat{\chi}_n^+)_{n\geqslant 0}$. Moreover, it appears in the proof of Proposition 24 we give now that $\delta\chi_0^R$ being non-observable for the system linearized on $(\hat{\chi}_n, \hat{\chi}_n^+)_{n\geqslant 0}$ is highly unlikely, as soon as updates are involved.

Proof. $\delta\chi_0^R$ is a non-observable direction of the linearized system if for any $m \geqslant 0$ equation (8.8) is verified:

$$H_{m+1} \left(\Pi_{n=0}^m F_n\right) \delta\chi_0^R = 0.$$

Let us develop this expression:

$$\begin{aligned} \Pi_{n=0}^m F_n &= \Pi_{n=0}^m \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ JR\left(\hat{\theta}_n^+\right)\bar{v}_n & I_2 & 0_2 \\ 0_{2,1} & 0_2 & I_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J\sum_{n=0}^m R\left(\hat{\theta}_n^+\right)\bar{v}_n & I_2 & 0_2 \\ 0_{2,1} & 0_2 & I_2 \end{pmatrix}, \end{aligned}$$

where we recursively used the easily verified property:

$$\begin{pmatrix} 1 & 0 & 0 \\ a_1 & 1 & 0 \\ b_1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ a_2 & 1 & 0 \\ b_2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ a_1 + a_2 & 1 & 0 \\ b_1 + b_2 & 0 & 1 \end{pmatrix}. \tag{8.9}$$

We have thus:

$$\begin{aligned} \Pi_{n=0}^{m} F_n &= \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J\sum_{n=0}^{m}\left(\hat{X}_{n+1} - \hat{X}_n^+\right) & I_2 & 0_2 \\ 0_{2,1} & 0_2 & I_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J\left(\hat{X}_{m+1} - \hat{X}_0 - \sum_{n=1}^{m}\hat{X}_n^+ - \hat{X}_n\right) & I_2 & 0_2 \\ 0_{2,1} & 0_2 & I_2 \end{pmatrix}, \end{aligned}$$

where the last equality is a mere re-ordering of the terms appearing in the sum(Note that $X_0^+ = X_0$ as the observations start at $n = 1$). Re-injecting in (8.8) we obtain:

$$\begin{aligned} H_{m+1}\left(\Pi_{n=0}^{m} F_n\right)\delta\chi_0^R &= H_{m+1}\begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J\left(\hat{X}_{m+1} - \hat{X}_0 - \sum_{n=1}^{m}\hat{X}_n^+ - \hat{X}_n\right) & I_2 & 0_2 \\ 0_{2,1} & 0_2 & I_2 \end{pmatrix}\begin{pmatrix} 1 \\ J\hat{X}_0 \\ J\hat{p}_0 \end{pmatrix}\delta\alpha \\ &= H_{m+1}\begin{pmatrix} 1 \\ J\left(\hat{X}_{m+1} - \sum_{n=1}^{m}\hat{X}_n^+ - \hat{X}_n\right) \\ J\hat{p}_0 \end{pmatrix}\delta\alpha \\ &= \left(Dh_{\hat{\chi}_{m+1}}\right)\begin{pmatrix} (\hat{p}_{m+1} - \hat{X}_{m+1})^T R(\hat{\theta}_n) J \\ -R(\hat{\theta}_{m+1}) \\ R(\hat{\theta}_{m+1}) \end{pmatrix}^T \begin{pmatrix} 1 \\ J\left(\hat{X}_{m+1} - \sum_{n=1}^{m}\hat{X}_n^+ - \hat{X}_n\right) \\ J\hat{p}_0 \end{pmatrix}\delta\alpha \\ &= -\left(Dh_{\hat{\chi}_{m+1}}\right) J R(\hat{\theta}_{m+1})\left(\hat{p}_{m+1} - \hat{p}_0 + \sum_{n=1}^{m}\hat{X}_n^+ - \hat{X}_n\right)\delta\alpha. \end{aligned}$$

In one particular case, the last parenthesis on the right side of this equation is zeros and the expression is ensured to be null: if the updates of $\hat{X}$ and $\hat{p}$ are always zero. This proves the first result. This is logical: the observability properties of the system are defined on a trajectory without updates. If no assumption is made on the updates (which can take any value depending on the noise), the latter expression can be non-zero and an initial rotation of the system is observable on the linearized system. □

This update issue is the basic mechanism of false observability making EKF acquire false information. Blue plot of Figure 8.2 displays the heading covariance estimated by an EKF for this problem where h is taken as Identity and one landmark is observed at each time step. As can be seen, the filter believes it improves its accuracy over time although it has physically no way to do better than its first guess. Red plot is not important for now.

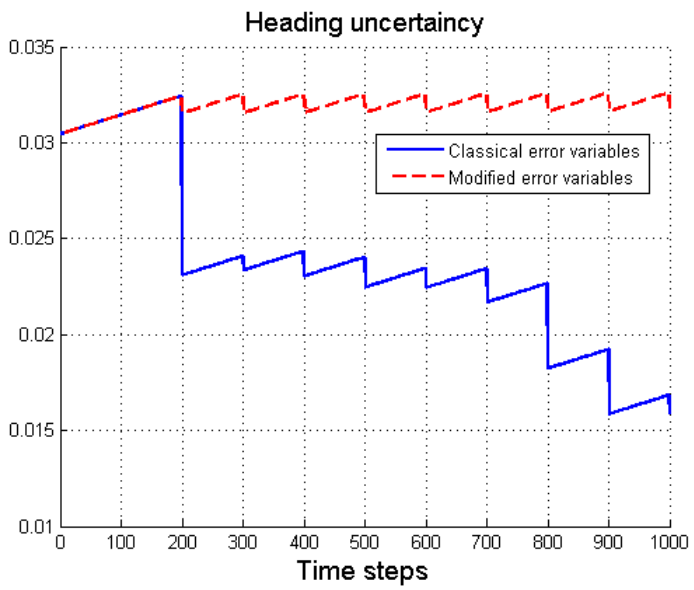


Figure 8.2 – Plot of the estimated heading variance (in rad^2) over time for a vehicle with known position, unknown heading, and one landmark. A sequence of estimates has been computed using the conventional EKF-SLAM algorithm. The blue solid curve corresponds the covariance matrix computed by this EKF. On the same sequence of estimates, the system has been linearized using a non-linear error variable (red dashed line). With a linear error variable, the second measurement at $t_1 = 200$ induces large decrease in the heading uncertainty, and the following measurements every 100 time steps keep decreasing the heading variance although the system has physically no way to acquire this information. On the other hand the heading variance remains at level when using the alternative nonlinear error variable as advocated in this paper. This simulation confirms again that the error variable chosen does change the result of the Riccati equation as soon as non-zero updates are involved.

8.4 Solving false observability through the choice of the error variable

We will show here that the way the error is defined has a dramatic impact on the Riccati equation as soon as the trajectory involves updates (if not, obviously, all the error variables are equivalent). Previous derivation linearized the evolution of an error defined as:

$$e_n = \begin{pmatrix} \theta_n - \hat{\theta}_n \\ X_n - \hat{X}_n \\ p - \hat{p}_n \end{pmatrix}.$$

Now we perform the same computation for an error inspired by the theory of symmetry-preserving observers [15, 19] and introduced for the SLAM problem in [22] (to derive an autonomous error equation but without observability considerations) and which has a

slightly different definition:

$$e_n = \begin{pmatrix} \theta_n - \hat{\theta}_n \\ R\left(\hat{\theta}_n - \theta_n\right) X_n - \hat{X}_n \\ R\left(\hat{\theta}_n - \theta_n\right) p - \hat{p}_n \end{pmatrix}. \tag{8.10}$$

Considering again a sequence of estimates $(\hat{\chi}_n, \hat{\chi}_n^+)$, the firs-order system is defined by the first-order approximations around this trajectory:

$$\begin{pmatrix} \theta_{n+1} - \hat{\theta}_{n+1} \\ R(\hat{\theta}_{n+1} - \theta_{n+1})X_{n+1} - \hat{X}_{n+1} \\ R(\hat{\theta}_{n+1} - \theta_{n+1})p - \hat{p}_{n+1} \end{pmatrix} \approx F_n \begin{pmatrix} \theta_n - \hat{\theta}_n^+ \\ R(\hat{\theta}_n^+ - \theta_n)X_n - \hat{X}_n^+ \\ R(\hat{\theta}_n^+ - \theta_n)p - \hat{p}_n^+ \end{pmatrix},$$

$$h[R(\theta_n)^T(p - X_n)] - h[R(\hat{\theta}_n)^T(\hat{p} - \hat{X}_n)] \approx H_n \begin{pmatrix} \theta_n - \hat{\theta}_n \\ R(\hat{\theta}_n - \theta_n)X_n - \hat{X}_n \\ R(\hat{\theta}_n - \theta_n)p - \hat{p}_n \end{pmatrix}.$$

This leads to the following values of F_n and H_n:

$$F_n = Id,$$

$$H_n = \left(Dh_{\hat{\chi}}\right) \begin{pmatrix} 0_{1,2} \\ -R(\hat{\theta}_n) \\ R(\hat{\theta}_n) \end{pmatrix}^T,$$

where $Dh_{\hat{\chi}}$ is the Jacobian of h computed at $R(\hat{\theta}_n)^T(\hat{p}_n - \hat{X}_n)$. The expression of the linearized system has become much simpler, and showing preservation of the non-observable directions is going to be easy. First we have to write the infinitesimal error corresponding to a small rotation of angle α of the whole system.

Proposition 25. *Let* $\hat{\chi} = \begin{pmatrix} \hat{\theta} \\ \hat{X} \\ \hat{p} \end{pmatrix}$ *be an estimate of the state. The first-order perturbation of the error* (8.10) *around zero corresponding to an infinitesimal rotation of the system through an angle* $\delta\alpha$ *around the origin reads* $\begin{pmatrix} 1 \\ 0_{2,1} \\ 0_{2,1} \end{pmatrix} \delta\alpha$.

Proof. The first-order perturbation is computed considering a rotation through an angle $\delta\alpha$ of the overall system then keeping only the first-order terms. As shown by Proposition 23, the rotated system reads(up to the first order):

$$\begin{pmatrix} \hat{\theta} + \delta\alpha \\ \hat{X} + \delta\alpha J\hat{X} \\ \hat{p} + \delta\alpha J\hat{p} \end{pmatrix}.$$

The error (defined by (8.10)) with the non-perturbed system reads:

$$\begin{pmatrix} \delta\alpha \\ R(-\delta\alpha)(\hat{X} + \delta\alpha J\hat{X}) - \hat{X} \\ R(-\delta\alpha)(\hat{p} + \delta\alpha J\hat{p}) - \hat{p} \end{pmatrix} \approx \begin{pmatrix} \delta\alpha \\ (I_2 - \delta\alpha J)(\hat{X} + \delta\alpha J\hat{X}) - \hat{X} \\ (I_2 - \delta\alpha J)(\hat{p} + \delta\alpha J\hat{p}) - \hat{p} \end{pmatrix}.$$

And removing the second-order terms:

$$\begin{pmatrix} \delta\alpha \\ R(-\delta\alpha)(X+\delta\alpha JX)-X \\ R(-\delta\alpha)(p+\delta\alpha Jp)-p \end{pmatrix} \approx \begin{pmatrix} \delta\alpha \\ 0_{2,1} \\ 0_{2,1} \end{pmatrix}.$$

With this modified error variable, an initial perturbation corresponding to an infinitesimal rotation is thus simply

$$\delta\chi_0^R = \begin{pmatrix} 1 \\ 0_{2,1} \\ 0_{2,1} \end{pmatrix} \delta\alpha.$$

□

Now we can study the observability of such a perturbation for the linearized system using condition (8.8). For any number m of observations we have immediately:

$$H_n \Pi_{n=0}^m F_n \delta\chi_0^R = \left(Dh_{\hat{\chi}_n}\right) \begin{pmatrix} 0 \\ -R(\hat{\theta}_n)^T \\ R(\hat{\theta}_n)^T \end{pmatrix}^T \begin{pmatrix} 1 \\ 0_{2,1} \\ 0_{2,1} \end{pmatrix} = 0_{2,1}.$$

As well, an infinitesimal translation of the whole system has the form $\delta\chi_0^u = \begin{pmatrix} 0 \\ u \\ u \end{pmatrix}$ with $u \in \mathbb{R}^2$ and we have:

$$H_{m+1} \left(\Pi_{n=0}^m F_n\right) \delta\chi_0^u = \left(Dh|_{\hat{\chi}_{m+1}}\right) \begin{pmatrix} 0_{1,2} \\ -R(\hat{\theta}_n) \\ R(\hat{\theta}_n) \end{pmatrix}^T \begin{pmatrix} 0 \\ u \\ u \end{pmatrix} = 0_{2,1}.$$

This proves the following Theorem:

Theorem 15. *The SLAM problem defined by equations* (8.5)-(8.6) *and linearized on a trajectory* $(\chi_n, \chi_n^+)_{n\geqslant 0}$ *using error variable*

$$e_n = \begin{pmatrix} \theta_n - \hat{\theta}_n \\ R(\hat{\theta}_n - \theta_n)X_n - \hat{X}_n \\ R(\hat{\theta}_n - \theta_n)p_n - \hat{p}_n \end{pmatrix},$$

boils down to the following propagation and observation matrices:

$$F_n = I_3 \qquad , \qquad H_n = \left(Dh_{\hat{\chi}}\right) \begin{pmatrix} 0_{1,2} \\ -R(\hat{\theta}_n) \\ R(\hat{\theta}_n) \end{pmatrix}^T .$$

Regardless of the linearization points $(\chi_n, \chi_n^+)_{n\geqslant 0}$*, the non-observability condition*

$$\forall m \geqslant 0, H_{m+1}\left(\Pi_{n=0}^m F_n\right)\delta\chi_0 = 0$$

is verified for the following values of the initial perturbation $\delta\chi_0$*:*

$$\delta\chi_0 = \delta\chi_0^R = \begin{pmatrix} 1 \\ J\hat{X}_{t_n} \\ J\hat{p}_{t_n} \end{pmatrix}, \qquad \delta\chi_0 = \delta\chi_0^1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \delta\chi_0 = \delta\chi_0^2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

They correspond to rotations and translations of the whole system.

Application of Proposition 22 gives then immediately

Theorem 16. *Let $(P_n)_{n\geqslant 0}$ denote the solution of the Riccati equation obtained linearizing the SLAM problem* (8.5)-(8.6) *around any trajectory $(\chi_n, \chi_n^+)_{n\geqslant 0}$ with error variable*

$$e_n = \begin{pmatrix} \theta_n - \hat{\theta}_n \\ R(\hat{\theta}_n - \theta_n)X_n - \hat{X}_n \\ R(\hat{\theta}_n - \theta_n)p_n - \hat{p}_n \end{pmatrix},$$

and choosing arbitrary noise covariance matrices Q_n and R_n.

$(P_n)_{n\geqslant 0}$ verifies:

$$\forall n \geqslant 0, \qquad \begin{aligned} P_n &= F_{n-1}P_{n-1}F_{n-1} + Q_{n-1}, \\ K_n &= P_n H_n^T \left(H_n P_n H_n^T + R_n\right)^{-1}, \\ P_n^+ &= (I - K_n H_n)P_n, \end{aligned} \tag{8.11}$$

with matrices F_n and H_n defined as in Theorem 15 above. Let $\delta\chi_0$ denote a linear combination of the infinitesimal rotations and translations $\delta\chi_0^R, \delta\chi_0^1, \delta\chi_0^2$ of the whole system, defined as in Theorem 15 above. Then the non-observability of $\delta\chi_0$ is preserved in the sense of Definition 10, regardless of the linearization points. We recall the meaning of this property, which takes a very simple form here due to the trivial form of F_n:

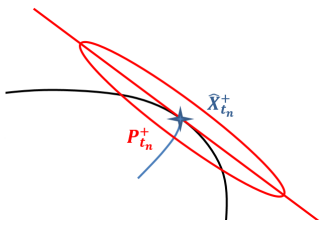
$(\delta\chi_0)^T P_n^{-1} (\delta\chi_0)$ *decreases over time.*

Red plot of Figure 8.2 displays the heading covariance computed on the same sequence of estimates as the blue one, but with the non-linear error variable. As proved by Theorem 15 the false observability issue disappeared. Before going further into the building of an EKF based on a non-linear error variable we give in the next section an alternative way to solve false observability problems, looking more natural than the use of a non-linear error variable but turning out to be strictly equivalent.

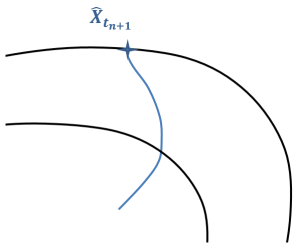
8.5 A different approach to the same result

8.5.1 A visual approach to the problem

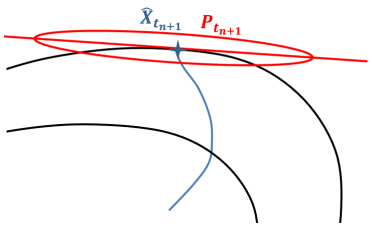
Previous section explained the problem of false observability through computations. Figures 8.3 and 8.4 illustrate the same phenomenon through a drawing for a 2-dimensional non-linear system. It shows the fundamental difference between propagation step and update step in the EKF. During the propagation, the estimate changes (following equation $\frac{d}{dt}\chi_t = f_u(\chi_t)$ or its discrete form) while the covariance matrix changes in a coherent way (involving the Jacobian F_n of the flow of f). This preserves the non-observability of a move along the black line representing a non-linear non-observable transformation, because the Jacobian of the flow maps non-observable directions of an instant t_n to the non-observable directions of an instant t_{n+1}. During the update step, the covariance matrix is updated using a first-order expansion around the current estimate $\chi_{t_{n+1}}$ and the result $P_{t_{n+1}}^+$ is coherent with the non-observable directions *around* $\hat{\chi}_{t_{n+1}}$. But the state is



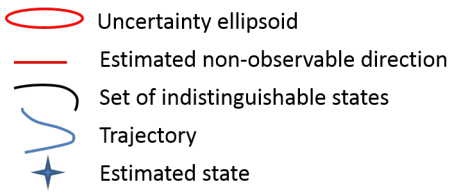
1) Estimate and uncertainty ellipsoid at time t_n if the filter has worked correctly: the ellipsoid is elongated in the non-observable direction.



2) Propagation of the estimate until t_{n+1}.

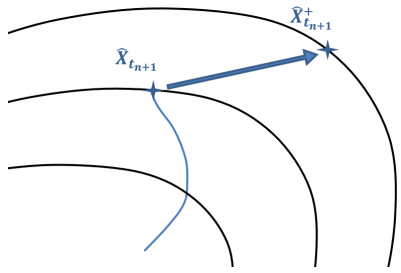


3) First-order propagation of the covariance: the direction estimated as "uncertain" sticks to the true non-observable direction.

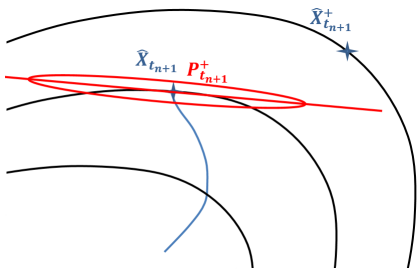


Legend.

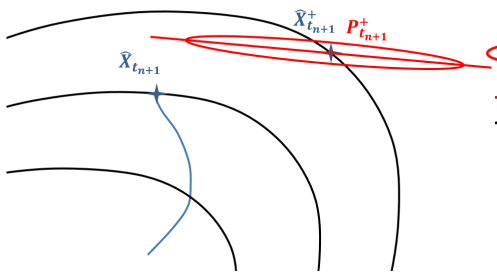
Figure 8.3 – Basic mechanism of false observability: propagation step. The estimate changes, but the Riccati equation propagates the covariance matrix so that the non-observable directions are mapped to non-observable directions. No false observability is created here, unlike in the update step described on Figure 8.4.



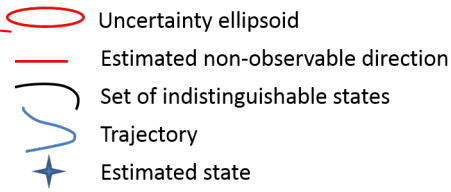
4) The updated state $\hat{X}^+_{t_{n+1}}$ is computed using linearizations at $\hat{X}_{t_{n+1}}$.



5) The updated covariance matrix $P^+_{t_{n+1}}$ is computed using linearizations at $\hat{X}_{t_{n+1}}$: its elongation expresses the non-observability *at* $\hat{X}_{t_{n+1}}$.



6) The non-observable direction at $\hat{X}^+_{t_{n+1}}$ has no reason to correspond to the non-observable direction at $\hat{X}_{t_{n+1}}$.



Legend.

Figure 8.4 – Basic mechanism of false observability: update step. The estimate moves to $\hat{X}^+_{t_{n+1}}$ and the covariance is updated. But all the computations performed to update P_{n+1} rely on a linearization at $\hat{X}_{t_{n+1}}$. Thus, the filter cannot "guess" the non-observable direction around $\hat{X}^+_{t_{n+1}}$. This is the moment where false observability is created.

updated and obviously, as shown by the last drawing, there is no reason for the non-observable directions around $\chi^+_{t_{n+1}}$ to coincide with the non-observable directions around $\chi_{t_{n+1}}$.

8.5.2 Mathematical derivation

The natural reaction at the sight of the problem occurring in the update step is: "the ellipsoid should turn !". This is precisely the simple idea discussed now. The short-hand subscript n instead of t_n will be used again. Mimicking the propagation step, we introduce a matrix $L(\hat{\chi}_n, \hat{\chi}^+_n)$ mapping the non-observable directions around $\hat{\chi}_n$ to the non-observable directions around the updated step $\hat{\chi}^+_n$ to transform P^+_n into a re-updated matrix P^{++}_n. The Riccati equation becomes for any $n \geqslant 0$:

$$\begin{gathered} P_n = F_{n-1}P_{n-1}F_{n-1} + Q_{n-1}, \\ P^+_n = (I - K_nH_n)P_n \qquad \text{with} \qquad K_n = P_nH_n^T\left(H_nP_nH_n^T + R_n\right)^{-1}, \\ P^{++}_n = L(\hat{\chi}_n, \hat{\chi}^+_n)P^+_{t_n}L(\hat{\chi}_n, \hat{\chi}^+_n)^T, \end{gathered} \tag{8.12}$$

where the function $L(\hat{\chi}_n, \hat{\chi}^+_n) : \mathbb{R}^{\dim\chi} \to \mathbb{R}^{\dim\chi}$ has been chosen to map non-observable directions around χ_n to non-observable directions around χ^+_n. Applied to our SLAM problem, this principle will lead to Theorem 17:

Theorem 17. *The SLAM problem* (8.5), (8.6) *linearized using the classical linear error variable gives the propagation and update matrices F_n and H_n defined by* (8.7)*. Assume they are used in Equation 8.12 and let P_n denote the obtained solution, where L is defined as:*

$$L\left(\hat{\chi}_n, \chi^+_n\right) = \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J\left(\hat{X}^+_n - \hat{X}_n\right) & I_2 & 0_{2,2} \\ J\left(\hat{p}^+_n - \hat{p}_n\right) & 0_{2,2} & I_2 \end{pmatrix}.$$

Let $\delta\chi_0$ denote a linear combination of the infinitesimal rotations (defined in Proposition 23) and translations of the whole system. Then the non-observability of $\delta\chi_0$ is preserved in the sense of Definition 10, regardless of the linearization points. We recall the meaning of this property:

$$(\delta\chi_n)^T P_n^{-1}(\delta\chi_n) \text{ decreases over time,}$$

where $\delta\chi_n$ is the propagation of χ_0 defined by:

$$\forall n \geqslant 0, \qquad \delta\chi_{n+1} = F_n\delta\chi_n.$$

The rest of this section shows how L is found and proves Theorem 15. The first-order rotations $\delta\chi_0^R$ (see Proposition 23) and translations χ_0^1, χ_0^1 are defined as follows:

$$\delta X_R = \begin{pmatrix} 1 \\ J\hat{X}_{t_n} \\ J\hat{p}_{t_n} \end{pmatrix}, \qquad \delta X_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \delta X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

The reader can verify that the matrix $L(\hat{\chi}_n, \hat{\chi}_n^+)$ defined below maps $\delta X_0^R, \delta X_0^1, \delta X_0^2$ on the same vectors defined around $\hat{\chi}_n^+$. Note that δX_0^1 and δX_0^2 stay the same but δX_0^R becomes $\delta X_R = \begin{pmatrix} 1 \\ J\hat{X}_n^+ \\ J\hat{p}_n^+ \end{pmatrix}$.

$$L(\hat{\chi}_n, \chi_n^+) = \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_n^+ - \hat{X}_n) & I_2 & 0_{2,2} \\ J(\hat{p}_n^+ - \hat{p}_n) & 0_{2,2} & I_2 \end{pmatrix}.$$

The modified Riccati equation resulting of the introduction of L is easily studied, as L, used after the classical covariance update and before the next propagation step, plays the exact same role as the Jacobian of the propagation. Thus, considering again a sequence of estimates (χ_n, χ_n^+), the linear system "seen" by the EKF is defined by the modified propagation matrix $\tilde{F}_n$ and the observation matrix H_n associated to the linear error (see 8.7):

$$\begin{aligned} \tilde{F}_n &= F_n . L(\hat{\chi}_n, \hat{\chi}_n^+) \\ &= \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ JR(\hat{\theta}_n^+)\bar{v}_n & I_2 & 0_{2,2} \\ 0_{2,1} & 0_{2,2} & I_2 \end{pmatrix} \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_n^+ - \hat{X}_n) & I_2 & 0_{2,2} \\ J(\hat{p}_n^+ - \hat{p}_n) & 0_{2,2} & I_2 \end{pmatrix}, \end{aligned}$$

$$H_n = \left(Dh_{\hat{\chi}_n}\right) \begin{pmatrix} (\hat{p}_n - \hat{X}_n)^T R(\hat{\theta}_n) J \\ -R(\hat{\theta}_n) \\ R(\hat{\theta}_n) \end{pmatrix}^T.$$

To prove preservation of the non-observability of infinitesimal rotations and translations we want to show for any $m \geqslant 0$:

$$H_{m+1} \left(\Pi_{n=1}^m \tilde{F}_n\right) F_0 \delta X_0 = 0,$$

where δX_0 is either an infinitesimal rotation or translation. Although looking maybe complicated at the beginning, the computation is going to simplify as out of magic, to the

contrary of the Riccati equation without correction studied in 8.3. We have:

$$
\begin{aligned}
(\Pi_{n=1}^{m}\tilde{F}_n)F_0 &= \left(\Pi_{n=1}^{m}F_n L\left(\hat{\chi}_n,\hat{\chi}_n^+\right)\right)F_0 \\
&= F_m\Pi_{n=0}^{m-1}L(\hat{\chi}_{n+1},\hat{\chi}_{n+1}^+)F_n \\
&= F_m\Pi_{n=0}^{m-1}\begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_{n+1}^+-\hat{X}_{n+1}) & I_2 & 0_{2,2} \\ J(\hat{p}_{n+1}^+-\hat{p}_{n+1}) & 0_{2,2} & I_2 \end{pmatrix}\begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ JR(\hat{\theta}_n^+)\bar{v}_n & I_2 & 0_{2,2} \\ 0_{2,1} & 0_{2,2} & I_2 \end{pmatrix} \\
&= F_m\Pi_{n=0}^{m-1}\begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_{n+1}^+-\hat{X}_{n+1}+R(\hat{\theta}_n^+)\bar{v}_n) & I_2 & 0_{2,2} \\ J(\hat{p}_{n+1}^+-\hat{p}_{n+1}) & 0_{2,2} & I_2 \end{pmatrix} \\
&= F_m\Pi_{n=0}^{m-1}\begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_{n+1}^+-\hat{X}_n^+) & I_2 & 0_{2,2} \\ J(\hat{p}_{n+1}^+-\hat{p}_n^+) & 0_{2,2} & I_2 \end{pmatrix}. \qquad (8.13)
\end{aligned}
$$

To obtain the last equality we used the propagation equations of $\hat{X}_n$ ($\hat{X}_{n+1}=X_{n+1}^+ + R(\hat{\theta}_n^+)\bar{v}_n$) and $\hat{p}$ ($\hat{p}_{n+1}=p_n^+$). An iterated application of (8.9) on the result (8.13) gives:

$$
\begin{aligned}
(\Pi_{n=1}^{m}\tilde{F}_n)F_0 &= F_m\begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_m^+-\hat{X}_0^+) & I_2 & 0_{2,2} \\ J(\hat{p}_m^+-\hat{p}_0^+) & 0_{2,2} & I_2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_{m+1}-\hat{X}_0^+) & I_2 & 0_{2,2} \\ J(\hat{p}_{m+1}-\hat{p}_0^+) & 0_{2,2} & I_2 \end{pmatrix}.
\end{aligned}
$$

And finally:

$$
\begin{aligned}
H_{m+1}(\Pi_{n=1}^{m}\tilde{F}_n)F_0\delta\chi_0 &= \left(Dh_{\hat{\chi}_{m+1}}\right)\begin{pmatrix} (\hat{p}_{m+1}-\hat{X}_{m+1})^T R(\hat{\theta}_{m+1})J \\ -R(\hat{\theta}_{m+1}) \\ R(\hat{\theta}_{m+1}) \end{pmatrix}^T \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} \\ J(\hat{X}_{m+1}-\hat{X}_0^+) & I_2 & 0_{2,2} \\ J(\hat{p}_{m+1}-\hat{p}_0^+) & 0_{2,2} & I_2 \end{pmatrix}\delta\chi_0 \\
&= \left(Dh_{\hat{\chi}_{m+1}}\right)R(\theta_{m+1})\begin{pmatrix} -\left(\hat{X}_0-\hat{p}_0\right)^T J \\ -I_2 \\ I_2 \end{pmatrix}^T \delta\chi_0.
\end{aligned}
$$

We recall the three values of δX for which we want to obtain $H_{m+1}(\Pi_{n=1}^{m}\tilde{F}_n)F_0\delta\chi_0 = 0$ are:

$$
\delta\chi_0^R = \begin{pmatrix} 1 \\ J\hat{X}_{t_n} \\ J\hat{p}_{t_n} \end{pmatrix}, \qquad \delta\chi_0^1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \delta\chi_0^2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.
$$

Verification is obvious. We just proved Theorem 17.

8.6 Equivalence of the two approaches

We provided two ways to solve the problem of false observability of global translations and rotations in the SLAM problem. The second one is the most intuitive but could look artificial. The first one inspires more confidence as the procedure is exactly the same as in the classical EKF, only the definition of the error changes, and the problem being non-linear there is no reason why the linear error $\chi_t - \hat{\chi}_t$ should prevail, on the contrary we advocate throughout the present thesis that it may be an inappropriate choice. But the physical sense of changing the error variable is not necessarily obvious. The answer is Proposition 26: choosing an error variable means precisely choosing a way to make the confidence ellipsoids "turn".

Proposition 26. *Solutions proposed in Sections 8.4 and 8.5 are equivalent in the following sense:*

Let us denote by $(\tilde{Q}_n, \tilde{P}_n)$ and (Q_n, P_n) the covariance matrices appearing in Equation (8.11) *and* (8.12) *respectively, and by $De_{\hat{\chi}}$ the Jacobian matrix of the non-linear error defined by the first-order expansion:*

$$e_n = \begin{pmatrix} \theta_n - \hat{\theta}_n \\ R(\hat{\theta} - \theta)X_n - \hat{X}_n \\ R(\hat{\theta} - \theta)p_n - \hat{p}_n \end{pmatrix} \approx De_{\hat{X}} \begin{pmatrix} \theta_n - \hat{\theta}_n \\ X_n - \hat{X}_n \\ p_n - \hat{p}_n \end{pmatrix}.$$

If Q_n and $\tilde{Q}_n$ are coherent regarding the definition of the non-linear error e_n, i.e., linked by the relation:

$$\tilde{Q}_n = \left(De_{\hat{\chi}_n}\right) Q_n \left(De_{\hat{\chi}_n}\right)^T,$$

then the solutions of Riccati equations (8.11) *and* (8.12) *are always linked by the relations:*

$$\tilde{P}_n = \left(De_{\hat{\chi}_n}\right) P_n \left(De_{\hat{\chi}_n}\right)^T,$$

$$\tilde{P}_n^+ = \left(De_{\hat{\chi}_n^+}\right) P_n^{++} \left(De_{\hat{\chi}_n^+}\right)^T.$$

Proof. While no update is performed the equivalence is obvious: the same differential equation is simply written in two different bases. Consider the update step. The Jacobian matrices corresponding to the linear and non-linear variable are denoted respectively by H_n and $\tilde{H}_n$. We write the update of P_n and $\tilde{P}_n$ in the information form:

$$\begin{aligned}(P_n^{++})^{-1} &= L(\hat{\chi}_n, \hat{\chi}_n^+)^{-T}(P_n^+)^{-1}L(\hat{\chi}_n, \hat{\chi}_n^+)^{-1} \\ &= L(\hat{\chi}_n, \hat{\chi}_n^+)^{-T}(P_n^{-1} + H_n^T R_n^{-1} H_n)L(\hat{\chi}_n, \hat{chi}_n^+)^{-1},\end{aligned}$$

$$(\tilde{P}_n^+)^{-1} = \tilde{P}_n^{-1} + \tilde{H}_n^T \tilde{R}_n^{-1} \tilde{H}_n.$$

We want to show that if we have $\tilde{P}_n = \left(De_{\hat{\chi}_n}\right) P_n \left(De_{\hat{\chi}_n}\right)^T$ before the update then we want to obtain after the update:

$$\tilde{P}_n^+ = \left(De_{\hat{\chi}_n^+}\right) P_n^{++} \left(De_{\hat{\chi}_n^+}\right). \tag{8.14}$$

This becomes easy if we notice the relations:

$$H_n = \tilde{H}_n De_{\hat{\chi}_{t_n}^+} \qquad \text{and} \qquad L(\hat{\chi}_{t_n}^+, \hat{\chi}_{t_n}) = (De_{\hat{\chi}_{t_n}^+})^{-1} De_{\hat{\chi}_{t_n}}. \tag{8.15}$$

To prove these relations we have to compute De_{χ_n} :

$$\begin{pmatrix} \theta_n - \hat{\theta}_n \\ R(\hat{\theta} - \theta)X_n - \hat{X}_n \\ R(\hat{\theta} - \theta)p_n - \hat{p}_n \end{pmatrix} \approx \begin{pmatrix} \left(\theta_n - \hat{\theta}_n\right) \\ \left(X_n - \hat{X}_n\right) - \left(\theta_n - \hat{\theta}_n\right) J\hat{X}_n \\ \left(p_n - \hat{p}_n\right) - \left(\theta_n - \hat{\theta}_n\right) J\hat{p}_n \end{pmatrix},$$

i.e. : $De_{\chi_n} = \begin{pmatrix} 1 & 0_{1,2} & 0_{1,2} & & \\ -J\hat{X}_n & 0_{1,2} & 0_{1,2} - J\hat{p}_n & 0_{1,2} & 0_{1,2} \end{pmatrix}$.

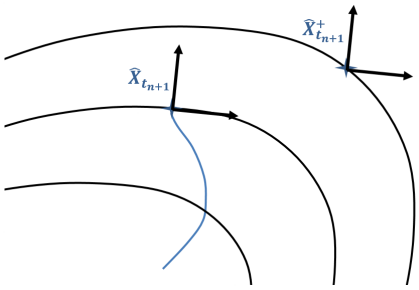
We only have to replace $P_{t_n}^{++}$ by its value in (8.14) and use 8.15 to obtain the result. □

Remark 15. *The same proof works for any system and any error variable: a non-linear error variable can be used virtually, adding only a correction $L(\hat{\chi}_n, \hat{\chi}_n^+)$ at each update.*

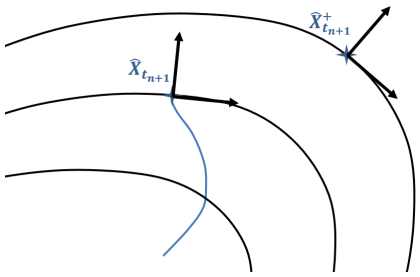
Figure 8.5 explains differently why choosing an error variable is equivalent to choosing a way to make the covariance matrix "turn" after each update.

8.7 Conclusion

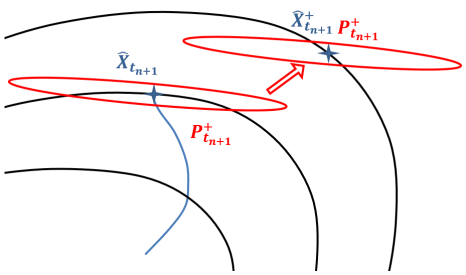
In this chapter we proposed an analysis of the classical problem of EKF SLAM inconsistency, and showed it could be solved by the mere use of a non-linear error variable. As this statement could seem surprising, we explained the physical meaning of using a non-linear error variable: this is a way to make the estimation error covariance matrices "turn" after each update to match the new non-observable subspace. Note that for now, we only discussed the Riccati equation associated to a given sequence of linearization points. Next chapter builds upon the important results obtained to discuss the properties of an EKF relying on a non-linear error variable.



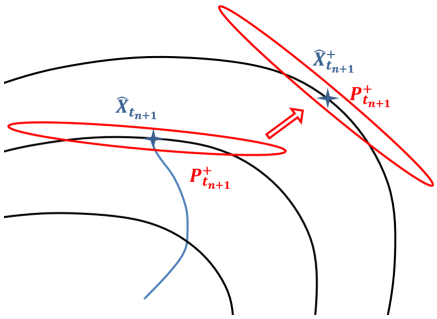
(a) A linear error variable implicitly defines a basis of the state space, this basis is independent from the estimate



(b) A non-linear error variable implicitly defines a basis of the state space, this basis depends on the estimate



(c) With a linear error variable the covariance of the state is the same as the covariance of the error. As $P^+_{t_{n+1}}$ represents the covariance of the error, it also represents the covariance of the state



(d) With a non-linear error variable the covariance of the state is not the covariance of the error. For a given covariance of the error the covariance of the state depends on the estimate. As a consequence, each time the filter updates its estimate, an implicit modification of the state covariance is performed.

Figure 8.5 – Relation between non-linear error variables and correction of the covariance matrix. A first-order perturbation of the state around a given value creates a first-order perturbation of the error around zero. Thus, a correspondence can be established between a basis of the error space and a basis of the state space. For a linear error, the basis of the state space depends only of the basis of the error space. For a non-linear error, it depends also on the point of the state space around which the perturbations are considered. It means that a non-linear error variable implicitly maps a basis around a point of the state space to a *different* basis around another point of the state space. The covariance matrix $P^+_{t_{n+1}}$ is computed in the basis defined by the error. Thus, if the estimate is updated, the ellipsoid representing $P^+_{t_{n+1}}$ turns with the basis. If the error variable is linear, the same behavior can be emulated with the covariance correction presented in Section 8.5.

Chapter 9

EKF based on a non-linear error

Chapter abstract This chapter extends the results obtained for the problem of SLAM to a general non-linear system and discusses the possibility of design of an EKF based on a non-linear error variable. We also propose a non-linear update step inspired by the exponential map of Lie groups and show that deriving some *global* properties for the EKF is possible, assuming the error variable and update procedure have been properly chosen. The novel properties obtained are to our knowledge among the few dealing with the global behavior of an EKF in a general enough non-linear context.

9.1 Introduction

We generalize now the results of Chapter 8. Note the use of non-linear errors to devise EKF has already been explored in various contexts, such as attitude filtering [36, 81], but our goal here is not anymore to find an error variable allowing the design of a better Kalman-like method, but to the discuss the implications of such a choice. Section 9.2 builds an EKF based on a non-linear error variable and shows that its update step is not as straightforward as could be expected. A method inspired by the theory of Lie groups and called "exponential update" is introduced and compared with more natural approaches. In sections 9.3 and 9.4, novel global properties of EKFs based on non-linear error variables are derived. Section 9.5 links the error variable issues with filtering on manifolds and explains the hidden choices underlying the classical EKF. Finally, Section 9.6 proposes the systems studied in Part I as a guideline to find relevant error variables.

9.2 Building an EKF upon a non-linear error variable

9.2.1 Transposition of the classical equations

We transpose the procedure of the EKF, described in Section 8.2.2, to the case where the error variable has been defined by a non-linear function:

$$e_t = \varepsilon(\hat{\chi}_t, \chi_t),$$

with $\varepsilon : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ (n being the dimension of the state x) which verifies $\varepsilon(x,x) = 0$ and which is locally invertible with respect to each of its variables around $\varepsilon = 0$.

Indeed, in a nonlinear setting there is no reason why the linear error $\chi_t - \hat{\chi}_t$ should prevail. We thus define some kind of Jacobian matrices $A^e_\chi(\hat{\chi}_{t_n})$, $A^e_w(\hat{\chi}_{t_n})$, $H^e_\chi(\hat{\chi}_{t_n})$ and $H^e_\chi(\hat{\chi}_{t_n})$ through a first-order approximation of the error evolution and of the measurement error as follows:

$$\frac{d}{dt}e_t = A^e_\chi(\hat{\chi}_{t_n})\, e_t + A^e_w(\hat{\chi}_{t_n})\, w_t + \circ(e_t) + \circ(w_t)\,, \tag{9.1}$$

$$Y_n - h(\hat{\chi}_{t_n}, 0) = H^e_\chi(\hat{\chi}_{t_n})\, e_{t_n} + H^e_V(\hat{\chi}_{t_n})\, V_n + \circ(e_{t_n}) + \circ(V_n)\,. \tag{9.2}$$

Note that we used notations letting appear the dependency in the estimate $\hat{\chi}_t$ of matrices $A^e_\chi, A^e_w, H^e_\chi, H^V_\chi$. As previously, one can use the linear Kalman theory to derive an estimate $\hat{e}^+_{t_n}$ of the error as follows. Assuming an uncertainty matrix $P_{t_{n-1}}$ of e_t at time t_{n-1} is available:

Propagation:

$$\frac{d}{dt}P_t = A^e_\chi(\hat{\chi}_t)P_t + P_t A^e_\chi(\hat{\chi}_t)^T + A^e_w(\hat{\chi}_t)Q_n A^e_w(\hat{\chi}_t)^T, \qquad t_{n-1} < t < t_n, \tag{9.3}$$

and P_t represents an approximation to the covariance of the error variable according to the Kalman filter.

Update:

$$\begin{aligned} S_n &= H^e_\chi(\hat{\chi}_{t_n})P_{t_n}H^e_\chi(\hat{\chi}_{t_n})^T + H^e_V(\hat{\chi}_{t_n})R_n H^e_V(\hat{\chi}_{t_n})^T, \\ K_n &= P_{t_n}H^e_\chi(\hat{\chi}_{t_n})^T S_n^{-1}, \\ \hat{e}^+_{t_n} &= K_n\big[Y_n - h(\hat{\chi}_{t_n}, 0)\big]. \end{aligned} \tag{9.4}$$

This done, deriving a new estimate $\hat{\chi}^+_{t_n}$ consistent with the estimated error $\hat{e}^+_{t_n}$ is not straightforward as in the linear case. It requires a mapping ψ taking the previous estimate $\hat{\chi}_{t_n}$ and the estimated error as argument, and giving an updated estimate:

$$\hat{\chi}^+_{t_n} = \psi\left(\hat{\chi}_{t_n}, \hat{e}^+_{t_n}\right), \tag{9.5}$$

and the error attached to $\hat{\chi}^+_{t_n}$ has covariance $P^+_{t_n}$:

$$P^+_{t_n} = (I - K_n H^e_\chi(\hat{\chi}_{t_n}))P_{t_n}.$$

When the error is defined as the classical vector difference, the issue of the update function ψ does not have to be raised as adding the computed error to the estimated state looks quite natural. In the non-linear case the answer is much less clear as shown by the different possibilities discussed in the next section.

9.2.2 The issue of the update method

All the steps of the classical EKF but one have been straightforwardly transposed to the case of a non-linear error variable. The only problem arises with the computation of the new update. Assume an estimate $e_{t_n}^+$ of the error variable has been obtained after the observation step. An updated estimate $\hat{\chi}_{t_n}^+$ has to be computed out of $e_{t_n}^+$ and $\hat{\chi}_{t_n}$. The most natural way, but which will turn out to be the least relevant (see Proposition 27 and Figure 9.5), is the following:

Definition 11 (Implicit update)**.** *Given an estimate $\hat{\chi}_{t_n}$ and estimated error $e_{t_n}^+$, the implicit update consists in choosing as a new estimate $\hat{\chi}_{t_n}^+ = \psi\left(\hat{\chi}_{t_n}, \hat{e}_{t_n}^+\right)$ the one which best fits the error, i.e. solving for $\hat{\chi}_{t_n}^+$ the equation:*

$$e_{t_n}^+ = \varepsilon(\hat{\chi}_{t_n}^+, \hat{\chi}_{t_n}).$$

This method can look interesting, but has the major drawback of relying on the second-order terms of the error function. A consequence for our SLAM example as introduced in Chapter 8 is that the obtained EKF gives different results if the whole system (car and features) is shifted by a vector $U \in \mathbb{R}^2$. Indeed, assume a correction $e_{t_n}^+ = \begin{pmatrix} e_\theta \\ e_X \end{pmatrix}$ has been computed and is applied to two copies of the system separated by the shift U. We use the notations introduced in 8 and skip the parts of the equations related to p. Without the shift the implicit update is obtained solving:

$$\begin{pmatrix} \hat{\theta}_n^+ - \hat{\theta}_n \\ R(\hat{\theta}_n^+ - \hat{\theta}_n)X_{t_n}^+ - \hat{X}_n \end{pmatrix} = \begin{pmatrix} e_\theta \\ e_X \end{pmatrix}.$$

It gives:

$$\hat{X}_n^+ = R(e_\theta)(\hat{X}_n + e_X).$$

The same computation replacing $\hat{X}_n$ with $\hat{X}_n + U$ gives:

$$\hat{X}_n^+ = R(e_\theta)(\hat{X}_n + e_X) + R(e_\theta)U,$$

where we would like to have:

$$\hat{X}_n^+ = R(e_\theta)(\hat{X}_n + e_X) + U.$$

Note that the difference can be as large as wanted if U is chosen large enough. We showed the following result:

Proposition 27. *An EKF for the SLAM problem described in 8.4 relying on the non-linear variable* (8.10) *gives different results if the general system is shifted.*

This disturbing property makes the implicit update unacceptable in practice. Note that the issue is typically non-linear: it does not occur for linear system although the classical update is a linear form of implicit update. Thus, we introduce here two better choices:

Definition 12 (First-order update)**.** *Given an estimate $\hat{\chi}_{t_n}$ and estimated error $e^+_{t_n}$, the first-order update consists in choosing as a new estimate the one which best fits the first-order expansion of the error, i.e. solving for $\hat{\chi}^+_{t_n}$ the equation:*

$$e^+_{t_n} = De_{\hat{\chi}_{t_n}}(\hat{\chi}^+_{t_n} - \hat{\chi}_{t_n}),$$

with $De_{\hat{\chi}_{t_n}}$ being defined by the first-order expansion:

$$\varepsilon(\hat{\chi}_{t_n} + \delta\chi, \hat{\chi}_{t_n}) \approx De_{\hat{\chi}_{t_n}}\delta\chi.$$

This method makes our EKF SLAM translation-invariant but we can do better, as the sequel will show, using a more sophisticated update that we introduce here and that is inspired by the exponential update we introduced in [11]:

Definition 13 (exponential update)**.** *Given an estimate $\hat{\chi}_{t_n}$ and estimated error $e^+_{t_n}$, let $\tilde{\chi}_s$ denote the solution of the differential equation:*

$$\frac{d}{ds}\tilde{\chi}_s = Dx(\tilde{\chi}_s)e^+_{t_n},$$

with initialization $\tilde{\chi}_0 = \hat{\chi}_{t_n}$ and where Dx is defined by the first-order expansion:

$$\varepsilon(\chi_t, \chi_t + Dx(\chi_t)\delta e) \approx \delta e.$$

It is the first-order expansion of the true state about χ_t w.r.t. a variation of the error about 0.

The exponential update consists then in choosing as a new estimate $\hat{\chi}^+_{t_n} = \tilde{\chi}_1$ (solution of the equation for $s = 1$).

Another way to see the exponential update is the following:

Proposition 28. *Assume the implicit or first-order update has been chosen, but instead of applying directly the formula $\hat{\chi}^+_{t_n} = \psi\left(\hat{\chi}_{t_n}, e^+_{t_n}\right)$, the error is divided by n then the operation $\chi \to \psi\left(\chi, \frac{e^+_{t_n}}{n}\right)$ is iterated n times. For $n \to \infty$ the new estimate obtained is independent from the mapping ψ (implicit or first-order) and is precisely the one resulting from the exponential mapping. This is the reason why we propose this terminology: exponential update is the infinite iteration of an elementary update.*

Note that solving a differential equation at each update is in practice not required as in most situations a close-form solution is available. This is due to the relation between error variables and Lie groups emphasized in the next chapters. For our SLAM problem, the reader can verify that the exponential update, for $e^+_{t_n} = \begin{pmatrix} e_\theta \\ e_X \\ e_p \end{pmatrix}$, is equivalent to:

$$\hat{\theta}^+_{t_n} = \hat{\theta}_{t_n} + e_\theta,$$

$$\hat{X}_{t_n}^+ = \begin{pmatrix} -\frac{\sin(e_\theta)}{e_\theta} & \frac{\cos(e_\theta)-1}{e_\theta} \\ \frac{1-\cos(e_\theta)}{e_\theta} & -\frac{\sin(e_\theta)}{e_\theta} \end{pmatrix} e_X + R(e_\theta)\hat{X}_{t_n},$$

$$\hat{p}_{t_n}^+ = \begin{pmatrix} -\frac{\sin(e_\theta)}{e_\theta} & \frac{\cos(e_\theta)-1}{e_\theta} \\ \frac{1-\cos(e_\theta)}{e_\theta} & -\frac{\sin(e_\theta)}{e_\theta} \end{pmatrix} e_p + R(e_\theta)\hat{p}_{t_n}.$$

Figure 9.5 shows the difference between the three updates.

Section 9.4 emphases an interesting property of the exponential update, confirming figure 9.5 and making it possibly decisive in SLAM applications. But we first summarize the ideas proposed so far.

9.3 Non-linear error and false observability

This section generalizes the results obtained in Chapter 8. The chosen error variable is shown to determine the observability properties of the system linearized on a trajectory involving updates (i.e. jumps). Remark 16 explains again the idea illustrated on figure 8.5: the error variable basically is a local basis, and two bases turning differently when the estimate "jumps" lead to two different linearized systems.

Remark 16. *The nonlinear error $\varepsilon(\hat{\chi},\chi)$ induces a basis of dimension n about any point $\hat{\chi}$ given by the columns of the Jacobian matrix $\left(D\varepsilon(\hat{\chi},.)|_{\hat{\chi}}\right)^{-1}$. The non-linear error is expressed in this coordinate system in the sense that up to second order terms we have*

$$\delta\varepsilon := \varepsilon(\hat{\chi},\hat{\chi}+\delta x) = 0 + D\varepsilon(\hat{\chi},.)|_{\hat{\chi}}\delta x \quad \textit{and thus} \quad \delta x = \left(D\varepsilon(\hat{\chi},.)|_{\hat{\chi}}\right)^{-1}\delta\varepsilon.$$

Note also that, the matrices A_x^e and H_x^e are expressed in this basis, that is, they take as inputs vectors whose components are coordinates in this basis. Similarly, P_t is defined in this basis which depends on the estimate $\hat{\chi}_t$. This is the key feature of all this chapter (and actually of all this document): the error variable is not just a parameterization allowing to do the computations of the EKF, it is the choice of a correspondence between first-order perturbations about two different points χ_1 and χ_2. As well, the update function is a way to translate a perturbation about a given point χ_1 as a new point χ_2. Geometric interpretations of Section 9.5 will try to show that any attempt of an intrinsic definition of the EKF on a smooth manifold boils down to the same questions: - How to map the tangent plan about one point to the tangent plan about another point ? - How to map the tangent plan about one point to the state space.

We are now ready to prove the main result of this section, that characterizes a set of non-observable directions expressed at each point in the basis defined by a given non-linear error (see Remark 16) and then proves that the information decreases indeed along those directions for any EKF based on the corresponding modified nonlinear.

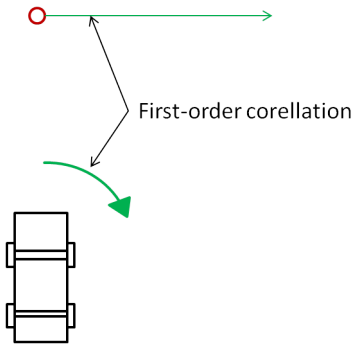


Figure 9.1 – The relative observations correlate the heading of the car with the position of the feature. If the heading is updated due to new information, the position of the feature is updated in a coherent way up to the first order.

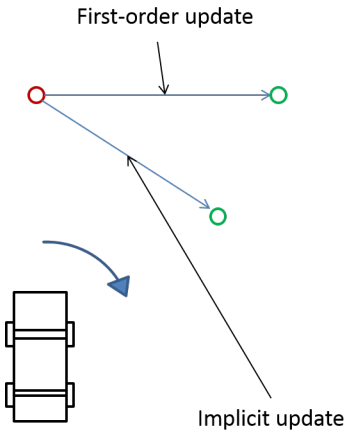


Figure 9.2 – If the update is large, implicit and first-order updates give different results. But none of them moves the feature along a circle as would be expected.

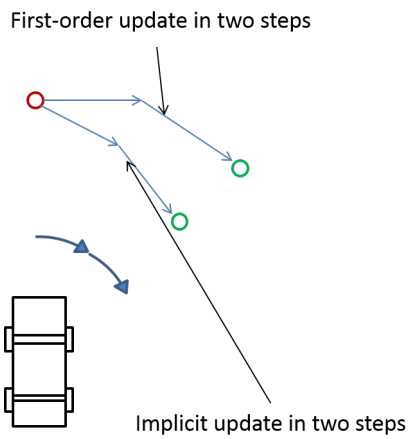


Figure 9.3 – If the update takes into account only an half of the error but is performed twice sequentially, the results of the two update methods come closer to each other, and closer to the circle.

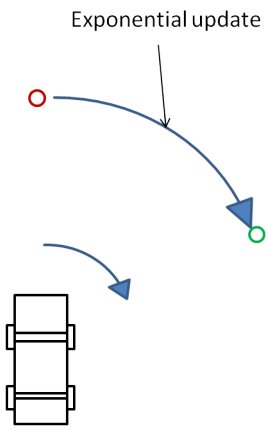


Figure 9.4 – If the error is divided by n instead of 2, the update functions ψ become equivalent when n becomes large. The limit value is equal to the exponential update.

Figure 9.5 – We consider here a car making relative range and bearing observations of a feature. After a few observations, the position of the feature is known in the reference frame of the car. Thus, if the heading is updated then the estimate of the feature should follow a circle around the car. In the case of SLAM, only the exponential update has this behavior

Theorem 18. *Consider a non-linear error e_t and the matrices $A^e_\chi(\chi)$ and $H^e_\chi(\chi)$ defined by* (9.1) *and* (9.2)*. Assume that a linear subspace $V \subset \mathbb{R}^n$ ensures $\forall \chi \in \mathbb{R}^n, A^e_\chi(\chi)V \subset V$ and $\forall \chi \in \mathbb{R}, V \subset \ker(H^e_\chi(\chi))$ then:*

- *V defines a set of directions of non-observability.*
- *Any EKF based on the error e_t, and with state error covariance matrix P_t, is such that the information along those directions always decreases i.e., for any $\xi \in V$ if we let $\xi_0 = \xi$, $\frac{d}{dt}\xi_t = A^e_\chi \xi_t$ the information along this direction $\xi_t^T (P_t)^{-1} \xi_t$ monotonically decreases.*

Proof. The result is trivial: if a vector space V is in $\ker H^e_\chi$ for any χ and stable by F^e_χ for any χ then it is true in particular on the estimated trajectory and V is a non-observable subspace of the linearized system and Proposition 22 can be applied. □

The EKF defined in this section has been applied to the problem of a car navigating using range-and-bearing measurements (Algorithm 6), and compared to the classical EKF (Algorithm 5). The latter is known to fail even in this convenient situation [61], which is confirmed by the results displayed on Figure 9.6 where the uncertainty ellipsoids are over-optimistic. The figure also shows that the EKF with modified error variable does not encounter the same problem as could be expected from the properties previously derived. Figure 9.7 is also interesting, as it precisely illustrates the results of this chapter. The information over the non-observable direction (rotation of the full system) is displayed and the difference between the two filters is striking: this quantity is decreasing only if the proper error variable is chosen.

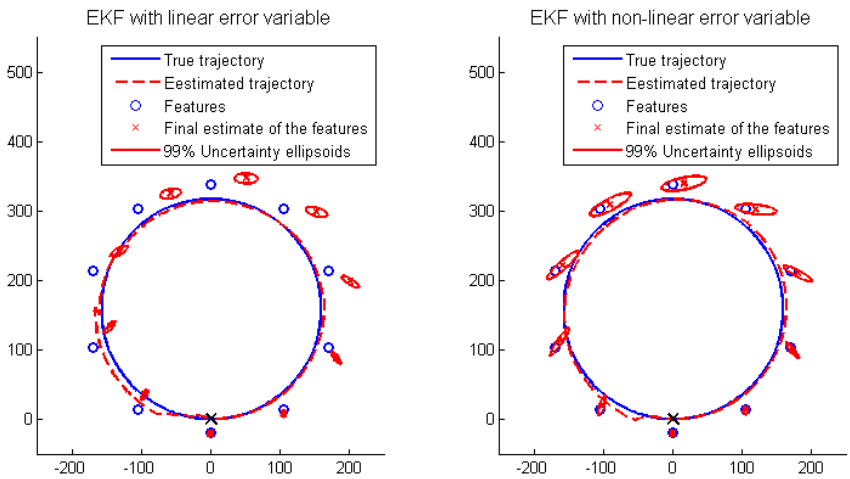


Figure 9.6 – Comparison for the metric SLAM problem (8.5)-(8.6) ($h = Id$) of the estimations of a standard EKF-SLAM described by Algorithm 5 and an alternative EKF-SLAM algorithm based on a different nonlinear estimation error variable and described by Algorithm 6.

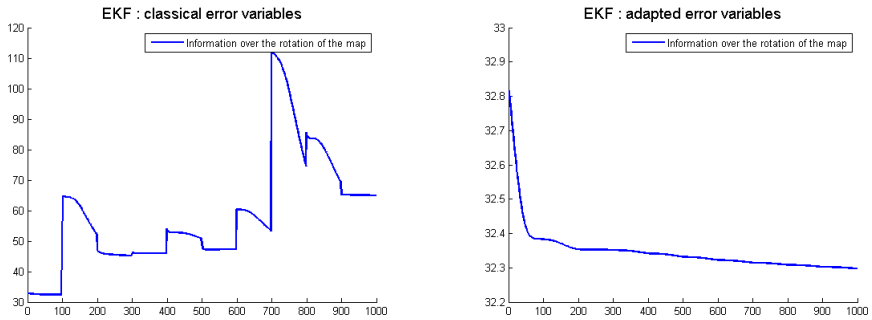


Figure 9.7 – Coefficient of the information matrix corresponding to heading for classical (left plot) and modified (right plot) error equation. Theorem 18 is illustrated: the information decreases over time as expected if the proper non-linear error variable is chosen.

Next section explores another issue: the ability of the filter to include certain information, that is, to handle singular covariance matrices. This is not at all a degenerated situation due to numerical error, as some variables of a system can come to be known with high precision while others are still unknown.

9.4 Nonlinear exponential update and singular information

In a filtering problem, one can be confronted to the issue of an information becoming extremely precise. It happens for instance if a robot observes a large number of times the same set of features with centered noise (with heading and range information). As the features don't move, the knowledge of the local map in the robot frame eventually becomes exact. This kind of information takes the form of a function of the state $\tilde{h}(\chi)$ to be known, that is:

$$\tilde{h}(\chi_t) \equiv c, t \geqslant t_0.$$

If an EKF has worked properly during the observation phase, its absolute confidence in the value of $\tilde{h}(\chi)$ is encoded by a first-order relation:

$$\tilde{H}^e_{\hat{\chi}_{t_n}} P_{t_n} \tilde{H}^{eT}_{\hat{\chi}_{t_n}} = 0, \tag{9.6}$$

where $\tilde{H}^e_{\hat{\chi}_t}$ is defined by:

$$\tilde{h}(\chi) \approx \tilde{h}(\hat{\chi}) + \tilde{H}^e_{\hat{\chi}} e + \circ(e). \tag{9.7}$$

$e = \varepsilon(\chi_t, \hat{\chi}_t)$ being the non linear error associated with the covariance matrix P_t. $\tilde{H}^e_{\hat{\chi}}$ is the classical Jacobian of $\tilde{h}$ if the error is linear. Two properties seem essential to ensure then a sound behavior of the filter:

1. No update can change the relation $\tilde{H}_{\hat{\chi}_{t_n}} P_{t_n} \tilde{H}^T_{\hat{\chi}_{t_n}} = 0$. Indeed, this would mean the new observation made the filter "forget" the value of $\tilde{h}$. Worse, as $rank\ P^+_{t_n} \leqslant rank\ P_{t_n}$, the filter would still consider some directions as certain. But not the right ones.

2. No update can change the value of $\tilde{h}(\hat{\chi}_{t_n})$

Algorithm 5 EKF with linear error variable $(\theta - \hat{\theta}, X - \hat{X}, p - \hat{p})$

R is known, $Q = \begin{pmatrix} q_\omega^2 & 0 & 0 \\ 0 & & \\ 0 & & Q_v \end{pmatrix}$ are known parameters

$\theta_0 = 0, x_0 = (0,0), P_0 = \begin{pmatrix} \sigma_{cap}^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

for $i = 2$ **to** N **do**

$\hat{x}_i = \hat{x}_{i-1} + \Delta t . R(\hat{\theta}_{i-1}) v_i$

$\hat{\theta}_i = \hat{\theta}_{i-1} + \Delta t . \omega_i$

$$F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\Delta t \sin(\hat{\theta}) v_1 - \Delta t \cos(\hat{\theta}) v_2 & 1 & 0 & 0 & 0 \\ \Delta t \cos(\hat{\theta}) v_1 - \Delta t \sin(\hat{\theta}) v_2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$P_i = F . P_{i-1} . F^T + Q$

if Observation Y_i available **then**

$H = \left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} R(\hat{\theta})^T (\hat{p} - \hat{x}), -R(\hat{\theta})^T, R(\hat{\theta})^T \right)$

$S = HPH^T + R$

$K = P_i H^T S^{-1}$

$e^+ = K \left[Y_i - R(\hat{\theta})^T (\hat{p} - \hat{x}) \right]$

$d\theta = e_1^+, \quad dx = (e_2^+; e_3^+)^T, \quad dp = (e_4^+; e_5^+)^T$

$\hat{\theta}_i^+ = \hat{\theta}_i + d\theta$

$\hat{x}_i^+ = \hat{x}_i + dx$

$\hat{p}_i^+ = \hat{p}_i + dp$

$P_i^+ = (I - KH) P_i$

end if

end for

Algorithm 6 EKF with error variable $\left(\theta-\hat{\theta}, X-R(\theta-\hat{\theta})\hat{X}), p-R(\theta-\hat{\theta})\hat{p}\right)$ and exponential update

R is known, q_ω^2 and Q_v are known parameters

$\theta_0 = 0, x_0 = (0,0), P_0 = \begin{pmatrix} \sigma_{cap}^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

for $i = 2$ **to** N **do**

  $\hat{x}_i = \hat{x}_{i-1} + \Delta t . R(\hat{\theta}_{i-1}) v_i$

  $\hat{\theta}_i = \hat{\theta}_{i-1} + \Delta t . \omega_i$

  $F = I_5, \hat{Q} = \begin{pmatrix} q_\omega^2 & 0_{1,2} & 0_{1,2} \\ 0_{2,1} & R(\hat{\theta}) Q_v R(\hat{\theta})^T & 0_{2,2} \\ 0_{2,1} & 0_{2,2} & 0_{2,2} \end{pmatrix}$

  $P_i = F . P_{i-1} . F^T + Q$

  if Observation Y_i available **then**

    $H = \begin{pmatrix} 0_{2,1} & -I_2 & I_2 \end{pmatrix}$

    $S = HPH^T + R$

    $K = P_i H^T S^{-1}$

    $e^+ = K\left[R(\hat{\theta}) Y_i - (\hat{p} - \hat{x})\right]$

    $d\theta = e_1^+, \quad dx = (e_2^+; e_3^+)^T, \quad dp = (e_4^+; e_5^+)^T$

    $\delta x = \frac{1}{d\theta} \begin{pmatrix} \sin(d\theta) dx_1 + (\cos(d\theta) - 1) dx_2 \\ \sin(d\theta) dx_2 - (\cos(d\theta) - 1) dx_1 \end{pmatrix}$

    $\delta p = \frac{1}{d\theta} \begin{pmatrix} \sin(d\theta) dp_1 + (\cos(d\theta) - 1) dp_2 \\ \sin(d\theta) dp_2 - (\cos(d\theta) - 1) dp_1 \end{pmatrix}$

    $\hat{\theta}_i^+ = \hat{\theta}_i^+ + d\theta$

    $\hat{x}_i^+ = R(d\theta) \hat{x}_i + \delta x$

    $\hat{p}_i^+ = R(d\theta) \hat{p}_i + \delta p$

    $P_i = (I - KH) P_i$

  end if

end for

Note that these properties are verified in the linear case: a Kalman filter runs perfectly with singular covariance matrix, the estimate being confined inside a subspace of the state space. They also turn out to be verified in some non-linear situations, but require the use of an exponential update as we are about to prove.

Theorem 19. *Assume we have built an EKF based on a non linear error $\varepsilon(.,.)$ through the equations* (9.3)*,* (9.4)*,* (9.5)*, such that for a given function $\tilde{h}$ and a time t_n the two following relations are verified:*

$$\tilde{h}(\hat{\chi}_{t_n}) = c,$$
$$\forall \chi \in \mathbb{R}^n, \tilde{H}^e(\chi) P_{t_n} \tilde{H}^e(\chi)^T = 0,$$

where $\tilde{H}^e$ is defined by (9.7)*. Then, if any new measurement is processed through an EKF update* (9.4)*,* (9.5) *where ψ is exponential (see Definition* (13)*), the same relations hold after the update:*

$$\tilde{h}\left(\hat{\chi}_{t_n}^+\right) = c,$$
$$\forall \chi \in \mathbb{R}^n, \tilde{H}^e(\chi) P_{t_n}^+ \tilde{H}^e(\chi)^T = 0.$$

Proof. Property 1 is verified as follows. For $\chi \in \mathbb{R}^n$, as P_{t_n} is symmetric, $\tilde{H}^e_\chi P_{t_n} H^e_\chi = 0$ implies $\tilde{H}^e_\chi P_{t_n} = 0$. Thus:

$$\tilde{H}^e_\chi P_{t_n}^+ \tilde{H}^{eT}_\chi = \tilde{H}^e_\chi (I - K_n H_n) P_{t_n} \tilde{H}^{eT}_\chi = 0.$$

Property 2 is obtained using simply definition 13. With the same notations we have:

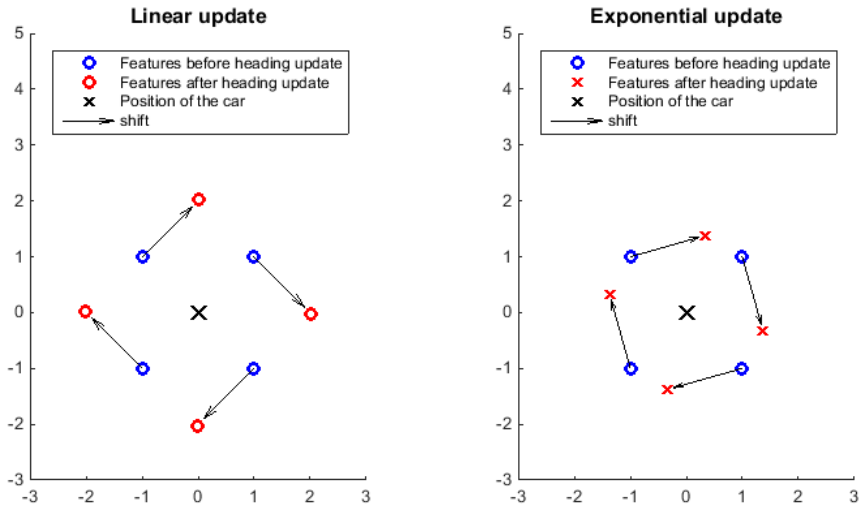
$$\frac{d}{ds}\tilde{h}(\tilde{\chi}_s) = (Dh)_{\tilde{\chi}_s}\left(Dx_{\tilde{\chi}_s}\right) P_{t_n} H^{eT}_{\hat{\chi}_{t_n}} S^{-1} = \tilde{H}_{\tilde{\chi}_s} P_{t_n} H^{eT}_{\hat{\chi}_{t_n}} S^{-1} = 0.$$

Thus, for $s = 1$, $h(\hat{\chi}^+) = h(\tilde{\chi}_0) = h(\hat{\chi})$. □

An immediate implication is preservation of a local map. Assume the software of the car has been able to build a very accurate local map in the reference frame of the car, when a large heading update is applied (due to loop closure for example). A desirable property of the EKF SLAM would be the local map to stay unchanged in the frame of the car, and the non-zero directions of its covariance P_t to be still aligned with the level sets of the local map. To show this is verified for the EKF SLAM with error variable (8.10), all we need is to check according to Theorem 19 that the "Jacobian" of the function $h(\theta, X, p) = R(\theta)^T(p - X)$ w.r.t. the error variable e has the same kernel on any point of the state space. We have:

$$\begin{aligned} R(\theta)^T(p - X) &\approx R(\hat{\theta})^T(\hat{p} - \hat{X}) - \delta\theta R(\hat{\theta})^T J(\hat{p} - \hat{X}) + R(\hat{\theta})^T \delta p - R(\hat{\theta})^T \delta X \\ &= R(\hat{\theta})^T [e_p - e_X], \end{aligned}$$

with (e_θ, e_p, e_X) the components of e (we recall the first-order expansions $e_p = \delta p - J\hat{p}\delta\theta$ and $e_X = \delta X - J\hat{X}\delta\theta$). Thus $H^e_\chi = R(\hat{\theta})^T [0_{2,1}, -I_2, I_2]$ has a kernel independent from χ. Plot 9.8 displays the result of a heading update after acquisition of a local map, for an EKF SLAM based on a linear error variable and on error (8.10).



Heading measurement and Linear update: the estimated features move tangentially to a circle centered on the car. But they follow straight lines and the distances are modified.

Heading measurement exponential heading update: the estimated features follow the circle centered on the car and distances are preserved.

Figure 9.8 – Preservation of local maps by an EKF SLAM algorithm based on error variable (8.10), depending on the update method used. A static car (black cross) maps its environment with no information about its heading. Four features, arranged as a square around the car, have been observed a large number of times. Thus, their positions are known up to the heading of the car to which they are highly correlated. We assume here this correlation is correctly encoded by the covariance matrix P_t. The blue circles indicate the estimated positions of the features. Then, an accurate heading observation is provided to the system, producing a large counter-clockwise variation of the estimated heading. Because of the correlation encoded by P_t between heading and features positions, those move in a way coherent with the heading update, i.e., they "turn" with the car. In other words, the car has accurate information in its own reference frame, which has no reason to be modified by a heading update: the estimated local map should turn as information is obtained regarding the general orientation of the system. On the left plot, the red circles are the new features estimates if a first-order update is used. They turned counter-clockwise as wanted, but only up to the first order. Then they followed straight lines, tangentially to the circle centered on the car. As a consequence, the estimated map underwent a more complicated transformation than the simple rotation we expected. On the right plot, the red circles are the new features estimates if an exponential update is used. The result is very different: the transformation is this time a true rotation. The positions of the estimates in the estimated reference frame of the car are the same as before the update, only the general orientation has changed. This property is the preservation of a local map mentioned at the end of Section 9.4. Note that a classical EKF encounters in this situation the exact same problem as the EKF using non-linear error variable but first-order update.

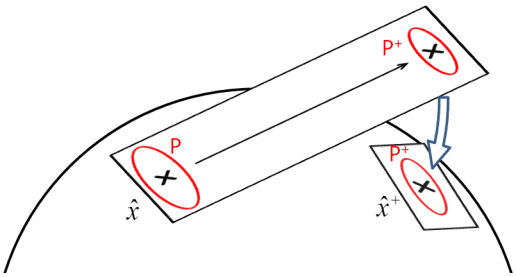


Figure 9.9 – EKF on the sphere: the updated state and covariance matrix are both defined in the tangent plan to the previous estimate. Thus, defining a mapping is necessary to bring them back to the state space.

9.5 A geometric interpretation

The reason why the question of the error variable chosen for SLAM has not been studied before is probably that working in a vector space gives the illusion of a "natural" way to associate perturbations about a point χ_1 with perturbations about another point χ_2 (the vector translation). As a toy, yet illustrative example, we now assume the state to be estimated lives on a sphere (as can be the case when one tries to estimate a direction of motion in the 3D space for instance). Recall the update step of the conventional EKF is defined by:

$$\hat{\chi}_{t_n}^+ = \hat{\chi}_{t_n} + \hat{e}_{t_n}^+ = \hat{\chi}_{t_n} + K_n\left[Y_n - h(\hat{\chi}_{t_n}, 0)\right], \qquad P_{t_n}^+ = (I - K_n H(\hat{\chi}_{t_n}))P_{t_n}$$

Its implementation is usually considered as straightforward, but we see it is actually not. On the sphere, no addition is defined, so the update step which is inherently an addition, is thus naturally performed in the tangent plane to $\hat{\chi}$, which is a vector space, as illustrated on Figure 9.9. The updated state $\hat{\chi}_{t_n}^+$ has a covariance matrix $P_{t_n}^+$ attached to it, and both are now defined in the tangent plan to $\hat{\chi}_{t_n}$. In order to define the updated state $\hat{\chi}_{t_n}^+$ as an element of the sphere (that is, the state space) with a covariance matrix attached to it (that is, in geometric terms an uncertainty ellipsoid defined in the tangent plane, see Fig. 9.9), at least two additional tools are needed: a function ψ mapping the new estimate from the tangent plane at $\hat{\chi}_{t_n}$ to the sphere, then a correspondence between vectors tangent to $\hat{\chi}_{t_n}$ and those tangents to $\hat{\chi}_{t_n}^+$ allowing to define the new uncertainty ellipsoid. The latter is equivalent to having a basis of the tangent plane at $\hat{\chi}_{t_n}^+$ in which $P_{t_n}^+$ is defined. Defining this new basis is precisely the sense to be given to the design of an error variable. The crucial issue is that this choice *does* have an influence on the filter. If the sphere (or a part of it) was parameterized, the problem would become 2-dimensional and the issues raised above would apparently disappear: a classical EKF could be implemented straightforwardly, and a filter devised using Euler angles would be wholly different from a filter based on e.g. stereographic projection. Of course, the problems are only hidden. The choice of a parameterization implicitly defines a choice of mapping ψ and a basis around any point (the one inherited from differentiation with respect to each coordinate). This is the basic influence of the coordinate system on the results of an EKF. But there is no reason a priori to assume a given parameterization

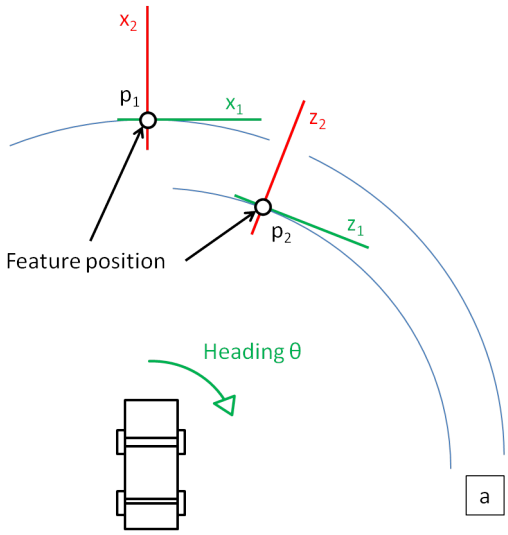


Figure 9.10 – Influence of a measurement on the trusted covariance matrix P, depending whether the system is linearized at $\hat{p} = p_1$ or $\hat{p} = p_2$ (the position of the car is assumed known for simplicity). Assume the position $p_1$of the landmark is measured in the robot's frame. Considering the linearized system variables at $\hat{p} = p_1$, the measurement yields an information about the projection of p onto the vertical axis x_2, and neither the heading θ nor the projection of p onto the horizontal axis x_1 are observed, but both variables become correlated (a future observation of the heading will yield information about the position of the landmark in the x_1 direction). In a similar way, considering the system linearized around $\hat{p} = p_2$ we see a measurement will yield an information gain in the projection of p onto z_2, and will infer a correlation of the projection of p_2 over z_1.

defines a *good* mapping and *good* local bases. For different non-linear systems, the issue of mapping and local bases still exists, and we believe is the major cause of the failure of EKF SLAM. We think the new EKF-like SLAM algorithm introduced previously is based on the best possible such choices. We close this chapter with a visual explanation of false observability in SLAM (Figure 9.10 then 9.11) illustrating again the problem of bad shift of a vector.

9.6 Link with Invariant Kalman filtering on Lie groups

We described until now the properties we can expect from an EKF based on a non-linear error variable. But the most difficult issue is the choice of this error variable. This question will not be fully answered, but Remark 17 below suggests that the situations studied in Part I can give a hint.

Remark 17. *The linearized propagation and observation matrices used in the IEKF are the same as those computed on any trajectory of the system. Thus, the observability properties of the linearized system are precisely the first-order observability properties of*

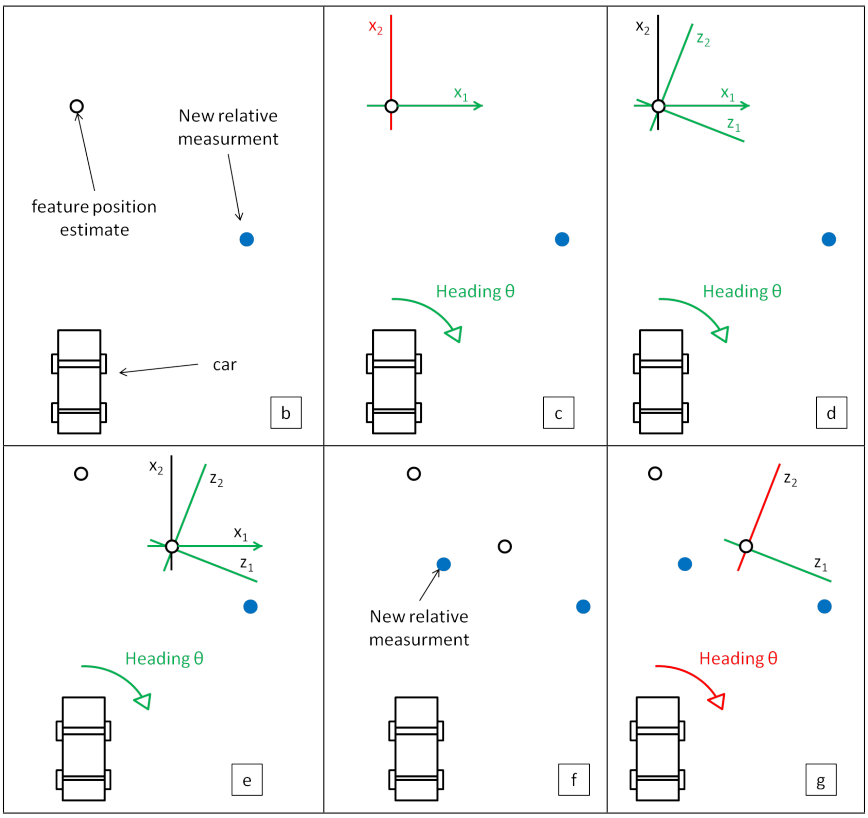


Figure 9.11 – Mechanism of metric SLAM inconsistency illustrated. The position of the car is assumed known for simplicity, but the heading is not. Each time a relative position to landmark measurement is available, the system is linearized on the *current estimate*. Refer to Figure 9.10 for the probabilistic consequences of one update. Here, this operation is repeated and creates false heading uncertainty reduction. We let x_1, x_2, z_1, z_2 denote the projections of p over the different plotted directions.**(b)** A new measurement is provided. **(c)** The system is linearized on the previous estimate. As explained on Fig. 9.10, after the update the Kalman covariance P encodes a reduced variance of x_2 and a correlation between x_1 and the heading. **(d)** As a consequence, the projection of p over the axis z_2 is (weakly) correlated with the heading, as z_2 is not orthogonal to x_2. **(e)** Now, a new estimate is computed through the Kalman state update. **(f)** A new relative measurement is available. **(g)** The system is linearized on the last estimate: as explained on Fig. 9.10, the variable z_2 is directly observed. As P encodes a correlation between z_2 and θ, the information obtained along z_2 reduces the variance of θ after the update. But as only relative measurements are available, no information should be ever gained on the orientation θ with respect to the fixed frame.

the true system. In particular, false observability cannot occur.

But finding the good error variable from scratch is difficult for the reason explained in remark 18:

Remark 18. *For the systems described in Chapter 4, an EKF built upon their multiplicative error variable is not exactly an IEKF. The difference is the innovation term. Considering for example the observation has the form $\chi_{t_n} b$ the innovation term is:*

- $z_n = Y_n - \chi_{t_n} b$ *for the EKF based on the error variable* $\chi_{t_n}^{-1}\hat{\chi}_{t_n}$.
- $z_n = \chi_{t_n}^{-1} Y_n - b$ *for the IEKF.*

Yet, the resulting covariance matrices and estimates are strictly equal: the modification of the innovation term is automatically compensated by a modification of the gain. But the consequence of this difference is that the matrix H depends on the estimate in the first case and the independence of the Riccati equation from the estimate is not visible.

As a consequence of this remark, finding a good error variable without starting from one of the perfect cases described in Chapter 4 is out of reach. Fortunately, most of navigation problems take the generalized form (5.1) and inherit some properties of the system they are based on as shown by Theorem 20 below:

Theorem 20. *If the simplified system:*

$$\frac{d}{dt}\chi_t = f_{u_t}(\chi_t),$$

$$Y_{t_n} = \chi_{t_n} b,$$

has a space V verifying the hypotheses of Theorem 18 for any sequence of inputs (as in SLAM for instance) then these directions are still non-observable in the full system and are preserved by the imperfect IEKF.

Proof. We have to verify that $\begin{pmatrix} V \\ 0_{p,1} \end{pmatrix}$ is stable through the flow and is in $\ker H$. We have:

$$F = \begin{pmatrix} F_\chi & F_{\Theta,\chi} \\ 0 & F_\Theta \end{pmatrix} \begin{pmatrix} \xi \\ 0_{p,1} \end{pmatrix} = \begin{pmatrix} F_\chi \xi \\ 0_{p,1} \end{pmatrix},$$

thus $\begin{pmatrix} V \\ 0_{p,1} \end{pmatrix}$ is stable. We have also:

$$H \begin{pmatrix} \xi \\ 0_{p,1} \end{pmatrix} = Dh[H_\chi, H_\Theta] \begin{pmatrix} \xi \\ 0_{p,1} \end{pmatrix} = DhH_\chi \xi = 0.$$

Thus Theorem 18 applies and V is a non-observable subspace preserved by the filter. □

Example 6. *The SLAM problem* (8.5)-(8.6) *studied in chapter 9 has the form* (5.1) *where the core system is the simplified car studied in section 4.3.1 with observation* (4.21)*. Thus the non-observability of general rotations and translations is preserved by an EKF using right-invariant error variable.*

Example 7. *Consider following the monocular SLAM problem:*

$$\frac{d}{dt}R_t = \alpha R_t \omega_t, \qquad \frac{d}{dt}x_t = \beta R_t(v_t + b), \qquad \frac{d}{dt}p_t = 0,$$
$$Y = angle(R_t^T(x_t - p_t)).$$

The rotations and translations of the whole system are non-observable, as well as a general scaling of x_t, p and β. An EKF based upon the error variable

$$e_t = \left(\hat{R}_t R^{-1}, (\hat{x}_t - \hat{R}_t R_t^{-1} x_t), (\hat{p}_t - \hat{R}_t R_t^{-1} p_t), \frac{\hat{\alpha}}{\alpha}, \frac{\hat{\beta}}{\beta}, (\hat{b} - b)\right),$$

preserves the non-observability of global translations and rotations, but not the non-observability of the scaling factor.

9.7 Conclusion

We detailed in this chapter the difficulties related to the design of an EKF relying on a non-linear error variable, and gave a physical meaning to these methods (in an EKF context at least): choosing an error variable is choosing a local basis around each point of the system. This choice is unimportant as long as the linearization trajectory is continuous. If "jumps" in the estimate occur, due to the updates, different error variables produce different observability properties. A result preventing the classical problem of false observability to occur was derived. The exponential update, inspired by the theory of Lie groups, was introduced and proved to ensure novel global properties, namely preservation of some deterministic side information about the state. The latter will be "experimentally" verified in the next chapter.

Chapter 10

Illustration by a tutorial example

Chapter abstract This Chapter clarifies the articulation between the methods described previously using the problem described in Section 4.3.1 (GPS-aided odometry for the non-holonomic car). Departing from the IEKF, we remove its specificities one by one to obtain 3 algorithms ranging from the IEKF to the conventional EKF. The superiority of the IEKF over the EKF even for this extremely simple low-dimensional and barely non-linear example is puzzling, both in terms of convergence and (especially) in terms of "rationality" of the behavior. Beyond proving the benefits of the IEKF, this tutorial example has been chosen to provide the reader with a physical/geometrical understanding of the limits of the EKF in high-performance navigation.

10.1 Introduction

In this chapter we study a navigation problem involving perfect knowledge of the dynamics (the process noise Q is turned off in the true world and in the filter parameters). This artificial situation is of decisive importance as a limit case of high-accuracy navigation. The IMUs used on aircrafts and ships give almost perfect propagation for several minutes, and even some cheap odometers can also achieve very accurate estimation of the trajectory on hundreds of meters. Regarding Kalman filtering, it means the gains can be large at the beginning if the prior information is inaccurate, but then decrease to zero. It does not prevent convergence of the estimate in the linear case but it is very different if the system is non-linear. In the example presented here the EKF converges for no initialization, except if the error is zero from the beginning. Although it manages to make the error decrease, the convergence is slowed down by non-linearities, and eventually not fast enough to counterbalance the decrease of the gains: the error stabilizes at a non-zero value. This extreme situation does not occur on true systems as the process noise Q is never zero. But the convergence speed can be so deteriorated that the result does not match the requirements of the system anymore. In particular, this is what happens during the transitory phase of an inertial navigation, known as alignment, where the algorithm starts with no information about the state of the system. Complicated procedures mentioned in Chapter 5 are then required to ensure the EKF starts with a good first guess of the state of the system. Note that the solution consisting in adding artificial process noise

(Q) also greatly degrades performance. Besides, we consider only EKF-like methods. Any attempt to explicitly use the fact that the dynamics is exact would be irrelevant here as what we are interested in is the limit behavior of these methods when the process noise Q goes to zero.

This chapter is organized as follows. Section 10.2 introduces the system, Section 10.3 derives the equations of the classical EKF, Section 10.4 those of the IEKF, Section 10.5 discusses the differences between these filters and proposes to test intermediary methods allowing to understand what is the most important difference between EKF and IEKF. One of them is discussed in the same section, the other one in Section 10.6. All the methods are compared in Section 10.7, and finally the results displayed are explained in Section 10.8 through theoretical arguments.

10.2 Perfect dynamics

The system we consider is the one described in Section 4.3.1 where the noise on the dynamics has been put to zero:

$$\begin{aligned} \frac{d}{dt}\theta_t &= u_t v_t, \\ \frac{d}{dt}x_t^1 &= \cos(\theta_t) v_t, \\ \frac{d}{dt}x_t^2 &= \sin(\theta_t) v_t, \end{aligned} \tag{10.1}$$

where θ_t is the heading of the car, x_t^1, x_t^2 are the components of the position vector, u_t is (a function of) the steering angle and v_t is the velocity computed through odometry. The GPS observation takes the form:

$$Y_n = \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} + V_n, \tag{10.2}$$

where V_n is an i.i.d. Gaussian noise with covariance matrix R. We focus thus on the behavior of the EKF in the degenerated case where the initial position of the car is perfectly known but its heading is not. This is modeled by a singular initial covariance matrix (the first variable is heading, the two others are position):

$$P_0 = \begin{pmatrix} \pi/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Of course, after a few time steps the position becomes also uncertain due to the initial heading error. The next sections describe different filters for this dynamics. The results of EKF and IEKF in terms of position error are displayed on Figure 10.1. They show that as soon as the initial heading error is non-zero, the EKF can improve the position error but never bring it to zero as opposed to the IEKF (left plot). We verified this letting the filter run long enough (right plot). Although the observation covariance matrix R is non-zero we fed the filter with perfect observations (whitch is a more favorable situation for the EKF than using noisy data) to obtain clean curves.

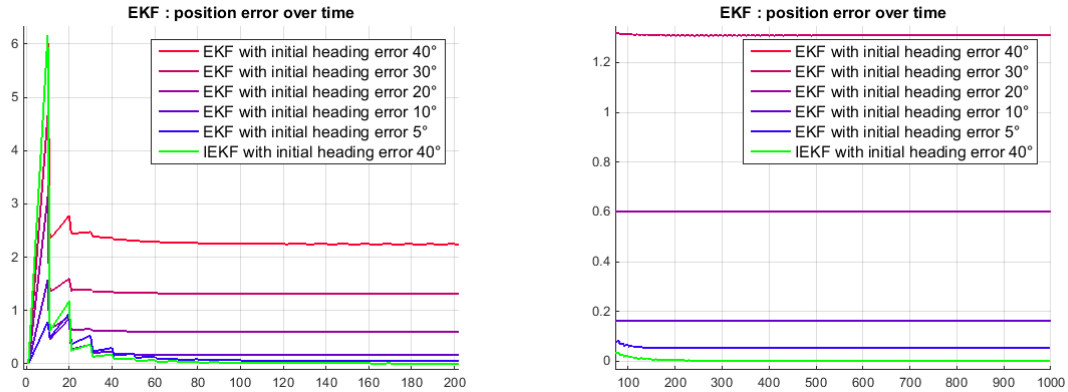


Figure 10.1 – Influence of singular information. Left plot : position error of the EKF for GPS and perfect odometry, with initial position known and initial heading unknown. We see the estimate is never converging, whatever be the initial heading error (except exactly zero). Right plot: after a very long time the results of the EKF have stabilized at a non-zero value while those of the IEKF converge even with maximum initial heading error.

10.3 Classical EKF equations

The EKF with classical error variable $e_t = \begin{pmatrix} \theta_t - \hat{\theta}_t \\ x_t^1 - \hat{x}_t^1 \\ x_t^2 - \hat{x}_t^2 \end{pmatrix}$ for our system is defined by the following propagation and update steps:

Propagation:

$$\begin{aligned} \frac{d}{dt}\hat{\theta}_t &= u_t v_t, \\ \frac{d}{dt}\hat{x}_t^1 &= \cos(\hat{\theta}_t) v_t, \qquad t_{n-1} < t < t_n, \\ \frac{d}{dt}\hat{x}_t^2 &= \sin(\hat{\theta}_t) v_t, \\ \frac{d}{dt}P_t &= A_t P_t + P_t A_t^T, \qquad t_{n-1} < t < t_n, \end{aligned} \tag{10.3}$$

with

$$A_t = \begin{pmatrix} 0 & 0 & 0 \\ -\sin(\hat{\theta}_t) v_t & 0 & 0 \\ \cos(\hat{\theta}_t) v_t & 0 & 0 \end{pmatrix}.$$

Note that Equation (10.3) contains no process noise (usually denoted by Q_t). It has been put to zero here to study the extreme case of a very accurate dynamics.

Update: The update step consists of the following operations:

- Computation of the gains:

$$K_n = P_{t_n} H^T \left(H P_{t_n} H^T + R \right)^{-1},$$

with $H = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

- Computation of the innovation:

$$z = Y - \begin{pmatrix} \hat{x}^1_{t_n} \\ \hat{x}^2_{t_n} \end{pmatrix}.$$

- Computation of the new estimate:

$$\begin{pmatrix} \hat{\theta}_{t_n}{}^{+} \\ (\hat{x}^1_{t_n})^+ \\ (\hat{x}^2_{t_n})^+ \end{pmatrix} = \begin{pmatrix} \hat{\theta}_{t_n} \\ \hat{x}^1_{t_n} \\ \hat{x}^1_{t_n} \end{pmatrix} + K_n z.$$

- Update of the covariance matrix:

$$P^+_{t_n} = (I - K_n H) P_{t_n}.$$

10.4 IEKF equations

The Left-invariant IEKF for our system uses the error variable:

$$\begin{pmatrix} \hat{\theta} - \theta \\ R(\theta)^T \begin{pmatrix} \hat{x}^1_t - x^1_t \\ \hat{x}^2_t - x^2_t \end{pmatrix} \end{pmatrix}.$$

The algorithm is defined by the propagation and update steps derived in Section 4.3.1 and recalled here:

Propagation:

$$\begin{aligned} \frac{d}{dt}\hat{\theta}_t &= u_t v_t, \\ \frac{d}{dt}\hat{x}^1_t &= \cos(\hat{\theta}_t) v_t, \\ \frac{d}{dt}\hat{x}^2_t &= \sin(\hat{\theta}_t) v_t, \qquad t_{n-1} < t < t_n \\ \frac{d}{dt}P_t &= A_t P_t + P_t A_t^T, \qquad t_{n-1} < t < t_n, \end{aligned} \tag{10.4}$$

with

$$A_t = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & u_t v_t \\ -v_t & -u_t v_t & 0 \end{pmatrix}.$$

Again, the equation contains no process noise.

Update: The update step consists of the following operations:

- Computation of the gains:

$$K_n = PH^T \left(HP_tH^T + R\right)^{-1},$$

with $H = -\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

- Computation of the innovation:

$$z = R(\hat{\theta})^T \left(Y_n - \begin{pmatrix} \hat{x}^1_{t_n} \\ \hat{x}^2_{t_n} \end{pmatrix}\right).$$

- Computation of the new estimate:

$$\begin{pmatrix} R(\hat{\theta}^+) & \hat{x}^+ \\ 0_{1\times 2} & 1 \end{pmatrix} = \begin{pmatrix} R(\hat{\theta}) & \hat{x} \\ 0_{1\times 2} & 1 \end{pmatrix} \exp(-K_n z)$$

where $\exp : \mathbb{R}^3 \to \mathbb{R}^3$ is the exponential map of $SE(2)$ defined in Chapter 2.

- Update of the covariance matrix:

$$P^+_{t_n} = (I - K_n H) P_{t_n}.$$

10.5 Differences between the two filters

There are essentially three differences between the EKF and IEKF described here:

- The definition of the innovation is different.
- The computation of the new estimate is different.
- The error variable is different.

An interesting question now is to study the influence of each of these differences on the results. To isolate them, we build in the next section a third filter: an EKF having the same (non-linear) error variable as the IEKF, but using the implicit update instead of the exponential update of the IEKF. But before, we settle the issue of the innovation term by Proposition 29 below.

Proposition 29. *Applying a linear function to the innovation term of an EKF before computing the gains does not change the results of the filter.*

Proof. The innovation term z_n is a function of the error e_{t_n} and the observation noise V_n. It defines the matrices H^e_n and H^V_n through the first-order expansion $z_n = H^e_n e_n + H_V V_n$. If a different innovation term is defined as $z'_n = L_n z_n$ with L_n being any invertible matrix possibly depending on the estimate $\hat{X}_{t_n}$ and having the dimension of z_n, matrices H^e_n and H^V_n become $H^{e\prime}_n = L_n H^e_n$ and $H^{V\prime}_n = L_n H^V_n$. The covariance matrix of z'_n becomes $S'_n = H^{e\prime}_n P_n H^{e\prime T}_n + H^{V\prime}_n R H^{V\prime T}_n = L_n S_n L^T_n$ with S_n the covariance matrix of z_n. The new gains are $K'_n = P_{t_n} H^{e\prime T}_n S'^{-1}_n = P_{t_n} H^{eT}_n S^{-1}_n L^{-1}_n = K_n L^{-1}_n$ and we finally obtain $K'_n z'_n = K_n z_n$. Thus the updated state is going to be the same (regardless of the update method). The updated covariance matrix is defined as $(I - K'_n H^e_n) P_n = \left(I - K_n L^{-1}_n L_n H^e_n\right) P_n$ and this step is also unchanged by the use of z'_n. □

10.6 Non-linear EKF using implicit update

We give here the equations of an "intermediate" filter. It uses the error variable of the IEKF $\begin{pmatrix} \hat{\theta} - \theta \\ R(\theta)^T \begin{pmatrix} \hat{x}_t^1 - x_t^1 \\ \hat{x}_t^2 - x_t^2 \end{pmatrix} \end{pmatrix}$ along with an implicit update. The algorithm is defined by the following propagation and update steps:

Propagation:

$$\begin{aligned} \frac{d}{dt}\hat{\theta}_t &= u_t v_t, \\ \frac{d}{dt}\hat{x}_t^1 &= \cos(\hat{\theta}_t) v_t, \\ \frac{d}{dt}\hat{x}_t^2 &= \sin(\hat{\theta}_t) v_t, \qquad t_{n-1} < t < t_n \\ \frac{d}{dt}P_t &= A_t P_t + P_t A_t^T, \qquad t_{n-1} < t < t_n, \end{aligned} \tag{10.5}$$

with

$$A_t = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & u_t v_t \\ -v_t & -u_t v_t & 0 \end{pmatrix}.$$

Again, the equation contains no process noise.

Update: The update step consists of the following operations:

- Computation of the gains:

$$K_n = PH^T \left(HP_t H^T + R\right)^{-1},$$

with $H = -\begin{pmatrix} 0 & \cos\left(\hat{\theta}_{t_n}\right) & -\sin\left(\hat{\theta}_{t_n}\right) \\ 0 & \sin\left(\hat{\theta}_{t_n}\right) & \cos\left(\hat{\theta}_{t_n}\right) \end{pmatrix}$.

- Computation of the innovation:

$$z = Y_n - \begin{pmatrix} \hat{x}_{t_n}^1 \\ \hat{x}_{t_n}^2 \end{pmatrix}.$$

- Computation of the new estimate (implicit update):

$$\hat{\theta}_{t_n}^+ = \hat{\theta}_{t_n} - \hat{e}_\theta, \qquad \begin{pmatrix} \hat{x}^1 \\ \hat{x}^2 \end{pmatrix}^+ = \begin{pmatrix} \hat{x}^1 \\ \hat{x}^2 \end{pmatrix} - R\left(\hat{\theta}_{t_n}^+\right) \begin{pmatrix} \hat{e}_{x^1} \\ \hat{e}_{x^2} \end{pmatrix},$$

with $\begin{pmatrix} \hat{e}_\theta \\ \hat{e}_{x^1} \\ \hat{e}_{x^2} \end{pmatrix} = K_n z_n$.

- Update of the covariance matrix:

$$P_{t_n}^+ = (I - K_n H) P_{t_n}.$$

10.7 Experimental comparison

The error variable used in the IEKF leads to a Riccati Equation independent from the estimated trajectory. As a consequence, a false estimate does not create false gain matrices. In particular, the observability properties which can be read on the covariance matrix P_t are those of the true system. This is true independently from the chosen update function. A natural question here is whether a good error variable is sufficient to improve performance and the exponential update is a detail, or the exponential update is required also. From a theoretical point of view, propagating correctly singular information requires exponential update (see Section 9.4). This aspect will be developed in Section 10.8. For now we add an experimental argument implementing the three filters to study the improvement. The results of the EKF with the two different error equations are displayed on Figure 10.2. The initial error is here 90 degrees to emphasis the difference. Left plot shows the true trajectory of the car and the estimates of the two filters over time. Right plot shows the heading error over time. Improvement is clear. Then, the EKF with

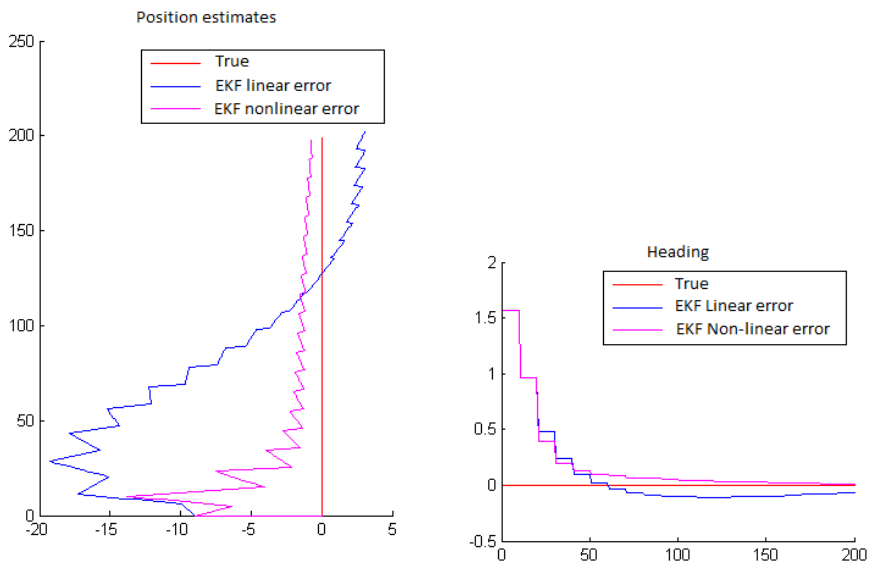


Figure 10.2 – Result of an EKF tracking the car, with linear and non-linear error variables. The true trajectory of the car is a straight line (red), going from the bottom to the top of the left plot. The initial position estimates are the true initial position, but the headings are directed to the left instead of the top (90 degrees initial error). We see the non-linear error variable improves the results, especially regarding the heading estimation, although the exponential update recommended in Chapter 9 is not yet used on this figure.

non-linear error variable is compared with the same algorithm where only the update method has been changed from implicit to exponential update. Figure 10.3 displays the estimates given by the two filters: trajectories on the left plot and heading error on the right plot. A second clear improvement is obtained. The interesting conclusion we can draw from this comparison is that exponential is required for the EKF to fully benefit from a non-linear error variable. Next section explains the physical phenomenon behind the

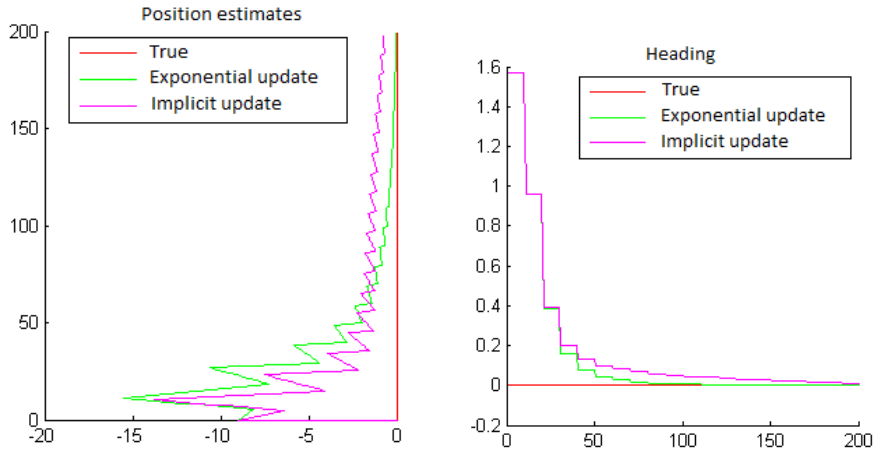


Figure 10.3 – Result of an EKF tracking the car, with non-linear error variable, depending if the exponential update recommended in Chapter 9 is used or not. The true trajectory of the car is a straight line (red), going from the bottom to the top of the left plot. The initial position estimates are the true initial position, but the headings are directed to the left instead of the top (90 degrees initial error). We see the introduction of the exponential update here brings a new dramatic improvement of the performance.

failure of the EKF, here and in most situations where an accurate navigation system has to be initialized.

10.8 Explanation

The specificity of a navigation using accurate dynamics and known initial position but unknown initial heading is that the true dimensionality of the problem is much less than the dimension of the state space. Regarding our car, the state space dimension is 3 (heading and the two components of position). But the only unknown being the initial heading, the set of possible states of the car is only a one-dimensional manifold at any time. For example, if the car drives a straight line from a known point, its position at any time is on a circle centered on the starting point. If the estimate leaves this manifold (the circle) its probability to reach it again is zero. We will see in the sequel (figure 10.5) that the EKF brings the estimate "outside the circle",i.e., too far from the starting point. The EKF with the right non-linear error but implicit update, to the opposite, brings the estimate too close to the starting point. But in both cases, nothing can bring them back to the right distance. What we will show in this section using Theorem 19, is that the IEKF keeps the estimate in the manifold of physically possible states (the circle). It means the search space of the IEKF is one-dimensional and contains the true state. In this situation, having the update steps pushing to the right side is almost sufficient to converge. Figures 10.4 and 10.5 illustrate this result: the distance between the IEKF estimate and the starting point is the same as for the true state, and the estimate has always the starting point exactly in its back, as the true space has. Proposition 30 below gives a mathematical

sense to these considerations. Its proof relies on Theorem 19 of Chapter 9.

Proposition 30. *Let b_t be defined by the differential equation*

$$b_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \frac{d}{dt} b_t = -\begin{pmatrix} 0 & -u_t v_t \\ u_t v_t & 0 \end{pmatrix} b_t - \begin{pmatrix} v_t \\ 0 \end{pmatrix}.$$

The relation

$$R(\theta_t) b_t + \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{10.6}$$

is always verified by the true state of the car, but also by the estimate of the IEKF. Thus, these two variables are restricted in a one-dimensional space which changes at each time step. In some sense, the IEKF understands it has to work in a reduced-dimension search space, what the classical EKF does not do.

The states verifying Equation (10.6) constitute, at a given time step, a one-dimensional manifold containing the true state regardless of the true value of the initial heading (note that this manifold varies over time). Proposition 30 means the estimate of the IEKF is confined inside this manifold. In the specific case of an odometer indicating straight line and constant velocity, the estimate of the IEKF always verifies the property "the starting point is exactly behind the car, at a distence proportional to the time gone by". This links the position and heading, the remaining uncertainty being in what direction the car is going. Figures 10.4 and 10.5 illustrate this result, but also show it is neither verified for EKF using linear error variable nor the proper non-linear error variable if it is not going along with an exponential update. This is extremely important. The relevant search space if the odometer indicates straight line is a circle and an EKF-like method involving singular covariance matrix cannot afford to leave it. Indeed, singularity of P_t makes the states accessible to the filter one-dimensional at each step and nothing can ensure *a priori* that the true state is within this space, or even close to this space. For the IEKF, it is ensured by Proposition 30. We give now its proof.

Proof. The perfect knowledge of the initial position takes the following form:

$$\begin{pmatrix} x_0^1 \\ x_0^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

we have:

$$\frac{d}{dt}\left[R(\theta_t) b_t + \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix}\right] = R(\theta_t)\left[-u_t v_t J b_t - \begin{pmatrix} v_t \\ 0 \end{pmatrix}\right] + \left[u_t v_t R(\theta_t) J\right] b_t + R(\theta_t)\begin{pmatrix} v_t \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. This means that, due to deterministic dynamics, this constraint on the initial state propagates over time to become $R(\theta_t) b_t + \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} = 0$. We denote now this known function by $\tilde{h}_t(\theta_t, x_t^1, x_t^2) = R(\theta_t) b_t + \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix}$ and show the two following properties are verified for any time $t > 0$:

- $\tilde{h}_t(\theta_t, x_t^1, x_t^2) = (0,0)^T$.
- $\tilde{H}_t^e P_t (\tilde{H}_t^e)^T = 0$, where $\tilde{H}_t^e$ is defined by the first-order expansion:

$$\tilde{h}_t(\theta_t, x_t^1, x_t^2) = \tilde{h}_t(\hat{\theta}_t, \hat{x}_t^1, \hat{x}_t^2) + \tilde{H}_t^e e_t + \circ(e_t),$$

  e_t being the error variable of the IEKF.

The proof is in two steps: propagation and update. But first, we need the expression of $\tilde{H}_t^e$. We have:

$$\tilde{h}_t(\theta_t, x_t^1, x_t^2) - \tilde{h}_t(\hat{\theta}_t, \hat{x}_t^1, \hat{x}_t^2) = -JR(\hat{\theta}_t) b_t \left(\hat{\theta}_t - \theta_t\right) - R(\hat{\theta}_t) R(\theta_t)^T \begin{pmatrix} \hat{x}_t^1 - x_t^1 \\ \hat{x}_t^2 - x_t^2 \end{pmatrix} + \circ(e_t)$$

Thus: $H_t^e = -R(\hat{\theta}_t)\,[Jb_t, I_2]$.

Propagation: Preservation of $\tilde{h}_t(\hat{\theta}, \hat{x}^1, \hat{x}^2) = (0,0)^T$ is obtained deriving this function, the computation is the same as the one already done for the true state. Preservation of $\tilde{H}_t^e P_t (\tilde{H}_t^e)^T = 0$ can be verified deriving $\tilde{H}_t^e P_t \left(\tilde{H}_t^e\right)^T$:

$$\begin{aligned}
\frac{d}{dt}\left[\tilde{H}_t^e P_t \left(\tilde{H}_t^e\right)^T\right] &= \left(\frac{d}{dt}\tilde{H}_t^e\right) P_t \left(\tilde{H}_t^e\right)^T + \left(\tilde{H}_t^e\right)\left(\frac{d}{dt}P_t\right)\left(\tilde{H}_t^e\right)^T + \left(\tilde{H}_t^e\right) P_t \left(\frac{d}{dt}\tilde{H}_t^e\right)^T \\
&= \left[\frac{d}{dt}\tilde{H}_t^e + \tilde{H}_t^e A_t\right] P_t \left(\tilde{H}_t^e\right)^T + \left(\tilde{H}_t^e\right) P_t \left[\frac{d}{dt}\tilde{H}_t^e + \tilde{H}_t^e A_t\right]^T.
\end{aligned}$$

we compute each term:

$$\frac{d}{dt}\tilde{H}_t^e = -u_t v_t R(\hat{\theta}_t) J\,[Jb_t, I_2] - R(\hat{\theta}_t)\left[-J^2 u_t v_t b_t - J\begin{pmatrix} v_t \\ 0 \end{pmatrix}, 0_2\right] = R(\hat{\theta}_t) J \left[\begin{pmatrix} v_t \\ 0 \end{pmatrix}, -u_t v_t I_2\right]$$

$$H_t^e A_t = -R(\hat{\theta}_t)\,[Jb_t, I_2] \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -u_t v_t \\ -v_t & u_t v_t & 0 \end{pmatrix} = R(\hat{\theta}_t)\left[J\begin{pmatrix} v_t \\ 0 \end{pmatrix}, -Ju_t v_t\right].$$

We obtain $\frac{d}{dt} H_t^e P_t \left(H_t^e\right)^T = 0$.

Update: This part is shown using Theorem 19. The hypothesis we have to verify is that $\ker H_t^e$ is independent from the point $\hat{\theta}_t, \hat{x}_t^1, \hat{x}_t^2$ where it is computed. But we have $\ker H_t^e = \ker\,[Jb_t, I_2]$ thus the result is obvious.

□

we conclude this study plotting the position estimates and 50% uncertainty ellipsoids at the moment of the first update for all the three filters (Figure 10.6). To avoid having totally flat ellipsoids we added a small uncertainty on the initial position. The result is a mere confirmation of the results of this chapter but gives a good visual interpretation of the benefits of IEKF. On both plots a circle has been drawn, centered on the initial position and having the traveled distance (given by the odometer) as radius. The left plot is obtained just before the first update. All the filters give the exact same estimate

and uncertainty ellipsoid (they are superposed on the plot so only one of them is visible). This confirms what was told throughout Chapter 8: in the absence of update all the error variables are equivalent. The position uncertainty is mainly due to heading, thus the ellipsoids are stretched tangentially to the circle. But from the first update, the filters have very different reactions. The classical EKF (blue) moves its estimate tangentially to the circle and keeps an ellipsoid encoding the uncertainty of the system linearized on the last estimate before the update. The EKF having the right error variable but implicit update has a more complicated evolution, but still leaves the circle. Yet, its uncertainty ellipsoid is consistent with its new estimate as the non-linear error variable made it "turn" properly. The IEKF finally, stays on the circle and has an uncertainty ellipsoid making sense with the physics of the problem. Here actually, it is not exactly on the circle as the uncertainty over initial position authorizes updates in any direction. But it simply confirms that the nice properties derived in this chapter for the limit case of a known initial position give a good idea of what happens on a true system.

10.9 Conclusion

This chapter illustrated on a very simple example a phenomenon at play in high-precision navigation: the set of likely values of the true state reduces, due to the information obtained, to the neighborhood of a non-linear sub-manifold of the state space. The estimate of an EKF not using both a relevant error variable and the exponential update introduced in Chapter 9 can very easily leave this region, and never reach it again. This issue makes the IEKF decisive in some applications like those presented in Chapter 5.3. In order to make sure the choice of the error variable was not the only ingredient making the performance of the IEKF, we implemented this method without the exponential update. The results obtained allow to conclude that the benefits of the exponential update are not simply theoretical.

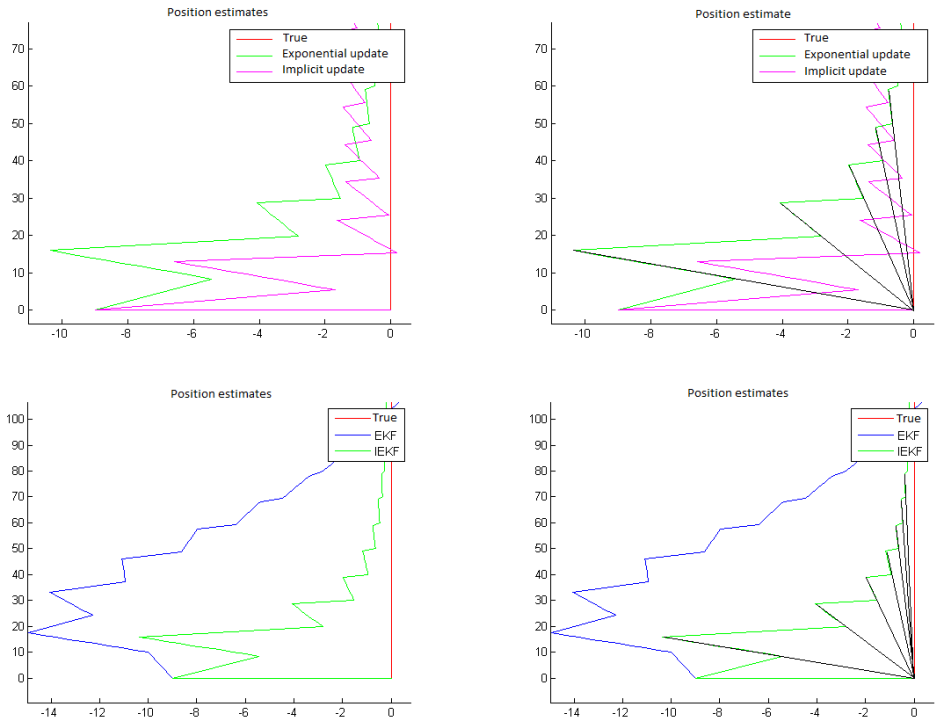


Figure 10.4 – Illustration of Proposition 30. The true trajectory of the car is a straight line (red), going from the bottom to the top of each plot. The estimates over time of several filters are displayed. Their initial position is the true one, but their heading is directed to the left (90 degrees initial error). Between two updates, the position is propagated using the odometer, which indicates straight line. Each time an update is performed, the estimate jumps. The initial position error covariance matrix is zero: the filters know exactly their initial position. Top left and bottom left: position of the different estimates over time. As the odometer indicates straight line and the initial position is known, the estimated heading and (thus) the propagation of the estimate should always be radial with respect to the starting point. The same plots are reproduced on the top and bottom right, but for the estimate of the EKF using non-linear error variable and exponential update, the trajectory of the estimate between consecutive updates has been extended (black lines) to show it always reaches the starting point. This is obviously not verified by the other filters.

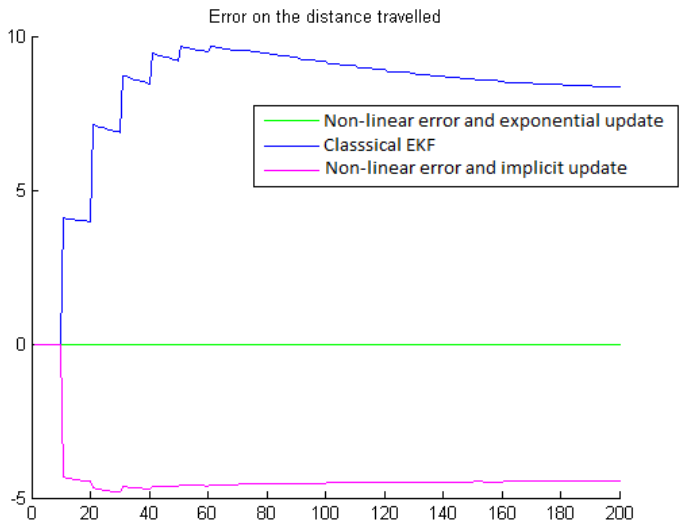


Figure 10.5 – Illustration of Illustration of Proposition 30. Difference between the distance of each estimate to the starting point, and the odometric distance. As the odometer is not noisy it should always be zero. We see that the IEKF ensures this property on the whole estimation process but the other filters do only until the first update.

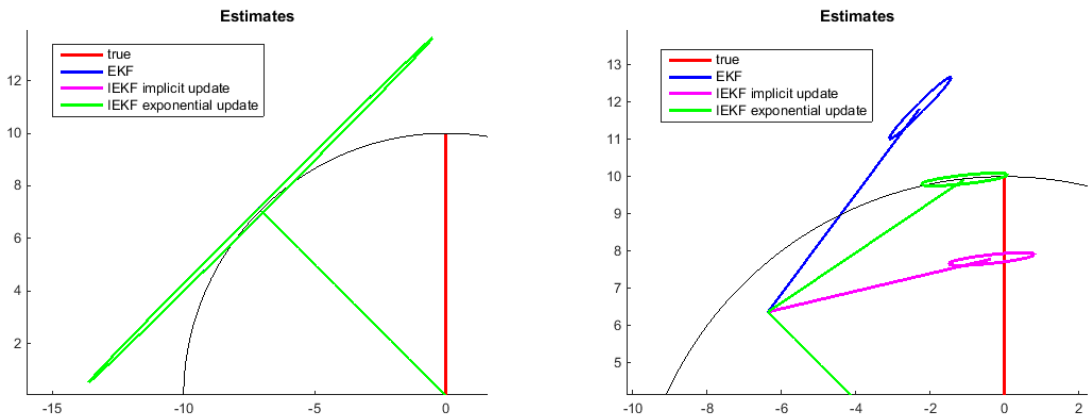


Figure 10.6 – Estimates and uncertainty ellipsoids for the three filters considered in this chapter. The circle is the set of possible states given the measures of the odometer. Up to the small uncertainty over initial position we added for this example, the true state has to be on this circle. Left plot: before the first update all the filters give the same results. Right plot: after the first update, both classical EKF (blue) and EKF using the right non-linear error variable (purple) but not the exponential update have left the circle. The EKF having non-linear error variable keeps an uncertainty ellipsoid consistent with its new position, to the opposite of the classical EKF which put itself in a very bad situation. The EKF with proper non-linear error variable and exponential update (IEKF) sticks naturally to the circle and moves its uncertainty ellipsoid in accordance with the update of the state.

Chapter 11

Conclusion of the thesis

This work investigated the role of the error variable in EKF and proved it to be of decisive importance for several classical problems related to navigation.

A new class of dynamical systems defined on a Lie group G has been characterized, considered as "perfect" in the sense that an error variable having autonomous equation can be built. It contains in particular linear equations, left-invariant dynamics, right-invariant dynamics, combinations of both, and some other equations appearing in inertial navigation. The algebraic logarithm of the error is shown to have linear evolution for these systems, which is also a novel result. This property is leveraged to prove the stability of the IEKF in this case even with changing gains. The hypotheses (first-order uniform observability) are the same as those required for linear systems, which contrasts with most previous results on the convergence of the EKF. The theory is applied to a long sequence of classical navigation problems and brings each time a stability guarantee.

The success of this approach led to a general study of non-linear errors in the EKF. The main question "What does choosing a non-linear error variable physically means ?" has been answered by the interpretation of the error as 1) a local basis and 2) an operation making the covariance matrix of an EKF "turn" after each update. This suggests in particular that there is no reason to approach EKF on linear spaces differently from EKF on manifolds. The classical problem of "false observability" in SLAM has been revisited, linked to the general topic of this work and proved to be solvable by the mere choice of a non-linear error variable. The implications of such a result in practice constitute probably the subject on which we will next devote our best efforts. The questions raised by the idea of an EKF based on a non-linear error variable have been discussed. This led to the introduction of the "exponential update" inspired by the theory of Lie groups. The issue of global behavior of the EKF, i.e., with large estimation errors, has been approached not in terms of global convergence, but in terms of preservation of physical constraints. To the question "My robot has good information of its close environment, is the EKF able to preserve it after an update ?" the answer is "yes" for linear systems and "No" in general. But novel properties have been derived regarding this issue, when non-linear error variable and exponential update are jointly used. Their possibly decisive importance has been illustrated through simulations.

These reflections having focused on the value of the estimate in the context of EKF, the properties of the whole p.d.f. of the estimation error have then been tackled in a sim-

plified framework. This required to discuss stochastic differential equations on Lie groups, whose rigorous meaning is non-trivial. Some sampling-based methods have been proposed, either to reduce a complicated system to a simpler one where the present theory applies, or to improve the fidelity of the covariance matrices used by Kalman-like approaches. These ideas are transpositions of existing methods: the Rao-blackwellized particle filter and Ensemble Kalman filter respectively. The study of fixed-gains filters has been related to the theory of Harris chains, which allowed to prove, under proper conditions that the p.d.f. of the estimation error converges to a limit distribution not depending on the initial density. This novel result allowed designing and optimizing highly non-linear gains for an outlier-resistant observer.

Much time was also devoted to implementation and tests for application to commercial navigation systems. Methods derived from the results summed up above were shown to allow substantial improvement in terms of simplicity of the development phase, final performance and operational flexibility. The results displayed in this document include real-data experiments, comparison with tried and tested softwares, and a campaign of realistic simulations. A program developing a navigation system for military purposes already shifted to this technology.

Further work will focus on SLAM applications as a priority. Due to the large amount of time devoted to inertial navigation, the properties derived for EKF SLAM have been the subject of minimalist simulations. They were only intended to illustrate the main result: non-linear error variables solve false observability issues. It gives a reasonable argument to come back to EKF SLAM, it does not prove the method can compete with state-of-the-art algorithms involving batch optimization or particle sampling. There is still much work before answering this question, especially as our approach can be combined with one of them, for example through a back-and-forth smoothing [6] based on non-linear error variable. Anyway, SLAM is a household where real experiments prevail : no applicative conclusion can be drawn before having achieved extensive simulation and real-data testing, which will be part of our future work.

Appendix A

Proofs of the results of Chapter 4

A.1 Proof of theorem 7

A.1.1 Review of existing linear results

Consider a linear time-varying Kalman filter and let $\Psi^t_{t_0}$ denote the flow of the error variable ξ_s. It is proved in [38] that if the parameters of the Riccati equation verify conditions (1) - (5) then there exist $\gamma_{\max} > 0$ and $\gamma_{\min} > 0$ such that $\gamma_{\max} I \succeq P_t \succeq \gamma_{\min} I$. This pivotal property allows to prove the solution of the *linear* error equation $\Psi^t_s \xi_s$ verifies for $V(P,\xi) = \xi^T P^{-1} \xi$:

$$V\left(P^+_{t_{n+N}}, \Psi^{t_{n+N}}_{t_n}(\xi^+_{t_n})\right) \leq V(P^+_{t_n}, \xi^+_{t_n}) - \beta_3 ||\Psi^{t_{n+N}}_{t_n}(\xi^+_{t_n})||^2 \tag{A.1}$$

where β_3 only depends on $\alpha_1, \alpha_2, \beta_1, \beta_2, \delta_1, \delta_2, \delta_3, N$. Of course, the proof given in [38] holds if the inequalities are only verified on an interval $[0,T]$. We will also use the direct consequence:

$$V\left(P^+_{t_{n+1}}, \Psi^{t_{n+1}}_{t_n}(\xi^+_{t_n})\right) < V\left(P^+_{t_n}, \xi^+_{t_n}\right) \tag{A.2}$$

A.1.2 Preliminary lemmas

The proof of Theorem 7 is displayed in the next subsection. It relies on the final Lemma 8, which is proved step by step in this section through lemmas 5, 6 and 7. The time interval between two successive observations will be denoted Δt. $\tilde{P}(t)$ will denote the Kalman covariance about the true state trajectory.

Lemma 5. [modified constants for closeby trajectories] *If the conditions* (1) *to* (5) *are satisfied about the* true *trajectory, then for any* $k > 1$ *there exists a radius* ε *such that the bound* $\forall s \in [0,t], ||\xi_{t_0+s}|| < \varepsilon$ *ensures the conditions* (1) *to* (5) *are also verified about the* estimated *trajectory, with the modified constants* $\hat{\delta}_1 = \delta_1, \hat{\delta}_2 = \frac{1}{k^2}\delta_2, \hat{\delta}_3 = \frac{1}{k^2}\delta_3, \hat{\alpha}_1 = \frac{1}{k^2}\alpha_1, \hat{\alpha}_2 = k^2\alpha_2, \hat{\beta}_1 = \frac{1}{k^2}\beta_1, \hat{\beta}_2 = k^2\beta_2$. *Moreover, if* $\frac{1}{k}\tilde{P}_{t_0} \leq P_{t_0} \leq k\tilde{P}_{t_0}$ *then* $\frac{1}{k}\tilde{P}_{t_0+s} \leq P_{t_0+s} \leq k\tilde{P}_{t_0+s}$ *holds on* $[0,t]$.

Proof. The Riccati equation computed about the true trajectory reads:

$$\frac{d}{dt}\tilde{P}(t) = A_t\tilde{P}(t) + \tilde{P}(t)A_t + Q_1 + Ad_{\chi^{-1}}Q_2Ad_{\chi^{-1}}^T$$
$$\tilde{P}^+(t) = \tilde{P}(t) - \tilde{P}(t)H^T\left(H\tilde{P}(t)H^T + N_1 + \chi^{-1}N_2\chi^{-T}\right)^{-1}H\tilde{P}(t)$$

This formulation allows to cover both cases of left and right observations. The Riccati equation computed on the estimated trajectory is obtained replacing χ_t with $\hat{\chi}_t$. Recalling the error η_t and the properties of the Ad, the idea of the proof is simply to rewrite the Riccati equation computed about $\hat{\chi}_t$ as a perturbation of the Riccati equation computed about χ_t:

$$\frac{d}{dt}P(t) = A_tP(t) + P(t)A_t + Q_1 + Ad_{\eta^{-1}}\left[Ad_{\chi^{-1}}Q_2Ad_{\chi^{-1}}^T\right]Ad_{\eta^{-1}}^T$$
$$P^+(t) = P(t) - P(t)H^T\left(HP(t)H^T + N_1 + \eta^{-1}\left[\chi^{-1}N_2\chi^{-T}\right]\eta^{-T}\right)^{-1}HP(t)$$

Controlling the perturbation is easy: let $L_\xi : x \to e^{-\xi}x$ and $A_\xi : \mathbb{R}^{\dim\mathfrak{g}} \to \mathbb{R}^{\dim\mathfrak{g}}, x \to Ad_{e^{-\xi}}x$. As these functions are continuous and equal to I_d for $\xi = 0$ there exists a real $\varepsilon > 0$ depending only on k such that $||\xi|| \leq \varepsilon$ ensures $\frac{1}{k}N_2 \preceq L_\xi^T N_2 L_\xi \preceq kN_2$ and $\frac{1}{k}Q_2 \preceq A_\xi^T Q_2 A_\xi \preceq kQ_2$. It ensures consequently $\frac{1}{k}\left(Q_1 + Ad_{\chi^{-1}}Q_2Ad_{\chi^{-1}}^T\right) \preceq Q_1 + Ad_{\hat{\chi}^{-1}}Q_2Ad_{\hat{\chi}^{-1}}^T \preceq k\left(Q_1 + Ad_{\chi^{-1}}Q_2Ad_{\chi^{-1}}^T\right)$ and $\frac{1}{k}\left(N_1 + \chi^{-1}N_2\chi^{-T}\right) \preceq N_1 + \hat{\chi}^{-1}N_2\hat{\chi}^{-T} \preceq k\left(N_1 + \chi^{-1}N_2\chi^{-T}\right)$, and a mere look at the definitions of the constants of Theorem 6 yields the modified constants.

The inequality $\frac{1}{k}\tilde{P}_{t_0+s} \leq P_{t_0+s} \leq k\tilde{P}_{t_0+s}$ follows from the matrix inequalities above on the covariance matrices, by writing the Riccati equation verified by kP_t and $\frac{1}{k}P_t$ and using simple matrix inequalities.

□

Lemma 6. [first-order control of growth] *Under the same conditions as in Lemma 5 (including $\frac{1}{k}\tilde{P}_{t_0} \leq P_{t_0} \leq k\tilde{P}_{t_0}$) and $||\xi_{t_0+s}||$ bounded by the same ε for $s \in [0, 2M\Delta T]$ (i.e. over $2M$ time steps, where M is defined as in Theorem 6), there exists a continuous function l_1 depending* only *on k ensuring $||\xi_{t_0+s}|| \leq l_1(||\xi_{t_0}||)$ for any $s \in [0, 2M\Delta T]$ and $l_1(x) = O(x)$.*

Proof. Using Lemma 5 and then Theorem 6 we know there exist two constants $\gamma_{\min} > 0$ and $\gamma_{\max} > 0$ such that $\gamma_{\max}I \succeq P_t \succeq \gamma_{\min}I$. The non-linear rest $r_{t_n}(\xi)$ introduced in (4.19) is defined by $\exp(\xi)\exp(L_n(e^\xi b - b - H\xi)) = \exp((I - L_nH)\xi + r_{t_n}(\xi))$. The Baker-Campbell-Hausdorff (BCH) formula gives $r_{t_n}(\xi) = O(||\xi||.||L_n\xi||)$ but L_n is uniformly bounded over time by $\frac{\gamma_{\max}||H||}{\delta_3}$ as an operator. Thus $||r_n||$ is *uniformly* dominated over time by a second order: there exists a continuous function $\tilde{l}^k$ (depending only on k and on the *true* trajectory) such that $\tilde{l}^k(x) = O(x^2)$ and $||r_{t_n}(\xi)|| \leq \tilde{l}^k(||\xi||)$ for any n such that $t_n \leq 2M\Delta T$.

Now we can control the evolution of the error using $\tilde{l}^k$. The propagation step is linear, thus we have the classical result $\frac{d}{dt}V_t(\xi_t) < 0$. It ensures $||\xi_{t+s}|| < \sqrt{\frac{\gamma_{\max}}{\gamma_{\min}}}||\xi_t||$ as long as there is no update on $[t, t+s]$. At each update step we have $V_{t_n}^+(\xi_{t_n}^+)^{1/2} = V_{t_n}^+\Big([I - L_nH]\xi_{t_n} +$

$r_n(\xi_{t_n})\Big)^{1/2} \leq V_{t_n}^+\Big([I-L_nH]\xi_{t_n}\Big)^{1/2} + V_{t_n}^+\Big(r_n(\xi_{t_n})\Big)^{1/2} \leq V_{t_n}\Big(\xi_{t_n}\Big)^{1/2} + V_{t_n}^+\Big(r_n(\xi_{t_n})\Big)^{1/2}$ using the triangular inequality. Thus: $||\xi_{t_n}^+|| \leq \sqrt{\frac{\gamma_{\max}}{\gamma_{\min}}}\Big(||\xi_{t_n}|| + ||r_n(\xi_{t_n})||\Big) \leq \sqrt{\frac{\gamma_{\max}}{\gamma_{\min}}}\Big(||\xi_{t_n}|| + \tilde{l}^k(||\xi_{t_n}||)\Big)$. Reiterating over successive propagations and updates over $[0, 2M\Delta T]$, we see $||\xi_{t+s}||$ is uniformly bounded by a function $l_1(||\xi_t||)$ that is first order in $||\xi_t||$.

□

Lemma 7. [second-order control of the Lyapunov function] *Under the same conditions as in Lemma 5 (including $\frac{1}{k}\tilde{P}_{t_0} \leq P_{t_0} \leq k\tilde{P}_{t_0}$) and for $||\xi_{t_0+s}||$ bounded by the same ε for $s \in [0, 2M\Delta T]$ (2M time steps), there exists a continuous function l_2 depending* only *on k ensuring $V_{t_0+s}\big(\xi_{t_0+s}\big) \leq V_{t_0+s}(\Psi_{t_0}^{t_0+s}\xi_{t_0}) + l_2(||\xi_{t_0}||) \leq V_{t_0}(\xi_{t_0}) + l_2(||\xi_{t_0}||)$ for any $s \in [0, 2M\Delta t]$ with $l_2(x) = O(x^2)$. We also have $V_{t_n}\big(\xi_{t_n}^+\big) \leq V_{t_n}^+((\Psi_{t_0}^{t_n})^+\xi_{t_0}) + l_2(||\xi_{t_0}||) \leq V_{t_0}(\xi_{t_0}) + l_2(||\xi_{t_0}||)$ for $t_0 \leq t_n \leq t_0 + 2M\Delta T$.*

Proof. The result stems from the decomposition (4.19) as:

$$
\begin{aligned}
V_{t_0+s}(\xi_{t_0+s})^{1/2} &= V_{t_0+s}\Big(\Psi_{t_0}^{t_0+s}\xi_{t_0} + \sum_{t_0<t_n<t_0+s} \Psi_{t_n}^{t_0+s} r_n(\xi_{t_n})\Big)^{1/2} \\
&\leq V_{t_0+s}\Big(\Psi_{t_0}^{t_0+s}\xi_{t_0}\Big)^{1/2} + \sum_{t_0<t_n<t_0+s} V_{t_0+s}\Big(\Psi_{t_n}^{t_0+s} r_n(\xi_{t_n})\Big)^{1/2} \quad \text{(triangular ineq.)} \\
&\leq V_{t_0+s}\Big(\Psi_{t_0}^{t_0+s}\xi_{t_0}\Big)^{1/2} + \sum_{t_0<t_n<t_0+s} V_{t_n}\Big(r_n(\xi_{t_n})\Big)^{1/2} \quad \text{(using (A.2))} \\
&\leq V_{t_0+s}\Big(\Psi_{t_0}^{t_0+s}\xi_{t_0}\Big)^{1/2} + \sum_{t_0<t_n<t_0+s} \sqrt{\frac{\gamma_{\max}}{\gamma_{\min}}}||r_n(\xi_{t_n})||^{1/2} \quad \text{(from the def. of } V) \\
&\leq V_{t_0+s}\Big(\Psi_{t_0}^{t_0+s}\xi_{t_0}\Big)^{1/2} + \sum_{t_0<t_n<t_0+s} \sqrt{\frac{\gamma_{\max}}{\gamma_{\min}}}(\tilde{l}^k \circ l_1^k)(||\xi_{t_0}||)^{1/2} \quad \text{(from Lemma 6)}
\end{aligned}
$$

As we have $(\tilde{l}^k \circ l_1^k)(x) = O(x^2)$, we obtain the result squaring the inequality and using $V_{t_0+s}\Big(\Psi_{t_0}^{t_0+s}\xi_{t_0+s}\Big) \leq V_{t_0}\Big(\xi_{t_0}\Big) \leq \frac{\gamma_{\max}}{\gamma_{\min}}||\xi_{t_0}||$ to control the crossed terms. □

Lemma 8. [final second order growth control] *Under the same conditions as in Lemma 5 (including $\frac{1}{k}\tilde{P}_{t_0} \leq P_{t_0} \leq k\tilde{P}_{t_0}$) and for $||\xi_{t_0+s}||$ bounded by the same ε for $s \in [0,t]$, there exist two functions $l_1^k(\xi) = O(||\xi||^2)$ and $l_2^k = o(||\xi||^2)$ and a constant β^k ensuring the relation:*

$$
V_{t_0+s}(\xi_{t_0+s}) \leq V_{t_0}(\xi_{t_0}) + l_1^k(\xi_{t_0}) - \sum_{i=0}^{J-1}\Big[\beta^k||\xi_{t_{n_0}+iM\Delta t}||^2 - l_2^k(\xi_{t_{n_0}+iM\Delta t})\Big] + l_1^k(||\xi_{t_{n_{\max}}}^+||) \qquad \text{(A.3)}
$$

where $n_{\max}$ is the last update before t_0+s (i.e. $n_{\max} = \max\{n, t_n \leq t_0+s\}$), J is the maximum number of successive sequences of M updates in $[t_0+M\Delta t, t_{n_{\max}}]$ (i.e. $J = \max\{j, t_{n_{\max}-jM} \geq t_0\} - 1$) and $n_0 = n_{\max} - JM$. If $t_0+s = t_{n_{\max}}$ the last term can be removed.

Proof. For l_1^k we choose the same function as in Lemma 7. There is nothing more to prove for $s < 2M\Delta t$. Let $s \geq 2M\Delta t$. We have $V_{t_0+s}(\xi_{t_0+s}) - V_{t_0}^+(\xi_{t_0}^+) = \Big(V_{t_0+s}(\xi_{t_0+s}) -$

$V^+_{t_{n_{\max}}}(\xi^+_{t_{n_{\max}}})\Big) + \Big(V^+_{t_{n_{\max}}}(\xi^+_{t_{n_{\max}}}) - V^+_{t_{n_0}}(\xi^+_{t_{n_0}})\Big) + \Big(V^+_{t_{n_0}}(\xi^+_{t_{n_0}}) - V_{t_0}(\xi_{t_0})\Big)$. The first and third terms are upper bounded using Lemma 7. The second term is controlled as follows:

$$\begin{aligned} V^+_{t_{n_{\max}}}(\xi^+_{t_{n_{\max}}}) - V^+_{t_{n_0}}(\xi^+_{t_{n_0}}) &= \sum_{i=0}^{J-1}\left[V^+_{t_{n_0+(i+1)M}}(\xi^+_{t_{n_0+(i+1)M}}) - V^+_{t_{n_0+iM}}(\xi^+_{t_{n_0+iM}})\right] \\ &\leq \sum_{i=0}^{J-1}\left[V^+_{t_{n_0+(i+1)M}}(\Psi^{t_{n_0+(i+1)M}+}_{t_{n_0+iM}}\xi^+_{t_{n_0+iM}}) - V^+_{t_{n_0+iM}}(\xi^+_{t_{n_0+iM}}) + l_2(\xi^+_{t_{n_0+iM}})\right] \end{aligned}$$

And we conclude using (see [38]):

$$\begin{aligned} V^+_{t_{n_0+(i+1)M}}(\Psi^{t_{n_0+(i+1)M}+}_{t_{n_0+iM}}\xi^+_{t_{n_0+iM}}) - V^+_{t_{n_0+iM}}(\xi^+_{t_{n_0+iM}}) &\leq -\tilde{\beta}^k||\Psi^{t_{n_0+(i+1)M}+}_{t_{n_0+iM}}\xi^+_{t_{n_0+iM}}||^2 \\ &\leq -\tilde{\beta}^k(\frac{\gamma_{\min}}{\gamma_{\max}}\delta_1)^M||\xi^+_{t_{n_0+iM}}||^2 \end{aligned}$$

for a $\tilde{\beta}^k$ depending only on the modified constants of Lemma 5. The last inequality is obtained using $\Psi^{t_n+}_{t_0} = (P^+_n P^{-1}_n)\Psi^{t_n}_{t_0}$ and an obvious recursion over M time steps. We finally set $\beta = \tilde{\beta}^k(\frac{\gamma_{\min}}{\gamma_{\max}}\delta_1)^M$. □

Remark 19. *The control we have obtained on ξ_{t_0+s} is verified if $||\xi_{t_0+s}||$ is* already *in a ball of radius ε over the whole interval $[t_0, t_0+t]$. We now prove the result holds assuming* only *that ξ_{t_0} is sufficiently small.*

A.1.3 Proof of theorem 7

Applying Lemma 8 with $t_0+s = t_{n_{\max}}$ gives for $\frac{1}{k}\tilde{P}_{t_0} \leq P_{t_0} \leq k\tilde{P}_{t_0}$ and $||\xi_{t_0+s}|| < \varepsilon$ on $[0,t]$:

$$||\xi^+_{t_{n_{\max}}}||^2 \leq \frac{\gamma_{\max}}{\gamma_{\min}}||\xi_{t_0}||^2 + \gamma_{\max}l_1^k(||\xi_{t_0}||) - \frac{\gamma_{\max}}{\gamma_{\min}}\sum_{i=0}^{J-1}\left[\beta^k||\xi^+_{t_{n_0+iM}}||^2 - l_2(\xi^+_{t_{n_0+iM}})\right]$$

There exist $K > 0$ and $\varepsilon' > 0$ such that for $x < \varepsilon'$, we have $l_2(x) < \frac{\beta^k}{2}x$ and $\gamma_{\max}l_1^k(x) < Kx$ (as $l_1(x) = O(x)$) which gives: $||\xi^+_{t_{n_{\max}}}||^2 \leq \left(\frac{\gamma_{\max}}{\gamma_{\min}} + K\right)||\xi_0||^2$. Thus, for $||\xi_{t_0}|| < \frac{\varepsilon'}{\sqrt{\frac{\gamma_{\max}}{\gamma_{\min}}+K}}$:

$$||\xi_{t_0+s}||^2 \leq \left(\frac{\gamma_{\max}}{\gamma_{\min}} + K + K(\frac{\gamma_{\max}}{\gamma_{\min}} + K)\right)||\xi_0||^2 - \frac{\gamma_{\max}}{\gamma_{\min}}\sum_{i=0}^{J-1}\frac{\beta^k}{2}||\xi^+_{t_{n_0+iM}}||^2 \tag{A.4}$$

which finally ensures

$$||\xi_{t_0}|| < \frac{1}{2}\varepsilon'/\left(\frac{\gamma_{\max}}{\gamma_{\min}} + K + K(\frac{\gamma_{\max}}{\gamma_{\min}} + K)\right) \Rightarrow ||\xi_{t_0+s}|| \leq \varepsilon'/2$$

Reducing ε' if necessary to have $\varepsilon' \leq \varepsilon$, we have obtained $||\xi_{t_0+s}|| < \varepsilon'$ for $s \in [0,t] \Rightarrow ||\xi_{t_0+s}|| \leq \varepsilon'/2$ for sufficiently small $||\xi_{t_0}||$ (as Lemma 8 applies). Letting $t = \inf\{s, ||\xi_{t_0+s}|| \geq \frac{3}{4}\varepsilon'\}$ for sufficiently small $||\xi_{t_0}||$ we end up with a contradiction if we suppose $t < +\infty$, which proves $t = +\infty$. All the previous results thus hold *only* for sufficiently small $||\xi_{t_0}||$.

Moreover, (A.4) shows that $\sum_{i=0}^{J-1}\frac{\beta^k}{2}||\xi^+_{t_{n_0+iM}}||^2$ is bounded and has positive terms thus $||\xi^+_{t_{n_0+iM}}||^2$ goes to zero. Note also that $||P_t - \tilde{P}_t|| \underset{t\to+\infty}{\longrightarrow} 0$ as a byproduct.

A.2 Proof of proposition 8

Only conditions (1) and (5) are non-trivial. Let Φ denote the flow of the dynamics. We have $\frac{d}{dt}\left[(\Phi^t_{t_n})^T\Phi^t_{t_n}\right] = (\Phi^t_{t_n})^T \begin{pmatrix} 0 & 0 & -v_t \\ 0 & 0 & 0 \\ -v_t & 0 & 0 \end{pmatrix} \Phi^t_{t_n} \preceq v_{\max}(\Phi^t_{t_n})^T\Phi^t_{t_n}$ thus $(\Phi^{t_{n+1}}_{t_n})^T\Phi^{t_{n+1}}_{t_n} \succeq \exp\big(-v_{\max}(t_{n+1}-t_n)\big)$ and (1) is verified. The difficult part of (5) is the lower bound. Denoting $Cov(V_n)$ by N we will show:

$$\exists\beta_1, \forall n \in \mathbb{N}, \beta_1 I_3 \leq \hat{R}^T_{t_n} N^{-1} \hat{R}_{t_n} + (\Phi^{t_{n+1}}_{t_n})^T \begin{pmatrix} 0_{1,2} & I_2 \end{pmatrix}^T \hat{R}^T_{t_{n-1}} N^{-1} \hat{R}_{t_{n-1}} \begin{pmatrix} 0_{1,2} & I_2 \end{pmatrix} \Phi^{t_{n+1}}_{t_n}$$

That is to say that we want a lower bound on the quadratic form:

$$\begin{aligned} M\begin{pmatrix}\theta \\ u\end{pmatrix} &= \begin{pmatrix}\theta \\ u\end{pmatrix}^T \hat{R}^T_{t_n} N^{-1} \hat{R}_{t_n} \begin{pmatrix}\theta \\ u\end{pmatrix} \\ &\quad + \begin{pmatrix}\theta \\ u\end{pmatrix}^T (\Phi^{t_{n+1}}_{t_n})^T \begin{pmatrix} 0_{1,2} & I_2 \end{pmatrix}^T \hat{R}^T_{t_{n-1}} N^{-1} \hat{R}_{t_{n-1}} \begin{pmatrix} 0_{1,2} & I_2 \end{pmatrix} \Phi^{t_{n+1}}_{t_n} \begin{pmatrix}\theta \\ u\end{pmatrix} \end{aligned}$$

We decompose $\Phi^{t_{n+1}}_{t_n}$ as $\Phi^{t_{n+1}}_{t_n} = \begin{pmatrix} 1 & 0 & 0 \\ \delta V_n & & T_n \end{pmatrix}$. To simplify the writing we introduce the norms $||x||^2_N = x^T N^{-1} x$ and the associated scalar product $<.,.>_N$. There exists $\alpha > 0$ such that $\forall x \in \mathbb{R}^2, ||x||_N \geq \alpha||x||$. For any $\begin{pmatrix}\theta \\ u\end{pmatrix} \in \mathbb{R}^3$ we have $M\begin{pmatrix}\theta \\ u\end{pmatrix} = ||\hat{R}_{t_n}u||^2_N + ||\theta\hat{R}_{t_{n-1}}\delta V_n + \hat{R}_{t_{n-1}}T_n u||^2_N = ||\hat{R}_{t_n}u||^2_N + \theta^2||\hat{R}_{t_{n-1}}\delta V_n||^2_N + 2\theta < \hat{R}_{t_{n-1}}\delta V_n, \hat{R}_{t_{n-1}}T_n u >_N + ||\hat{R}_{t_{n-1}}T_n u||^2_N$ and for $\lambda \in]0,1]$ we have:

$$\begin{aligned} M\begin{pmatrix}\theta \\ u\end{pmatrix} =& ||\hat{R}_{t_n}u||^2_N + (1-\lambda^2)\theta^2||\hat{R}_{t_{n-1}}\delta V_n||^2_N \\ &+ \lambda^2\theta^2||\hat{R}_{t_{n-1}}\delta V_n||^2_N + 2\theta < \hat{R}_{t_{n-1}}\delta V_n, \hat{R}_{t_{n-1}}T_n u >_N + ||\hat{R}_{t_{n-1}}T_n u||^2_N \\ =& ||\hat{R}_{t_n}u||^2_N + (1-\lambda^2)\theta^2||\hat{R}_{t_{n-1}}\delta V_n||^2_N \\ &+ ||\lambda\theta\hat{R}_{t_{n-1}}\delta V_n + \frac{1}{\lambda}\hat{R}_{t_{n-1}}T_n u||^2_N + (1-\frac{1}{\lambda^2})||\hat{R}_{t_{n-1}}T_n u||^2_N \\ \geq& \alpha\Bigg[||\hat{R}_{t_n}u||^2 + (1-\lambda^2)\theta^2||\hat{R}_{t_{n-1}}\delta V_n||^2 \\ &+ ||\lambda\theta\hat{R}_{t_{n-1}}\delta V_n + \frac{1}{\lambda}\hat{R}_{t_{n-1}}T_n u||^2 + (1-\frac{1}{\lambda^2})||\hat{R}_{t_{n-1}}T_n u||^2\Bigg] \\ \geq& \alpha\left[||u||^2 + (1-\lambda^2)\theta^2||\delta V_n||^2 + (1-\frac{1}{\lambda^2})||u||^2\right] \\ \geq& \alpha\left[(2-\frac{1}{\lambda^2})||u||^2 + \theta^2 + [\frac{1-\lambda^2}{2-\frac{1}{\lambda^2}}||\delta V_n||^2 - 1]\theta^2\right] \\ \geq& \alpha(2-\frac{1}{\lambda^2})\left(||u||^2 + \theta^2 + [\frac{1-\lambda^2}{2-\frac{1}{\lambda^2}}v^2_{\min} - 1]\theta^2\right) \end{aligned}$$

As $\frac{1-\lambda^2}{2-\frac{1}{\lambda^2}} \underset{\lambda \to \frac{1}{\sqrt{2}}^-}{\longrightarrow} +\infty$ there exists λ_0 such that: $M\begin{pmatrix}\theta \\ u\end{pmatrix} \geq \alpha(2-\frac{1}{\lambda_0^2})\left|\left|\begin{matrix}\theta \\ u\end{matrix}\right|\right|^2$ and the result is true for $\beta_1 = \alpha(2-\frac{1}{\lambda_0^2})$.

Appendix B

Proofs of the results of Chapter 6

B.1 Proofs of the results of Section 6.4.1

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

Definition 14. *(First hitting time) Let S be a measurable space and $(X_n)_{n\geqslant 0}$ a Markov chain taking values in S. The first hitting time τ_C of a subset $C \subset S$ is defined by $\tau_C = \inf_{n\geqslant 0} \{X_n \in C\}$.*

Definition 15. *(Recurrent aperiodic Harris chain) Let S be a measurable space and $(X_n)_{n\geqslant 0}$ a Markov chain taking values in S. $(X_n)_{n\geqslant 0}$ is said to be a recurrent aperiodic Harris chain if there exist two sets $A, B \subset S$ satisfying the following properties:*

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

2. *There exists a probability measure ρ on B, and $\varepsilon > 0$ such that if $x \in A$ and $D \subset B$ then $\mathbb{P}(X_1 \in D | X_0 = x) > \varepsilon\rho(D)$.*
3. *There exists an integer $N \geqslant 0$ such that: $\forall x \in A, \forall n \geqslant N, \mathbb{P}(X_n \in A | X_0 = x) > 0$.*

The somewhat technical property 2 means that any given area of B can be reached from each point of A with non-vanishing probability.

Theorem 21. *[Harris, 1956] A recurrent aperiodic Harris chain admits a unique stationary distribution ρ_∞ and the density of the state X_n converges to ρ_∞ in T.V. norm for any distribution of X_0.*

Proof. See [44], Theorems 6.5 and 6.8 □

In other words, if for any initialization we can ensure that the process will come back with probability 1 to a central area A where mixing occurs, we have a convergence property. The following technical result is a cornerstone to demonstrate the theorems of Subsection 6.4.1.

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

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

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

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

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

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

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

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

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

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

3. $(\eta_n^+)_{n\geq 0}$ is an aperiodic recurrent Harris chain with $A = U_1$ and $B = U_0'$.

The conclusion will then immediately follow from Theorem 21. We give first a qualitative explanation of the approach. We have by assumption a sequence of sets $(U_n)_{n\geqslant 0}$. Intermediate result 1) states that starting from a set U_{n+1} the chain has a non-vanishing chance to jump to the "smaller" set U_n at the next time step. Intermediate result 2) states that starting from U_{n+1}, because of the non-vanishing property 1), the chain cannot avoid U_n forever and eventually reaches U_n, U_{n-1} and finally U_1. Intermediate result 3) brings out the fact that, $A = U_0'$ being totally included in the ball of radius α centered on any of its points, the chain can go from any given point of its pre-image $B = U_1$ to any area of A with controlled probability, which is the last property needed to obtain a recurrent aperiodic Harris chain.

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

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

3. Define $A = U_1$ and $B = U_0' = \mathscr{B}_o(x_0, \frac{\alpha}{2}) \cap C$. For any $V \subset B$ and $x \subset A$ we have $V \subset \mathscr{B}_o(q_0(x),\alpha)$, thus $Q(x,V) \geqslant \varepsilon \mu_G(V)$ by assumption (1). Thus $(\eta_n^+)_{n \geqslant 0}$ verifies property 2) of Definition 15. As $A = U_1$ the intermediate result 2) shows that $(\eta_n^+)_{n \geqslant 0}$ verifies the property 1) of Definition 15. As there exists a.s. a rank k such that $\eta_k \in U_1$ by intermediate result 2), there exists an integer M such that $\mathbb{P}(\eta_M \in U_1) > 0$. As $U_1 \cap U_0' \subset \mathscr{B}_o(q(x),\alpha)$ for any $x \in U_1$ (by definition of U_0' and U_1) Assumption (1) gives immediately $\mathbb{P}(\eta_{M+1} \in U_1 \cap U_0') > \mathbb{P}(\eta_M \in U_1)\varepsilon\mu_G(U_1 \cap U_0')$. Using only the definition of U_0' and U_1, and Assumption (2), we see easily that both contain x_0. Thus the set $U_0' \cap U_1$ is non-empty and open. We have then $\mu_G(U_1 \cap U_0') > 0$ and $\mathbb{P}(\eta_{M+1} \in U_1 \cap U_0') > 0$. By an obvious induction on m, $\mathbb{P}(\eta_m \in U_1 \cap U_0') > 0$ for any $m > M$. As $U_1 \cap U_0' \subset A$ we get $\mathbb{P}(\eta_m \in A) > 0$ for any $m > M$ and the property iii) of Definition 15 is also verified. We finally obtain that $(\eta_n^+)_{n \geqslant 0}$ is an aperiodic recurrent Harris chain.

Using Theorem 21 we can conclude that the density of η_n^+ converges in T.V. norm to its unique equilibrium distribution for any initialization. □

Building upon the previous results, the proof of Theorem 11 is as follows. The Markov chain defined by equations (6.10) and (6.11) with fixed Υ can be written under the form $\eta_{n+1}' = q_z(\eta_n') = q_{1,z} \circ q_{2,z}(\eta_n')$ with $z = (W_n, V_n) \in G \times \mathbb{R}^p$, $q_{1,z}(x) = \Upsilon W_n x \Upsilon^{-1}$ and $q_{2,z}(x) = xK(h(x,V_n))^{-1}$. Let $0 = (I_d, 0)$. We will show that $(\eta_n')_{n \geqslant 0}$ has all the properties required in Lemma 2.

Let $\alpha, \varepsilon, \varepsilon'$ be defined as in 6.4.1. Let α' such as $\mathscr{B}_o(I_d,\alpha') \subset \Upsilon \mathscr{B}_o(I_d, \frac{\alpha}{2})\Upsilon^{-1}$ and $\varepsilon'' = \varepsilon' |Ad_{\Upsilon^{-1}}| \varepsilon$, where $|Ad_{\Upsilon^{-1}}|$ is the determinant of the adjoint operator on $\mathfrak{g}$. For any

$x \in C$ and $V \subset \mathscr{B}_o(q_0(x), \alpha') = \mathscr{B}_o(I_d, \alpha')q_0(x)$ we have:

$$\begin{aligned} P(q_z(x) \in V) &\geqslant P(q_z(x) \in V, q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x), \frac{\alpha}{2})) \\ &\geqslant P(q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x), \frac{\alpha}{2}))\times \\ &P(\Upsilon W_n q_{2,z}(x)\Upsilon^{-1} \in V | q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x), \frac{\alpha}{2})) \end{aligned}$$

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

$$\begin{aligned} \Upsilon^{-1} V \Upsilon q_{2,z}(x)^{-1} &\subset \Upsilon^{-1} \mathscr{B}_o(I_d, \alpha') q_0(x) \Upsilon q_{2,z}(x)^{-1} \\ &\subset \mathscr{B}_o(I_d, \frac{\alpha}{2}) \Upsilon^{-1} q_0(x) \Upsilon q_{2,z}(x)^{-1} \\ &\subset \mathscr{B}_o(I_d, \frac{\alpha}{2}) q_{2,0}(x) q_{2,z}(x)^{-1} \\ &\subset \mathscr{B}_o(I_d, \frac{\alpha}{2} + \frac{\alpha}{2}) \end{aligned}$$

As W_n is independent from the other variables we obtain:

$$\begin{aligned} P(W_n \in \Upsilon^{-1} V \Upsilon q_{2,z}(x)^{-1} | q_{2,z}(x) \in \mathscr{B}_o(q_{2,0}(x), \frac{\alpha}{2})) &\geqslant \varepsilon \mu_G(\Upsilon^{-1} V \Upsilon q_{2,z}(x)^{-1}) \\ &\geqslant \varepsilon \mu_G(\Upsilon^{-1} V \Upsilon) \\ &\geqslant \varepsilon \mu_G(V) |Ad_{\Upsilon^{-1}}| \end{aligned}$$

And finally:

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

As q_0 has I_d as a fix point we only have to verify that the sets U_n as defined in Lemma 2 eventually cover the whole set C. It suffices to consider for any $n \geqslant 0$ the set $D_n = \{x \in C, \forall k \geqslant n, q_0^k(x) \in U_0'\}$. As we have $q_0^n(x) \to I_d$ almost-everywhere on C we get: $\mu_G(\cup_{n \geqslant 0} D_n) = \mu_G(C)$. As the sequence of sets $(D_n)_{n \geqslant 0}$ increases we have: $\mu_G(D_n) \underset{n \to \infty}{\longrightarrow} \mu_G(C)$. We introduce here the quantity $v_{\min} = \min_{x \in C} \mu_G(\mathscr{B}_o(x, \frac{\alpha'}{2}) \cap C)$ (the property $C = cl(C^o)$ and the regularity of μ_G ensure that we have $v_{\min} > 0$). Let $N \in \mathbb{N}$ be such that $\mu_G(D_N) > \mu_G(C) - v_{\min}$. We have then: $\forall y \in C, d(y, D_N) < \frac{\alpha'}{2}$ (otherwise we would have $\mu_G(D_N) \leqslant \mu_G(C) - v_{\min}$). As we have $D_N \subset U_N$ we obtain $\forall y \in C, d(y, D_N) < \frac{\alpha'}{2}$ thus $U_N' = C$ and $U_{N+1} = q_0^{-1}(U_N') = q_0^{-1}(C) = C$. So all the conditions of Lemma 2 hold and we can conclude convergence in T.V. norm to a stationary distribution which doesn't depend on the prior, provided that its support lies in C. Theorem 11 is proved.

Corollary 2 follows directly from the case $C = G$.

Theorem 12 can be proved as follows. Let $C = G$. Let $K = \{x \in G, h(x,0) = h(I_d, 0)\}$. Note that the left-right equivariance assumption ensures that K is a subgroup of G. First we show that there exists an integer N_1 such that $K \subset U_{N_1}$. As we have $\forall x \in K \; q_0(x) = \Upsilon x \Upsilon^{-1}$ the sequence of sets $Q_n = \Upsilon^{-n} U_n \Upsilon^n \cap K$ is growing and we have: $\{x \in K, \; d(x, Q_n) < \frac{\alpha}{2}\} \subset Q_{n+1}$.The set $Q_\infty = \cup_{n=1}^{\infty} Q_n$ is open in K as an union of open sets, but we have $\forall x \in Q_\infty, \mathscr{B}_o(x, \frac{\alpha}{2}) \cap K \subset Q_\infty$ thus $\forall x \in K \setminus Q_\infty, \mathscr{B}_o(x, \frac{\alpha}{4}) \cap K \subset K \setminus Q_\infty$. This implies that Q_∞ and $K \setminus Q_\infty$ are both open in K. As K is connected (see assumptions of Theorem 12) we

obtain $Q_\infty = K$. As K is compact (as a closed subset of a compact) the open cover $\cup_{i=1}^{\infty} Q_n$ has a finite subcover and there exists an integer N_1 such that $Q_{N_1} = K$, i.e. $\Upsilon^{N_1} K \Upsilon^{-N_1} = U_{N_1}$. As K is a subgroup of G (see above) containing Υ we obtain $K \subset U_{N_1}$. Now we have to prove that the sets $(U_n)_{n \geqslant 0}$ eventually cover the whole set C. For any $x \in G$ we have $h(q_0^n(x), 0) \to h(I_d, 0)$. Thus there exists a rank n such that $\forall k \geqslant n, q_0^k(x) \in U'_N$ (otherwise we could extract a subsequence from $(q_0^n(x))_{n \geqslant 0}$ which stays at a distance $> \frac{\alpha}{2}$ from K, as G is compact we could extract again a convergent subsequence and its limit $q_0^\infty(x)$ would be outside K, thus we would have $h(q_0^n(x)) \not\to h(I_d, 0)$). Here we can define as in the proof of Theorem 11 the sets $D_n = \{x \in C, \forall k \geqslant n, q_0^k(x) \in U'_{N_1}\}$ (note that U'_1 has been replaced by U'_{N_1} in this definition of D_n). As for almost any $x \in G$ there exists a rank n such that $x \in D_n$ we have $\mu_G(D_n) \to \mu_G(C)$ and as in the proof of Theorem 11 the sets U_n eventually cover the set C.

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

$$\begin{aligned}\zeta_{n+1} = \tilde{\Upsilon}\eta_{n+1} = \tilde{\Upsilon}\Upsilon W_n \eta_n^+ \Upsilon^{-1} = \Upsilon\tilde{\Upsilon} W_n \eta_n^+ \Upsilon^{-1} &\overset{\mathscr{L}}{=} \Upsilon W_n \tilde{\Upsilon}\eta_n^+ \Upsilon^{-1} \\ &\overset{\mathscr{L}}{=} \Upsilon W_n \zeta_n^+ \Upsilon^{-1}\end{aligned}$$

$$\begin{aligned}\zeta_{n+1}^+ &= \tilde{\Upsilon}\eta_{n+1} K_{n+1}(h(\eta_{n+1}, V_{n+1}))^{-1} \\ &\overset{\mathscr{L}}{=} \tilde{\Upsilon}\eta_{n+1} K_{n+1}(h(\tilde{\Upsilon}\eta_{n+1}, V_{n+1}))^{-1} \\ &\overset{\mathscr{L}}{=} \zeta_{n+1} K_{n+1}(h(\zeta_{n+1}, V_{n+1}))^{-1}\end{aligned}$$

where we have used the property:

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

We see that the law of the error process is invariant under left multiplication by $\tilde{\Upsilon}$, thus the asymptotic distribution π inherits this property.

B.2 Proofs of the results of Section 6.4.2

B.2.1 Proof of Proposition 18

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

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

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

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

$$\begin{aligned}
|\frac{d}{dt}\langle k(\tilde{\gamma}_t), k(\tilde{\gamma}_n)\rangle| &= |\langle \frac{d}{dt}k(\tilde{\gamma}_t), k(\tilde{\gamma}_n)\rangle| \\
&\leqslant |\frac{d}{dt}k(\tilde{\gamma}_t)||k(\tilde{\gamma}_n)| \\
&\leqslant |\frac{d}{dt}\tilde{\gamma}_t||k(\tilde{\gamma}_n)| \text{ (due to } |\frac{\partial k}{\partial x}| \leqslant 1 \text{)} \\
&\leqslant |k(\tilde{\gamma}_n)|^2 \text{ (as } \frac{d}{dt}\tilde{\gamma}_t = -\tilde{\gamma}_t k(\tilde{\gamma}_n) \text{)}
\end{aligned}$$

Thus:

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

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

B.2.2 Proof of Proposition 17

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

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

Let $k(\gamma) = k_1(\gamma^T b_1) \times b_1 + k_2(\gamma^T b_2) \times b_2$ and $E : \gamma \to k_1||\gamma^T b_1 - b_1||^2 + k_2||\gamma^T b_2 - b_2||^2$. To apply Proposition 18 we will first verify that 1) $\forall \gamma \in SO(3), \gamma.k(\gamma) = grad_E(\gamma)$ 2) $|\frac{\partial k}{\partial \gamma}| \leqslant 1$. To prove 1) consider the dynamics in $\in SO(3)$ defined by $\frac{d}{dt}\gamma_t = \gamma_t(\psi_t)_\times$ for some rotation vector path ψ_t. We have $\frac{d}{dt}E(\gamma_t) = k_1(\gamma_t^T b_1 - b_1)^T \frac{d}{dt}\gamma_t^T b_1 + k_2(\gamma_t^T b_2 - b_2)^T \frac{d}{dt}\gamma_t^T b_2$. Using triple product equalities this is equal to $\langle k_1(\gamma_t^T b_1)_\times b_1 + k_2(\gamma_t^T b_2)_\times b_2, \psi_t\rangle = \langle k(\gamma_t), \psi_t\rangle = \langle \gamma_t.k(\gamma_t), \frac{d}{dt}\gamma_t\rangle$. Thus $\gamma.k(\gamma) = grad_E(\gamma)$. To prove 2) we analogously see that $\frac{d}{dt}k(\gamma_t) = -[k_1(b_1)_\times(\gamma_t^T b_1)_\times + k_2(b_2)_\times(\gamma_t^T b_2)_\times]\psi_t$. Thus $|\frac{d}{dt}k(\gamma_t)| \leqslant (k_1+k_2)|\psi_t|$ and finally $|\frac{\partial k}{\partial \gamma}| \leqslant 1$. Now, except if initially γ_0 is the rotation of axis $b_1 \times b_2$ and angle π, the function E is strictly decreasing, and I_d is the only point in the sublevel set $\{\gamma \in SO(3) \mid E(\gamma) \leq E(\gamma_0)\}$ such that $grad_E(\gamma) = 0$. Applying Proposition 18 allows to prove Proposition 17.

Bibliography

[1] N. Aghannan and P. Rouchon. On invariant asymptotic observers. In *41st IEEE Conference on Decision and Control*, pages 1479–1484, 2002.

[2] V. Andrieu and L. Praly. On the existence of a kazantzis-kravaris/luenberger observer. *SIAM Journal on Control and Optimization*, 45:432–446, 2006.

[3] Vincent Andrieu and Laurent Praly. A unifying point of view on output feedback designs for global asymptotic stabilization. *Automatica*, 45(8):1789–1798, 2009.

[4] Vincent Andrieu, Laurent Praly, and Alessandro Astolfi. Asymptotic tracking of a reference trajectory by output-feedback for a class of non linear systems. *Systems & Control Letters*, 58(9):652–663, 2009.

[5] Vincent Andrieu, Laurent Praly, and Alessandro Astolfi. High gain observers with updated gain and homogeneous correction terms. *Automatica*, 45(2):422–428, 2009.

[6] Didier Auroux and Jacques Blum. Back and forth nudging algorithm for data assimilation problems. *Comptes Rendus Mathematique*, 340(12):873–878, 2005.

[7] Albert-Jan Baerveldt and Robert Klang. A low-cost and low-weight attitude estimation system for an autonomous helicopter. In *Intelligent Engineering Systems, 1997. INES'97. Proceedings., 1997 IEEE International Conference on*, pages 391–395, 1997.

[8] Tim Bailey, Juan Nieto, Jose Guivant, Michael Stevens, and Eduardo Nebot. Consistency of the ekf-slam algorithm. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 3562–3568. IEEE, 2006.

[9] Martin Barczyk, Silvere Bonnabel, Jean-Emmanuel Deschaud, and François Goulette. Invariant ekf design for scan matching-aided localization. 2015.

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

[11] A. Barrau and S. Bonnabel. Intrinsic filtering on lie groups with applications to attitude estimation. *arXiv preprint arXiv:1310.2539*, 2013. IEEE Transactions on Automatic Control. In press.

[12] Axel Barrau and Silvère Bonnabel. The invariant extended kalman filter as a stable observer. *arXiv preprint arXiv:1410.1465*, 2014.

[13] Axel Barrau and Silvere Bonnabel. Invariant particle filtering with application to localization. In *IEEE Conference on Decision and Control*, 2014.

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

[15] S. Bonnabel. *Observateurs asymptotiques invariants: théorie et exemples*. PhD thesis, Ecole des Mines de Paris, 2007.

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

[17] S. Bonnabel, P. Martin, and E. Salaun. Invariant Extended Kalman Filter: Theory and application to a velocity-aided attitude estimation problem. In *IEEE Conference on Decision and Control*, 2009.

[18] S. Bonnabel, Ph. Martin, and P. Rouchon. A non-linear symmetry-preserving observer for velocity-aided inertial navigation. In *American Control Conference (ACC06)*, pages 2910–2914, June 2006.

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

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

[21] S. Bonnabel, Ph. Martin, and E. Salaun. Invariant extended Kalman filter: theory and application to a velocity-aided attitude estimation problem. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 1297–1304. IEEE, 2009.

[22] Silvere Bonnabel. Symmetries in observer design: Review of some recent results and applications to ekf-based slam. In *Robot Motion and Control 2011*, pages 3–15. Springer, 2012.

[23] Silvere Bonnabel and Jean-Jacques Slotine. A contraction theory-based analysis of the stability of the extended kalman filter. *arXiv preprint arXiv:1211.6624*, 2012.

[24] M. Boutayeb, H. Rafaralahy, and M. Darouach. Convergence analysis of the extended Kalman filter used as an observer for nonlinear deterministic discrete-time systems. *IEEE transactions on automatic control*, 42, 1997.

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

[26] Francesco Bullo. *Geometric control of mechanical systems*, volume 49. Springer Science & Business Media, 2005.

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

[28] Olivier Cappé, Simon J Godsill, and Eric Moulines. An overview of existing methods and recent advances in sequential monte carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.

[29] Francois Caron, Emmanuel Duflos, Denis Pomorski, and Philippe Vanheeghe. Gps/imu data fusion using multisensor kalman filtering: introduction of contextual aspects. *Information Fusion*, 7(2):221–230, 2006.

[30] Luis Rodolfo García Carrillo, Alejandro Enrique Dzul López, Rogelio Lozano, and Claude Pégard. Combining stereo vision and inertial navigation system for a quadrotor uav. *Journal of Intelligent & Robotic Systems*, 65(1-4):373–387, 2012.

[31] José A Castellanos, José Neira, and Juan D Tardós. Limits to the consistency of ekf-based slam 1. 2004.

[32] José A Castellanos, Juan D Tardós, and Günther Schmidt. Building a global map of the environment of a mobile robot: The importance of correlations. In *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, volume 2, pages 1053–1059. IEEE, 1997.

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

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

[35] John L Crassidis. Sigma-point kalman filtering for integrated gps and inertial navigation. *Aerospace and Electronic Systems, IEEE Transactions on*, 42(2):750–756, 2006.

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

[37] Frank Dellaert and Michael Kaess. Square root sam: Simultaneous localization and mapping via square root information smoothing. *The International Journal of Robotics Research*, 25(12):1181–1203, 2006.

[38] J. Deyst and C. Price. Conditions for asymptotic stability of the discrete minimum-variance linear estimator. *IEEE Trans. on Automatic Control*, 13:702–705, 1968.

[39] A. Doucet, N. De Freitas, K. Murphy, and S. Russell. Rao-blackwellised particle filtering for dynamic bayesian networks. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 176–183. Morgan Kaufmann Publishers Inc., 2000.

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

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

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

[43] Hugh F Durrant-Whyte. An autonomous guided vehicle for cargo handling applications. *The International Journal of Robotics Research*, 15(5):407–440, 1996.

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

[45] Jay Farrell. *Aided navigation: GPS with high rate sensors*. McGraw-Hill, Inc., 2008.

[46] Matthieu Fruchard, Pascal Morin, and Claude Samson. A framework for the control of nonholonomic mobile manipulators. *The International Journal of Robotics Research*, 25(8):745–780, 2006.

[47] N Ganganath and H Leung. Mobile robot localization using odometry and kinect sensor. In *Emerging Signal Processing Applications (ESPA), 2012 IEEE International Conference on*, pages 91–94. IEEE, 2012.

[48] Jean-Paul Gauthier and Ivan AK Kupka. Observability and observers for nonlinear systems. *SIAM Journal on Control and Optimization*, 32(4):975–994, 1994.

[49] J.P. Gauthier and I. Kupka. *Deterministic Observation Theory and Applications*. Cambridge University Press, 2001.

[50] Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET, 1993.

[51] Tarek Hamel, Robert Mahony, Jochen Trumpf, Pascal Morin, and Minh-Duc Hua. Homography estimation on the special linear group based on direct point correspondence. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 7902–7908. IEEE, 2011.

[52] H. Hammouri and JP. Gauthier. Global time-varying linearization up to output injection. *SIAM J. Control Optim.*, 30:1295–1310, 1992.

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

[54] Thibault Hervier, Silvère Bonnabel, and François Goulette. Accurate 3D maps from depth images and motion sensors via nonlinear Kalman filtering. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 5291–5297. IEEE, 2012.

[55] Ch. Hide, T. Moore, and M. Smith. Adaptive kalman filtering for low-cost ins/gps. *The Journal of Navigation*, 56(01):143–152, 2003.

[56] Albert S Huang, Abraham Bachrach, Peter Henry, Michael Krainin, Daniel Maturana, Dieter Fox, and Nicholas Roy. Visual odometry and mapping for autonomous flight using an rgb-d camera. In *International Symposium on Robotics Research (ISRR)*, pages 1–16, 2011.

[57] Guoquan P Huang, Anastasios I Mourikis, and Stergios I Roumeliotis. Analysis and improvement of the consistency of extended kalman filter based slam. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 473–479. IEEE, 2008.

[58] Guoquan P Huang, Anastasios I Mourikis, and Stergios I Roumeliotis. Observability-based rules for designing consistent ekf slam estimators. *The International Journal of Robotics Research*, 29(5):502–528, 2010.

[59] Shoudong Huang and Gamini Dissanayake. Convergence and consistency analysis for extended kalman filter based slam. *Robotics, IEEE Transactions on*, 23(5):1036–1049, 2007.

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

[61] Simon J Julier and Jeffrey K Uhlmann. A counter example to the theory of simultaneous localization and map building. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 4, pages 4238–4243. IEEE, 2001.

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

[63] Velimir Jurdjevic. *Geometric control theory*. Cambridge university press, 1997.

[64] Velimir Jurdjevic and Héctor J Sussmann. Control systems on lie groups. *Journal of Differential equations*, 12(2):313–329, 1972.

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

[66] Christian Kerl, Jurgen Sturm, and Daniel Cremers. Robust odometry estimation for rgb-d cameras. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 3748–3754. IEEE, 2013.

[67] Alexander A Kirillov. *An introduction to Lie groups and Lie algebras*, volume 113. Cambridge University Press Cambridge, 2008.

[68] Kurt Konolige, Motilal Agrawal, and Joan Sola. Large-scale visual odometry for rough terrain. In *Robotics Research*, pages 201–212. Springer, 2011.

[69] Arthur J Krener. The convergence of the extended kalman filter. In *Directions in mathematical systems theory and optimization*, pages 173–182. Springer, 2003.

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

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

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

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

[74] D. Luenberger. An introduction to observers. *IEEE Trans. on Automatic Control*, 16(6):596–602, 1971.

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

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

[77] Robert Mahony, Tarek Hamel, Pascal Morin, and Ezio Malis. Nonlinear complementary filters on the special linear group. *International Journal of Control*, 85(10):1557–1573, 2012.

[78] Robert Mahony, Tarek Hamel, and J-M Pflimlin. Complementary filter design on the special orthogonal group SO(3). In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 1477–1484. IEEE, 2005.

[79] Robert Mahony, Tarek Hamel, and Jean-Michel Pflimlin. Nonlinear complementary filters on the special orthogonal group. *Automatic Control, IEEE Transactions on*, 53(5):1203–1218, 2008.

[80] Ezio Malis, Tarek Hamel, Robert Mahony, and Pascal Morin. Estimation of homography dynamics on the special linear group. In *Visual Servoing via Advanced Numerical Methods*, pages 133–150. Springer, 2010.

[81] F Landis Markley. Attitude error representations for kalman filtering. *Journal of guidance, control, and dynamics*, 26(2):311–317, 2003.

[82] Ph. Martin, E. Salaün, et al. Generalized multiplicative extended kalman filter for aided attitude and heading reference system. In *Proc. AIAA Guid., Navigat., Control Conf*, pages 1–13, 2010.

[83] Sergey Matyunin, Dmitriy Vatolin, Yury Berdnikov, and Maxim Smirnov. Temporal filtering for depth maps generated by kinect depth camera. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, pages 1–4. IEEE, 2011.

[84] Michael Montemerlo, Sebastian Thrun, Daphne Koller, Ben Wegbreit, et al. Fastslam: A factored solution to the simultaneous localization and mapping problem. In *AAAI/IAAI*, pages 593–598, 2002.

[85] Pascal Morin and Claude Samson. Time-varying exponential stabilization of a rigid spacecraft with two control torques. *Automatic Control, IEEE Transactions on*, 42(4):528–534, 1997.

[86] Pascal Morin and Claude Samson. Practical stabilization of driftless systems on lie groups: the transverse function approach. *Automatic Control, IEEE Transactions on*, 48(9):1496–1508, 2003.

[87] Pascal Morin and Claude Samson. Trajectory tracking for nonholonomic vehicles. In *Robot Motion and Control*, pages 3–23. Springer, 2006.

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

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

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

[91] Honghui Qi and John B Moore. Direct kalman filtering approach for gps/ins integration. *Aerospace and Electronic Systems, IEEE Transactions on*, 38(2):687–693, 2002.

[92] F. Sonnemann Reif, K. and R. Unbehauen. An ekf-based nonlinear observer with a prescribed degree of stability. *Automatica*, 34:1119–1123, 1998.

[93] Alessandro Saccon, Jochen Trumpf, Robert Mahony, and A Pedro Aguiar. Second-order-optimal filters on lie groups. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 4434–4441. IEEE, 2013.

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

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

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

[97] Glauco Garcia Scandaroli, Pascal Morin, and Geraldo Silveira. A nonlinear observer approach for concurrent estimation of pose, imu bias and camera-to-imu rotation. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 3335–3341. IEEE, 2011.

[98] Korbinian Schmid, Teodor Tomic, Felix Ruess, Heiko Hirschmuller, and Marianna Suppa. Stereo vision based indoor/outdoor navigation for flying robots. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 3955–3962. IEEE, 2013.

[99] Max Schuler. Die störung von pendel und kreiselapparaten durch die beschleunigung des fahrzeuges'. *Physikalische Zeitschrift*, 24(16):344–350, 1923.

[100] M Simandl and O. Straka. Sampling densities of particle filter: a survey and comparison. In *American Control Conference, 2007. ACC'07*, pages 4437–4442. IEEE, 2007.

[101] Randall Smith, Matthew Self, and Peter Cheeseman. Estimating uncertain spatial relationships in robotics. In *Autonomous robot vehicles*, pages 167–193. Springer, 1990.

[102] Y.K. Song and J.W. Grizzle. The extended Kalman filter as a local asymptotic observer. *Estimation and Control*, 5:59–78, 1995.

[103] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*. MIT press, 2005.

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

[105] Ivan Vasconcelos, Paul Sava, and Huub Douma. Nonlinear extended images via image-domain interferometry. *Geophysics*, 75(6):SA105–SA115, 2010.

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

[107] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. IEEE, 2000.

[108] Jan Wendel, Oliver Meister, Christian Schlaile, and Gert F Trommer. An integrated gps/mems-imu navigation system for an autonomous helicopter. *Aerospace Science and Technology*, 10(6):527–533, 2006.

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

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

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

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

[113] Junlan Yang, Dan Schonfeld, and Magdi Mohamed. Robust video stabilization based on particle filter tracking of projected camera motion. *Circuits and Systems for Video Technology, IEEE Transactions on*, 19(7):945–954, 2009.

Acknowledgements

Most people involved in this work being French-speaking, I'm shifting to this language in the present section.

Ces travaux de thèse n'auraient pas rencontré le même succès sans l'implication de mon directeur de thèse Silvère Bonnabel. Peu de doctorants ont eu comme moi la possibilité de discuter à tout moment avec leur encadrant, ni bénéficié d'une aide aussi conséquente pour rédiger leurs publications.

Mon responsable au sein de Sagem, Xavier Bissuel, a également joué un rôle essentiel dans ces recherches. Son expérience a été précieuse pour mettre en relation la théorie développée dans ce document avec les applications industrielles les plus prometteuses.

Emmanuel Robert et son équipe, en particulier Thierry Perrot et Olivier Guitton, m'ont fourni une aide précieuse dans la mise en oeuvre de mes algorithmes sur des produits commerciaux. Thierry Perrot a de plus contribué à la rédaction de ce document par ses relectures attentives.

Le stagiaire que j'ai eu la joie d'encadrer, Thomas Bernal, a également fournit un travail remarquable qui m'a fait économiser un temps précieux.

La collaboration entre Sagem et l'Ecole des Mines sur les techniques de filtrage non linéaire, sans laquelle cette thèse n'aurait pas eu lieu, a été initiée par MM. Erwan Salaun, Daniel Duclos, Yann Le Guilloux, Xavier Bissuel et Emmanuel Robert.

Je tiens aussi à remercier Sébastien Marcade et Loic Davain pour avoir proposé d'appliquer les méthodes de navigation que j'ai conçues aux projets dont ils étaient en charge. Ce sont eux qui rendu possible la transition d'un travail purement théorique vers une mise en oeuvre concrète profitable d'un point de vue industriel.

Enfin, je suis reconnaissant à toutes les personnes ayant fréquenté le centre de robotique de l'Ecole des Mines de Paris, en particulier le bureau V2O (Manu Alibay, Florent

Taralle, David Aura, Xiangjun Qian, Lucille Callebert, Raphaelle Ducret) et son annexe du 7 rue Royer-Collard (Cyril Joly, Étienne Servais), pour ces trois années très agréables.

Titre

Résumé : Cette thèse étudie l'utilisation de variables d'erreurs non linéaires dans la conception de filtres de Kalman étendus (EKF). La théorie des observateurs invariants sur les groupes de Lie sert de point de départ au développement d'un cadre plus général mais aussi plus simple, fournissant des variables d'erreur non linéaires assurant la propriété nouvelle et surprenante de suivre une équation différentielle (partiellement) linéaire. Ce résultat est mis à profit pour prouver, sous des hypothèses naturelles d'observabilité, la stabilité de l'EKF invariant (IEKF) une fois adapapté à la classe de systèmes (non-invariants) introduite. Le gain de performance remarquable par rapport à l'EKF classique est illustré par des applications à des problèmes industriels réels, réalisées en partenariat avec l'entreprise SAGEM. Dans une seconde approche, les variables d'erreurs sont étudiées en tant que processus stochastiques. Pour les observateurs convergeant globalement si les bruits sont ignorés, on montre que les ajouter conduit la variable d'erreur à converger en loi vers une distribution limite indépendante de l'initialisation. Ceci permet de choisir des gains à l'avance en optimisant la densité asymptotique. La dernière approche adoptée consiste à prendre un peu de recul vis-à-vis des groupes de Lie, et à étudier les EKF utilisant des variables d'erreur non linéaires de façon générale. Des propriété globales nouvelles sont obtenues. En particulier, on montre que ces méthodes permettent de résoudre le célèbre problème de fausse observabilité créé par l'EKF s'il est appliqué aux questions de localisation et cartographie simultanées (SLAM).
Mots clés : Estimation non linéaire, Filtrage de Kalman étendu, Navigation, Localisation et Cartographie Simultanées, Méthodes géométriques.

Title

Abstract: The present thesis explores the use of non-linear state errors to devise extended Kalman filters (EKFs). First we depart from the theory of invariant observers on Lie groups and propose a more general yet simpler framework allowing to obtain non-linear error variables having the novel unexpected property of being governed by a (partially) linear differential equation. This result is leveraged to ensure local stability of the invariant EKF (IEKF) under standard observability assumptions, when extended to this class of (non-invariant) systems. Real applications to some industrial problems in partnership with the company SAGEM illustrate the remarkable performance gap over the conventional EKF. A second route we investigate is to turn the noise on and consider the invariant errors as stochastic processes. Convergence in law of the error to a fixed probability distribution, independent of the initialization, is obtained if the error with noise turned off is globally convergent, which in turn allows to assess gains in advance that minimize the error's asymptotic dispersion. The last route consists in stepping back a little and exploring general EKFs (beyond the Lie group case) relying on a non-linear state error. Novel mathematical (global) properties are derived. In particular, these methods are shown to remedy the famous problem of false observability created by the EKF if applied to simultaneous localization and mapping (SLAM), which is a novel result.
Keywords: Non-linear estimation, extended Kalman filtering, Navigation, SLAM, Geometric methods

MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY