Observers on linear Lie groups with linear estimation error dynamics

by

Mikhail Koldychev

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A major motivation for Lie group observers is their application as sensor fusion algorithms for an inertial measurement unit which can be used to estimate the orientation of a rigid-body. In the first part of this thesis we propose several types of nonlinear, deterministic, locally exponentially convergent, state observers for systems with all, or part, of their states evolving on the general linear Lie group of invertible $n \times n$ matrices. Our proposed Lie group observer with full-state measurement is applicable to left-invariant systems on linear Lie groups and yields linear estimation error dynamics. We also propose a way to extend our full-state observer, to build observers with partial-state measurement, i.e., only a proper subset of the states are available for measurement. Our proposed Lie group observer with partial-state measurement is applicable when the measured states are evolving on a Lie group and the rest of the states are evolving on the Lie algebra of this Lie group. We illustrate our observer designs on various examples, including rigid-body orientation estimation and dynamic homography estimation.

In the second part of this thesis we propose a nonlinear, deterministic state observer, for systems that evolve on real, finite-dimensional vector spaces. This observer uses the property of high-gain observers, that they are approximate differentiators of the output signal of a plant. Our new observer is called a composite high-gain observer because it consists of a chain of two or more sub-observers. The first sub-observer in the chain differentiates the output of the plant. The second sub-observer in the chain differentiates a certain function of the states of the first sub-observer. Effectiveness of the composite observer is demonstrated via simulation.

Acknowledgements

I would like to thank Dr. Christopher Nielsen, my supervisor, for being an excellent mentor and for providing not only valuable research assistance, but also helpful encouragement and motivation. Thank you for giving me the chance to explore a topic that interests me very much in control theory.

Dedication

I dedicate this to all my family and friends.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Feedback control for finite dimensional systems can be broadly catagorized into full state feedback and output feedback configurations. In full state feedback control the input signal to the plant is calculated based on present and/or past states of the plant. In output feedback control, the input signal is calculated based on present and/or past measured outputs of the plant.

Every feedback control system uses sensors that measure either the full state or the output of the plant. All physical sensors have their measurements corrupted by "noise", which could be caused by sensor imperfections and external disturbances. Therefore, any feedback controller will necessarily be affected by noise and uncertainty. Observers provide a method to reduce the effect of noise on feedback control systems. Observers also provide a framework for designing output feedback controllers. A control engineer first designs the controller under the assumption that the full state of the plant is available for feedback. Subsequently, an observer is designed to estimate the plant state and the state estimates are used to implement the controller. Provided that some version of the separation principle holds, the closed-loop system will behave as if the states were directly measured.

An observer is a data processor that performs state estimation by measuring the inputs and outputs of a plant. Observers can greatly improve the quality of sensor measurements, because an observer knows the model of the plant and can use this model to its advantage. Typically, both the plant and the observer are sets of differential equations, however there is a fundamental difference between the two. The plant is a "physical" process, that we can only affect through its input. On the other hand, the observer is a "virtual" process in the sense that it runs on a computer.

1.1 Motivation

Suppose we are given a signal of time $y(t) = x(t) + d(t)$, which is the sum of a "useful" signal $x \in \mathbb{R}^n$ and a noise/disturbance signal $d \in \mathbb{R}^n$. Our goal is to apply a filter to the signal $y(t)$, such that $d(t)$ is attenuated while $x(t)$ is preserved. If we do not know the model of the system from which the signal $x(t)$ originated, then we can try to use a band-pass filter to attenuate frequencies in $y(t)$, which are not expected to be present in $x(t)$. If we do know the model of the system from which the signal $x(t)$ originated, then we can build a more complicated type of filter, called a full-state observer, to attenuate $d(t)$ and preserve $x(t)$.

The scenario above is made more difficult by removing some of the output channels from y, i.e., making $y(t) = x_1(t) + d_1(t)$, where $x_1, d_1 \in \mathbb{R}$ are the first components of the vectors $x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^\top$ and $d = \begin{bmatrix} d_1 & \dots & d_n \end{bmatrix}^\top$. Suppose we know the model of the system, from which $x(t)$ originated. We would like to estimate the state vector $x(t_0)$, at time $t_0 > 0$, from knowledge of previous measurements of the output $\{y(\tau), 0 \leq \tau \leq t_0\}$. One type of algorithm which can be used for this task is a partial-state observer.

Let us demonstrate, via simulation, how sensor noise affects controller performance and how a simple full-state observer can be used to reduce the negative effect of sensor noise on the tracking performance of the controller. To this end, consider the simple control system

$$\begin{aligned} \dot{x} &= u \\ y &= x, \end{aligned} \tag{1.1}$$

where $x \in \mathbb{R}$ is the state, $u \in \mathbb{R}$ is the control input. Suppose we want the output, $y(t)$, of the system (1.1), to track a piece-wise constant reference trajectory, $y_{ref}(t)$. This task can be achieved by using a simple state feedback control law

$$u = K(y_{ref} - x) \tag{1.2}$$

where $K > 0$ is a controller gain that determines the rate at which y approaches y_{ref}. The controller (1.2) uses the current state $x(t)$ to calculate the current control $u(t)$. The signal $x(t)$ is directly available from the output $y = x$, so (1.2) is implemented as

$$u = K(y_{ref} - y). \tag{1.3}$$

The signal y is measured by a sensor and therefore contains measurement noise. Intuitively one expects that, the more noise present in the signal y, the worse the controller performs.

To reduce the deterioration of controller performance, we can augment the controller (1.3) so that it uses an estimated version of x, denoted $\hat{x}$, obtained via the observer

$$\dot{\hat{x}} = u + a_0(y - \hat{x}) \tag{1.4}$$

where $a_0 > 0$ is the observer gain. The observer (1.4) behaves similar to a low-pass filter. Namely, when $u = 0$, the transfer function from y to $\hat{x}$ is given by

$$\frac{\hat{X}(s)}{Y(s)} = \frac{a_0}{s + a_0}$$

which is the transfer function of a low-pass filter. However, the observer (1.4) is much more powerful than a low-pass filter, because it uses an extra piece of information, namely the input u, to filter out noise from y. Intuitively, when the state of the plant (1.1) changes, it is because the input u has changed it. Since the observer has access to u, it can "mimic" this change of state, by applying the same input u to its own state (1.4). Thus, any high-frequency movements in the state of the plant are copied by the observer, without using the output y, and therefore without low-pass filtering. The reason why the transfer function of the observer (1.4) is that of a low-pass filter, when $u = 0$, is because the state of the plant is constant when $u = 0$, so there is no useful high-frequency content in y. This type of observer is sometimes called a complementary filter, because the signals y and u "complement" each other in such a way that allows us to filter out high-frequency noise from y, while preserving much of the useful high-frequency information in y.

When we connect the static controller (1.3) with the observer (1.4), we get the dynamic control law

$$\begin{aligned} \dot{\hat{x}} &= u + a_0(y - \hat{x}) \\ u &= K(y_{ref} - \hat{x}). \end{aligned} \tag{1.5}$$

We simulate the two controllers: (1.3) and (1.5), for tracking a piece-wise constant signal $y_{ref}(t)$, under increasing sensor noise. The output is taken to be $y = x + d$, where d is normally distributed with zero mean and with standard deviation taking on different values $\sigma \in \{0.1, 0.5, 1.0, 1.5\}$, to simulate increasing sensor noise. The controller gain is taken to be $K = 6$ and the observer gain is taken to be $a_0 = 0.4$. The results are shown in Figure (1.1).

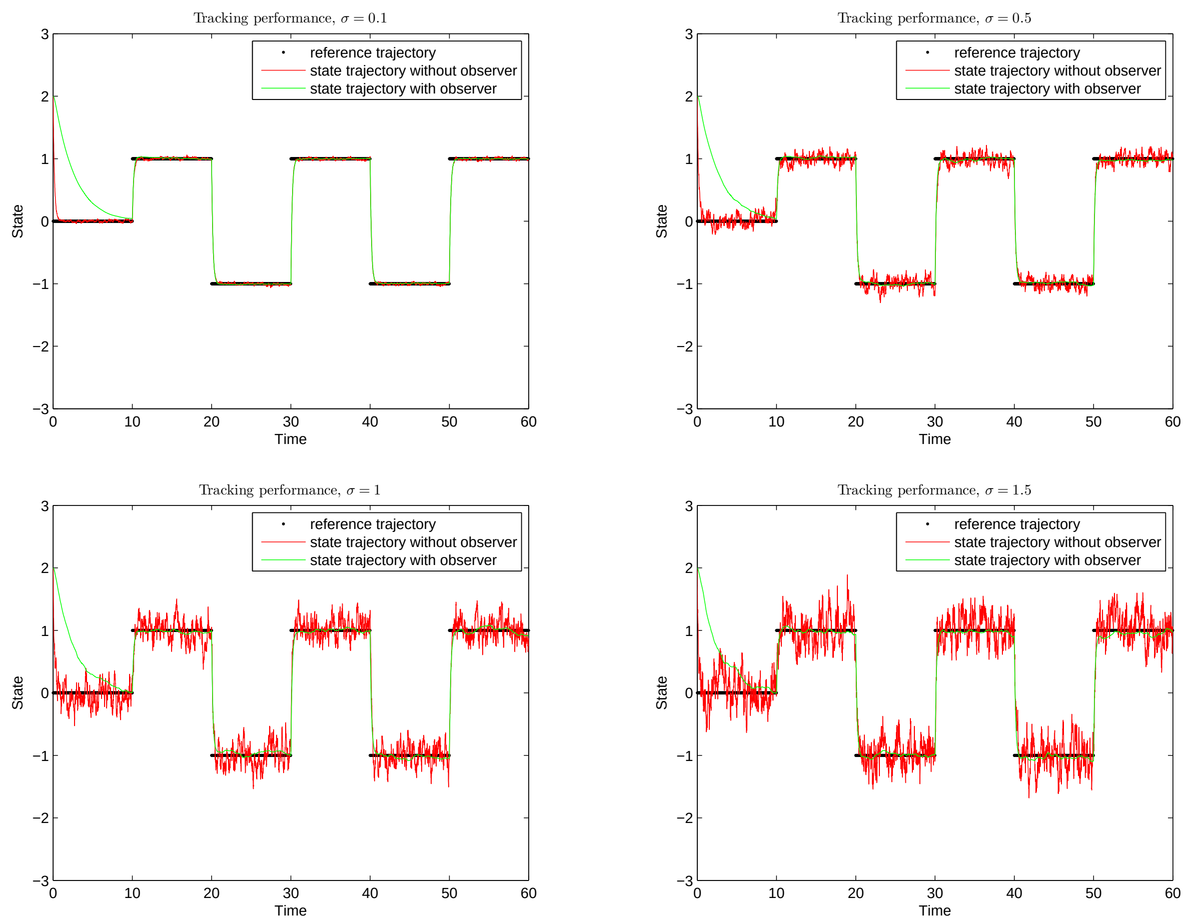

Figure 1.1: Comparison of controller (1.3) versus the dynamic controller (1.5), in tracking a piece-wise constant reference trajectory.

Figure (1.1) illustrates that the dynamic controller (1.5) takes a bit of time to converge (about $t = 10$ is when $\hat{x}$ becomes sufficiently close to x). But once $\hat{x}$ becomes sufficiently close to x, the dynamic controller (1.5) is much more resistant to noise than the static controller (1.3). Resistance to noise is not achieved by temporal smoothing of $y(t)$ by a low-pass filter. This ensures that the high-frequency information in $y(t)$ is preserved. This is reflected in the fact that both controllers react with equal vigor to step changes in the reference signal y_{ref}. Even though we did not do any formal noise analysis in this section, the advantage of using a full-state observer is illustrated by the simulation results.

1.2 Literature Review

The first subject of this thesis deals with nonlinear observers on Lie groups, which have been the subject of much recent research. A series of recent papers propose observers, which comprise of a copy of the system's model, plus an "innovation" term that is based on the projection of the measurement error from the Lie group to the Lie algebra. These projection based observers work well for some Lie groups. For example, recent work on full-state observers for systems on $\mathsf{SO}(3)$, describing rigid-body rotational kinematics, was done in [1], [2]. For systems on $\mathsf{SE}(3)$, describing the pose of a rigid-body, full-state observers were proposed in [3], [4]. For systems on $\mathsf{SL}(3)$, describing a homography transformation of a moving camera which is looking at a planar image, partial-state observers were proposed in [5]. However there has not been proposed any way to extend the projection based observers to work on the general linear Lie group, $\mathsf{GL}(n, \mathbb{R})$. This appears to be a limitation of the projection based observer design.

A recent breakthrough in observer design on the general linear Lie group was achieved in [6], where exponential observers are proposed for left-invariant and right-invariant systems on arbitrary finite dimensional, connected Lie groups. The authors propose exponential observer design by using gradient-like driving terms derived from cost functions of the Lie group measurement error. Gradient-like observers, proposed in [6], overcome some of the limitations of the projection based observers, by allowing observer design for systems on $\mathsf{GL}(n, \mathbb{R})$. In Chapter 3, we propose an alternative to gradient-like observers, which also works on $\mathsf{GL}(n, \mathbb{R})$. The weakness of our result is that we only show local exponential stability of our observers and not global stability.

One important work on partial-state observers on $\mathsf{SO}(3)$ is [7], which proposes a partial-state observer, that uses measurement of the orientation and of the torque to estimate the angular velocity of the rigid-body. A slightly different rigid-body dynamic rotation model is used in [7], than what we use in Chapter 4. It is not clear to us, how the partial-state observers, proposed in [7], are related to our partial-state observers.

A vector field on a connected Lie group is said to be linear if its flow is a one-parameter subgroup, see [8]. Recent work on linear vector fields and linear systems on Lie groups was done in [8] and [9]. The estimation error dynamics of our Lie group full-state observers are shown to be linear vector fields on Lie group, because their flow is a one-parameter subgroup. We also show a stronger property of our estimation error dynamics. Namely, we show that the estimation error dynamics are differentially equivalent, by means of the logarithm map, to a linear vector field on the Lie algebra.

Interesting research on observers for systems on Lie groups was also done in [10]. This article considers a class of systems which is much more general than the left-invariant systems considered in this thesis. Namely, the article considers systems that evolve on a vector space $\mathbb{R}^n$, but are such that a certain Lie group action leaves the system equations unchanged (this property is called invariance). Although the class of systems of [10] includes the left-invariant systems on Lie groups with full-state measurement, the observer design of [10] is meant for a much more general problem. The research done in [10] is interesting, because the authors show that, if the plant is invariant under the action of some Lie group, then part of the states of the plant can be redefined as evolving on this Lie group, at least locally. This article also proposes a rigid-body attitude estimation algorithm, but the algorithm uses measurements from magnetometer and accelerometer sensors directly, instead of converting them first to a rotation matrix. So, this observer is difficult to objectively compare with our proposed observer on $\mathsf{SO}(3)$.

In all of our orientation estimation problems, we assume that the entire orientation is directly measurable. This direct measurement of the orientation can be obtained from measurement of two linearly independent vectors in the inertial and mobile frame. The assumption that the entire orientation is measured is what makes our work very different from work such as [11], where the authors assume that only one vector measurement is made, but that the vector in the inertial frame changes over time. We do not think it is possible to compare our proposed observers with the observer proposed in [11], because they are meant for solving completely different problems.

The second subject of this thesis deals with nonlinear observers on vector spaces, which is a now classical topic. Much research has been done and a large variety of nonlinear observer design techniques exist in this field of research. A comprehensive summary of recent results on nonlinear observers on vector spaces can be found in [12]. In Chapter 5, we will work with one type of nonlinear observer, called the high-gain observer. A comprehensive summary of high-gain observer results is found in [13]. High-gain observers were originally applied to fully-feedback-linearizable plants, since this type of plant can be transformed into a chain of integrators from the input to the output. A major breakthrough in high-gain observer design was achieved in [14], [15], [16] and related papers, where exponential high-gain observers are proposed for single input, single output systems, which are uniformly observable for any input.

High-gain observers are well suited for output feedback stabilization of nonlinear systems, due to their property of being approximate differentiators. A very important work in this direction is [17], where the authors show that a fully-feedback-linearizable plant, under certain assumptions, can be stabilized by output feedback, using high-gain observers as approximate differentiators of the output. Other works in this area, such as [18] and [19], extend this to non-fully-feedback-linearizable systems. They rely on defining new states as derivatives of the output and of the input, and then using the high-gain observer to estimate these derivatives. For systems which are non-fully-feedback-linearizable and also non-minimum-phase, some interesting ideas have appeared in [20], [21], which also rely on using high-gain observers as approximate differentiators. A number of other works on high-gain observers underline their main property of being approximate differentiators of smooth signals, see [13], [22], [23]. This property of high-gain observers, being approximate differentiators, is very important for our analysis, because the interconnection of several approximate differentiators is the main idea of the Composite High-Gain Observer, discussed in Chapter 5.

1.3 Contributions

The main contribution of this thesis is a local exponential full-state observer for systems on the general linear Lie group, which has linear estimation error dynamics.

Summary of practical contributions:

1. Section 3.3 proposes a an exponential full-state observer for left-invariant systems on the general linear Lie group. Our observer yields linear estimation error dynamics, which is what distinguishes it from other observers that are proposed in the literature.

2. Section 3.7 proposes an Inertial Measurement Unit (IMU) algorithm, which uses angular rate gyroscopes to filter out noise from a noisy measurement of the rigid-body orientation. This filter is based on the results of Chapter 3 and yields linear estimation error dynamics.

3. Section 4.3 proposes a new type of local exponential partial-state observer, for a class of systems that is a generalization of the left-invariant systems considered in Chapter 3. This new class of systems has only a proper subset of its states evolving on the Lie group, while the rest of the states evolve on the Lie algebra. Our proposed partial-state observer is an extension of the observer proposed in Section 3.3 to the case when only a proper subset of the states of the system is directly available for measurement. Our partial-state observer does not yield linear estimation error dynamics.

4. Section 4.7.2 proposes another Inertial Measurement Unit algorithm, which is a modification of the IMU algorithm proposed in Section 3.7. This new algorithm uses

angular accelerometers, instead of gyroscopes, to filter out noise from a noisy measurement of the rigid-body orientation and to also estimate the angular velocity of the rigid-body.

5. Section 4.7.3 proposes a partial-state observer for estimation of homography dynamics of a moving camera. Our proposed observer is used for solving the same problem as the observer proposed in [24].

6. Chapter 5 introduces a new method for building high-gain observers for nonlinear systems, by connecting several high-gain observers into a "chain" of sub-observers. The first sub-observer in the chain is observing the plant, the second sub-observer in the chain is observing the first sub-observer, as if it was a "virtual plant", and so on.

7. Section 5.5 designs a composite high-gain observer for the translational oscillator with a rotating actuator (TORA) system. The observer is not shown to be convergent, but is tested in simulation, with and without measurement noise.

Summary of theoretical contributions:

1. Section 3.5 gives the exact solution of the nonlinear differential equation $\dot{E} = E\log(E)$, with $E \in \mathsf{GL}(n, \mathbb{R})$. It also shows that this differential equation transforms into a linear differential equation on Lie algebra, given by $\dot{e} = e$, where $e := \log(E)$.

2. Corollary 3.6.1 proves that the nonlinear system on Lie group, given by $\dot{E} = [E, u]$, transforms into $\dot{e} = [e, u]$, where $e := \log(E)$.

Chapter 2

Preliminaries

2.1 General Notation

In Chapters 3 and 4, the state of the plant is one or more matrices in $\mathbb{R}^{n\times n}$. While in Chapter 5, the state of the plant is a vector in $\mathbb{R}^n$. Table 2.1 lists the abbreviated observer names used in this thesis. We keep the following notation consistent throughout the whole thesis

(i) The state of the plant is labeled by a letter without a hat, eg. x, z, X.

(ii) The measured output of the plant is labeled by Y if it is a matrix, or by y if it is a vector or a scalar.

(iii) The input to the system is labeled by u, which may be a matrix, vector, or scalar.

(iv) The state of the observer is labeled by the same letter as the state of the plant, but with a hat on top, eg. $\hat{x}$, $\hat{z}$, $\hat{X}$.

(v) For systems on Lie groups, the estimation error between the plant and the observer is labeled by E_l, E_r, e_l, or e_r. For systems on vector spaces, the estimation error is labeled by the same symbol as the state of the plant, but with a tilde symbol on top, eg. $\tilde{x}$, $\tilde{z}$.

(vi) If $x \in \mathbb{R}^n$, then x_i refers to the i^{th} component of x. When X is a matrix, then $X_{i,j}$ refers to the $(i,j)^{\text{th}}$ element of X.

Table 2.1: List of abbreviated observer names

HGO	High-Gain Observer
LFSO	Lie-group Full-State Observer
LPSO	Lie-group Partial-State Observer
CHGO	Composite High-Gain Observer

At times, we describe the asymptotic behavior of variables, as we make a small parameter approach zero. Order of magnitude notation is the standard tool used for this purpose, see [22], [25].

Definition 2.1.1. *Let x be an $\mathbb{R}^n$ valued signal, which is a function of both the time, and of a small parameter ϵ, i.e., $x : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^n$, and $x(t, \epsilon) \in \mathbb{R}^n$. We say x is of order $O(\epsilon)$ pointwise in time if*

$$(\forall t > 0)\ (\exists k, c > 0)\ (\forall \epsilon < c) \qquad \|x(t, \epsilon)\| \leq k\epsilon.$$

Definition 2.1.2. *Let x be an $\mathbb{R}^n$ valued signal, which is a function of both the time, and of a small parameter ϵ, i.e., $x : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^n$, and $x(t, \epsilon) \in \mathbb{R}^n$. We say x is of order $O(\epsilon)$ uniformly in time if*

$$(\exists k, c > 0)\ (\forall t > 0)\ (\forall \epsilon < c) \qquad \|x(t, \epsilon)\| \leq k\epsilon.$$

Notice that x being order $O(\epsilon)$ uniformly in time is a much stronger statement than x being order $O(\epsilon)$ pointwise in time. In this thesis, we will mostly be concerned with order of magnitude statements pointwise in time.

Definition 2.1.3. *Consider a smooth function $h : \mathbb{R}^n \to \mathbb{R}^m$, which is given by m real-valued component functions $h_1(x_1, \ldots, x_n)$, $\ldots$, $h_m(x_1, \ldots, x_n)$. We define the differential of h, denoted by dh, to be the $m \times n$ Jacobian matrix of the function h, i.e.,*

$$dh := \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}.$$

Definition 2.1.4. *Consider a real valued function $h : \mathbb{R}^n \to \mathbb{R}$ and a vector field $f : \mathbb{R}^n \to \mathbb{R}^n$. We define the directional or Lie derivative of h, in the direction of f, at the point $x \in \mathbb{R}^n$, as*

$$L_f h(x) := (dh(x))\,(f(x))\,.$$

The repeated k-times Lie derivative is defined as

$$L_f^k h(x) := \left(dL_f^{k-1} h(x)\right)(f(x)).$$

Definition 2.1.5. *Consider a system with state $x \in \mathbb{R}^n$ and output $y \in \mathbb{R}$*

$$\begin{aligned} \dot{x} &= f(x) \\ y &= h(x), \end{aligned}$$

where the maps $f : \mathbb{R}^n \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}$ are assumed to be smooth. We define the observability matrix as

$$\begin{bmatrix} dh(x) \\ dL_f h(x) \\ \vdots \\ dL_f^{n-1} h(x) \end{bmatrix},$$

and we say that the observability rank condition holds at $x_0 \in \mathbb{R}^n$ if the rank of this observability matrix is equal to n, at $x = x_0$, i.e.,

$$\operatorname{rank} \begin{bmatrix} dh(x_0) \\ dL_f h(x_0) \\ \vdots \\ dL_f^{n-1} h(x_0) \end{bmatrix} = n.$$

2.2 Lie Groups

In this section, we give a brief introduction to Lie groups as subgroups of matrices. The main mathematical reference is [26]. All the results in this section are well-known.

We denote by $\mathbb{R}^+$ the set of real numbers, equipped with the additive group structure. We denote by $\mathbb{R}^{n\times n}$ the set of $n \times n$ matrices with real entries. We denote by $\mathsf{GL}(n, \mathbb{R})$ the general linear Lie group of all invertible $n \times n$ matrices with real entries. We denote by $\mathsf{M}(n, \mathbb{R})$ the algebra of all $n \times n$ matrices with real entries. The bilinear product that makes $\mathsf{M}(n, \mathbb{R})$ an algebra is the matrix commutator, i.e., given $A, B \in \mathbb{R}^{n\times n}$, the product of A and B is defined as $[A, B] := AB - BA$. We denote by I_n the $n \times n$ identity matrix. We denote by 0_n the $n \times n$ zero matrix. For matrices $A \in \mathsf{M}(n, \mathbb{R})$ and $X \in \mathsf{GL}(n, \mathbb{R})$, we define the adjoint map as $\mathrm{Ad}_X(A) := XAX^{-1}$.

For a vector $x \in \mathbb{R}^n$, we let $\|x\|$ denote the Euclidean norm. For a matrix $A \in \mathbb{R}^{n\times n}$, we define the corresponding induced matrix norm as

$$\|A\| := \max\left\{\|Ax\| : x \in \mathbb{R}^n, \|x\| \leq 1\right\}.$$

Induced norms on $\mathbb{R}^{n\times n}$ are submultiplicative [27], i.e., for any A, $B \in \mathbb{R}^{n\times n}$

$$\|AB\| \leq \|A\|\|B\|.$$

Definition 2.2.1. *A linear Lie group G is a closed subgroup of $\mathsf{GL}(n, \mathbb{R})$, with matrix multiplication as the group operation.*

A linear Lie group is not the same thing as a matrix Lie group. A matrix Lie group is defined to be a subgroup of $\mathsf{GL}(n, \mathbb{R})$ (see [28, Definition 5.13], for example), but not necessarily closed as a set. The fact that linear groups are additionally restricted to be closed sets, ensures that the Lie group is an embedded submanifold of $\mathbb{R}^{n\times n}$, see [28]. Thus, we do not apply our analysis to Lie groups which are immersed submanifolds, but not embedded submanifolds of $\mathbb{R}^{n\times n}$. All the "interesting" or "practical" Lie groups are embedded submanifolds, so we do not really lose much by excluding the non-embedded submanifolds.

When we speak of Lie groups in this thesis we refer to linear Lie groups, unless otherwise specified. By working with Lie groups that are submanifolds of the space of real $n \times n$ matrices, rather than abstract manifolds, we are able to use functional calculus, i.e., Taylor series expansions of matrix functions, to prove results about our observer design.

Given a matrix $A \in \mathbb{R}^{n\times n}$ and a real scalar $r > 0$, we define the following open ball

$$B(A, r) := \left\{X \in \mathbb{R}^{n\times n} : \|A - X\| < r\right\}.$$

Proposition 2.2.2. *Let $X \in B(I_n, 1)$, then the series $\sum_{k=0}^{\infty}(-1)^k(X - I_n)^k$ converges in norm and*

$$X^{-1} = \sum_{k=0}^{\infty}(-1)^k(X - I_n)^k. \tag{2.1}$$

Proof. Let $X \in B(I_n, 1)$ and define $M := I_n - X$ so that $\|M\| < 1$. Taking the norm of the right hand side of (2.1)

$$\begin{aligned}
\left\|\sum_{k=0}^{\infty}(-1)^k(X-I_n)^k\right\| &= \left\|\sum_{k=0}^{\infty} M^k\right\| \\
&\leq \sum_{k=0}^{\infty}\|M^k\| \\
&\leq \sum_{k=0}^{\infty}\|M\|^k \\
&= (1-||M||)^{-1}.
\end{aligned}$$

Hence the right hand side of (2.1) is a convergent series. To see that its sum equals X^{-1}, since X is a square matrix, it is enough to left-multiply by X and to check that the result is the identity matrix. To this end

$$\begin{aligned}
X\sum_{k=0}^{\infty}(-1)^k(X-I_n)^k &= (I_n-M)\sum_{k=0}^{\infty} M^k \\
&= \sum_{k=0}^{\infty}\left(M^k-M^{k+1}\right) \\
&= I_n+\sum_{k=1}^{\infty}\left(M^k\right)-\sum_{k=0}^{\infty}\left(M^{k+1}\right) \\
&= I_n+M-M+M^2-M^2+\cdots \\
&= I_n.
\end{aligned}$$

$\square$

A consequence of Proposition 2.2.2 is that the set $B(I_n, 1)$ is contained in $\mathsf{GL}\,(n,\mathbb{R})$.

Definition 2.2.3. *Given a matrix* $A \in \mathsf{M}\,(n,\mathbb{R})$*, the matrix exponential* $\exp : \mathsf{M}\,(n,\mathbb{R}) \to \mathsf{GL}\,(n,\mathbb{R})$ *is defined to be*

$$\exp(A) := \sum_{k=0}^{\infty}\frac{A^k}{k!}. \tag{2.2}$$

Since $||A^k|| \leq ||A||^k$, the series (2.2) converges in norm for every matrix $A \in \mathbb{R}^{n\times n}$.

Definition 2.2.4. *Given a linear Lie group G, we define the Lie algebra of G, denoted by* $\mathrm{Lie}(G)$*, to be the set*

$$\mathrm{Lie}(G) := \{A \in \mathsf{M}(n,\mathbb{R}) : \quad \forall t \in \mathbb{R}, \exp(tA) \in G\}.$$

It is shown in [26, Theorem 3.2.1] that $\mathrm{Lie}(G)$ is a subalgebra of $\mathsf{M}(n,\mathbb{R})$.

Definition 2.2.5. *Given a matrix $X \in B(I_n, 1)$, the matrix logarithm map,* $\log : B(I_n, 1) \to \mathsf{M}(n,\mathbb{R})$*, is defined as*

$$\log(X) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}(X - I_n)^k. \tag{2.3}$$

The above series converges in norm for every matrix $X \in B(I_n, 1) \subset \mathsf{GL}(n,\mathbb{R})$ because $\sum_k \frac{1}{k}\|X - I_n\|^k$ converges for $\|X - I_n\| < 1$.

The series definition of the matrix logarithm map (2.3) has a relatively small domain of convergence. It is possible to extend the matrix logarithm map to a much larger domain. By using Gregory's series (1668), we obtain the following series definition [29, Section 11.3]

$$\log(X) = -2\sum_{k=0}^{\infty} \frac{1}{2k+1}\left((I_n - X)(I_n + X)^{-1}\right)^{2k+1},$$

which converges if all the eigenvalues of X have strictly positive real parts. In this thesis, for clarity, we only use definition (2.3). Regardless of which series is used, the proposed observers remain the same, only the proven region of convergence may change.

Lemma 2.2.6. *The exponential and the logarithm maps have the following properties*

(a) $\forall X \in B(I_n, 1) : \quad \exp(\log(X)) = X$,

(b) $\forall A \in B(0_n, \log(2)) : \quad \log(\exp(A)) = A$,

(c) $\forall A \in \mathsf{M}(n,\mathbb{R})$*, and* $\forall X \in \mathsf{GL}(n,\mathbb{R})$,

$$\exp(XAX^{-1}) = X\exp(A)X^{-1},$$

(d) $\forall X \in B(I_n, 1)$*, and* $\forall A \in \mathsf{GL}(n,\mathbb{R})$ *such that* $AXA^{-1} \in B(I_n, 1)$,

$$\log(AXA^{-1}) = A\log(X)A^{-1}.$$

Proof. The proof of properties (a) and (b) is standard and can be found in [26, Theorem 2.2.1] or [30].

To see that (c) holds, note that $(XAX^{-1})^k = XA^kX^{-1}$. Substituting this identity into the series definition (2.2) of the exponential map gives the desired result.

To see that (d) holds, note that $(AXA^{-1} - I_n)^k = (A(X - I_n)A^{-1})^k = A(X - I_n)^k A^{-1}$. Substituting this identity into the series definition (2.3) of the logarithm map gives the desired result. ☐

A simple consequence of Lemma 2.2.6, part (d), is that on the set $B(I_n, 1)$, the logarithm of a matrix commutes with the matrix itself, i.e., for all $X \in B(I_n, 1)$,

$$X \log(X) = X \log(X) X^{-1} X = \log(X) X. \tag{2.4}$$

This elementary property of the logarithm map will play an important role in our later analysis.

Definition 2.2.7. *Let G be a linear Lie group. A one-parameter subgroup of G is defined to be a continuous group homomorphism $\gamma : \mathbb{R}^+ \to G$.*

Lemma 2.2.8 ([26, Theorem 3.1.1])**.** *If $\gamma : \mathbb{R}^+ \to \mathsf{GL}(n, \mathbb{R})$ is a one-parameter subgroup of $\mathsf{GL}(n, \mathbb{R})$, then γ is real-analytic and $\gamma(t) = \exp(tA)$, with $A = \gamma'(0)$.*

Lemma 2.2.9. *Let $X : \mathbb{R} \to \mathsf{GL}(n, \mathbb{R})$ be a smooth parameterized curve. Then*

$$\frac{d}{d\tau} X^{-1} = -X^{-1} \frac{dX}{d\tau} X^{-1}. \tag{2.5}$$

Proof. For any smooth parameterized curve $X : \mathbb{R} \to \mathsf{GL}(n, \mathbb{R})$, and any $\tau \in \mathbb{R}$, $X(\tau)X^{-1}(\tau) = I_n$ where τ is the curve parameter. Differentiating both sides of this identity with respect to τ, while using the product rule, we get

$$\begin{aligned} 0 = \frac{d}{d\tau} I_n &= \frac{d}{d\tau}\left(XX^{-1}\right) \\ &= \frac{dX}{d\tau} X^{-1} + X \frac{d}{d\tau} X^{-1}, \end{aligned}$$

from which (2.5) immediately follows. ☐

Definition 2.2.10. *Given a smooth parameterized curve $X : \mathbb{R} \to \mathsf{GL}(n, \mathbb{R})$, the body-velocity of X is the curve $v_X : \mathbb{R} \to \mathsf{M}(n, \mathbb{R})$, defined by $v_X(t) := X^{-1}(t)\dot{X}(t)$.*

In addition to $\mathsf{GL}(n,\mathbb{R})$, the following linear Lie groups are used in this thesis

$$\begin{aligned}\mathsf{SL}(n,\mathbb{R}) &:= \{X \in \mathbb{R}^{n\times n} : \det(X) = 1\}\\ \mathsf{SO}(n,\mathbb{R}) &:= \{X \in \mathbb{R}^{n\times n} : XX^\top = I_n, \det(X) = 1\}.\end{aligned}$$

The corresponding Lie algebras are

$$\begin{aligned}\mathrm{Lie}(\mathsf{GL}(n,\mathbb{R})) &:= \mathsf{M}(n,\mathbb{R})\\ \mathrm{Lie}(\mathsf{SL}(n,\mathbb{R})) &:= \{A \in \mathsf{M}(n,\mathbb{R}) : \mathrm{tr}(A) = 0\}\\ \mathrm{Lie}(\mathsf{SO}(n,\mathbb{R})) &:= \{A \in \mathsf{M}(n,\mathbb{R}) : A + A^\top = 0_n\}.\end{aligned}$$

If $G \subseteq \mathsf{GL}(n,\mathbb{R})$ is any linear Lie group, then as a consequence of Lemma 2.2.6, part (a), and of the definition of $\mathrm{Lie}(G)$, the matrix logarithm map takes elements of $G \cap B(I_n, 1)$ to elements of $\mathrm{Lie}(G)$, i.e., $\log : G \cap B(I_n, 1) \to \mathrm{Lie}(G)$.

Definition 2.2.11. *If G is a linear Lie group and $X \in G$ is any element of G, then the tangent space to G at X is defined as*

$$T_X G := \{A \in \mathsf{M}(n,\mathbb{R}) : A = \dot\gamma(0), \gamma : (-\delta,\delta) \subset \mathbb{R} \to G \textit{ is smooth and } \gamma(0) = X\}.$$

Proposition 2.2.12. *For $X \in G$, we have*

$$\begin{aligned}T_X G &= X\,\mathrm{Lie}(G) = \{XA : A \in \mathrm{Lie}(G)\}\\ &= \mathrm{Lie}(G)X = \{AX : A \in \mathrm{Lie}(G)\}.\end{aligned}$$

Proof. By Lemma 2.2.6, part (a), we have that

$$(\forall X \in G)\ \ (\forall A \in \mathrm{Lie}(G))\ \ (\forall t \in \mathbb{R}) \quad X\exp(tA)X^{-1} = \exp(tXAX^{-1}).$$

Thus $X\,\mathrm{Lie}(G)X^{-1} \subseteq \mathrm{Lie}(G)$. Using an identical argument, replacing X with X^{-1}, one can verify that $X^{-1}\,\mathrm{Lie}(G)X \subseteq \mathrm{Lie}(G)$, which implies $\mathrm{Lie}(G) \subseteq X\,\mathrm{Lie}(G)X^{-1}$. Therefore, we have shown that $X\,\mathrm{Lie}(G)X^{-1} = \mathrm{Lie}(G)$, which proves that $X\,\mathrm{Lie}(G) = \mathrm{Lie}(G)X$.

Now, for any $A \in \mathrm{Lie}(G)$, the curve $\gamma(t) = X\exp(tA)$ is smooth with $\gamma(0) = X$, hence $\dot\gamma(0) \in T_X G$. Computing the derivative of the curve, we get

$$\dot\gamma(0) = \frac{d}{dt}X\exp(tA)\Big|_{t=0} = XA,$$

which shows that $XA \in T_X G$. Since our choice of $A \in \mathrm{Lie}(G)$ was arbitrary, we have that $X\,\mathrm{Lie}(G) \subseteq T_X G$.

Conversely, for any $B \in T_X G$, there is a smooth curve γ_B, with $\gamma_B(0) = X$ and $\dot{\gamma}_B(0) = B$. For small $|t|$, let us define

$$\beta(t) := \log(X^{-1}\gamma_B(t)),$$

which is a curve in $\mathrm{Lie}(G)$, and is such that $\beta(0) = 0$. Furthermore, since $\mathrm{Lie}(G)$ is a vector space, we have that $\dot{\beta}(0) \in T_{\beta(0)} \mathrm{Lie}(G) \cong \mathrm{Lie}(G)$. For small $|t|$, the curve γ_B is given by

$$\gamma_B(t) = X \exp(\beta(t)).$$

Computing $\dot{\gamma}_B(0)$, and using the fact that $\beta(0) = 0$, we get

$$\begin{aligned}
\dot{\gamma}_B(0) &= X \frac{d}{dt} \left[\sum_{k=0}^{\infty} \frac{\beta(t)^k}{k!} \right] \Bigg|_{t=0} \\
&= X \frac{d}{dt} \left[I_n + \beta(t) + \frac{\beta(t)^2}{2} + \cdots \right] \Bigg|_{t=0} \\
&= X \left[\dot{\beta}(t) + \frac{\dot{\beta}(t)\beta(t) + \beta(t)\dot{\beta}(t)}{2} + \cdots \right] \Bigg|_{t=0} \\
&= X\dot{\beta}(0).
\end{aligned}$$

Thus, we have $B = \dot{\gamma}_B(0) = X\dot{\beta}(0)$, but we have seen before that $\dot{\beta}(0) \in \mathrm{Lie}(G)$, therefore $B \in X\,\mathrm{Lie}(G)$. Since our choice of $B \in T_X G$ was arbitrary, we have that $T_X G \subseteq X\,\mathrm{Lie}(G)$. □

2.3 High-Gain Observer

Consider the following nonlinear, single-input, single-output system on $\mathbb{R}^n$

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_n &= \phi(x, u) \\
y &= x_1,
\end{aligned} \tag{2.6}$$

where y is the measured output and $\phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a smooth, globally Lipschitz and uniformly bounded function.

The high-gain observer (HGO) allows us to estimate all the states of system (2.6), by using the signal y and possibly also u. One useful feature of the HGO is that it can estimate the states of the system (2.6) without full knowledge of the function $\phi(x,u)$, i.e., HGO is robust to modeling uncertainties. If we do not know the function $\phi(x,u)$ exactly, but only know some nominal model of this function, call it $\phi_0(x,u)$, then we can build the following HGO:

$$\begin{aligned}
\dot{\hat{x}}_1 &= \hat{x}_2 + \frac{1}{\epsilon} a_{n-1}(y - \hat{x}_1) \\
\dot{\hat{x}}_2 &= \hat{x}_3 + \frac{1}{\epsilon^2} a_{n-2}(y - \hat{x}_1) \\
&\vdots \\
\dot{\hat{x}}_{n-1} &= \hat{x}_n + \frac{1}{\epsilon^{n-1}} a_1(y - \hat{x}_1) \\
\dot{\hat{x}}_n &= \phi_0(\hat{x}, u) + \frac{1}{\epsilon^n} a_0(y - \hat{x}_1),
\end{aligned} \tag{2.7}$$

where the constants $a_0, \ldots, a_{n-1}$ are design parameters of the HGO, chosen so that the polynomial $p(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$ is Hurwitz and the constant $\epsilon > 0$ is a small design parameter.

The estimation error between (2.6) and (2.7)

$$\tilde{x} := x - \hat{x} = \begin{bmatrix} x_1 - \hat{x}_1, & x_2 - \hat{x}_2, & \ldots, & x_n - \hat{x}_n \end{bmatrix}^\top$$

satisfies the dynamics

$$\dot{\tilde{x}} = \begin{bmatrix} -\frac{1}{\epsilon}a_{n-1} & 1 & 0 & \cdots & 0 \\ -\frac{1}{\epsilon^2}a_{n-2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{\epsilon^{n-1}}a_1 & 0 & 0 & \cdots & 1 \\ -\frac{1}{\epsilon^n}a_0 & 0 & 0 & \cdots & 0 \end{bmatrix} \tilde{x} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \delta(x, \hat{x}, u), \tag{2.8}$$

where $\delta(x, \hat{x}, u) = \phi(x,u) - \phi_0(\hat{x}, u)$ is the perturbation, which arises due to modeling uncertainty ($\phi_0 \neq \phi$) and due to estimation error ($\hat{x} \neq x$). The system (2.8) can be viewed as a linear system, forced by a nonlinear term $\delta(x, \hat{x}, u)$. The unforced, linear system has the characteristic polynomial $q(s) = s^n + \frac{a_{n-1}}{\epsilon}s^{n-1} + \cdots + \frac{a_1}{\epsilon^{n-1}}s + \frac{a_0}{\epsilon^n}$. For any $\epsilon > 0$, the polynomial $q(s)$ is Hurwitz if and only if the polynomial $p(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$, is Hurwitz. Indeed, if $p(s)$ factors as

$$p(s) = (s - r_1) \cdots (s - r_n),$$

where $\{r_1, \dots, r_n\}$ is the set of repeating roots of $p(s)$, then it is easy to check that $q(s)$ factors as

$$q(s) = \left(s - \frac{r_1}{\epsilon}\right) \cdots \left(s - \frac{r_n}{\epsilon}\right).$$

Since we chose $a_0, \dots, a_{n-1}$ such that $p(s)$ is Hurwitz, we must have that all the roots of the characteristic polynomial, $q(s)$, have strictly negative real parts. Therefore all the eigenvalues of the unforced, linear part of (2.8) also have negative real parts. Therefore the unforced, linear system is exponentially stable, by Lemma A.3.1, if the coefficients $a_0, \dots, a_{n-1}$ are chosen as discussed above.

In the absence of modeling uncertainty, i.e., when $\phi_0 = \phi$, the perturbation term $\delta(x, \hat{x}, u)$ is a vanishing perturbation, because it equals 0 when $\tilde{x} = 0$, i.e., for any $x \in \mathbb{R}^n$, $\delta(x, x, u) = 0$. In this case, the perturbed error system (2.8) has an equilibrium point at $\hat{x} = 0$, which is exponentially stable for ϵ sufficiently close to zero [22].

In the presence of modeling uncertainty, i.e., $\phi_0 \neq \phi$, the perturbation term $\delta(x, \hat{x}, u)$ is non-vanishing at $\tilde{x} = 0$. With modeling uncertainty, the perturbed error system (2.8) does not have an equilibrium point at $\tilde{x} = 0$. Thus, we can not talk about the stability of $\tilde{x} = 0$ and the HGO (2.7) can not converge to the system states. However, using some perturbation theory analysis, it is shown in [17] that the estimation error $\tilde{x}(t)$ decays to order $O(\epsilon)$ values after a short transient period. This transient period, also called the "peaking period", has time duration $[0, T_1(\epsilon)]$ where $T_1(\epsilon)$ tends to zero as ϵ tends to zero. During the "peaking period", the estimation error may temporarily peak to order $O(1/\epsilon)$ values, before decaying towards zero. Thus, by choosing ϵ sufficiently small, the state estimates can be made arbitrarily accurate, at the expense of increased peaking in the estimation error, see also [22], [23], [31].

When noise is present in the measurement y, decreasing ϵ causes amplification of this measurement noise. So there will be a practical limit to how small we can make ϵ without the observer becoming useless. If the noise is low-frequency with bounded time derivatives, then the HGO can still work fairly well [23], [32].

HGOs were successfully used in [17] to stabilize fully feedback linearizable systems. Much research has been done on the use of HGO in feedback control of nonlinear systems, for a comprehensive summary, see [13].

2.4 High-Gain Observers as Differentiators

Consider the following special case of the high-gain observer (2.7), with $n = 2$ and $\phi_0(x, u) = 0$,

$$\begin{aligned}\dot{\hat{x}}_1 &= \hat{x}_2 + \frac{1}{\epsilon}a_1(y - \hat{x}_1)\\ \dot{\hat{x}}_2 &= \frac{1}{\epsilon^2}a_0(y - \hat{x}_1).\end{aligned} \tag{2.9}$$

The high-gain observer (2.9) is a linear system, computing the transfer function from y to $\hat{x}_2$, we get

$$\frac{\hat{X}_2(s)}{Y(s)} = \frac{sa_0}{s^2\epsilon^2 + s\epsilon a_1 + a_0}.$$

In the limit, as $\epsilon \to 0$, we get $\frac{\hat{X}_2(s)}{Y(s)} \to s$. Thus, we see that the observer (2.9) approaches the ideal differentiator, shown in Figure 2.1.

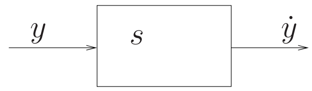

Figure 2.1: The ideal differentiator, a non-causal system

Thus, the high-gain observer should be viewed as an approximate differentiator of the signal $y(t)$. This viewpoint is discussed in [13], [22].

Chapter 3

Lie Group Full-State Observers

This chapter proposes a local exponential observer for left-invariant systems on linear Lie groups, where the full state of the system is available for measurement. We propose two different observer designs for this class of system. We show that, depending on the observer chosen, local exponential stability of one of the two estimation error dynamics, left-invariant, or right-invariant error dynamics is obtained. These results are developed by showing that the estimation error dynamics are differentially equivalent to a stable, linear differential equation on a vector space.

We illustrate our observer design on the attitude estimation problem on $\mathsf{SO}(3)$, which has direct application to mobile robotics. Where the orientation of the mobile robot and its angular velocities are measured by means of on-board sensors.

3.1 Introduction

Nonlinear, full-state observers for left-invariant systems on Lie groups find many applications in mobile robotics. One well-developed application is the estimation of the orientation of a rigid-body using low-cost on-board sensors. A very effective filter for this purpose was proposed in [2]. Small autonomous robots usually undergo a lot of vibration and other disturbances, while being restricted to carrying only a basic, light-weight sensor package. For this reason, a lot of high-frequency noise is usually present in the sensor measurements of these robots. The noise has a detrimental effect on the robot's performance. Nonlinear observers for left-invariant systems on Lie groups are useful because, in certain cases, they can be used to filter out the sensor noise. Depending on the specific application and the type of sensor package available, it could be possible to use the nonlinear observer as a sensor fusion algorithm that eliminates a lot of the sensor noise.

Left-invariant systems on Lie groups arise in kinematic modeling of rigid-body rotation. The orientation of a rigid-body can be represented by an orthogonal matrix with unit magnitude, $R \in \mathsf{SO}(3)$. As the rigid-body undergoes rotation, the matrix R changes with time, according to the following left-invariant equation on $\mathsf{SO}(3)$,

$$\dot{R} = Ru,$$

where $u \in \text{Lie}(\mathsf{SO}(3))$ is a skew-symmetric matrix, that encodes the angular velocity of the rigid-body. The idea is that R and u can be measured by low-cost sensors, which are mounted on the rigid-body. We will return to this kinematic rigid-body rotation model in Section 3.7.

In this chapter we consider left-invariant systems on the general linear Lie group, i.e., the group of all non-singular real $n \times n$ matrices. In other words we consider systems whose state is an invertible, square matrix with real coefficients and whose vector field is left-invariant. The output of the system is taken to be the entire state matrix, i.e., we require that the full state of the system be available for measurement. We choose to mostly talk about left-invariant systems, to avoid frequent repetition of the same essential concept, however it is possible to extend analogous results to right-invariant systems on Lie groups.

At first glance it may seem unnecessary to design an observer to estimate the state of a system if that state is already measurable. We are motivated to study this problem for two reasons. First such an observer provides noise filtering and can be used as a sensor fusion algorithm, as we have seen in Section 1.1. The second reason is that by looking at this simple case we hope to gain insight into the state estimation problem when the output is not equal to the entire state, which we will explore in detail in Chapter 4.

When analyzing the convergence properties of our observers, we treat the plant as a deterministic process and the measurements as perfect signals, uncorrupted by noise. In this mathematical framework, the noise filtering properties of a full-state observer are not apparent. For this reason, we do not formally analyze the effects of measurement noise in this thesis, but rather show that the estimation error converges to zero if the observer is given noiseless measurements. Insight into the behavior of our observer under noise is tested in simulation.

Many current algorithms for state estimation on Lie groups, such as those proposed in [1], [2], [3] , [4] and [5], rely on projecting the measurement error from the Lie group to its corresponding Lie algebra. The projected vector in the Lie algebra is then used to drive the observer to converge to the system trajectory. This projection based approach is nice because the projection operator is numerically simple to implement. However this approach has a drawback in that it does not work for the general linear Lie group, $\mathsf{GL}(n, \mathbb{R})$. Since the Lie algebra of $\mathsf{GL}(n, \mathbb{R})$ is $\mathsf{M}(n, \mathbb{R})$, it is not clear how to project vectors from $\mathsf{GL}(n, \mathbb{R})$ to $\mathsf{M}(n, \mathbb{R})$.

Our proposed observer is similar in form to the above mentioned observers, but instead of using a projection operator to convert the measurement error from the Lie group to the Lie algebra, we use the matrix logarithm map. The advantage of our observer is that it works for the general linear Lie group, $\mathsf{GL}(n, \mathbb{R})$, and our observer also yields estimation error dynamics that are differentially equivalent to a linear and stable vector field. The matrix logarithm map transforms the Lie group error dynamics, in a neighbourhood of the identity matrix, into linear dynamics, defined on the Lie algebra.

Our full-state observers are meant to solve the same type of problem as the observers proposed in [6]. One possible advantage of our observer design, as compared to [6] is that we give an explicit observer equation. Another possible advantage of our observer design is that it yields linear estimation error dynamics. The major disadvantage of our analysis is that we do not show global convergence.

3.2 Problem Statement

Let $G \subseteq \mathsf{GL}(n, \mathbb{R})$ be any linear Lie group. Consider the following system on G

$$\begin{aligned} \dot{X} &= Xu \\ Y &= X, \end{aligned} \tag{3.1}$$

where $u : \mathbb{R} \to \mathrm{Lie}(G)$ is the control input to the system, and $Y \in G$ is the measured output of the system.

The system (3.1) is left-invariant. This means, for any constant matrix $A \in G$, if we define a new variable $Z := AX$, then Z will satisfy the same differential equation as X. Computing the dynamics of Z, we get

$$\begin{aligned} \dot{Z} &= \frac{d}{dt} AX \\ &= A\dot{X} \\ &= AXu \\ &= Zu, \end{aligned}$$

which is the same as the dynamics $\dot{X} = Xu$ in (3.1). See [10] for more information on invariant systems.

In this thesis we assume that the input signal u is admissible for the system (3.1). This means, for any initial condition $X(0) \in G$, the corresponding solution of (3.1) with the admissible input signal u is unique, continuously differentiable and exists for all time.

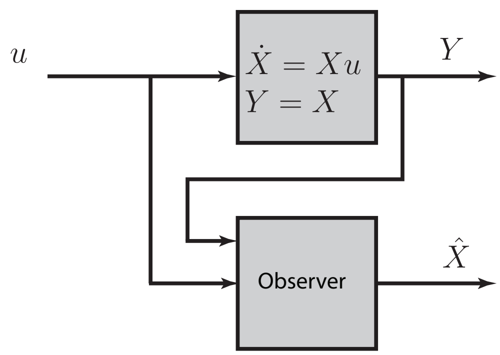

Figure 3.1: Block diagram of the system and observer setup.

Assumption 1. *The input u to system (3.1) is such that for any initial condition $X(0) \in G$, there exists a compact set $\mathcal{G} \subseteq G$, such that $X(t) \in \mathcal{G}$, for all $t \geq 0$.*

Note that this assumption is automatically satisfied if the group G is compact, for example $G = \mathsf{SO}(3)$. This means that when working with compact Lie groups, such as $\mathsf{SO}(3)$, we do not have to impose uniform boundedness of the input signal, u.

Our objective is to build an observer, having state $\hat{X} \in G$, for the system (3.1). The observer has access to the measured output $Y \in G$ and to the input $u \in \mathrm{Lie}(G)$. We will design the observer such that, if Assumption 1 holds, then for $\hat{X}(0)$ sufficiently close to $X(0)$, we have $\hat{X}(t) \to X(t)$ exponentially as $t \to \infty$.

Results of this chapter can be extended to right-invariant systems on Lie groups, which have the form

$$\dot{X} = uX.$$

However, we choose to only talk about left-invariant systems to avoid repetition and for clarity.

3.3 Proposed Observers

For the left-invariant system (3.1), we propose two different observers, which we call LFSOs (Lie-group Full-State Observers). The first is the passive LFSO, given by

$$\dot{\hat{X}} = \hat{X}u - a_0\hat{X}\log(Y^{-1}\hat{X}). \tag{3.2}$$

The second is the direct LFSO, given by

$$\dot{\hat{X}} = YuY^{-1}\hat{X} - a_0\hat{X}\log(Y^{-1}\hat{X}). \tag{3.3}$$

In the above two observers, the constant $a_0 > 0$ is a design parameter that, as we will show, can be used to change the rate of observer convergence.

Following the terminology of [6], we call the term $\alpha(\hat{X}, Y) := -a_0\hat{X}\log(Y^{-1}\hat{X})$, appearing in (3.2) and (3.3), the innovation term of the observer. It can be checked that this term α satisfies the conditions for being an innovation term, given in [6, Definition 15].

An intuitive and informal explanation for taking this particular form of $\alpha(\hat{X}, Y) = -a_0\hat{X}\log(Y^{-1}\hat{X})$ is that the matrix $Y^{-1}\hat{X}$ represents a "measurement error" on the Lie group G. The Lie group G is not a vector space and therefore $Y^{-1}\hat{X}$ is not a vector, so it should not be added or subtracted with other matrices. To make the "group-like" matrix $Y^{-1}\hat{X}$ more "vector-like", we take its logarithm, which maps it to the vector space $\mathrm{Lie}(G)$, while preserving all its information (since $\log : G \to \mathrm{Lie}(G)$ is a local diffeomorphism). We then push-forward this vector, $\log(Y^{-1}\hat{X})$, from $T_{I_n}G$ to $T_{\hat{X}}G$, the tangent space at $\hat{X}$, by applying the push-forward map of left translation. By Proposition 2.2.12, left translation from $T_{I_n}G$ to $T_{\hat{X}}G$ is the same as left multiplication by $\hat{X}$. The result of the left translation is a vector in $T_{\hat{X}}G$, which is a vector-like estimation error in the tangent space at $\hat{X}$. This vector-like estimation error is then multiplied by the gain $a_0 > 0$ to adjust the rate of observer convergence.

We call the term $\hat{X}u$ appearing in (3.2) and the term $YuY^{-1}\hat{X}$ appearing in (3.3), the synchronization terms of the observer. The choice of synchronization term is what distinguishes the passive observer from the direct observer.

3.4 Estimation Error Functions

An estimation error function, as used in this thesis, is a map from $G \times G$ to some error space. In this thesis, the error space will be either the Lie group G, or its Lie algebra $\mathrm{Lie}(G)$.

In [6] it is shown that two canonical choices of Lie group estimation error functions are available, when studying observers for left-invariant systems on Lie groups.

Definition 3.4.1. *Given a system on a linear Lie group G, with state $X \in G$, and an observer for this system, with state $\hat{X} \in G$, the canonical left-invariant error, $E_l : G \times G \to G$, is*

$$E_l(X, \hat{X}) := X^{-1}\hat{X} \tag{3.4}$$

and the canonical right-invariant error, $E_r : G \times G \to G$, is

$$E_r(X, \hat{X}) := \hat{X}X^{-1}. \tag{3.5}$$

The reason why E_l (respectively, E_r) is called left-invariant (right-invariant) is because it does not change if both of its arguments are multiplied on the left (on the right) by an arbitrary element of G,

$$\begin{aligned} (\forall A, X, \hat{X} \in G) \qquad & E_l(AX, A\hat{X}) = X^{-1}A^{-1}A\hat{X} = X^{-1}\hat{X} = E_l(X, \hat{X}) \\ (\forall A, X, \hat{X} \in G) \qquad & E_r(XA, \hat{X}A) = \hat{X}AA^{-1}X^{-1} = \hat{X}X^{-1} = E_r(X, \hat{X}). \end{aligned}$$

It is apparent from the definitions of E_l and E_r that they are related by

$$(\forall X, \hat{X} \in G) \qquad E_r = XE_lX^{-1} = \mathrm{Ad}_X(E_l),$$

where $\mathrm{Ad}_X : G \to G$ is a global diffeomorphism for any fixed $X \in G$, because it is smooth and its inverse, $\mathrm{Ad}_{X^{-1}}$, is also smooth. The canonical estimation errors always satisfy $E_l = I_n \iff E_r = I_n$. However in general, convergence of E_l to I_n is not sufficient to deduce convergence of E_r to I_n and vice-versa. This is because the diffeomorphism Ad_X changes with time as X changes with time. So, if X grows unbounded very fast, then the diffeomorphism Ad_X might "stretch" the coordinates $E_r = \mathrm{Ad}_X(E_l)$ faster than E_l is converging to I_n, thus preventing E_r from converting to I_n. To eliminate such pathological system trajectories, it is sufficient to impose Assumption 1.

The goal of our observer design is to have $\|\hat{X} - X\| \to 0$ exponentially. To this end, we have the following result.

Proposition 3.4.2. *Suppose that Assumption 1 holds. If either $E_r \to I_n$ exponentially, or if $E_l \to I_n$ exponentially as $t \to \infty$, then $\hat{X} \to X$ exponentially, as $t \to \infty$.*

Proof. For any $X, \hat{X} \in G$ the following identities hold

$$\begin{aligned} \hat{X} - X &= X(X^{-1}\hat{X} - I_n) = X(E_l - I_n) \\ \hat{X} - X &= (\hat{X}X^{-1} - I_n)X = (E_r - I_n)X. \end{aligned}$$

Taking the norms of the above identities, we get the following inequalities

$$\begin{aligned} \|\hat{X} - X\| &\leq \|X\|\|E_l - I_n\| \\ \|\hat{X} - X\| &\leq \|X\|\|E_r - I_n\|. \end{aligned} \tag{3.6}$$

In addition, we also have that, for any $X, \hat{X} \in G$ the following identities hold

$$\begin{aligned} (E_l - I_n) &= X^{-1}(\hat{X} - X) \\ (E_r - I_n) &= (\hat{X} - X)X^{-1}. \end{aligned}$$

Taking the norms of the above identities, we get the following inequalities

$$\begin{aligned} \|E_l - I_n\| &\leq \|X^{-1}\|\|\hat{X} - X\| \\ \|E_r - I_n\| &\leq \|X^{-1}\|\|\hat{X} - X\|. \end{aligned} \tag{3.7}$$

By Assumption 1, there exists a compact subset $\mathcal{G} \subset \mathsf{GL}(n, \mathbb{R})$, such that $X(t) \in \mathcal{G}$ for all $t \geq 0$. This implies that $\|X(t)\|$ is uniformly bounded, i.e., $\sup_{t\geq 0} \|X(t)\| = K_1 < \infty$. Furthermore, since the matrix inverse map is continuous, the image of $\mathcal{G}$ under the matrix inverse map is also a compact subset of $\mathsf{GL}(n, \mathbb{R})$. Therefore, $\|X^{-1}(t)\|$ is also uniformly bounded, i.e., $\sup_{t\geq 0} \|X^{-1}(t)\| = K_2 < \infty$.

From (3.6) and (3.7), we then easily see that the exponential convergence of $E_r \to I_n$, or of $E_l \to I_n$, implies the exponential convergence of $\hat{X} \to X$. Indeed, suppose that $\|E_r(t) - I_n\| \to 0$ exponentially, as $t \to \infty$, then by definition of exponential stability, we have

$$\begin{gathered} (\exists \delta > 0)\ \ (\exists m, \alpha > 0)\ \ (\forall E_r(0) \in B(I_n, \delta)) \Rightarrow \\ (\forall t \geq 0)\ \ \|E_r(t) - I_n\| < me^{-\alpha t}\|E_r(0) - I_n\| \end{gathered}$$

By the inequalities (3.6), and uniform boundedness of $\|X\|$, we have that

$$\|E_r - I_n\| < m \Rightarrow \|\hat{X} - X\| < K_1 m.$$

By the inequalities (3.7), and uniform boundedness of $\|X^{-1}\|$, we have that

$$\|\hat{X} - X\| < \frac{\delta}{K_2} \Rightarrow \|E_r - I_n\| < \delta.$$

Combining the above results, we have exponential convergence of $\|\hat{X} - X\| \to 0$,

$$\begin{gathered} (\exists \delta, m, \alpha > 0) \left(\|\hat{X}(0) - X(0)\| < \frac{\delta}{K_2} \right) \Rightarrow \\ (\forall t \geq 0)\ \ \|\hat{X}(t) - X(t)\| < K_1 K_2 m e^{-\alpha t}\|\hat{X}(0) - X(0)\| \end{gathered}$$

The proof for E_l is identical. $\square$

In addition to using the error functions E_l and E_r, we introduce two other, closely related error functions.

Definition 3.4.3. *For $E_l \in B(I_n, 1)$, we define the log left-invariant error $e_l : G \times G \to \mathrm{Lie}(G)$ to be*

$$e_l(X, \hat{X}) := \log(E_l(X, \hat{X})) = \log(X^{-1}\hat{X}) \tag{3.8}$$

and for $E_r \in B(I_n, 1)$, we define the log right-invariant error $e_r : G \times G \to \mathrm{Lie}(G)$ to be

$$e_r(X, \hat{X}) := \log(E_r(X, \hat{X})) = \log(\hat{X}X^{-1}). \tag{3.9}$$

Since e_l is a function of only E_l, and since E_l is left-invariant, it follows that e_l is also left-invariant, i.e., $\forall A \in G : \quad e_l(AX, A\hat{X}) = e_l(X, \hat{X})$. For the same reason, it follows that e_r is right-invariant, i.e., $\forall A \in G : \quad e_r(XA, \hat{X}A) = e_r(X, \hat{X})$.

The variables e_l and e_r are useful because they are vectors in $\mathrm{Lie}(G)$ (which is a real vector space), so they allow us to convert a differential equation on a Lie group into a differential equation on a vector space. The disadvantage of e_l and e_r is that they are only defined for $E_l, E_r \in B(I_n, 1)$, which means that we can only use e_l and e_r when the error between the system and the observer is not too great.

Lemma 3.4.4. *If $E_l, E_r \in B(I_n, 1)$, then*

$$e_r = X e_l X^{-1}.$$

Proof. By direct computation and with the help of Lemma 2.2.6 (d), we obtain

$$\begin{aligned} e_r &= \log(E_r) \\ &= \log(X E_l X^{-1}) \\ &= X \log(E_l) X^{-1} \\ &= X e_l X^{-1}. \end{aligned}$$

□

In this chapter, when we write one of the four possible estimation error functions, i.e., one of E_l, E_r, e_l, or e_r, we always use it to denote the error between the system (3.1) and one of the two observers (3.2) or (3.3). Whether we are referring to observer (3.2) or (3.3) will be made obvious from the context.

3.5 A Differential Equation on $\mathsf{GL}(n, \mathbb{R})$

Here we restrict our attention to examining the behaviour of the following differential equation, evolving on $\mathsf{GL}(n, \mathbb{R})$

$$\dot{E} = -a_0 E \log(E), \tag{3.10}$$

where $E \in \mathsf{GL}(n, \mathbb{R})$ and $a_0 \in \mathbb{R}$ is a positive constant. The differential equation (3.10) arises in the analysis of the error dynamics associated with the observers (3.2), (3.3). Note that the above equation can also be written as $\dot{E} = -a_0 \log(E)E$, due to identity (2.4).

Since we have only defined the log map on the domain $B(I_n, 1)$, the differential equation (3.10) is only defined for $E \in B(I_n, 1)$. We will see in Proposition 3.5.3, that $E(t)$ will stay in $B(I_n, 1)$ for $t \geq 0$, as long as $E(0)$ is initialized sufficiently close to I_n.

We defined (3.10) to evolve on the set of all invertible matrices, $\mathsf{GL}(n, \mathbb{R})$. If $G \subseteq \mathsf{GL}(n, \mathbb{R})$ is any linear Lie group, then the vector field (3.10) is tangent to the submanifold G. Therefore, the submanifold G is positively invariant for (3.10). To see that the vector field (3.10) is tangent to any linear Lie group G, suppose that $E(t_0) \in G$ at some time $t_0 \in \mathbb{R}$. Then $\log(E(t_0)) \in \mathrm{Lie}(G)$ and left-translation of this vector takes it to the tangent space to G at $E(t_0)$, i.e., $E(t_0)\log(E(t_0)) \in T_{E(t_0)}G$, by Proposition 2.2.12. Thus, the vector field (3.10) is such that $\dot{E}(t_0) \in T_{E(t_0)}G$.

The crucial property of the differential equation (3.10) is that the matrices $\dot{E}$ and E commute, i.e., $E\dot{E} = \dot{E}E$. This property is a consequence of matrices E and $\log(E)$ commuting. Commutativity of $\dot{E}$ and E, combined with the product rule, gives us the following result.

Lemma 3.5.1. *Let $E : \mathbb{R} \to \mathsf{GL}(n, \mathbb{R})$ be a curve in $\mathsf{GL}(n, \mathbb{R})$, such that E and $\dot{E}$ commute, i.e., $E\dot{E} = \dot{E}E$. Then for all positive integers k, we have*

$$\begin{aligned} \frac{d}{dt}\left[(E - I_n)^k\right] &= k\dot{E}(E - I_n)^{k-1} \\ &= k(E - I_n)^{k-1}\dot{E}. \end{aligned}$$

Proof. By straight-forward computation, using the product rule and commutativity of $\dot{E}$

and E, we get

$$
\begin{aligned}
\frac{d}{dt}\left[(E-I_n)^k\right] &= \frac{d}{dt}\left[\overbrace{(E-I_n)(E-I_n)\cdots(E-I_n)}^{\text{k times}}\right] \\
&= \overbrace{\dot{E}(E-I_n)^{k-1} + (E-I_n)\dot{E}(E-I_n)^{k-2} + \cdots + (E-I_n)^{k-1}\dot{E}}^{\text{k times}} \\
&= k\dot{E}(E-I_n)^{k-1} \\
&= k(E-I_n)^{k-1}\dot{E}.
\end{aligned}
$$

□

Lemma 3.5.1 is the key reason why we will be able to convert the differential equation (3.10) into a linear differential equation via a certain change of coordinates. This change of coordinates is the matrix logarithm map defined on $B(I_n, 1) \subset \mathsf{GL}\,(n, \mathbb{R})$. By Lemma 2.2.6, the matrix logarithm map $\log : B(I_n, 1) \to \mathsf{M}\,(n, \mathbb{R})$ is a diffeomorphism onto its image. Furthermore, the codomain of the log map is the set $\mathsf{M}\,(n, \mathbb{R})$, which is isomorphic to $\mathbb{R}^{n^2}$, as a vector space. Therefore, the log map is a local coordinate transformation on $\mathsf{GL}\,(n, \mathbb{R})$, defined on the ball $B(I_n, 1)$.

We denote by $e \in \mathsf{M}\,(n, \mathbb{R})$ the log coordinates of the matrix $E \in B(I_n, 1)$

$$e := \log(E). \tag{3.11}$$

To express the differential equation (3.10) in log coordinates we differentiate e with respect to time, making use of Lemma 3.5.1 and Proposition 2.2.2

$$
\begin{aligned}
\dot{e} &= \frac{d}{dt}\log(E) \\
&= \frac{d}{dt}\left[\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}(E-I_n)^k\right] \\
&= \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{d}{dt}\left[(E-I_n)^k\right] \\
&= \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\left[k\dot{E}(E-I_n)^{k-1}\right] \\
&= \sum_{k=0}^{\infty}(-1)^k(E-I_n)^k\dot{E} \\
&= E^{-1}\dot{E} \\
&= -a_0E^{-1}E\log(E) \\
&= -a_0e.
\end{aligned}
$$

The above equation, rewritten $\dot{e} = -a_0e$, is linear and its poles are real and located at $-a_0$, with multiplicity n^2. Thus, for any positive constant $a_0 > 0$, the point $e = 0_n$ is an exponentially stable equilibrium of $\dot{e} = -a_0e$, by Lemma A.3.1. Therefore, the equilibrium point $E = I_n$ is locally exponentially stable for (3.10), by Lemma A.2.1. The above discussion proves the following.

Lemma 3.5.2. *On the set $E \in B(I_n, 1)$, the vector field (3.10) is differentially equivalent to the vector field*

$$\dot{e} = -a_0e, \tag{3.12}$$

where $e := \log(E)$. As a result, the equilibrium point $E = I_n$ of (3.10) is locally exponentially stable.

The solution, $E(t)$, of the differential equation (3.10) can be expressed in closed form. This is useful in obtaining an intuitive understanding of the equation (3.10), but is not necessary for our main argument.

Proposition 3.5.3. *Let $E_0 \in B(I_n, 1)$ be sufficiently close to I_n so that $\log(E_0) \in B(0_n, \log(2))$ and let $a_0 > 0$ be arbitrary. Then the solution of (3.10) with initial condition $E(0) = E_0$ is defined for all $t \geq 0$ and is given by*

$$E(t) = \exp(\exp(-a_0t)\log(E_0)). \tag{3.13}$$

Proof. First, we will show that the candidate solution (3.13) is a solution to the differential equation (3.10) with initial condition E_0. First, we check the initial condition. The value at $t = 0$ of $E(t)$ is $E(0) = E_0$ as required.

Next, let us check that $E(t)$ satisfies (3.10). Differentiate $E(t)$ with respect to time

$$\begin{aligned}\frac{d}{dt}E(t) &= \frac{d}{dt}\left[\exp(\exp(-a_0 t)\log(E_0))\right] \\ &= -a_0 \exp(-a_0 t)\log(E_0)\exp(\exp(-a_0 t)\log(E_0)) \\ &= -a_0 \log(E)E,\end{aligned}$$

where we have used the identity

$$\log(E) = \log(\exp(\exp(-a_0 t)\log(E_0))) = \exp(-a_0 t)\log(E_0),$$

which follows from Lemma 2.2.6 (b) and the assumption that $\log(E_0) \in B(0_n, \log(2))$.

To finish the proof, let us check that the solution $E(t)$ of (3.10), with initial condition E_0 as stated in the proposition, is such that for all future times $t > 0$, we have $E(t) \in B(I_n, 1)$. The solution of the linear differential equation (3.12) is $e(t) = e(0)\exp(-a_0 t)$, where $e(0) = \log(E_0)$. Using the fact that $E(t) = \exp(e(t))$, we compute upper bound on $\|E(t) - I_n\|$, for $t \geq 0$:

$$\begin{aligned}\|E(t) - I_n\| &= \left\|\sum_{k=0}^{\infty}\frac{1}{k!}e(0)^k \exp(-a_0 t k) - I_n\right\| \\ &= \left\|\sum_{k=1}^{\infty}\frac{1}{k!}e(0)^k \exp(-a_0 t k)\right\| \\ &\leq \sum_{k=1}^{\infty}\frac{1}{k!}\|e(0)\|^k \exp(-a_0 t k) \\ &\leq \sum_{k=1}^{\infty}\frac{1}{k!}\|e(0)\|^k \\ &= \exp(\|e(0)\|) - 1 \\ &< 2 - 1 = 1.\end{aligned}$$

The last inequality follows from the assumption that $\log(E_0) \in B(0_n, \log(2))$. Thus, $E(t) \in B(I_n, 1)$ for all $t \geq 0$. $\square$

Let $E(t)$ be a solution of (3.10), which is initailized at $E_0 = E(0)$, such that the conditions of Proposition 3.5.3 are satisfied. Then for all $t \geq 0$, $E(t)$ stays on the same one-parameter subgroup, on which it was initialized at time $t = 0$. Indeed, by Proposition 3.5.3, we have

$$(\forall t \geq 0) \ (\exists \tau \in \mathbb{R}) \quad E(t) = \exp(\tau \log(E_0)).$$

Thus, in a neighbourhood of $I_n \in \mathsf{GL}\,(n, \mathbb{R})$, the vector field (3.10) is a linear vector field on Lie group, since its flow is a one-parameter subgroup [8], [9].

3.6 Estimation Error Dynamics

In this section, we analyze the dynamics of the error functions E_l and E_r. When looking at the observer equations (3.2) and (3.3), we substitute X for the variable Y. Also, we make frequent use of Lemma 2.2.9.

In this section, we make the assumption that $\hat{X}$ is initialized sufficiently close to X, so that $E_l, E_r \in B(I_n, 1)$. This assumption is sufficient to ensure that our series definitions, using (2.3), of $\log(E_r)$ and $\log(E_l)$ are convergent series, and also sufficient for properties like for example, $X \log(E_l) X^{-1} = \log(E_r)$, to hold. Actually, the assumption that $E_l, E_r \in B(I_n, 1)$ is not necessary to ensure that $\log(E_l)$ and $\log(E_r)$ are convergent series, as discussed in Section 2.2.

3.6.1 Passive Observer

Consider what happens when the passive observer (3.2) is applied to estimate the state of the system (3.1). First, we examine the dynamics of the right-invariant error, E_r, making use of Lemma 2.2.9:

$$\begin{aligned}
\dot{E}_r &= \frac{d}{dt}\left[\hat{X}X^{-1}\right] \\
&= \dot{\hat{X}}X^{-1} - \hat{X}X^{-1}\dot{X}X^{-1} \\
&= \hat{X}uX^{-1} - a_0\hat{X}\log(X^{-1}\hat{X})X^{-1} - \hat{X}uX^{-1} \\
&= -a_0\hat{X}\log(X^{-1}\hat{X})X^{-1} \\
&= -a_0\hat{X}X^{-1}X\log(X^{-1}\hat{X})X^{-1} \\
&= -a_0\hat{X}X^{-1}\log(\hat{X}X^{-1}) \\
&= -a_0E_r\log(E_r).
\end{aligned} \tag{3.14}$$

The above differential equation is the same as the equation (3.10), if we identify E with E_r. This means that if $\hat{X}$ is sufficiently close to X so that $E_r \in B(I_n, 1)$, then we can use Lemma 3.5.2 to rewrite the above equation in e_r-coordinates

$$\dot{e}_r = -a_0 e_r, \tag{3.15}$$

which is a linear differential equation. Lemma 3.5.2 also tells us that if we choose $a_0 > 0$ to be any positive constant, then the equilibrium point $E_r = I_n$ is locally exponentially stable for the dynamics (3.14).

Next, we examine the dynamics of the left-invariant error, E_l

$$\begin{aligned} \dot{E}_l &= \frac{d}{dt}\left[X^{-1}\hat{X}\right] \\ &= -X^{-1}\dot{X}X^{-1}\hat{X} + X^{-1}\dot{\hat{X}} \\ &= -uX^{-1}\hat{X} + X^{-1}\hat{X}u - a_0 X^{-1}\hat{X}\log(X^{-1}\hat{X}) \\ &= -uE_l + E_l u - a_0 E_l \log(E_l) \\ &= \delta_P(u, E_l) - a_0 E_l \log(E_l), \end{aligned} \tag{3.16}$$

where $\delta_P(u, E_l) := E_l u - uE_l$ is a perturbation term that vanishes when $E_l = I_n$. In the differential equation (3.16), the matrices E_l and $\dot{E}_l$ do not commute in general, because E_l and $\delta_P(u, E_l)$ do not commute in general. For this reason, Lemma 3.5.1 does not hold for the vector field (3.16). This makes it difficult to compute the derivative of the Taylor series expansion of the matrix logarithm map, like we did in the proof of Lemma 3.5.2. This makes the differential equation (3.16) harder to analyze than the differential equation (3.10).

Our goal is to transform the error dynamics (3.16) into log coordinates, i.e., to find the expression for $\dot{e}_l$. To perform the transformation into log coordinates, we recall Lemma 3.4.4 which gives us the equality $e_l = X^{-1}e_r X$. Thus, to compute $\dot{e}_l$, we can just differentiate both sides of this relation and use the product rule. We know derivatives of X^{-1}, of e_r and of X, which gives us enough information to compute the expression of $\dot{e}_l$. To be able to do this, it is sufficient that the conditions of Lemma 3.4.4 are satisfied, i.e., that $E_l, E_r \in B(I_n, 1)$.

$$\begin{aligned} \dot{e}_l &= \frac{d}{dt}\left[X^{-1}e_r X\right] \\ &= -X^{-1}\dot{X}X^{-1}e_r X + X^{-1}\dot{e}_r X + X^{-1}e_r\dot{X} \\ &= -ue_l - a_0 e_l + e_l u \\ &= -a_0 e_l + \left[e_l, u\right]. \end{aligned} \tag{3.17}$$

The above system, rewritten $\dot{e}_l = -a_0 e_l + [e_l, u]$, is not linear, but is bilinear. Indeed, this vector field is linear in e_l and linear-affine in u, because the matrix commutator $[e_l, u]$ is linear in e_l and in u, therefore the system (3.17) is bilinear.

3.6.2 Direct Observer

We now consider what happens when the direct observer (3.3) is applied to estimate the state of the system (3.1). We first examine the dynamics of the left-invariant error E_l, making use of Lemma 2.2.9,

$$\begin{aligned} \dot{E}_l &= \frac{d}{dt}\left[X^{-1}\hat{X}\right] \\ &= -X^{-1}\dot{X}X^{-1}\hat{X} + X^{-1}\dot{\hat{X}} \\ &= -uX^{-1}\hat{X} + uX^{-1}\hat{X} - a_0 X^{-1}\hat{X}\log(X^{-1}\hat{X}) \\ &= -uE_l + uE_l - a_0 E_l \log(E_l) \\ &= -a_0 E_l \log(E_l). \end{aligned} \tag{3.18}$$

The above equation (3.18) is the same as the equation (3.10), if we identify E_l with E. This means that if $\hat{X}$ is sufficiently close to X so that $E_l \in B(I_n, 1)$, then we can use Lemma 3.5.2 to convert the above equation into e_l-coordinates

$$\dot{e}_l = -a_0 e_l, \tag{3.19}$$

which is a linear differential equation. Lemma 3.5.2 also tells us that if we choose $a_0 > 0$ to be any positive constant, then the equilibrium point $E_l = I_n$ is locally exponentially stable for the dynamics (3.18).

Next, we examine the dynamics of the right-invariant error E_r, when the direct observer is used

$$\begin{aligned} \dot{E}_r &= \frac{d}{dt}\left[\hat{X}X^{-1}\right] \\ &= \dot{\hat{X}}X^{-1} - \hat{X}X^{-1}\dot{X}X^{-1} \\ &= XuX^{-1}\hat{X}X^{-1} - a_0\hat{X}\log(X^{-1}\hat{X})X^{-1} - \hat{X}uX^{-1} \\ &= XuX^{-1}E_r - E_r XuX^{-1} - a_0 E_r X \log(E_l) X^{-1} \\ &= XuX^{-1}E_r - E_r XuX^{-1} - a_0 E_r \log(E_r) \\ &= \delta_D(u, X, E_r) - a_0 E_r \log(E_r), \end{aligned} \tag{3.20}$$

where $\delta_D(u, X, E_r) := XuX^{-1}E_r - E_rXuX^{-1}$ is a perturbation term that vanishes when $E_r = I_n$. The above equation (3.20) has the same problem that we encountered when trying to analyze equation (3.16). The problem with (3.20) is that matrices E_r and $\dot{E}_r$ do not commute in general, because E_r and $\delta_D(u, X, E_r)$ do not commute in general. Fortunately, we can use the same trick that we used to transform equation (3.16) into log coordinates. Namely, we differentiate the equation $e_r = Xe_lX^{-1}$ and use our knowledge of derivatives of X, of e_l and of X^{-1}. To be able to do this, it is sufficient that the conditions of Lemma 3.4.4 are satisfied, i.e., that $E_l, E_r \in B(I_n, 1)$.

$$\begin{aligned}\dot{e}_r &= \frac{d}{dt}\left[Xe_lX^{-1}\right] \\ &= \dot{X}e_lX^{-1} + X\dot{e}_lX^{-1} - Xe_lX^{-1}\dot{X}X^{-1} \\ &= Xue_lX^{-1} - a_0Xe_lX^{-1} - Xe_luX^{-1} \\ &= -a_0e_r + \left[XuX^{-1}, e_r\right].\end{aligned} \tag{3.21}$$

The above system, rewritten $\dot{e}_r = -a_0e_r + [XuX^{-1}, e_r]$, is not linear, but is bilinear. Indeed, this vector field is linear in e_r and linear-affine in u, because the matrix commutator $[XuX^{-1}, e_r]$ is linear in e_r and in u. However, unlike (3.17), dynamics (3.21) are not autonomous, because X appears as part of its vector field. Since X changes with time, it means that the dynamics (3.21) are not time invariant.

3.6.3 Discussion

We summarize the results of this section in Table 3.1. It is interesting to note that neither

Table 3.1: Summary of Lie-group full-state observer convergence results

System	Observer	Error	Error Dynamics
Left-invariant	Passive	e_r	Linear, Autonomous
Left-invariant	Passive	e_l	Bilinear, Autonomous
Left-invariant	Direct	e_r	Bilinear, Non-autonomous
Left-invariant	Direct	e_l	Linear, Autonomous

the system (3.1), nor the passive LFSO (3.2) are linearizable via the change of coordinates $\log(X)$ and $\log(\hat{X})$, respectively. However, the right-invariant Lie group estimation error between them is linearizable via the matrix logarithm map, i.e., via $\log(E_r)$. The same can be said of the direct LFSO (3.3).

We have seen that the passive observer (3.2) produces exponential convergence of the right-invariant error, $E_r \to I_n$, while the direct observer (3.3) produces exponential convergence of the left-invariant error, $E_l \to I_n$. Therefore, if Assumption 1 holds true, then both of these observers are such that $\hat{X} \to X$, by Proposition 3.4.2. Unfortunately, without invoking Assumption 1, convergence of either one of $E_l \to I_n$ or of $E_r \to I_n$ is not sufficient to deduce convergence of $X \to \hat{X}$ (see Proposition 3.4.2).

For example, if we use the direct observer and get $E_l \to I_n$, then we can have a situation where the input u to the system (3.1) makes $\|X\|$ grow unbounded. This can cause the diffeomorphism Ad_X to "stretch" the coordinates $E_r = \mathrm{Ad}_X(E_l)$ faster than E_l is converging to I_n. This can prevent the convergence of E_r to I_n, even as E_l does converge to I_n. This problem is eliminated if we make Assumption 1 hold, because it forces $\|X\|$ to be uniformly bounded for all time (see Proposition 3.4.2).

Another way to look at this convergence problem is to consider the perturbation term from the passive observer, $\delta_P(u, E_l) := E_l u - u E_l$. If this perturbation term is not identically equal to zero for all u, then we can make the norm of this perturbation term arbitrarily large by some choice of the input u. For example, suppose that E_l and u do not commute, so that we have $\|\delta_P(u, E_l)\| = r > 0$. Then if we set $\bar{u} = ku$, where $k > 0$ is an arbitrarily large, positive constant, we will have $\delta_P(\bar{u}, E_l) = k\delta_P(u, E_l)$. If u is an admissible input to the system (3.1), then so is $\bar{u} = ku$. So we see that $\|\delta_P(u, E_l)\|$ can be made arbitrarily large if it is not identically equal to zero for all u. Analogous result holds for the perturbation term of the direct observer: $\delta_D(ku, X, E_r) = k\delta_D(u, X, E_r)$.

By showing that the system (3.16) is diffeomorphic to the system (3.17), we have found an easy way to prove the following result.

Corollary 3.6.1. *Let $G \subseteq \mathsf{GL}(n, \mathbb{R})$ be a linear Lie group. Consider the following system on G*

$$\dot{E} = [E, u], \tag{3.22}$$

where $E \in G \subseteq \mathsf{GL}(n, \mathbb{R})$ is the state and $u \in \mathrm{Lie}(G) \subseteq \mathsf{M}(n, \mathbb{R})$ is an admissible input signal. The system (3.22) is diffeomorphic to the following dynamical system on the Lie algebra of G

$$\dot{e} = [e, u], \tag{3.23}$$

where $e = \log(E)$ is defined for all $E \in B(I_n, 1)$.

Proof. Rewrite (3.22) as a difference of two vector fields

$$\begin{aligned}\dot{E} &= ([E,u] + E\log(E)) - (E\log(E)) \\ &= f(E,u) - g(E),\end{aligned}$$

where $f(E,u) := [E,u] + E\log(E)$ and $g(E) := E\log(E)$. Since the system (3.16) transforms into the system (3.17), we know that the vector field $f(E,u)$ transforms into $[e,u]+e$. Also, since the dynamics (3.10) transform into the dynamics (3.12), we know that the vector field $g(E)$ transforms into e. This means that the vector field $f(E,u)-g(E)$ transforms into $[e,u] + e - e = [e,u]$. □

Recall, that in the equation (3.10), we were able to easily differentiate the Taylor series expansion of $\log(E)$, because the matrices E and $\dot{E} = -a_0 E\log(E)$ were mutually commuting, i.e., $E\dot{E} = \dot{E}E$. However, in the equation (3.22), the matrices E and $\dot{E} = [E,u]$ are not mutually commuting, i.e., in general $E\dot{E} \neq \dot{E}E$. This non-commutativity makes it very difficult to differentiate the Taylor series expansion of $\log(E)$, when $\dot{E} = [E,u]$, as in (3.22). Thus, it seems that the result of Corollary 3.6.1 would be difficult to obtain by directly differentiating the series expansion of $\log(E)$ and substituting $\dot{E} = [E,u]$. However, our analysis of equation (3.22) is made much easier by "splitting" the equation (3.22) into a pair consisting of "system" (3.1), with state X, and "observer" (3.2), with state $\hat{X}$. The splitting is done as $E = X^{-1}\hat{X}$, and allows us to convert the differential equation (3.22) into log coordinates.

3.7 Example: Kinematic Rigid-Body Orientation Estimation on $\mathsf{SO}(3)$

The LFSO, described in this section, can be used to accurately estimate the orientation of any object by means of two sets of on-board sensors: an orientation sensor and angular-rate gyroscopes. An orientation sensor is a device that can approximately determine the orientation of the rigid-body, possibly with a large amount of zero-mean measurement noise. The orientation sensor can be a number of different things, depending on the specific application. For an outdoor flying vehicle, the orientation can be measured by a combination of a triaxial accelerometer and a triaxial magnetometer. For an indoor flying vehicle, the orientation can be measured by an on-board camera, possibly combined with a triaxial accelerometer. Measurements made by the orientation sensor, such as an accelerometer and magnetometer pair, are usually corrupted by a large amount of high-frequency, zero-mean noise. In the accelerometer, measurement noise is caused by vibration of the rigid-body. In the magnetometer, measurement noise is caused by high-frequency electric currents, flowing near the magnetometer.

As we have seen in Section 1.1, feeding sensor measurements directly into a state feedback control law makes the controller very sensitive to high-frequency, zero-mean measurement noise. To make any rigid-body control law more resistant to measurement noise, one solution is to use a full-state observer, as was shown in Figure 3.1. To this end, we define the "input" to our system to be the angular velocity of the rigid-body, which can be measured by angular rate gyroscopes. Knowledge of the angular velocity is the extra piece of information that lets our LFSO efficiently filter out noise from the orientation measurement. This type of full-state observer has also been called "complementary filter" in some of the literature, such as [1] and [2].

Each element $R \in \mathsf{SO}(3)$ can be used to represent a "relative rotation", between any two "orientations", or frames, $\{A\}$ and $\{B\}$, which have common origin. Any vector $v_{\{B\}} \in \mathbb{R}^3$, expressed in frame $\{B\}$ coordinates, is transformed into frame $\{A\}$ coordinates via multiplication on the left by R, i.e., $v_{\{A\}} = Rv_{\{B\}}$. We assume that the frame $\{A\}$ remains stationary while the frame $\{B\}$ rotates with known angular velocities. The angular velocities of $\{B\}$, expressed in the $\{B\}$ coordinates, are encoded into a skew-symmetric matrix $u \in \text{Lie}(\mathsf{SO}(3))$. In this case, the rotation matrix R satisfies the left-invariant kinematic equations

$$\begin{aligned} \dot{R} &= Ru \\ Y &= R, \end{aligned} \tag{3.24}$$

where $Y \in \mathsf{SO}(3)$ is a direct measurement of R. The reason why this equation is useful is because Y and u can be measured or estimated by low-cost sensors, mounted on the mobile frame $\{B\}$. In fact, u can be measured using angular rate gyroscopes, while the rotation matrix Y can be calculated from sensor measurements, made by an accelerometer and magnetometer pair.

An accelerometer tells us which direction is "up" and a magnetometer tells us the direction of the Earth's magnetic field, which tells us the direction of the magnetic North. Both of these sensors output 3-dimensional vectors, as measured in the mobile frame $\{B\}$, they can then be compared with the same 3-dimensional vectors, as measured in the inertial frame $\{A\}$, to uniquely compute the corresponding measured rotation, $Y \in \mathsf{SO}(3)$. One convenient way to convert these vector measurements into a corresponding measured rotation matrix Y, is to first convert these vector measurements into Euler angles representation of Y. For example, the accelerometer measurement can tell us the roll and pitch of the rigid-body, all that remains is to use the magnetometer measurement to determine the yaw. Once Euler angles are known, they are converted into the matrix Y, by using standard formulae.

For the kinematic system (3.24), our passive LFSO is

$$\dot{\hat{R}} = \hat{R}u - a_0\hat{R}\log(Y^\top\hat{R}) \tag{3.25}$$

and our direct LFSO is

$$\dot{\hat{R}} = YuY^\top\hat{R} - a_0\hat{R}\log(Y^\top\hat{R}). \tag{3.26}$$

We would like to briefly discuss the connection between our two observers and the filters proposed in [1], [2]. The filters proposed in these two papers use the anti-symmetric projection operator in matrix space, $\pi_a : \mathsf{M}(n, \mathbb{R}) \to \mathrm{Lie}(\mathsf{SO}(3))$, in their innovation terms. The anti-symmetric projection operator is defined as

$$\pi_a(A) := \frac{1}{2}(A - A^\top).$$

The anti-symmetric projection operator satisfies: $\pi_a(A^\top) = -\pi_a(A)$. The passive filter proposed in [1], [2] is

$$\begin{aligned}\dot{\hat{R}} &= \hat{R}u + k\hat{R}\pi_a(\hat{R}^\top Y)\\ &= \hat{R}u - k\hat{R}\pi_a(Y^\top\hat{R})\end{aligned} \tag{3.27}$$

and the direct filter proposed in [1], [2] is

$$\begin{aligned}\dot{\hat{R}} &= \mathrm{Ad}_Y(u)\hat{R} + k\,\mathrm{Ad}_{\hat{R}}(\pi_a(\hat{R}^\top Y))\hat{R}\\ &= YuY^\top\hat{R} + k\hat{R}\pi_a(\hat{R}^\top Y)\\ &= YuY^\top\hat{R} - k\hat{R}\pi_a(Y^\top\hat{R}),\end{aligned} \tag{3.28}$$

where $k \in \mathbb{R}$ is a positive constant, which adjusts the rate of observer convergence, similar to the positive constant a_0 in our proposed observers.

To compare our proposed observers (3.25) and (3.26) to the filters (3.27) and (3.28), we will rewrite the equations of our obsevers (3.25) and (3.26), so that they use the anti-symmetric matrix projection, instead of the matrix logarithm map. We can express the matrix logarithm map on $\mathsf{SO}(3)$ in terms of the anti-symmetric projection operator by using the following equation from [1, Section III. C.]

$$\pi_a(A) = \frac{\sin(\theta_A)}{\theta_A}\log(A), \qquad A \in \mathsf{SO}(3), \tag{3.29}$$

where θ_A is the rotation angle associated to the rotation matrix $A \in \mathsf{SO}(3)$, which can be computed as follows [33]

$$\theta_A := \arccos\left(\frac{\operatorname{tr}(A) - 1}{2}\right)$$
$$= \arccos\left(\frac{1}{2}\left[A_{11} + A_{22} + A_{33} - 1\right]\right).$$

Using the identity (3.29), we can immediately rewrite our proposed observers (3.25) and (3.26) so that they use the anti-symmetric projection operator, rather than the logarithm map. For $\theta_{Y^\top \hat{R}} \neq \pm\pi$, our passive LFSO (3.25) is rewritten as

$$\dot{\hat{R}} = \hat{R}u - a_0 \frac{\theta_{Y^\top \hat{R}}}{\sin(\theta_{Y^\top \hat{R}})} \hat{R}\pi_a(Y^\top \hat{R}) \tag{3.30}$$

and our direct LFSO (3.26) is rewritten as

$$\dot{\hat{R}} = YuY^\top \hat{R} - a_0 \frac{\theta_{Y^\top \hat{R}}}{\sin(\theta_{Y^\top \hat{R}})} \hat{R}\pi_a(Y^\top \hat{R}). \tag{3.31}$$

If we identify the observer gains to be equal $k = a_0$, then our observers (3.30), (3.31) differ from the observers (3.27), (3.28) only by the scalar quantity $\frac{\theta_{Y^\top \hat{R}}}{\sin(\theta_{Y^\top \hat{R}})}$, which appears in the innovation terms of (3.30), (3.31), but does not appear in the innovation terms of (3.27), (3.28). This scalar quantity, $\frac{\theta_{Y^\top \hat{R}}}{\sin(\theta_{Y^\top \hat{R}})}$, approaches 1 as $\theta_{Y^\top \hat{R}}$ approaches 0. Thus, on $\mathsf{SO}(3)$ and for small $\theta_{Y^\top \hat{R}}$, our observers behave similar to the observers of [1], [2].

It is computationally costly and inefficient to compute the matrix logarithm map by using the series definitions. Fortunately, by rewriting our LFSOs in the form (3.30) and (3.31), we avoid using the logarithm map and only use the anti-symmetric projection operator, which is much faster to compute. Thus, we have found a computationally fast method to implement our proposed LFSOs for state estimation on the Lie group $\mathsf{SO}(3)$.

We now simulate our direct and passive LFSOs. The initial conditions for the plant and the observer are chosen (somewhat arbitrarily) to be the following matrices

$$R(0) = \begin{pmatrix} 0.6330 & -0.1116 & -0.7660 \\ 0.7128 & -0.3020 & 0.6330 \\ -0.3020 & -0.9467 & -0.1116 \end{pmatrix}, \qquad \hat{R}(0) = I_3.$$

The angular velocity input is chosen (somewhat arbitrarily) to be

$$u(t) = \begin{pmatrix} 0 & -2\sin(t) & \cos(t) \\ 2\sin(t) & 0 & -\sin(t) \\ -\cos(t) & \sin(t) & 0 \end{pmatrix}.$$

The observer gain is taken to be

$$a_0 = 1.$$

Figure 3.2 shows the simulation results when there is no noise, i.e., when Y is exactly equal to R. From these plots we see that, when there no measurement noise, the direct and passive LFSOs have similar performance.

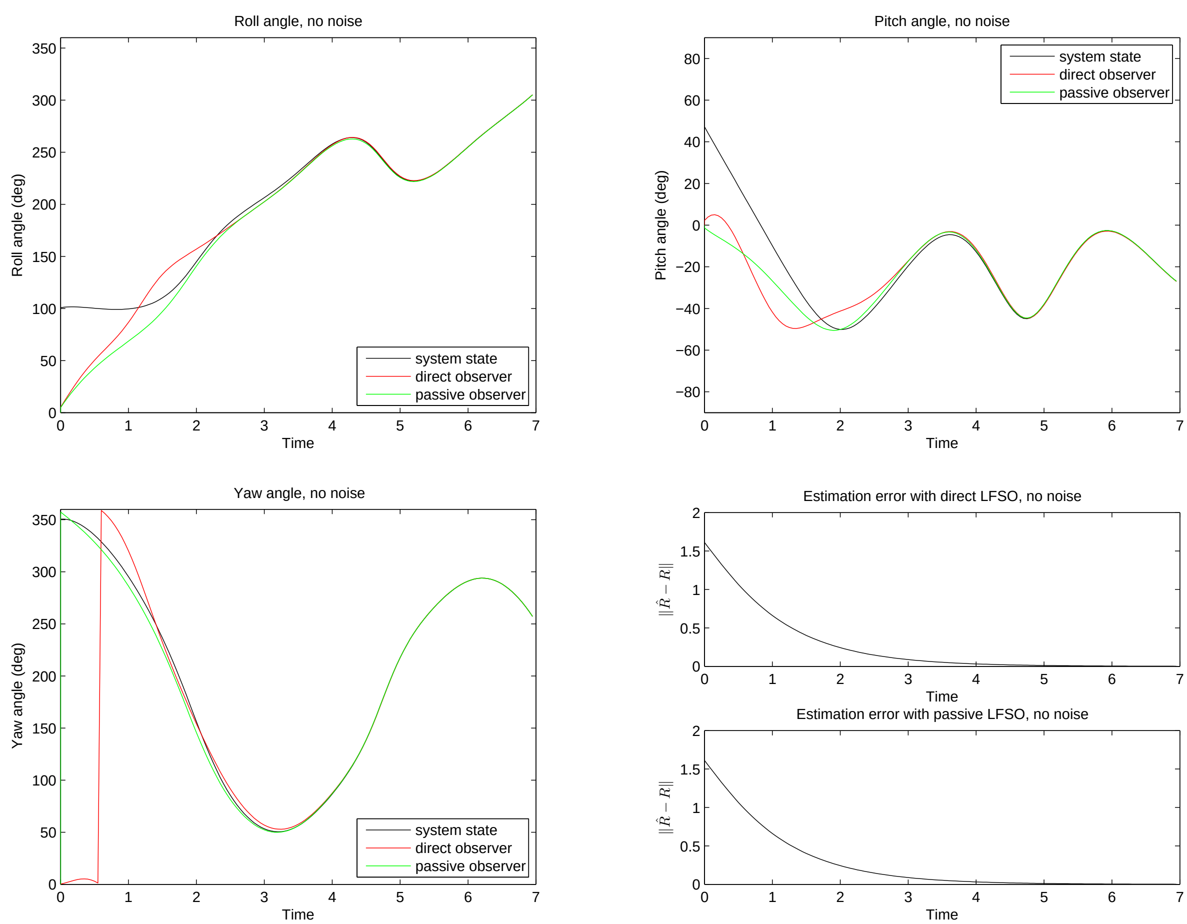

Figure 3.2: Direct and Passive LFSOs, without measurement noise.

The Lie group estimation errors, E_l and E_r, are initially (at time $t = 0$)

$$E_l(0) = R^{-1}(0)\hat{R}(0) = R^{-1}(0)$$
$$E_r(0) = \hat{R}(0)R^{-1}(0) = R^{-1}(0).$$

Calculating the distance between $E_l(0)$ or $E_r(0)$ and I_3, we get

$$\begin{aligned}
\|E_l(0) - I\| &= \|E_r(0) - I\| \\
&= \|R^{-1}(0) - I\| \\
&= \left\| \begin{pmatrix} 0.6330 & 0.7128 & -0.3020 \\ -0.1116 & -0.3020 & -0.9467 \\ -0.7660 & 0.6330 & -0.1116 \end{pmatrix} - I_3 \right\| \\
&= 1.6675,
\end{aligned}$$

therefore $E_l(0), E_r(0) \notin B(I_3, 1)$. Our analysis of the Lie group estimation error dynamics was performed on the ball $B(I_n, 1)$, in particular the differential equivalence of (3.10) to (3.12) was shown to hold for $E \in B(I_n, 1)$. So the fact that $E_l(0), E_r(0) \notin B(I_3, 1)$ means that the results of this chapter do not allow us to deduce the convergence of our LFSO, with the initial conditions as chosen in this simulation. However, our simulation shows that that our LFSOs are indeed convergent. This suggests that the local exponential stability of our LFSO is a conservative statement, and that our LFSOs could have a fairly large (perhaps global) region of convergence.

Next, we add measurement noise to our simulation. Measurement noise is simulated by multiplying R on the right by a randomly generated rotation matrix N, to obtain the noise corrupted output: $Y = RN$. The matrix N represents uncertainty that arises from the inaccuracy of measuring sensors. To generate the random rotation matrix $N \in \mathsf{SO}(3)$, we first generate a random skew-symmetric matrix, $n \in \mathrm{Lie}(\mathsf{SO}(3)) \cong \mathbb{R}^3$, so that it is normally distributed, with zero-mean and standard deviation of $\sigma = 0.4$. The random rotation matrix $N \in \mathsf{SO}(3)$ is then computed as $N := \exp(n)$.

Since both R and N are elements of $\mathsf{SO}(3)$, their product, $Y = RN$, is also an element of $\mathsf{SO}(3)$. The noisy output, $Y = RN$, is fed into the observer and the observer tries to reconstruct the state R. Figure 3.3 shows the simulation results when there a significant amount of measurement noise added to the output. From the noisy simulation, we see that the passive LFSO appears to work better than the direct LFSO, in the presence of significant amounts of noise.

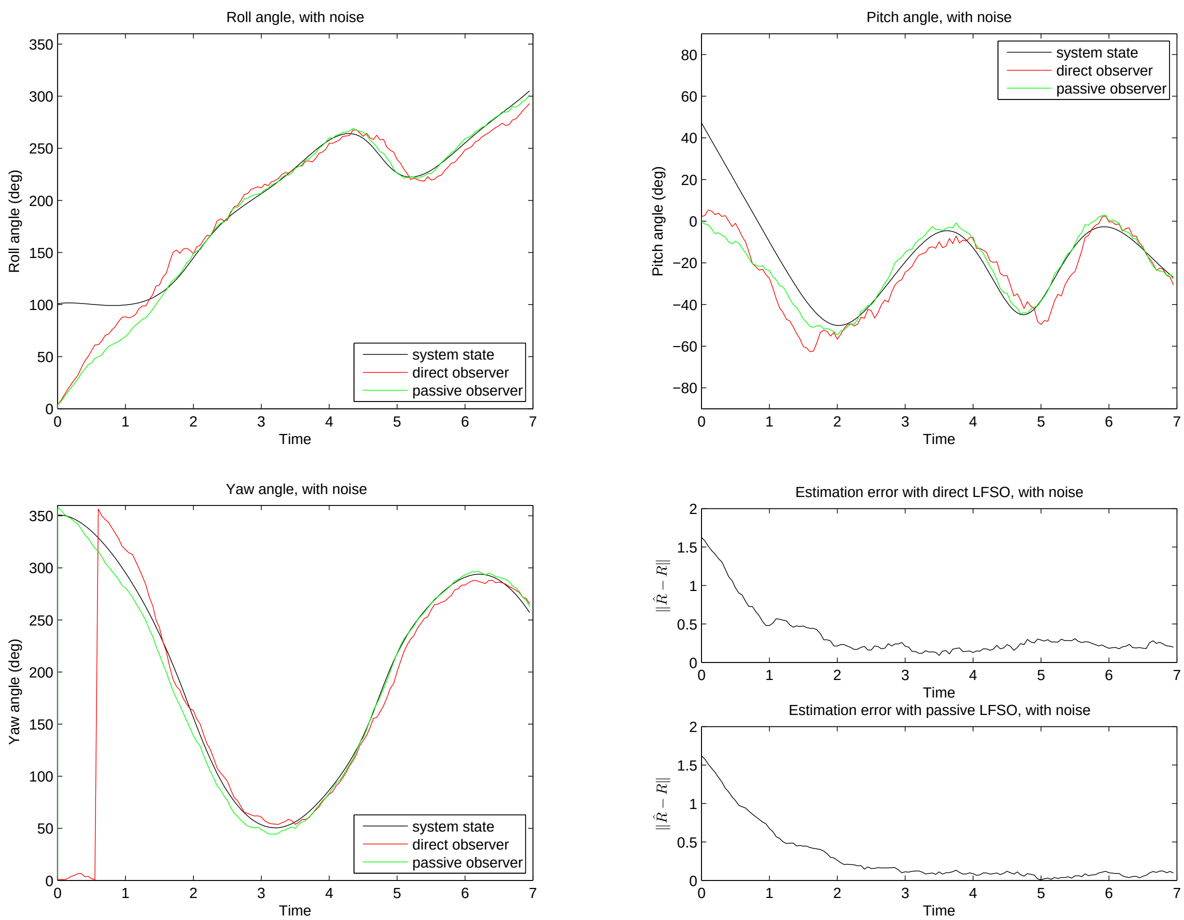

Figure 3.3: Direct and Passive LFSOs, with measurement noise.

3.8 Application to Control

The results of Section 3.5, can be used to design control laws for left-invariant systems on Lie group, such as for the system (3.1). Suppose that we know the state X exactly, and our goal is to set the input $u \in \mathrm{Lie}(G)$ such that $X \to X_{ref}$, where $X_{ref} \in G$ is the constant target state. To achieve this task, we propose to use the following "proportional" state feedback control law:

$$u = -k_0 \log(X_{ref}^{-1} X),$$

where $k_0 \in \mathbb{R}$ is a positive constant, which is the controller gain. With such a controller, the closed loop system is:

$$\dot{X} = -k_0 X \log(X_{ref}^{-1} X).$$

Consider the quantity: $X_{ref}^{-1} X$, which is a way to express "control error" between X and X_{ref}. This quantity satisfies the differential equation:

$$\begin{aligned} \frac{d}{dt} X_{ref}^{-1} X &= X_{ref}^{-1} \dot{X} \\ &= -k_0 X_{ref}^{-1} X \log(X_{ref}^{-1} X). \end{aligned} \tag{3.32}$$

The differential equation (3.32) is the same as (3.10), if we identify $X_{ref}^{-1} X$ with E. Therefore, Lemma 3.5.2 tells us that $X_{ref}^{-1} X = I_n$ is a locally exponentially stable equilibrium point of (3.32). Which implies that if X is initialized sufficiently close to X_{ref}, then $X \to X_{ref}$ exponentially, since $X_{ref} \in G$ is constant.

Chapter 4

Lie Group Partial-State Observers

In this chapter, we propose an extension of the results presented in Chapter 3 to a broader class of systems. The class of systems considered in this chapter has only a proper subset of its states evolving on the Lie group. The rest of the states are evolving on the Lie algebra of the Lie group. The states that are evolving on the Lie group are the states that are available for measurement from the output. The states that are evolving on the Lie algebra are not available for measurement from the output. This class of systems arises when dynamics are considered for a left-invariant system on Lie group. We propose observer design for this class of dynamical systems. Since not all the states are available for measurement, this type of observer falls under the class of partial-state observers. We show that the estimation error dynamics are locally exponentially stable at the zero equilibrium point. We also give some examples where such observers can be used.

4.1 Introduction

In this chapter, we generalize the results of Chapter 3 to the case when the output does not include all the states. To motivate our research, we return to the example of kinematic rigid-body orientation estimation on $\mathsf{SO}(3)$, from Section 3.7, where we used the kinematic rigid-body rotation model (3.24), given here again

$$\begin{aligned} \dot{R} &= Ru \\ Y &= R, \end{aligned}$$

where the input, $u \in \mathrm{Lie}(\mathsf{SO}(3))$, is the angular velocity of the mobile frame and the output, $Y \in \mathsf{SO}(3)$, is its angular position. The reason for using this kinematic model was originally to fuse measurements from two sensors that are mounted on the mobile frame: 1) an orientation sensor, which measures the angular position and 2) a set of tri-axial laser ring gyroscopes, which measure angular velocity. The LFSO's knowledge of the angular velocity of the mobile frame was the extra piece of information that allowed it to efficiently filter out noise from the measurement of the orientation.

The angular velocity of the mobile frame can be measured by laser ring gyroscopes, mounted on the mobile frame. As long as the laser ring gyroscopes can accurately measure the angular velocity of the mobile frame, our proposed LFSO of Section 3.7 can be used to estimate the orientation of the mobile frame. Unfortunately, laser ring gyroscopes become progressively less accurate, as the angular velocity of the mobile frame increases. All laser ring gyroscopes typically have a "saturation" angular speed, after which the gyroscope becomes inaccurate. Therefore, when the mobile frame is spinning very fast, the LFSO of Section 3.7 will become inaccurate. In such a scenario, we can use another type of sensor - angular accelerometer, which can measure angular acceleration accurately even when the mobile frame is spinning very fast.

Thus, suppose our observer does not have knowledge of the angular velocity of the mobile frame, but it does have knowledge of the angular acceleration. To solve this sensor fusion problem, we can no longer rely on the kinematic rigid-body model (3.24). We must modify this model in such a way that the input to the system is the angular acceleration and the output of the system is the angular position. This is done by adding a state to encode the angular velocity, call it $\omega \in \mathrm{Lie}(\mathsf{SO}\,(3))$. The input $u \in \mathrm{Lie}(\mathsf{SO}\,(3))$ is then defined to be the time derivative of ω, i.e., the angular acceleration of the mobile frame. Doing this we get

$$\begin{aligned}\dot{R} &= R\omega \\ \dot{\omega} &= u \\ Y &= R.\end{aligned}$$

This new rigid-body model is no longer kinematic, but is now dynamic. The part of the state space that is directly measurable is the rotation matrix $R \in \mathsf{SO}\,(3)$, which can be obtained using exteroseptive sensors, such as on-board cameras. In this context our observer receives a measurement of the orientation $Y = R$ and a measurement of the angular acceleration $u \in \mathrm{Lie}(\mathsf{SO}\,(3))$. The job of the observer is now to fuse measurements Y and u, to estimate the angular velocity ω, and also to filter out noise from Y. Since the angular velocity is not directly measurable, this type of observer falls under the class of partial-state observers. We call it a Lie group partial-state observer (LPSO), since part of its state evolves on a Lie group.

The above example is not the only application of our proposed LPSOs, it is only a special case of our LPSO when the Lie group is taken to be $G = \mathsf{SO}(3)$. However the example of dynamic orientation estimation allows us to more easily visualize the problem that is being solved and to see the importance of the problem. In fact, the LPSOs proposed in this chapter work not only for $\mathsf{SO}(3)$, but for any linear Lie group.

4.2 Problem Statement

Let G be any linear Lie group, then as we have already seen, the Lie algebra, $\mathrm{Lie}(G)$, is a vector space and a subspace of $\mathsf{M}(n, \mathbb{R})$, therefore its tangent space at any point is isomorphic to the Lie algebra itself. So if $x : \mathbb{R} \to \mathrm{Lie}(G)$ is any smooth curve, then its derivative, $\dot{x}$, is also a curve in $\mathrm{Lie}(G)$, i.e., $\dot{x} : \mathbb{R} \to \mathrm{Lie}(G)$. Now, consider the following system with dynamics

$$\begin{aligned} \dot{X} &= X x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_d &= u \\ Y &= X, \end{aligned} \tag{4.1}$$

where $X \in G$ is the state that evolves on Lie group, $x_2 \in \mathrm{Lie}(G)$ is the body-velocity of X, $x_3 \in \mathrm{Lie}(G)$ is the body-acceleration of X, and so on. The input to (4.1) is $u : \mathbb{R} \to \mathrm{Lie}(G)$ which we assume to be a smooth, uniformly bounded and globally Lipschitz signal of time. The output of (4.1) is $Y = X \in G$ which will, in practice, be corrupted by noise.

The following assumption, almost identical to Assumption 1 from Chapter 3, will be assumed to hold throughout this chapter.

Assumption 2. *The input u to system (4.1) is such that for any initial conditions $X(0) \in G$ and $x_2(0), \ldots, x_d(0) \in \mathrm{Lie}(G)$, there exists a compact set $\mathcal{G} \subseteq G$ such that $X(t) \in \mathcal{G}$, for all $t \geq 0$.*

Note that this assumption is automatically satisfied, for all admissible inputs, if the group G itself is compact, for example if $G = \mathsf{SO}(3)$.

Our objective is to build an observer, with states $\hat{X} \in G$ and $\hat{x}_2, \ldots, \hat{x}_d \in \mathrm{Lie}(G)$, for the plant (4.1). The observer has access to the measured output $Y \in G$ and to the input $u \in \mathrm{Lie}(G)$. We will design the observer such that, under Assumption 2, if $\|\hat{X}(0) - X(0)\|$, $\|\hat{x}_2(0) - x_2(0)\|$, $\ldots$, $\|\hat{x}_d(0) - x_d(0)\|$ are sufficiently small, then $\|\hat{X}(t) - X(t)\| \to 0$, $\|\hat{x}_2(t) - x_2(t)\| \to 0$, $\ldots$, $\|\hat{x}_d(t) - x_d(t)\| \to 0$ exponentially, as $t \to \infty$.

4.3 Proposed Observers

For the system (4.1), we propose two different partial-state observers. The first is the direct LPSO, given by

$$\begin{aligned} \dot{\hat{X}} &= Y\hat{x}_2 Y^{-1}\hat{X} - a_{d-1}\hat{X}\log(Y^{-1}\hat{X}) \\ \dot{\hat{x}}_2 &= \hat{x}_3 - a_{d-2}\log(Y^{-1}\hat{X}) \\ &\vdots \\ \dot{\hat{x}}_{d-1} &= \hat{x}_d - a_1\log(Y^{-1}\hat{X}) \\ \dot{\hat{x}}_d &= u - a_0\log(Y^{-1}\hat{X}) \end{aligned} \tag{4.2}$$

and the second is the passive LPSO, given by

$$\begin{aligned} \dot{\hat{X}} &= \hat{X}\hat{x}_2 - a_{d-1}\hat{X}\log(Y^{-1}\hat{X}) \\ \dot{\hat{x}}_2 &= \hat{x}_3 - a_{d-2}\log(Y^{-1}\hat{X}) \\ &\vdots \\ \dot{\hat{x}}_{d-1} &= \hat{x}_d - a_1\log(Y^{-1}\hat{X}) \\ \dot{\hat{x}}_d &= u - a_0\log(Y^{-1}\hat{X}). \end{aligned} \tag{4.3}$$

In the above two observers, the constants $a_0, \ldots, a_{d-1} \in \mathbb{R}$ are design parameters, chosen such that the polynomial $p(s) = s^d + a_{d-1}s^{d-1} + \cdots + a_1 s + a_0$ is Hurwitz. These design parameters can be used to modify the rate of convergence of the estimation error.

We will show that the direct LPSO (4.2) is locally exponentially stable for the system (4.1), but unfortunately we are not able to show exponential stability of the passive LPSO (4.3). However, we will see in simulation that the passive LPSO appears to work quite well in the presence of measurement noise.

4.4 Estimation Error Functions

To quantify the estimation error between X and $\hat{X}$, we use the Lie group canonical invariant error functions E_l and E_r, which we used in Chapter 3. The matrices x_i and $\hat{x}_i$, for $i = 2, \ldots, d$ are vectors in $\mathrm{Lie}(G)$, so to quantify the error between x_i and $\hat{x}_i$, we can subtract them as vectors, i.e.,

$$e_i := x_i - \hat{x}_i, \quad i = 2, \ldots, d. \tag{4.4}$$

Since the matrices x_i and $\hat{x}_i$ are elements of the vector space $\mathrm{Lie}(G)$, the estimation error e_i is also an element of $\mathrm{Lie}(G)$.

4.5 A Differential Equation on $\mathsf{GL}(n, \mathbb{R})$ and $\mathsf{M}(n, \mathbb{R})$

Here we restrict our attention to examining the behaviour of the following differential equation, which is a natural extension of the differential equation (3.10),

$$\begin{aligned} \dot{E} &= e_2 E - a_{d-1} E \log(E) \\ \dot{e}_2 &= e_3 - a_{d-2} \log(E) \\ &\vdots \\ \dot{e}_{d-1} &= e_d - a_1 \log(E) \\ \dot{e}_d &= -a_0 \log(E), \end{aligned} \tag{4.5}$$

where $E \in \mathsf{GL}(n, \mathbb{R})$, $e_i \in \mathsf{M}(n, \mathbb{R})$ for $i = 2, \dots, d$ and $a_0, \dots, a_{d-1} \in \mathbb{R}$ are constants such that the polynomial $p(s) = s^d + a_{d-1}s^{d-1} + \cdots + a_1 s + a_0$ is Hurwitz. The differential equation (4.5) arises in the analysis of the error dynamics associated with the direct LPSO (4.2). We have only defined the log map on the domain $B(I_n, 1)$. Therefore the differential equation (4.5) is only defined when $E \in B(I_n, 1)$.

Remark 4.5.1. *Let $G \subseteq \mathsf{GL}(n, \mathbb{R})$ be any linear Lie group, then the embedded submanifold, $S := G \times \mathrm{Lie}(G) \times \cdots \times \mathrm{Lie}(G)$, in the state space of (4.5) is positively invariant under the dynamics (4.5). To see this, we just check that, if $p = (E, e_2, \dots, e_d)$ is any point in S, then the vector field (4.5), evaluated at p, lies in the tangent space to S at p.*

Indeed, by Proposition 2.2.12, we have that $E \log(E) \in T_E G$ and that $e_2 E \in T_E G$. Therefore, the vector $\dot{E} = e_2 E - a_{d-1} E \log(E)$ is in the tangent space to G at E, i.e., $\dot{E} \in T_E G$. Furthermore we have that, for $i = 2, \dots, d$, $\dot{e}_i \in \mathrm{Lie}(G) \cong T_{e_i} \mathrm{Lie}(G)$, because $\mathrm{Lie}(G)$ is a vector space. So the following holds

$$\begin{gathered} E \in G, \quad e_2 \in \mathrm{Lie}(G), \quad \dots, \quad e_d \in \mathrm{Lie}(G) \\ \Downarrow \\ \dot{E} \in T_E G, \quad \dot{e}_2 \in T_{e_2} \mathrm{Lie}(G), \quad \dots, \quad \dot{e}_d \in T_{e_d} \mathrm{Lie}(G), \end{gathered}$$

Therefore, the vector field (4.5) is tangent to the submanifold S.

The matrices E and $\dot{E}$ in (4.5) are generally non-commuting matrices, i.e., $[E, \dot{E}] \neq 0$. This is because E and e_2 are generally non-commuting matrices, i.e.,

$$\begin{aligned}\left[E, \dot{E}\right] &= [E, e_2 E - a_1 E \log(E)] \\ &= [E, e_2 E] \\ &= E e_2 E - e_2 E^2 \\ &= [E, e_2]\, E.\end{aligned}$$

When we try to convert the differential equation (4.5) into log coordinates, i.e., define $e_1 := \log(E)$ and find $\dot{e}_1$, we are not able to do it as easily as we did with equation (3.10) in Section 3.5, due to the non-commutativity of E and $\dot{E}$. Recall that in Section 3.5, we were able to easily differentiate the series expansion of $\log(E)$, when $E(t)$ was a solution of (3.10), because the matrices E and $\dot{E}$ were commuting in that case. When we try to repeat the same approach with $E(t)$ being a solution of (4.5) we are not able to get a closed form expression for $\dot{e}_1$, because E and $\dot{E}$ do not commute. We do not know how to convert the differential equation (4.5) into log coordinates, the following provides a clue.

Proposition 4.5.2. *For $E \in B(I_n, 1)$, the differential equation (4.5) is differentially equivalent to*

$$\begin{aligned}\dot{e}_1 &= e_2 - a_{d-1} e_1 + K(e_1, e_2) \\ \dot{e}_2 &= e_3 - a_{d-2} e_1 \\ &\vdots \\ \dot{e}_{d-1} &= e_d - a_1 e_1 \\ \dot{e}_d &= -a_0 e_1,\end{aligned}$$

where $e_1 := \log(E)$ and $K : \mathrm{Lie}(G) \times \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ is a function, which vanishes for any commuting e_1 and e_2, i.e.,

$$[e_1, e_2] = 0 \quad \Rightarrow \quad K(e_1, e_2) = 0.$$

Proof. Since $e_1 = \log(E)$ is a Taylor series only in E, when we differentiate this Taylor series we find that, $\dot{e}_1$ only depends on E and $\dot{E}$. Furthermore, from (4.5), we know that $\dot{E}$ only depends on E and e_2. Thus, using $E = \exp(e_1)$, we have that $\dot{e}_1$ only depends on e_1 and e_2. Let $K(e_1, e_2) := \dot{e}_1 - e_2 + a_{d-1} e_1$.

Assume that e_1 and e_2 commute, this implies that $E = \exp(e_1)$ and e_2 also commute and this implies that E and $\dot{E}$ commute. Since $E\dot{E} = \dot{E}E$, we can repeat almost the same analysis that we used in Section 3.5, doing this we get

$$\begin{aligned}\dot{e}_1 &= \dot{E}E^{-1}\\ &= e_2 - a_{d-1}e_1,\end{aligned}$$

therefore $K(e_1, e_2) = 0$ for any commuting e_1 and e_2.

The expressions of $\dot{e}_i$ for $i = 2, \ldots, d$ are computed by substituting $\log(E) = e_1$ into (4.5). □

This is as far as we want to go in trying to convert (4.5) into log coordinates, from now on we focus on the simpler task of analyzing the stability of (4.5).

Lemma 4.5.3. *The equilibrium point $(E, e_2, \ldots, e_d) = (I_n, 0, \ldots, 0)$ of the differential equation (4.5) is locally exponentially stable if the constants $a_0, \ldots, a_{d-1} \in \mathbb{R}$ are chosen such that the polynomial $p(s) = s^d + a_{d-1}s^{d-1} + \cdots + a_1 s + a_0$ is Hurwitz.*

Proof. Adapting the proof of [5, Theorem 3.1 (ii)] and using Lemma A.3.2, we will show that (4.5) is locally exponentially stable at the equilibrium point, by showing that its linearization, around the equilibrium point, is exponentially stable.

We define $\delta E, \delta e_2, \ldots, \delta e_d$ to be elements of $\mathrm{Lie}(G)$ corresponding to the first order approximations of $E, e_2, \ldots, e_d$, respectively, around the equilibrium point $(I_n, 0_n, \ldots, 0_n)$:

$$\delta E := E - I_n, \qquad \delta e_2 := e_2 - 0_n, \qquad \ldots, \qquad \delta e_d := e_d - 0_n.$$

Using the series definition of the matrix logarithm (2.3)

$$\log(E) = (E - I_n) - \frac{1}{2}(E - I_n)^2 + \cdots$$

we deduce that near $\delta E = 0$

$$\log(E) \approx \delta E.$$

Similarly, using $E = \delta E + I_n$, and dropping higher order terms in δE, we get

$$E\log(E) = (\delta E + I_n)\left((\delta E) - \frac{1}{2}(\delta E)^2 + \cdots\right) \approx \delta E.$$

Similarly, using $e_2 = \delta e_2$, and dropping higher order terms in δE and δe_2, we get

$$e_2 E = (\delta e_2)(\delta E + I_n) \approx \delta e_2.$$

Substituting these approximations into the differential equation (4.5), we get the linearization of (4.5) at the equilibrium $(I_n, 0_n, \ldots, 0_n)$

$$\frac{d}{dt}\begin{bmatrix} \delta E \\ \delta e_2 \\ \delta e_3 \\ \vdots \\ \delta e_{d-1} \\ \delta e_d \end{bmatrix} = \begin{pmatrix} -a_{d-1}I_n & I_n & 0_n & \ldots & 0_n & 0_n \\ -a_{d-2}I_n & 0_n & I_n & \ldots & 0_n & 0_n \\ -a_{d-3}I_n & 0_n & 0_n & \ldots & 0_n & 0_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_1 I_n & 0_n & 0_n & \ldots & 0_n & I_n \\ -a_0 I_n & 0_n & 0_n & \ldots & 0_n & 0_n \end{pmatrix} \begin{bmatrix} \delta E \\ \delta e_2 \\ \delta e_3 \\ \vdots \\ \delta e_{d-1} \\ \delta e_d \end{bmatrix}.$$

The above linear system has its eigenvalues located at the roots of the polynomial $p(s) = s^d + a_{d-1}s^{d-1} + \cdots + a_1 s + a_0$, with multiplicity n, for each (possibly repeating) root of $p(s)$. Since all the eigenvalues have negative real parts, the linearization above is stable by Lemma A.3.1. Therefore the equilibrium point, $(E, e_2, \ldots, e_d) = (I_n, 0_n, \ldots, 0_n)$, of (4.5) is locally exponentially stable, by Lemma A.3.2. □

4.6 Estimation Error Dynamics

4.6.1 Direct Observer

Consider what happens when we apply the direct LPSO (4.2) to estimate the state of the system (4.1). For the direct observer, we use the left-invariant error $E_l = X^{-1}\hat{X}$ to quantify the estimation error between X and $\hat{X}$. Let us begin by computing the dynamics of E_l

$$\begin{aligned} \dot{E}_l &= \frac{d}{dt}\left[X^{-1}\hat{X}\right] \\ &= -X^{-1}\dot{X}X^{-1}\hat{X} + X^{-1}\dot{\hat{X}} \\ &= -x_2 X^{-1}\hat{X} + \hat{x}_2 X^{-1}\hat{X} - a_{d-1}X^{-1}\hat{X}\log(X^{-1}\hat{X}) \\ &= (\hat{x}_2 - x_2)X^{-1}\hat{X} - a_{d-1}X^{-1}\hat{X}\log(X^{-1}\hat{X}) \\ &= e_2 E_l - a_{d-1}E_l\log(E_l). \end{aligned} \tag{4.6}$$

Next, we compute the dynamics of e_i, for $i = 2, \ldots, d-1$

$$\begin{aligned} \dot{e}_i &= \frac{d}{dt}\left[\hat{x}_i - x_i\right] \\ &= \hat{x}_{i+1} - a_{d-i}\log(X^{-1}\hat{X}) - x_{i+1} \\ &= e_{i+1} - a_{d-i}\log(E_l). \end{aligned} \tag{4.7}$$

Finally, we compute the dynamics of e_d

$$\begin{aligned}
\dot{e}_d &= \frac{d}{dt}\left[\hat{x}_d - x_d\right] \\
&= u - a_0 \log(X^{-1}\hat{X}) - u \\
&= -a_0 \log(E_l).
\end{aligned} \tag{4.8}$$

Putting together results (4.6) , (4.7) and (4.8), we obtain the following complete estimation error dynamics between the system (4.1) and the direct LPSO (4.2),

$$\begin{aligned}
\dot{E}_l &= e_2 E_l - a_{d-1} E_l \log(E_l) \\
\dot{e}_2 &= e_3 - a_{d-2} \log(E_l) \\
&\ \vdots \\
\dot{e}_{d-1} &= e_d - a_1 \log(E_l) \\
\dot{e}_d &= -a_0 \log(E_l).
\end{aligned} \tag{4.9}$$

The above differential equation is the same as the equation (4.5), if we identify E with E_l. This means that, by Lemma 4.5.3, the equilibrium point $(E_l, e_2, \ldots, e_d) = (I_n, 0_n, \ldots, 0_n)$ of (4.9) is locally exponentially stable.

4.6.2 Passive Observer

Consider what happens when we apply the passive observer (4.3) to estimate the state of the system (4.1). For the passive observer, we use the right-invariant error $E_r = \hat{X}X^{-1}$ to quantify the estimation error between X and $\hat{X}$. Let us compute the dynamics of E_r

$$\begin{aligned}
\dot{E}_r &= \frac{d}{dt}\left[\hat{X}X^{-1}\right] \\
&= \dot{\hat{X}}X^{-1} - \hat{X}X^{-1}\dot{X}X^{-1} \\
&= \hat{X}\hat{x}_2X^{-1} - a_{d-1}\hat{X}\log(X^{-1}\hat{X})X^{-1} - \hat{X}x_2X^{-1} \\
&= \hat{X}(\hat{x}_2 - x_2)X^{-1} - a_{d-1}\hat{X}X^{-1}X\log(X^{-1}\hat{X})X^{-1} \\
&= \hat{X}X^{-1}X(\hat{x}_2 - x_2)X^{-1} - a_{d-1}\hat{X}X^{-1}\log(\hat{X}X^{-1}) \\
&= E_r X e_2 X^{-1} - a_{d-1}E_r \log(E_r) \\
&= E_r \operatorname{Ad}_X(e_2) - a_{d-1}E_r \log(E_r).
\end{aligned} \tag{4.10}$$

The dynamics of e_i, for $i = 2, \ldots, d$ are computed just as we did in (4.7) and (4.8). Therefore we have the following complete estimation error dynamics between the system (4.1) and the passive LPSO (4.3)

$$\begin{aligned}
\dot{E}_r &= E_r \operatorname{Ad}_X(e_2) - a_{d-1} E_r \log(E_r) \\
\dot{e}_2 &= e_3 - a_{d-2} \log(E_r) \\
&\vdots \\
\dot{e}_{d-1} &= e_d - a_1 \log(E_r) \\
\dot{e}_d &= -a_0 \log(E_r).
\end{aligned} \tag{4.11}$$

The above differential equation is not the same as the equation (4.5), because of the term $\operatorname{Ad}_X(e_2)$. Therefore, we can not apply Lemma 4.5.3 to deduce the stability of the equilibrium point $(E_r, e_2, \ldots, e_d) = (I_n, 0_n, \ldots, 0_n)$. Unfortunately, we are not able to prove the stability of this error dynamics, but we will see in simulation that the passive LPSO appears to work quite well and in fact appears to work better than the direct observer, when a large amount of measurement noise is present in Y.

4.6.3 Discussion

We have seen that the direct LPSO (4.2) produces local exponential convergence of the left-invariant Lie group estimation error, $E_l \to I_n$, and of the Lie algebra estimation errors, $e_i \to 0_n$, for $i = 2, \ldots, d$. Therefore, by Proposition 3.4.2, if Assumption 2 holds, then the direct observer is such that $\hat{X} \to X$ and $\hat{x}_i \to x_i$, for $i = 2, \ldots, d$, if the observer is initialized sufficiently close to the system. Unfortunately we are not able to prove exponential stability of the error dynamics E_r for the passive LPSO (4.3). However we conjecture that the passive LPSO is locally exponentially convergent if Assumption 2 holds. This conjecture will be supported by simulation, where the passive LPSO seems to perform better than the direct LPSO, when a large amount of measurement noise is present in Y.

4.7 Examples

4.7.1 Scalar Lie Group $\mathsf{GL}(1, \mathbb{R})$

The Lie group $\mathsf{GL}(1, \mathbb{R})$ is the group of non-zero real numbers, with the group operation being multiplication. Consider the following system on $\mathbb{R}^2$, which arises when calculating continously compounded interest,

$$\begin{aligned}\dot{x}_1 &= x_1 x_2 \\ \dot{x}_2 &= u \\ y &= x_1,\end{aligned} \tag{4.12}$$

where $x_1 \in \mathbb{R}/\{0\} = \mathsf{GL}(1,\mathbb{R})$, $x_2 \in \mathbb{R} = \mathrm{Lie}(\mathsf{GL}(1,\mathbb{R}))$ and $u \in \mathbb{R} = \mathrm{Lie}(\mathsf{GL}(1,\mathbb{R}))$. The state x_1 represents the instantaneous amount of money invested and the state x_2 represents the instantaneous interest rate, that the investment generates. It is unclear to us how building observers for the above system can be useful in solving any practical problem, however this system offers an interesting example of a particularly simple LPSO.

The system (4.12) is nonlinear and it can be viewed as a system of class (4.1), with $G = \mathsf{GL}(1,\mathbb{R})$ and $d = 2$. Therefore, the system (4.12) admits the following LPSO

$$\begin{aligned}\dot{\hat{x}}_1 &= \hat{x}_1\hat{x}_2 - a_1\hat{x}_1 \log(y^{-1}\hat{x}_1) \\ \dot{\hat{x}}_2 &= u - a_0 \log(y^{-1}\hat{x}_1),\end{aligned} \tag{4.13}$$

which is both a direct and a passive LPSO, since $\mathsf{GL}(1,\mathbb{R})$ is a commutative Lie group.

The system (4.12) can be converted into the form (2.6) by a change of coordinates:

$$z_1 := x_1 \qquad \text{and} \qquad z_2 := x_1 x_2.$$

In z-coordinates, the system (4.12) takes on the form

$$\begin{aligned}\dot{z}_1 &= z_2 \\ \dot{z}_2 &= \frac{z_2^2}{z_1} + z_1 u \\ y &= z_1.\end{aligned}$$

The above system admits an HGO of type (2.7), in z-coordinates,

$$\begin{aligned}\dot{\hat{z}}_1 &= \hat{z}_2 + \frac{a_1}{\epsilon}(y - \hat{z}_1) \\ \dot{\hat{z}}_2 &= \frac{\hat{z}_2^2}{\hat{z}_1} + \hat{z}_1 u + \frac{a_0}{\epsilon^2}(y - \hat{z}_1).\end{aligned}$$

We convert this HGO back into x-coordinates

$$\begin{aligned}\dot{\hat{x}}_1 &= \hat{x}_1\hat{x}_2 + \frac{a_1}{\epsilon}(y - \hat{x}_1) \\ \dot{\hat{x}}_2 &= u + \left(\frac{1}{\hat{x}_1}\frac{a_0}{\epsilon^2} - \frac{\hat{x}_2}{\hat{x}_1}\frac{a_1}{\epsilon}\right)(y - \hat{x}_1).\end{aligned} \tag{4.14}$$

Thus, we have obtained two nonlinear observers for the system (4.12), the first observer is the LPSO (4.13) and the second observer is the HGO (4.14). Let us compare these two observers in simulation. For the observer parameters, we take the following constants

$$a_0 = 1, \qquad a_1 = 2, \qquad \epsilon = 1.$$

The results of the simulation are shown in Figure 4.1. From the simulation, it appears that the two observers have very similar performance, we can not conclusively say which one of the observers is better.

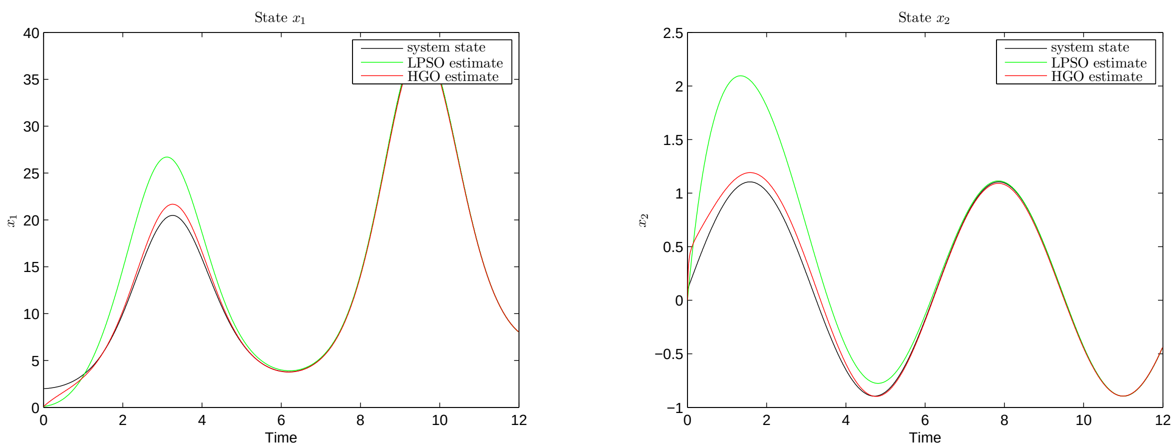

Figure 4.1: Estimation of system (4.12) by means of LPSO (4.13) and HGO (4.14).

4.7.2 Dynamic Rigid-Body Orientation Estimation on $\mathsf{SO}(3)$

The system that we use to model the dynamics of rigid-body orientation is

$$\begin{aligned} \dot{R} &= R\omega \\ \dot{\omega} &= u \\ Y &= R, \end{aligned} \tag{4.15}$$

where $R \in \mathsf{SO}(3)$ is the rotation matrix, which encodes the orientation of the rigid-body, $\omega \in \mathrm{Lie}(\mathsf{SO}(3))$ is the skew-symmetric matrix, which encodes the angular velocity and $u \in \mathrm{Lie}(\mathsf{SO}(3))$ is the skew-symmetric matrix, which encodes the angular acceleration.

A similar type of system was mentioned by Brockett in [34]. In that paper, Example 2 gives a model for rigid-body dynamics, with torque as the input and orientation as the output. If we simplify this rigid-body model, so that input is taken to be angular acceleration, rather than torque, then then we get the model (4.15). Building an observer for this system allows us to estimate rigid-body orientation and angular velocity by using angular accelerometer sensors and an orientation sensor, such as a camera.

Applying the results of this chapter, we have the following direct LPSO

$$\begin{aligned} \dot{\hat{R}} &= Y\hat{\omega}Y^{-1}\hat{R} - a_1\hat{R}\log(Y^{-1}\hat{R}) \\ \dot{\hat{\omega}} &= u - a_0\log(Y^{-1}\hat{R}) \end{aligned} \tag{4.16}$$

and the following passive LPSO

$$\begin{aligned} \dot{\hat{R}} &= \hat{R}\hat{\omega} - a_1\hat{R}\log(Y^{-1}\hat{R}) \\ \dot{\hat{\omega}} &= u - a_0\log(Y^{-1}\hat{R}). \end{aligned} \tag{4.17}$$

We will now simulate the direct and the passive LPSOs, with increasing amounts of noise in the output. The initial conditions for the plant and the observer are chosen to be the following matrices

$$R(0) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \omega(0) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}, \quad \hat{R}(0) = I_3, \quad \hat{\omega}(0) = 0_3.$$

The angular acceleration input is chosen (somewhat arbitrarily) to be

$$u(t) = \begin{pmatrix} 0 & -2\sin(t) & \cos(t) \\ 2\sin(t) & 0 & -\sin(t) \\ -\cos(t) & \sin(t) & 0 \end{pmatrix}.$$

The observer gains are chosen as

$$a_0 = 1, \qquad a_1 = 2.$$

Noise is injected into the output via the random rotation matrix $N \in \mathsf{SO}(3)$, by setting $Y = RN$. The matrix N is generated as in Section 3.7. The simulation results are shown in Figure 4.2. From the simulation we can see that, when there is no measurement noise, the direct LPSO appears to converge faster than the passive LPSO. However, with a large amount of measurement noise, the passive LPSO appears to be comparable to the direct LPSO.

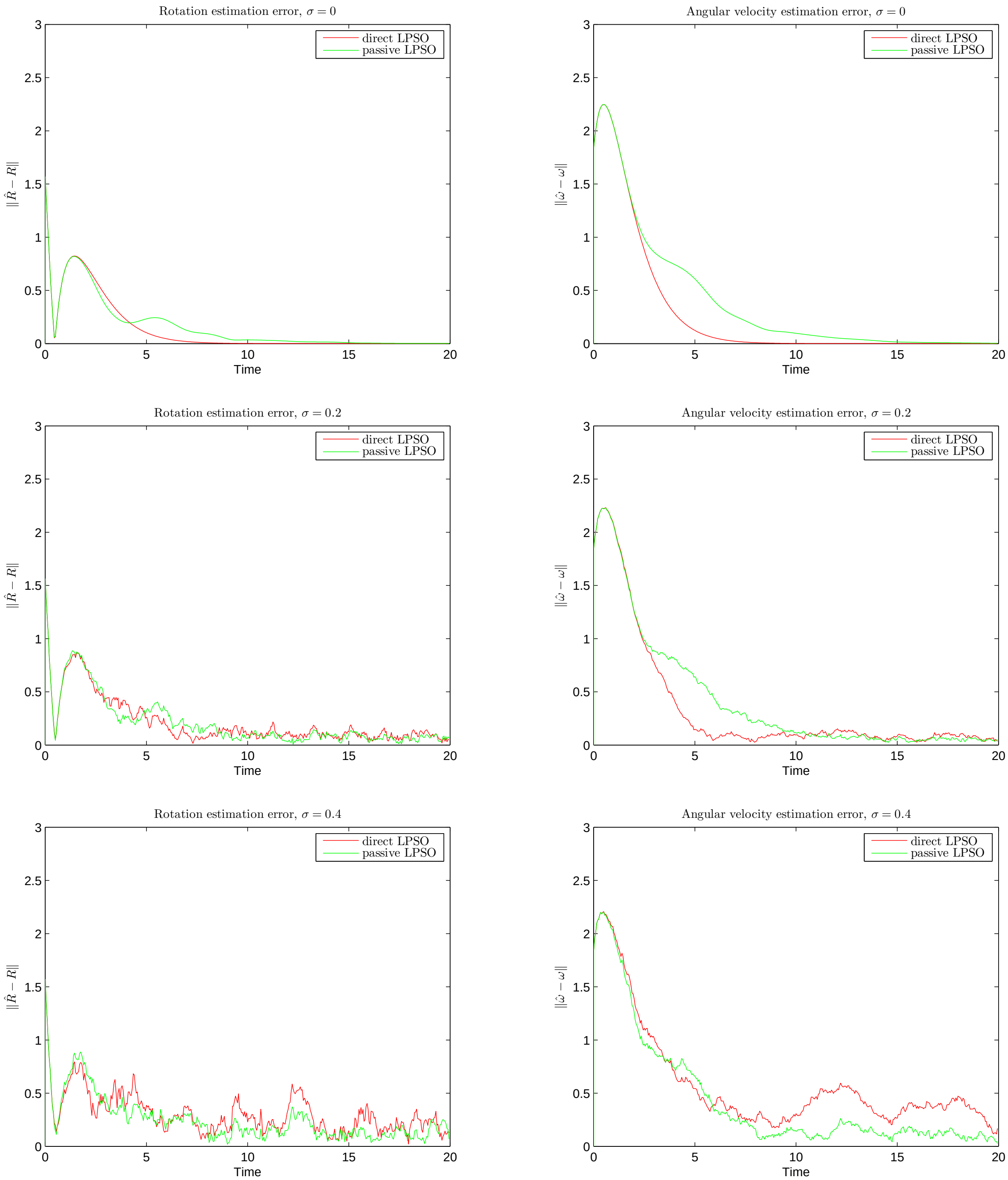

Figure 4.2: Direct and Passive LPSOs, with increasing amounts of measurement noise.

4.7.3 Dynamic Homography Estimation on $\mathsf{SL}(3)$

Many control laws for mobile robots are designed based on visual information provided by a camera. A homography is an invertible mapping between two images, made by two different cameras, looking at the same planar scene. Relative homography between any two cameras can be estimated based on the images these two cameras take of the same planar scene. Algorithms exist to decompose a homography into the corresponding camera pose (rotation and translation). The details of these decomposition algorithms are beyond the scope of this chapter, see [35] and [36] for details. As an alternative to doing this decomposition, it was proposed in [37] to design control laws based directly on the homography itself. The set of all homographies forms a Lie group, with composition as the group operation. Intuitively this implies that if there are three cameras: A, B and C, then the homography from A to B, composed with the homography from B to C, is the same as the homography from A to C. A key component of the work in [37] is the existence of a group isomorphism between the Lie group of homographies and the group of invertible 3×3 matrices with unit determinant. These matrices form the special linear Lie group, $\mathsf{SL}(3)$. In this section, we propose an LPSO on $\mathsf{SL}(3)$ that solves exactly the same problem as the nonlinear observer designed in [5], [24].

Consider the homography dynamics, evolving on the Lie group $\mathsf{SL}(3)$ and on its Lie algebra, using the notation of [24],

$$\begin{aligned} \dot{H} &= HA \\ \dot{A} &= 0 \\ Y &= H, \end{aligned} \tag{4.18}$$

where $H \in \mathsf{SL}(3)$ describes the homography transformation between two cameras, $A \in \mathrm{Lie}(\mathsf{SL}(3))$ is the body-velocity of H, this body-velocity is unknown, but is assumed to be constant. The model (4.18) fits the class (4.1), with $X = H$, $x_2 = A$, $u = 0$, and $d = 2$. We do not know the physical meaning of the body-velocity, A, or of the body-acceleration, u, that is why we assume that A is constant and that u is equal to zero (as do the authors of [24]). If we had some way of knowing u, then we would not have to assume that $u = 0$ and our homography estimator would be improved.

The idea is that we have some way to compute H by analyzing the images taken by the cameras, hence the output of the system (4.18) is the homography $Y = H$. The computation of the homography, based on image data, will not be perfect and there will be some measurement noise in Y. So the first goal of our LPSO is to filter out noise from Y. The second goal of our LPSO is to estimate the body-velocity, A, which is assumed to be constant.

Applying the results of this chapter, we have the following direct LPSO, for the system (4.18),

$$\begin{aligned}\dot{\hat{H}} &= Y\hat{A}Y^{-1}\hat{H} - a_1\hat{H}\log(Y^{-1}\hat{H})\\ \dot{\hat{A}} &= -a_0\log(Y^{-1}\hat{H}),\end{aligned} \tag{4.19}$$

which is similar to the projection based observer of [24]. In addition, we also have the following passive LPSO, for the system (4.18),

$$\begin{aligned}\dot{\hat{H}} &= \hat{H}\hat{A} - a_1\hat{H}\log(Y^{-1}\hat{H})\\ \dot{\hat{A}} &= -a_0\log(Y^{-1}\hat{H}).\end{aligned} \tag{4.20}$$

As usual, the constants a_0, a_1 in the above two observers are design parameters, chosen such that the polynomial $p(s) = s^2 + a_1 s + a_0$ is Hurwitz.

We now use numerical simulation to compare our two LPSOs with the projection based observer of [24], which is written in expanded form as

$$\begin{aligned}\dot{\hat{H}} &= Y\hat{A}Y^{-1}\hat{H} - a_1 Y\mathbb{P}(Y^\top\hat{H}^{-\top}(I_3 - \hat{H}^{-1}Y))Y^{-1}\hat{H}\\ \dot{\hat{A}} &= -a_0\mathbb{P}(Y^\top\hat{H}^{-\top}(I_3 - \hat{H}^{-1}Y)),\end{aligned} \tag{4.21}$$

where $\mathbb{P}(H)$ is the projection from the Lie group $\mathsf{SL}(3)$ to its Lie algebra Lie($\mathsf{SL}(3)$), with respect to the Euclidean matrix inner product $\ll A, B \gg = \operatorname{tr}(A^\top B)$,

$$\mathbb{P}(H) := \left(H - \frac{\operatorname{tr}(H)}{3}I_3\right) \in \mathrm{Lie}(\mathsf{SL}(3)).$$

Due to our assumption that $\dot{A} = 0$, we have that $\forall t > 0$:

$$A(t) = A(0) \qquad \text{and} \qquad H(t) = H(0)\exp(A(0)t).$$

The gains, a_0 and a_1, of the observers are chosen as

$$a_0 = 1, \qquad a_1 = 2.$$

The initial conditions of the plant are chosen as

$$H(0) = \begin{pmatrix} 0.2189 & 1.3770 & -0.7845\\ -1.3770 & 4.9422 & -1.3770\\ -0.7845 & 1.3770 & 0.2189\end{pmatrix}, \quad A(0) = \begin{pmatrix} 0.1 & -0.1 & 0.1\\ 0.1 & -0.2 & 0.1\\ 0.1 & -0.1 & 0.1\end{pmatrix},$$

which are chosen such that the state H of the plant starts far away from the identity matrix, then passes close to the identity matrix during the middle of the simulation (at $t = 10$) and then goes off far away from the identity matrix again. It was discovered, during the course of the simulation, that if the state H of the plant is very far away from the identity, then measurement noise starts to have a very detrimental effect on all three of the observers. For this reason, we designed the system trajectory such that H does not get extremely far away from the identity matrix during the time period from $t = 0$ to $t = 20$. The initial conditions of the observer are chosen as

$$\hat{H}(0) = I_3, \quad \hat{A}(0) = 0_3.$$

We model measurement noise by setting $Y = HN$ where $N \in \mathsf{SL}(3)$ is a randomly generated $\mathsf{SL}(3)$ matrix, $N = \exp(n)$, where $n \in \mathrm{Lie}(\mathsf{SL}(3))$ is Gaussian distributed. This ensures that our output, Y, is always an element of $\mathsf{SL}(3)$. We model the system as being a deterministic process and the only place where randomness appears is in the measurement of the output. The results of the simulation are shown in Figure 4.3. From the plots, we can see that our proposed LPSOs appear to work better than the projection based observer of [24] in the presence of measurement noise. Furthermore, it was found that the projection based observer of [24] has a tendency to escape to infinity when noise is increased further from what is shown in the plots.

The Lie group estimation errors, E_l and E_r, are initially (at time $t = 0$)

$$\begin{aligned} E_l(0) &= H^{-1}(0)\hat{H}(0) = H^{-1}(0) \\ E_r(0) &= \hat{H}(0)H^{-1}(0) = H^{-1}(0). \end{aligned}$$

Calculating the distance between $E_l(0)$ or $E_r(0)$ and I_3, we get

$$\begin{aligned} \|E_l(0) - I\| &= \|E_r(0) - I\| \\ &= \|H^{-1}(0) - I\| \\ &= \left\| \begin{pmatrix} xx & xx & xx \\ xx & xx & xx \\ xx & xx & xx \end{pmatrix} - I_3 \right\| \\ &= xx, \end{aligned}$$

which shows that $E_l(0), E_r(0)$ are initially quite far from I_3. In spite of this, the simulation shows that our LPSOs still converge to the system trajectory, which seems to suggest that local exponential stability of the estimation error is actually a very conservative statement.

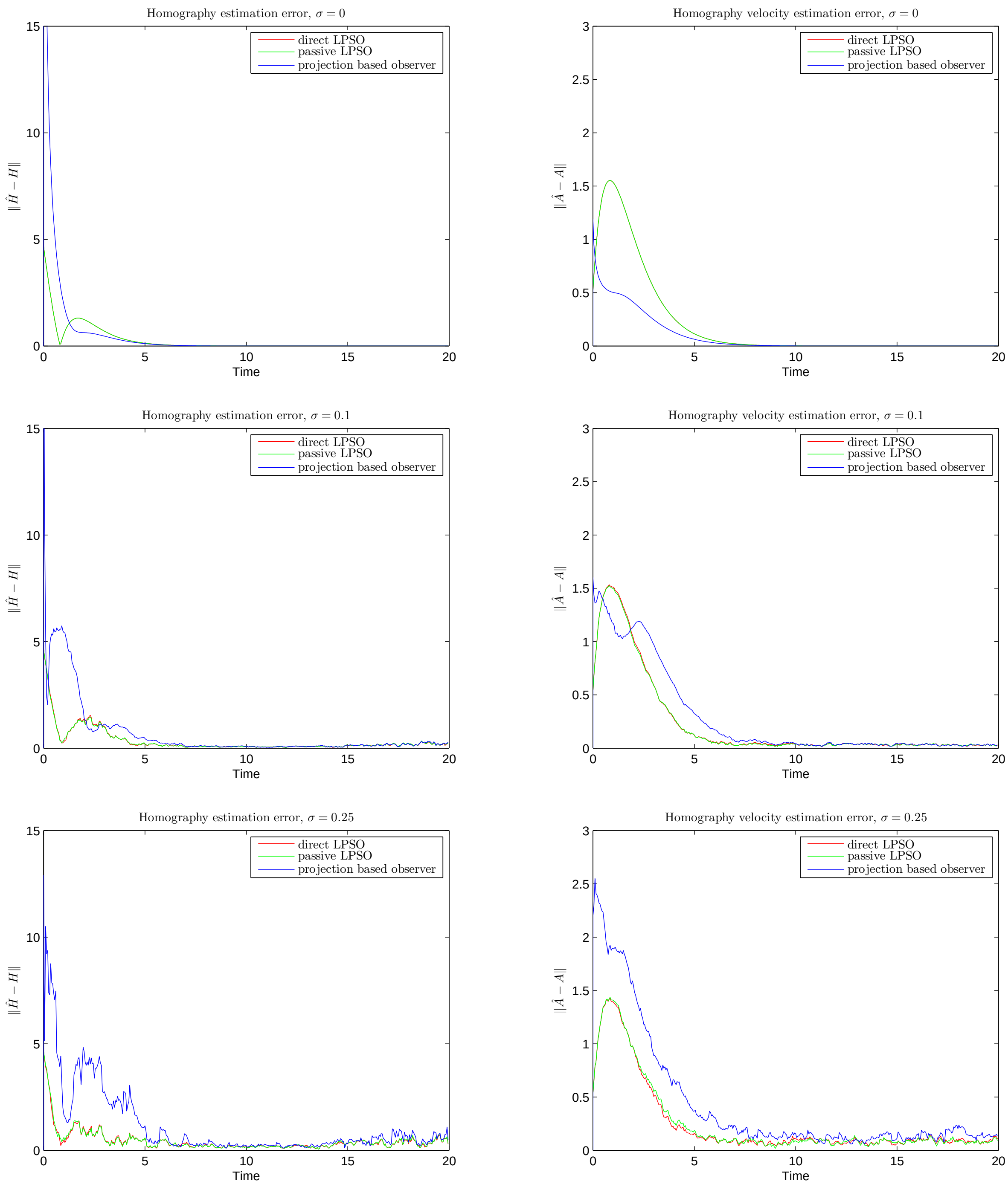

Figure 4.3: Direct and Passive LPSOs versus the Projection observer.

Chapter 5

Composite High-Gain Observers

This chapter is a departure from systems on Lie groups, in this chapter we consider systems whose state is a vector in $\mathbb{R}^n$. A new approach to observer design for nonlinear, input-affine, single-input, single-output systems, evolving on $\mathbb{R}^n$ is proposed. Our approach works by interconnecting two or more approximate differentiators (high-gain observers) to build a composite high-gain observer. We show that this approach to observer design is effective through simulation and do not prove the converge of the estimation error.

Designing the composite high-gain observer involves adding one, or more, redudant state variables to the system's differential equations in such a way that the augmented system takes on a special form. In this special form, the augmented system may be viewed as an interconnection of two, or more, sub-systems. The interconnection is such that it is possible to build two, or more, interconnected high-gain observers, which together behave as a single high-gain observer for the augmented system.

5.1 Introduction

In Section 2.3, we introduced the high-gain observer and in Section 2.4, we mentioned that high-gain observers act as approximate differentiators of a smooth signal, $y(t)$. If we connect two exact single differentiators (transfer function $H(s) = s$) in series, the resulting system will be an exact double differentiator (transfer function $H(s) = s^2$). Thus, we expect that if we connect two approximate single differentiators in series, the resulting system will be an approximate double differentiator.

To illustrate the above discussion, let us take two high-gain observers of type (2.7) with $n = 2$ and $\phi_0(\hat{x}, u) = 0$ and connect them in series. We get the following composite high-gain observer

$$
\begin{aligned}
\dot{\hat{x}}_1 &= \hat{x}_2 + \frac{1}{\epsilon} a_1 (y - \hat{x}_1) \\
\dot{\hat{x}}_2 &= \frac{1}{\epsilon^2} a_0 (y - \hat{x}_1) \\
\dot{\hat{x}}_3 &= \hat{x}_4 + \frac{1}{\epsilon} b_1 (\hat{x}_2 - \hat{x}_3) \\
\dot{\hat{x}}_4 &= \frac{1}{\epsilon^2} b_0 (\hat{x}_2 - \hat{x}_3),
\end{aligned}
\tag{5.1}
$$

where $\epsilon > 0$ is a small design parameter, and the constants a_0, a_1 and b_0, b_1 are chosen such that the polynomials $p_1(s) = s^2 + a_1 s + a_0$ and $p_2(s) = s^2 + b_1 s + b_0$ are Hurwitz. Notice that the $(\hat{x}_3, \hat{x}_4)$-subsystem of (5.1) does not use a direct measurement of y. Instead, the role of y is played by the state estimate $\hat{x}_2$. Hence we can view the $(\hat{x}_1, \hat{x}_2)$-subsystem as providing a "virtual measured output" to the $(\hat{x}_3, \hat{x}_4)$-subsystem.

The system (5.1) is linear, so we can compute the transfer function from y to $\hat{x}_4$. To this end, we treat the signal y as the input and rewrite the observer (5.1) as follows

$$\dot{\hat{x}} = A\hat{x} + By$$

where

$$A = \begin{pmatrix} -\frac{a_1}{\epsilon} & 1 & 0 & 0 \\ -\frac{a_0}{\epsilon^2} & 0 & 0 & 0 \\ 0 & \frac{b_1}{\epsilon} & -\frac{b_1}{\epsilon} & 1 \\ 0 & \frac{b_0}{\epsilon^2} & -\frac{b_0}{\epsilon} & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} \frac{a_1}{\epsilon} \\ \frac{a_0}{\epsilon^2} \\ 0 \\ 0 \end{pmatrix}.$$

Computing the transfer function from y to $\hat{x}_4$, we get

$$\frac{\hat{X}_4(s)}{Y(s)} = \frac{s^2 a_0 b_0}{(s^2\epsilon^2 + s\epsilon a_1 + a_0)(s^2\epsilon^2 + s\epsilon b_1 + b_0)}.$$

In the limit, as $\epsilon \to 0$, we get $\frac{\hat{X}_4(s)}{Y(s)} \to s^2$. Thus, we see that connecting two approximate single differentiators in series, yields an approximate double differentiator. In fact, the observer (5.1) is one simple example of a composite high-gain observer. The observer (5.1) happens to be linear, but in general our composite high-gain observers will not be linear.

The CHGO (5.1) can be viewed as a "chain" of two sub-HGOs. The first sub-HGO is

$$\begin{aligned}\dot{\hat{x}}_1 &= \hat{x}_2 + \frac{1}{\epsilon}a_1(y - \hat{x}_1)\\ \dot{\hat{x}}_2 &= \frac{1}{\epsilon^2}a_0(y - \hat{x}_1)\end{aligned}$$

and the second sub-HGO is

$$\begin{aligned}\dot{\hat{x}}_3 &= \hat{x}_4 + \frac{1}{\epsilon}b_1(\hat{x}_2 - \hat{x}_3)\\ \dot{\hat{x}}_4 &= \frac{1}{\epsilon^2}b_0(\hat{x}_2 - \hat{x}_3).\end{aligned}$$

The first sub-HGO differentiates the signal y, while the second sub-HGO differentiates the signal $\hat{x}_2$, which comes from the first sub-HGO. In the general case, the second sub-HGO can differentiate not just $\hat{x}_2$, but any function of the states of the first sub-HGO. We can also have more than two sub-HGO connected into a chain, so that each sub-HGO in the chain differentiates some function of the states of the sub-HGOs which are up the chain from itself and the first sub-HGO in the chain differentiates the signal y.

5.2 Composition of High-Gain Observers

Recall from Section 2.3, that to observe the system (2.6) we could use the HGO (2.7). We now consider building observers for the following, slightly more general system

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= x_3\\ &\vdots\\ \dot{x}_r &= \phi_r(x_1, \dots x_r, x_{r+1}, u)\\ \dot{x}_{r+1} &= x_{r+2}\\ &\vdots\\ \dot{x}_n &= \phi_n(x, u)\\ y &= h(x) = x_1,\end{aligned} \tag{5.2}$$

where ϕ_r and ϕ_n are globally Lipschitz and bounded. We write the above system as $\dot{x} = f(x, u)$. The following result tells us that ϕ_r is non-singular in its x_{r+1} argument if and only if the system (5.2) is observable.

Proposition 5.2.1. *The observability rank condition for the system (5.2) holds at $x = 0$ and for $u = u_0$ if and only if*

$$\left[\frac{\partial \phi_r}{\partial x_{r+1}}\right]\Big|_{x=0,u=u_0} \neq 0.$$

Proof. Assume that the system (5.2) is observable, at $x = 0$ and for $u = u_0$, i.e., assume that:

$$\text{rank}\{dh, dL_f h(x), dL_f^2 h(x), \dots, dL_f^{n-1} h(x)\} = n$$

at $x = 0$ and for $u = u_0$. It is clear that

$$dh = dx_1, \quad dL_f h(x) = dx_2, \quad \dots, \quad dL_f^{r-1} h(x) = dx_r, \quad dL_f^r h(x) = d\phi_r \tag{5.3}$$

and since $\text{rank}\{dh(x), \dots, dL_f^r h(x)\} = r + 1$, the differential $d\phi_r$ must be linearly independent from the set of differentials $\{dx_1, dx_2, \dots, dx_r\}$. Since ϕ_r is only a function of the states $x_1, \dots, x_{r+1}$, the row vector $d\phi_r$ can be written as

$$d\phi_r = \frac{\partial \phi_r}{\partial x_1} dx_1 + \dots + \frac{\partial \phi_r}{\partial x_r} dx_r + \frac{\partial \phi_r}{\partial x_{r+1}} dx_{r+1}.$$

If $\frac{\partial \phi_r}{\partial x_{r+1}}$ was equal to 0, then from the above equation, we would have that $d\phi_r$ is linearly dependent with the row vectors $dx_1, \dots dx_r$, which is a contradiction of what we have said above. Therefore we must have that $\left[\frac{\partial \phi_r}{\partial x_{r+1}}\right]\Big|_{x=0,u=u_0} \neq 0$.

For the converse, assume that $\frac{\partial \phi_r}{\partial x_{r+1}} = p \neq 0$. The observability matrix of the system (5.2) is given by

$$\begin{bmatrix} dh(x) \\ dL_f h(x) \\ \vdots \\ dL_f^{n-1} h(x) \end{bmatrix} = \left[\begin{array}{c|cccc} I_r & & 0_{r \times n-r} & & \\ \hline * & p & 0 & \dots & 0 \\ * & * & p & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \dots & p \end{array}\right],$$

which is full-rank, since $p \neq 0$. $\square$

From (5.3) we see that $x_1, \dots, x_r, \phi_r$ are just derivatives of the output up to order r. The quantity $\phi_r(x, u)$ is not a state but it can be estimated by means of an observer. Let us begin by building an approximate differentiator to estimate $x_1, \dots, x_r$ and $\phi_r(x,u)$:

$$\begin{aligned}
\dot{\hat{x}}_1 &= \hat{x}_2 + \frac{a_r}{\epsilon}(y - \hat{x}_1) \\
\dot{\hat{x}}_2 &= \hat{x}_3 + \frac{a_{r-1}}{\epsilon^2}(y - \hat{x}_1) \\
&\vdots \\
\dot{\hat{x}}_r &= \hat{\gamma} + \frac{a_1}{\epsilon^r}(y - \hat{x}_1) \\
\dot{\hat{\gamma}} &= \frac{a_0}{\epsilon^{r+1}}(y - \hat{x}_1),
\end{aligned} \tag{5.4}$$

where $\hat{x}_i$ approximates the state x_i, and $\hat{\gamma}$ approximates the quantity $\phi_r(x,u)$. If the constants $a_0, \dots, a_r$ are chosen such that the polynomial $p(s) = s^{r+1} + a_r s^r + \cdots + a_1 s + a_0$ is Hurwitz, then the error between x and $\hat{x}$ becomes of order $O(\epsilon)$ after a short transient peaking period, whose duration tends to zero, as ϵ tends to zero.

At this point, in light of Proposition 5.2.1, we have enough information to obtain an estimate of x_{r+1}, based on the states of the sub-HGO (5.4). To do this, note that the output of the system (5.2) satisfies

$$y^{(r)} = \phi_r(y, \dots, y^{(r-1)}, x_{r+1}, u).$$

Since ϕ_r is non-singular in x_{r+1} and is C^∞ in all of its arguments, by the inverse function theorem there exists a C^∞ function λ, such that in an open neighbourhood of $\mathbb{R}^{r+2}$ containing $(y, \dots, y^{(r-1)}, y^{(r)}, u) = (0, \dots, 0, \phi_r(0, \dots, 0, u_0), u_0)$,

$$\lambda(y, \dots, y^{(r-1)}, y^{(r)}, u) = x_{r+1}. \tag{5.5}$$

Notice that all arguments to λ are either known (i.e., input u), or are estimated by sub-HGO (5.4). In other words, we can estimate the quantities $y, \dots, y^{(r-1)}, y^{(r)}$, by states of the sub-HGO (5.4) $\hat{x}_1, \dots, \hat{x}_r, \hat{\gamma}$, to any desired accuracy. Therefore we evaluate λ with the states of (5.4)

$$\lambda(\hat{x}_1(t), \dots, \hat{x}_r(t), \hat{\gamma}(t), u(t)) \xrightarrow{\epsilon \to 0} \lambda(x_1(t), \dots, x_r(t), y^{(r)}(t), u(t)) = x_{r+1}(t).$$

This convergence is pointwise in time, i.e., it holds for any particular fixed time $t > 0$, but may fail to hold uniformly for all $t > 0$. This means that for any fixed time $t > 0$, we can make the estimates arbitrarily accurate at this t by decreasing ϵ towards 0 [25].

The next step is to treat the function $\lambda(\hat{x}_1(t), \dots, \hat{x}_r(t), \hat{\gamma}(t), u(t))$ as a "virtual output", or "virtual measurement" of x_{r+1}, and to build a second sub-HGO that approximately differentiates the "virtual output"

$$
\begin{aligned}
\dot{\hat{x}}_{r+1} &= \hat{x}_{r+2} + \frac{b_{n-r-1}}{\epsilon}(\lambda - \hat{x}_{r+1}) \\
\dot{\hat{x}}_{r+2} &= \hat{x}_{r+3} + \frac{b_{n-r-2}}{\epsilon^2}(\lambda - \hat{x}_{r+1}) \\
&\vdots \\
\dot{\hat{x}}_{n-1} &= \hat{x}_{n} + \frac{b_{1}}{\epsilon^{n-r-1}}(\lambda - \hat{x}_{r+1}) \\
\dot{\hat{x}}_{n} &= \frac{b_{0}}{\epsilon^{n-r}}(\lambda - \hat{x}_{r+1}),
\end{aligned}
\tag{5.6}
$$

where the constants $b_0, \ldots, b_{n-r-1}$ are chosen such that the polynomial $p(s) = s^{n-r} + b_{n-r-1}s^{n-r-1} + \cdots + b_1 s + b_0$ is Hurwitz. The function λ in the above HGO is understood to be evaluated as $\lambda = \lambda(\hat{x}_1, \ldots, \hat{x}_r, \hat{\gamma}, u)$. The two sub-HGOs (5.4) and (5.6) are connected by means of the function λ, present in the second sub-HGO. These two sub-HGOs, taken together, form a Composite High-Gain Observer (CHGO) for the system (5.2).

During the peaking period, the function $\lambda(\hat{x}_1, \ldots, \hat{x}_r, \hat{\gamma}, u)$ may fail to be defined. However, if λ is globally defined, then we conjecture that, as $\epsilon \to 0$, the composite system (5.4), (5.6) will behave more and more like the composition of ideal differentiators, shown in Figure 5.1.

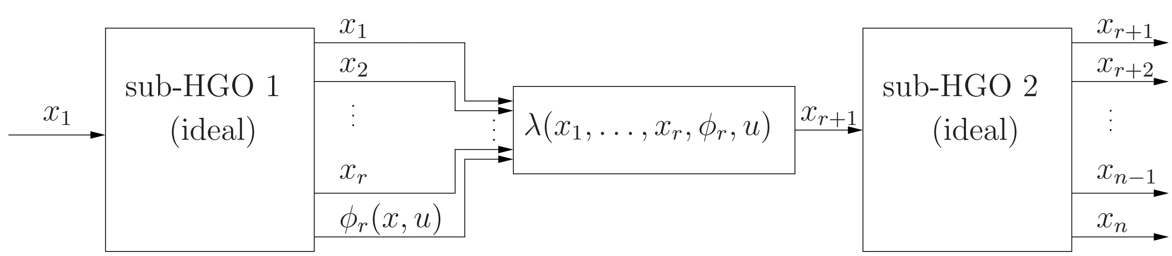

Figure 5.1: Composition of ideal differentiators, a non-causal system.

Notice that both of our sub-HGOs (5.4) and (5.6) are linear systems and do not include the full model of the system. In particular, the sub-HGO (5.4) does not include a model for how $\hat{\gamma}$ changes with time and the sub-HGO (5.6) does not include a model for how $\hat{x}_n$ changes with time. Therefore, when we compose the two sub-HGOs, we do not expect the estimation error dynamics of the resulting observer to have an equilibrium point at 0. We do not include the full model in the sub-HGO (5.4), because doing so would mean that we need to know the derivative of u, because the quantity $\phi_r(x, u)$ depends on u. We can still include the full model in the sub-HGO (5.6), but we choose not to do this right now, because we simply want to illustrate interconnection of linear high-gain observers. In later sections of this chapter, we will show how to make the estimation error have an equilibrium at 0 without knowing the derivative of the input, we can do this by using a more powerful high-gain observer, which will be introduced in the next section.

Let us summarize the process that we used to design the CHGO:

1. Build the first sub-HGO to estimate derivatives of y up to order r.

2. Construct the C^∞ function λ, which satisfies (5.5). The function λ is meant to be evaluated at the states of the first sub-HGO, thus providing a "virtual measurement" of x_{r+1}. Ideally, the function λ will be globally, or at least semi-globally defined.

3. Build the second sub-HGO to estimate derivatives of $\lambda(\hat{x}_1, \ldots, \hat{x}_r, \hat{\gamma}, u)$ up to order $n - r$.

The reader might be wondering why we have chosen to compose specifically high-gain observers and not some other type of observers, such as sliding mode observers, or extended Kalman filters. The answer lies in the fact that high-gain observers are approximate differentiators, as discussed in Section 2.4. Approximate differentiators are robust to modeling uncertainties, because it is possible to approximately differentiate a signal without knowing the full model of the system from which this signal originated. The robustness to modeling uncertainties is crucial for CHGO to work, because the observation error of the first sub-HGO must decay to order $O(\epsilon)$ values, while the estimation error of the second sub-HGO is still large. This is where the robustness of the high-gain observer is useful - the HGO can approximately differentiate the signal $y(t)$ without knowing the full model of the system from which $y(t)$ originated. This allows the first sub-HGO to get order $O(\epsilon)$ close to $\{x_1, \ldots, x_r, y^{(r)}\}$, without using the full model of the system.

Once the first sub-HGO's observation error has decayed to order $O(\epsilon)$ values, the function $\lambda(\hat{x}_1, \ldots, \hat{x}_r, \hat{\gamma}, u)$ becomes close to the value of x_{r+1} and the second sub-HGO will begin receiving correct "virtual measurement" of x_{r+1}, which it approximately differentiates.

5.3 Generalized High-Gain Observer

In this section, we briefly present a more powerful type of HGO, which works for a much larger class of systems than (2.6). This type of high-gain observer is analyzed in [14], [15], [16] and other related papers.

Consider the problem of partial-state observer design for the following nonlinear, input affine, single-input, single-output system defined on $\mathbb{R}^n$

$$\begin{aligned} \dot{x} &= f(x) + g(x)u \\ y &= h(x) = x_1, \end{aligned} \tag{5.7}$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $f, g : \mathbb{R}^n \to \mathbb{R}^n$ are assumed to be C^∞ maps. The variable y is the measured output of the system and the variable u is the input applied to the system, also measured.

Definition 5.3.1. *For the system (5.7), define the region of observability for any input, Ω_x, to be the subset of the state space of (5.7), on which the observability rank condition holds for any input, i.e.,*

$$\Omega_x := \left\{ x \in \mathbb{R}^n : \operatorname{rank}\{dh(x), dL_{f+gu}h(x), \ldots, dL_{f+gu}^{n-1}h(x)\} = n,\ \forall u \in \mathbb{R} \right\}.$$

The authors of [14] suggest a local change of coordinates, $z = \Psi(x)$, which is defined for $x \in \Omega_x$, and is computed as follows

$$\begin{aligned} z_1 &:= h(x) = x_1 \\ z_2 &:= L_f h(x) \\ &\vdots \\ z_n &:= L_f^{n-1} h(x). \end{aligned} \tag{5.8}$$

The result [14, Theorem 2], implies that $\Psi : \Omega_x \to \mathbb{R}^n$ is a local diffeomorphism and the change of coordinates $z = \Psi(x)$ transforms the system (5.7) into the following triangular normal form

$$\begin{aligned} \dot{z}_1 &= z_2 + g_1(z_1)u \\ \dot{z}_2 &= z_3 + g_2(z_1, z_2)u \\ &\vdots \\ \dot{z}_{n-1} &= z_n + g_{n-1}(z_1, \ldots, z_{n-1})u \\ \dot{z}_n &= \phi(z) + g_n(z)u \\ y &= z_1. \end{aligned} \tag{5.9}$$

It is shown in [15] and [16] that, if the functions ϕ, g_i, $i = 1, \ldots, n$ are globally Lipschitz and bounded then, for $\epsilon > 0$ sufficiently small, the following is an exponential observer for (5.9),

$$\begin{aligned}
\dot{\hat{z}}_1 &= \hat{z}_2 + g_1(\hat{z}_1)u + \frac{1}{\epsilon}a_{n-1}(y - \hat{z}_1) \\
\dot{\hat{z}}_2 &= \hat{z}_3 + g_2(\hat{z}_1, z_2)u + \frac{1}{\epsilon^2}a_{n-2}(y - \hat{z}_1) \\
&\vdots \\
\dot{\hat{z}}_{n-1} &= \hat{z}_n + g_{n-1}(\hat{z}_1, \ldots, \hat{z}_{n-1})u + \frac{1}{\epsilon^{n-1}}a_1(y - \hat{z}_1) \\
\dot{\hat{z}}_n &= \phi(\hat{z}) + g_n(\hat{z})u + \frac{1}{\epsilon^n}a_0(y - \hat{z}_1),
\end{aligned} \tag{5.10}$$

where the constants $a_0, \ldots, a_{n-1}$ are design parameters, chosen such that the polynomial $p(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$ is Hurwitz. These results of [14], [15] and [16] are important because they allow high-gain observer design for a much broader class of systems than the "standard" HGO (2.7), which we introduced in Section 2.3. The HGO (5.10) is still very similar in structure to the HGO (2.7), because it is just a copy of the vector field of the system (5.9), plus the same corrective high-gain term used in the HGO (2.7).

To construct an HGO for the nonlinear system (5.7) using the normal form (5.9), one normally follows the following steps:

1. Use (5.8) to compute the change of coordinates, $z = \Psi(x)$, which transforms (5.7) into the normal form (5.9).

2. The mapping $z = \Psi(x)$ has a full-rank Jacobian matrix for all $x \in \Omega_x$, so $\Psi : \Omega_x \to \mathbb{R}^n$ is a local diffeomorphism around each point $x \in \Omega_x$. However Ψ may fail to be a global diffeomorphism onto its image, because it may fail to be injective. This means that some global information may be lost when we change the coordinates to $z = \Psi(x)$.

3. Build the HGO (5.10) in z-coordinates.

4. Transform the HGO (5.10) back into x-coordinates using the inverse coordinate transformation $x = \Psi^{-1}(z)$. After doing this, one has obtained the HGO equations in x-coordinates. The observer equations obtained using the procedure above will be undefined at $x \notin \Omega_x$.

Example 1. Consider the following system, given in its original coordinates,

$$\begin{aligned}
\dot{x}_1 &= (x_1 - 1)x_2 + u \\
\dot{x}_2 &= u \\
y &= h(x) = x_1.
\end{aligned} \tag{5.11}$$

Write the above system as $\dot{x} = f(x, u)$, then the observability matrix for this system is

$$\begin{bmatrix} dh(x) \\ dL_f h(x) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ x_2 & (x_1 - 1) \end{bmatrix}. \tag{5.12}$$

Therefore, the region of observability for any input of the above system is

$$\Omega_x = \{(x_1, x_2) \in \mathbb{R}^2 : \quad x_1 \neq 1\}.$$

To convert the above system into the normal form (5.9), define a coordinate transformation $z = \Psi(x)$, according to (5.8), to be $z_1 := x_1$ and $z_2 := (x_1 - 1)x_2$. In the z-coordinates, the system (5.11) takes on the normal form (5.9)

$$\begin{aligned} \dot{z}_1 &= z_2 + u \\ \dot{z}_2 &= \frac{z_2(z_2 + u)}{z_1 - 1} + (z_1 - 1)u \\ y &= z_1. \end{aligned} \tag{5.13}$$

Next, we build the HGO in the z-coordinates:

$$\begin{aligned} \dot{\hat{z}}_1 &= \hat{z}_2 + u + \frac{a_1}{\epsilon}(y - \hat{z}_1) \\ \dot{\hat{z}}_2 &= \frac{\hat{z}_2(\hat{z}_2 + u)}{\hat{z}_1 - 1} + (\hat{z}_1 - 1)u + \frac{a_0}{\epsilon^2}(y - \hat{z}_1). \end{aligned} \tag{5.14}$$

Finally, we convert the HGO into the x-coordinates:

$$\begin{aligned} \dot{\hat{x}}_1 &= (\hat{x}_1 - 1)\hat{x}_2 + u + a_1 \frac{1}{\epsilon}(y - \hat{x}_1) \\ \dot{\hat{x}}_2 &= u + \left(\frac{a_0}{\epsilon^2} - \frac{a_1}{\epsilon}\hat{x}_2 \right) \frac{y - \hat{x}_1}{\hat{x}_1 - 1}. \end{aligned} \tag{5.15}$$

Notice that as $\hat{x}_1 \to 1$, we have $\dot{\hat{x}}_2 \to \infty$. This is because the state x_2 becomes progressively "less observable" as x_1 approaches 1 and x_2 becomes unobservable when $x_1 = 1$.

5.4 Simple Design Example

We will design a CHGO, based on the observers from Section 5.3, for the system from Example 1. By composing HGOs of type (5.10) we will be able to obtain estimation error dynamics that have an equilibrium point at the zero error $\tilde{x} = 0$, unlike the case when observers of Section 5.2 are used.

5.4.1 Observer Design

Step 1: Augment the states of the system (5.11) with one additional redundant state variable $\gamma \in \mathbb{R}$, which is defined to be drift of x_1 under zero input

$$\gamma := \frac{d}{dt}x_1\Big|_{u=0} = (x_1 - 1)x_2.$$

We defined γ to be the drift of x_1 under zero input, rather than the full derivative of x_1 as in Section 5.2, because this way γ is not dependent on u and therefore $\dot{u}$ does not appear in the equation of $\dot{\gamma}$. Since $\dot{u}$ does not appear in $\dot{\gamma}$, we will later be able to include a full model of γ in our observer, without knowing $\dot{u}$.

If we know x_1 and γ, we can uniquely solve for x_2 if and only if $x_1 \neq 1$. For future use, define the function that solves for x_2, given x_1 and γ

$$\lambda(x_1, \gamma) := \frac{\gamma}{x_1 - 1}.$$

Notice that this function $\lambda(x_1, \gamma)$ is singular when $x_1 = 1$, i.e., when $x \notin \Omega_x$. This is not surprising because when $x_1 = 1$, the system loses observability and the state x_2 can not be obtained from the knowledge of the output and its derivatives.

Step 2: Compute the expression for the derivative of γ with respect to time

$$\begin{aligned}\dot{\gamma} &= \frac{d}{dt}\left[(x_1 - 1)x_2\right] \\ &= (x_1 - 1)x_2^2 + (x_1 + x_2 - 1)u \\ &=: M(x_1, x_2, u).\end{aligned}$$

Now we can rewrite the system (5.11), with the redundant variable γ included, as

$$\begin{aligned}\dot{x}_1 &= \gamma + u \\ \dot{\gamma} &= (x_1 - 1)x_2^2 + (x_1 + x_2 - 1)u \\ \dot{x}_2 &= u \\ y &= x_1.\end{aligned} \tag{5.16}$$

The idea is to treat the augmented system (5.16) as an interconnection of two sub-systems. The first sub-system has states x_1 and γ, while the second sub-system has state x_2. The first sub-system is

$$\begin{aligned}\dot{x}_1 &= \gamma + u \\ \dot{\gamma} &= (x_1 - 1)x_2^2 + (x_1 + x_2 - 1)u \\ y &= h_1(x) = x_1,\end{aligned} \tag{5.17}$$

where x_1, γ are states and x_2 is treated as an unknown function of time, i.e., a modeling uncertainty. The second sub-system is

$$\begin{aligned} \dot{x}_2 &= u \\ y &= h_2(x) = x_2. \end{aligned} \tag{5.18}$$

The output of the second sub-system (5.18) is $h_2(x) = x_2$, but this output is also equal to $\lambda(x_1, \gamma) = \frac{\gamma}{x_1 - 1}$, which is a function entirely of the states of the first sub-system (5.17). This happens because $\frac{\gamma}{x_1-1} = x_2$, as long as $x \in \Omega_x$.

Step 3: Both the sub-systems (5.17) and (5.18) are in the triangular normal form (5.9). So we design two sub-HGOs of type (5.10), one for each sub-system. The first sub-HGO uses the measured output y and the input u to estimate the states x_1 and γ

$$\begin{aligned} \dot{\hat{x}}_1 &= \hat{\gamma} + u + \frac{a_1}{\epsilon_1}(y - \hat{x}_1) \\ \dot{\hat{\gamma}} &= (\hat{x}_1 - 1)\hat{x}_2^2 + (\hat{x}_1 + \hat{x}_2 - 1)u + \frac{a_0}{\epsilon_1^2}(y - \hat{x}_1), \end{aligned} \tag{5.19}$$

where $\epsilon_1 > 0$ is a small design parameter and $a_0, a_1 \in \mathbb{R}$ are design parameters, chosen such that the polynomial $p_1(s) = s^2 + a_1 s + a_0$ is Hurwitz.

The first sub-HGO (5.19) can provide a "virtual measurement" of x_2, via the function $\lambda(\hat{x}_1, \hat{\gamma})$, once the estimation error between $(\hat{x}_1, \hat{\gamma})$ and (x_1, γ) has decayed to order $O(\epsilon_1)$ values. The second sub-HGO uses this "virtual measurement", $\lambda(\hat{x}_1, \hat{\gamma})$, which is a function entirely of the states of the first sub-HGO, as if it was the signal $h_2(x) = x_2$. Thus, the second sub-HGO uses $\lambda(\hat{x}_1, \hat{\gamma})$ and the input u to estimate the state x_2:

$$\dot{\hat{x}}_2 = u + \frac{b_0}{\epsilon_2}\left(\lambda(\hat{x}_1, \hat{\gamma}) - \hat{x}_2\right), \tag{5.20}$$

where $\epsilon_2 > 0$ is a small design parameter and $b_0 > 0$ is a design parameter, chosen such that the polynomial $p_2(s) = s + b_0$ is Hurwitz.

Combining the two sub-HGOs (5.19) and (5.20) into a single CHGO, we get

$$\begin{aligned} \dot{\hat{x}}_1 &= \hat{\gamma} + u + \frac{a_1}{\epsilon_1}(y - \hat{x}_1) \\ \dot{\hat{\gamma}} &= (\hat{x}_1 - 1)\hat{x}_2^2 + (\hat{x}_1 + \hat{x}_2 - 1)u + \frac{a_0}{\epsilon_1^2}(y - \hat{x}_1) \\ \dot{\hat{x}}_2 &= u + \frac{b_0}{\epsilon_2}(\lambda - \hat{x}_2), \end{aligned} \tag{5.21}$$

where the function λ is understood to be evaluated as $\lambda(\hat{x}_1, \hat{\gamma})$. The two sub-HGOs (5.19) and (5.20), which make up the CHGO (5.21), are interconnected in a rather complicated way, because $\hat{x}_2$ appears in (5.19). However, as $\epsilon_1, \epsilon_2 \to 0$, the dominant interconnection is through the λ function from the first sub-HGO to the second sub-HGO.

We have not yet explained how to choose the design parameters ϵ_1 and ϵ_2, other than saying that they should be made small. We will later argue that these parameters should be chosen such that the first sub-HGO has higher gain than the second sub-HGO, i.e., $\epsilon_1 << \epsilon_2$. For example, we can take $\epsilon_1 = \epsilon^2$ and $\epsilon_2 = \epsilon$, where $\epsilon > 0$ is a single design parameter that replaces ϵ_1 and ϵ_2. The intuitive explanation for taking the gain of the first sub-HGO to be higher than the gain of the second sub-HGO is that the "virtual output" $\lambda(\hat{x}_1, \hat{\gamma})$ that goes into the second sub-HGO must become approximately accurate, i.e., sufficiently close to x_2, before it can becomes useful to the second sub-HGO. In other words, the first sub-HGO's estimation error must decay to small values before the second sub-HGO can even begin to do anything useful.

5.4.2 Estimation Error Dynamics

We define the estimation errors to be $\tilde{x}_i := x_i - \hat{x}_i$ for $i = 1, 2$ and $\tilde{\gamma} := \gamma - \hat{\gamma}$. Differentiating these variables, we obtain the following estimation error dynamics

$$
\begin{aligned}
\dot{\tilde{x}}_1 =& \tilde{\gamma} - a_1 \frac{1}{\epsilon_1} \tilde{x}_1 \\
\dot{\tilde{\gamma}} =& (x_1 - 1)x_2^2 - (\hat{x}_1 - 1)\hat{x}_2^2 + (\tilde{x}_1 + \tilde{x}_2)u - \frac{a_0}{\epsilon_1^2}(\tilde{x}_1) \\
\dot{\tilde{x}}_2 =& -\frac{b_0}{\epsilon_2}(\lambda - \hat{x}_2),
\end{aligned}
\tag{5.22}
$$

where λ is understood to be evaluated as $\lambda = \lambda(\hat{x}_1, \hat{\gamma})$.

We introduce the virtual measurement error, $\tilde{\lambda}$, which is defined to be the difference between the true state, x_2, and its virtual measurement, $\lambda(\hat{x}_1, \hat{\gamma})$

$$
\begin{aligned}
\tilde{\lambda}(\tilde{x}_1, \tilde{\gamma}, x_1, \gamma) :=& \lambda(x_1, \gamma) - \lambda(x_1 - \tilde{x}_1, \gamma - \tilde{\gamma}) \\
=& \lambda(x_1, \gamma) - \lambda(\hat{x}_1, \hat{\gamma}) \\
=& x_2 - \lambda(\hat{x}_1, \hat{\gamma}).
\end{aligned}
$$

We will suppress the arguments to $\tilde{\lambda}$ so that when we write $\tilde{\lambda}$, it is understood that we mean $\tilde{\lambda}(\tilde{x}_1, \tilde{\gamma}, x_1, \gamma)$. Using this definition, the term $\lambda - \hat{x}_2$ from the error dynamics (5.22) is written as

$$\begin{aligned}\lambda(\hat{x}_1,\hat{\gamma})-\hat{x}_2 &= \lambda(\hat{x}_1,\hat{\gamma})-\lambda(x_1,\gamma)+\lambda(x_1,\gamma)-\hat{x}_2\\ &= \lambda(\hat{x}_1,\hat{\gamma})-\lambda(x_1,\gamma)+x_2-\hat{x}_2\\ &= \lambda(\hat{x}_1,\hat{\gamma})-\lambda(x_1,\gamma)+\tilde{x}_2\\ &= \tilde{x}_2-\tilde{\lambda}.\end{aligned}$$

Using this, we rewrite the error dynamics (5.22) to make it look like a linear, stable system, which is perturbed by a nonlinear perturbation,

$$\frac{d}{dt}\begin{pmatrix}\tilde{x}_1\\ \tilde{\gamma}\\ \tilde{x}_2\end{pmatrix}=\begin{pmatrix}-a_1\frac{1}{\epsilon_1} & 1 & 0\\ -a_0\frac{1}{\epsilon_1^2} & 0 & 0\\ 0 & 0 & -b_0\frac{1}{\epsilon_2}\end{pmatrix}\begin{pmatrix}\tilde{x}_1\\ \tilde{\gamma}\\ \tilde{x}_2\end{pmatrix}+\delta+\Delta, \tag{5.23}$$

where δ and Δ are perturbation terms, given by

$$\delta=\begin{pmatrix}0\\ (x_1-1)x_2^2-(\hat{x}_1-1)\hat{x}_2^2+(\tilde{x}_1+\tilde{x}_2)u\\ 0\end{pmatrix},$$

$$\Delta=\begin{pmatrix}0\\ 0\\ b_0\frac{1}{\epsilon_2}\tilde{\lambda}\end{pmatrix}.$$

The good thing about the perturbation terms δ and Δ is that they are vanishing perturbations, meaning that when the estimation error is zero (i.e., when $\hat{x}=x$ and $\hat{\gamma}=\gamma$, or equivalently, when $\tilde{x}=0$ and $\tilde{\gamma}=0$), then the perturbation terms are equal to zero:

$$\left(\forall x\in\mathbb{R}^2\right)\ (\forall\gamma\in\mathbb{R})\ (\forall u\in\mathbb{R})\qquad \delta\big|_{\tilde{x}=0,\tilde{\gamma}=0}=0\ \text{ and }\ \Delta\big|_{\tilde{x}=0,\tilde{\gamma}=0}=0.$$

This means that the point $(\tilde{x},\tilde{\gamma})=0$ is an equilibrium point of the error dynamics.

The bad thing about the perturbation term Δ is that it contains a term multiplied by $1/\epsilon_2$. So Δ increases in magnitude as $\epsilon_2\to 0$. Fortunately, the high-gain perturbation only depends on the estimation error of the first sub-HGO, which decreases to order $O(\epsilon_1)$ values.

5.4.3 Simulation

We simulate the CHGO (5.21) and the HGO (5.15), both applied to the system (5.11) from Example 1. The control input is chosen to be a fixed function of time

$$u=-\sin(t/2).$$

The system and the observer are initialized at

$$x_1 = -2, \qquad x_2 = -2, \qquad \hat{x}_1 = 0, \qquad \hat{x}_2 = 0.$$

The results of the simulation are shown in Figure 5.2. From the simulation, we can see that the CHGO and the HGO have similar performance in the absence of measurement noise.

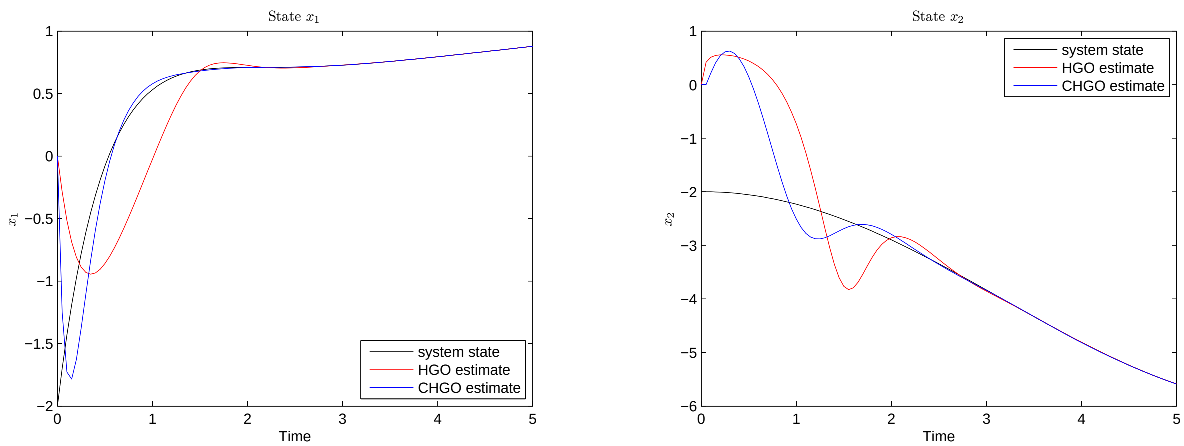

Figure 5.2: Simulation of the CHGO (5.21) and the HGO (5.15), both applied to the system (5.11), without measurement noise.

5.5 TORA System

In this section, we design a CHGO to estimate the states of a particular simple mechanical system, called the translational oscillator with a rotating actuator (TORA) system, shown in Figure 5.3. The TORA system consists of a rotating pendulum on a cart, the cart is free to move back and forth in one dimension and is attached to a fixed wall by a spring. To our knowledge, there has been no work on successfully applying high-gain observers to estimate the states of the TORA system. In [21], the authors stabilize the TORA system without constructing an explicit observer for all of its states. We are not going to attempt to stabilize the TORA system to any equilibrium point.

For simplicity, we assume there is no gravity and no friction in any of the components.

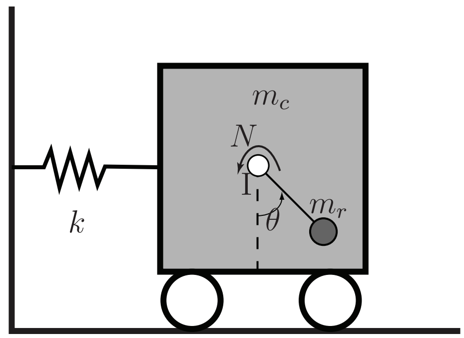

Figure 5.3: The translational oscillator with a rotating actuator (TORA) System

The equations of motion of the TORA system are given by [38]

$$(m_r + m_c)\ddot{d} + m_r l_r(\ddot{\theta}\cos(\theta) - \dot{\theta}^2 \sin(\theta)) + kd = 0,$$
$$(I + m_r l_r^2)\ddot{\theta} + m_r l_r \ddot{d}\cos(\theta) = N.$$

Table 5.1 gives a list of physical parameters of the TORA system.

Table 5.1: List of Parameters of the TORA System

m_c	mass of the cart
m_r	mass of the rotating actuator
l_r	length of the actuator
N	torque on the actuator
I	inertia of the disk
θ	angle of the actuator away from horizontal
d	displacement of the cart

Following the procedure of [38], we introduce the dimensionless variables

$$A := \frac{m_r l_r}{\sqrt{(I + m_r l_r^2)(m_r + m_c)}}, \qquad \tau := \sqrt{\frac{k}{m_r + m_c}}t,$$
$$x_3 := \sqrt{\frac{m_r + m_c}{I + m_r l_r^2}}d, \qquad u := \frac{m_r + m_c}{k(I + m_r l_r^2)}N.$$

We define the state variables as follows

$$x = (x_1, x_2, x_3, x_4) := (\theta, \dot{\theta}, x_3, \dot{x}_3),$$

where the dot and double dot now denote derivatives with respect to τ. The vector field of the TORA system in these coordinates, is given by

$$\begin{aligned}
\dot{x}_1 =& x_2\\
\dot{x}_2 =& \frac{A\cos(x_1)\left(x_3 - Ax_2^2\sin(x_1)\right)}{1 - A^2\cos^2(x_1)} + \frac{1}{1 - A^2\cos^2(x_1)}u\\
\dot{x}_3 =& x_4\\
\dot{x}_4 =& \frac{Ax_2^2\sin(x_1) - x_3}{1 - A^2\cos^2(x_1)} - \frac{A\cos(x_1)}{1 - A^2\cos^2(x_1)}u\\
y =& h(x) = x_1.
\end{aligned} \tag{5.24}$$

Note that the denominator $1 - A^2\cos^2(x_1)$ is never zero, because $A < 1$ by definition. Writing the system (5.24) as $\dot{x} = f(x, u)$, then the observability matrix is

$$\begin{bmatrix} dh(x) \\ dL_f h(x) \\ dL_f^2 h(x) \\ dL_f^3 h(x) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ * & * & \frac{A\cos(x_1)}{1-A^2\cos^2(x_1)} & 0 \\ * & * & * & \frac{A\cos(x_1)}{1-A^2\cos^2(x_1)} \end{bmatrix}. \tag{5.25}$$

Since $1 - A^2\cos^2(x_1)$ is never zero, the observability matrix of the TORA system is full rank, for any input and at points $x \in \mathbb{R}^4$, such that $\cos(x_1) \neq 0$. Therefore the region of observability for any input is

$$\Omega_x := \{x \in \mathbb{R}^4 : \cos(x_1) \neq 0\}.$$

We see that the TORA system loses observability when $\cos(x_1) = 0$. Physically, the configuration $\cos(x_1) = 0$ corresponds to the rotating actuator (pendulum) being in the horizontal position. In this configuration, any acceleration of the cart does not produce torque on the pendulum. On the other hand, when $\cos(x_1) \neq 0$, acceleration of the cart induces a torque on the pendulum. This torque appears in the second derivative of the output.

Theoretically, it is possible to convert the TORA system into the triangular normal form (5.9) and then apply the high-gain observer (5.10). Unfortunately, this change of coordinates is very cumbersome to work with. When we design a CHGO for the TORA system, we avoid doing this coordinate transformation and design the CHGO in the original coordinates of the TORA system (5.24), by introducing a redundant state variable.

5.5.1 Observer Design

We will design a CHGO for the TORA system by following the same steps we took in Section 5.4.

Step 1: Augment the states of the TORA system (5.24) with one additional redundant state variable $\gamma \in \mathbb{R}$, which is defined to be the drift of x_2 under zero input, the expression for which is easy to compute,

$$\gamma := \frac{d}{dt}\left[x_2\right]\Big|_{u=0} = \frac{A\cos(x_1)\left(x_3 - Ax_2^2\sin(x_1)\right)}{1 - A^2\cos^2(x_1)}. \tag{5.26}$$

If we know x_1, x_2 and γ, we can uniquely solve for x_3, by rearranging the equation (5.26), if and only if $\cos(x_1) \neq 0$. For future use, we define the "virtual output" function, that solves for x_3 given x_1, x_2 and γ

$$\lambda(x_1, x_2, \gamma) := \frac{\gamma(1 - A^2\cos^2(x_1))}{A\cos(x_1)} + Ax_2^2\sin(x_1) \tag{5.27}$$

As expected, the function $\lambda(x_1, x_2, \gamma)$ is undefined when $\cos(x_1) = 0$, i.e., when $x \notin \Omega_x$. When $\cos(x_1) = 0$, the TORA system loses observability and the states x_3 and x_4 can not be estimated from the output.

Step 2: Compute the expression for the derivative of γ with respect to time

$$\begin{aligned}
&\frac{d}{dt}\gamma = \frac{d}{dt}\left[\frac{A\cos(x_1)\left(x_3 - Ax_2^2\sin(x_1)\right)}{1 - A^2\cos^2(x_1)}\right] \\
&= \frac{-Ax_2x_3\sin(x_1) + A^2x_2^3\sin^2(x_1) - A^2x_2^3\cos^2(x_1) + A\cos(x_1)x_4}{1 - A^2\cos^2(x_1)} + \\
&- 4\,\frac{A^3x_2x_3\cos^2(x_1)\sin(x_1) - A^4x_2^3\sin^2(x_1)}{\left(1 - A^2\cos^2(x_1)\right)^2} - 2\,\frac{A^2\cos(x_1)\,x_2\,\sin(x_1)}{\left(1 - A^2\cos^2(x_1)\right)^2}u \\
&=: M(x_1, x_2, x_3, x_4, u).
\end{aligned} \tag{5.28}$$

Now we can rewrite the TORA system (5.24), with the redundant variable γ included, as

$$\begin{aligned}
\dot{x}_1 =& x_2 \\
\dot{x}_2 =& \gamma + \frac{1}{1 - A^2 \cos^2(x_1)} u \\
\dot{\gamma} =& M(x_1, x_2, x_3, x_4, u) \\
\dot{x}_3 =& x_4 \\
\dot{x}_4 =& \frac{A x_2^2 \sin(x_1) - x_3}{1 - A^2 \cos^2(x_1)} - \frac{A \cos(x_1)}{1 - A^2 \cos^2(x_1)} u \\
y =& h(x) = x_1.
\end{aligned} \tag{5.29}$$

The idea is to treat the augmented system (5.29) as an interconnection of two sub-systems. The first sub-system has states x_1, x_2 and γ, while the second sub-system has states x_3 and x_4. The first sub-system is

$$\begin{aligned}
\dot{x}_1 =& x_2 \\
\dot{x}_2 =& \gamma + \frac{1}{1 - A^2 \cos^2(x_1)} u \\
\dot{\gamma} =& M(x_1, x_2, x_3, x_4, u) \\
y =& h_1(x) = x_1,
\end{aligned} \tag{5.30}$$

where x_1, x_2, γ are states and x_3, x_4 are treated as unknown functions of time, i.e., modeling uncertainties. The second sub-system is

$$\begin{aligned}
\dot{x}_3 =& x_4 \\
\dot{x}_4 =& \frac{A x_2^2 \sin(x_1) - x_3}{1 - A^2 \cos^2(x_1)} - \frac{A \cos(x_1)}{1 - A^2 \cos^2(x_1)} u \\
y =& h_2(x) = x_3,
\end{aligned} \tag{5.31}$$

where x_3 and x_4 are states and x_1, x_2 are treated as unknown functions of time, i.e., modeling uncertainties. The output of the second sub-system is $h_2(x) = x_3$, but this output is also equal to $\lambda(x_1, x_2, \gamma) = x_3$, which is a function entirely of the states of the first sub-system (5.30).

Step 3: Both the sub-systems (5.30) and (5.31) are in the triangular normal form (5.9). So we design two sub-HGOs of type (5.10), one for each sub-system. The first sub-HGO uses the measured output y and input u to estimate the states x_1, x_2 and γ

$$\begin{aligned}\dot{\hat{x}}_1 =& \hat{x}_2 + a_2 \frac{1}{\epsilon_1}(y - \hat{x}_1) \\ \dot{\hat{x}}_2 =& \hat{\gamma} + \frac{1}{1 - A^2\cos^2(\hat{x}_1)}u + a_1\frac{1}{\epsilon_1^2}(y - \hat{x}_1) \\ \dot{\hat{\gamma}} =& M(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, u) + a_0\frac{1}{\epsilon_1^3}(y - \hat{x}_1),\end{aligned} \tag{5.32}$$

where $\epsilon_1 > 0$ is a small design parameter and $a_0, a_1, a_2 \in \mathbb{R}$ are design parameters, chosen such that the polynomial $p_1(s) = s^3 + a_2 s^2 + a_1 s + a_0$ is Hurwitz.

The first sub-HGO (5.32) provides a "virtual measurement" of x_3, via the function $\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma})$, after the estimation error between $(\hat{x}_1, \hat{x}_2, \hat{\gamma})$ and (x_1, x_2, γ) has decayed to order $O(\epsilon_1)$ values. The second sub-HGO uses the "virtual measurement", $\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) \approx x_3$, which is a function entirely of the states of the first sub-HGO, as if it was the signal $h_2(x) = x_3$. Thus, the second sub-HGO uses $\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma})$ and the input u to estimate the states x_3 and x_4

$$\begin{aligned}\dot{\hat{x}}_3 =& \hat{x}_4 + b_1\frac{1}{\epsilon_2}(\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) - \hat{x}_3) \\ \dot{\hat{x}}_4 =& \frac{A\hat{x}_2^2\sin(\hat{x}_1) - \hat{x}_3}{1 - A^2\cos^2(\hat{x}_1)} - \frac{A\cos(\hat{x}_1)}{1 - A^2\cos^2(\hat{x}_1)}u + b_0\frac{1}{\epsilon_2^2}(\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) - \hat{x}_3),\end{aligned} \tag{5.33}$$

where $\epsilon_2 > 0$ is a small design parameter and $b_0, b_1 \in \mathbb{R}$ are design parameters, chosen such that the polynomial $p_2(s) = s^2 + b_1 s + b_0$ is Hurwitz.

Combining the two sub-HGOs (5.32) and (5.33), we get

$$\begin{aligned}\dot{\hat{x}}_1 =& \hat{x}_2 + a_2 \frac{1}{\epsilon_1}(y - \hat{x}_1) \\ \dot{\hat{x}}_2 =& \hat{\gamma} + \frac{1}{1 - A^2\cos^2(\hat{x}_1)}u + a_1\frac{1}{\epsilon_1^2}(y - \hat{x}_1) \\ \dot{\hat{\gamma}} =& M(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, u) + a_0\frac{1}{\epsilon_1^3}(y - \hat{x}_1) \\ \dot{\hat{x}}_3 =& \hat{x}_4 + b_1\frac{1}{\epsilon_2}(\lambda - \hat{x}_3) \\ \dot{\hat{x}}_4 =& \frac{A\hat{x}_2^2\sin(\hat{x}_1) - \hat{x}_3}{1 - A^2\cos^2(\hat{x}_1)} - \frac{A\cos(\hat{x}_1)}{1 - A^2\cos^2(\hat{x}_1)}u + b_0\frac{1}{\epsilon_2^2}(\lambda - \hat{x}_3),\end{aligned} \tag{5.34}$$

where the function λ in the above CHGO is understood to be evaluated as $\lambda = \lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma})$. The two sub-HGOs (5.32) and (5.33), which make up the CHGO (5.34) are interconnected in a rather complicated way, because $\hat{x}_3$ and $\hat{x}_4$ appear in (5.32) and $\hat{x}_1$ appears in (5.33). However, as $\epsilon_1, \epsilon_2 \to 0$, the dominant interconnection is through the λ function from the first sub-HGO to the second sub-HGO.

5.5.2 Estimation Error Dynamics

We define the estimation errors to be $\tilde{x}_i := x_i - \hat{x}_i$ for $i = 1, 2, 3, 4$ and $\tilde{\gamma} := \gamma - \hat{\gamma}$. Differentiating these variables, we obtain the following error dynamics

$$
\begin{aligned}
\dot{\tilde{x}}_1 =& \tilde{x}_2 - a_2 \frac{1}{\epsilon_1} \tilde{x}_1 \\
\dot{\tilde{x}}_2 =& \tilde{\gamma} + \frac{1}{1 - A^2 \cos^2(x_1)} u - \frac{1}{1 - A^2 \cos^2(\hat{x}_1)} u - a_1 \frac{1}{\epsilon_1^2} \tilde{x}_1 \\
\dot{\tilde{\gamma}} =& M(x, u) - M(\hat{x}, u) - a_0 \frac{1}{\epsilon_1^3} \tilde{x}_1 \\
\dot{\tilde{x}}_3 =& \tilde{x}_4 - b_1 \frac{1}{\epsilon_2} \left(\lambda - \hat{x}_3\right) \\
\dot{\tilde{x}}_4 =& \left(\frac{A x_2^2 \sin(x_1) - x_3}{1 - A^2 \cos^2(x_1)} - \frac{A \cos(x_1)}{1 - A^2 \cos^2(x_1)} u \right) - \\
& - \left(\frac{A \hat{x}_2^2 \sin(\hat{x}_1) - \hat{x}_3}{1 - A^2 \cos^2(\hat{x}_1)} - \frac{A \cos(\hat{x}_1)}{1 - A^2 \cos^2(\hat{x}_1)} u \right) - b_0 \frac{1}{\epsilon_2^2} \left(\lambda - \hat{x}_3\right),
\end{aligned}
\tag{5.35}
$$

where λ is understood to be evaluated as $\lambda = \lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma})$.

We introduce the virtual measurement error, $\tilde{\lambda}$, which is defined to be the difference between the true state, x_3, and its virtual measurement, $\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma})$,

$$
\begin{aligned}
\tilde{\lambda}(\tilde{x}_1, \tilde{x}_2, \tilde{\gamma}, x_1, x_2, \gamma) :=& \lambda(x_1, x_2, \gamma) - \lambda(x_1 - \tilde{x}_1, x_2 - \tilde{x}_2, \gamma - \tilde{\gamma}) \\
=& \lambda(x_1, x_2, \gamma) - \lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) \\
=& x_3 - \lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}).
\end{aligned}
$$

We will suppress the arguments to $\tilde{\lambda}$ so that when we write $\tilde{\lambda}$, it is understood that we mean $\tilde{\lambda}(\tilde{x}_1, \tilde{x}_2, \tilde{\gamma}, x_1, x_2, \gamma)$. Using this definition, the term $\lambda - \hat{x}_3$ from the error dynamics (5.35) is written as

$$
\begin{aligned}
\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) - \hat{x}_3 &= \lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) - \lambda(x_1, x_2, \gamma) + \lambda(x_1, x_2, \gamma) - \hat{x}_3 \\
&= \lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) - \lambda(x_1, x_2, \gamma) + x_3 - \hat{x}_3 \\
&= \lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma}) - \lambda(x_1, x_2, \gamma) + \tilde{x}_3 \\
&= \tilde{x}_3 - \tilde{\lambda}.
\end{aligned}
$$

Using this, we rewrite the error dynamics (5.35) to make it look like a linear, stable system, which is perturbed by a nonlinear perturbation

$$\frac{d}{dt}\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{\gamma} \\ \tilde{x}_3 \\ \tilde{x}_4 \end{pmatrix} = \begin{pmatrix} -a_2\frac{1}{\epsilon_1} & 1 & 0 & 0 & 0 \\ -a_1\frac{1}{\epsilon_1^2} & 0 & 1 & 0 & 0 \\ -a_0\frac{1}{\epsilon_1^3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -b_1\frac{1}{\epsilon_2} & 1 \\ 0 & 0 & 0 & -b_0\frac{1}{\epsilon_2^2} & 0 \end{pmatrix}\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{\gamma} \\ \tilde{x}_3 \\ \tilde{x}_4 \end{pmatrix} + \delta + \Delta, \tag{5.36}$$

where δ and Δ are perturbation terms, given by

$$\begin{aligned} \delta &= \begin{pmatrix} 0 \\ \frac{1}{1-A^2\cos^2(x_1)}u - \frac{1}{1-A^2\cos^2(\hat{x}_1)}u \\ M(x,u) - M(\hat{x},u) \\ 0 \\ \left(\frac{Ax_2^2\sin(x_1)-x_3}{1-A^2\cos^2(x_1)} - \frac{A\cos(x_1)}{1-A^2\cos^2(x_1)}u\right) - \left(\frac{A\hat{x}_2^2\sin(\hat{x}_1)-\hat{x}_3}{1-A^2\cos^2(\hat{x}_1)} - \frac{A\cos(\hat{x}_1)}{1-A^2\cos^2(\hat{x}_1)}u\right) \end{pmatrix}, \\ \Delta &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ b_1\frac{1}{\epsilon_2}\tilde{\lambda} \\ b_0\frac{1}{\epsilon_2^2}\tilde{\lambda} \end{pmatrix}. \end{aligned} \tag{5.37}$$

The good thing about the perturbation terms δ and Δ is that they are vanishing perturbations, meaning that when the estimation error is zero (i.e., when $\hat{x} = x$ and $\hat{\gamma} = \gamma$, or equivalently, when $\tilde{x} = 0$ and $\tilde{\gamma} = 0$), then the perturbation terms are equal to zero

$$\left(\forall x \in \mathbb{R}^4\right)\ (\forall \gamma \in \mathbb{R})\ (\forall u \in \mathbb{R}) \qquad \delta\big|_{\tilde{x}=0,\tilde{\gamma}=0} = 0 \ \text{ and } \ \Delta\big|_{\tilde{x}=0,\tilde{\gamma}=0} = 0.$$

This means that the point $(\tilde{x}, \tilde{\gamma}) = 0$ is an equilibrium point of the error dynamics.

The bad thing about the perturbation term Δ is that the last two components of Δ contain terms multiplied by $1/\epsilon_2$ and $1/\epsilon_2^2$. Namely the two terms: $b_1\frac{1}{\epsilon_2}\tilde{\lambda}$ and $b_0\frac{1}{\epsilon_2^2}\tilde{\lambda}$ are problematic because they increase in magnitude as $\epsilon_2 \to 0$. Fortunately, these high-gain perturbation terms only depend on the estimation error of the first sub-HGO, which decreases to order $O(\epsilon_1)$ values.

The two high-gain perturbation terms $\frac{b_1}{\epsilon_2}\tilde{\lambda}$ and $\frac{b_0}{\epsilon_2^2}\tilde{\lambda}$ are not unexpected. Indeed, if the estimation errors: $\tilde{x}_1$, $\tilde{x}_2$ and $\tilde{\gamma}$, of the first sub-HGO are not small then the virtual measurement error, $\tilde{\lambda}$, is not guaranteed to be small. This means that the second sub-HGO will receive an incorrect virtual output signal, $\lambda(\hat{x}_1, \hat{x}_2, \hat{\gamma})$, which is not close to $\lambda(x_1, x_2, \gamma) = x_3$. When the virtual output is incorrect, i.e., when $\tilde{\lambda}$ is not small, then increasing the gain of the second sub-HGO will not help it in reducing estimation error, since it will simply be differentiating an incorrect virtual output signal. Thus, we see that for the CHGO (5.34) to work, the estimation errors: $\tilde{x}_1$, $\tilde{x}_2$, $\tilde{\gamma}$, of the first sub-HGO must decay to sufficiently small values, so that $\tilde{\lambda}$ becomes sufficiently small. Only after this has happened, can the second sub-HGO begin to work properly.

5.5.3 Simulation

We simulate the CHGO (5.34), applied to the TORA system (5.24). The physical parameters of the TORA system are taken to be

$$k = 10, \qquad m_r = 15, \qquad m_c = 5, \qquad l_r = 7, \qquad I = 1.$$

With these parameter values, the transformation into dimensionless variables is given by

$$A = 0.8654, \qquad \tau = 0.7071t, \qquad x_3 = 0.1648d, \qquad u = 0.0027N.$$

The following gains are chosen for the CHGO

$$a_0 = 1, \qquad a_1 = 3, \qquad a_2 = 3, \qquad b_0 = 1, \qquad b_1 = 2, \qquad \epsilon_1 = 0.49, \qquad \epsilon_2 = 0.70.$$

The TORA system (5.24) is initialized at

$$x_1 = -0.5, \qquad x_2 = -0.1, \qquad x_3 = 0.5, \qquad x_4 = 0.$$

The CHGO (5.34) is initialized at

$$\hat{x}_1 = 0, \qquad \hat{x}_2 = 0, \qquad \hat{\gamma} = 0, \qquad \hat{x}_3 = 0, \qquad \hat{x}_4 = 0.$$

The control input, u, is chosen in such a way that the state x does not leave the set Ω_x during the simulation, i.e., such that $\cos(x_1) \neq 0$ holds,

$$u = \begin{cases} 0.06 & \text{for } \tau < 9 \\ 0.07 & \text{for } \tau \geq 9 \text{ and } \tau < 15 \\ 0.1 & \text{for } \tau \geq 15 \end{cases}$$

Figure 5.4 shows simulation of the CHGO without measurement noise, i.e., taking $y = x_1$. Figure 5.5 shows simulation of the CHGO with additive measurement noise, i.e., taking $y = x_1 + n$, where n is normally distributed.

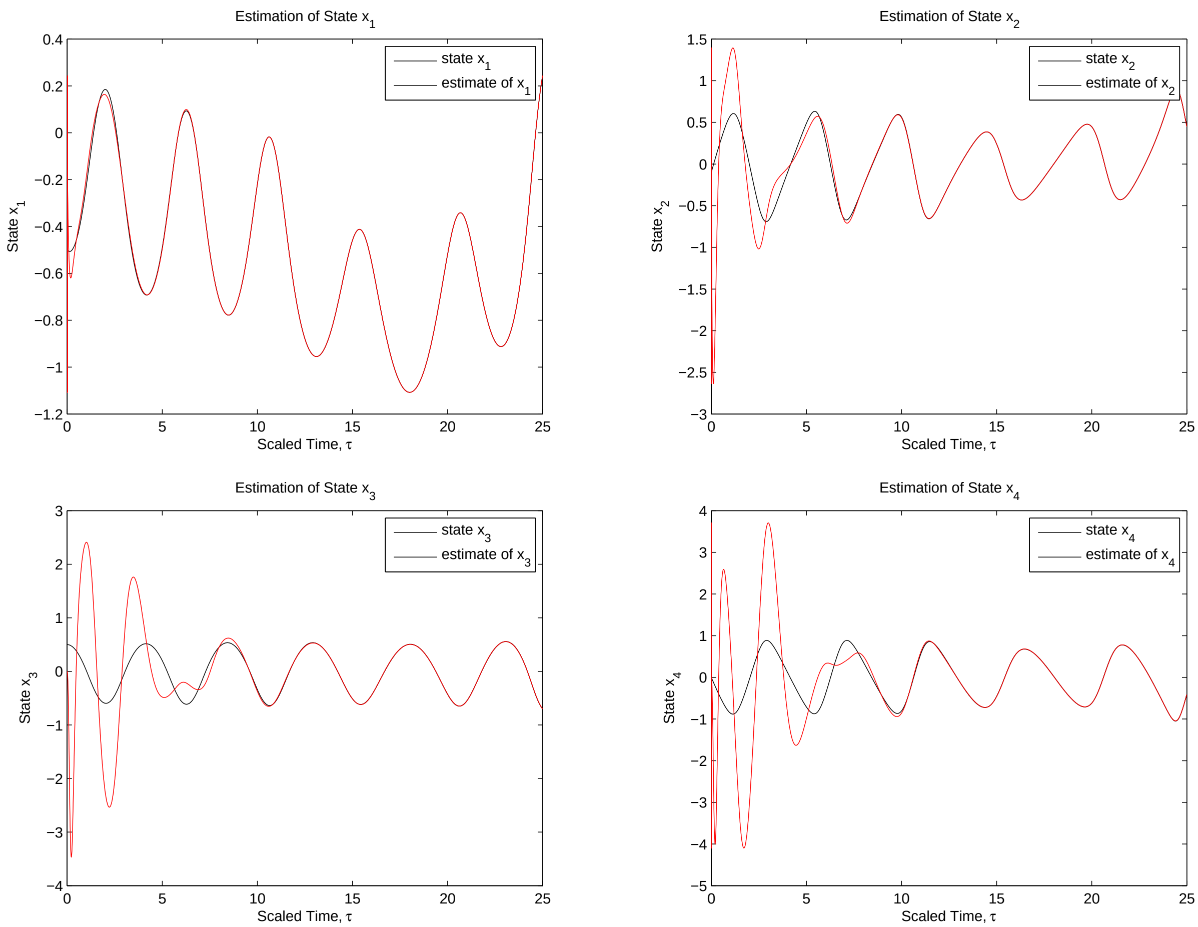

Figure 5.4: Simulation of the CHGO (5.34) applied to the TORA system (5.24), without measurement noise.

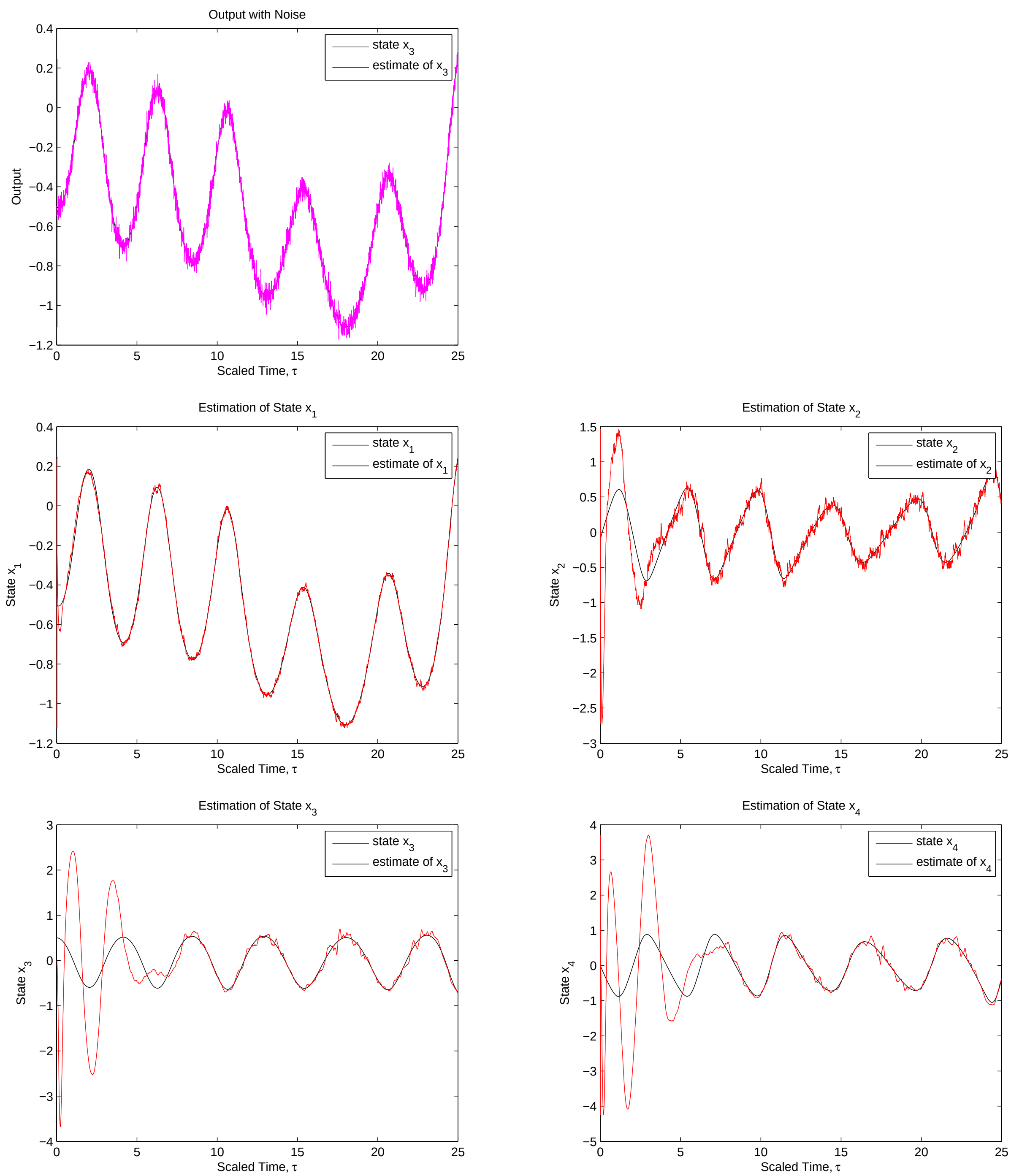

Figure 5.5: Simulation of the CHGO (5.34) applied to the TORA system (5.24), with measurement noise.

Chapter 6

Future Research

The following future research directions could be considered:

(i) **Relativistic observer on $\mathsf{SO}(3,1)$.** The Lorentz group $\mathsf{SO}(3,1)$ is the Lie group of proper Lorentz transformations of Minkowski space-time. The group $\mathsf{SO}(3,1)$ is 6 dimensional, 3 of the dimensions account for rotations, while the other 3 dimensions account for Lorentz boosts. Recall from Section 3.7, that an element $X \in \mathsf{SO}(3)$ "encodes" orientation, while an element $u \in \mathrm{Lie}(\mathsf{SO}(3))$ "encodes" angular velocities of rigid-body $\{A\}$ relative to rigid-body $\{B\}$. In much the same way, an element $X \in \mathsf{SO}(3,1)$ "encodes" orientation and relativistic velocity (a.k.a. "boost"), while an element $u \in \mathrm{Lie}(\mathsf{SO}(3,1))$ "encodes" angular velocities and linear accelerations of rigid-body $\{A\}$ relative to rigid-body $\{B\}$. For objects that move at relativistic speeds, Newtonian mechanics become less accurate due to effects of Special Relativity. To build a sensor fusion algorithm that takes into account the effects of Special Relativity, one could try to use the LFSO proposed in this chapter, applied to Lie group $\mathsf{SO}(3,1)$. Measurement of the boost between $\{A\}$ and $\{B\}$ one could likely rely on the physical phenomenon called "red-shift", where the color of an object changes depending on how fast this object is moving towards or away from the observer. Obviously such a measurement of boosts would be highly noisy, which is where the full-state observer would become useful in filtering out the noise. The input u encodes linear accelerations (in 3 dimensions) and angular velocities (in 3 axis) of frame $\{A\}$ relative to frame $\{B\}$, which is something that can be measured with a set of accelerometers and angular-rate gyroscopes.

(ii) **Efficient discrete-time implementation on SU**(2) representing the orientation of a flying robot by means of a rotation matrix is not always convenient, because the rotation matrix needs to be "orthogonalized" periodically to make sure that it does not drift off the $\mathsf{SO}(3)$ submanifold due to numerical errors. It is well known that the Lie group $\mathrm{SU}(2) \subset \mathsf{GL}(2,\mathbb{C}) \subset \mathsf{GL}(4,\mathbb{R})$ is a double cover for the Lie group $\mathsf{SO}(3)$, which implies that SU(2) can also be used to represent the orientation of a rigid-body. The group SU(2) can be shown to be diffeomorphic to a 3-sphere embedded into 4 dimensions. Just like any sphere, it has the notion of a "geodesic along the sphere", which is the shortest path along the sphere between any two orientations. There is a simple formula to compute the motion along this geodesic. The formula is called spherical linear interpolation (SLERP) and it is derived in [39]. We conjecture that the differential equation (3.10) always evolves on the same geodesic curve that it started on. In other words, we conjecture that, for the Lie group SU(2), the one-parameter subgroup coincides with the geodesic along the sphere. Using the SLERP formula, a very efficient and accurate attitude estimation algorithm can be built for discrete time implementation of the algorithm proposed in this chapter. This algorithm treats the system's state (orientation) as evoling on a sphere in 4 dimensions.

(iii) **Control of systems on Lie groups.** Extend the results of Section 3.8 to design controllers that can converge to non-constant reference trajectories. It would be useful to design a Lie group analogue of the Proportional, Integral, Derivative (PID) controller that is often used for controlling linear systems. Also, one could try to design controllers for "dynamical" systems on Lie groups, which were discussed in Chapter 4.

(iv) **Proof of CHGO convergence.** Find out under what conditions the CHGO converges and how the gains of the sub-observers should be chosen to have better convergence.

APPENDICES

Appendix A

Stability of Equilibria

A.1 Definitions

Consider the system

$$\dot{x} = f(x,t), \tag{A.1}$$

where $x \in \mathbb{R}^n$ is the state, and $f(x,t)$ is the system vector field. Assume that the system has an equilibrium point at $x^* = 0$, i.e., assume that $(\forall t \in \mathbb{R}) : f(0,t) = 0$.

Definition A.1.1. *Equilibrium point $x^* = 0$ of (A.1) is (Lyapunov) stable if any solution $x(t)$ of (A.1) satisfies*

$$(\forall t_0 \in \mathbb{R})\ (\forall \epsilon > 0)\ (\exists \delta > 0)\ (\forall \|x(t_0)\| < \delta)\ (\forall t \geq t_0) \quad \|x(t)\| < \epsilon.$$

Definition A.1.2. *Equilibrium point $x^* = 0$ of (A.1) is asymptotically stable if it is stable and if any solution $x(t)$ of (A.1) satisfies*

$$(\forall t_0 \in \mathbb{R})\ (\exists \delta > 0)\,(\|x(t_0)\| < \delta) \Rightarrow (x(t) \to 0,\ as\ t \to \infty)\,.$$

Definition A.1.3. *Equilibrium point $x^* = 0$ of (A.1) is exponentially stable if it is asymptotically stable and if any solution $x(t)$ of (A.1) satisfies*

$$(\forall t_0 \in \mathbb{R})\ (\exists \delta > 0)\ (\exists m, \alpha > 0)\,(\|x(t_0)\| < \delta) \Rightarrow (\forall t \geq t_0)\ \|x(t)\| < m e^{-\alpha(t-t_0)} \|x(t_0)\|.$$

A.2 Stability Under Diffeomorphism

In this section, we will show that stability of equilibria is preserved under diffeomorphisms. To this end, consider a change of coordinates $z = \Psi(x)$, where $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ is a local diffeomorphism onto its image at $x = 0$ and is such that $\Psi(0) = 0$. In the z-coordinates, the system (A.1) transforms to

$$\dot{z} = g(z,t), \tag{A.2}$$

where $z \in \mathbb{R}^n$ is the state, and $g(z,t)$ is the vector field. The point $z^* = 0$ is an equilibrium point of (A.2), i.e., $(\forall t \in \mathbb{R}) : g(0,t) = 0$.

Lemma A.2.1. *(i) $x^* = 0$ is stable if and only if $z^* = 0$ is stable.*

(ii) $x^* = 0$ is asymptotically stable if and only if $z^* = 0$ is asymptotically stable.

(iii) $x^* = 0$ is exponentially stable if and only if $z^* = 0$ is exponentially stable.

Proof. Choose a neighbourhood $U \subseteq \mathbb{R}^n$ of $x^* = 0$, and a neighbourhood $V \subseteq \mathbb{R}^n$ of $z^* = 0$, sufficiently small, so that $\Psi : U \to V$ is a diffeomorphism.

(i) ($\Leftarrow$) Assume that $x^* = 0$ is stable. Let $t_0 \in \mathbb{R}$ and $\epsilon_z > 0$ be arbitrary. Using continuity of Ψ, choose $\epsilon_x > 0$ sufficiently small so that

$$\|x\| < \epsilon_x \Rightarrow \|\Psi(x)\| < \epsilon_z$$

and using stability of $x^* = 0$, choose $\delta_x > 0$ sufficiently small so that

$$\|x(t_0)\| < \delta_x \Rightarrow (\forall t \geq t_0 : \|x(t)\| < \epsilon_x),$$

now using continuity of Ψ^{-1}, choose $\delta_z > 0$ sufficiently small so that

$$\|z\| < \delta_z \Rightarrow (z \in V, \|\Psi^{-1}(z)\| < \delta_x).$$

Then, we have that

$$\|z(t_0)\| < \delta_z \Rightarrow (\forall t \geq t_0 : \|z(t)\| < \epsilon_z).$$

The proof of the converse is symmetrical and therefore omitted.

(ii) ($\Leftarrow$) Assume that $x^* = 0$ is asymptotically stable and let $t_0 \in \mathbb{R}$ be arbitrary, then

$$(\exists \delta > 0): \quad (\|x(t_0)\| < \delta) \Rightarrow (x(t) \to 0, \text{ as } t \to \infty).$$

Since $\Psi^{-1} : V \to U$ is smooth, it is also locally Lipschitz, so there exists a neighbourhood, V_0 of $0 \in V$ and a Lipschitz constant $L > 0$, such that

$$(\forall z \in V_0): \quad \|\Psi^{-1}(z)\| \leq L\|z\|.$$

Thus, if $\|z(t_0)\| < \frac{\delta}{L}$, then $\|\Psi^{-1}(z(t_0))\| < \delta$. By the asymptotic stability of $x^* = 0$, this implies that $\Psi^{-1}(z(t)) \to 0$ and therefore $z(t) \to 0$ as $t \to \infty$. Proof of the converse is symmetrical and therefore omitted.

(iii) ($\Leftarrow$) Assume that $x^* = 0$ is exponentially stable and let $t_0 \in \mathbb{R}$ be arbitrary, then

$$(\exists \delta > 0), (\exists m, \alpha > 0): \quad (\|x(t_0)\| < \delta) \Rightarrow \left((\forall t \geq t_0) : \|x(t)\| < me^{-\alpha(t-t_0)}\|x(t_0)\|\right).$$

Since $\Psi^{-1} : V \to U$ is smooth, it is also locally Lipschitz, so there exists a neighbourhood, V_0 of $0 \in V$ and a Lipschitz constant $L > 0$, such that

$$(\forall z \in V_0): \quad \|\Psi^{-1}(z)\| \leq L\|z\|$$

and since $\Psi : U \to V$ is smooth, it is also locally Lipschitz, so there exists a neighbourhood, U_0 of $0 \in U$ and a Lipschitz constant $K > 0$, such that

$$(\forall x \in U_0): \quad \|\Psi(x)\| \leq K\|x\|.$$

Thus, if $\|z(t_0)\| < \frac{\delta}{L}$, then $\|\Psi^{-1}(z(t_0))\| < \delta$. By the exponential stability of $x^* = 0$, this implies that

$$\|\Psi^{-1}(z(t))\| < me^{-\alpha(t-t_0)}\|\Psi^{-1}(z(t_0))\|)$$

and therefore

$$\|z(t)\| < KLme^{-\alpha(t-t_0)}\|z(t_0)\|.$$

Proof of the converse is symmetrical and therefore omitted.

□

A.3 Linear Systems and Linearization

Consider the nonlinear, time-invariant system

$$\dot{x} = f(x), \tag{A.3}$$

where $x \in \mathbb{R}^n$ and $f : D \to \mathbb{R}^n$ is continuously differentiable and D is a neighbourhood of $x^* = 0$. Assume that the Jacobian matrix $\frac{\partial f}{\partial x}$ is bounded and Lipschitz on D. Let

$$A := \frac{\partial f}{\partial x}(x)\Big|_{x=0},$$

then the following system is the linearization of (A.3), around the equilibrium point $x^* = 0$,

$$\dot{x} = Ax. \tag{A.4}$$

In this thesis, we will require the following simplified versions of [22, Theorem 4.5] and [22, Theorem 4.13].

Lemma A.3.1 ([22, Theorem 4.5]). *The equilibrium point $x^* = 0$ of the linear system (A.4) is globally exponentially stable (we also say that the system (A.4) is exponentially stable), if and only if all the eigenvalues of A have strictly negative real parts.*

Lemma A.3.2 ([22, Theorem 4.7]). *The equilibrium point $x^* = 0$ of the nonlinear system (A.3) is locally exponentially stable if the linearization (A.4) is exponentially stable.*

Appendix B

Symbols and Abbreviations

HGO High-gain observer, page 18.

CHGO Composite high-gain observer, page 66.

LFSO Lie-group full-state observer, page 25.

LPSO Lie-group partial-state observer, page 49.

$O(\epsilon)$ A function of ϵ, which is upper bounded by ϵ, as $\epsilon \to 0$, page 10.

$dh(x)$ Differential of a vector-valued function $h : \mathbb{R}^n \to \mathbb{R}^m$, page 10.

$L_f h(x)$ Lie derivative of a scalar-valued function $h : \mathbb{R}^n \to \mathbb{R}$, along the vector field $f : \mathbb{R}^n \to \mathbb{R}^n$, page 10.

$L_f^k h(x)$ Repeated k-times Lie derivative of a scalar-valued function $h : \mathbb{R}^n \to \mathbb{R}$, along the vector field $f : \mathbb{R}^n \to \mathbb{R}^n$, page 10.

$\mathbb{R}^+$ Real numbers, with addition as the group operation, page 11.

$\mathsf{GL}(n, \mathbb{R})$ The general linear Lie group of all invertible $n \times n$ matrices, with real coefficients, page 11.

$\mathsf{M}(n, \mathbb{R})$ Lie algebra of $n \times n$ matrices, with real coefficients, page 11.

$[A, B]$ Lie bracket of two $n \times n$ matrices, A and B, page 11.

I_n The $n \times n$ identity matrix, page 11.

0_n The $n \times n$ zero matrix, page 11.

$\mathrm{Ad}_X(A)$ Matrix adjoint of A by X, given by $X^{-1}AX$, page 11.

$\|x\|$ The Euclidean norm of a vector $x \in \mathbb{R}^n$, page 12.

$\|A\|$ The induced operator norm of matrix $A \in \mathbb{R}^{n \times n}$, page 12.

$B(A, r)$ The open ball of radius $r > 0$, centered at $A \in \mathbb{R}^{n \times n}$, page 12.

$\exp(A)$ Matrix exponential of $A \in \mathbb{R}^{n \times n}$, page 13.

$\mathrm{Lie}(G)$ The Lie algebra of a linear Lie group G, page 14.

$\log(X)$ Matrix logarithm of $X \in B(I_n, 1)$, page 14.

$T_X G$ Tangent space to Lie group $G \subseteq \mathsf{GL}(n, \mathbb{R})$, at $X \in G$, page 16.

$\mathsf{SO}(n)$ The special orthogonal Lie group of $n \times n$, orthogonal matrices with unit determinant, page 16.

$\mathsf{SL}(n)$ The special linear Lie group of $n \times n$ matrices with unit determinant, page 16.

E_l Left-invariant Lie group error, page 26.

E_r Right-invariant Lie group error, page 26.

e_l Log left-invariant Lie algebra error, page 28.

e_r Log right-invariant Lie algebra error, page 28.

References

[1] R. Mahony, T. 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, dec. 2005.

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

[3] Grant Baldwin, Robert Mahony, Jochen Trumpf, Tarek Hamel, and Thibault Cheviron. Complementary filter design on the special Euclidean group SE(3). In *European Control Conference*, 2007.

[4] Minh-Duc Hua, Mohammad Zamani, Jochen Trumpf, Robert Mahony, and Tarek Hamel. Observer design on the special Euclidean group SE(3). In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 8169 –8175, dec. 2011.

[5] E. Malis, T. Hamel, R. Mahony, and P. Morin. Dynamic estimation of homography transformations on the special linear group for visual servo control. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 1498 –1503, may 2009.

[6] C. 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, feb. 2010.

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

[8] Philippe Jouan. On the existence of observable linear systems on Lie groups. *Journal of Dynamical and Control Systems*, 15:263–276, 2009. 10.1007/s10883-009-9063-2.

[9] P. Jouan. Controllability of linear systems on Lie groups. *Journal of Dynamical and Control Systems*, 17:591–616, 2011. 10.1007/s10883-011-9131-2.

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

[11] D. Firoozi and M. Namvar. Noise analysis in satellite attitude estimation using angular rate and a single vector measurement. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 7476 –7481, dec. 2011.

[12] G. Besançon. *Nonlinear Observers and Applications.* Lecture Notes in Control and Information Sciences. Springer-Verlag Berlin Heidelberg, 2007.

[13] H.K. Khalil. High-gain observers in nonlinear feedback control. In *Control, Automation and Systems, 2008. ICCAS 2008. International Conference on*, pages xlvii –lvii, oct. 2008.

[14] J.P. Gauthier, H. Hammouri, and S. Othman. A simple observer for nonlinear systems applications to bioreactors. *Automatic Control, IEEE Transactions on*, 37(6):875 – 880, jun 1992.

[15] F. Deza, E. Busvelle, J.P. Gauthier, and D. Rakotopara. High gain estimation for nonlinear systems. *Systems & Control Letters*, 18(4):295 – 299, 1992.

[16] G. Bornard and H. Hammouri. A high gain observer for a class of uniformly observable systems. In *Decision and Control, 1991., Proceedings of the 30th IEEE Conference on*, pages 1494 –1496 vol.2, dec 1991.

[17] F. Esfandiari and H. K. Khalil. Output feedback stabilization of fully linearizable systems. *International Journal of Control*, 56(15):1007–1037, 1992.

[18] Andrew Teel and Laurent Praly. Global stabilizability and observability imply semi-global stabilizability by output feedback. *Systems &; Control Letters*, 22(5):313 – 325, 1994.

[19] A.N. Atassi and H.K. Khalil. A separation principle for the stabilization of a class of nonlinear systems. *Automatic Control, IEEE Transactions on*, 44(9):1672 –1687, sep 1999.

[20] A. Isidori. A tool for semi-global stabilization of uncertain non-minimum-phase nonlinear systems via output feedback. *IEEE Transactions on Automatic Control*, 45(10):1817 – 1827, October 2000.

[21] S. Nazrulla and H.K. Khalil. Robust stabilization of non-minimum phase nonlinear systems using extended high-gain observers. *Automatic Control, IEEE Transactions on*, 56(4):802 –813, april 2011.

[22] H. K. Khalil. *Nonlinear Systems.* Prentice Hall, Upper Saddle River, NJ, 3 edition, 2002.

[23] H. K. Khalil. Robust servomechanism output feedback controllers for feedback linearizable systems. *Automatica*, 30(10):1587–1599, October 1994.

[24] Ezio Malis, Tarek Hamel, Robert Mahony, and Pascal Morin. Estimation of homography dynamics on the special linear group. In Graziano Chesi and Koichi Hashimoto, editors, *Visual Servoing via Advanced Numerical Methods*, volume 401 of *Lecture Notes in Control and Information Sciences*, pages 133–150. Springer Berlin / Heidelberg, 2010.

[25] Kokotović Petar, Khalil Hassan K., and O'Reilly John. *Singular Perturbation Methods in Control: Analysis and Design.* Society for Industrial and Applied Mathematics, 1999.

[26] J. Faraut. *Analysis on Lie Groups: An Introduction.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2008.

[27] M. Vidyasagar. *Nonlinear systems analysis.* SIAM, 2^{nd} edition, 2002.

[28] F. Bullo and Andrew D. Lewis. *Geometric control of mechanical systems: modeling, analysis, and design for simple mechanical control systems.* Texts in applied mathematics. Springer, 2005.

[29] Nicholas J. Higham. *Functions of Matrices: Theory and Computation.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

[30] W. Rossmann. *Lie groups: An introduction through linear groups.* Oxford, 2002.

[31] A. Dabroom and H.K. Khalil. Numerical differentiation using high-gain observers. In *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, volume 5, pages 4790 –4795 vol.5, dec 1997.

[32] Luma K. Vasiljevic and Hassan K. Khalil. Error bounds in differentiation of noisy signals by high-gain observers. *Systems & Control Letters*, 57(10):856 – 862, 2008.

[33] M. Vidyasagar Mark W. Spong, Seth Hutchinson. *Robot Modeling and Control.* John Wiley & Sons., Hoboken, NJ, 2006.

[34] R. W. Brockett. System theory on group manifolds and coset spaces. *Society for Industrial and Applied Mathematics*, 1972.

[35] Olivier Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. Technical Report RR-0856, INRIA, June 1988.

[36] Ezio Malis and Manuel Vargas. Deeper understanding of the homography decomposition for vision-based control. Rapport de recherche RR-6303, INRIA, 2007.

[37] S. Benhimane and E. Malis. Homography-based 2D visual tracking and servoing. *Int. J. Rob. Res.*, 26(7):661–676, July 2007.

[38] Chih-Jian Wan, Dennis S. Bernstein, and Vincent T. Coppola. Global stabilization of the oscillating eccentric rotor. *Nonlinear Dynamics*, 10:49–62, 1996.

[39] Ken Shoemake. Animating rotation with quaternion curves. *SIGGRAPH Comput. Graph.*, 19(3):245–254, July 1985.