N°d'ordre NNT : **2021LYSE1085**

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale 160
Electronique, Electrotechnique, Automatique

Spécialité de doctorat : **Automatique**
Discipline : **Mathématiques**

Soutenue publiquement le 19/05/2021, par :
Lucas Brivadis

Stabilisation des systèmes de contrôle non-uniformément observables et observateurs de dimension infinie

---

Devant le jury composé de :

Maschke, Bernhard, Professeur des universités, UCBL1, LAGEPP	Président
Prieur, Christophe, Directeur de recherche CNRS, GIPSA-lab	Rapporteur
Wirth, Fabian, Professeur des universités, Université de Passau	Rapporteur
Bernard, Pauline, Maître de conférences, MINES ParisTech, CAS	Examinatrice
Coron, Jean-Michel, Professeur des universités, UPMC, LJLL	Examinateur
Jacob, Birgit, Professeur des universités, Université de Wuppertal	Examinatrice
Andrieu, Vincent, Directeur de recherche CNRS, LAGEPP	Directeur de thèse
Serres, Ulysse, Maître de conférences, UCBL1, LAGEPP	Co-encadrant de thèse
Gauthier, Jean-Paul, Professeur émérite, UTLN, LIS	Co-encadrant de thèse

Université Claude Bernard – LYON 1

Administrateur provisoire de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Pierre ROLLAND

COMPOSANTES SANTE

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne : Mme Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen : M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur : M. Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directrice : Mme Christine VINCIGUERRA

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOISNE
Institut National du Professorat et de l'Education	Administrateur Provisoire : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI

Lucas BRIVADIS

LAGEPP, Université Lyon 1, 43 Boulevard du 11 Novembre 1918, Bâtiment CPE, 69622 Villeurbanne Cedex, France

Email: *lucas.brivadis@gmail.com*
Homepage: *https://sites.google.com/view/lucas-brivadis*

Keywords. Stabilization, Output feedback, Observability, Dissipative systems, Luenberger observers, Infinite-dimensional systems, Back and Forth Nudging, Crystallization processes.

Mots clés. Stabilisation, Bouclage de sortie, Observabilité, Systèmes dissipatifs, Observateurs de Luenberger, Systèmes de dimension infinie, Back and Forth Nudging, Procédés de cristallisation.

Stabilization of non-uniformly observable control systems and infinite-dimensional observers

Ô mathématiques concises, par l'enchaînement rigoureux de vos propositions tenaces et la constance de vos lois de fer, vous faites luire, aux yeux éblouis, un reflet puissant de cette vérité suprême dont on remarque l'empreinte dans l'ordre de l'univers.

Comte de Lautréamont, *Les Chants de Maldoror*, Chant deuxième

Remerciements

Je remercie Christophe Prieur et Fabian Wirth pour le travail qu'ils ont accompli en tant que rapporteurs de cette thèse ainsi que Pauline Bernard, Jean-Michel Coron et Birgit Jacob pour leur participation au jury de soutenance et Bernhard Maschke d'avoir présidé ce jury.

Je tiens à remercier ceux qui m'ont guidé tout au long du doctorat: Vincent Andrieu, Ulysse Serres et Jean-Paul Gauthier. Travailler à vos côtés est un plaisir et une chance. Merci pour votre enthousiasme scientifique et votre disponibilité sans faille. Si chacun d'entre vous est un chercheur exceptionnel, c'est votre complémentarité qui m'a le plus inspiré. Merci d'avoir partagé avec moi votre passion et vos connaissances. Grâce à vous, je garderai en mémoire un souvenir heureux de ces années de thèse, des travaux exploratoires au tableau jusqu'au minutieux réglage des arguments techniques. Et même si, dans les années qui viennent, nos rencontres se font plus rares, je compte bien continuer à travailler avec vous trois, sur les mêmes questions ou sur d'autres sujets.

Je remercie également Ludovic Sacchelli, frère d'arme dans la guerre contre les singularités d'observabilité. J'ai beaucoup appris à tes côtés. Collaborer avec toi fut une opportunité formidable, et je me félicite que cela perdure ! Je te souhaite le meilleur pour les épreuves à venir. Durant cette thèse, j'ai également pu collaborer avec Daniele Astolfi et Swann Marx. Merci de m'avoir proposé de me joindre à vous. Ce fut une expérience enrichissante, et je serais heureux de travailler à nouveau avec vous deux ! Je remercie Élodie Chabanon, Émilie Gagnière et Noureddine Lebaz qui m'ont initié, dès mon stage au LAGEPP, aux problématiques de la cristallisation. C'est toute une partie de cette thèse qui est née de nos échanges, avec Vincent et Ulysse. Merci à vous.

Un grand merci à tous les membres du LAGEPP, qui m'ont accueilli avec bienveillance (M. Nadri, P. Dufour, B. Maschke, B. Hamroun, C.-Z. Xu, N. Chapel), et à tous les doctorants avec qui j'ai partagé ces années (Alexandre, Greta, Maroua, Maya, et tous les autres Lageppiens !). Une mention particulière pour les membres du café-contrôle (Mattia, Samuele, Bertrand, Johan, Mohammed, Steeven, Quentin). Merci à vous pour ces moments d'échange qui élargissent notre curiosité scientifique.

Je pense également aux professeurs de mathématiques dont j'ai eu la chance de suivre les enseignements: P. Paccini et D. Janey d'abord, qui allumèrent la flamme; F. Ridde et D. Choimet ensuite, deux mentors; G. Vial et M. Marion enfin, qui m'ont donné les outils pour entamer sereinement ce doctorat. Merci à vous.

Je remercie chaleureusement tous mes amis qui m'ont accompagné durant cette thèse, et certains depuis bien plus longtemps ! Mes plus vieux amis d'abord: Maxime (vieux frère !), Benjamin, Raphaël, j'ai hâte vous retrouver pour de nouvelles sessions de gaming intensif ! Je pense aussi à tous ceux que je n'ai pas recroisés depuis

trop longtemps (Thibaut, François, Clément, Clémence, etc.). Mes amis du Parc ensuite: Nathanaël, Thibault (un trio légendaire), Pierre, Benjamin, Thomas (vivement que tu me racontes tes aventures hongkongaises), Bérénice, Pierre, Antonin (hâte de vous revoir enfin), et tous les autres ! Mes amis centraliens enfin: Antoine, Antonio, Thierry, j'ai hâte de vous retrouver. Maintenant c'est à vous de préparer la soutenance, courage !

Enfin, ces dernières lignes sont pour ma famille. Si le reste de la thèse vous semble nébuleux, vous trouverez ici l'essentiel de ma pensée: merci à vous tous ! Merci pour votre soutien, vos encouragements, votre présence. Merci à mes grands-parents pour tout ce que vous m'avez transmis: le goût des arts, de la philosophie, de la générosité et de la sagesse. Merci à mes parents pour votre amour indéfectible. Je suis très fier d'être votre fils. Merci à toi, ma sœur. Être ton frère est une joie immense et une intarissable source de fierté. Je vous aime fort.

Abstract

Context

This thesis deals with two different but related topics. In the first part, we are concerned with the problem of dynamic output feedback stabilization. When only part of the state of a control system is known, a stabilizing state feedback cannot be directly implemented. Hence, a common strategy to stabilize the state to some target point is to design an observer system to asymptotically estimate the state by filtering the output online, and to use as an input the stabilizing state feedback applied to the observer. This approach is known to be efficient on uniformly observable systems, that are observable for all inputs. However, it is not generic for nonlinear systems to be uniformly observable when the dimension of the output is less or equal than the dimension of the input. Hence, in the presence of observability singularities, new techniques need to be developed.

In the second part, we focus on the problem of observer design for linear time-varying infinite-dimensional systems. The goal is to design a dynamical system learning the state from the output dynamics. The finite-dimensional notion of observability may be extended in several ways. In particular, one distinguishes exact and approximate observability assumptions. While exponential convergence of Luenberger observers can be proved on exactly observable systems, much less is known for approximate observability-like hypotheses, on which we focus. These observers can also be used in the context of offline reconstruction of initial data. The procedure is based on iterations of forward and backward observers, and named Back and Forth Nudging (BFN). Such methods can be applied to a batch crystallization process, where the state to be estimated is the Particle Size Distribution (PSD).

Main contributions

In Chapter 1, we state the dynamic output feedback stabilization problem and give some necessary conditions. Some results of the existing literature are recalled. We distinguish two main classes of non-uniformly observable systems, depending on whether or not their target point corresponds to an observable input. In Chapter 2, we focus on systems with observable target. The challenge lies in the existence of inputs making the system unobservable, that the closed-loop system may produce during the stabilization. Avoiding these inputs would be a first step in towards the achievement of a generic separation principle. Our main contribution on this problem is stated in [Bri+21b].

Contribution 1. On Single-Input Single-Output (SISO) bilinear systems with observable target, generic perturbations of the feedback law guarantee that the inputs

produced by the closed-loop system render the system observable.

In Chapter 3, we highlight the usefulness of dissipativity properties in the context of output feedback stabilization. Dissipative systems are such that the distance between two trajectories is always non-increasing. They allow to build Luenberger observers with non-increasing error, regardless of the observability properties. Our results are stated in [Sac+20], and a corollary on state feedback stabilizability is proved in [Bri+21c].

Contribution 2. On state-affine dissipative systems that are state feedback stabilizable, 0-detectability is a necessary and sufficient condition for the existence of a globally stabilizing dynamic output feedback.

Chapter 4 investigates systems with unobservable target. We coalesce the insights provided by the previous chapters to come up with guidelines to attack the problem, and illustrate them on examples of linear systems with nonlinear output. In order to take advantage of dissipativity properties on a wider class of systems, we propose an embedding-based strategy: the observer is designed for the embedded system, admitting an observer with dissipative error. This work has resulted in [Bri+20b].

Contribution 3. On examples of nonlinear systems, we illustrate three main guidelines for output feedback stabilization at an unobservable target:

- additive perturbations of the state feedback law yield new observability properties without preventing the stabilization process;
- observers with dissipative error system are robust to observability singularities;
- embeddings into finite or infinite-dimensional systems allow to design Luenberger observers with dissipative error systems.

In Chapter 4, we use infinite-dimensional Luenberger observers under approximate observability hypotheses. This naturally lead us to the second part of this thesis. Our main theoretical results of this part are stated in Chapters 5 and 6.

Contribution 4. Up to a weak detectability assumption, infinite-dimensional Luenberger observers estimate the observable part of the state in the weak topology of the state space. Strong convergence can be obtained with additional assumptions on the error system.

Contribution 5. The convergence results of Contribution 4 can be adapted to the BFN context.

These works have resulted in the publication [Bri+21a]. During this thesis, the problem of infinite-dimensional observer design has been considered on a batch crystallization process. The state to be estimated from different measurements is the Particle Size Distribution (PSD), which satisfies a transport equation. We propose three strategies of estimation, that are stated in Chapter 7 and have resulted in the articles [Bri+20a, BS20, Bri+21a].

Contribution 6. In the context of a batch crystallization process, we propose several strategies to reconstruct the PSD:

- a direct approach based on a Tikhonov regularization, using the knowledge of the Chord Length Distribution (CLD);
- a Kazantzis-Kravaris/Luenberger (KKL) observer, using the knowledge of temperature and solute concentration;
- an infinite-dimensional Luenberger observer, based on Contributions 4 and 5, using the knowledge of the CLD.

Each chapter begins with an independent abstract. A detailed abstract in French is provided at the end of the thesis.

Publications

During this thesis, the following articles have been published or submitted. The submitted articles [MBA20a] and [MBA20b] deal with the problem of output regulation for coupled ODE/PDE systems. This topic is not covered in this thesis, but is related to approximate observability properties investigated in Chapter 5. The recently submitted conference papers [BS21a] and [BS21b] are not discussed in the thesis.

- Journal articles:

  1. L. Brivadis, V. Andrieu, É. Chabanon, É. Gagnière, N. Lebaz, and U. Serres. "New dynamical observer for a batch crystallization process based on solute concentration". *Journal of Process Control* 87 (2020), pp. 17–26. ISSN: 0959-1524. DOI: 10.1016/j.jprocont.2019.12.012.
  2. L. Brivadis, L. Sacchelli, V. Andrieu, J.-P. Gauthier, and U. Serres. "From local to global asymptotic stabilizability for weakly contractive control systems". *Automatica J. IFAC* 124 (2021). ISSN: 0005-1098. DOI: 10.1016/j.automatica.2020.109308.
  3. L. Brivadis, V. Andrieu, U. Serres, and J.-P. Gauthier. "Luenberger observers for infinite-dimensional systems, Back and Forth Nudging, and application to a crystallization process". *SIAM Journal on Control and Optimization* 59.2 (2021), pp. 857–886. DOI: 10.1137/20M1329020.
  4. L. Brivadis, J.-P. Gauthier, L. Sacchelli, and U. Serres. "Avoiding observability singularities in output feedback bilinear systems". *SIAM Journal on Control and Optimization* 59.3 (2021), pp. 1759–1780. DOI: 10.1137/19M1272925.
  5. S. Marx, L. Brivadis, and D. Astolfi. "Forwarding techniques for the global stabilization of dissipative infinite-dimensional systems coupled with an ODE". Submitted to *Mathematics of Control, Signals, and Systems*. Under review. Sept. 2020.
  6. L. Brivadis, J.-P. Gauthier, L. Sacchelli, and U. Serres. "New perspectives on output feedback stabilization at an unobservable target". Submitted to *ESAIM. Control, Optimisation and Calculus of Variations.* Under review. Nov. 2020.

7. L. Brivadis and L. Sacchelli. "New inversion methods for the single/ multi-shape CLD-to-PSD problem with spheroid particles". Submitted to *Journal of Process Control*. Under review. Dec. 2020.

- Conference papers:

1. L. Brivadis, V. Andrieu, and U. Serres. "Luenberger observers for discrete-time nonlinear systems". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. 2019, pp. 3435–3440. DOI: 10.1109/CDC40024.2019.9029220.
2. S. Marx, L. Brivadis, and D. Astolfi. "Forwarding design for stabilization of a coupled transport equation-ODE with a cone-bounded input nonlinearity". In: *2020 59th IEEE Conference on Decision and Control (CDC)*. 2020, pp. 640–645. DOI: 10.1109/CDC42340.2020.9304178.
3. L. Sacchelli, L. Brivadis, V. Andrieu, U. Serres, and J.-P. Gauthier. "Dynamic output feedback stabilization of non-uniformly observable dissipative systems". *IFAC-PapersOnLine* 53.2 (2020). 21th IFAC World Congress, pp. 4923–4928. ISSN: 2405-8963. DOI: https://doi.org/10.1016/j.ifacol.2020.12.1071.
4. L. Brivadis and L. Sacchelli. "A switching technique for output feedback stabilization at an unobservable target". Submitted to *2021 60th IEEE Conference on Decision and Control (CDC)*. Under review. Mar. 2021.
5. L. Brivadis and L. Sacchelli. "Approximate observability and back and forth observer of a PDE model of crystallisation process". Submitted to *2021 60th IEEE Conference on Decision and Control (CDC)*. Under review. Mar. 2021.

Contents

Notations

General notations

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$	non-negative integers, integers, rational numbers, real numbers, complex numbers
$\mathbb{K}_+$, $\mathbb{K}^*$	non-negative/non-zero numbers in $\mathbb{K}$
$\lvert x\rvert$	Euclidean norm of x
$\lVert A\rVert$	induced matrix norm of A
A'	transpose of the matrix A
$\mathbb{S}^{n-1}$	unit sphere of $\mathbb{R}^n$
$\partial\Omega$	boundary of the set Ω
$\bar{\Omega}$	closure of the set Ω
$\mathcal{O}$, o	big/little-o notation
$\Re z$, $\Im z$	real and imaginary part of the complex number z
$\dot{x}$	derivative of $t \mapsto x(t)$
f'	derivative of f
$f^{(k)}$	k-derivative of f
$\frac{\partial f}{\partial x}$	partial derivative of f with respect to x
$Df(x)[v]$	differential at $x \in \mathbb{R}^n$ applied to the vector $v \in \mathbb{R}^n$ of the function f
$L_f V$	Lie derivative of V with respect to f
$f * g$	convolution product of f with g

Linear algebra

$\langle \cdot, \cdot \rangle_X$	scalar product on X
$\lVert \cdot \rVert_X$	norm on X
$B_X(x_0, r) = \{x \in X \ : \ \lVert x - x_0 \rVert_X < r\}$	open ball centered at x_0 of radius r in X

$\overset{w}{\searrow}$	weak convergence
$\mathscr{L}(X,Y)$	bounded linear operators from X to Y
$\mathscr{L}(X) = \mathscr{L}(X,X)$	bounded linear endomorphisms on X
X^*	dual space of X
Id_X	identity operator on X
$\ker A$, $\mathrm{Im}\, A$	kernel and image of the operator A
$\rho(A)$	resolvent set of the operator A
$\sigma(A)$	spectrum of the operator A
A^*	adjoint operator of A
Π_E	orthogonal projection onto E
$E^\perp$	orthogonal of E
$E \oplus F$	direct sum of E and F

Function spaces

$C^k(\Omega;X)$	space of k-times continuously differentiable X-valued functions on Ω, $k \geqslant 0$

$C^\infty(\Omega;X) = \bigcap_{k\in\mathbb{N}} C^k(\Omega;X)$

$L^p(\Omega;X) = \{f : \Omega \to X \text{ measurable } : \ \|f\|_X^p \text{ is integrable over } \Omega\}$, $1 \leqslant p < \infty$

$L^\infty(\Omega;X) = \{f : \Omega \to X \text{ measurable } : \ \|f\|_X \leqslant C \text{ a.e. in } \Omega \text{ for some constant } C\}$

$H^1(\Omega;X) = \{f \in L^2(\Omega;X) \ : \ f' \in L^2(\Omega;X)\}$

$H^m(\Omega;X) = \{f \in H^{m-1}(\Omega;X) \ : \ f' \in H^{m-1}(\Omega;X)\}$, $2 \leqslant m < \infty$

Acronyms

BFN	Back and Forth Nudging
CLD	Chord Length Distribution
FBRM	Focused Beam Reflectance Measurement
KKL	Kazantzis-Kravaris/Luenberger
PATs	Process Analytical Technologies
PSD	Particle Size Distribution
SISO	Single-Input Single-Output

Part I

Output feedback stabilization

Chapter 1

Problem statement

A beginning is the time for taking the most delicate care that the balances are correct.

F. Herbert, *Dune*

Abstract. *This chapter serves as an introduction to the problem of dynamic output feedback stabilization. Essentials concepts are defined, and some necessary conditions are stated. Usual separation principles for uniformly observable systems are recalled, and some existing strategies for non-uniformly observable ones are introduced.*

Contents

Introduction

Stabilizing the state of a dynamical system to a target point is a classical problem in control theory. However, in many physical problems, only part of the state, named the output, is known. Hence a state feedback cannot be directly implemented. Only the output and the state of a dynamical system fed by the output can be used to stabilize the state of the original system. This problem, known as *dynamic output feedback stabilization*, has been extensively studied (see, *e.g.,* [GB81, EK92, GK92, KE93, Cor94a, TP94, JG95, TP95, AK99, MPI07, AP09]). When a state stabilizing feedback can be designed, a common strategy to achieve dynamic output feedback stabilization is to build an observer of the system, that is a dynamical system fed by the output that asymptotically learns the actual state, and to apply the state feedback to the estimation obtained by the observer. This strategy is known to be efficient for uniformly observable systems since [TP94, TP95] and [JG95]. The *observability* of a control system for some fixed input qualifies the ability to estimate the state using its output. It characterizes the fact that two trajectories of the system can be distinguished by their respective output over a given time interval.

This crucial notion constitutes a field of study in itself (see, *e.g.,* [GK01, AP09, Ber+17, Ber19]). A system is uniformly observable if it is observable for all inputs. However, as shown in [GK01], it is not generic for a dynamical system to be uniformly observable when the dimension of its input is greater or equal to the one of its output. There may exist singular inputs for the system, that are inputs that make the system unobservable, and the output feedback may produce such singular inputs. This defeats the purpose of output feedback stabilization, which is still an open problem when such inputs exist. The first part of this thesis is devoted to this issue.

One can distinguish two main contexts, depending on whether or not the value of the feedback law at the target point is a constant input makes the system observable. Chapter 2 is devoted to the first case, while Chapter 4 deals with the second one. Chapter 3 contains an intermediate result bridging them.

1.1 Definition

Let n, m and p be positive integers, $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^m$. For all $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$, consider a general observation-control system:

$$\begin{cases} \dot{x} = f(x, u) \\ y = h(x) \end{cases} \tag{1.1}$$

where x is the state of the system, u is the control (or input) and y is the observation (or output, or measurement).

The first part of the thesis deals with the problem of dynamic output feedback stabilization of (1.1). The goal is to use the online measurement of y to stabilize, by acting on the control u, the state x to some target point $x^\star \in \mathbb{R}^n$. Stabilization to more general target sets in $\mathbb{R}^n$ is beyond the scope of this thesis. Up to a change of coordinates, we assume without loss of generality that $x^\star = 0$ and $h(0) = 0$.

To guarantee the well-posedness of the Cauchy problem associated to the open-loop system (1.1), assume that f is continuous and uniformly locally Lipschitz with respect to x. According to the Cauchy-Lipschitz theorem, for any $u \in C^0(\mathbb{R}_+, \mathbb{R}^n)$ and any $x_0 \in \mathbb{R}^n$, there exists exactly one maximal solution $\varphi_t(x_0, u)$ defined for $t \in [0, T_u(x_0))$ such that $\varphi_0(x_0, u) = x_0$ and $\frac{\partial \varphi_t(x_0,u)}{\partial t} = f(\varphi_t(x_0, u), u(t))$. The map φ is continuous and called the flow of (1.1).

Definition 1.1 (Dynamic output feedback stabilizability)**.** System (1.1) is said to be *locally* (resp. *globally*) *stabilizable by means of a dynamic output feedback* if and only if the following holds.

There exist two continuous maps $\nu : \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^q$ and $\varpi : \mathbb{R}^q \times \mathbb{R}^m \to \mathbb{R}^p$ for some non-negative integer q such that $(0, 0) \in \mathbb{R}^n \times \mathbb{R}^q$ is a locally (resp. globally) asymptotically stable equilibrium point of the following system:

$$\begin{cases} \dot{x} = f(x, u) \\ y = h(x) \end{cases}, \qquad \begin{cases} \dot{w} = \nu(w, u, y) \\ u = \varpi(w, y). \end{cases} \tag{1.2}$$

Additionally, if for any compact set $\mathcal{K}_x \subset \mathbb{R}^n$, there exist two continuous maps $\nu : \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^q$ and $\varpi : \mathbb{R}^q \times \mathbb{R}^m \to \mathbb{R}^p$ for some non-negative integer q, and a compact set $\mathcal{K}_w \subset \mathbb{R}^q$ such that $(0, 0) \in \mathbb{R}^n \times \mathbb{R}^q$ is an asymptotically stable equilibrium point of (1.2) with basin of attraction containing $\mathcal{K}_x \times \mathcal{K}_w$, then (1.1) is said to be *semi-globally stabilizable by means of a dynamic output feedback.*

Remark 1.2. Please pay attention to the fact that the notation $\mathcal{K}_x$ does not mean that the set depends on some variable x, but rather that the variables belonging to the set are usually denoted by x, namely, $x \in \mathcal{K}_x$. Similar notations are used on compact sets throughout the thesis to immediately indicate to the reader what kind of variables will belong to some set at the moment it is defined.

Remark 1.3. Clearly, we have following implications about the stabilizability by means of dynamic output feedback of (1.1):

$$\text{Global} \implies \text{Semi-global} \implies \text{Local}. \tag{1.3}$$

Remark 1.4. In most of engineering applications, one must achieve semi-global stabilization. Local stabilization is not always sufficient, depending on the size of the basin of attraction. But global stabilization is not useful if one knows an order of magnitude of the initial conditions. We particularly focus on semi-global results in this part of the thesis.

Remark 1.5. Clearly, a necessary condition to the local asymptotic stability of (1.2) at $(0,0)$ is that $f(0,u^\star)=0$ where $u^\star=\varpi(0,0)$. Without loss of generality, we assume that if (1.2) is locally asymptotically stable at $(0,0)$, then the value of the control at the target point is zero: $\varpi(0,0)=0\in\mathbb{R}^p$ and $f(0,0)=0$.

Remark 1.6. The uniqueness of solutions of the closed-loop system (1.2) is not guaranteed. Hence, let us recall that a dynamical system is said to be asymptotically stable at an equilibrium point with some basin of attraction if and only if each initial condition in the basin of attraction yields at least one solution to the corresponding Cauchy problem, each solution converges to the equilibrium point, and the equilibrium point is Lyapunov stable.

Several generalizations of Definition 1.1 can be considered. The dynamical system fed by the output may be time-varying, that is,

$$\begin{cases}\dot{x}=f(x,u)\\ y=h(x)\end{cases}, \qquad \begin{cases}\dot{w}=\nu(w,u,y,t)\\ u=\varpi(w,y,t).\end{cases} \tag{1.4}$$

In particular, the *periodic time-varying* output feedback stabilizability defined in [Cor94a] falls within this framework (see Definition 1.24). In this thesis, we rather focus on *autonomous* dynamic output feedback stabilization. The problem of semi-global autonomous dynamic output feedback stabilization is known to be of deep interest in control theory as in applications, and still very little is known for generic maps f and h.

In Chapter 4, we will allow the additional variable w to lie in an infinite-dimensional space, which requires to extend Definition 1.1 (in this chapter only). Our strategy and functional framework is very different from the infinite-dimensional controller introduced in [MP93]. Somehow counter-intuitive, this proposition will be justified in the corresponding chapter.

Let us now introduce some necessary conditions to the dynamic output feedback stabilizability of a system.

1.2 Necessary conditions

The problem of dynamic state feedback stabilization of (1.1) is equivalent to the dynamic output feedback stabilization in the case where $h(x) = x$. Therefore, dynamic state feedback stabilizability of (1.1) is a necessary condition for dynamic output feedback stabilizability. One may wonder if *static* state feedback stabilizability of (1.1) is a necessary condition for dynamic output feedback stabilizability. In [AP09], the authors answered by the positive if a sufficiently regular selection function can be found. We recall their result below.

Definition 1.7 (State feedback stabilizability)**.** System (1.1) is said to be *locally* (resp. *globally*) *stabilizable by means of a (static) state feedback* if and only if there exists a continuous map $\phi : \mathbb{R}^n \to \mathbb{R}^p$ such that $0 \in \mathbb{R}^n$ is a locally (resp. globally) asymptotically stable equilibrium point of

$$\begin{cases} \dot{x} = f(x, u) \\ u = \phi(x). \end{cases} \tag{1.5}$$

Additionally, if for any compact set $\mathcal{K}_x \subset \mathbb{R}^n$, there exists a continuous map $\phi : \mathbb{R}^n \to \mathbb{R}^p$ such that $0 \in \mathbb{R}^n$ is an asymptotically stable equilibrium point of (1.5) with basin of attraction containing $\mathcal{K}_x$, then (1.1) is said to be *semi-globally stabilizable by means of a static state feedback*.

Remark 1.8. Clearly, the implications (1.3) hold for the stabilizability by means of static state feedback of (1.1).

Theorem 1.9 ([AP09, Lemma 1, (1)])**.** *Assume that* (1.2) *is locally asymptotically stable at* $(0, 0)$ *with basin of attraction*[1] $\mathcal{U}_x \times \mathcal{U}_w$*. Let* V *be a* $C^\infty(\mathcal{U}_x \times \mathcal{U}_w, \mathbb{R}_+)$ *strict proper Lyapunov function of* (1.2)*. If there exists a selection map* $\mathcal{U}_x \ni x \mapsto \phi(x) \in \operatorname{argmin}_{\mathcal{U}_w} V(x, \cdot)$ *which is locally Hölder of order strictly larger than* $\frac{1}{2}$*, then* (1.5) *is locally asymptotically stable at* 0 *with basin of attraction containing* $\mathcal{U}_x$*.*

Therefore, up to the existence of a sufficiently regular selection map, this result implies that the following local (resp. semi-global, global) condition is necessary for the local (resp. semi-global, global) stabilizability of (1.1) by means of a dynamic output feedback.

Condition 1.10 (State feedback stabilizability — local, semi-global, global)**.** System (1.1) is locally (resp. semi-globally, globally) stabilizable by means of a static state feedback.

In [Cor94a], J.-M. Coron stated two additional conditions that he proved to be sufficient when local static state feedback stabilizability holds to ensure local dynamic output feedback stabilizability, provided that one allows the output feedback to depend on time (which we do *not* allow in this thesis). The two following conditions are weaker versions of the ones of [Cor94a]. We prove that these two conditions are necessary to ensure dynamic output feedback stabilizability. The first one, known as 0-*detectability* is also used by E. Sontag in [Son81] in the context of abstract nonlinear regulation theory.

[1] In [AP09, Lemma 1, (1)], the authors state only a global version of the result, that is, $\mathcal{U}_x = \mathbb{R}^n$ and $\mathcal{U}_w = \mathbb{R}^q$. However, the proof remains identical in the other cases.

Condition 1.11 (0-detectability — local, global)**.** Let $\mathcal{X}_0 = \{x_0 \in \mathbb{R}^n \ : \ \forall t \in [0, T_0(x_0)), h(\varphi_t(x_0, 0)) = 0\}$. Then $0 \in \mathcal{X}_0$ is a locally (resp. globally) asymptotically stable equilibrium point of the vector field $\mathcal{X}_0 \ni x \mapsto f(x, 0)$.

Theorem 1.12. *If* (1.1) *is locally (resp. semi-globally, globally) stabilizable by means of a dynamic output feedback, then Condition 1.11 holds locally (resp. globally, globally).*

Proof. The set $\mathcal{X}_0$ is forward invariant for the vector field $x \mapsto f(x, 0)$ and $0 \in \mathcal{X}_0$. Let $x_0 \in \mathcal{X}_0$. Assume that (1.1) is locally stabilizable by means of a dynamic output feedback, and that $(x_0, 0)$ is in the basin of attraction of $(0, 0)$ for (1.2).

Then $t \mapsto (\varphi_t(x_0, 0), 0)$ is a trajectory of (1.2) with initial condition $(x_0, 0)$. Hence $\varphi_t(x_0, 0)$ is well-defined for all $t \geqslant 0$ and tends towards 0 as t goes to infinity. Moreover, for all $R > 0$, there exists $r > 0$ such that, if $x_0 \in B_{\mathbb{R}^n}(0, r)$, then $\varphi_t(x_0, 0) \in B_{\mathbb{R}^n}(0, R)$ for all $t \geqslant 0$.

If we assume that (1.1) is globally stabilizable by means of a dynamic output feedback, then the arguments still hold for any $x_0 \in \mathbb{R}^n$. If (1.1) is only semi-globally stabilizable by means of a dynamic output feedback, we first define $\mathcal{K}_x$ as in Definition 1.1 containing x_0. ■

Condition 1.13 (Indistinguishability $\implies$ common stabilizability — local, global)**.** For all x_0, $\tilde{x}_0$ in some neighborhood of $0 \in \mathbb{R}^n$ (resp. for all x_0, $\tilde{x}_0$ in $\mathbb{R}^n$), if for all $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ such that $T_u(x_0) = +\infty$ it holds that $h(\varphi_t(x_0, u)) = h(\varphi_t(\tilde{x}_0, u))$ for all $t \in [0, T_u(\tilde{x}_0))$, then there exists $v \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ such that $\varphi_t(x_0, v)$ and $\varphi_t(\tilde{x}_0, v)$ are well-defined for all $t \in \mathbb{R}_+$ and tend towards 0 as t goes to infinity.

Theorem 1.14. *If* (1.1) *is locally (resp. semi-globally, globally) stabilizable by means of a dynamic output feedback, then Condition 1.13 holds locally (resp. globally, globally).*

Proof. Let $x_0, \tilde{x}_0 \in \mathbb{R}^n$ be such that for all $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ such that $T_u(x_0) = +\infty$ it holds that $h(\varphi_t(x_0, u)) = h(\varphi_t(\tilde{x}_0, u))$ for all $t \in [0, T_u(\tilde{x}_0))$. Assume that (1.1) is locally stabilizable by means of a dynamic output feedback, and that $(x_0, 0)$, $(\tilde{x}_0, 0)$ are in the basin of attraction of $(0, 0)$ for (1.2).

Let (x, w) be a solution of (1.2) starting from $(x_0, 0)$. Set $v = \varpi(w, h(x))$. Then $T_v(x_0) = +\infty$ and $\varphi_t(x_0, v) \to 0$ as $t \to +\infty$. Let $\tilde{x}(t) = \varphi_t(\tilde{x}_0, v)$ for all $t \in [0, T_v(\tilde{x}_0))$. Since $h(\varphi_t(x_0, v)) = h(\varphi_t(\tilde{x}_0, v))$ for all $t \in [0, T_v(\tilde{x}_0))$, $(\tilde{x}, w)$ is a solution of (1.2) starting from $(\tilde{x}_0, 0)$. Hence $T_v(\tilde{x}_0) = +\infty$ and $\varphi_t(\tilde{x}_0, v) \to 0$ as $t \to +\infty$.

If we assume that (1.1) is globally stabilizable by means of a dynamic output feedback, then the arguments still hold for any $x_0, \tilde{x}_0 \in \mathbb{R}^n$. If (1.1) is only semi-globally stabilizable by means of a dynamic output feedback, we first define $\mathcal{K}_x$ as in Definition 1.1 containing x_0 and $\tilde{x}_0$. ■

Conditions 1.10, 1.11 and 1.13 are known to be insufficient for the dynamic output feedback stabilizability of a system. Indeed, consider

$$\dot{x} = u, \qquad y = x^2. \tag{1.6}$$

Clearly, the system is state feedback stabilizable (with $u = -x$), 0-detectable ($x^2 = 0 \implies x = 0$), and has no indistinguishable initial conditions (they are

all distinguished by $u = 1$). Hence, Conditions 1.10, 1.11 and 1.13 are satisfied, but the system is not locally stabilizable by means of a dynamic output feedback (see [Cor94a] and Chapter 4, Corollary 4.2).

Many techniques have been developed to achieve the output feedback stabilization, leading to different sufficient conditions. In this thesis, we mainly focus on separation principles.

1.3 Separation principle, uniform observability

In the survey [AP09], the authors proposed to classify the output feedback designs in two categories, each of them requiring the knowledge of a stabilizing state feedback law ϕ:

- the direct approach, in which the goal is to directly estimate a stabilizing control $u = \phi(x)$ by using the output y, without necessarily estimating the full state x;
- the indirect approach, in which the goal is to estimate the full state x of the system by using the output y, and then to apply ϕ to this estimation.

The direct approach requires robustness properties of the system to perturbations of the input. Although it has lead to important results for specific classes of systems (see, *e.g.,* [AK01, PP04, And05, PQ05, AP08]), this route has hardly been followed, and the indirect approach is more common.

In this thesis, we focus on the indirect approach. Our goal is to build an observer $\hat{x}$ of the state x, based on the measurement y, and to apply the control $u = \phi(\hat{x})$. This technique is also known as observer-based control. For linear systems, it is equivalent to the separation principle, which consists of designing "separately" a stabilizing state feedback law and a state observer. Then the coupled system provides a suitable dynamic output feedback. For nonlinear systems, the existence of both a stabilizing state feedback law and a state observer is in general not sufficient to guarantee the asymptotic stability of the closed-loop system, unless additional assumptions are made. Even in this case, the observer may not always be designed "separately" from the state feedback: most of the time, parameters of the observer system depend on ϕ. For this reason, this approach is also called observer-based output feedback design instead of separation principle for nonlinear systems.

In order to design nonlinear separation principles, most authors rely on observability hypotheses on the system. Let us introduce some of the hypotheses considered in the literature.

Definition 1.15 (Observability)**.** System (1.1) is said to be *observable* in time T for an input $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ if and only if, for all initial conditions $x_0 \neq \tilde{x}_0 \in \mathbb{R}^n$, the set

$$\Big\{ t \in [0, \min(T, T_u(x_0), T_u(\tilde{x}_0))) \ : \ h(\varphi_t(x_0, u)) \neq h(\varphi_t(\tilde{x}_0, u)) \Big\} \tag{1.7}$$

has positive measure. If system (1.1) is observable in any time $T > 0$ for all inputs u, then it is said to be *uniformly observable* in small time.

A stronger notion is the complete uniform observability defined in [TP94]

Definition 1.16 (Complete uniform observability)**.** Assume that f and h are sufficiently smooth. System (1.1) is said to be *completely uniformly observable* if and only if there exist two non-negative integers n_y and n_u and a smooth function $\eta : \mathbb{R}^{m(n_y+1)} \times \mathbb{R}^{p(n_u+1)} \to \mathbb{R}^n$ such that, for all smooth inputs $u : \mathbb{R}_+ \to \mathbb{R}^p$ and all solutions (x, y) of (1.1), we have for all $t \in [0, T_u(x(0)))$,

$$x(t) = \eta((y^{(i)}(t))_{0\leqslant i\leqslant n_y}, (u^{(i)}(t))_{0\leqslant i\leqslant n_u}) \tag{1.8}$$

where $(y^{(i)}(t))_{0\leqslant i\leqslant n_y}$ and $(u^{(i)}(t))_{0\leqslant i\leqslant n_u}$ denote the n_y and n_u first derivatives at time t of y and u, respectively.

A related notion is the strong differential observability of [GK01].

Definition 1.17 (Strong differential observability)**.** Assume that f and h are sufficiently smooth. System (1.1) is said to be *strongly differentially observable* of order N if and only if there exists a non-negative integer N such that the mapping[2]

$$\begin{aligned} \mathbb{R}^n \times \mathbb{R}^{pN} &\longrightarrow \mathbb{R}^{mN} \times \mathbb{R}^{pN} \\ (x, (u^{(i)})_{0\leqslant i\leqslant N}) &\longmapsto ((y^{(i)})_{0\leqslant i\leqslant N}, (u^{(i)})_{0\leqslant i\leqslant N}), \end{aligned}$$

where $(y^{(i)})_{0\leqslant i\leqslant N}$ denotes the N first derivatives at $t = 0$ of the output y corresponding to the initial condition x and an input whose N first derivatives at $t = 0$ are given by $(u^{(i)})_{0\leqslant i\leqslant N}$, is an injective immersion.

For such systems, a separation principle can be achieved according to [TP94, TP95] and [JG95].

Theorem 1.18 ([TP94])**.** *If system* (1.1) *is*

- *semi-globally stabilizable by means of a smooth state feedback,*
- *completely uniformly observable,*

then it is semi-globally stabilizable by means of a dynamic output feedback.

Theorem 1.19 ([JG95])**.** *If system* (1.1) *is*

- *semi-globally stabilizable by means of a smooth state feedback,*
- *strongly differentially observable,*

then it is semi-globally stabilizable by means of a dynamic output feedback.

According to [GK01], when $m > p$, *i.e.,* when there are more outputs than inputs, it is generic for nonlinear systems to be strongly differentially observable.

Theorem 1.20 ([GK01, Chapter 4, Theorem 2.2])**.** *Assume that $m > p$. Let $\Sigma = \{(f, h) \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}^n) \times C^\infty(\mathbb{R}^n, \mathbb{R}^p)\}$ be endowed with the Whitney C^∞ topology[3]. Then the set of pairs (f, h) such that* (1.1) *is strongly differentially observable of order $N \geqslant 2n + 1$ contains a residual subset of Σ, i.e., a countable intersection of dense open sets.*

[2]This mapping is well-defined since (1.1) is such that the N first derivatives of y at $t = 0$ depend only on x and the N first derivative of u at $t = 0$. The reader may refer to [GK01] for more properties of this mapping.

[3]See Definition 2.26.

On the contrary, this genericity property does not hold when $m \leqslant p$. In particular, in the case of Single-Input Single-Output (SISO) bilinear systems, uniformly observable systems actually have a normal form, hence non-uniformly observable ones are generic (see Theorems 2.6 and 2.8 for a more precise statement). Therefore, the question of dynamic output feedback stabilization for non-uniformly observable systems remains an open and important question, that the first part of this thesis is dealing with.

1.4 Non-uniformly observable systems

Non-uniformly observable systems can be split in two classes, depending on whether or not their target corresponds to an observable input.

Definition 1.21 (Observability at the target)**.** System (1.1) is *observable at the target* in some time $T > 0$ if it is observable in time T for the constant input $u \equiv 0$. Otherwise, (1.1) is *unobservable at the target* in time T.

Note that the input $u \equiv 0$ is precisely the value of the input at the target point of the closed-loop system (1.2) since $\varpi(0, 0) = 0$ (see Remark 1.5). Each case leads to different difficulties in the design of a separation principle.

If the target is observable, and if the state tends towards the target, then the input of the closed-loop system will eventually tend towards an observable one. Hence, observability issues occur only during the transient response. The main difficulty is that the input of the closed-loop system may be one of the unobservable ones. In that case, the observer system will not be able to estimate the state, and the stabilization strategy will fail.

If the target is unobservable, then the observability singularity is somehow unavoidable. Indeed, if stabilization is achieved, then the input tends to render the system less and less observable as the state tends towards the target. Therefore, proving the observer convergence as the state approaches the target is challenging.

In the existing literature, less attention has been paid to non-uniformly observable systems than to uniformly observable ones, for which efficient tried-and-tested methods exist. However, observability singularities occur in various practical engineering systems (see [HPR14, Com+16, Fla19, Sur+19, Aja+20, RD20, Sur+20, AGS21]), leading to a renewal of interest in the issue in recent years. In the following, we recall some existing results of output feedback stabilization dealing with observability singularities.

A popular technique, in particular for systems with unobservable target, is to modify the input (*i.e.,* not to apply directly the state feedback to the observer) in order to gain new observability properties. This way have been paved by the seminal paper [Cor94a], in which local dynamic time-varying periodic output feedback stabilization is achieved up to a Lie null-observability condition. The precise result is the following.

Assume that f and h are smooth. For any multi-index $\alpha \in \mathbb{N}^p$ and constant input $u \in \mathbb{R}^p$, let $f_u^\alpha \in C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ be defined for all $x \in \mathbb{R}^n$ by

$$f_u^\alpha(x) = \frac{\partial^{|\alpha|} f}{\partial u^\alpha}(x, u) \tag{1.9}$$

Denote by $\mathcal{O}$ the linear subspace of $C^\infty(\mathbb{R}^n \times \mathbb{R}^p; \mathbb{R}^m)$ spanned by the maps ω such that

$$\omega(x,u) = L_{f_u^{\alpha_r}} \dots L_{f_u^{\alpha_r}} h(x) \tag{1.10}$$

for some multi-indices $\alpha_1, \dots, \alpha_r$. For $k \in \mathbb{N}$, denote by $L^k_{f_u^\alpha}$ the iterated Lie derivative with respect to the vector field f_u^α.

Definition 1.22 (Lie null-observability)**.** Assume that f and h are smooth. System (1.1) is locally Lie null-observable if there exists $\varepsilon > 0$ such that:

(i) For all $x_0 \in B_{\mathbb{R}^n}(0,\varepsilon) \setminus \{0\}$, there exists $N \in \mathbb{N}$ such that $L^N_{f_0} h(x_0) \neq 0$;

(ii) For all $x_0, \tilde{x}_0 \in B_{\mathbb{R}^n}(0,\varepsilon) \setminus \{0\}$, and all $u \in B_{\mathbb{R}^p}(0,\varepsilon)$, if $x_0 \neq \tilde{x}_0$, then there exists $\omega \in \mathcal{O}$ such that $\omega(x_0,u) \neq \omega(\tilde{x}_0,u)$.

Note that complete uniform observability implies Lie null-observability, but the converse is not true. In particular, system (1.6) is Lie null-observable.

In [Cor94a], both the state and output feedback laws may be time-varying, which requires the following definitions.

Definition 1.23 (Small time reachability)**.** The origin of (1.1) is *locally continuously reachable in small time* if for all $T > 0$ there exist $\varepsilon > 0$ and a map $\phi \in C^0(B_{\mathbb{R}^n}(0,\varepsilon); L^1((0,T);\mathbb{R}^p))$ such that $\sup_{t\in(0,T)} |\phi(x)(t)| \underset{x\to 0}{\longrightarrow} 0$ and for all $x_0 \in B_{\mathbb{R}^n}(0,\varepsilon)$, any corresponding solution of $\dot{x} = f(x,\phi(x)(t))$ is such that $x(T) = 0$.

Definition 1.24 (Dynamic periodic time-varying output feedback stabilizability)**.** System (1.1) is *locally stabilizable in small time by means of a dynamic continuous periodic time-varying output feedback law* if, for all $T > 0$, there exist $\varepsilon > 0$ and two continuous maps $\nu : \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}^q$ and $\varpi : \mathbb{R}^q \times \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}^p$ for some non-negative integer q such that

(i) $\nu(0,0,0,t) = \varpi(0,0,t) = 0$ for all $t \in \mathbb{R}_+$;

(ii) $\nu(w,u,y,t+T) = \nu(w,u,y,t)$ and $\varpi(w,y,t+T) = \varpi(w,y,t)$ for all $t \in \mathbb{R}_+$;

(iii) Any solution of (1.4) such that $(x(s),w(s)) = (0,0)$ for some $s \in \mathbb{R}_+$ is such that $(x(t),w(t)) = (0,0)$ for all $t \geqslant s$;

(iv) Any solution of (1.4) such that $|(x(s),w(s))| \leqslant \varepsilon$ for some $s \in \mathbb{R}_+$ is such that $(x(t),w(t)) = (0,0)$ for all $t \geqslant s + T$.

Then we have the following theorem.

Theorem 1.25 ([Cor94a, Theorem 2.4])**.** *If system* (1.1) *is*

- *locally continuously reachable in small time,*
- *Lie null-observable,*

then it is locally stabilizable in small time by means of a dynamic continuous periodic time-varying output feedback law.

The proof of the result relies on a two-steps strategy:

(Observation phase) On the time interval $[0, T]$, the system is "excited" with a time-varying input making the system observable, but vanishing once the state is stabilized. Thanks to this input, an observer estimating the state in finite-time is designed.

(Stabilization phase) On the time interval $[T, 2T]$, the state is stabilized in finite time thanks to the exact estimation provided by the observer at time T and a small time stabilizing state feedback.

This method has also been used in [Son81] in the context of output regulation, in [NS98] for systems with positive outputs. In [ST03], the authors follow a similar strategy, except that the input making the system observable (during the observation phase) does not vanishes when the state tends towards the target. Hence, the target point is not an equilibrium point, but a *practical* stabilization is still obtained, *i.e.,* the state is asymptotically stabilized to an arbitrary small neighborhood of the target when iterating the observation and stabilization phases. Using Lyapunov arguments to contain the error between the observer and the actual state of the system during the stabilization phase, a semi-global result is obtained rather than a local one.

More recently, some authors have proposed to accomplish both observation and stabilization with the same input, chosen as a slight modification of the control obtained by applying the stabilizing state feedback to the observer. The signal must be sufficiently close to the original one to achieve stabilization, but the modulation allows to obtain new observability properties to guarantee observer convergence. The procedure based on "virtual measurements", developed in [Com+16, Sur+19, Sur+20] for various examples, is based on this paradigm. In this approach, a small-amplitude high-frequency periodic signal is superimposed on the input in order to access, with averaging techniques, to a new "virtual" output. This new output can be used to design the dynamic output feedback.

On a different line of reflection, the authors of [LSG17] proposed to consider additive perturbations of the feedback law instead of the input. One of the main feature of this strategy, compared with all those mentioned above, is that the dynamic output feedback remains autonomous. On an example of a bilinear system borrowed from quantum control, the authors build an explicit perturbation of the stabilizing state feedback law to obtain an "almost" global stabilization result. More precisely, their statement is the following.

Consider the observation-control the system

$$\begin{cases} \dot{x} = A(u)x \\ y = Cx \end{cases}, \quad A(u) = \begin{pmatrix} 0 & 1 & u_1 \\ -1 & 0 & u_2 \\ -u_1 & -u_2 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \tag{1.11}$$

where x in the unit sphere $\mathbb{S}^2$ is the state, u in $\mathbb{R}^2$ is the control and y in $\mathbb{R}$ is the output. Since $A(u)$ is skew-symmetric for all $u \in \mathbb{R}^2$, trajectories of (1.11) starting from the unit sphere $\mathbb{S}^2$ remain on it. The goal is to stabilize (1.11) at the target point $x_{\mathrm{t}} = (0, 0, -1)$. Clearly, the system is unobservable for the control $u = 0$, which is the value of the control at the target point. Moreover, $\phi(x) = (x_1, x_2)$ is a stabilizing[4] state feedback with basin of attraction $\mathbb{S}^2 \setminus \{-x_{\mathrm{t}}\}$. Then, by choosing a

[4]This can be checked by applying LaSalle's invariance principle on the candidate Lyapunov function $V(x) = x_3$.

feedback perturbation of the form $\delta(\hat{x}_3^2 - 1)$, we obtain the following result.

Theorem 1.26 ([LSG17])**.** *There exists $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$, the system*

$$\begin{cases} \dot{x} = A(u)x \\ y = Cx \end{cases}, \qquad \begin{cases} \dot{\hat{x}} = A(u)\hat{x} - C'(C\hat{x} - y) \\ u = \phi(\hat{x}) + \delta(\hat{x}_3^2 - 1) \end{cases} \tag{1.12}$$

is locally asymptotically stable at $(x_\mathrm{t}, x_\mathrm{t})$ with a basin of attraction that is open, dense, and of full measure in $\mathbb{S}^2 \times \mathbb{R}^3$.

Remark 1.27. An important feature of the additive perturbation $\delta(\hat{x}_3^2 - 1)$ is that it vanishes at the target point x_t. Moreover, near the target, it is negligible compared to the feedback law, which guarantees that local asymptotic stability is still achieved.

Inspired by [LSG17], this feedback perturbation strategy is one of the key tools used in this part of the thesis to stabilize non-uniformly systems at an observable (Chapter 2) or unobservable (Chapter 4) target point.

Chapter 2

Observable target

Not all those who wander are lost;

J. R. R. Tolkien, *The Fellowship of the Ring*

Abstract. *We address the problem of dynamic output feedback stabilization at an observable target point of SISO bilinear systems. The challenge lies in the existence of inputs making the system unobservable. During the stabilization strategy, the control generated by the closed-loop system may be one of these inputs, hence the observer convergence will not be guaranteed. To tackle this phenomenon, we propose an autonomous perturbation strategy of the feedback law. The perturbed feedback is still stabilizing when applied to the state, but prevents the closed-loop input to render the system unobservable. We prove, under genericity assumptions on the system, the existence of a dense open set of such perturbations. We apply the results on both the Luenberger and Kalman observers. We discuss how this strategy may pave the way to a generic separation principle for SISO bilinear systems.*

Contents

Introduction

In this chapter, we restrict ourselves to the class of Single-Input Single-Output (SISO) bilinear systems with linear observation that are state feedback stabilizable at some target point, which, with no loss of generality, is chosen to be 0. We also assume the system to be observable at the target, that is, the constant input obtained by evaluation of the feedback at 0 is not singular. This class of systems is a natural choice of study for two reasons. Firstly, the uniform observability hypothesis is still not generic in this case when the dimension of the input is greater or equal to the one of the output. Secondly, according to [FK83], any control-affine system with finite-dimensional observation space may be immersed in such a system.

The existence of inputs making the system unobservable renders the problem of dynamic output feedback stabilization difficult, and no general strategy exists, even if the target is observable. The main obstacle is that the input generated by the closed-loop system may be one of these singular inputs. In [LSG17], the authors propose to introduce a perturbation of the feedback law to avoid this phenomenon. Guided by this previous work, a question to ask is: "Can we ensure that only observable inputs are produced by the dynamics when the output feedback is obtained as a combination of an observer and a stabilizing state feedback?" This question falls within the more general and unsolved problem of building a smooth separation principle for systems with observability singularities. One cannot hope for generic bilinear systems that all stabilizing state feedback laws ensure the observability of the closed-loop system. However, we show that for any stabilizing state feedback law, there exist small additive perturbations to this feedback that satisfy this observability property and conserve its locally stabilizing property. Transversality theory is used to prove the existence of such an open and dense class of perturbations. In particular, for almost all considered systems, almost any locally stabilizing feedback law ensures observability of the closed-loop system. Actually achieving output feedback stabilization is beyond the scope of this work, which focuses only on the observability issue. Yet, the obtained results may pave the way to the construction of a generic separation principle. For the results to hold, some properties of the dynamical observer are needed. The problem is tackled with a general observer design, and it is shown in a closing section that the classical Luenberger and Kalman observers fit the considered hypotheses.

2.1 Problem statement

2.1.1 SISO Bilinear systems

We restrict our analysis to the case of SISO bilinear systems. Let n be a positive integer, $A, B \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{1\times n}$, $b \in \mathbb{R}^n$ and $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$. Set $A_u = A + uB$ and consider the following observation-control bilinear system:

$$\begin{cases} \dot{x} = A_u x + bu \\ y = Cx. \end{cases} \tag{2.1}$$

Bilinear systems are used to model various physical phenomena (see [MK80] for a review on the subject). In particular, the Ćuk converter and the heat exchanger

investigated in Examples 3.3.1 and 3.3.2 are bilinear. Moreover, control-affine systems with finite-dimensional observation space can be immersed into bilinear ones, as we recall in the following.

Definition 2.1 (Control-affine systems)**.** A C^∞ (resp. analytic) *control-affine* system is a system of the form

$$\begin{cases} \dot{x} = f(x) + \sum_{i=1}^{p} u_i g_i(x) \\ y = h(x) \end{cases} \tag{2.2}$$

where f and $(g_i)_{1\leqslant i\leqslant p}$ are C^∞ (resp. analytic) vector fields, h is a C^∞ (resp. analytic) map, $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^p$ is the control and $y \in \mathbb{R}^m$ is the observation.

Definition 2.2 (Observation space)**.** The observation space of a C^∞ (resp. analytic) control-affine systems is the smallest vector subspace of $C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ (resp. $C^\omega(\mathbb{R}^n, \mathbb{R}^n)$ containing the observation function h and closed under the Lie derivation along elements of $\mathcal{F} := \{f + \sum_{i=1}^{p} u_i g_i, u \in \mathbb{R}^p\}$.

Definition 2.3 (Immersion)**.** Let

$$\begin{cases} \dot{x}_1 = f_1(x_1, u) \\ y_1 = h_1(x_1) \end{cases} \tag{2.3}$$

and

$$\begin{cases} \dot{x}_2 = f_2(x_2, u) \\ y_2 = h_2(x_2) \end{cases} \tag{2.4}$$

be two C^∞ (resp. analytic) control systems where f_i and h_i are C^∞ (resp. analytic) maps, $x_i \in \mathbb{R}^{n_i}$ are the states, $u \in \mathbb{R}^p$ is the control and $y_i \in \mathbb{R}^m$ are the observations for $i \in \{1, 2\}$. Denote by $\varphi_{1,t}(x_1, u)$ (resp. $\varphi_{2,t}(x_2, u)$) the flow of (2.3) (resp. (2.4)) defined for $t \in [0, T_{1,u}(x_1))$ (resp. $t \in [0, T_{2,u}(x_2))$). We shall say that (2.3) can be *immersed* into (2.4) if there exists a C^∞ (resp. analytic) map $\tau : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ (called the *immersion*) such that:

(i) $\forall x_1, \tilde{x}_1 \in \mathbb{R}^n$, $h_1(x_1) \neq h_1(\tilde{x}_1) \implies h_2(\tau(x_1)) \neq h_2(\tau(\tilde{x}_1))$;

(ii) $\forall u \in C^0(\mathbb{R}_+, \mathbb{R}^p), \forall x_1 \in \mathbb{R}^{n_1}, \forall t \in [0, \min(T_{1,u}(x_1), T_{2,u}(\tau(x_1)))),$
$h_1(\varphi_{1,t}(x_1, u)) = h_2(\varphi_{2,t}(\tau(x_1), u))$.

Remark 2.4. This definition of immersion does not coincide with the usual notion of immersion in differential topology: the differential of τ is not supposed to be everywhere injective.

Theorem 2.5 ([FK83, Theorem 1])**.** *A C^∞ (resp. analytic) control-affine system can be immersed into a bilinear one if and only if its observation space is finite-dimensional.*

Let $u \in \mathbb{R}$ be some constant input. Then (2.1) is a linear system. Hence, if (2.1) is observable in some time $T > 0$, then it is also observable in any time $T > 0$, and

we say that the pair (C, A_u) is observable. According to the Kalman rank condition, (C, A_u) is observable if and only if the rank of the following observability matrix

$$\mathcal{O}(C, A_u) = \begin{pmatrix} C \\ CA_u \\ \vdots \\ CA_u^{n-1} \end{pmatrix} \tag{2.5}$$

is equal to n.

An important property of SISO bilinear systems is that they preserve the genericity of observability singularities. Hence the output feedback stabilization problem remains challenging. Indeed, uniformly observable SISO bilinear systems are characterized by the following normal form.

Theorem 2.6 ([GK92, Theorem 2])**.** *System* (2.1) *is observable for any bounded input u if and only if there exists an invertible matrix $T \in \mathbb{R}^{n\times n}$ such that x is a solution of* (2.1) *if and only if $\tilde{x} := Tx$ is a solution of*

$$\begin{cases} \dot{\tilde{x}} = (\tilde{A} + u\tilde{B})\tilde{x} + \tilde{b}u \\ y = \tilde{C}\tilde{x} \end{cases} \tag{2.6}$$

where

$$\tilde{A} = TAT^{-1} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} & \alpha_n \end{pmatrix} \text{ for some } (\alpha_i)_{1\leqslant i\leqslant n}, \tag{2.7}$$

$$\tilde{B} = TBT^{-1} = \begin{pmatrix} \beta_{1,1} & 0 & \cdots & 0 \\ \vdots & \beta_{2,2} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ \beta_{n,1} & \cdots & \cdots & \beta_{n,n} \end{pmatrix} \text{ for some } (\beta_{i,j})_{1\leqslant j\leqslant i\leqslant n}, \tag{2.8}$$

$$\tilde{C} = CT^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}, \tag{2.9}$$

$$\text{and } \tilde{b} = Tb. \tag{2.10}$$

Remark 2.7. The proof of Theorem 2.6 relies on the following strategy. If the pair (C, A) is observable, then there exists a linear change of coordinates T such that A and C have the above normal form (see, *e.g.,* [Jur96, Chapter 4, Theorem 1] in its dual form[1]). Moreover, T given in the proof of [Jur96, Chapter 4, Theorem 1] depends continuously on (A, C). Then, if B has a non-zero coefficient above its diagonal in these coordinates, it is easy to design a time-varying control u making the system unobservable.

Theorem 2.8. *The set Σ of matrices $(A, B, C) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{1\times n}$ such that* (2.1) *is not uniformly observable is dense with nonempty interior.*

[1]By *dual form*, we refer to the usual duality existing between observation and control problems for linear systems.

Proof. First, we show that Σ is dense. Let $(A, B, C) \notin \Sigma$. Then according to Theorem 2.6, there exists an invertible $T \in \mathbb{R}^{n\times n}$ such that $(TAT^{-1}, TBT^{-1}, CT^{-1})$ is in the form of (2.7)-(2.8)-(2.9). Let $1 \leqslant i < j \leqslant n$. For all $k \in \mathbb{N}^*$, let $B_k = B + \frac{1}{k}T^{-1}E_{i,j}T$ where $E_{i,j}$ is the matrix having coefficient 1 at (i, j) and 0 everywhere else. Then $B_k \to B$ as $k \to +\infty$, and $TB_kT^{-1} = TBT^{-1} + \frac{1}{k}E_{i,j}$. Hence $(A, B_k, C) \in \Sigma$ for all $k \in \mathbb{N}^*$ according to Remark 2.7. Hence Σ is dense in $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{1\times n}$.

Second, we show that Σ contains an open set U. Let $U = \{(A, B, C) \in \Sigma : (C, A) \text{ is observable}\}$. Let $(A, B, C) \in U$. Then according to [Jur96, Chapter 4, Theorem 1], there exists an invertible matrix $T \in \mathbb{R}^{n\times n}$ such that TAT^{-1} and CT^{-1} are in the form of (2.7)-(2.9). But according to Theorem 2.6, since $(A, B, C) \in \Sigma$, TBT^{-1} is not in the form of (2.8), *i.e.,* there exists $\beta_{i,j} \neq 0$, $1 \leqslant i < j \leqslant n$, a non-zero coefficient of TBT^{-1} in its upper-triangular part. Hence, there exists a neighborhood V of (A, B, C) such that the (i, j)-coefficient of $T_vB_vT_v^{-1}$ is non-zero for all $(A_v, B_v, C_v) \in V$, where T_v is the change of coordinates depending continuously (see Remark 2.7 and the proof of [Jur96, Chapter 4, Theorem 1]) on the observable pair (C_v, A_v). Thus U is open. ■

Let $\mathbb{S}^{n-1}$ be the unit sphere of $\mathbb{R}^n$. Due to the bilinear structure (2.1), we also have the following observability characterization.

Proposition 2.9. *System* (2.1) *is observable in time T for some control $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ if and only if for every $\omega_0 \in \mathbb{S}^{n-1}$ the unique solution of $\dot\omega = A_u\omega$ initiated from ω_0 satisfies $C\omega|_{[0,T]} \not\equiv 0$.*

Proof. Let $T > 0$ and $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$. Assume that (2.1) is observable in time T for the input u. Let $\omega_0 \in \mathbb{S}^{n-1}$, and denote by ω the unique solution of the Cauchy problem $\dot\omega = A_u\omega$, $\omega(0) = \omega_0$. Assume for the sake of contradiction that $C\omega(t) = 0$ for all $t \in [0, T]$. Denote by x and $\tilde x$ the unique solutions of (2.1) starting from $x(0) = \omega_0$ and $\tilde x(0) = 0$, respectively. Then $x(0) - \tilde x(0) = \omega_0$ and $\dot x - \dot{\tilde x} = A_u(x - \tilde x)$. By uniqueness of solutions of the Cauchy problem, $\omega(t) = x(t) - \tilde x(t)$ for all $t \in \mathbb{R}_+$. Hence $C(x - \tilde x)|_{[0,T]} \equiv 0$, *i.e.,* $Cx|_{[0,T]} \equiv C\tilde x|_{[0,T]}$. Since (2.1) is observable in time T for the input u, $x(t) = \tilde x(t)$, *i.e.,* $\omega(t) = 0$, for all $t \in [0, T]$. In particular, $\omega_0 = 0$, which contradicts $\omega_0 \in \mathbb{S}^{n-1}$.

Conversely, assume that for every $\omega_0 \in \mathbb{S}^{n-1}$ the unique solution of $\dot\omega = A_u\omega$ initiated from ω_0 satisfies $C\omega|_{[0,T]} \not\equiv 0$. Let $x_0 \neq \tilde x_0 \in \mathbb{R}^n$, and denote by x and $\tilde x$ the corresponding solutions of (2.1). Let $\omega(t) = \frac{x(t)-\tilde x(t)}{|x_0-\tilde x_0|}$. Then $\omega(0) = \frac{x_0-\tilde x_0}{|x_0-\tilde x_0|} \in \mathbb{S}^{n-1}$ and $\dot\omega = A_u\omega$. Hence $C\omega|_{[0,T]} \not\equiv 0$. Since x and $\tilde x$ are continuous, there exists an open interval $\mathcal{I}$ such that $C\omega(t) \neq 0$, *i.e.,* $Cx(t) \neq C\tilde x(t)$, for all $t \in \mathcal{I}$. Thus (2.1) is observable in time T for the input u. ■

Remark 2.10. Roughly speaking, ω in Proposition 2.9 stands for the difference between two solutions of (2.1). With no loss of generality (see the proof of Proposition 2.9 above), we have assumed that $\omega_0 \in \mathbb{S}^{n-1}$. But note that $\omega(t)$ lies in $\mathbb{R}^n$ (not necessarily in $\mathbb{S}^{n-1}$) for $t > 0$.

2.1.2 State feedback

State feedback stabilization is an important issue for SISO bilinear systems and various strategies have been developed (see, *e.g.,* [Qui80, Gut81, BB91] for quadratic feedback laws, [Bac90] and references therein, or [CV00] more recently). In the

context of dynamic output feedback stabilization, we assume the existence of a smooth locally stabilizing state feedback: let $\lambda \in C^\infty(\mathbb{R}^n, \mathbb{R})$ be such that 0 is an asymptotically stable equilibrium point of the vector field $x \mapsto A_{\lambda(x)}x + b\lambda(x)$ for some open domain of attraction $\mathcal{D}(\lambda)$. As stated in Theorem 1.9, up to the existence of a sufficiently regular selection map, it is a necessary condition. We further assume that $\lambda(0) = 0$, which is true up to a substitution of A with $A + \lambda(0)B$.

2.1.3 Observer design

Following the indirect approach described in [AP09], our output feedback stabilization strategy relies on an observer $\hat{x}$ of the state. We fix the observer structure as follows. Let $\mathcal{S}_n \subset \mathbb{R}^{n\times n}$ be the manifold of real positive-definite symmetric matrices and let $L : \mathcal{S}_n \to \mathbb{R}^{n\times 1}$. For all $u \in \mathbb{R}$, let $f(\cdot, u)$ be a vector field over $\mathcal{S}_n$. Denoting $\varepsilon = \hat{x} - x$, we introduce a dynamical observer system depending on the pair (f, L):

$$\begin{cases} \dot{\hat{x}} = A_u\hat{x} + bu - L(\xi)C\varepsilon \\ \dot{\varepsilon} = (A_u - L(\xi)C)\,\varepsilon \\ \dot{\xi} = f(\xi, u). \end{cases} \tag{2.11}$$

This structure matches the usual Luenberger and Kalman observers by setting $f(\xi, u) = 0$ (Luenberger observer) or $f_Q^{\text{Kalman}}(\xi, u) = \xi A_u' + A_u\xi + Q - \xi C'C\xi$ for some $Q \in \mathcal{S}_n$ (Kalman observer) and $L(\xi) = \xi C'$.

2.1.4 Closed-loop system

If (2.1) is not uniformly observable (which is generic), then a natural question to ask, and a first step to achieve output feedback stabilization, is: "Can we ensure that only observable inputs are produced by the dynamics when the output feedback is obtained as a combination of the observer and the stabilizing state feedback?" The stabilizing state feedback λ does not necessarily satisfy this property: there may exist initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0)$ such that the control $u = \lambda \circ \hat{x}$ (where $\hat{x}$ follows (2.11)) make system (2.1) unobservable. Hence, we consider a small perturbation $\lambda + \delta$ of it. For all $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$, we consider the coupled system

$$\begin{cases} \dot{\hat{x}} = A_{(\lambda+\delta)(\hat{x})}\hat{x} + b(\lambda + \delta)(\hat{x}) - L(\xi)C\varepsilon \\ \dot{\varepsilon} = \left(A_{(\lambda+\delta)(\hat{x})} - L(\xi)C\right)\varepsilon \\ \dot{\xi} = f(\xi, (\lambda + \delta)(\hat{x})) \\ \dot{\omega} = A_{(\lambda+\delta)(\hat{x})}\omega \end{cases} \tag{2.12}$$

where $(\hat{x}, \varepsilon, \xi, \omega)$ lies in $\mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n \times \mathbb{R}^n$.

Remark 2.11. In system (2.12), the dynamics of $(\hat{x}, \varepsilon, \xi)$ do not depend on ω. However, the dynamics of ω are included in (2.12) as they are crucial for the observability analysis of (2.1) with input $u = \lambda(\hat{x})$, as stated in Proposition 2.9. We will sometimes consider $(\hat{x}, \varepsilon, \xi)$ to be the first coordinates of a solution of (2.12) without fixing any initial condition for ω.

From now on, we denote by $\mathcal{K} = \mathcal{K}_x \times \mathcal{K}_\varepsilon \times \mathcal{K}_\xi$ a semi-algebraic compact subset of $\mathcal{D}(\lambda) \times \mathbb{R}^n \times \mathcal{S}_n$, which stands for a subset of the space of initial conditions of system (2.11). For all $R > 0$, let

$$\mathcal{V}_R = \{\delta \in C^\infty(\mathbb{R}^n, \mathbb{R}) \ : \ \forall x \in B(0,R), \quad \delta(x) = 0\}.$$

In order to establish our observability results, we make the following important assumptions on the observer given by (f, L):

(FC) (Forward completeness.) For all $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$, the time-varying vector field $f(\cdot, u)$ is forward complete. Moreover, for all $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathcal{K} \times \mathbb{S}^{n-1}$ and for all $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$ bounded over $\mathcal{D}(\lambda)$, the coupled system (2.12) has a unique solution $(\hat{x}, \varepsilon, \xi, \omega) \in C^\infty(\mathbb{R}_+, \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n \times \mathbb{R}^n)$ defined on $[0, +\infty)$.

(NFOT) (No flat observer trajectories.) For all $R > 0$, there exists $\eta > 0$ such that for all $\delta \in \mathcal{V}_R$ satisfying $\sup\{|\delta(x)| \ : \ x \in \mathcal{K}_x\} < \eta$ and all $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathcal{K} \times \mathbb{S}^{n-1}$ such that $(\hat{x}_0, \varepsilon_0) \neq (0,0)$, there exists a positive integer k such that the solution of (2.12) with initial condition $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0)$ satisfies $\hat{x}^{(k)}(0) \neq 0$.

In particular, we show that the classical Luenberger and Kalman observers fit these hypotheses so that the main results may be applied to these observers. For all $k \in \mathbb{N}$, $K \subset \mathbb{R}^n$ and $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$, let

$$\|\delta\|_{k,K} = \sup\left\{\left|\frac{\partial^\ell \delta}{\partial x_{i_1} \cdots \partial x_{i_\ell}}(x)\right| \ : \ 0 \leqslant \ell \leqslant k, \quad 1 \leqslant i_1 \leqslant \cdots \leqslant i_\ell \leqslant n, \quad x \in K\right\}.$$

For any $k \in \mathbb{N}$, any compact subset $K \subset \mathbb{R}^n$ and any $\eta > 0$, $k \in \mathbb{N}$, let

$$\mathscr{N}(k, K, \eta) = \left\{\delta \in C^\infty(\mathbb{R}^n, \mathbb{R}) \ : \ \|\delta\|_{k,K} < \eta\right\}.$$

Remark 2.12. For any open subset $U \subset \mathcal{D}(\lambda)$ relatively compact in $\mathcal{D}(\lambda)$, for all $R > 0$, there exists $\eta > 0$ such that for all $\delta \in \mathcal{V}_R$ bounded by η on U, the feedback $\lambda + \delta$ is such that 0 is asymptotically stable with domain of attraction containing U. This can be easily checked, for example, by choosing a strict proper Lyapunov function V (thanks to a converse Lyapunov theorem such as [TP00, Theorem 1]) corresponding to $\psi : x \mapsto A_{\lambda(x)}x + b\lambda(x)$ satisfying $\frac{\partial V}{\partial x}(x)[\psi(x)] < -V(x)$ for $x \in D(\lambda)$. Then V is also a strict proper Lyapunov function for $\psi_\delta : x \mapsto A_{(\lambda+\delta)(x)}x + b(\lambda+\delta)(x)$ on $B(0,R)$ (since $\delta \equiv 0$ in $B(0,R)$) and also on $U \setminus B(0,R)$ (by selecting η such that $\eta\frac{\partial V}{\partial x}(x)[Bx + b] \leqslant \frac{1}{2}V(x)$, so that $\frac{\partial V}{\partial x}(x)[\psi_\delta(x)] < -\frac{1}{2}V(x)$ for $x \in U \setminus B(0,R)$). Hence, the remaining question is: among these small perturbations (that are easy to design), how many do ensure the observability of the closed-loop system?

The main problem on which we focus is the following.

Problem 2.13. Let $T > 0$. Under genericity assumptions on (A, B, C), does there exist $R, \eta > 0$, a positive integer k and a residual set $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ such that the following property holds? *"For all $\delta \in \mathscr{O} \cap \mathcal{V}_R$ and for all $(\hat{x}_0, \varepsilon_0, \xi_0) \in \mathcal{K}$, system* (2.1) *is observable in time T for the control $u = (\lambda + \delta) \circ \hat{x}$, where $\hat{x}$ follows* (2.12) *with initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0)$ and feedback perturbation δ."*

2.2 Main observability results

2.2.1 Statement of the results

We first state our main theorem, that deals with the observability of system (2.12). Its proof is the most technical part of the chapter, and heavily relies on transversality theory.

Theorem 2.14. *Assume that the pairs (C, A) and (C, B) are observable. Assume that $0 \notin \mathcal{K}_x$. Then there exist $\eta > 0$, a positive integer k and a dense open (in the Whitney C^∞ topology) subset $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ such that the solution to (2.12) with $\delta \in \mathscr{O}$ and initial condition $(\hat{x}(0), \varepsilon(0), \xi(0), \omega(0)) \in \mathcal{K} \times \mathbb{S}^{n-1}$ satisfies*

$$\exists k_0 \in \{0, \dots, k\} \quad : \quad \left.\frac{\mathrm{d}^{k_0}}{\mathrm{d}t^{k_0}}\right|_{t=0} C\omega(t) \neq 0. \tag{2.13}$$

The proof of this theorem can be found in Section 2.3.4. The Whitney C^∞ topology is recalled in Definition 2.26.

Remark 2.15. Property (2.13) is stronger than observability of (2.12) in any time $T > 0$. This implication is shown in Corollary 2.35. Pay attention to the assumption $0 \notin \mathcal{K}_x$. In Section 2.3.5, this assumption is removed, while only slightly weakening our observability result.

Theorem 2.14 leads to the following corollary which states that under genericity assumptions on the system, there exists a generic class of perturbations δ such that the feedback $\lambda + \delta$ makes (2.12) observable.

Corollary 2.16. *Assume that the pairs (C, A) and (C, B) are observable. Assume that 0 is in the interior of $\mathcal{K}_x$. Let $T > 0$. Then there exist $R, \eta > 0$, a positive integer k and a dense open subset $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_R$ such that the solution to (2.12) with $\delta \in \mathscr{O}$ and initial condition $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathcal{K} \times \mathbb{S}^{n-1}$ satisfies*

$$\exists t \in [0, T] \quad : \quad C\omega(t) \neq 0,$$

that is, system (2.1) is observable in time T for the control $u = (\lambda + \delta) \circ \hat{x}$, where $\hat{x}$ follows (2.12) with initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0)$ and feedback perturbation δ.

This result also implies a generic observability property directly on the stabilizing state feedback law λ.

Corollary 2.17. *Assume that the pairs (C, A) and (C, B) are observable. Assume that 0 is in the interior of $\mathcal{K}_x$. Denote by Λ the set of feedbacks $\lambda \in C^\infty(\mathbb{R}^n, \mathbb{R})$ such that 0 is a locally asymptotically stable equilibrium point of the vector field $x \mapsto A_{\lambda(x)}x + b\lambda(x)$. Let $T > 0$ and $\Lambda_T \subset \Lambda$ be the set of feedbacks $\lambda \in \Lambda$ such that (2.1) is observable in time T for the control $u = \lambda \circ \hat{x}$, where $\hat{x}$ follows (2.12) with $\delta \equiv 0$ and initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0)$ in $\mathcal{K}$. Then Λ_T is a dense open subset of Λ.*

This last corollary is an important step toward the achievement of a generic separation principle for SISO bilinear systems. Indeed, it states that if a system is state feedback stabilizable, then generically on the feedback and the system, the inputs produced by the closed-loop system make it observable. However, we will see in Section 2.2.2 that a gap is still to be filled.

The proof of these two corollaries can be found in Section 2.3.5.

Remark 2.18. Because $\mathcal{V}_R$ is not open in the Whitney C^∞ topology, the set $\mathscr{O}$ defined in Corollary 2.16 is not open in the Whitney C^∞ topology, but it is open in the induced topology on $\mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_R$. Also, the set of matrices $(A, B, C) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{1\times n}$ such that (C, A) and (C, B) are both observable is open and dense. As a consequence, "(C, A) and (C, B) are observable" is a generic hypothesis. Contrarily to the strategy followed in [LSG17] or in Chapter 4 for some specific examples, the results of this chapter do not explicitly design any perturbation $\delta \in \mathscr{O}$, but rather state that for almost all bilinear system, almost all perturbation $\delta \in \mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_R$ belongs to $\mathscr{O}$ (in a topological sense).

Finally, the next theorem shows that the classical Luenberger and Kalman observers fit hypotheses (FC) and (NFOT). Hence, our results may be applied to these well-known observers.

Theorem 2.19. *Assume that (C, A) is observable. Assume that λ is bounded over $\mathcal{D}(\lambda)$. Let $Q \in \mathcal{S}_n$. For all $\xi \in \mathcal{S}_n$ and all $u \in \mathbb{R}$, consider the following well-known observers:*

$$f^{\text{Luenberger}}(\xi, u) = 0 \qquad \text{(Luenberger observer)}$$
$$f_Q^{\text{Kalman}}(\xi, u) = \xi A_u' + A_u \xi + Q - \xi C' C \xi \qquad \text{(Kalman observer)}$$

and $L(\xi) = \xi C'$. Then the coupled system (2.12) *given by (f, L) satisfies the hypotheses* (FC) *and* (NFOT) *for any $f \in \{f^{\text{Luenberger}}, f_Q^{\text{Kalman}}\}$.*

The proof of this theorem can be found in Section 2.4.

Remark 2.20. If λ is unbounded over $\mathcal{D}(\lambda)$, then for any open subset U relatively compact in $\mathcal{D}(\lambda)$, we can obtain by smooth saturation of λ a new bounded feedback law λ_{sat} such that $\lambda_{\text{sat}|U} = \lambda_{|U}$, for which the previous statement holds. (In particular $U \subset \mathcal{D}(\lambda_{\text{sat}})$.)

2.2.2 Towards a separation principle?

In Section 1.3 we recalled some existing separation principles for nonlinear uniformly observable systems. For bilinear uniformly observable systems, the Kalman observer is known to be an exponential observer, with arbitrary decay rate by tuning the smallest eigenvalue of Q (see [Hd90, BBH96, GK01, Bes07]). This is crucial in the context of semi-global output feedback stabilization: by choosing a sufficiently fast observer, the trajectories of the state are retained in some compact set. Similarly, one can also consider the Kalman-like observer defined by

$$f_\theta^{\text{Kalman-like}}(\xi, \bar{A}, \bar{C}) = \xi \bar{A}' + \bar{A}\xi + \theta\xi - \xi \bar{C}' \bar{C} \xi$$

where θ is some positive parameter tuning the convergence rate. The (NFOT) hypothesis is still satisfied for the Kalman-like observer, with a proof identical to the Kalman observer (see Section 2.4).

Using Corollary 2.17, one can prove the existence of a perturbation δ such that the new feedback law $\lambda + \delta$ makes the closed-loop system observable. Then, using this new perturbed feedback (still denoted by λ in the next theorem), it is possible to achieve a separation principle? Actually, one can prove the convergence of bounded trajectories.

Theorem 2.21. *Assume that λ is bounded over $\mathcal{D}(\lambda)$ and that all the trajectories $(\hat{x}, \varepsilon, \xi)$ of (2.12) starting from $\mathcal{K}$, with $\delta \equiv 0$, $f = f_\theta^{\text{Kalman-like}}$ and $L(\xi) = \xi C'$, remain in $\mathcal{K}$ and (2.1) is observable in any positive time for the control $u = \lambda \circ \hat{x}$. Then $(0, 0, \xi_\infty)$ is a locally asymptotically stable equilibrium point of (2.12) with basin of attraction containing $\mathcal{K}$, where ξ_∞ is the unique solution of $f_\theta^{\text{Kalman-like}}(\xi_\infty) = 0$.*

Proof. The proof is similar to the strategy used in [GK92] in the uniformly observable case. Let $(\hat{x}, \varepsilon, \xi)$ be a trajectory and $u = \lambda \circ \hat{x}$. Let $\zeta = \xi^{-1}$. Then

$$\dot{\zeta} = -A_u'\zeta - \zeta A_u - \theta\zeta + C'C. \tag{2.14}$$

Hence

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\varepsilon'\zeta\varepsilon &= 2\varepsilon'\zeta\dot{\varepsilon} + \varepsilon'\dot{\zeta}\varepsilon \\ &= 2\varepsilon'\zeta(A_u - \xi C'C)\varepsilon - 2\varepsilon'\zeta A_u\varepsilon - \theta\varepsilon'\zeta\varepsilon + \varepsilon'C'C\varepsilon \\ &= -|C\varepsilon|^2 - \theta\varepsilon'\zeta\varepsilon \\ &\leqslant -\theta\varepsilon'\zeta\varepsilon.\end{aligned}$$

Hence $\varepsilon'\zeta\varepsilon(t) \leqslant e^{-\theta t}(\varepsilon'\zeta\varepsilon)(0)$ for all $t \geqslant 0$. Moreover,

$$\zeta(t) = e^{-\theta t}(\Phi_u'(t))^{-1}\zeta_0(\Phi_u(t))^{-1} + \int_0^t e^{-\theta(t-s)}(\Phi_u'(t))^{-1}\Phi_u'(s)C'C\Phi_u(s)(\Phi_u(t))^{-1}\mathrm{d}s$$

where $\Phi_u(t)$ is the resolvent matrix of $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_u(t) = (A + uB)\Phi_u(t)$. Hence

$$\zeta(t) = e^{-\theta t}(\Phi_u'(t))^{-1}\zeta_0(\Phi_u(t))^{-1} + W_u(t)$$

where $W_u(t)$ is the Gramian-like observability matrix defined by

$$\begin{aligned}W_u(t) &:= \int_0^t e^{-\theta(t-s)}(\Phi_{u(\cdot+s)}'(t-s))^{-1}C'C(\Phi_{u(\cdot+s)}(t-s))^{-1}\mathrm{d}s \\ &\geqslant e^{-\theta\tau}\int_{t-\tau}^t (\Phi_{u(\cdot+s)}'(t-s))^{-1}C'C(\Phi_{u(\cdot+s)}(t-s))^{-1}\mathrm{d}s\end{aligned}$$

for any $\tau \in (0, t)$. For all $(\hat{x}_0, \varepsilon_0, \xi_0) \in \mathcal{K}$, the corresponding input $u = \lambda \circ \hat{x}$ is such that

$$\int_0^\tau |C(\Phi_u(\tau - s))^{-1}x|^2\mathrm{d}s \geqslant \alpha_u|x|^2 \tag{2.15}$$

for all $x \in \mathbb{R}^n$ for some positive constant α_u since u makes (2.1) observable in any positive time. The function $(\hat{x}_0, \varepsilon_0, \xi_0) \mapsto \alpha_{\lambda\circ\hat{x}}$ has a positive minimum α over $\mathcal{K}$ since it is continuous (see [GK92]). Note that if $u = \lambda \circ \hat{x}$, then $u(\cdot + t)$ can also be written as $\lambda \circ \hat{x}$ with initial conditions $(\hat{x}, \varepsilon, \xi)(t) \in \mathcal{K}$. Hence, $\zeta(t) \geqslant W_u(t) \geqslant e^{-\theta\tau}\alpha\mathrm{Id}$, which yields

$$|\varepsilon(t)|^2 \leqslant \frac{1}{\alpha}e^{-\theta(t-\tau)}(\varepsilon'\zeta\varepsilon)(0). \tag{2.16}$$

Thus ε is exponentially converging towards zero. The rest of the proof is identical to [GK92, Theorem 3] and we dot not recall it here, but the arguments are the same than those developed in Sections 3.4.1 and 3.4.3: in the ω-limit set of any trajectory, $\varepsilon \equiv 0$, hence the stabilizing property of λ makes $\hat{x}$ tends towards zero, and ζ to ζ_∞. The local asymptotic stability is obtained by the center manifold theorem. ∎

Hence, to achieve semi-global output feedback stabilization, the difficulty lies in proving that the trajectories of (2.12) are bounded. In the uniformly observable case, it is sufficient to choose θ sufficiently large. Then the exponential decrease of $\varepsilon'\zeta\varepsilon$ and the uniform lower bound α on the observability Gramian yields boundedness of trajectories. However, in the non-uniformly observable case, one need to invoke Corollary 2.16 in order to find a perturbation δ such that $\lambda + \delta$ makes the system observable. But this perturbation depends on θ, and the lower bound of the observability Gramian depends on δ. Therefore, when increasing θ, the lower bound of ζ could tend towards zero, hence nothing shows that increasing θ actually increases the rate of convergence of ε towards 0. This precise question remains open. It is the main obstacle to obtain to generic separation principle for SISO bilinear systems.

This difficulty leads us to consider a more restrictive class of systems, for which at least the observer error ε remains bounded, independently of the observability assumptions. The question is then: for dissipative systems, are we able to use the perturbation strategy developed in this section to set up a separation principle? We will see in Chapter 3 that for such systems, no perturbation is needed to achieve this goal.

2.3 Proofs of the observability results

Ἀγεωμέτρητος μηδεὶς εἰσίτω

Πλάτων

2.3.1 Reminder on transversality theory

As the proof of the observability results rely on transversality, let us first recall some definitions, and the main theorem to be applied. All these results are from [GM88, Part I, Section 1.3], [GG74, Chapter 2] and [Hir94].

Definition 2.22 (Transversality)**.** Let X and Y be two smooth manifolds and $f \in C^\infty(X,Y)$. If Z is a submanifold of Y, we say that f is transversal to Z at $x \in X$ if

$$\Big(f(x) \notin Z\Big) \quad \text{or} \quad \Big(f(x) \in Z \text{ and } T_{f(x)}Z + \operatorname{Im} Df_x = T_{f(x)}Y\Big). \tag{2.17}$$

We say that f is transversal to Z if it is transversal to Z at all point $x \in X$.

Remark 2.23 (Submersion)**.** If $f \in C^\infty(X,Y)$ is a submersion, *i.e.,* if its differential is everywhere surjective, then f is transversal to all submanifold of Y.

Remark 2.24 (When transversal means to avoid)**.** If $\dim Z + \dim X < \dim Y$, then $f \in C^\infty(X,Y)$ is transversal Z if and only if $f(X) \cap Z = \varnothing$.

These two remarks are at the heart of the strategy of the proof (see Section 2.3.2).

Definition 2.25 (Jet bundles)**.** Let X and Y be two smooth manifolds and let $k \in \mathbb{N}$. A k-jet $j^k_x f$ is an equivalence class of (x, f, U, φ) where (U, φ) is a chart on X, $x \in U$ and $f \in C^\infty(U,Y)$. The equivalence relation is given by $j^k_{x_1} f_1 =$

$j^k_{x_2} f_2$ if and only if $x_1 = x_2$ and f_1 and f_2 have the same derivatives at x_1 up to order k. The set of all k-jets from X to Y is denoted by $J^k(X,Y)$. For each $f \in C^\infty(X,Y)$, the mapping $j^k f : X \to J^k(X,Y)$ is defined by $j^k f(x) = j^k_x f$. The mapping $\sigma : J^k(X,Y) \to X$ given by $\sigma : j^k_x f \mapsto x$ is called the source map and the mapping $\tau : J^k(X,Y) \to Y$ given by $\tau : j^k_x f \mapsto f(x)$ is called the target map. Put $J^k_x(X,Y) = \sigma^{-1}(x)$, $J^k(X,Y)_y = \tau^{-1}(y)$ and $J^k_x(X,Y)_y = \sigma^{-1}(x) \bigcap \tau^{-1}(y)$. Then $J^k(X,Y) = \coprod_{x\in X} J^k_x(X,Y) = \coprod_{y\in X} J^k(X,Y)_y = \coprod_{(x,y)\in Y} J^k_x(X,Y)_y$

Set $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. Then $j^k_x f$ is canonically identified to the Taylor polynomial of f of order k at x. Hence $J^k(\mathbb{R}^n, \mathbb{R}^m)$ is canonically identified to $\mathbb{R}^n \times \prod_{i=1}^k L^i_{\text{sym}}(\mathbb{R}^n, \mathbb{R}^m)$, where $L^k_{\text{sym}}(\mathbb{R}^n, \mathbb{R}^m)$ denotes the vector space of symmetric k-linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. In particular it is a finite-dimensional vector space.

If X and Y are smooth manifolds of dimension n and m respectively, (U,φ) and (V,ψ) are charts of X and Y, then $J^k(\varphi(U), \psi(V))$ is an open subset of $J^k(\mathbb{R}^n, \mathbb{R}^m)$ and the map $\theta : J^k(U,V) \to J^k(\varphi(U), \psi(V))$ that sends each jet to its local representation is a bijection. Therefore $J^k(X,Y)$ is a smooth manifold, $(\theta, J^k(U,V))$ can be viewed as a chart of $J^k(X,Y)$, and the topology of $J^k(X,Y)$ is induced by these charts.

The topology considered in the results of Section 2.2 is the Whitney C^∞ topology.

Definition 2.26 (Whitney C^∞ topology)**.** If X and Y are two smooth manifolds, then the Whitney C^∞ topology on the space $C^\infty(X,Y)$ of smooth maps, is the topology whose basis consists of open sets

$$M(U) = \{f \in C^\infty(X,Y) \ : \ j^k f(X) \subset U\}$$

where $0 \leqslant k < +\infty$ and U is an open subset of $J^k(X,Y)$.

Remark 2.27. Let d be a metric on $J^k(X,Y)$ (compatible with its topology). Set $f \in C^\infty(X,Y)$ and let $\delta : X \to \mathbb{R}_+$ be a continuous mapping. Then

$$B_\delta(f) = \{g : C^\infty(X,Y) \ : \ \forall x \in X, \ d(j^k_x f, j^k_x f) < \delta(x)\}$$

is an open set. On compact manifolds, we may find a countable neighborhood basis of f by taking $\left(B_{x\mapsto \frac{1}{n}}(f)\right)_{n\geqslant 1}$.

In Section 2.3.4, we apply the next transversality theorem to prove our main results.

Theorem 2.28 (Goresky-MacPherson theorem, [GM88, Part I, Section 1.3.2])**.** *Let X and Y be two smooth manifolds. If $Z_1 \subset X$ and $Z_2 \subset Y$ are closed subsets with Whitney stratifications, then*

$$\{f \in C^\infty(X,Y) \ : \ f|_{Z_1} \text{ is transversal to } Z_2\} \tag{2.18}$$

is open and dense (in the Whitney C^∞ topology) in $C^\infty(X,Y)$.

Remark 2.29. For any $k \in \mathbb{N}$, closed semi-algebraic sets in $\mathbb{R}^k$ are closed sets that are finite unions of sets defined by finitely many algebraic equalities and inequalities. Closed semi-algebraic sets in $\mathbb{R}^k$ being also subanalytic, they are Whitney stratified according to [GM88, Part I, Section 1.2]. Moreover, according to the Tarski–Seidenberg theorem (see, *e.g.,* [Dri98, Chapter 2, (2.10)]), semi-algebraic sets are closed under projection. Having in mind to apply the result on arbitrary compact sets $\mathcal{K}_x \times \mathcal{K}_\varepsilon \times \mathcal{K}_\xi$ of $\mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n$, we will consider only semi-algebraic compact sets with no loss of generality. A semi-algebraic compact set $\mathcal{K}_\xi$ of $\mathcal{S}_n$ is simply a semi-algebraic compact set of $\mathbb{R}^{n\times n}$ included in $\mathcal{S}_n$.

2.3.2 Strategy of the proof

In order to prove our main Theorem 2.14 and its Corollary 2.16, we need a series of preliminary results that we state and prove below. The main results will appear as corollaries of these subsequent lemmas.

Before we start the more technical elements of the section, let us present the method we follow in order to prove the main results. Theorem 2.14 is an application of transversality theory to our particular problem (see Theorem 2.28 for the statements we rely on; see also [AR67, GG74] for similar but different transversality theorems). Consider a solution to (2.12) for a given perturbation δ of the feedback law, and a set of initial conditions in $\mathcal{K}\times\mathbb{S}^{n-1}$. We set $h : C^\infty(\mathbb{R}^n,\mathbb{R})\times(\mathcal{K}\times\mathbb{S}^{n-1})\times\mathbb{R}^+ \to \mathbb{R}$ to be the smooth map given by

$$h(\delta,(\hat{x}_0,\varepsilon_0,\xi_0,\omega_0),t) = C\omega(t).$$

As stated in Proposition 2.9, to get observability after perturbation of the feedback, we would like to show that there exists δ, preferably small, such that

$$\Big(t \mapsto h(\delta,z_0,t)\Big) \not\equiv 0, \qquad \forall z_0 = (\hat{x}_0,\varepsilon_0,\xi_0,\omega_0) \in \mathcal{K}\times\mathbb{S}^{n-1}. \tag{2.19}$$

A sufficient condition for δ to satisfy (2.19) is that for each $z_0 \in \mathcal{K}\times\mathbb{S}^{n-1}$, there exists an integer k such that $\frac{\mathrm{d}^k}{\mathrm{d}t^k}\Big|_{t=0}(h(\delta,z_0,t)) \neq 0$. In other words, our goal will be achieved if we can prove that there exist δ and a finite set $\mathcal{I}\subset\mathbb{N}$ such that the map $H : C^\infty(\mathbb{R}^n,\mathbb{R})\times(\mathcal{K}\times\mathbb{S}^{n-1}) \to \mathbb{R}^{|\mathcal{I}|}$ given by

$$H(\delta,z_0) = \left(\frac{\mathrm{d}^k}{\mathrm{d}t^k}\bigg|_{t=0} h(\delta,z_0,t)\right)_{k\in\mathcal{I}},$$

never vanishes. This is where transversality theory comes into play. Let N denote the dimension of the surrounding space of $\mathcal{K}\times\mathbb{S}^{n-1}$. We can ensure that there exists δ satisfying (2.19) if we can prove that for some choice of $\mathcal{I}$, with $|\mathcal{I}| > N$, H is *transversal* to $\{0\}$ at $\delta = 0$. That is to say, if we can prove that the rank of the map $H(0,\cdot)$ is maximal, equal to $|\mathcal{I}| > N$, at any of its vanishing points (at which point $H(0,\cdot)$ is then a submersion).

Now it should be noted that in general, proving that a map is transversal to a point is a major hurdle, especially if the dimensions n and N of the spaces are unspecified. As a general rule, considering more orders of derivation of h greatly increases the degrees of freedom of the map H (by including higher order derivatives of v, as jet spaces grow exponentially in dimension), while only slightly increasing the size of the target space. This points towards an augmentation of the rank of H, making a proof of transversality achievable.

The difficulty lies however in producing a "rank increasing property" on H as $|\mathcal{I}|$ increases. That is, finding a symmetry in the successive derivatives of h that proves that for any dimension, a set $\mathcal{I}$ can be found by differentiating h sufficiently many times. The symmetry we use to prove the rank condition on the map H can be described as follows. For $k\in\mathbb{N}$, let

$$h^k(\delta,z_0,t) = CB^k\omega(t).$$

It turns out that if $h^{k+1}(0,z_0,\cdot)$ has a non-zero derivative of any order (including order 0), then we automatically get the rank condition for $h^k(0,z_0,\cdot)$ (this statement will be made precise in Corollary 2.32).

Here the hypothesis that (C, B) is an observable pair becomes crucial. Indeed, observe that $h^k(0, z_0, 0) = CB^k\omega_0$. Hence, for any $\omega_0 \in \mathbb{S}^{n-1}$ there exists a $k \in \{0, \dots, n-1\}$ such that

$$h^k(0, z_0, 0) \neq 0.$$

This in turns induces a partition of $\mathcal{K} \times \mathbb{S}^{n-1}$ into n subsets on each of which at least one of the maps $h^0, \dots, h^{n-1}$ never vanishes. Since $h^{k+1}(0, z_0, \cdot)$ not vanishing implies that the rank condition is satisfied for $h^k(0, z_0, \cdot)$, we chain-apply n successive transversality theorems to prove the existence of a δ such that $h(\delta, z_0, \cdot)$ has always at least one non-zero time derivative at any point $z_0 \in \mathcal{K} \times \mathbb{S}^{n-1}$.

Section 2.3.3 is aimed at making explicit the connection between the rank condition and the family of maps $(h^k)_{k\in\mathbb{N}}$. Section 2.3.4 is dedicated to the effective application of the principles presented in this introduction, which leads to the proof of Theorem 2.14. Section 2.3.5 concludes the proof of the observability statements by taking into account the behavior of the system near the target 0.

2.3.3 Preliminary results

Let $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$ and consider the ordinary differential equation

$$\dot{\omega} = (A + u(t)B)\,\omega. \tag{2.20}$$

For all $k, m \in \mathbb{N}$, let $F_k^m : C^\infty(\mathbb{R}_+, \mathbb{R}) \times \mathbb{R}^n \to \mathbb{R}$ be the function such that

$$F_k^m(u, \omega_0) = CB^m\omega^{(k)}(0)$$

where $t \mapsto \omega(t)$ is the solution of (2.20) with initial condition ω_0.

Let us introduce the $n \times n$ matrix valued polynomials in the indeterminates $X_0, \dots, X_{k-1}$ by:

$$\mathbb{R}^{n\times n}[X_0, \dots X_{k-1}] = \begin{cases} \mathbb{R}^{n\times n} & \text{if } k = 0 \\ \mathbb{R}^{n\times n}[X_0, \dots X_{k-2}][X_{k-1}] & \text{otherwise,} \end{cases}$$

and set

$$\mathbb{R}^{n\times n}\left[(X_k)_{k\in\mathbb{N}}\right] = \bigcup_{k\in\mathbb{N}} \mathbb{R}^{n\times n}[X_0, \dots X_{k-1}].$$

Let $\Psi : \mathbb{R}^{n\times n}\left[(X_k)_{k\in\mathbb{N}}\right] \to \mathbb{R}^{n\times n}\left[(X_k)_{k\in\mathbb{N}}\right]$ be the linear map defined by

$$\Psi(P)(X_0, \dots, X_k) = P(X_0, \dots, X_{k-1})(A + X_0B) + \sum_{i=0}^{k-1} \frac{\partial P}{\partial X_i}(X_0, \dots, X_{k-1})\, X_{i+1},$$

where $k = \min\{\ell \in \mathbb{N} \ : \ P \in \mathbb{R}^{n\times n}[X_0, \dots X_{\ell-1}]\}$.

Finally, let us define the family $(P_k)_{k\in\mathbb{N}}$ of matrix valued polynomials such that $P_0 \in \mathbb{R}^{n\times n}$ and $P_k \in \mathbb{R}^{n\times n}[X_0, \dots X_{k-1}]$, for all $k \geqslant 1$, by

$$P_0 = \mathrm{Id}, \qquad P_{k+1} = \Psi(P_k), \qquad \forall k \in \mathbb{N}. \tag{2.21}$$

It is clear[2] that for all $m \in \mathbb{N}$,

$$F_k^m(u, \omega_0) = \begin{cases} CB^m\omega_0 & \text{if } k = 0 \\ CB^m P_k\left(u^{(0)}, u^{(1)}, \dots, u^{(k-1)}\right)\omega_0 & \text{otherwise,} \end{cases}$$

[2]Note that, for $k \neq 0$, the function F_k^m actually acts on $(k-1)$-jets at zero of functions and not on functions themselves. Consequently, the restriction $F_k^m|_{J_0^\ell(\mathbb{R},\mathbb{R})\times\mathbb{R}^n}$ is well-defined as soon as $\ell \geqslant k-1$. Of course, for $k = 0$, the restriction $F_0^m|_{J_0^\ell(\mathbb{R},\mathbb{R})\times\mathbb{R}^n}$ makes sense only if $\ell \geqslant 0$. In summary, the restriction $F_k^m|_{J_0^\ell(\mathbb{R},\mathbb{R})\times\mathbb{R}^n}$ is well-defined as soon as $\ell \geqslant k$.

where $u^{(i)}$ is shorthand for $\frac{\mathrm{d}^i u}{\mathrm{d}t^i}(0)$ for all $i \in \mathbb{N}$. For all $k \in \mathbb{N}$ and $i \in \mathbb{N}$, $1 \leqslant i \leqslant k$, let $Q_i^k = \dfrac{\partial P_k}{\partial X_{k-i}}$.

Lemma 2.30. *For all $i \in \mathbb{N} \setminus \{0\}$, there exist $R_i^0, \dots, R_i^{i-1} \in \mathbb{R}^{n\times n}[X_0, \dots X_{i-1}]$ such that*[3]

$$Q_i^{i+k} = \sum_{j=0}^{i-1} k^j R_i^j, \qquad \forall k \geqslant 0.$$

Furthermore, $R_i^{i-1} = \dfrac{BP_{i-1}}{(i-1)!}$.

The proof of this technical lemma is postponed to Appendix A.1.

Corollary 2.31. *Let $i, m \in \mathbb{N}$, $i \geqslant 1$. Let $v \in \mathbb{R}^i$ and $\omega_0 \in \mathbb{R}^n$. Either there exists $k_0 \geqslant i$ such that $CB^m Q_i^k(v)\omega_0 \neq 0$ for all $k \geqslant k_0$ or $CB^m Q_i^k(v)\omega_0 = 0$ for all $k \geqslant i$.*

Proof. By Lemma 2.30, we have $Q_i^k = \sum_{j=0}^{i-1}(k-i)^j R_i^j$ for all integers $k \geqslant i$. If $CB^m R_i^j(v)\omega_0 = 0$ for all $j \in \{0, \dots, i-1\}$, then $CB^m Q_i^k(v)\omega_0 = 0$ for all $k \geqslant i$. Otherwise, there exists $j \in \{0, \dots, i-1\}$ such that $CB^m R_i^j(v)\omega_0 \neq 0$. Let $(\ell_0, \dots \ell_{i-1}) \in \mathbb{N}^i$ with $\ell_0 < \dots < \ell_{i-1}$. We have

$$CB^m \begin{pmatrix} Q_i^{i+k_0}(v) \\ \vdots \\ Q_i^{i+k_{i-1}}(v) \end{pmatrix} \omega_0 = \begin{pmatrix} 1 & \ell_0 & \dots & \ell_0^{i-1} \\ \vdots & \vdots & & \vdots \\ 1 & \ell_{i-1} & \dots & \ell_{i-1}^{i-1} \end{pmatrix} CB^m \begin{pmatrix} R_i^0(v) \\ \vdots \\ R_i^{i-1}(v) \end{pmatrix} \omega_0.$$

Since $\ell_0, \dots \ell_{i-1}$ are pairwise different, the Vandermonde matrix is invertible. Consequently, there exists $j \in \{0, \dots, i-1\}$ such that $CB^m Q_i^{i+\ell_j}(v)\omega_0 \neq 0$. Hence, there exist at most $i-1$ positive integers ℓ_j such that $CB^m Q_i^{i+\ell_j}(v)\omega_0 = 0$. Thus, there exists $k_0 \geqslant i$ such that $CB^m Q_i^k(v)\omega_0 \neq 0$ for all $k \geqslant k_0$. ■

For all $P \in \mathbb{R}^{n\times n}[X_0, \dots X_{k-1}]$ and all $v \in \mathbb{R}^{\mathbb{N}}$, we set $P(v) = P(v_0, \dots, v_{k-1})$. The next corollary is the last preliminary result of this section. Proposition 2.34, in which we apply a transversality theorem, critically relies on this full rank property.

Corollary 2.32. *Let $v \in \mathbb{R}^{\mathbb{N}}$, $\omega_0 \in \mathbb{R}^n$ and $m \in \mathbb{N}$. If there exists $i \in \mathbb{N} \setminus \{0\}$ such that $CB^{m+1} P_{i-1}(v)\omega_0 \neq 0$, then there exists $k_0 \in \mathbb{N}$ such that, for all $N \in \mathbb{N} \setminus \{0\}$, the mapping* [4] *$\phi : J_0^{k_0+N-1}(\mathbb{R}, \mathbb{R}) = \mathbb{R}^{k_0+N} \to \mathbb{R}^N$ defined by*

$$\phi(\cdot) = (CB^m P_{k_0}(\cdot)\omega_0, \dots, CB^m P_{k_0+N-1}(\cdot)\omega_0)$$

has a rank N differential at $(v_0, \dots, v_{k_0+N-1})$.

Proof. Assume that there exists $i \geqslant 1$ such that $CB^{m+1} P_{i-1}(v)\omega_0 \neq 0$. Since, according to Lemma 2.30, $R_i^{i-1} = BP_{i-1}/(i-1)!$, this is equivalent to $CB^m R_i^{i-1}(v)\omega_0 \neq 0$. Thus, reasoning as in the proof of Corollary 2.31, the sequence $\left(CB^m Q_i^k(v)\omega_0\right)_{k\geqslant i}$ is not constant equal to zero. Set

$$i_0 = \min\left\{ i \in \mathbb{N} \setminus \{0\} : \left(CB^m Q_i^k(v)\omega_0\right)_{k\geqslant i} \not\equiv 0 \right\}. \tag{2.22}$$

[3] Actually, we can show that $R_i^0, \dots, R_i^{i-1} \in \mathbb{R}^{n\times n}[X_0, \dots X_{i-2}]$.

[4] Note that $\phi(\cdot) = F^m_{\{k_0,\dots,k_0+N-1\}}(\cdot, \omega_0)$, with $F^m_{\{k_0,\dots,k_0+N-1\}}$ defined as in Section 2.3.4.

As a consequence of Corollary 2.31, there exists $k_0 \in \mathbb{N}$ such that $CB^m Q_{i_0}^k(v)\omega_0 \neq 0$ for all $k \geqslant k_0$, *i.e.,*

$$\frac{\partial\left(CB^m P_k \omega_0\right)}{\partial X_{k-i_0}}(v_0, \ldots, v_{k_0+N-1}) = \frac{\partial\left(CB^m P_k \omega_0\right)}{\partial X_{k-i_0}}(v) \neq 0, \qquad \forall k \geqslant k_0,$$

and (by construction of i_0)

$$\frac{\partial\left(CB^m P_k \omega_0\right)}{\partial X_{\ell}}(v_0, \ldots, v_{k_0+N-1}) = \frac{\partial\left(CB^m P_k \omega_0\right)}{\partial X_{\ell}}(v) = 0, \qquad \forall \ell > k - i_0.$$

In other words,

$$D\phi(v_0, \ldots, v_{k_0+N-1}) = \begin{pmatrix} * & \ldots & * & a_0(v) & 0 & & \ldots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & & \vdots \\ * & & \ldots & & * & a_{N-1}(v) & 0 & \ldots 0 \end{pmatrix}, \tag{2.23}$$

with $a_i(v) = CB^m Q_{i_0}^{k_0+i}(v)\omega_0$. The statement follows. ■

2.3.4 Observability away from the target

Using the results of the previous section, we are now able to prove our main Theorem 2.14. In this section, we assume that $0 \notin \mathcal{K}_x$. From now on $t \mapsto (\hat{x}(t), \varepsilon(t), \xi(t), \omega(t))$, or simply $(\hat{x}, \varepsilon, \xi, \omega)$, denotes the solution to (2.12) with initial condition $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0)$.

Let us introduce some new notation. For any $k \in \mathbb{N}$, define the map G^k by:

$$\begin{aligned} G^k : J^k(\mathbb{R}^n, \mathbb{R}) \times \mathcal{K}_\varepsilon \times \mathcal{K}_\xi &\longrightarrow J_0^k(\mathbb{R}, \mathbb{R}) \\ \Big(j^k\delta(\hat{x}_0), \varepsilon_0, \xi_0\Big) &\longmapsto j^k\Big((\lambda + \delta) \circ \hat{x}\Big)(0). \end{aligned}$$

For any finite subset $\mathcal{I} \subset \mathbb{N}$ and any $m \in \mathbb{N}$, set $k_{\mathcal{I}} = \max \mathcal{I}$ and define the maps, $F_{\mathcal{I}}^m$ and $H_{\mathcal{I}}^m$ as follows:

$$\begin{aligned} F_{\mathcal{I}}^m : J_0^{k_{\mathcal{I}}}(\mathbb{R}, \mathbb{R}) \times \mathbb{S}^{n-1} &\longrightarrow \mathbb{R}^{|\mathcal{I}|} \\ (v, \omega_0) &\longmapsto \Big(CB^m P_k(v)\omega_0\Big)_{k \in \mathcal{I}}, \end{aligned}$$

$$H_{\mathcal{I}}^m = F_{\mathcal{I}}^m \circ \left(G^{k_{\mathcal{I}}} \times \mathrm{Id}_{\mathbb{S}^{n-1}}\right).$$

Remark 2.33. Notice that for any $m, k_0 \in \mathbb{N}$ and any $N \in \mathbb{N} \setminus \{0\}$ such that $\mathcal{I} \subset \{k_0, \ldots, k_0 + N - 1\}$, the map $F_{\mathcal{I}}^m$ satisfies

$$F_{\mathcal{I}}^m = \pi_{\mathcal{I}} \circ F_{\{k_0, \ldots, k_0+N-1\}}^m,$$

where $\pi_{\mathcal{I}} : J_0^{k_0+N-1}(\mathbb{R}, \mathbb{R}) = \mathbb{R}^{k_0+N} \to \mathbb{R}^{|\mathcal{I}|}$ denotes the canonical projection onto the factors that correspond to indices in $\mathcal{I}$.

Now we state the following proposition, which leads directly to Theorem 2.14.

Proposition 2.34. *For all $m \in \{0, \dots, n-1\}$, define*

$$E_m = \begin{cases} \mathbb{S}^{n-1} & \text{if } m = 0 \\ \{\omega_0 \in \mathbb{S}^{n-1} \ : \ CB^i\omega_0 = 0, \quad \forall i \in \{0, \dots, m-1\}\} & \text{otherwise.} \end{cases}$$

Suppose (C, A) and (C, B) are observable pairs. Then for every $m \in \{0, \dots, n-1\}$, there exist $k \in \mathbb{N}$, a positive real number η and a dense open subset $\mathscr{O}_m \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ such that for all $(\delta, \hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathscr{O}_m \times \mathcal{K} \times E_m$

$$H^m_{\{0,\dots,k\}}(j^k\delta(\hat{x}_0), \varepsilon_0, \xi_0, \omega_0) \neq 0.$$

Proof. The proof strongly relies on the results of Section 2.3.3 and on the Goresky-MacPherson transversality theorem (see Theorem 2.28). We prove the proposition by finite descending induction on m. Note that since the pair (C, B) is observable, we have $\varnothing = E_n \subset E_{n-1} \subset \cdots \subset E_1 \subsetneq E_0 = \mathbb{S}^{n-1}$.

For $m = n-1$, the result is immediate because, by observability of the pair (C, B), $CB^{n-1}\omega_0 \neq 0$ for all $\omega_0 \in E_{n-1}$. Hence, for $k = 0$ and any positive real number η, we have for all $(\delta, \hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathscr{N}(k, \mathcal{K}_x, \eta) \times \mathcal{K} \times E_{n-1}$,

$$H^{n-1}_{\{0\}}(j^0\delta(\hat{x}_0), \varepsilon_0, \xi_0, \omega_0) = CB^{n-1}\omega_0 \neq 0.$$

Now suppose $1 \leqslant m \leqslant n-1$. Note that, by definition of $E_{m-1} \setminus E_m$,

$$CB^{m-1}\omega_0 \neq 0, \quad \forall \omega_0 \in E_{m-1} \setminus E_m. \tag{2.24}$$

Assume that we are given a $k \in \mathbb{N}$, a positive real number η and a dense open subset $\mathscr{O}_m \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ such that

$$H^m_{\{0,\dots,k\}}(j^k\delta(\hat{x}_0), \varepsilon_0, \xi_0, \omega_0) \neq 0, \qquad \forall (\delta, \hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathscr{O}_m \times \mathcal{K} \times E_m. \tag{2.25}$$

Choose $(\delta, \hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathscr{O}_m \times \mathcal{K} \times E_m$ and put $u(t) = (\lambda + \delta)\big(\hat{x}(t)\big)$. Equation (2.25) implies that $CB^mP_i(u^{(0)}, \dots, u^{(k)})\omega_0 \neq 0$ for an integer $i \in \{0, \dots, k\}$, so, by Corollary 2.32 there exists $k_0 \in \mathbb{N}$ such that, for any positive integer k_1, the map $F^{m-1}_{\{k_0,\dots,k_0+k_1-1\}}$ has a rank k_1 differential at $(u^{(0)}, \dots, u^{(k_0+k_1-1)})$.

Let $i_0 \in \mathbb{N}$ be defined as in the proof of Corollary 2.32. Let $p \in \mathbb{N} \setminus \{0\}$ be such that $\hat{x}^{(p)} \neq 0$ and $\hat{x}^{(q)} = 0$ for all $q < p$ (which exists by hypothesis (NFOT) and $0 \notin \mathcal{K}_x$), and choose $\ell \in \{1, \dots n\}$ so that $\hat{x}^{(p)}_\ell \neq 0$. Put[5]

$$j_0 = \min\Big\{j \geqslant k_0 \ : \ j - i_0 \equiv 0 \pmod p\Big\} \text{ and } \mathcal{I} = \Big\{j_0 + rp \ : \ r \in \{0, \dots, N-1\}\Big\},$$

where N is a positive integer. The (partial) differential of $G^m_{\mathcal{I}}$ with respect to

$$w = \left(\delta, \frac{\partial}{\partial x_\ell}\delta, \dots, \left(\frac{\partial}{\partial x_\ell}\right)^{k_{\mathcal{I}}}\delta\right)\Bigg|_{x=\hat{x}_0}$$

at $X_0 = (j^{k_{\mathcal{I}}}\delta(\hat{x}_0), \varepsilon_0, \xi_0, \omega_0)$ is the submatrix $D_wG^m_{\mathcal{I}}(X_0)$ obtained from $DG^m_{\mathcal{I}}(X_0)$ by deleting all columns that do not correspond to partial derivatives with respect to w. In other words,

$$D_wG^m_{\mathcal{I}}(X_0) = \Big(\mathtt{col}(0) \cdots \mathtt{col}(k_{\mathcal{I}} - 1)\Big).$$

[5]Index j_0 corresponds to the smallest index $j \geqslant k_0$ such that $\hat{x}^{(p)}_\ell$ appears in $u^{(j-i_0)}$.

Each column $\texttt{col}(i)$, $i \in \{0, \dots, k_{\mathcal{I}} - 1\}$ of $D_w G^m_{\mathcal{I}}(X_0)$ satisfies

$$\texttt{col}(i)' = \Big(0 \cdots 0 \; b_i(X_0) \; * \cdots *\Big)', \qquad b_i(X_0) \neq 0,$$

where the non zero coefficient $b_i(X_0)$ appears at the ip th row. According to the Faà di Bruno formula, we have

$$b_i(X_0) = n_i \left(\hat{x}^{(p)}_\ell\right)^i,$$

n_i being a positive integer for each $i \in \{0, \dots, k_{\mathcal{I}} - 1\}$.

It is clear from the definition of $F^m_{\mathcal{I}}$ and Remark 2.33 thereafter that $DF^m_{\mathcal{I}}$ is the submatrix of $DF^m_{\{k_0,\dots,k_{\mathcal{I}}\}}$ (see equation (2.23)) obtained by keeping the i th rows for $i \in \mathcal{I}$. Therefore,

$$\begin{aligned}\operatorname{rank}\left(DH^m_{\mathcal{I}}(X_0)\right) &\geqslant \operatorname{rank}\Big(D_v F^m_{\mathcal{I}}\left(G^{k_{\mathcal{I}}}(X_0), \omega_0\right) \circ D_w G^{k_{\mathcal{I}}}(X_0)\Big) \\ &= \operatorname{rank}\begin{pmatrix} * \cdots * & c_0(X_0) & 0 & & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ * \quad \cdots & & * & c_{N-1}(X_0) & 0 & \cdots 0 \end{pmatrix},\end{aligned}$$

where $c_r(X_0) = a_{j_0+rp}\left(G^{k_{\mathcal{I}}}(X_0), \omega_0\right) b_{j_0+rp}(X_0)$, $r \in \{0, \dots, N-1\}$. Hence $H^{m-1}_{\mathcal{I}}$ has a rank N differential at X_0.

For any $k \in \mathbb{N}$, any compact subset $K \subset \mathbb{R}^n$ and any $\eta > 0$, $k \in \mathbb{N}$, define

$$\mathcal{M}(k, K, \eta) = \left\{\alpha \in J^k(\mathbb{R}^n, \mathbb{R}) : \exists f \in \mathscr{N}(k, K, \eta), \quad \exists a \in K, \quad \alpha = j^k f(a)\right\}.$$

Clearly, $\mathcal{M}(k, K, \eta)$ is an open submanifold of $J^k(\mathbb{R}^n, \mathbb{R})$.

Since the rank is a semi-continuous map, there exists a neighborhood $V \subset \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \eta) \times \mathcal{K}_\varepsilon \times \mathcal{K}_\xi \times E_m$ of $(j^{k_{\mathcal{I}}}_0(\hat{x}_0), \varepsilon_0, \xi_0, \omega_0)$ such that $H^{m-1}_{\mathcal{I}}$ has a rank N on V. Let $\rho \in (0, \eta)$ and $\mathcal{C}(\rho) = \mathcal{C}_x \times \mathcal{C}_\varepsilon \times \mathcal{C}_\xi \times \Omega_m$ be a semi-algebraic compact subset of $\mathcal{K} \times E_m$ such that

$$W := \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \rho) \times \mathcal{C}_\varepsilon \times \mathcal{C}_\xi \times \Omega_m \subset V.$$

Let $B = \left(H^{m-1}_{\mathcal{I}}|_W\right)^{-1}(0)$ and $Z = \pi(B)$, where π is the projection that is parallel to $\mathcal{C}_\varepsilon \times \mathcal{C}_\xi \times \Omega_m$. Then, and because $\mathcal{C}_\varepsilon \times \mathcal{C}_\xi \times \Omega_m$ is compact, $Z \subset \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \rho)$ is a closed semi-algebraic subset. Hence, according to Theorem 2.28, the set

$$\tilde{\mathscr{O}}(\rho) = \Big\{f \in C^\infty\Big(\mathbb{R}^n, \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \rho)\Big) \; : \; f|_{\mathcal{C}_x} \text{ is transversal to } Z\Big\}$$

is open and dense (in the Whitney C^∞ topology) in $C^\infty\Big(\mathbb{R}^n, \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \rho)\Big)$. Moreover, since $H^{m-1}_{\mathcal{I}}|_W$ is a submersion, we have $\operatorname{codim}_{\mathcal{M}(k_{\mathcal{I}},\mathcal{K}_x,\rho)} Z \geqslant \operatorname{codim}_{\mathbb{R}^N}\{0\} - \dim(\mathcal{C}(\rho) \times E_m) = N - \dim(\mathcal{C}(\rho) \times E_m)$. Picking N sufficiently large, we have

$$\operatorname{codim}_{\mathcal{M}(k_{\mathcal{I}},\mathcal{K}_x,\rho)} Z > n$$

in which case, transversal necessarily means to avoid. It follows that

$$\begin{aligned}&\tilde{\mathscr{O}}(\rho) \\ &= \Big\{f \in C^\infty\Big(\mathbb{R}^n, \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \rho)\Big) \; : \; \forall \hat{x} \in \mathcal{C}_x, f(\hat{x}) \notin Z\Big\} \\ &= \Big\{f \in C^\infty\Big(\mathbb{R}^n, \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \rho)\Big) \; : \; \forall (\hat{x}, \varepsilon, \xi, \omega) \in \mathcal{C}(\rho), \Big(f(\hat{x}), \varepsilon, \xi, \omega\Big) \notin B\Big\} \\ &= \Big\{f \in C^\infty\Big(\mathbb{R}^n, \mathcal{M}(k_{\mathcal{I}}, \mathcal{K}_x, \rho)\Big) \; : \; \forall (\hat{x}, \varepsilon, \xi, \omega) \in \mathcal{C}(\rho), H^{m-1}_{\mathcal{I}}\Big(f(\hat{x}), \varepsilon, \xi, \omega\Big) \neq 0\Big\}.\end{aligned}$$

By compactness of $\mathcal{K} \times E_m$, there exists $q \in \mathbb{N}$ such that

$$\mathcal{K} \times E_m = \bigcup_{i=1}^{q} \mathcal{C}(\rho_i). \tag{2.26}$$

Set $\eta = \min\{\rho_i \ : \ i = 1, \dots, q\} > 0$, $k = \max\{k_{\mathcal{I}}(\rho_i) \ : \ i = 1, \dots, q\}$ and define $\tilde{\mathscr{O}} = \bigcap_{i=1}^{q} \tilde{\mathscr{O}}(\rho_i)$. According to (2.26),

$$\tilde{\mathscr{O}} = \Big\{ f \in C^\infty\Big(\mathbb{R}^n, \mathcal{M}(k, \mathcal{K}_x, \eta)\Big) \ : \ \forall (\hat{x}, \varepsilon, \xi, \omega) \in \mathcal{K} \times E_m, \\ H^{m-1}_{\{0,\dots,k\}}\Big(f(\hat{x}), \varepsilon, \xi, \omega\Big) \neq 0 \Big\}.$$

Also, by definition of E_{m-1} and E_m, $H^{m-1}_{\{0\}}(\omega) = CB^{m-1}\omega \neq 0$ for all $\omega \in E_{m-1} \setminus E_m$. Thus,

$$\tilde{\mathscr{O}} = \Big\{ f \in C^\infty\Big(\mathbb{R}^n, \mathcal{M}(k, \mathcal{K}_x, \eta)\Big) \ : \ \forall (\hat{x}, \varepsilon, \xi, \omega) \in \mathcal{K} \times E_{m-1}, \\ H^{m-1}_{\{0,\dots,k\}}\Big(f(\hat{x}), \varepsilon, \xi, \omega\Big) \neq 0 \Big\}$$

is an open dense subset of $C^\infty(\mathbb{R}^n, \mathcal{M}(k, \mathcal{K}_x, \eta))$. Then $\mathscr{O}_{m-1} := \{\tau \circ f \ : \ f \in \tilde{\mathscr{O}}\}$ where τ is the target map is an open dense subset of $\mathscr{N}(k, \mathcal{K}_x, \eta)$ and

$$\mathscr{O}_{m-1} = \Big\{ \delta \in \mathscr{N}(k, \mathcal{K}_x, \eta) \ : \ \forall (\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathcal{K} \times E_{m-1}, \\ H^{m-1}_{\{0,\dots,k\}}(j^k\delta(\hat{x}_0), \varepsilon_0, \xi_0, \omega_0) \neq 0 \Big\}.$$

This concludes the induction and the proof. ■

Proof of Theorem 2.14. Applying Proposition 2.34 to $m = 0$ and recalling the definition of $H^0_{\{0,\dots,k\}}$, we immediately get the main Theorem 2.14. ■

A straightforward consequence of Theorem 2.14 is the following corollary, that deals with the observability of (2.1), as announced in Remark 2.15.

Corollary 2.35. *Assume that (C, A) and (C, B) are observable pairs. Assume that $0 \notin \mathcal{K}_x$. Then there exist $\eta > 0$, $k \in \mathbb{N}$ and an open dense subset $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ such that for all $(\delta, \hat{x}_0, \varepsilon_0, \xi_0) \in \mathscr{O} \times \mathcal{K}$, system (2.1) is observable in any time $T > 0$ for the control $u = (\lambda + \delta) \circ \hat{x}$, where $\hat{x}$ follows (2.12) with initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0)$ and feedback perturbation δ.*

Proof. Applying Proposition 2.34 to $m = 0$, we find that there exist $\eta > 0$, $k \in \mathbb{N}$ and an open dense subset $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ such that for all $(\delta, \hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathscr{O} \times \mathcal{K} \times E_0$, $H^0_{\{0,\dots,k\}}(j^k\delta(\hat{x}_0), \varepsilon_0, \xi_0, \omega_0) \neq 0$. Let $(\delta, \hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathscr{O} \times \mathcal{K} \times \mathbb{S}^{n-1}$, and let $(\hat{x}, \varepsilon, \xi, \omega)$ denote the solution of (2.12) with initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0)$. From the definition of $H^0_{\{0,\dots,k\}}$ it follows that there exists $i \in \mathbb{N}$ such that $C\omega^{(i)}(0) \neq 0$. Consequently, $C\omega|_{[0,T]} \not\equiv 0$, which was to be proved. ■

As stated in Remark 2.15, we now want to complete the compact $\mathcal{K}_x$ with a neighborhood of zero as in Corollary 2.16. We do so in the following section.

2.3.5 Observability near the target

We use Theorem 2.14 to prove Corollary 2.16. In order to do so, we need the following notations and lemmas. For any control $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$, let $\Phi_u : \mathbb{R}_+ \to \mathbb{R}^{n\times n}$ be the flow of the time-varying linear ordinary differential equation (2.20). So $\Phi_u(t)\omega_0$ is the solution of (2.20) at time $t \in \mathbb{R}_+$ with initial condition $\omega_0 \in \mathbb{R}^n$. Notice for instance that $\Phi_0(t) = e^{At}$. Recall that an input $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$ is said to make system (2.1) observable in time $T > 0$ if for all $\omega_0 \in \mathbb{S}^{n-1}$ there exists $t \in [0, T]$ such that $C\Phi_u(t)\omega_0 \neq 0$.

Lemma 2.36. *Let $T > 0$, $\eta_0 = \max\{|C\Phi_0(t)\omega_0| \ : \ t \in [0,T],\ \omega_0 \in \mathbb{S}^{n-1}\}$ and $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$. If*

$$\forall t \in [0,T], \forall \omega_0 \in \mathbb{S}^{n-1}, \quad |C\Phi_u(t)\omega_0 - C\Phi_0(t)\omega_0| < \eta_0, \tag{2.27}$$

then u makes system (2.1) observable in time T.

Proof. Let $t \in [0,T]$ and $\omega_0 \in \mathbb{S}^{n-1}$ be such that $|C\Phi_0(t)\omega_0| = \eta_0$. Using (2.27), we get

$$|C\Phi_u(t)\omega_0| \geqslant |C\Phi_0(t)\omega_0| - |C\Phi_u(t)\omega_0 - C\Phi_0(t)\omega_0| > 0,$$

which shows that u makes system (2.1) observable in time T. ■

Lemma 2.37. *Let $T > 0$. Let $M = \sup\{\|\Phi_0(t)\| \ : \ t \in [0,T]\}$. Let $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$ and let $u_M = \sup\{|u(t)| \ : \ t \in [0,T]\}$. Then there exists a constant $K > 0$ such that for all $t \in [0,T]$ and all $\omega_0 \in \mathbb{S}^{n-1}$,*

$$|\Phi_u(t)\omega_0 - \Phi_0(t)\omega_0| < MKu_M e^{Ku_M}. \tag{2.28}$$

Proof. By the variation of constants formula, for all $t \in [0,T]$ and all $\omega_0 \in \mathbb{S}^{n-1}$,

$$\Phi_u(t)\omega_0 - \Phi_0(t)\omega_0 = \int_0^t \Phi_0(t-s)Bu(s)\Phi_u(s)\mathrm{d}s\,\omega_0.$$

Iterating integrals, we get a (formal) series expansion

$$\int_0^{s_0} \Phi_0(s_0 - s_1)Bu(s)\Phi_u(s)\mathrm{d}s_1 = \sum_{k=0}^{+\infty} J_k \tag{2.29}$$

where

$$J_k = \int_0^{s_0} \cdots \int_0^{s_k} \Psi_k(s_0, \ldots, s_{k+1})\Phi_0(s_{k+1})u(s_0)\cdots u(s_{k+1})\mathrm{d}s_1 \cdots \mathrm{d}s_{k+1}$$

with $\Psi_k(s_0, \ldots, s_{k+1}) = \Phi_0(s_0 - s_1)B \cdots \Phi_0(s_k - s_{k+1})B$.

Then $\|\Psi_k(s_0, \ldots, s_{k+1})\| \leqslant M^{k+1}\|B\|^{k+1}$ and

$$\|J_k\| \leqslant M^{k+2}\|B\|^{k+1}u_M^{k+1}\int_0^{s_0} \cdots \int_0^{s_k} \mathrm{d}s_1 \cdots \mathrm{d}s_{k+1} \leqslant M^{k+2}\|B\|^{k+1}u_M^{k+1}\frac{T^{k+1}}{(k+1)!}.$$

Thus

$$\begin{aligned}\sum_{k=0}^{+\infty} \|J_k\| &\leqslant \sum_{k=0}^{+\infty} M^{k+2}\|B\|^{k+1}u_M^{k+1}\frac{T^{k+1}}{(k+1)!} \\ &\leqslant M^2\|B\|u_M T\sum_{k=0}^{+\infty} M^k\|B\|^k u_M^k\frac{T^k}{k!}\end{aligned}$$

which proves the convergence of the series expansion (2.29) and inequality (2.28) with $K = M\|B\|T$. ■

Proposition 2.38. *Assume that the pair (C, A) is observable. Assume that 0 is in the interior of $\mathcal{K}_x$. Let $T > 0$. Then there exists $R > 0$ such that $B(0, R) \subset \mathcal{K}_x$ and $\eta_1 > 0$ such that the following property holds:*

Let $(\hat{x}, \varepsilon, \xi, \omega)$ be the solution of (2.12) with initial condition $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in B(0, R) \times \mathbb{R}^n \times \mathcal{S}_n \times \mathbb{S}^{n-1}$. Let $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$ such that $\delta(0) = 0$ and $\sup\{|\delta(x)| \ : \ x \in \mathcal{K}_x\} < \eta_1$. If $\hat{x}(t) \in B(0, R)$ for all $t \in [0, T]$, then the control $u : t \mapsto (\lambda + \delta)(\hat{x}(t))$ makes system (2.1) observable in time T.

Proof. Let $T > 0$ and η_0 be as in the statement of Lemma 2.36. The observability of the pair (C, A) yields $\eta_0 > 0$. Let $\eta_1 > 0$ be such that $MK\eta_1 e^{K\eta_1} < \eta_0$. For all $R > 0$ and all $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$ satisfying $\delta(0) = 0$ and $\sup\{|\delta(x)| \ : \ x \in \mathcal{K}_x\} < \eta_1$, let $u_M(R, \delta) = \sup\{|(\lambda + \delta)(x)| \ : \ x \in B(0, R)\}$. Since $\lambda + \delta$ is continuous and $\lambda(0) = \delta(0) = 0$, $u_M(\cdot, \delta)$ is a continuous non decreasing function on $\mathbb{R}_+$ such that $u_M(0, 0) = 0$ and $u_M(R, \delta) \leqslant u_M(R, 0) + \eta_1$. Then, we can choose $R > 0$ such that $MK(u_M(R, 0) + \eta_1)e^{K(u_M(R,0)+\eta_1)} < \eta_0$. Since $u_M(\cdot, 0)$ is non decreasing, it is possible to choose R such that $B(0, R) \subset \mathcal{K}_x$.

Now, fix $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$ satisfying $\delta(0) = 0$ and $\sup\{|\delta(x)| \ : \ x \in \mathcal{K}_x\} < \eta_1$. Let $(\hat{x}, \varepsilon, \xi, \omega)$ be the solution of (2.12) with initial condition $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in B(0, R) \times \mathbb{R}^n \times \mathcal{S}_n \times \mathbb{S}^{n-1}$. Then $MKu_M(R, \delta)e^{Ku_M(R,\delta)} < \eta_0$. Hence, from Lemmas 2.36 and 2.37, if $\hat{x}(t) \in B(0, R)$ for all $t \in [0, T]$, then the control $u : t \mapsto (\lambda + \delta)(\hat{x}(t))$ makes system (2.1) observable in time T. ∎

Proof of Corollary 2.16. Let $R > 0$ and η_1 be as in Proposition 2.38. Let $r \in (0, R)$ and $\rho \in (0, r)$. We apply Corollary 2.35 to the compact $\mathcal{K}_x \setminus B(0, r)$. Since the statement holds for some η small enough, we assume without loss of generality that $\eta < \eta_1$: there exist $\eta \in (0, \eta_1)$, $k \in \mathbb{N}$ and an open dense subset $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x \setminus B(0, r), \eta)$ such that for all $(\delta, \hat{x}_0, \varepsilon_0, \xi_0) \in \mathscr{O} \times (\mathcal{K}_x \setminus B(0, r)) \times \mathcal{K}_\varepsilon \times \mathcal{K}_\xi$, system (2.1) is observable in any time $T > 0$ for the control $u = (\lambda + \delta) \circ \hat{x}$, where $\hat{x}$ follows (2.12) with initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0)$ and feedback perturbation δ.

Let

$$\mathscr{O}' = \left\{\tilde{\delta} \in \mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_\rho \ : \ \exists \delta \in \mathscr{O}, \forall x \in \mathcal{K}_x \setminus B(0, r), \tilde{\delta}(x) = \delta(x)\right\}.$$

Then $\mathscr{O}'$ is open and dense in $\mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_\rho$ (in the Whitney C^∞ induced topology) since $\mathscr{O}$ is open and dense in $\mathscr{N}(k, \mathcal{K}_x \setminus B(0, r), \eta)$. Moreover, if $\tilde{\delta} \in \mathscr{O}'$, then system (2.1) is still observable in any time $T > 0$ for the control $u = (\lambda + \tilde{\delta}) \circ \hat{x}$ with initial conditions $(\hat{x}_0, \varepsilon_0, \xi_0)$ in $(\mathcal{K}_x \setminus B(0, r)) \times \mathcal{K}_\varepsilon \times \mathcal{K}_\xi$.

Let $(\tilde{\delta}, \hat{x}_0, \varepsilon_0, \xi_0) \in \mathscr{O}' \times \mathcal{K}$. If $\hat{x}_0 \notin B(0, r)$, then the result holds from above. On the other hand, assume that $\hat{x}_0 \in B(0, r)$. If $\hat{x}(t) \in B(0, R)$ for all $t \in [0, T]$, then according to Proposition 2.38, (2.1) is observable in time T for the control $u = (\lambda + \tilde{\delta}) \circ \hat{x}$. Otherwise, there exists $t_0 \in (0, T)$ such that $\hat{x}(t_0) \notin B(0, r)$. Apply Corollary 2.35 with the new initial condition $(\hat{x}(t_0), \varepsilon(t_0), \xi(t_0))$ and with the same perturbation $\tilde{\delta}$. Then (2.1) is observable in time $T > t_0$ for the control $u = (\lambda + \tilde{\delta}) \circ \hat{x}$. ∎

Proof of Corollary 2.17. Let $T > 0$ and $\lambda \in \Lambda$. Let R, η, k and $\mathscr{O}$ be as in Corollary 2.16. Since $\mathscr{O}$ is dense (in the Whitney C^∞ topology) in $\mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_R$, for all neighborhoods $\mathscr{U}$ of $\lambda \in \Lambda$, there exists $\delta \in \mathscr{O}$ such that $\lambda + \delta \in \mathscr{U} \cap \Lambda_T$. Hence,

Λ_T is a dense subset of Λ. Moreover,

$$\begin{aligned}\Lambda_T &= \left\{\lambda \in \Lambda \ : \ \forall (\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathcal{K} \times \mathbb{S}^{n-1}, \exists t \in [0,T], C\omega(t) \neq 0\right\} \\ &= \bigcap_{(\hat{x}_0,\varepsilon_0,\xi_0,\omega_0)\in\mathcal{K}\times\mathbb{S}^{n-1}} h^{-1}_{\hat{x}_0,\varepsilon_0,\xi_0,\omega_0}(\mathcal{C}^\infty([0,T],\mathbb{R}) \setminus \{0\})\end{aligned}$$

where $h_{\hat{x}_0,\varepsilon_0,\xi_0,\omega_0} : \Lambda \to C^\infty([0,T],\mathbb{R})$ is given by $h_{\hat{x}_0,\varepsilon_0,\xi_0,\omega_0}(\lambda) = C\omega|_{[0,T]}$ where ω is the solution of (2.12) with initial condition $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0)$ and $\delta \equiv 0$. The map h is continuous, the set $C^\infty([0,T],\mathbb{R}) \setminus \{0\}$ is open and the set $\mathcal{K} \times \mathbb{S}^{n-1}$ is compact. Thus Λ_T is open in Λ. ■

2.4 Application to classical observers

In this section, we show that there exist observers such that the key hypotheses (FC) and (NFOT) are satisfied. In particular, we show that both the Luenberger observer and the Kalman observer satisfy these hypotheses, as stated in Theorem 2.19. Hence, the main Theorem 2.14 and its Corollary 2.16 apply to these observers. While (FC) has already been studied for such observers (see *e.g.,* [GK01, Bes07]), (NFOT) is more difficult to check, and relies on the fact that the observer dynamics is somehow compatible with the Kalman observability decomposition.

For the sake of generality, we state the results of this section for an arbitrary output dimension m (*i.e.,* $C \in \mathbb{R}^{m\times n}$).

Regarding hypothesis (FC), the following result is well-known.

Proposition 2.39. *Assume that λ is bounded over $\mathcal{D}(\lambda)$. Let $Q \in \mathcal{S}_n$. For all $\xi \in \mathcal{S}_n$ and all $u \in \mathbb{R}$, consider the following well-known observers:*

$$f^{\text{Luenberger}}(\xi, u) = 0 \qquad \text{(Luenberger observer)}$$
$$f^{\text{Kalman}}_Q(\xi, u) = \xi A_u' + A_u\xi + Q - \xi C'C\xi \qquad \text{(Kalman observer)}$$

and $L(\xi) = \xi C'$. Then the coupled system (2.12) *given by (f, L) satisfies the hypothesis* (FC) *for any $f \in \{f^{\text{Luenberger}}, f^{\text{Kalman}}_Q\}$.*

Let us investigate hypothesis (NFOT). First, we state sufficient conditions for it to hold, and then show that they are satisfied by both the Kalman and Luenberger observers.

For all $A_0 \in C^\infty(\mathbb{R}_+, \mathbb{R}^{n\times n})$ and for all $C_0 \in \mathbb{R}^{m\times n}$, let $f(\cdot, A_0, C_0)$ be a forward complete time-varying vector field over $\mathcal{S}_n$. Let $L : \mathcal{S}_n \times \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times m}$. For all invertible matrices $T \in \mathbb{R}^{n\times n}$, for all $(\bar{A}, \bar{C}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times n}$ and for all $\xi \in \mathcal{S}_n$, let $(\bar{f}, \bar{L})$ be defined by

$$\begin{cases} \bar{f}(T\xi T', T\bar{A}T^{-1}, \bar{C}T^{-1}) = Tf(\xi, \bar{A}, \bar{C})T' \\ \bar{L}(T\xi T', \bar{C}T^{-1}) = TL(\xi, \bar{C}). \end{cases} \tag{2.30}$$

For all $(\bar{A}, \bar{C}, \bar{b}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times n} \times \mathbb{R}^n$, we consider the following dynamical observer system

$$\begin{cases} \dot{\hat{x}} = \bar{A}\hat{x} + \bar{b} - \bar{L}(\xi, \bar{C})\bar{C}\varepsilon \\ \dot{\varepsilon} = \left(\bar{A} - \bar{L}(\xi, \bar{C})\bar{C}\right)\varepsilon \\ \dot{\xi} = \bar{f}(\xi, \bar{A}, \bar{C}). \end{cases} \tag{2.31}$$

For all $k \in \{1, \dots, n\}$, let $(\bar{A}, \bar{C}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times n}$ having the following structure:

$$\bar{A} = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \qquad \bar{C} = \begin{pmatrix} C_1 & 0 \end{pmatrix}, \tag{2.32}$$

with suitable matrices $A_{11} \in \mathbb{R}^{k\times k}$, $A_{21} \in \mathbb{R}^{(n-k)\times k}$, $A_{22} \in \mathbb{R}^{n-k}$ and $C_1 \in \mathbb{R}^{m\times k}$. For any solution of (2.31), set similarly

$$\hat{x} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \xi = \begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi_{12}' & \xi_{22} \end{pmatrix}.$$

Proposition 2.40. *Assume that the pair (C, A) is observable. Assume that for all invertible matrices $T \in \mathbb{R}^{n\times n}$, for all $(\bar{f}, \bar{L})$ as in (2.30), for all $k \in \{1, \dots, n\}$ and for all $(\bar{A}, \bar{C}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times n}$ as in (2.32), the following hypotheses hold.*

H1. *There exists (f_{11}, L_1) such that*

$$\begin{cases} \dot{\hat{x}}_1 = A_{11}\hat{x}_1 + b_1 - L_1(\xi_{11}, C_1)C_1\varepsilon_1 \\ \dot{\varepsilon}_1 = (A_{11} - L_1(\xi_{11}, C_1)C_1)\,\varepsilon_1 \\ \dot{\xi}_{11} = f_{11}(\xi_{11}, A_{11}, C_1) \end{cases} \tag{2.33}$$

where (f_{11}, L_1) is such that

$$\bar{f}(\xi, \bar{A}, \bar{C}) = \begin{pmatrix} f_{11}(\xi_{11}, A_{11}, C_1) & * \\ * & * \end{pmatrix}, \qquad \bar{L}(\xi, \bar{C}) = \begin{pmatrix} L_1(\xi_{11}, C_1) \\ * \end{pmatrix}.$$

H2. *If $(C_1, A_{11}) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{k\times k}$ is an observable pair, then the solutions of (2.31) are such that for any initial conditions, $L_{11}(\xi_{11}(t), C_1)C_1\varepsilon_1(t) \to 0$ as $t \to +\infty$.*

H3. *For all $\xi_{11} \in \mathcal{S}_k$ and all $C_1 \in \mathbb{R}^{m\times k}$, $\ker L_1(\xi_{11}, C_1) \cap \operatorname{Im} C_1 = \{0\}$.*

Then the coupled system (2.12) given by $(f(\cdot, A_u, C), L(\cdot, C))$ satisfies the hypothesis (NFOT).

Remark 2.41. In the case where T is the identity matrix and $k = n$, (H1) is clearly satisfied, (H2) means that the correction term $L(\xi, \bar{C})\bar{C}\varepsilon$ converges to zero for any observable pair $(\bar{A}, \bar{C})$, and (H3) means that the correction term is null if and only if $\bar{C}\varepsilon = 0$. We will see in Theorem 2.19 that these hypotheses are clearly satisfied for the Luenberger and Kalman observers.

Remark 2.42. Hypothesis (H1) can be seen as a compatibility condition between the observer dynamics and the Kalman observability decomposition: when $\bar{A}$ is of the standard form (2.32), the observer acts autonomously on the upper left matrix block, which will correspond to the observable part of the system.

This proposition is a consequence of the series of lemmas that follows. Until the end of the proof of Proposition 2.40, assume that its hypotheses are satisfied. For any $\mu : \mathbb{R}^n \to \mathbb{R}$, F_μ denotes the vector field over $\mathbb{R}^n$ given by $F_\mu(x) = A_{\mu(x)}x + b\mu(x)$.

Lemma 2.43. *For all $R > 0$, there exists $\eta > 0$ such that for all $\delta \in \mathcal{V}_R$ satisfying $\sup\{|\delta(x)| \ : \ x \in \mathcal{K}_x\} < \eta$, 0 is the unique equilibrium point of $F_{\lambda+\delta}$ lying in $\mathcal{K}_x$.*

Proof. Let $R > 0$ and $\delta \in \mathcal{V}_R$. Let $x \in \mathcal{K}_x$ be such that $F_{\lambda+\delta}(x) = 0$. Then,

$$0 = F_{\lambda+\delta}(x) = F_\lambda(x) + \delta(x)(Bx + b).$$

Then $|F_\lambda(x)| = |\delta(x)|\,|Bx + b|$. Set $C_1 = \inf\{|F_\lambda(x)| \ : \ x \in \mathcal{K}_x \backslash B(0, R)\}$. Since 0 is not in the closure of $\mathcal{K}_x \backslash B(0, R)$, we get by uniqueness of the equilibrium point of F_λ that $C_1 > 0$. Set also $C_2 = \sup\{|Bx + b| \ : \ x \in \mathcal{K}_x\}$. Since $\mathcal{K}_x$ is compact, $C_2 < +\infty$. Set $\eta = \frac{C_1}{C_2}$. Assume that $\sup\{|\delta(x)| : x \in \mathcal{K}_x\} < \eta$. Then,

$$F_\lambda(x) \leqslant \eta\,|Bx + b| \leqslant C_1.$$

Hence $x \in B(0, R)$ by definition of C_1. Then $\delta(x) = 0$. Hence $F_\lambda(x) = 0$. Thus, $x = 0$ since 0 is the unique equilibrium point of F_λ. Moreover, by definition of $\mathcal{V}_R$, $F_{\lambda+\delta}(0) = 0$. ■

Lemma 2.44. *Assume that the pair (C, A) is observable. Let $(u_0, \hat{x}_0, \varepsilon_0, \xi_0) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n$. Let $(\hat{x}, \varepsilon, \xi)$ be the solution of (2.11) given by the initial condition $(\hat{x}_0, \varepsilon_0, \xi_0)$ and the constant input $u \equiv u_0$. If $\hat{x}$ is constant, then for all $t \in \mathbb{R}_+$, $L(\xi(t), C)C\varepsilon(t) = 0$.*

Proof. Let $(u_0, \hat{x}_0, \varepsilon_0, \xi_0) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n$. Let $(\hat{x}, \varepsilon, \xi)$ be the solution of (2.11) given by the initial condition $(\hat{x}_0, \varepsilon_0, \xi_0)$ and the constant input $u \equiv u_0$. Assume that $\hat{x}$ is constant, *i.e.,* $\hat{x} \equiv \hat{x}_0$. Set $A_0 = A + u_0 B$ and $b_0 = bu_0$. Then $\dot{\hat{x}} \equiv 0$ yields

$$A_0\hat{x} + b_0 - L(\xi, C)C\varepsilon \equiv 0.$$

Since $\hat{x}$ is constant, so is $L(\xi)C\varepsilon$. Then, set $K = L(\xi, C)C\varepsilon$. It remains to show that $K = 0$.

Let $k = \operatorname{rank} \mathcal{O}(C, A_0)$ where $\mathcal{O}(C, A_0)$ is defined by (2.5) Since $C \neq 0$ (since (C, A_0) is observable), $k \geqslant 1$. According to the Kalman observability decomposition, there exists an invertible matrix $T \in \mathbb{R}^{n\times n}$ such that $\bar{A} = TA_0T^{-1}$ and $\bar{C} = CT^{-1}$ have the following structure:

$$\bar{A} = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \qquad \bar{C} = \begin{pmatrix} C_1 & 0 \end{pmatrix}, \tag{2.34}$$

with suitable matrices $A_{11} \in \mathbb{R}^{k\times k}$, $A_{21} \in \mathbb{R}^{(n-k)\times k}$, $A_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$ and $C_1 \in \mathbb{R}^{m\times k}$. Moreover, the pair (C_1, A_{11}) is observable. For the sake of readability, we omit the horizontal bars over the submatrices (for instance, A_{11} is a submatrix of $\bar{A}$ and not of A). Similarly, set

$$\bar{x} = Tx = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \bar{\hat{x}} = T\hat{x} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}, \quad \bar{\varepsilon} = T\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix},$$
$$\bar{b}_0 = Tb_0 = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \ \bar{K} = TK = \begin{pmatrix} K_1 \\ K_2 \end{pmatrix}, \ \bar{\xi} = T\xi T' = \begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi'_{12} & \xi_{22} \end{pmatrix}.$$

Then, according to (2.30), we have the following observed control system on $\bar{x}$, and the corresponding observer:

$$\begin{cases} \dot{\bar{x}} = \bar{A}\bar{x} + \bar{b}_0 \\ y = \bar{C}\bar{x} \\ \dot{\hat{\bar{x}}} = \bar{A}\hat{\bar{x}} + \bar{b}_0 - \bar{L}(\xi, \bar{C})\bar{C}\bar{\varepsilon} \\ \dot{\bar{\varepsilon}} = \left(\bar{A} - \bar{L}(\xi, \bar{C})\bar{C}\right)\bar{\varepsilon} \\ \dot{\bar{\xi}} = \bar{f}(\bar{\xi}, \bar{A}, \bar{C}). \end{cases} \tag{2.35}$$

Then, according to hypothesis (H1), we can write

$$\begin{cases} \dot{\xi}_{11} = f_{11}(\xi_{11}, A_{11}) \\ \dot{\hat{x}}_1 = A_{11}\hat{x}_1 + b_1 - L_1(\xi_{11}, C_1)C_1\varepsilon_1 \\ \dot{\varepsilon}_1 = (A_{11} - L_1(\xi_{11}, C_1)C_1)\,\varepsilon_1. \end{cases} \tag{2.36}$$

Since the pair (C_1, A_{11}) is observable, (H1) and (H2) yield $L_1(\xi_{11}(t), C_1)C_1\varepsilon_1(t) \to 0$ as $t \to +\infty$. The equality $K_1 = L_1(\xi_{11}(t), C_1)C_1\varepsilon_1(t)$ thus yields $K_1 = 0$. Then, by hypotheses (H1) and (H3), $\bar{C}\varepsilon \equiv C_1\varepsilon_1 \equiv 0$. Hence $K = 0$. Finally, we have $K = T^{-1}\bar{K} = 0$. ■

Lemma 2.45. *Let $(\delta, \hat{x}_0, \varepsilon_0, \xi_0) \in C^\infty(\mathbb{R}^n, \mathbb{R}) \times \mathcal{K}$. Let $(\hat{x}, \varepsilon, \xi)$ be the solution of (2.12) given by $(\delta, \hat{x}_0, \varepsilon_0, \xi_0)$. Set $u_0 = (\lambda + \delta)(\hat{x}_0)$. Let $(\hat{x}_\omega, \varepsilon_\omega, \xi_\omega)$ be the solution of (2.11) given by the initial condition $(\hat{x}_0, \varepsilon_0, \xi_0)$ and the constant input $u \equiv u_0$. If $\hat{x}^{(i)}(0) = 0$ for all $i \in \mathbb{N} \setminus \{0\}$, then $\hat{x}_\omega$ is constant and*

$$(\varepsilon_\omega^{(k)}(0), \xi_\omega^{(k)}(0)) = (\varepsilon^{(k)}(0), \xi^{(k)}(0)) \tag{2.37}$$

for all $k \in \mathbb{N}$.

Proof. Assume that $\hat{x}^{(i)}(0) = 0$ for all $i \in \mathbb{N} \setminus \{0\}$. Then, for all $i \in \mathbb{N} \setminus \{0\}$,

$$A^{(i)}_{(\lambda+\delta)(\hat{x})}(0) = 0. \tag{2.38}$$

According to the ODE version of the Cauchy-Kovalevskaya theorem, $(\hat{x}_\omega, \varepsilon_\omega, \xi_\omega)$ is analytic in a neighborhood of 0. Hence, it is sufficient to show that

$$(\hat{x}_\omega^{(k)}(0), \varepsilon_\omega^{(k)}(0), \xi_\omega^{(k)}(0)) = (\hat{x}^{(k)}(0), \varepsilon^{(k)}(0), \xi^{(k)}(0)) \tag{2.39}$$

for all $k \in \mathbb{N}$. By definition of $(\hat{x}, \varepsilon, \xi)$ and $(\hat{x}_\omega, \varepsilon_\omega, \xi_\omega)$, we have

$$(\hat{x}_\omega(0), \varepsilon_\omega(0), \xi_\omega(0)) = (\hat{x}_0, \varepsilon_0, \xi_0) = (\hat{x}(0), \varepsilon(0), \xi(0)).$$

Let $k \in \mathbb{N}$. Assume that for all $i \in \{0, \dots, k\}$, (2.39) is satisfied. Then we prove that (2.39) is also satisfied for $i = k + 1$. Using Faà di Bruno's formula and (2.38), we get

$$\begin{aligned} \xi^{(k+1)}(0) &= f\left(\xi, A_{(\lambda+\delta)(\hat{x})}, C\right)^{(k)}(0) \\ &= f\left(\xi, A_{(\lambda+\delta)(\hat{x}(0))}, C\right)^{(k)}(0) && \text{(by (2.38))} \\ &= f\left(\xi_\omega, A_{(\lambda+\delta)(\hat{x}(0))}, C\right)^{(k)}(0) && \text{(by induction hypothesis)} \\ &= \xi_\omega^{(k+1)}(0). \end{aligned}$$

Likewise, we obtain $\varepsilon^{(k+1)}(0) = \varepsilon_\omega^{(k+1)}(0)$ and $\hat{x}^{(k+1)}(0) = \hat{x}_\omega^{(k+1)}(0)$. ■

Lemma 2.46. *Assume that the pair (C, A) is observable. Let $(\hat{x}_0, \varepsilon_0, \xi_0) \in \mathcal{K}$. Let $R > 0$, $\eta > 0$ as in Lemma 2.43 and $\delta \in \mathcal{V}_R$ satisfying $\sup\{|\delta(x)| \ : \ x \in \mathcal{K}_x\} < \eta$. Let $(\hat{x}, \varepsilon, \xi)$ be the solution of (2.12) given by $(\delta, \hat{x}_0, \varepsilon_0, \xi_0)$. If for all $i \in \mathbb{N} \setminus \{0\}$, $\hat{x}^{(i)}(0) = 0$, then $\hat{x} \equiv \varepsilon \equiv 0$.*

Proof. Assume that for all $i \in \mathbb{N} \setminus \{0\}$, $\hat{x}^{(i)}(0) = 0$. Set $u_0 = (\lambda + \delta)(\hat{x}_0)$. Let $(\hat{x}_\omega, \varepsilon_\omega, \xi_\omega)$ be the solution of (2.12) given by the initial condition $(\hat{x}_0, \varepsilon_0, \xi_0)$ and the constant input $u \equiv u_0$. According to Lemma 2.45, $\hat{x}_\omega \equiv \hat{x}_0$ and for all $k \in \mathbb{N}$, $(\varepsilon_\omega^{(k)}(0), \xi_\omega^{(k)}(0)) = (\varepsilon^{(k)}(0), \xi^{(k)}(0))$. Then, by Lemma 2.44, we get that $L(\xi_\omega, C)C\varepsilon_\omega \equiv 0$. Hence, $A_{u_0}\hat{x}_\omega + bu_0 \equiv 0$ *i.e.,* $A_{(\lambda+\delta)(\hat{x}_0)}\hat{x}_\omega(t) + b(\lambda + \delta)(\hat{x}_0) = 0$ for all $t \in \mathbb{R}_+$. In particular, at $t = 0$ we have that $F_{\lambda+\delta}(\hat{x}_0) = 0$. Hence, from Lemma 2.43, $\hat{x}_0 = 0$. By uniqueness of the solution of (2.12) for a given initial condition, it remains to prove that $\varepsilon_0 = 0$ in order to get that $\hat{x} \equiv \varepsilon \equiv 0$. Since the pair (C, A) is observable, it is sufficient to prove that $CA^k\varepsilon_0 = 0$ for all $k \in \mathbb{N}$. We proceed by induction. From Lemma 2.44, $L(\xi_\omega(0), C)C\varepsilon_\omega(0) = 0$. Then, according to hypothesis (H3), $C\varepsilon_0 = C\varepsilon_\omega(0) = 0$. Let $k \in \mathbb{N}$. Assume that $CA^i\varepsilon_0 = 0$ for all $i \in \{0, \dots, k-1\}$. We prove in the following that $CA^k\varepsilon_0 = 0$. From Lemma 2.44, $(L(\xi_\omega, C)C\varepsilon_\omega)^{(i)}(0) = 0$ for all $i \in \mathbb{N}$. Hence, by Lemma 2.45, we get for all $i \in \mathbb{N}$, $(L(\xi, C)C\varepsilon)^{(i)}(0) = (L(\xi_\omega, C)C\varepsilon_\omega)^{(i)}(0) = 0$ and then $C\varepsilon^{(i)}(0) = CA^i_{u_0}\varepsilon_0 = CA^i\varepsilon_0$ since $u_0 = (\lambda + \delta)(\hat{x}_0) = (\lambda + \delta)(0) = 0$. Then,

$$
\begin{aligned}
0 &= (L(\xi_\omega, C)C\varepsilon_\omega)^{(k)}(0) && \text{(by Lemma 2.44)} \\
&= (L(\xi, C)C\varepsilon)^{(k)}(0) && \text{(by Lemma 2.45)} \\
&= \sum_{i=0}^{k} \binom{k}{i} L(\xi, C)^{(k-i)}(0)C\varepsilon^{(i)}(0) && \text{(by Leibniz rule)} \\
&= \sum_{i=0}^{k} \binom{k}{i} L(\xi, C)^{(k-i)}(0)CA^i\varepsilon_0 \\
&= L(\xi_0, C)CA^k\varepsilon_0. && \text{(by induction hypothesis)}
\end{aligned}
$$

Thus, by hypothesis (H3), $CA^k\varepsilon_0 = 0$, which concludes the induction and the proof. ■

This concludes the series of lemmas necessary to prove Proposition 2.40 and Theorem 2.19.

Proof of Proposition 2.40. The statement follows directly from the contrapositive of Lemma 2.46. ■

Proof of Theorem 2.19. Recall that, according to Proposition 2.39, the Luenberger observer and the Kalman observer satisfy (FC). It remains to show that the sufficient conditions stated in the Proposition 2.40 are satisfied by these observers to conclude the proof of Theorem 2.19.

Let $Q \in \mathcal{S}_n$. For all $(\bar{A}, \bar{C}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times n}$ and all $\xi \in \mathcal{S}_n$, let

$$
\begin{aligned}
&f^{\text{Luenberger}}(\xi, \bar{A}, \bar{C}) = 0 && \text{(Luenberger observer)} \\
&f_Q^{\text{Kalman}}(\xi, \bar{A}, \bar{C}) = \xi\bar{A}' + \bar{A}\xi + Q - \xi\bar{C}'\bar{C}\xi && \text{(Kalman observer)}
\end{aligned}
$$

and $L(\xi, C) = \xi\bar{C}'$. Let $f \in \{f^{\text{Luenberger}}, f_Q^{\text{Kalman}}\}$. According to Proposition 2.39, the time-varying vector field f is forward complete. For all invertible matrices $T \in \mathbb{R}^{n\times n}$, for all $(\bar{A}, \bar{C}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times n}$ and for all $\xi \in \mathcal{S}_n$, let $(\bar{f}, \bar{L})$ be defined by

$$\begin{cases} \bar{f}(T\xi T', T\bar{A}T^{-1}, \bar{C}T^{-1}) = Tf(\xi, \bar{A}, \bar{C})T' \\ \bar{L}(T\xi T', \bar{C}T^{-1}) = TL(\xi, \bar{C}). \end{cases} \tag{2.40}$$

Then

$$\bar{L}(T\xi T', \bar{C}T^{-1}) = TL(\xi, \bar{C}) = T\xi\bar{C}' = T\xi T'(\bar{C}T^{-1})' = L(T\xi T', \bar{C}T^{-1}).$$

Hence $\bar{L} = L$. Moreover, if $f = f^{\text{Luenberger}}$, then $\bar{f} = f = 0$. Otherwise, if $f = f_Q^{\text{Kalman}}$ and then

$$\begin{aligned} \bar{f}(T\xi T', T\bar{A}T^{-1}, \bar{C}T^{-1}) &= Tf(\xi, \bar{A}, \bar{C})T' \\ &= T\xi\bar{A}' + \bar{A}\xi + Q - \xi\bar{C}'\bar{C}\xi T' \\ &= T\xi T'(T\bar{A}T^{-1})' + (T\bar{A}T^{-1})T\xi T' \\ &\quad + TQT' - T\xi T'(\bar{C}T^{-1})'\bar{C}T^{-1}T\xi T' \\ &= f_{TQT'}^{\text{Kalman}}(T\xi T', T\bar{A}T^{-1}, \bar{C}T^{-1}), \end{aligned}$$

Hence it is sufficient to prove that, for all $(\bar{A}, \bar{C}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times n}$ satisfying (2.32), (f, L) satisfies hypotheses (H1), (H2) and (H3). Hypothesis (H1) requires some computations to check that if $(\bar{A}, \bar{C})$ is of the form (2.32), then (2.33) is satisfied with

$$f_{11}(\xi_{11}, \bar{A}_{11}, \bar{C}_1) = \begin{cases} 0 & \text{if } f = f^{\text{Luenberger}} \\ \xi_{11}\bar{A}_{11}' + \bar{A}_{11}\xi_{11} + Q_{11} - \xi_{11}\bar{C}_1'\bar{C}_1\xi_{11} & \text{if } f = f_Q^{\text{Kalman}} \end{cases} \tag{2.41}$$

and $L_1(\xi_{11}, \bar{C}_1) = \xi_{11}\bar{C}_1'$. Hence, for any $f \in \{f^{\text{Luenberger}}, f_Q^{\text{Kalman}}\}$, f_{11} is an observer of the same form than f acting on $\mathbb{R}^k$. Hypothesis (H2) follows from the fact that these well-known observers guaranty that the correction term $L_1(\xi_{11}, \bar{C}_1)\bar{C}_1\varepsilon_1$ goes to 0 as soon as the pair $(\bar{C}_1, \bar{A}_{11})$ is observable (see *e.g.,* [Bes07, Chapter 1, Theorems 3 and 4]). Hypothesis (H3) is clear: for all $\xi_{11} \in \mathcal{S}_k$ and all $\bar{C}_1 \in \mathbb{R}^{m\times k}$, if $\varepsilon_1 \in \mathbb{R}^k$ is such that $\xi_{11}\bar{C}_1'\bar{C}_1\varepsilon_1 = 0$, then $\bar{C}_1\varepsilon_1 = 0$ since ξ_{11} is invertible. Thus the conclusion of Proposition 2.40 holds. ■

Chapter 3

Dissipative systems

A light from the shadows shall spring;

J. R. R. Tolkien, *The Fellowship of the Ring*

Abstract. *The distance between two trajectories of a same state-affine dissipative system is always non-increasing. In this chapter, we show that this property is a powerful tool in the context of output feedback stabilization. Contrarily to the previous chapter, we do not assume that the target is observable, and we do not use any perturbation strategy of the feedback law. The* 0*-detectability condition is proved to be a necessary and sufficient condition to set up a separation principle for dissipative systems. The proof relies on a small gain Luenberger observer. The results are applied on a Ćuk converter and a heat exchanger. Numerical simulations are provided.*

Contents

Figures and Tables

Introduction

In Chapter 2, we have shown for SISO bilinear systems that observability at the target is sufficient to render the closed-loop system observable by means of a perturbation of the feedback law. However, as detailed in Section 2.2.2, this strategy is not sufficient to achieve semi-global dynamic output feedback stabilization. The main difficulty lies in the fact that trajectories of the closed-loop system may be unbounded, and this cannot be avoided by tuning the observer gain. To counter this phenomenon, one could consider the case of systems with bounded trajectories. In particular, if the bilinear system is of the form $\dot{x} = (A + uB)x$ with $A + uB$ a dissipative matrix for all $u \in \mathbb{R}$, then trajectories of the system are bounded. Then, one can apply the perturbation strategy of Chapter 2, and show that dynamic output feedback stabilization is achieved. However, the aim of this chapter is to show that for state-affine dissipative systems, a perturbation strategy is superfluous. For such systems, the necessary and sufficient condition to achieve a separation principle is 0-detectability (see Condition 1.11), which is weaker than observability at the target. The main results are stated and proved in Sections 3.2 and 3.4, respectively. Two examples of application on engineering systems are investigated in Section 3.3.

3.1 Problem statement

Definition 3.1 (State-affine systems)**.** A control system is said to be *state-affine* if it is of the form

$$\dot{x} = A(u)x + B(u) \tag{3.1}$$

where $x \in \mathbb{R}^n$ is the state of the system, $u \in C^0(\mathbb{R}_+, \mathcal{U})$ is the input, $\mathcal{U} \subset \mathbb{R}^p$ is the set of admissible controls and $A : \mathcal{U} \to \mathbb{R}^{n\times n}$ and $B : \mathcal{U} \to \mathbb{R}^n$ are continuous maps.

In particular, bilinear systems considered in Chapter 2 are state-affine. Note that in this chapter, we remove the SISO assumption that has prevailed up to now.

Definition 3.2 (Dissipative system)**.** The state-affine system (3.1) is said to be *dissipative* over an admissible set $\mathcal{U} \subset \mathbb{R}^p$ if there exists a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that for all $u \in \mathcal{U}$,

$$PA(u) + A(u)'P \leqslant 0. \tag{3.2}$$

Many physical systems satisfy such a dissipativity property. For example, it is the case for input-state-output port-Hamiltonian systems (see, *e.g.,* [SJ+14]). The key of this property relies in the following proposition, stating that the distance (in the metric associated to P) between two trajectories sharing the same input is non-increasing.

Proposition 3.3. *Let x_1 and x_2 be two solutions of a dissipative system* (3.1) *with the same continuous input $u : \mathbb{R}_+ \to \mathcal{U}$. Then $t \mapsto (x_1(t) - x_2(t))'P(x_1(t) - x_2(t))$ is non-increasing.*

Proof. For all $t \in \mathbb{R}_+$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(x_1 - x_2)'P(x_1 - x_2) &= (x_1 - x_2)'P(\dot{x}_1 - \dot{x}_2) + (\dot{x}_1 - \dot{x}_2)'P(x_1 - x_2) \\ &= (x_1 - x_2)'(PA(u) + A(u)'P)(x_1 - x_2) \\ &\leqslant 0. \qquad \text{(by (3.2))}\end{aligned}$$

■

The aim of this chapter is to show the interest of dissipativity in the context of output feedback stabilization. We show for dissipative systems with linear output that local asymptotic state feedback stabilizability and 0-detectability (see Condition 1.11) are sufficient to prove semi-global asymptotic dynamic output feedback stabilizability. The key point is that 0-detectability is a much weaker assumption than uniform observability. According to Theorem 1.12, it is even a necessary condition. Contrarily to the previous chapters, we do not follow any perturbation strategy of the feedback law: for dissipative systems, applying the observability results of Section 2.2 is not useful. In Section 2.4, we provide various examples and applications of the result.

3.2 Main results on dissipative systems

Let n, m and p be positive integers, $A : \mathbb{R}^p \to \mathbb{R}^{n\times n}$ and $B : \mathbb{R}^p \to \mathbb{R}^n$ be two locally Lipschitz maps, and $C \in \mathbb{R}^{m\times n}$. For all $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$, we consider the following observation-control system:

$$\begin{cases}\dot{x} = A(u)x + B(u) \\ y = Cx\end{cases} \tag{3.3}$$

where x is the state of the system, u is the input and y is the output.

Theorem 3.4. *Assume that* (3.3) *satisfies Conditions 1.10 (local) and 1.11 are satisfied. Let $\mathcal{D}(\lambda)$ be the basin of attraction of a stabilizing state feedback λ. Assume moreover that λ is locally Lipschitz. If* (3.3) *is dissipative over $\mathcal{U} = \lambda(\mathcal{D}(\lambda))$, then it is globally stabilizable by means of a dynamic output feedback.*

Moreover, the dynamic output feedback is designed with the following Luenberger observer with dynamic gain:

$$\begin{cases}\dot{x} = A(\lambda(\hat{x}))x + B(\lambda(\hat{x})) \\ \dot{\hat{x}} = A(\lambda(\hat{x}))\hat{x} + B(\lambda(\hat{x})) - \alpha(\hat{x}, C\varepsilon)P^{-1}C'C\varepsilon\end{cases} \tag{3.4}$$

where $\hat{x}(0)$ lies in $\mathcal{D}(\lambda)$ and α is a locally Lipschitz function given by (3.9).

By requiring the observer gain to be constant, it is still possible to obtain a semi-global result.

Theorem 3.5. *Assume that* (3.3) *satisfies Conditions 1.10 (local) and 1.11 are satisfied. Let $\mathcal{D}(\lambda)$ be the basin of attraction of a stabilizing state feedback λ. Assume moreover that λ is locally Lipschitz. If* (3.3) *is dissipative over $\mathcal{U} = \lambda(\mathcal{D}(\lambda))$, then*

for all compact sets $\mathcal{K}_x \times \mathcal{K}_{\hat{x}} \subset \mathbb{R}^n \times \mathcal{D}(\lambda)$*, there exits* $\alpha_0 > 0$ *such that for all* $\alpha \in (0, \alpha_0)$*,* $(0,0)$ *is a locally asymptotically stable equilibrium point of*

$$\begin{cases} \dot{x} = A(\lambda(\hat{x}))x + B(\lambda(\hat{x})) \\ \dot{\hat{x}} = A(\lambda(\hat{x}))\hat{x} + B(\lambda(\hat{x})) - \alpha P^{-1}C'C\varepsilon \end{cases} \tag{3.5}$$

with basin of attraction containing $\mathcal{K}_x \times \mathcal{K}_{\hat{x}}$*.*

Theorems 3.4 and 3.5 are proved in Section 3.4.

Remark 3.6. Condition 1.10 implies that $b(\lambda(0)) = 0$. In the following, we assume with no loss of generality that $\lambda(0) = 0$.

Remark 3.7 (0-detectability)**.** Set $A_0 = A(0)$. Then 0-detectability (*i.e.,* Condition 1.11) is equivalent to the detectability of the pair (C, A_0). In particular, the local and global versions of the condition are equivalent. Moreover, the set of pairs (C, A_0) that are detectable is open and dense in $\mathbb{R}^{m\times n} \times \mathbb{R}^{n\times n}$. In that sense, Theorem 3.4 is a generic separation principle for dissipative systems.

Remark 3.8 (Necessary and sufficient conditions)**.** Combining Theorem 3.4 with Theorems 1.9 and 1.12, we obtain somehow necessary and sufficient conditions for the output feedback stabilization of dissipative systems. Moreover, Theorem 1.14 claims that under the assumptions of Theorem 3.4, Condition 1.13 is also satisfied.

Remark 3.9 (Small gain)**.** Note that the observer gain α must be chosen sufficiently small, in both the semi-global and global results. This is a crucial step of the proof. In particular, it is not possible to choose α as large as desired to accelerate the convergence. The underlying phenomenon is the following. If α is small enough, the observer dynamics is close to a stabilizing one, since λ is a stabilizing state feedback. Hence, by usual Lyapunov arguments, $\hat{x}$ will not escape $\mathcal{D}(\lambda)$. Due to the dissipativity property and 0-detectability, ε will eventually tends towards 0. Then, x will also enter in $\mathcal{D}(\lambda)$, and its dynamics will be close enough to the stabilizing dynamics. Thus, both x and $\hat{x}$ tend towards the origin. Local asymptotic stability results from linearization.

Note that taking $p = n$ and C the identity matrix, 0-detectability is trivially satisfied, hence Theorem 3.4 implies the following corollary, which is an interesting result in itself about the stabilization of dissipative systems.

Corollary 3.10. *Any dissipative state-affine system that is locally asymptotically stabilizable by means of a locally Lipschitz feedback law is also globally asymptotically stabilizable by means of a dynamic feedback.*

We extended this corollary to the more general framework of nonlinear weakly contractive control systems in [Bri+21c]. A copy of this article is enclosed in Appendix D. The results of this chapter and of [Bri+21c], as well as their link with the Jurdjevic and Quinn approach (see [JQ78]) suggest that dissipativity is a powerful tool in the context of stabilization. Keeping this fact in mind, we set up similar strategies in Chapter 4.

Let $\varepsilon = \hat{x} - x$ be the error between the actual state of the system and the observer, so that (3.4) can be rewritten as

$$\begin{cases} \dot{\hat{x}} = A(\lambda(\hat{x}))\hat{x} + B(\lambda(\hat{x})) - \alpha P^{-1}C'C\varepsilon \\ \dot{\varepsilon} = \left(A(\lambda(\hat{x})) - \alpha(\hat{x}, C\varepsilon)P^{-1}C'C\right)\varepsilon. \end{cases} \tag{3.6}$$

In the proof, we focus on the stability properties of this $(\hat{x}, \varepsilon)$-system, which is equivalent to (3.4).

Remark 3.11 (On dissipativity)**.** Dissipativity of the system is the key point of the result. It implies that the function $V : \varepsilon \mapsto \varepsilon' P\varepsilon$ is a Lyapunov function for the ε-subsystem of (3.6) as long as $\hat{x} \in \mathcal{D}(\lambda)$. The proof is similar to Proposition 3.3. Indeed,

$$\begin{aligned} \frac{\mathrm{d}V(\varepsilon)}{\mathrm{d}t} &= \varepsilon' P\dot{\varepsilon} + \dot{\varepsilon}' P\varepsilon \\ &= \varepsilon' \left(PA(\lambda(\hat{x})) + A(\lambda(\hat{x}))'P\right)\varepsilon - 2\alpha\varepsilon' C'C\varepsilon \\ &\leqslant -2\alpha|C\varepsilon|^2 \qquad \text{(by (3.2))} \\ &\leqslant 0. \end{aligned}$$

3.3 Examples and applications

In this section, we provide some examples and applications to illustrate the main Theorem 3.5. We focus on the semi-global result to illustrate the role of the observer gain α. First, note that if $A(u) = (J(u) - R(u))\mathcal{H}$ for some positive-definite matrix $\mathcal{H}$ and positive semi-definite (resp. skew-symmetric) matrix $R(u)$ (resp. $J(u)$), B is linear and $C = B'\mathcal{H}$, then we recognize an input-state-output port-Hamiltonian system (see *e.g.* [SJ+14]). In that case, a static output stabilizing feedback is given by $u = -ky$ for any $k > 0$. However, for the same dynamics with a different linear output (*i.e.* such that $C \neq B'\mathcal{H}$), our result provides a methodology for semi-global dynamic output feedback stabilization when the pair $(C, A(0))$ is detectable. The examples of this section are in this form.

Example 3.12 (Harmonic oscillator)**.** Consider (3.3) with

$$A(u) = \begin{pmatrix} 0 & -(1+u) \\ 1+u & 0 \end{pmatrix}, \; B(u) = \begin{pmatrix} u \\ 0 \end{pmatrix} \text{ and } C = \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

Let $\lambda : \mathbb{R}^2 \ni (x_1, x_2) \mapsto -x_1$. Then $W : \mathbb{R}^2 \ni x \mapsto |x|^2$ is a Lyapunov function for the vector field $f : x \mapsto A(\lambda(x))x + B(\lambda)$. Indeed, for any solution x of (3.3),

$$\frac{\mathrm{d}W(x)}{\mathrm{d}t} = 2x'A(\lambda(x))x + 2x'B(\lambda(x)) = -2x_1^2$$

since $A(u)$ is skew-symmetric for all $u \in \mathbb{R}$. According to the LaSalle's invariance principle, the ω-limit set of the trajectory is the largest positively invariant set contained in $\{x \in \mathbb{R}^2 \; : \; x_1 \equiv 0\}$. Note that $\lambda \equiv 0$ and $\dot{x}_1 = -x_2$ on this set. Then $x \to 0$. Hence λ is a globally asymptotically stabilizing feedback law. The Kalman observability matrix of the pair $(C, A(0))$ is the full rank matrix

$$\begin{pmatrix} C \\ CA(0) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Hence $(C, A(0))$ is observable, and *a fortiori* detectable. Thus, Conditions 1.10 and 1.11 are satisfied, and we may apply Theorem 3.5 to find a semi-globally asymptotically stabilizing dynamic output feedback: for all compact sets $K_1 \times K_2 \subset \mathbb{R}^n \times \mathbb{R}^n$, there exists $\alpha_0 > 0$ such that for all $\alpha \in (0, \alpha_0)$, $(0, 0)$ is an asymptotically stable equilibrium point with basin of attraction containing $K_1 \times K_2$ of (3.5).

3.3.1 Ćuk converter

The averaged model of the Ćuk converter given in Figure 3.2 can be written as follow:

$$\dot{x} = \begin{pmatrix} 0 & -(1-u) & 0 & 0 \\ 1-u & 0 & u & 0 \\ 0 & -u & 0 & -1 \\ 0 & 0 & 1 & -\frac{1}{R} \end{pmatrix} Px + \begin{pmatrix} E \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{3.7}$$

where x_1 and x_3 are the fluxes in the inductances L_1 and L_3, x_2 and x_4 are the charges in the capacitors C_2 and C_4, R is the load resistance, E is the voltage source and $P = \operatorname{diag}\left(\frac{1}{L_1}, \frac{1}{C_2}, \frac{1}{L_3}, \frac{1}{C_4}\right)$. As in [ROE01], the goal is to stabilize the system at

$$x^\star = \left(\frac{L_1}{REV_d^2}, C_2V_d + E, -\frac{L_3}{R}V_d, -C_4V_d\right)'$$

for some output capacitor voltage V_d, which is attained for $u^\star = \frac{V_d}{V_d+E}$. Assume that only the charge x_2 is measured, and address the problem of output feedback stabilization. In order to match (3.7) and (3.3), we set $\bar{x} = x - x^\star$ and $\bar{u} = u - u^\star$. Then (3.7) can be rewritten as (3.3) by replacing x by $\bar{x}$ and u by $\bar{u}$ and with

$$A(\bar{u}) = \begin{pmatrix} 0 & -(1-u^\star-\bar{u}) & 0 & 0 \\ 1-u^\star-\bar{u} & 0 & u^\star+\bar{u} & 0 \\ 0 & -u^\star-\bar{u} & 0 & -1 \\ 0 & 0 & 1 & -\frac{1}{R} \end{pmatrix} P,$$

$$B(\bar{u}) = \bar{u}b \text{ with } b = \begin{pmatrix} C_2x_2^\star \\ L_3x_3^\star - L_1x_1^\star \\ -C_2x_2^\star \\ 0 \end{pmatrix} \text{ and } C = \begin{pmatrix} 0, & 1, & 0, & 0 \end{pmatrix}.$$

Remark that $u \equiv 1$ and $u \equiv 0$ renders (3.3) unobservable, since the Kalman observability matrices of the pairs $(C, A(1-u^\star))$ and $(C, A(-u^\star))$ are not invertible. So the well-known results for dynamic output feedback stabilization of uniformly observable systems do not apply. Theorem 3.5 may overcome this difficulty.

The system is dissipative since $PA(\bar{u}) + A(\bar{u})'P$ is negative semi-definite for all inputs $\bar{u}$. The pair $(C, A(0))$ is observable, and *a fortiori* detectable, since its Kalman observability matrix is full rank as soon as $u^\star \neq 1$ and $u^\star \neq 0$ *i.e.* $E \neq 0$ and $V_d \neq 0$. Consider the saturated feedback law $\lambda(\bar{x}) = \operatorname{sat}(-\beta b'P\bar{x})$, where $\beta > 0$ is a tuning parameter and sat is a saturation function such that $u^\star + \lambda$ lies in $(0, 1)$, which is always possible since $u^\star \in (0, 1)$. Then $x \mapsto x'Px$ is a Lyapunov function of the vector field $f : x \mapsto A(\lambda(x))x + B(\lambda(x))$, and according to the LaSalle's invariance principle, the ω-limit set of any trajectory is the largest

positively invariant set contained in $\{x \in \mathbb{R}^2 \,:\, b'Px \equiv 0\}$, which gives $x \to 0$ when $(b'P, A(0))$ is observable. Hence, for almost all choice of parameters, λ is a globally asymptotically stabilizing feedback law. One may also choose any other locally asymptotically stabilizing feedback law, for example the one given in [ROE01].

Then, Theorem 3.5 applies, and (3.5) gives a semi-globally asymptotically stabilizing dynamic output feedback. In Figures 3.3 and 3.4, we provide numerical simulations for the following choice of parameters (as in [ROE01]):

L_1	C_2	L_3	C_4
$10.9\,\mathrm{mH}$	$22.0\,\mu\mathrm{F}$	$10.9\,\mathrm{mH}$	$22.9\,\mu\mathrm{F}$
R	E	V_d	β
$22.36\,\Omega$	$12\,\mathrm{V}$	$25\,\mathrm{V}$	10^{-4}

Table 3.1 – Numerical values for the simulation of the Ćuk converter.

For these values, the pair $(b', A(0))$ is observable, hence λ is a stabilizing state feedback law. We choose the initial conditions $x(0) = 0$ and $\hat{x}(0) = x^\star$. In Figure 3.3, we plot the output voltage $\frac{x_4}{C_4}$ that we want to stabilize at V_d for the state feedback law λ and for the dynamic output feedback based on the Luenberger observer for $\alpha = 1$, $\alpha = 10$ and $\alpha = 100$. In Figure 3.4, we plot the error between the actual state of the system and the observer for the same values of α. When α is larger, the observer converges faster to the state of the system. For $\alpha = 100$, $\hat{x}$ converges quickly to x, and then the dynamics of x obtained via the dynamic output feedback is close to the one obtained via state feedback. On the contrary $\alpha = 1$ leads to a slow convergence of the observer. Then, the state dynamics is very close to the one with the constant control $u \equiv \lambda(\hat{x}(0)) = u^\star$ Finally, $\alpha = 10$ is a compromise between these two behaviors: the state dynamics is similar to the case where $\alpha = 1$ at the beginning, and to the case where $\alpha = 100$ at the end of the simulation.

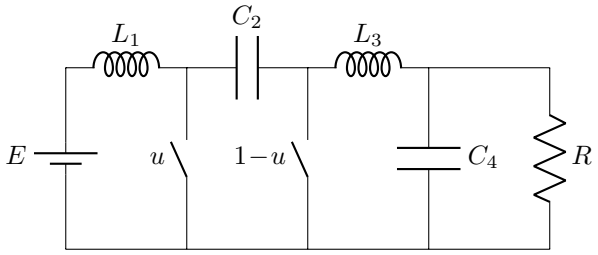


Figure 3.2 – Ideal Ćuk converter.

Remark 3.13. The matrix $A(0)$ is Hurwitz for any $u^\star \in (0, 1)$. Hence, the constant control $\bar{u} = 0$ *i.e.* $u = u^\star$ stabilizes the system at the target point. This phenomenon is due to the load resistance R. However, the user does not have any control on R, so this strategy potentially leads to a very slow stabilization. Indeed, taking $R \to +\infty$ or $R \to 0$, some eigenvalues of $A(0)$ converge to the imaginary axis. In this case, the damping assignment state feedback is much more efficient, and that is why we build a dynamic output feedback based on this state feedback. A similar remark holds for the next example.

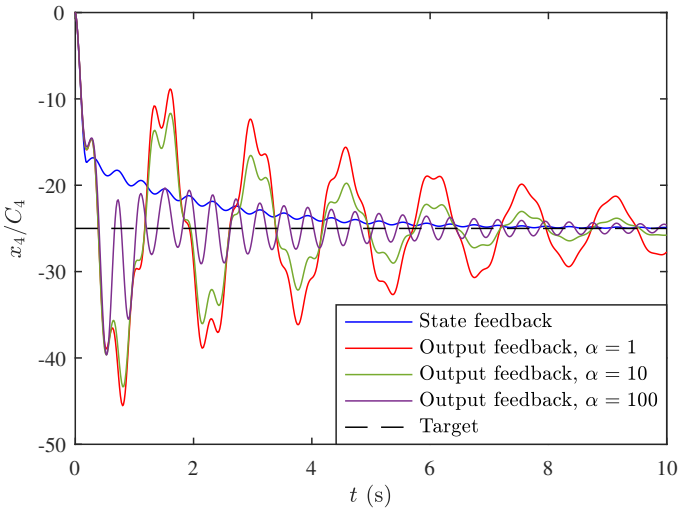


Figure 3.3 – Output voltage of the Ćuk converter with the state feedback law λ and with the corresponding dynamic output feedback law based on the Luenberger observer for different values of α.

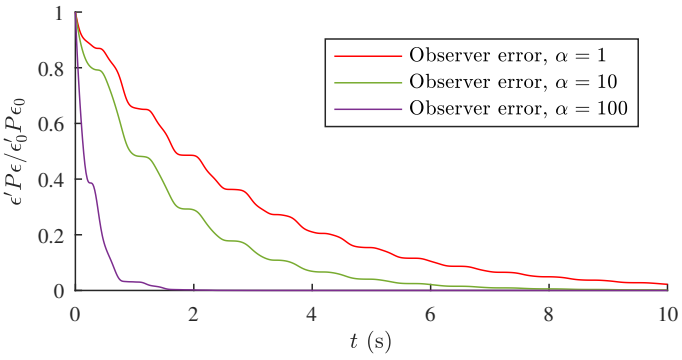


Figure 3.4 – Evolution of the error between the actual state of the Ćuk converter and the observer for different values of α.

3.3.2 Heat exchanger

In [Zit+20] (which we refer reader to for details), a model of a counter-current heat exchanger is introduced. The system is 6-dimensional, and each component x_i of the state represents the temperature of one exchanger's compartment. After a change of coordinates and control (as in the previous Section 3.3.1), the system can be rewritten in form of (3.3) with

$$A(\bar{u}) = \begin{pmatrix} -k\mathrm{Id}_3 + \gamma_1(u^\star + \bar{u})J & k\mathrm{Id}_3 \\ k\mathrm{Id}_3 & -k\mathrm{Id}_3 + \gamma_2 J' \end{pmatrix}, \; B(\bar{u}) = \bar{u}b$$
$$\text{with } b = \Big(E - \gamma_1 x_1^\star,\, \gamma_1(x_1^\star - x_2^\star),\, \gamma_1(x_2^\star - x_3^\star),\, 0,\, 0,\, 0\Big)',$$

and $C = \Big(0,\, 0,\, 0,\, 1,\, 0,\, 0\Big)$ where Id_3 is the 3×3 identity matrix, k, γ_1, γ_2, E are positive physical constants, and

$$J = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

With G a positive physical constant of the system, each control $u^\star > 0$ leads to exactly one equilibrium state $x^\star$ such that $A(0)x^\star = \Big(Eu^\star,\, 0,\, 0,\, 0,\, 0,\, G\Big)'$. The matrix $A(0)$ is invertible according to [Zit+20]. Again, this system is not uniformly observable. Indeed, the determinant of the Kalman observability matrix of the pair $(C, A(\bar{u}))$ is $k^3\gamma_2^6(k^2 - \gamma_1\gamma_2(\bar{u}+u^\star))^3$. Hence, the constant input $\bar{u} \equiv \frac{k^2}{\gamma_1\gamma_2} - u^\star$ renders (3.3) unobservable.

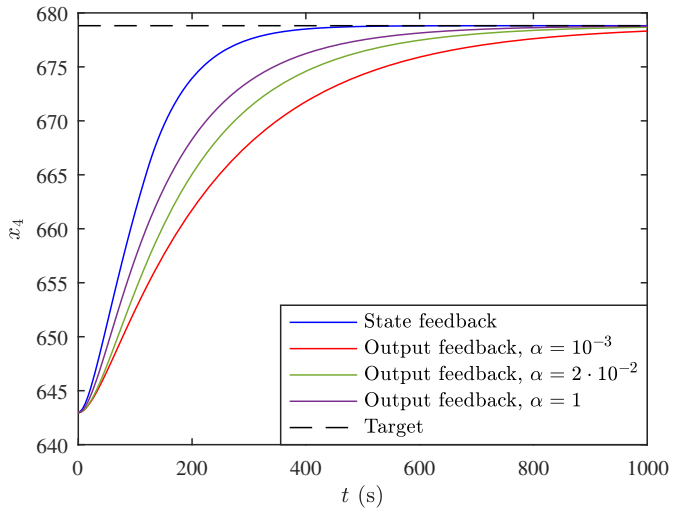


Figure 3.5 – Output enthalpy of the heat exchanger with the state feedback law λ and with the corresponding dynamic output feedback law based on the Luenberger observer for different values of α.

However, Theorem 3.5 can be applied. Choose $\lambda(\bar{x}) = \mathrm{sat}\,(-\beta b'\bar{x})$, where $\beta > 0$ is a tuning parameter and sat is a saturation function such that $u^\star + \lambda$ lies in an interval $(0, u_M)$, which is always possible if $u^\star \in (0, u_M)$. If the pair $(b', A(0))$ is detectable, we apply the LaSalle's invariance principle to the Lyapunov function $x \mapsto x'x$, and get that x converge towards 0. Moreover, $A(\bar{u}) + A(\bar{u})'$ is negative

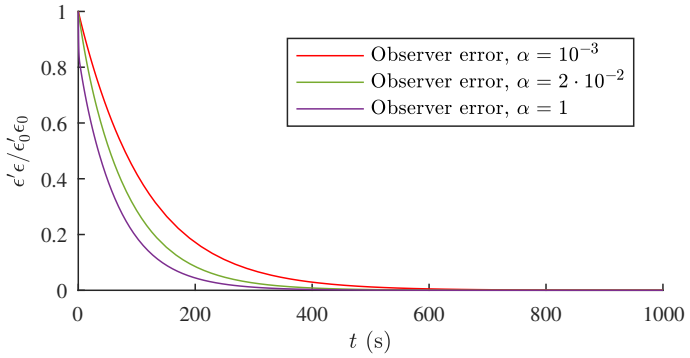


Figure 3.6 – Evolution of the error between the actual state of the heat exchanger and the observer for different values of α.

definite when $u^\star + \bar{u} > 0$ according to the Gershgorin circle theorem. The pair $(C, A(0))$ is observable, and *a fortiori* detectable, if and only if $u^\star \neq \frac{k^2}{\gamma_1 \gamma_2}$. We fix the following parameters, that satisfy all the previous assumptions.

k	γ_1	γ_2	E
$1.20 \cdot 10^{-2}$ s^{-1}	$5.06 \cdot 10^{-1}$ kg^{-1}	$1.00 \cdot 10^{-2}$ s^{-1}	360 K
G	u_M	$u^\star$	β
300 K	0.05 kg $\cdot$ s^{-1}	$0.5u_M$	1

Table 3.7 – Numerical values for the simulation of the heat exchanger.

Set $\hat{x}(0) = x^\star$, and let $x(0)$ be the steady state that corresponds to the constant input $u \equiv 0.17u_M$. Then Theorem 3.5 build a dynamic output feedback based on λ and a Luenberger observer. In Figure 3.5, we plot the evolution of the output x_4 (that we intend to stabilize as in [Zit+20]) for the state feedback law λ and for the dynamic output feedback based on the observer for $\alpha = 10^{-3}$, $\alpha = 2 \cdot 10^{-2}$ and $\alpha = 1$. The error between the state and the observer is given in Figure 3.6 for the same values of α. As in Section 3.3.1, the convergence of the observer to the state of the system is faster when α is larger, and then the stabilization of the state with dynamic output feedback gets closer to the one obtained by state feedback.

3.4 Proof of asymptotic stability

In this section, we suppose that the assumptions of Theorem 3.4 are satisfied. Let us first define the dynamic gain α. For all $(\hat{x}, y) \in \mathbb{R}^n \times \mathbb{R}^m$, let $k(\hat{x}, y) = -\alpha(\hat{x}, y)P^{-1}C'y$. Let f be the vector field defined over $\mathbb{R}^n$ by $f : \mathbb{R}^n \ni x \mapsto A(\lambda(x))x + B(\lambda(x))$. Since λ is a stabilizing state feedback law, $0 \in \mathbb{R}^n$ is a locally asymptotically stable equilibrium point of f, with basin of attraction $\mathcal{D}(\lambda)$. According to the converse Lyapunov theorem (see e.g. [TP00]), there exists a proper function $W \in C^\infty(\mathcal{D}(\lambda), \mathbb{R}_+)$ such that $W(0) = 0$ and

$$\frac{\partial W}{\partial x}(x) f(x) \leqslant -W(x), \quad \forall x \in \mathcal{D}(\lambda). \tag{3.8}$$

For all $r > 0$, set $D(r) = \{x \in \mathbb{R}^n \ : \ V(x) \leqslant r\}$ which is a compact subset of $\mathcal{D}(\lambda)$. Let $\alpha : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ be the function defined for all $(\hat{x}, y) \in \mathbb{R}^n \times \mathbb{R}^m$ by

$$\alpha(\hat{x}, y) = -\frac{\max\{W(\hat{x}), 1\}}{2\left(1 + \left|\frac{\partial W}{\partial x}(\hat{x})\right|\right)(1 + |P^{-1}C'y|)}. \tag{3.9}$$

Note that α is locally Lipschitz and $\alpha(\hat{x}, y) > 0$ for all $(\hat{x}, y) \in \mathbb{R}^n \times \mathbb{R}^m$. Also, it yields

$$|k(\hat{x}, y)| \leqslant \frac{\max\{W(\hat{x}), 1\}}{2\left(1 + \left|\frac{\partial W}{\partial x}(\hat{x})\right|\right)}, \quad \forall (\hat{x}, y) \in \mathbb{R}^n \times \mathbb{R}^m. \tag{3.10}$$

The feedback law λ being also locally Lipschitz, the Cauchy-Lipschitz theorem guarantees the existence of a unique maximal solution to the Cauchy problem associated to (3.6) with initial conditions $(\hat{x}_0, \varepsilon_0)$ in $\mathcal{D}(\lambda) \times \mathbb{R}^n$.

The proof of Theorem 3.4 relies on the three following lemmas.

3.4.1 Local asymptotic stability

Lemma 3.14. *System* (3.6) *is locally asymptotically stable at* $(0, 0)$.

Proof. Let $A_0 = A(0)$ and $\alpha_0 = \alpha(0, 0) > 0$. Consider the linearization of (3.6) at the origin:

$$\begin{cases} \dot{\hat{x}} = A_0\hat{x} - \alpha_0 P^{-1}C'C\varepsilon \\ \dot{\varepsilon} = \left(A_0 - \alpha_0 P^{-1}C'C\right)\varepsilon. \end{cases} \tag{3.11}$$

This system is upper triangular. Let us first focus on the ε part of the system. Consider the function $V : \varepsilon \mapsto \varepsilon' P \varepsilon$. Then V is a Lyapunov function for the ε-subsystem. Indeed,

$$\begin{aligned} \frac{\mathrm{d}V(\varepsilon)}{\mathrm{d}t} &= \varepsilon' P\dot{\varepsilon} + \dot{\varepsilon}' P\varepsilon \\ &= \varepsilon'\left(PA_0 + A_0'P\right)\varepsilon - 2\alpha_0\varepsilon' C'C\varepsilon \\ &\leqslant -2\alpha_0|C\varepsilon|^2 \qquad \text{(by (3.2))} \\ &\leqslant 0. \end{aligned}$$

We denote by $\Omega(\varepsilon_0)$ the ω-limit set of the the ε-subsystem with initial condition $\varepsilon_0 \in \mathbb{R}^n$. Then, by LaSalle's invariance principle, $\Omega(\varepsilon_0) \subset \{\varepsilon_0 \in \mathbb{R}^n \ : \ C\varepsilon \equiv 0\}$. Since the pair (C, A_0) is detectable by 0-detectability (see Remark 3.7), we have $\varepsilon \to 0$. Since the system is linear, this implies that all eigenvalues of $A_0 - \alpha_0 P^{-1}C'C$ have negative real part. Now let us consider the $\hat{x}$-subsystem. Since 0 is asymptotically stable for the vector field f, all the eigenvalues of A_0 have non-positive real part. Moreover, $\{\varepsilon_0 \in \mathbb{R}^n \ : \ C\varepsilon \equiv 0\}$ is invariant under the dynamics of the $\hat{x}$-subsystem. Then, applying the center manifold theorem (see *e.g.* [GK01, Appendix, Theorem 4.2]), (3.6) is locally asymptotically stable at 0. ■

3.4.2 All trajectories are bounded

Lemma 3.15. *All the trajectories of* (3.6) *with initial conditions in* $\mathcal{D}(\lambda) \times \mathbb{R}^n$ *remain in a compact subset of* $\mathcal{D}(\lambda) \times \mathbb{R}^n$. *In particular, solutions are complete in positive time.*

Proof. Consider the function $V : \varepsilon \mapsto \varepsilon' P \varepsilon$. For all initial conditions $(\hat{x}_0, \varepsilon_0) \in \mathcal{D}(\lambda) \times \mathbb{R}^n$, any corresponding solution of the closed-loop system (3.6) denoted by $(\hat{x}(\cdot), \varepsilon(\cdot))$ satisfies

$$\begin{aligned}\frac{\mathrm{d}V(\varepsilon)}{\mathrm{d}t} &= \varepsilon' P \dot{\varepsilon} + \dot{\varepsilon}' P \varepsilon \\ &= \varepsilon' \left(P A(\lambda(\hat{x})) + A(\lambda(\hat{x}))' P \right) \varepsilon - 2\alpha \varepsilon' C' C \varepsilon \\ &\leqslant -2\alpha |C\varepsilon|^2 \qquad \text{(by (3.2))} \\ &\leqslant 0.\end{aligned}$$

Hence, ε remains in a compact set. For all $r > 0$, set $D(r) = \{x \in \mathbb{R}^n \ : \ W(x) < r\} \subset \mathcal{D}$, where W is the Lyapunov function defined in (3.8). Moreover, for all $(\hat{x}, y) \in \mathbb{R}^n \times \mathbb{R}^p$,

$$\begin{aligned}\frac{\partial W}{\partial x}(\hat{x})[f(\hat{x}, \lambda(\hat{x})) + k(\hat{x}, y)] &\leqslant -W(\hat{x}) + \frac{\partial W}{\partial x}(\hat{x}) k(\hat{x}, y) \\ &\leqslant -W(\hat{x}) + \left| \frac{\partial W}{\partial x}(\hat{x}) \right| |k(\hat{x}, y)| \\ &\leqslant -W(\hat{x}) + \left| \frac{\partial W}{\partial x}(\hat{x}) \right| \frac{\max\{W(\hat{x}), 1\}}{2\left(1 + \left|\frac{\partial W}{\partial x}(\hat{x})\right|\right)} \\ &\leqslant -W(\hat{x}) + \frac{1}{2} \max\{W(\hat{x}), 1\}.\end{aligned}$$

Hence, if $\hat{x} \in \mathcal{D}(\lambda) \setminus D(1)$,

$$\frac{\partial W}{\partial x}(\hat{x})(f(\hat{x}, \lambda(\hat{x})) + k(\hat{x}, y)) \leqslant -\frac{1}{2} W(\hat{x}). \tag{3.12}$$

Thus

$$W(\hat{x}) \leqslant \max\{W(\hat{x}_0), 1\},$$

In other words, $\hat{x}$ remains in $D(1) \cup D(W(\hat{x}_0))$ which is a compact subset of $\mathcal{D}(\lambda)$. Thus, solutions of (3.6) are complete in positive time. ■

3.4.3 All trajectories converge to 0.

Lemma 3.16. *All the trajectories of* (3.6) *with initial conditions in* $\mathcal{D}(\lambda) \times \mathbb{R}^n$ *tends towards 0 as time goes to infinity.*

Proof. For all $(\hat{x}_0, \varepsilon_0) \in \mathcal{D}(\lambda) \times \mathbb{R}^n$, let $t \mapsto (\hat{X}(t, \hat{x}_0, \varepsilon_0), E(t, \hat{x}_0, \varepsilon_0))$ be the semi-trajectory of (3.6) with initial conditions $(\hat{x}_0, \varepsilon_0)$. Fix $(\hat{x}_0, \varepsilon_0) \in \mathcal{D}(\lambda) \times \mathbb{R}^n$. Let $(\hat{x}, \varepsilon)$ be the semi-trajectory of (3.6) starting from $(\hat{x}_0, \varepsilon_0)$, and $\Omega(\hat{x}_0, \varepsilon_0)$ the ω-limit set of this semi-trajectory. According to Lemma 3.15, $(\hat{x}, \varepsilon)$ is bounded. We prove that $(\hat{x}, \varepsilon)$ converges to $(0, 0)$ as a consequence of Lemma 3.14, by proving that the semi-trajectory enters the basin of attraction of $(0, 0)$ in finite time. It is sufficient to prove that $(0, 0) \in \Omega(\hat{x}_0, \varepsilon_0)$ since this implies that $(\hat{x}, \varepsilon)$ enters any open set containing $(0, 0)$ in finite time. We prove this in three steps: first $\Omega(\hat{x}_0, \varepsilon_0) \subset \{(\hat{x}_1, \varepsilon_1) \in \mathcal{D}(\lambda) \times \mathbb{R}^n \ : \ CE(\cdot, \hat{x}_1, \varepsilon_1) \equiv 0\}$, then $\Omega(\hat{x}_0, \varepsilon_0) \cap (\{0\} \times \mathbb{R}^n) \neq \varnothing$ and finally $(0, 0) \in \Omega(\hat{x}_0, \varepsilon_0)$. Recall that $\frac{\mathrm{d}V(\varepsilon)}{\mathrm{d}t} \leqslant -2\alpha |C\varepsilon|^2$ by (3.2). Then, according to LaSalle's invariance principle, $\Omega(\hat{x}_0, \varepsilon_0) \subset \{(\hat{x}_1, \varepsilon_1) \in \mathcal{D}(\lambda) \times \mathbb{R}^n \ : \ CE(\cdot, \hat{x}_1, \varepsilon_1) \equiv 0\}$.

Let $(\hat{x}_1, \varepsilon_1) \in \Omega(\hat{x}_0, \varepsilon_0)$. The set $\Omega(\hat{x}_0, \varepsilon_0)$ is compact and invariant under the dynamics of the system, hence $\hat{X}(t, \hat{x}_1, \varepsilon_1) \in \Omega(\hat{x}_0, \varepsilon_0)$ for all $t \geqslant 0$. This further implies that $\Omega(\hat{x}_1, \varepsilon_1)$ is a non-empty compact subset of $\Omega(\hat{x}_0, \varepsilon_0)$. Since λ is a stabilizing state feedback, $\hat{X}(t, \hat{x}_1, \varepsilon_1) \to 0$ as $t \to +\infty$. Hence $\Omega(\hat{x}_1, \varepsilon_1) \subset \{0\} \times \mathbb{R}^n$ and thus

$$\Omega(\hat{x}_0, \varepsilon_0) \cap (\{0\} \times \mathbb{R}^n) \neq \varnothing.$$

Then there exists $\varepsilon_2 \in \mathbb{R}^n$ such that $(0, \varepsilon_2) \in \Omega(\hat{x}_0, \varepsilon_0) \subset \{(\hat{x}_1, \varepsilon_1) \in \mathcal{D}(\lambda) \times \mathbb{R}^n \ : \ CE(\cdot, \hat{x}_1, \varepsilon_1) \equiv 0\}$. Hence $\hat{X}(\cdot, 0, \varepsilon_2) \equiv 0$. Then $E(\cdot, 0, \varepsilon_2)$ is solution of

$$\dot{\varepsilon} = A_0 \varepsilon, \qquad C\varepsilon = 0. \tag{3.13}$$

Since the pair (C, A_0) is detectable (by 0-detectability), $E(\cdot, 0, \varepsilon_2) \to 0$. Hence $\{(0,0)\} = \Omega(0, \varepsilon_2) \subset \Omega(\hat{x}_0, \varepsilon_0)$. By local asymptotic stability of $(0,0)$, it follows that the semi-trajectory $(\hat{x}, \varepsilon)$ converges towards 0. ■

Proof of Theorem 3.4. Combining stability from Lemma 3.14 and semi-global convergence towards $(0,0)$ from Lemma 3.16, we get the result. ■

Proof of Theorem 3.5. The proof is very similar to the global version of the result. In particular, Lemmas 3.14 and 3.16 remain unchanged. However, the proof of Lemma 3.15 must be adapted.

Let $K_{\hat{x}} \times K_\varepsilon \subset \mathcal{D} \times \mathbb{R}^n$ be a compact set. Let $R = \mu_{\max} \sup_{K_\varepsilon} V < +\infty$, where $\mu_{\max}$ denotes the largest eigenvalue of P. Let W be the Lyapunov function defined in (3.8). For all $r > 0$, set $D(r) = \{x \in \mathbb{R}^n \ : \ W(x) < r\} \subset \mathcal{D}$ and denote by $\partial D(r)$ its boundary. Let $\rho > 0$ be such that $K_{\hat{x}} \subset D(\rho)$ and the closure of $D(\rho)$ lies in $\mathcal{D}$. Set $M_1 = \sup_{\partial D(\rho)} L_f W < 0$ and $M_2 = 1 + \sup_{\partial D(\rho)} |\nabla W| < +\infty$ where L_f denotes the usual Lie derivative along f and ∇ stands for the Euclidean gradient.

Let $\alpha_0 = \frac{-M_1}{RM_2|P^{-1}||C|^2} > 0$ and take $\alpha \in (0, \alpha_0)$. Take $(\hat{x}_0, \varepsilon_0) \in K_{\hat{x}} \times K_\varepsilon$ and denote $(\hat{x}, \varepsilon)$ the semi-trajectory of (3.6) starting from $(\hat{x}_0, \varepsilon_0)$. Since $V : \varepsilon \mapsto \varepsilon' P \varepsilon$ is a Lyapunov function for the ε-subsystem of (3.6), we have $|\varepsilon| \leqslant R$. Assume there exists $t > 0$ such that $W(\hat{x}(t)) = \rho$. Then

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} W(\hat{x}(t)) &= L_f W(\hat{x}(t)) - \alpha \left(\nabla W(\hat{x}(t))\right)' P^{-1} C' C \varepsilon(t) \\ &\leqslant M_1 + \alpha M_2 |P^{-1}||C|^2 R \\ &< 0. \end{aligned}$$

Hence $\hat{x}(t) \in D(\rho)$ for all $t \geqslant 0$. Thus, for all $\alpha \in (0, \alpha_0)$, all the trajectories of (3.6) with initial conditions in $K_{\hat{x}} \times K_\varepsilon$ remain in a compact subset of $\mathcal{D} \times \mathbb{R}^n$.

■

Chapter 4

Unobservable target

To strive, to seek, to find, and not to yield.

Lord A. Tennyson, *Ulysses*

Abstract. *We address the problem of dynamic output feedback stabilization at an unobservable target point. The challenge lies in according the antagonistic nature of the objective and the properties of the system: the system tends to be less observable as it approaches the target. We illustrate two main ideas: well chosen perturbations of a state feedback law can yield new observability properties of the closed-loop system, and embedding systems into bilinear systems admitting observers with dissipative error systems allows to mitigate the observability issues. We apply them on a case of systems with linear dynamics and nonlinear observation map and make use of an* ad hoc *finite-dimensional embedding. More generally, we introduce a new strategy based on infinite-dimensional unitary embeddings. To do so, we extend the usual definition of dynamic output feedback stabilization in order to allow infinite-dimensional observers fed by the output. Infinite-dimensional Luenberger observers, studied in Chapter 5, are used. We show how this technique, based on representation theory, may be applied to achieve output feedback stabilization at an unobservable target.*

Contents

Figures and Tables

Introduction

Stabilizing a system at an unobservable target is a challenging issue occurring in practical engineering systems, where original strategies have been explored in recent years [HPR14, Com+16, Fla19, Sur+19, Aja+20, RD20, Sur+20, AGS21], leading to a renewal of interest in this problem. The challenge lies in according the antagonistic nature of the state estimation and stabilization: while the system approaches the target, observability properties vanish, hence the state estimation is getting worse, which in turn prevents stabilization.

General methods, based on time-varying feedback laws have been developed to deal with singular inputs. Let us mention the seminal article [Cor94a] by J.-M. Coron, in which local stabilization is achieved by means of a periodic time-varying feedback, up to a Lie null-observability condition. A "sample-and-hold" strategy was developed in [ST03] for the practical semi-global stabilization as well. Furthermore, perturbations of the input, such as high-frequency excitation [Com+16, Sur+19, Sur+20], stochastic noise [Fla19] appear to be a key tool to enhance the observability properties of the system. In [LSG17], the authors introduce an explicit feedback law perturbation for a specific bilinear system. This idea guided us in Chapter 2 to find generic perturbations in the case of systems that are observable at the target. This strategy makes the closed-loop system autonomous, which is of interest for engineering applications. In this chapter, we use such autonomous perturbations to obtain observability properties. Contrarily to Chapter 2, perturbations are explicitly designed.

Another important tool in stabilization theory is the use of systems with non-expanding flows, as the dissipative systems of Chapter 3 and weakly contractive systems in Appendix D. Indeed, in [LSG17], the strategy of feedback perturbation is used in conjunction with the contraction property of a quantum control system to achieve stabilization at an unobservable target. The crucial feature is that the error of the observer system is non-increasing, no matter the observability properties. Hence, the state estimation is not getting worse as the state approaches the target.

In the present chapter, we coalesce the insights provided by Chapters 2 and 3 in order to come up with solutions to attack the problem of dynamic output feedback stabilization at an unobservable target. Essentially, we consider systems with linear conservative dynamics and nonlinear observation maps. A simple example by J.-M. Coron in [Cor94a] highlights how some systems are not stabilizable by means of a finite-dimensional autonomous dynamic output feedback, even locally (Section 4.1.1). Similar examples, sharing the same unobservability issues, may nonetheless be stabilized, as we illustrate in Section 4.1.2.

In order to access the properties of dissipative systems, we look into embedding techniques. In [Cel+89], the authors propose an observer design strategy based on infinite-dimensional unitary embeddings. We rely on their approach in the context of output feedback stabilization, which leads to a coupling of the finite-dimensional original system with an infinite-dimensional dissipative observer system (Section 4.2.1). Interestingly, adding an infinite-dimensional virtual state fed by the output to the original system lifts the topological obstructions identified in [Cor94a]. We illustrate this strategy in Section 4.2.2 by focusing on examples with linear dynamics and nonlinear observation map, that we hope may pave the way to more general results in the future.

Section 4.1 is devoted to the study of an illustrative example in which the specific form of the output allows us to find a finite-dimensional embedding of the system into a bilinear system, and to design a Luenberger observer with dissipative error system. In Section 4.2, we show how tools from representation theory may help to embed a system into a bilinear unitary infinite-dimensional one and stabilize a larger class of systems by means of dynamic output feedback. In both cases, the stabilizing state feedback law is modified with a perturbation that vanishes at the target point.

4.1 An illustrative example

4.1.1 An obstruction by J.-M. Coron

Consider the case where (1.1) is single-input single-output and f is a linear map, so that it can be written in the form of

$$\begin{cases} \dot{x} = Ax + bu, \\ y = h(x). \end{cases} \tag{4.1}$$

where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^{n\times 1}$ and $h : \mathbb{R}^n \to \mathbb{R}$. If h is nonlinear and is not an invertible transformation of a linear map, then the usual theory of linear systems fails to be applied. Condition 1.10 reduces to the stabilizability of the pair (A, b). If it holds, then (4.1) is globally stabilizable by a linear static state feedback.

In [Cor94a], J.-M. Coron introduced the following illustrative one-dimensional example:

$$\dot{x} = u, \qquad y = x^2. \tag{4.2}$$

He proved that (4.2) is not locally stabilizable by means of a dynamic output feedback, unless introducing a time-dependent component in the feedback law. The difficulty with this system comes from the unobservability of the target point 0. Indeed, (4.2) is not observable for the constant input $u \equiv 0$ in any time $T > 0$ since the initial conditions x_0, $-x_0 \in \mathbb{R}$ are indistinguishable. In particular, the system is not uniformly observable, and consequently the results of [TP94, JG95, TP95] fail to be applicable. To overcome this issue, [Cor94a] introduced time-dependent output feedback laws, and proved by this means the local stabilizability of (4.2). This system can also be stabilized by means of "dead-beat" or "sample-and-hold" techniques (see [NS98], [ST03], respectively).

A generalization of (4.2) in higher dimension is

$$\begin{cases} \dot{x} = Ax + bu, \\ y = h(x) \end{cases} \tag{4.3}$$

for a skew-symmetric matrix A and h radially symmetric[1]. Again, the constant input $u \equiv 0$ makes the system unobservable in any time $T > 0$ since for any initial conditions x_0, $\tilde{x}_0$ in $\mathbb{R}^n$ satisfying $|x_0| = |\tilde{x}_0|$, $h(\varphi_t(x_0)) = h(\tilde{x}_0) = h(x_0) = h(\varphi_t(x_0))$ for all $t \in \mathbb{R}_+$. Condition 1.10 (global) reduces to the stabilizability of (A, b) and Condition 1.11 (global) is always satisfied. Let us state a necessary condition for the stabilizability of (4.3) by means of a dynamic output feedback.

[1]Up to a change of scalar product, one may also consider the case where $PA + A'P = 0$ for some positive definite matrix $P \in \mathbb{R}^{n\times n}$ and h such that $(x_1'Px_1 = x_2'Px_2) \Rightarrow (h(x_1) = h(x_2))$.

Theorem 4.1. *If* (4.3) *is locally stabilizable by means of a dynamic output feedback, then A is invertible.*

Proof. The proof is an adaptation of the one given in [Cor94a] in the one-dimensional context. Assume that $(0,0)$ is a locally asymptotically stable equilibrium point of

$$\begin{cases} \dot{x} = Ax + bu, \\ y = h(x) \end{cases}, \qquad \begin{cases} \dot{w} = \nu(w,u,y) \\ u = \varpi(w,y) \end{cases} \tag{4.4}$$

for some positive integer q and two continuous maps $\nu : \mathbb{R}^q \times \mathbb{R} \times \mathbb{R}$ and $\varpi : \mathbb{R}^q \times \mathbb{R}$. Set $F : \mathbb{R}^n \times \mathbb{R}^q \ni (x,w) \mapsto (Ax + b\varpi(w,h(x)), \nu(w,\varpi(w,h(x)),h(x)))$. Then, according to [KZ84, Theorem 52.1] (see [Cor94b] when one does not have uniqueness of the solutions to the Cauchy problem), the index of $-F$ at $(0,0)$ is 1. Assume, for the sake of contradiction, that A is not invertible. Let $\mathcal{N}$ be a one-dimensional subspace of $\ker A$. Denote by Σ the reflection through the hyperplane $\mathcal{N}^\perp$, that is, $\Sigma = \mathrm{Id}_n - 2vv'$ for some unitary vector $v \in \mathcal{N}$. Then $\det \Sigma = -1$, $A\Sigma = A$ and $h(\Sigma x) = h(x)$. Hence $(x,w) \mapsto -F(\Sigma x, w)$ has index -1 at $(0,0)$ and $F(\Sigma x, w) = F(x,w)$. Thus $1 = -1$ which is a contradiction. ∎

According to the spectral theorem, we have the following immediate corollary. If $n = 1$, we recover the result of J.-M. Coron in [Cor94a].

Corollary 4.2. *If n is odd and A is skew-symmetric, then* (4.3) *is not locally stabilizable by means of a dynamic output feedback.*

4.1.2 Converse theorem: a positive result

One of the main results of Section 4.1 is the following theorem which is the converse of Theorem 4.1 in the case where $h(x) = \frac{1}{2}|x|^2$. The proof relies on the guidelines described in introduction, that is, an embedding into a bilinear system, an observer design with dissipative error-system and a feedback perturbation.

Consider the special case for system (4.3):

$$\begin{cases} \dot{x} = Ax + bu, \\ y = h(x) = \dfrac{1}{2}|x|^2. \end{cases} \tag{4.3'}$$

Theorem 4.3. *If A is skew-symmetric and invertible and (A,b) is stabilizable, then* (4.3') *is semi-globally stabilizable by means of a dynamic output feedback.*

Remark 4.4. The dynamic output feedback is explicitly given in (4.12). It is easily implementable, and does not use time-dependent feedback laws.

The proof of Theorem 4.3 is the object of the section. We follow the same steps as in [AGS21], with a very similar embedding strategy. The main difference is the observability analysis developed in Section 4.1.2: here the target is unobservable, while in [AGS21] it was observable.

Embedding into a bilinear system of higher dimension

Consider the map

$$\begin{aligned} \tau : \mathbb{R}^n &\longrightarrow \mathbb{R}^{n+1} \\ x &\longmapsto \left(x, \tfrac{1}{2}|x|^2\right). \end{aligned} \tag{4.5}$$

For all $z = (z_1, \ldots, z_{n+1}) \in \mathbb{R}^{n+1}$, define $\bar{z}_n = (z_1, \ldots, z_n) \in \mathbb{R}^n$. If x is a solution of (4.3), then $\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|x|^2 = x'Ax + x'bu = x'bu$ since A is skew-symmetric. Hence $z = \tau(x)$ defines an embedding of (4.3) into

$$\begin{cases} \dot{z} = \mathcal{A}(u)z + \mathcal{B}u \\ y = \mathcal{C}z. \end{cases} \tag{4.6}$$

where $\mathcal{A}(u) = \begin{pmatrix} A & 0 \\ ub' & 0 \end{pmatrix}$, $\mathcal{B} = \begin{pmatrix} b \\ 0 \end{pmatrix}$ and $\mathcal{C} = \begin{pmatrix} 0 \cdots 0 & 1 \end{pmatrix}$ and with initial conditions in $\mathcal{T} = \tau(\mathbb{R}^n)$. Moreover, the semi-trajectory z remains in $\mathcal{T}$.

Observer design with dissipative error system

Let us introduce a Luenberger observer with dynamic gain for (4.6). In order to make the error system dissipative, set $\mathcal{L}_\alpha(u) = \begin{pmatrix} bu \\ \alpha \end{pmatrix} \in \mathbb{R}^{n+1}$ for some positive constant α to be fixed later. The corresponding observer system is given by

$$\begin{cases} \dot{\varepsilon} = (\mathcal{A}(u) - \mathcal{L}_\alpha(u)\mathcal{C})\,\varepsilon \\ \dot{\hat{z}} = \mathcal{A}(u)\hat{z} + \mathcal{B}u - \mathcal{L}_\alpha(u)\mathcal{C}\varepsilon \end{cases} \tag{4.7}$$

where $z = \hat{z} - \varepsilon$ satisfies (4.6), $\hat{z}$ is the estimation of the state made by the observer system and ε is the error between the estimation of the state and the actual state of the system. Note that for all $u \in \mathbb{R}$,

$$\mathcal{A}(u) - \mathcal{L}_\alpha(u)\mathcal{C} = \begin{pmatrix} A & -bu \\ ub' & -\alpha \end{pmatrix} = \begin{pmatrix} A & -bu \\ ub' & 0 \end{pmatrix} - \alpha\mathcal{C}'\mathcal{C}. \tag{4.8}$$

This implies that the ε-subsystem of (4.7) is dissipative, that is, for all inputs $u \in C^0(\mathbb{R}_+, \mathbb{R})$, the solutions of (4.7) satisfy

$$\frac{\mathrm{d}|\varepsilon|^2}{\mathrm{d}t} = 2\varepsilon'\left(\mathcal{A}(u) - \mathcal{L}_\alpha(u)\mathcal{C}\right)\varepsilon = -2\alpha|\mathcal{C}\varepsilon|^2 \leqslant 0. \tag{4.9}$$

This is the first key fact of the strategy applied below.

Feedback perturbation and closed-loop system

Because (A, b) is stabilizable, there exists $K \in \mathbb{R}^{1\times n}$ such that $A + bK$ is Hurwitz (in particular, (K, A) is detectable). Since A is skew-symmetric, its eigenvalues are purely imaginary. Hence, the spectral conditions of Hautus lemmas for stabilizability (resp. detectability) and controllability (resp. observability) of the pair (A, b) (resp. (K, A)) are equivalent[2]. Therefore, (A, b) is controllable and (K, A) is observable.

[2]Recall that the Hautus lemmas state that (A, b) is stabilizable (resp. controllable) if and only if $\mathrm{rank}(\mu\mathrm{Id}_{\mathbb{R}^n} - A, b) = n$ for all eigenvalues $\mu \in \mathbb{C}$ of A (resp. for all eigenvalues $\mu \in \mathbb{C}$ of A for which $\Re\mu \geqslant 0$). Since A is skew-symmetric, all its eigenvalues have non-negative real part (actually, they are purely imaginary). Hence conditions for stabilizability and controllability are clearly equivalent. A similar result holds for detectability and observability.

With a separation principle in mind, a natural strategy for dynamic output feedback stabilization of (4.3) would be to combine the Luenberger observer (4.7) with the state feedback law $\phi : x \mapsto Kx$. However, it appears that this strategy fails to be applied due to the unobservability at the target. To overcome this difficulty, we rather consider a perturbed feedback law $\phi_\delta : x \mapsto Kx + \frac{\delta}{2}|x|^2$ for some positive constant δ to be fixed later. This is the second key fact of the strategy. For all $\delta > 0$, denote by $\mathcal{D}_\delta$ the basin of attraction of $0 \in \mathbb{R}^n$ of the vector field $\mathbb{R}^n \ni x \mapsto Ax + b\phi_\delta(x)$. Since the linearization of this vector field at 0 is $x \mapsto (A + bK)x$, it is locally asymptotically stable at 0 for all $\delta > 0$. As stated in the following lemma, the drawback of this perturbation is to pass from a globally stabilizing state feedback to a semi-globally stabilizing one.

Lemma 4.5. *For any compact set $\mathcal{K}_x \subset \mathbb{R}^n$, there exists $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$, $\mathcal{K}_x \subset \mathcal{D}_\delta$.*

Proof. Let $\rho > 0$ be such that $\mathcal{K}_x \subset B_{\mathbb{R}^n}(0, \rho)$. Since $A + bK$ is Hurwitz, there exists $P \in \mathbb{R}^{n\times n}$ positive definite such that $P(A+bK) + (A+bK)'P < -2\mathrm{Id}_{\mathbb{R}^n}$. Set $V : \mathbb{R}^n \ni x \mapsto x'Px$. Then, for all $x \in \mathcal{K}_x$,

$$\begin{aligned}\frac{\partial V}{\partial x}(x)(Ax + b\phi_\delta(x)) &= 2x'P(A + bK)x + \delta|x|^2x'Pb \\ &\leqslant (-2 + \delta|x||Pb|)|x|^2 \\ &\leqslant (-2 + \delta\rho|Pb|)|x|^2.\end{aligned}$$

Set $\delta_0 = \frac{1}{\rho|Pb|}$ and let $\delta \in (0, \delta_0)$. Then V is positive definite and

$$\frac{\partial V}{\partial x}(x)(Ax + b\phi_\delta(x)) < -|x|^2$$

for all $x \in \mathcal{K}_x$. Hence, $0 \in \mathbb{R}^n$ is a locally asymptotically (even exponentially) stable equilibrium point of the vector field $\mathbb{R}^n \ni x \mapsto Ax + b\phi_\delta(x)$ with basin of attraction containing $\mathcal{K}_x$. ■

Hence, for all compact sets $\mathcal{K}_x \subset \mathbb{R}^n$ there exists $\delta_0 > 0$ such that if $\delta \in (0, \delta_0)$, then $\mathcal{K}_x \subset \mathcal{D}_\delta$. For system (4.6), we choose the feedback law

$$\lambda_\delta(z) = \begin{pmatrix} K & \delta \end{pmatrix} z, \tag{4.10}$$

which satisfies $\phi_\delta = \lambda_\delta \circ \tau$. The corresponding closed-loop system is given by

$$\begin{cases} \dot{\varepsilon} = (\mathcal{A}(\lambda_\delta(\hat{z})) - \mathcal{L}_\alpha(\lambda_\delta(\hat{z}))\mathcal{C})\,\varepsilon, \\ \dot{\hat{z}} = \mathcal{A}(\lambda_\delta(\hat{z}))\hat{z} + \mathcal{B}\lambda_\delta(\hat{z}) - \mathcal{L}_\alpha(\lambda_\delta(\hat{z}))\mathcal{C}\varepsilon. \end{cases} \tag{4.11}$$

By using $w = \hat{z}$ as the new dynamical system fed by the output y, we are now able to exhibit the coupled system that solves the semi-global dynamic output feedback stabilization problem of (4.3):

$$\begin{cases} \dot{x} = Ax + bu, \\ y = \dfrac{1}{2}|x|^2 \end{cases}, \qquad \begin{cases} \dot{\hat{z}} = \mathcal{A}(u)\hat{z} + \mathcal{B}u - \mathcal{L}_\alpha(u)\,(\mathcal{C}\hat{z} - y) \\ u = \lambda_\delta(\hat{z}). \end{cases} \tag{4.12}$$

It is now sufficient to prove the following theorem, which implies Theorem 4.3, in the next sections.

Theorem 4.6. *For any compact set $\mathcal{K}_x \times \mathcal{K}_w \subset \mathbb{R}^n \times \mathbb{R}^{n+1}$, there exist $\delta_0 > 0$ and $\alpha_0 > 0$ such that for all $\delta \in (0, \delta_0)$ and all $\alpha \in (\alpha_0, +\infty)$, $(0,0) \in \mathbb{R}^n \times \mathbb{R}^{n+1}$ is a locally asymptotically stable equilibrium point of (4.12) with basin of attraction containing $\mathcal{K}_x \times \mathcal{K}_w$.*

Boundedness of trajectories

Since $\mathbb{R}^n \ni x \mapsto \frac{1}{2}|x|^2$ and ϕ_δ are locally Lipschitz continuous functions, according to the Cauchy-Lipschitz theorem, for any initial condition $(x_0, \hat{z}_0) \in \mathbb{R}^n \times \mathbb{R}^{n+1}$, there exists exactly one maximal solution $(x, \hat{z})$ of (4.12) such that $(x(0), \hat{z}(0)) = (x_0, \hat{z}_0)$. Before going into the proof of Theorem 4.6, we need to ensure the existence of global solutions.

Lemma 4.7. *For any compact set $\mathcal{K}_x \times \mathcal{K}_w \subset \mathbb{R}^n \times \mathbb{R}^{n+1}$, there exist $\delta_0 > 0$ and $\alpha_0 > 0$ such that for all $\delta \in (0, \delta_0)$ and all $\alpha \in (\alpha_0, +\infty)$, (4.12) has a unique global solution $(x, \hat{z})$ for each initial condition $(x_0, \hat{z}_0) \in \mathcal{K}_x \times \mathcal{K}_w$. Moreover, $(x, \hat{z})$ is bounded and $\bar{\hat{z}}_n$ remains in a compact subset of $\mathcal{D}_\delta$.*

Proof. Let $(x_0, \hat{z}_0) \in \mathcal{K}_x \times \mathcal{K}_w$ and $(x, \hat{z})$ be the corresponding maximal solution of (4.12). Set $z = \tau(x)$ and $\varepsilon = \hat{z} - z$, so that $(\varepsilon, \hat{z})$ is the maximal solution of (4.11) starting from $(\varepsilon_0, \hat{z}_0)$. Then, it is sufficient to prove that $(\varepsilon, \hat{z})$ is a global solution, $(\varepsilon, \hat{z})$ is bounded and $\bar{\hat{z}}_n$ remains in a compact subset of $\mathcal{D}_\delta$. According to (4.9), ε is bounded since $|\varepsilon|$ is non-increasing. Moreover, $\hat{z}_{n+1} = \varepsilon_{n+1} + \frac{1}{2}|\bar{z}_n|^2 = \varepsilon_{n+1} + \frac{1}{2}|\bar{\hat{z}}_n - \bar{\varepsilon}_n|^2$. Then, it remains to show that there exist $\delta_0 > 0$ and $\alpha_0 > 0$ such that for all $\delta \in (0, \delta_0)$ and all $\alpha \in (\alpha_0, +\infty)$, for all initial conditions $(\varepsilon_0, \hat{z}_0) \in (\mathcal{K}_w - \tau(\mathcal{K}_x)) \times \mathcal{K}_w$, $\bar{\hat{z}}_n$ remains in a compact subset of $\mathcal{D}_\delta$.

Since $A + bK$ is Hurwitz, there exists $P \in \mathbb{R}^{n\times n}$ positive definite such that $P(A+bK) + (A+bK)'P < -2\mathrm{Id}_{\mathbb{R}^n}$. Then $V : \mathbb{R}^n \ni x \mapsto x'Px$ is a strict Lyapunov function for system (4.3) with feedback law ϕ. For all $r > 0$, set $D(r) = \{x \in \mathbb{R}^n \ : \ V(x) \leqslant r\}$. Let $\rho' > \rho > 0$ and $r' > r > 0$ be such that $B_{\mathbb{R}^{n+1}}(0, \rho)$ contains $(\mathcal{K}_w - \tau(\mathcal{K}_x))$ and $\mathcal{K}_w$ and $B_{\mathbb{R}^{n+1}}(0, \rho) \subset D(r) \subset D(r') \subset B_{\mathbb{R}^{n+1}}(0, \rho')$. According to Lemma 4.5, there exists $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$, $\mathcal{D}_\delta$ contains the closure of $B_{\mathbb{R}^n}(0, \rho')$. In the following, we show that there exists $\alpha_0 > 0$ such that, if $\alpha > \alpha_0$, then $\bar{\hat{z}}_n$ remains in $B_{\mathbb{R}^n}(0, \rho')$. For all $\hat{z}$, ε in $\mathbb{R}^{n+1}$, define

$$\begin{aligned}
\mu_\delta^1(\hat{z}) &= \mathcal{A}(\phi_\delta(\bar{\hat{z}}_n))\hat{z} + \mathcal{B}\phi_\delta(\bar{\hat{z}}_n),\\
\mu_\delta^2(\hat{z}) &= (\mathcal{A}(\lambda_\delta(\hat{z})) - \mathcal{A}(\phi_\delta(\bar{\hat{z}}_n)))\hat{z} + \mathcal{B}(\lambda_\delta(\hat{z}) - \phi_\delta(\bar{\hat{z}}_n)),\\
\mu_{\delta,\alpha}^3(\varepsilon, \hat{z}) &= -\mathcal{L}_\alpha(\lambda_\delta(\hat{z}))\mathcal{C}\varepsilon,
\end{aligned}$$

so that the solutions of (4.11) satisfy

$$\dot{\hat{z}} = \mu_\delta^1(\hat{z}) + \mu_\delta^2(\hat{z}) + \mu_{\delta,\alpha}^3(\varepsilon, \hat{z}). \tag{4.13}$$

In particular,

$$\dot{\bar{\hat{z}}}_n = A\bar{\hat{z}}_n + \lambda_\delta(\hat{z})b - \lambda_\delta(\hat{z})\varepsilon_{n+1}b.$$

By continuity of $(\hat{z}, \delta) \mapsto \lambda_\delta(\hat{z})$,

$$\overline{M} := \sup_{\substack{\varepsilon, \hat{z} \in B_{\mathbb{R}^{n+1}}(0,\rho') \\ \delta \in [0, \delta_0]}} |A\bar{\hat{z}}_n + \lambda_\delta(\hat{z})b - \lambda_\delta(\hat{z})\varepsilon_{n+1}b| < \infty.$$

Let $T_0 = \frac{\rho'-\rho}{\overline{M}}$. Since $|\varepsilon|$ is non-increasing, any trajectory of (4.11) starting in $B_{\mathbb{R}^{n+1}}(0,\rho) \times B_{\mathbb{R}^{n+1}}(0,\rho)$ will be such that $\bar{\hat{z}}_n$ remains in $B_{\mathbb{R}^n}(0,\rho')$ over the time interval $[0,T_0]$. It remains to show that $\bar{\hat{z}}_n$ does not exit $B_{\mathbb{R}^n}(0,\rho')$ after time T_0.

The projection operator on the first n coordinates $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$, *i.e.*, $\pi(\hat{z}) = \bar{\hat{z}}_n$, is a left-inverse of τ:

$$\pi(\tau(x)) = x, \qquad \forall x \in \mathbb{R}^n. \tag{4.14}$$

Note that $\mu^1_\delta(\hat{z}_1) = \mu^1_\delta(\hat{z}_2)$ if $\pi(\hat{z}_1) = \pi(\hat{z}_2)$. Then,

$$\underline{m} := - \max_{\substack{\bar{\hat{z}}_n \in \partial D(r') \\ \hat{z} \in B_{\mathbb{R}^{n+1}}(0,\rho')}} \left(L_{\mu^1_0} V \circ \pi\right)(\hat{z}) = - \max_{\substack{\pi(\hat{z}) \in \partial D(r') \\ \hat{z} \in B_{\mathbb{R}^{n+1}}(0,\rho')}} \frac{\partial V}{\partial x}(\pi(\hat{z}))\,(A+bK)\pi(\hat{z}) > 0.$$

Notice that $(\mu^1_\delta - \mu^1_0 + \mu^2_\delta)(\hat{z}) = \delta \hat{z}_{n+1} \begin{pmatrix} b \\ b' \bar{\hat{z}}_n \end{pmatrix}$. Hence, without loss of generality, one can assume that $\delta_0 > 0$ is (small enough) such that for all $\delta \in (0, \delta_0)$,

$$\max_{B_{\mathbb{R}^{n+1}}(0,\rho')} |L_{\mu^1_\delta - \mu^1_0 + \mu^2_\delta} V \circ \pi| \leqslant \frac{1}{3}\underline{m}.$$

Fix $\delta \in (0, \delta_0)$. Assume for the sake of contradiction that $\bar{\hat{z}}_n$ leaves $D(r')$ for the first time at $T_0' > T_0$. Then $\frac{\mathrm{d}}{\mathrm{d}t}_{|t=T_0'} V(\pi(\hat{z})) \geqslant 0$. We have

$$\begin{aligned}
0 &\leqslant \frac{\mathrm{d}}{\mathrm{d}t} V\left(\pi(\hat{z}(t))\right)\Big|_{t=T_0'} \\
&= (L_{\mu^1_0} V \circ \pi)(\hat{z}(T_0')) + (L_{\mu^1_\delta - \mu^1_0 + \mu^2_\delta} V \circ \pi)(\hat{z}(T_0')) + \frac{\partial V \circ \pi}{\partial \hat{z}}(\hat{z}(T_0'))\mu^3_{\delta,\alpha}(\varepsilon(T_0'), \hat{z}(T_0')) \\
&\leqslant -\frac{2}{3}\underline{m} + \frac{\partial V \circ \pi}{\partial \hat{z}}(\hat{z}(T_0'))\mu^3_{\delta,\alpha}(\varepsilon(T_0'), \hat{z}(T_0'))
\end{aligned}$$

Now, we show that there exists $\alpha_0 > 0$ large enough such that for all $\alpha > \alpha_0$,

$$\frac{\partial V \circ \pi}{\partial \hat{z}}(\hat{z}(T_0'))\mu^3_{\delta,\alpha}(\varepsilon(T_0'), \hat{z}(T_0')) \leqslant \frac{1}{3}\underline{m}, \tag{4.15}$$

which contradicts $\underline{m} > 0$. By definition of $\mathcal{L}_\alpha$, π and $\mu^3_{\delta,\alpha}$,

$$\frac{\partial V \circ \pi}{\partial \hat{z}}(\hat{z})\mu^3_{\delta,\alpha}(\varepsilon, \hat{z}) = -\varepsilon_{n+1}\lambda_\delta(\hat{z})\frac{\partial V}{\partial x}(\pi(\hat{z}))b.$$

Let $Q = \max_{\substack{(\hat{z}_2,\hat{z}_3) \in \partial B_{\mathbb{R}^n}(0,\rho') \\ \varepsilon, \hat{z} \in B_{\mathbb{R}^{n+1}}(0,\rho')}} |\lambda_\delta(\hat{z})\frac{\partial V}{\partial x}(\pi(\hat{z}))b|$, so that $|\lambda_\delta(\hat{z}(T_0'))\frac{\partial V}{\partial x}(\pi(\hat{z}(T_0')))b| \leqslant Q$. Recall that

$$\dot{\varepsilon}_{n+1} = -\alpha\varepsilon_{n+1} + \lambda_\delta(\hat{z})b'\bar{\varepsilon}_n$$

and thus, for all $t \geqslant 0$,

$$\varepsilon_{n+1}(t) = e^{-\alpha t}\varepsilon_{n+1}(0) + \int_0^t e^{-\alpha(t-s)}\lambda_\delta(\hat{z}(s))b'\bar{\varepsilon}_n(s)\mathrm{d}s.$$

Moreover, $\varepsilon(t)$ and $\bar{\hat{z}}_n(t)$ are in $B_{\mathbb{R}^{n+1}}(0,\rho')$ for all $t \in [0, T_0']$ and

$$\lambda_\delta(\hat{z}) = \begin{pmatrix} K & \delta \end{pmatrix}\hat{z} = K\bar{\hat{z}}_n + \delta\left(\varepsilon_{n+1} + \frac{1}{2}|\bar{\hat{z}}_n - \bar{\varepsilon}_n|^2\right).$$

Hence,

$$|\lambda_\delta(\hat{z})| \leqslant \rho' \left(|K| + \delta(1 + 2\rho') \right).$$

As a consequence, for all $t \in [0, T_0']$,

$$|\varepsilon_{n+1}(t)| \leqslant \rho' \left(e^{-\alpha t} + \frac{\rho'^2 |b|}{\alpha} \left(|K| + \delta(1 + 2\rho') \right) \right).$$

Thus there exists $\alpha_0 > 0$ such that if $\alpha > \alpha_0$, then $|\varepsilon_{n+1}(T_0')| \leqslant \dfrac{m}{3Q}$. Fix $\alpha > \alpha_0$. Then (4.15) holds, which concludes the proof of the lemma. ■

Observability analysis

The following lemma is a crucial step of the proof of Theorem 4.3 that emphasizes the usefulness of the feedback perturbation described above. Indeed, one can easily see that its proof fails if $\delta = 0$ (since the matrix $\mathcal{Q}$ defined below is not invertible in this case).

Lemma 4.8. *Let $(z_0, \hat{z}_0) \in (\mathcal{T} \times \mathbb{R}^{n+1}) \setminus \{(0,0)\}$. Let $(\varepsilon, \hat{z})$ be the semi-trajectory of (4.7) with initial condition $(\hat{z}_0 - z_0, \hat{z}_0)$. Then, for all $T > 0$, (4.6) is observable in time T for the input $u = \lambda_\delta(\hat{z})$.*

Proof. Let $\omega_0 \in \ker(\mathcal{C}) \setminus \{0\}$, and consider ω a solution of the dynamical system

$$\dot{\omega} = \mathcal{A}(\lambda_\delta(\hat{z}))\omega \tag{4.16}$$

with initial condition ω_0. To prove the result, it is sufficient to show that $\mathcal{C}\omega$ has a non-zero derivative of some order at $t = 0$ if $(\varepsilon_0, \hat{z}_0) \neq (0,0)$. Indeed, it implies that for all initial conditions $z_0 \neq \tilde{z}_0$ in $\mathbb{R}^{n+1}$, if z (resp. $\tilde{z}$) is the solution of (4.6) with initial condition z_0 (resp. $\tilde{z}_0$), then $\omega = z - \tilde{z}$ is a solution to (4.16) starting at $\omega_0 \neq 0$ and $\mathcal{C}\omega$ is not constantly equal to zero on any time interval $[0, T] \subset \mathbb{R}_+$. We prove this fact by contradiction: assume that

$$\mathcal{C}\omega^{(k)}(0) = \omega_{n+1}^{(k)}(0) = 0 \qquad \forall k \in \mathbb{N}, \tag{4.17}$$

for some $\omega(0) \neq 0$, and prove that $(z_0, \hat{z}_0) = (0,0)$. Let $u = \lambda_\delta(\hat{z})$. Then $\dot{\omega}_{n+1} = ub'\bar{\omega}_n$ and $\dot{\bar{\omega}}_n = A\bar{\omega}_n$. Hence

$$0 = \omega_{n+1}^{(k+1)}(0) = \sum_{i=0}^{k} \binom{k}{i} u^{(i)}(0) b' A^{k-i} \bar{\omega}_n(0) \tag{4.18}$$

for all $k \in \mathbb{N}$, where $\binom{k}{i}$ denote binomial coefficients. The proof goes through the following three steps.

Step 1: Show that $u^{(k)}(0) = 0$ for all $k \in \mathbb{N}$. Let $p \in \mathbb{N}$ be the smallest integer such that $u^{(p)}(0) \neq 0$ and look for a contradiction. Equation (4.18) yields

$$\sum_{i=0}^{k} \binom{p+k}{p+i} u^{(p+i)}(0) b' A^{k-i} \bar{\omega}_n(0) = 0 \tag{4.19}$$

for all $k \in \mathbb{N}$. Since (A, b) is controllable and $\bar{\omega}_n(0) \neq 0$, there exists $q \in \{0, \dots, n\}$ such that $b'A^q\bar{\omega}_n(0) \neq 0$ and $b'A^i\bar{\omega}_n(0) = 0$ for all $i \in \{0, \dots, q-1\}$. Then

$$0 = \sum_{i=0}^{q} \binom{p+q}{p+i} u^{(p+i)}(0) b'A^{q-i}\bar{\omega}_n(0) = \binom{p+q}{p} u^{(p)}(0) b'A^q\bar{\omega}_n(0). \tag{4.20}$$

which is a contradiction.

Step 2: Find $\mathcal{Q} \in \mathbb{R}^{(n+2)\times(n+2)}$ (invertible) such that $\mathcal{Q}\begin{pmatrix} \hat{\boldsymbol{z}}(\mathbf{0}) \\ \boldsymbol{\varepsilon_{n+1}(0)} \end{pmatrix} = \mathbf{0}$.
For all $k \in \mathbb{N}$,

$$0 = u^{(k)}(0) = \begin{pmatrix} K & \delta \end{pmatrix} \hat{z}^{(k)}(0).$$

Moreover,

$$\begin{pmatrix} \dot{\bar{\hat{z}}}_n \\ \dot{\hat{z}}_{n+1} \\ \dot{\varepsilon}_{n+1} \end{pmatrix} = \begin{pmatrix} A & -bu & 0 \\ b'u & 0 & -\alpha \\ 0 & 0 & -\alpha \end{pmatrix} \begin{pmatrix} \bar{\hat{z}}_n \\ \hat{z}_{n+1} \\ \varepsilon_{n+1} \end{pmatrix} + u \begin{pmatrix} b \\ 0 \\ b'\bar{\varepsilon}_n \end{pmatrix}.$$

Hence, for all $k \geqslant 1$, $\bar{\hat{z}}_n^{(k)}(0) = A^k\bar{\hat{z}}_n(0)$ and $\hat{z}_{n+1}^{(k)}(0) = \varepsilon_{n+1}^{(k)}(0) = (-\alpha)^k\varepsilon_{n+1}(0)$. Thus $\begin{pmatrix} KA^k & \delta(-\alpha)^k \end{pmatrix} \begin{pmatrix} \bar{\hat{z}}_n(0) \\ \varepsilon_{n+1}(0) \end{pmatrix} = 0$ for all $k \geqslant 1$. By setting

$$\mathcal{Q} = \begin{pmatrix} K & \delta & 0 \\ KA & 0 & -\delta\alpha \\ \vdots & \vdots & \vdots \\ KA^{n+1} & 0 & \delta(-\alpha)^{n+1} \end{pmatrix} \tag{4.21}$$

we get that $\mathcal{Q}\begin{pmatrix} \hat{z}(0) \\ \varepsilon_{n+1}(0) \end{pmatrix} = 0$.

Step 3: Conclusion. In Appendix A.2, we check the following lemma.

Lemma 4.9. *The matrix $\mathcal{Q}$ is invertible.*

Hence, $\hat{z}(0) = 0$ and $\varepsilon_{n+1}(0) = 0$. Thus, $\frac{1}{2}\left|\bar{z}_n(0)\right|^2 = z_{n+1}(0) = \hat{z}_{n+1}(0) - \varepsilon_{n+1}(0) = 0$ *i.e.* $(z_0, \hat{z}_0) = (0, 0)$ which is a contradiction. ■

On the basis of Lemmas 4.7 and 4.8, we are now in a position to prove Theorem 4.6. Let $\mathcal{K}_x \times \mathcal{K}_w \subset \mathbb{R}^n \times \mathbb{R}^{n+1}$ be a compact set, and $\delta_0 > 0$ and $\alpha_0 > 0$ be as in Lemma 4.7. Fix $\delta \in (0, \delta_0)$ and $\alpha \in (\alpha_0, +\infty)$. Let $(x_0, \hat{z}_0) \in \mathcal{K}_x \times \mathcal{K}_w$ and $(x, \hat{z})$ be the corresponding solution of (4.12). Set $z = \boldsymbol{\tau}(x)$, $\varepsilon = \hat{z} - z$ so that $(\varepsilon, \hat{z})$ is the solution of (4.11) starting from $(\varepsilon_0, \hat{z}_0)$, $\varepsilon_0 = \hat{z}_0 - \boldsymbol{\tau}(x_0)$. We need to show the two following statements:

1. (Stability) $(0, 0)$ is a stable equilibrium point of (4.12),

2. (Attractivity) and its basin of attraction contains $\mathcal{K}_x \times \mathcal{K}_w$.

We prove the former in Section 4.1.2 and the latter in Section 4.1.2.

Stability

Let $R > 0$. We seek $r > 0$ such that, if $|x_0|, |\hat{z}_0| \leqslant r$, then $|x(t)|, |\hat{z}(t)| \leqslant R$ for all $t \in \mathbb{R}_+$. We have

$$\begin{aligned}\dot{x} &= Ax + b\lambda_\delta(\hat{z}) \\ &= Ax + b\lambda_\delta(\tau(x) + \varepsilon) \\ &= Ax + b\phi_\delta(x) + b\begin{pmatrix} K & \delta \end{pmatrix}\varepsilon.\end{aligned}$$

Fix $\eta > 0$ such that $R - \eta\sqrt{1 + \frac{\eta^2}{2}} > 0$. Since $x \mapsto Ax + b\phi_\delta(x)$ is locally asymptotically stable, there exists a positive constant $r_\varepsilon \leqslant R - \eta\sqrt{1 + \frac{\eta^2}{2}}$ such that, if $|\varepsilon(t)| \leqslant r_\varepsilon$ for all $t \in \mathbb{R}_+$, then $|x(t)| \leqslant \eta$ for all $t \in \mathbb{R}_+$. Let $r > 0$ be such that $r + r\sqrt{1 + \frac{r^2}{2}} \leqslant r_\varepsilon$. Assume that $|x_0|, |\hat{z}_0| \leqslant r$. Then,

$$|\varepsilon_0| \leqslant |\hat{z}_0| + |\tau(x_0)| = |\hat{z}_0| + |x_0|\sqrt{1 + \frac{|x_0|^2}{2}} \leqslant r + r\sqrt{1 + \frac{r^2}{2}} \leqslant r_\varepsilon.$$

According to (4.9), $|\varepsilon|$ is non-increasing. Hence, for all $t \in \mathbb{R}_+$, $|x(t)| \leqslant \eta \leqslant R$ and

$$|\hat{z}(t)| \leqslant |\tau(x(t))| + |\varepsilon(t)| \leqslant \eta\sqrt{1 + \frac{\eta^2}{2}} + r_\varepsilon \leqslant R.$$

Attractivity

Recall that $(\varepsilon, \hat{z})$ is a solution of (4.11). According to (4.9), $\frac{\mathrm{d}|\varepsilon|^2}{\mathrm{d}t} = -2\alpha|\mathcal{C}\varepsilon|^2$. Hence, according to LaSalle's invariance principle, the ω-limit set of $(\varepsilon, \hat{z})$ is the largest subset of $(\ker \mathcal{C}) \times \mathbb{R}^n$ that is forward invariant under the dynamics of (4.11). If $(\varepsilon^\star, \hat{z}^\star)$ is a solution of (4.11) remaining in $(\ker \mathcal{C}) \times \mathbb{R}^n$, then $\varepsilon^\star$ is a solution of (4.16). Hence Lemma 4.8 guarantees that either $\varepsilon^\star \equiv 0$, or $(\varepsilon^\star(0), \hat{z}^\star(0)) = (0, 0)$, which also implies $\varepsilon^\star \equiv 0$. Therefore, the ω-limit set of $(\varepsilon, \hat{z})$ is a subset of $\{0\} \times \mathbb{R}^n$. Since $(\varepsilon, \hat{z})$ is bounded, $\varepsilon \to 0$.

Since $\hat{z}_{n+1} = \varepsilon_{n+1} + \frac{1}{2}|\bar{\hat{z}}_n - \bar{\varepsilon}_n|^2$, it remains to prove that $\bar{\hat{z}}_n \to 0$. First, notice that

$$|\mu_\delta^2(\hat{z})| = |\lambda_\delta(\hat{z}) - \phi_\delta(\hat{z})|\sqrt{|b|^2 + |b'\bar{\hat{z}}_n|^2}$$

and

$$|\mu_{\delta,\alpha}^3(\varepsilon, \hat{z})| = \sqrt{\alpha^2 + |b|^2\lambda_\delta(\hat{z})^2}|\mathcal{C}\varepsilon|.$$

Since $C\varepsilon \to 0$ and $\hat{z}$ is bounded, $|\mu_{\delta,\alpha}^3(\varepsilon, \hat{z})| \to 0$. Likewise,

$$\begin{aligned}\lambda_\delta(\hat{z}) - \phi_\delta(\hat{z}) &= \delta\left(\hat{z}_{n+1} - \frac{1}{2}|\bar{\hat{z}}_n|^2\right) \\ &= \delta\left(\varepsilon_{n+1} + z_{n+1} - \frac{1}{2}|\bar{\varepsilon}_n|^2 - \frac{1}{2}|\bar{z}_n|^2 + \bar{\varepsilon}_n'\bar{z}_n\right) \\ &= \delta\left(\varepsilon_{n+1} - \frac{1}{2}|\bar{\varepsilon}_n|^2 + \bar{\varepsilon}_n'\bar{z}_n\right).\end{aligned}$$

Since $\varepsilon \to 0$ and z is bounded, $\mu_\delta^2(\hat{z}) \to 0$.

According to the converse Lyapunov theorem [TP00, Theorem 1], there exists a strict proper Lyapunov function V_δ for system (4.3) with feedback law $\phi_\delta : x \mapsto$

$Kx + \frac{\delta}{2}|x|^2$ over the basin of attraction $\mathcal{D}_\delta$. For all $r > 0$, set $D(r) = \{x \in \mathcal{D}_\delta \ : V_\delta(x) \leqslant r\}$. In order to prove that $\bar{\hat{z}}_n \to 0$, we show that for all $r > 0$, there exists $T(r) \geqslant 0$ such that $\bar{\hat{z}}_n(t) \in D(r)$ for all $t \geqslant T(r)$. According to Lemma 4.7, there exists a compact set $\mathcal{K} \subset \mathcal{D}_\delta$ such that $\bar{\hat{z}}_n \in \mathcal{K}$. If $r > 0$ is such that $\mathcal{K} \subset D(r)$ then $T(r) = 0$ satisfies the statement. Let $0 < r < R$ be such that $\mathcal{K} \not\subset D(r)$ and $\mathcal{K} \subset D(R)$, then

$$\bar{m} := - \max_{D(R)\setminus D(r)} L_{\phi_\delta} V_\delta > 0.$$

Since $|\mu_\delta^2(\hat{z}(t))| \to 0$ and $|\mu_{\delta,\alpha}^3(\varepsilon(t), \hat{z}(t))| \to 0$, there exists $T_1(r) > 0$ such that for all $t \geqslant T_1(r)$, if $\bar{\hat{z}}_n(t) \notin D(r)$, then

$$\frac{\mathrm{d}}{\mathrm{d}t} V_\delta(\bar{\hat{z}}_n) < -\frac{\bar{m}}{2}.$$

First, this implies that if $\pi(\hat{z}(t)) \in D(r)$ for some $t \geqslant T_1(r)$, then $\pi(\hat{z}(s)) \in D(r)$ for all $s \geqslant t$. Second, for all $t \geqslant 0$,

$$\begin{aligned} V_\delta(\bar{\hat{z}}_n(T_1(r) + t)) &= V_\delta(\bar{\hat{z}}_n(T_1(r))) + \int_0^t \frac{\mathrm{d}}{\mathrm{d}t} V_\delta(\bar{\hat{z}}_n(T_1(r) + s))\mathrm{d}s \\ &\leqslant R - \frac{\bar{m}}{2} t \qquad \text{while } \bar{\hat{z}}_n(T_1(r) + t) \notin D(r). \end{aligned}$$

Set $T_2(r) = \frac{2R-r}{\bar{m}}$ and $T(r) = T_1(r) + T_2(r)$. Then for all $t \geqslant T(r)$, $\bar{\hat{z}}_n(t) \in D(r)$, which concludes the proof of convergence, and therefore the proof of Theorem 4.6.

4.1.3 Numerical simulations

In this section, we illustrate the output feedback stabilization given by (4.12) with numerical simulations. System parameters are given in Table 4.1. With the help of the Matlab® ode45 function (based on a Runge-Kutta formula), we plot the obtained result on Figure 4.2. For the initial condition given in Table 4.1, the trajectory $(x, \hat{z})$ seems to converge exponentially to the target point $(0, 0) \in \mathbb{R}^2 \times \mathbb{R}^3$. However, note that (4.12) cannot be exponentially stable at $(0, 0)$. Indeed, its linearization at the target point is given by

$$\begin{pmatrix} \dot{x} \\ \dot{\hat{z}} \end{pmatrix} = \begin{pmatrix} A & bK & b\delta \\ 0 & A + bK & b\delta \\ 0 & 0 & -\alpha \end{pmatrix} \begin{pmatrix} x \\ \hat{z} \end{pmatrix} \tag{4.22}$$

which is not exponentially stable since A has two purely imaginary eigenvalues.

$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$	$b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$K = \begin{pmatrix} 0 & -2 \end{pmatrix}$	$\alpha = 1$	$\delta = 1$	$x_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\hat{z}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

Table 4.1 – Parameters of the numerical simulation of system (4.12).

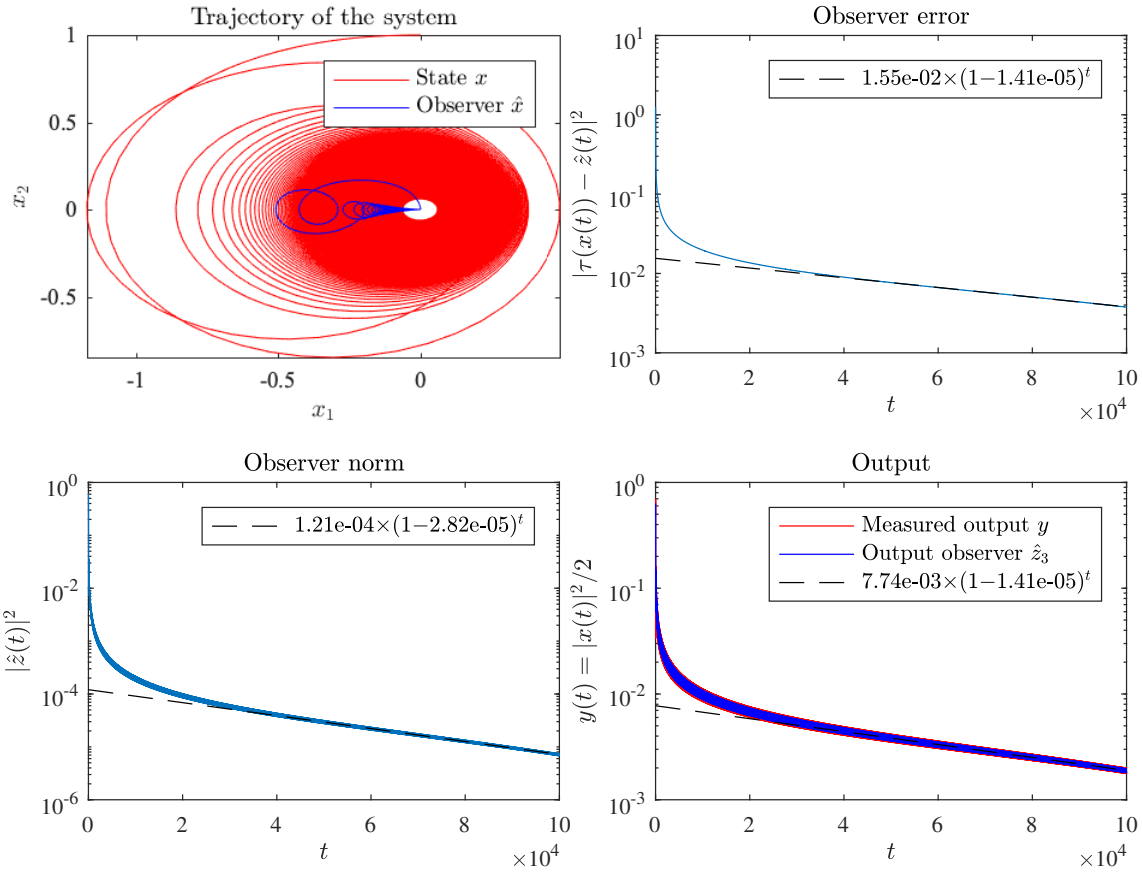


Figure 4.2 – Numerical simulation of system (4.12) with parameters given in Table 4.1. The map τ is the embedding defined by (4.5) *(Top left)* The trajectory $(x, \hat{x})$ of the system seems to converge to the origin. *(Top right)* The observer error is non-increasing, as stated by (4.9), but the convergence to 0 is slow. A linear regression indicates an exponential rate about -1.41×10^{-5}. *(Bottom left)* The observer tends towards the origin with exponential rate about -2.82×10^{-5}. At $t = 10^5$, the observer squared norm is about 10^{-5}. *(Bottom right)* The output, which is half the square of the state norm, has the same convergence rate as the observer. At $t = 10^5$, the state squared norm is about 10^{-3}.

4.2 An infinite-dimensional perspective

Guided by the illustrative example of Section 4.1, we aim to provide more general results, based on the same two principles: embedding into a dissipative system, and feedback perturbation. The embedding strategy used in Section 4.1.2 appears to be specific to this example, and hardly generalizable, since it relies mostly on the form of the observation map. A different strategy must be found. In [Cel+89], the authors introduce a technique for the synthesis of observers for nonlinear systems. The method is based on representation theory, and embedding into bilinear unitary systems. It is far more general than the embedding found in Section 4.1.2. The price to pay is that the observer system can be infinite-dimensional. In this section, we show how to use this strategy in the context of dynamic output feedback stabilization. After exhibiting some general results when such an embedding exists, we investigate

a case of systems with linear conservative dynamics and nonlinear observation maps. These results link the theory of infinite-dimensional linear time-varying observers of Part II with the output feedback stabilization issue of finite-dimensional systems studied in Part I.

4.2.1 Embedding into infinite-dimensional unitary systems

Embedding into unitary systems and observer design

Let $(X, \|\cdot\|_X)$ be a Hilbert space and $\mathcal{D}$ be a dense subspace of X. For all $u \in \mathbb{R}^p$, let $\mathcal{A}(u) : \mathcal{D} \to X$ be the skew-adjoint generator of a strongly continuous unitary group on X and $\mathcal{C} \in \mathscr{L}(X, \mathbb{C}^{\mathfrak{m}})$ for some positive integer $\mathfrak{m}$. Let $u \in C^1(\mathbb{R}_+, \mathbb{R}^p)$ and $z_0 \in X$. Consider the non-autonomous linear abstract Cauchy problem with measured output

$$\begin{cases} \dot{z} = \mathcal{A}(u(t))z \\ z(0) = z_0 \end{cases} \qquad \mathfrak{y} = \mathcal{C}z. \tag{4.23}$$

According to [Paz83, Chapter 5, Theorem 4.8], the family $(\mathcal{A}(u(t)))_{t\in\mathbb{R}_+}$ is the generator of a unique evolution system on X that we denote by $(\mathbb{T}_t(\cdot, u))_{t\in\mathbb{R}_+}$. For any $z_0 \in X$, (4.23) admits a unique solution $z \in C^0(\mathbb{R}_+, X)$ given by $z(t) = \mathbb{T}_t(z_0, u)$ for all $t \in \mathbb{R}_+$. Moreover, if $z_0 \in \mathcal{D}$, then $z \in C^0(\mathbb{R}_+, \mathcal{D}) \cap C^1(\mathbb{R}_+, X)$. The reader may refer to [Paz83, Chapter 5], [EN00, Chapter VI.9] or [IK02] for more details on the evolution equations theory.

For such systems, a Luenberger observer with constant gain $\alpha > 0$ can be built as follows:

$$\begin{cases} \dot{\hat{z}} = \mathcal{A}(u(t))\hat{z} - \alpha\mathcal{C}^*(\mathcal{C}\hat{z} - \mathfrak{y}) \\ \hat{z}(0) = \hat{z}_0 \in X. \end{cases} \tag{4.24}$$

The study of this observer is the main topic of Chapter 5, to which the reader may refer to find sufficient conditions of convergence of $\hat{z}$ to z. Here we simply recall some well-posedness results. Set $\varepsilon = \hat{z} - z$ and $\varepsilon_0 = \hat{z}_0 - z_0$. From now on, $\hat{z}$ represents the state estimation made by the observer system and ε the error between this estimation and the actual state of the system. Then $\hat{z}$ satisfies (4.24) if and only if ε satisfies

$$\begin{cases} \dot{\varepsilon} = (\mathcal{A}(u(t)) - \alpha\mathcal{C}^*\mathcal{C})\varepsilon \\ \varepsilon(0) = \varepsilon_0. \end{cases} \tag{4.25}$$

Since $\mathcal{C} \in \mathscr{L}(X, \mathbb{C}^{\mathfrak{m}})$, [Paz83, Chapter 5, Theorem 2.3] claims that $(\mathcal{A}(u(t)) - \alpha\mathcal{C}^*\mathcal{C})_{t\geqslant 0}$ is also a stable family of generators of strongly continuous semigroups, and generates an evolution system on X denoted by $(\mathbb{S}_t(\cdot, u))_{t\in\mathbb{R}_+}$. Then, systems (4.24) and (4.25) have respectively a unique solution $\hat{z}$ and ε in $C^0(\mathbb{R}_+, X)$. Moreover, $\hat{z}(t) = \mathbb{T}_t(z_0, u) + \mathbb{S}_t(\varepsilon_0, u)$ and $\varepsilon(t) = \mathbb{S}_t(\varepsilon_0, u)$ for all $t \in \mathbb{R}_+$. If $(\hat{z}_0, \varepsilon_0) \in \mathcal{D}^2$, then $\hat{z}$, $\varepsilon \in C^0(\mathbb{R}_+, \mathcal{D}) \cap C^1(\mathbb{R}_+, X)$.

This infinite-dimensional Luenberger observer is investigated in Chapter 5 and [Cel+89], in which it is proved that $\varepsilon(t) \overset{w}{\to} 0$ as t goes to infinity if u is a *regularly persistent input*. Our goal is to embed the original system (1.1) into a unitary system, and to use this observer design in the context of dynamic output feedback stabilization.

Definition 4.10 (Embedding)**.** An injective map $\tau : \mathbb{R}^n \mapsto X$ is said to be an embedding of (1.1) into the unitary system (4.23) if there exists $\mathfrak{h} : \mathbb{R}^m \to \mathbb{C}^{\mathfrak{m}}$ such that the following diagram is commutative for all $t \in \mathbb{R}_+$ and all $u \in C^1(\mathbb{R}_+, \mathbb{R}^p)$:

$$\begin{array}{ccccccc} \mathbb{R}^n & \xrightarrow{\varphi_t(\cdot,u)} & \mathbb{R}^n & \xrightarrow{h} & \mathbb{R}^m & \xrightarrow{\mathfrak{h}} & \mathbb{C}^{\mathfrak{m}} \\ \tau\downarrow & & \downarrow\tau & & & \nearrow_{\mathcal{C}} & \\ X & \xrightarrow[\mathbb{T}_t(\cdot,u)]{} & X & & & & \end{array} \tag{4.26}$$

i.e., , for all $x_0 \in \mathbb{R}^n$, $\tau(\varphi_t(x_0, u)) = \mathbb{T}_t(\tau(x_0), u)$ and $\mathfrak{h}(h(x_0)) = \mathcal{C}\tau(x_0)$.

Remark 4.11. This definition of embedding does not coincide with the usual notion of embedding in differential topology, even on finite-dimensional spaces. Moreover, in Definition 2.3, we introduced the notion of *immersion* of control systems. Even on finite-dimensional spaces, these two notions do not coincide. Indeed, commutativity of the diagram

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\varphi_t(\cdot,u)} & \mathbb{R}^n \\ \tau\downarrow & & \downarrow\tau \\ X & \xrightarrow[\mathbb{T}_t(\cdot,u)]{} & X \end{array} \tag{4.27}$$

is not required in Definition 2.3.

Here, the map $\mathfrak{h}$ is a degree of freedom that may be chosen to find an embedding of (1.1) into (4.23). Let $u \in C^1(\mathbb{R}_+, \mathbb{R}^p)$, $z_0, \varepsilon_0 \in \mathcal{D}$, $z(t) = \mathbb{T}_t(\hat{z}_0, u)$ and $\varepsilon(t) = \mathbb{S}_t(\varepsilon_0, u)$ for all $t \in \mathbb{R}_+$. For all $t \in \mathbb{R}_+$, $\mathcal{A}(u(t))$ is skew-adjoint, hence

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|z(t)\right\|_X^2 = \Re\left\langle \mathcal{A}(u(t))z(t), z(t)\right\rangle_X = 0, \tag{4.28}$$

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\varepsilon(t)\right\|_X^2 = \Re\left\langle \mathcal{A}(u(t))\varepsilon(t), \varepsilon(t)\right\rangle_X - \alpha\Re\left\langle \mathcal{C}^*\mathcal{C}\varepsilon(t), \varepsilon(t)\right\rangle_X = -\alpha\left|\mathcal{C}\varepsilon(t)\right|^2 \leqslant 0. \tag{4.29}$$

Thus $\left\|z\right\|_X$ is constant and $\left\|\varepsilon\right\|_X$ is non-increasing. If there exists a positive constant β such that for all $x \in \mathcal{D}$ and all $u \in \mathbb{R}$,

$$\left\|\mathcal{C}^*\mathcal{C}\mathcal{A}(u)x\right\|_X \leqslant \beta\left\|x\right\|_X, \tag{4.30}$$

then

$$\begin{aligned} \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|C\varepsilon(t)\right\|_Y^2 &= \left\langle \mathcal{C}\varepsilon(t), \mathcal{C}\dot{\varepsilon}(t)\right\rangle_{\mathbb{C}^{\mathfrak{m}}} \\ &= \left\langle \mathcal{C}\varepsilon(t), \mathcal{C}\mathcal{A}(u(t))\varepsilon(t)\right\rangle_{\mathbb{C}^{\mathfrak{m}}} - \alpha\left\langle \mathcal{C}\varepsilon(t), \mathcal{C}\mathcal{C}^*\mathcal{C}\varepsilon(t)\right\rangle_{\mathbb{C}^{\mathfrak{m}}} \\ &= \left\langle \varepsilon(t), \mathcal{C}^*\mathcal{C}\mathcal{A}(u(t))\varepsilon(t)\right\rangle_X - \alpha\left\|\mathcal{C}^*\mathcal{C}\varepsilon(t)\right\|_X^2 \\ &\leqslant \beta\left\|\varepsilon(0)\right\|_X^2 \end{aligned}$$

since $\left\|\varepsilon\right\|_X$ is non-increasing. Thus, $\left|C\varepsilon\right|^2$ is an integrable positive function, with bounded derivative. Hence, according to Barbalat's lemma, $\mathcal{C}\varepsilon(t) \to 0$ as $t \to +\infty$. Condition (4.30) is also discussed in Remark 5.43. Inequality (4.29) is similar to (4.9), and will be a key argument to achieve the dynamic output feedback stabilization.

Embedding inversion: from the embedded system's weak observer to the original system's observer

Let us recall the characterization of the strong and weak topologies on X. A sequence $(x_n)_{n\geqslant 0} \in X^{\mathbb{N}}$ is said to be strongly convergent to some $x^\star \in X$ if $\|x_n - x^\star\|_X \to 0$ as $n \to +\infty$, and we shall write $x_n \to x^\star$ as $n \to +\infty$. It is said to be weakly convergent to $x^\star$ if $\langle x_n - x^\star, \psi\rangle_X \to 0$ as $n \to +\infty$ for all $\psi \in X$, and we shall write $x_n \overset{w}{\to} x^\star$ as $n \to +\infty$. The strong topology on X is finer than the weak topology (see, *e.g.,* [Bre11] for more properties on these usual topologies).

In Section 4.1, a crucial argument was the existence of a left-inverse π to the embedding τ (see (4.14)). Now, X being infinite-dimensional, we must make the notion of left-inverse precise, and, moreover, the convergence of the observer $\hat{z}$ to the embedded state z will hold only in the weak topology of X, namely, $\varepsilon \overset{w}{\to} 0$.

This is an important issue, which causes difficulties in achieving output feedback stabilization. However, in this section, we show that if the original state x remains bounded, and if the embedding τ is injective and analytic, then $\hat{x} = \pi(\hat{z})$ is actually an observer of x in the usual topology of $\mathbb{R}^n$, namely, $\hat{x} - x \to 0$. This is summarized in Corollary 4.16.

Definition 4.12 (Strong left-inverse)**.** Let $(X, \|\cdot\|_X)$ be a normed vector space, $\mathcal{K}_x \subset \mathbb{R}^n$ and $\tau : \mathbb{R}^n \to X$. A map $\pi : X \to \mathcal{K}_x$ is called a *strong left-inverse* of τ on $\mathcal{K}_x$ if and only if there exists a class $\mathcal{K}_\infty$ function[3] ρ^* and $Q \in \mathscr{L}(X, \mathbb{C}^q)$ for some a positive integer q such that, for all $(x, \xi) \in \mathcal{K}_x \times X$,

$$|\pi(\xi) - x| \leqslant \rho^*(|Q(\xi - \tau(x))|). \tag{4.31}$$

Remark 4.13. If π is a strong left-inverse of τ on $\mathcal{K}_x$, then (4.31) implies that π is also a left-inverse in the usual sense: for all $x \in \mathcal{K}_x$, $\pi(\tau(x)) = x$. In particular, τ is injective over $\mathcal{K}_x$.

The reason for which we look for a strong left-inverse of τ is the following lemma, which follows directly from (4.31) and the fact that $Q \in \mathscr{L}(X, \mathbb{C}^q)$.

Lemma 4.14. *Let $(X, \|\cdot\|_X)$ be a normed vector space, $\mathcal{K}_x \subset \mathbb{R}^n$ and $\tau : \mathbb{R}^n \to X$. Let $\pi : X \to \mathcal{K}_x$ be a strong left-inverse of τ on $\mathcal{K}_x$. Let $(x_n)_{n\in\mathbb{N}}$ and $(\xi_n)_{n\in\mathbb{N}}$ be two sequences in $\mathcal{K}_x$ and X, respectively. If $\xi_n - \tau(x_n) \overset{w}{\to} 0$ as n goes to infinity, then $|\pi(\xi_n) - x_n| \to 0$ as n goes to infinity.*

This justifies the denomination of *strong* left-inverse, in the sense that it allows to pass from weak convergence in the infinite-dimensional space X to (usual) convergence in the finite-dimensional space $\mathbb{R}^n$. The following theorem states sufficient conditions for the existence of a strong left-inverse.

Theorem 4.15. *Let X be a separable Hilbert space, $\tau : \mathbb{R}^n \to X$ be an analytic map and $\mathcal{K}_x \subset \mathbb{R}^n$ be a compact set . If $\tau|_{\mathcal{K}_x}$ is injective, then τ has a continuous strong left-inverse on $\mathcal{K}_x$.*

Proof. Let $(e_k)_{k\in\mathbb{N}}$ be a Hilbert basis of X. For all $i \in \mathbb{N}$, let

$$E_i = \{(x_a, x_b) \in \mathbb{R}^n \times \mathbb{R}^n \ : \ \forall k \in \{0, \dots, i-1\},\ \langle \tau(x_a) - \tau(x_b), e_k\rangle_X = 0\}.$$

[3] A class $\mathcal{K}_\infty$ function is a continuous function $\rho^* : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\rho^*(0) = 0$, ρ^* is strictly increasing and tends to infinity at infinity.

Then $(E_i)_{i\in\mathbb{N}}$ is a non-increasing family of analytic sets. According to [Nar66, Chapter 5, Corollary 1], $(E_i \cap {\mathcal{K}_x}^2)_{i\in\mathbb{N}}$ is stationary, *i.e.*, there exists $q \in \mathbb{N}$ such that $E_q \cap {\mathcal{K}_x}^2 = E_i \cap {\mathcal{K}_x}^2$ for all $i \geqslant q$. Hence,

$$\begin{aligned} E_q \cap {\mathcal{K}_x}^2 &= \bigcap_{k\in\mathbb{N}} E_k \cap {\mathcal{K}_x}^2 \\ &= \{(x_a, x_b) \in {\mathcal{K}_x}^2 \ : \ \tau(x_a) = \tau(x_b)\} \\ &\qquad\qquad \text{(since } (e_k)_{k\in\mathbb{N}} \text{ is a Hilbert basis of } X) \\ &= \{(x_a, x_a) \ : \ x_a \in \mathcal{K}_x\}. \qquad \text{(since } \tau \text{ is injective on } \mathcal{K}_x) \end{aligned}$$

Let $Q : X \ni \xi \mapsto (\langle \xi, e_k\rangle_X)_{k\in\{0,\dots,q-1\}} \in \mathbb{C}^q$ and $\tilde{\tau} = Q \circ \tau$. Then $\tilde{\tau}$ is continuous and injective on $\mathcal{K}_x$. Indeed, for all $(x_a, x_b) \in {\mathcal{K}_x}^2$, if $\tilde{\tau}(x_a) = \tilde{\tau}(x_b)$, then $(x_a, x_b) \in E_q \cap {\mathcal{K}_x}^2$ which yields $x_a = x_b$. Hence, combining [Ber+17, Lemma 6] and [AP06, Theorem 1], there exists a continuous map $\tilde{\pi} : \mathbb{C}^q \to \mathcal{K}_x$ and a class $\mathcal{K}_\infty$ function ρ^* such that for all $(x, \mathfrak{z}) \in \mathcal{K}_x \times \mathbb{C}^q$, $|\tilde{\pi}(\mathfrak{z}) - x| \leqslant \rho^*(|\mathfrak{z} - \tilde{\tau}(x)|)$. Set $\pi = \tilde{\pi} \circ Q$. Then π is continuous and for all $(x, \xi) \in \mathcal{K}_x \times X$,

$$|\pi(\xi) - x| \leqslant \rho^*(|Q(\xi) - \tilde{\tau}(x)|) = \rho^*(|Q(\xi - \tau(x))|).$$

■

Applying Theorem 4.15, then Lemma 4.14, we get the following result in our context.

Corollary 4.16. *Let X be a separable Hilbert space, $\tau : \mathbb{R}^n \to X$ be an analytic embedding of* (1.1) *into the unitary system* (4.23) *and $\mathcal{K}_x$ be a compact subset of $\mathbb{R}^n$. Then τ has a continuous strong left-inverse π on $\mathcal{K}_x$.*

Let $x_0 \in \mathcal{K}_x$, $\hat{z}_0 \in X$ and $u \in C^1(\mathbb{R}_+, \mathbb{R}^p)$. Denote by x and $\hat{z}$ the corresponding solutions of (1.1) *and* (4.24)*, respectively. Set $\hat{x} = \pi(\hat{z})$. Assume that $x(t) \in \mathcal{K}_x$ for all $t \in \mathbb{R}_+$. If $\hat{z} - \tau(x) \xrightarrow{w} 0$, then $\hat{x} - x \to 0$.*

Remark 4.17. Beyond the problem of output feedback stabilization, Corollary 4.16 may be used in the context of observer design. In [Cel+89], after embedding the original finite-dimensional system into an infinite-dimensional unitary system, the authors investigate only the convergence of the infinite-dimensional observer. Corollary 4.16 states that if the infinite-dimensional observer converges and if the original system's state trajectory remains bounded, then an observer can be built for the original system, by using a strong left-inverse of the embedding.

Feedback perturbation and closed-loop system

In order to set up a separation principle to solve the dynamic output feedback stabilization problem of (1.1), let us assume that Condition 1.10 (semi-global) and the following assumption are satisfied.

Assumption 4.18 (Existence of an embedding)**.** System (1.1) admits an analytic embedding into the unitary system (4.23).

Let $\mathcal{K}_x$ be a compact subset of $\mathbb{R}^n$. Denote by ϕ a locally asymptotically stabilizing state feedback of (1.1) with basin of attraction containing $\mathcal{K}_x$ and by τ an embedding of (1.1) into (4.23). According to Theorem 4.15, there exists $\pi : X \to \mathcal{K}_x$,

a strong left-inverse of τ on $\mathcal{K}_x$. Then, a natural way to build a dynamic output feedback would be to combine (1.1)-(4.24) with the control input $u = \phi(\pi(\hat{z}))$, and to ensure that the state x of (1.1) remains in $\mathcal{K}_x$. However, due to the unobservability of the original system at the target, we propose, as in Section 4.1.2, to add a perturbation to this feedback law. In [Cel+89], the convergence of the error system (4.25) to 0, when it holds, is only in the weak topology of X. Therefore, the perturbation added to the feedback law must be chosen to vanish when the observer state $\hat{z}$ of (4.24) tends towards $\tau(0)$ in the weak topology. For this reason, let us define a weak norm on X.

Definition 4.19 (Weak norm)**.** Let $(e_k)_{k\in\mathbb{Z}}$ be a Hilbert basis of X. For all $\xi \in X$, set

$$\mathcal{N}(\xi) = \sqrt{\sum_{k\in\mathbb{Z}} \frac{|\langle \xi, e_k \rangle_X|^2}{k^2+1}}.$$

Then $\mathcal{N}$ defines a norm, we call the *weak norm*, on X.

Note that $\mathcal{N}$ is not equivalent to $\|\cdot\|_X$, but satisfies $\mathcal{N}(\cdot) \leqslant \nu\|\cdot\|_X$ with $\nu = \sqrt{\sum_{k\in\mathbb{Z}} \frac{1}{k^2+1}} < +\infty$. Moreover, $\mathcal{N}$ induces a metric on bounded sets of X endowed with the weak topology. More precisely, for any bounded sequence $(\xi_n)_{n\in\mathbb{N}}$ in X, $\mathcal{N}(\xi_n) \to 0$ as n goes to infinity if and only if $\xi_n \overset{w}{\to} 0$ as n goes to infinity. Now, for some positive constant δ to be fixed (small enough) later, we can add the perturbation $\hat{z} \mapsto \delta\mathcal{N}^2(\hat{z} - \tau(0))$ to the feedback law, and obtain the following full coupled system:

$$\begin{cases} \dot{x} = f(x,u) \\ y = h(x) \end{cases}, \qquad \begin{cases} \dot{\hat{z}} = \mathcal{A}(u)\hat{z} - \alpha\mathcal{C}^*(\mathcal{C}\hat{z} - \mathfrak{h}(y)) \\ u = \phi(\pi(\hat{z})) + \delta\mathcal{N}^2(\hat{z} - \tau(0)). \end{cases} \tag{4.32}$$

Since X is infinite-dimensional and $\hat{z}$ lies in X, Definition 1.1 of semi-global dynamic output feedback stabilization must be revised. Indeed, (4.32) does not exactly fit the form of (1.2).

Definition 4.20 (Infinite-dimensional embedding-based dynamic output feedback stabilizability)**.** Let $\mathcal{K}_x \subset \mathbb{R}^n$ be a compact set. System (1.1) is said to be *stabilizable over $\mathcal{K}_x$ by means of an infinite-dimensional embedding-based dynamic output feedback* if and only if the following holds.

There exists an embedding τ of (1.1) into (4.23), a strong left-inverse π of τ on $\mathcal{K}_x$, a map $\phi : \mathbb{R}^n \to \mathbb{R}^p$, two positive constants α and δ and a compact set $\mathcal{K}_w \subset \mathbb{R}^n$ such that:

(i) For all initial conditions $(x_0, \hat{z}_0) \in \mathcal{K}_x \times \tau(\mathcal{K}_w)$, (4.32) has at least one solution in X over $\mathbb{R}_+$.

(ii) For all $R_x, R_{\hat{z}} > 0$, there exist $r_x, r_{\hat{z}} > 0$ such that for all $(x_0, \hat{z}_0) \in \mathcal{K}_x \times \tau(\mathcal{K}_w)$, if $|x_0| < r_x$ and $\|\hat{z}_0 - \tau(0)\|_X < r_{\hat{z}}$, then any solution $(x, \hat{z})$ of (4.32) starting from $(x_0, \hat{z}_0)$ satisfies $|x(t)| < R_x$ and $\|\hat{z}(t) - \tau(0)\|_X < R_{\hat{z}}$ for all $t \geqslant 0$.

(iii) Any solution $(x, \hat{z})$ of (4.32) with initial condition in $\mathcal{K}_x \times \tau(\mathcal{K}_w)$ is such that $x(t) \to 0$ and $\hat{z}(t) \overset{w}{\to} \tau(0)$ as t goes to infinity.

If the previous conditions hold for any compact $\mathcal{K}_x \subset \mathbb{R}^n$, then system (1.1) is said to be *semi-globally stabilizable by means of an infinite-dimensional embedding-based dynamic output feedback*.

Remark 4.21. If X is finite-dimensional, then *(i)-(ii)-(iii)* is equivalent to the usual definition of asymptotic stability of (4.32) at $(0, \tau(0))$ with basin of attraction containing $\mathcal{K}_x \times \tau(\mathcal{K}_w)$. However, when X is infinite-dimensional (the case of interest in this section), the convergence of trajectories towards the equilibrium point holds only in the weak topology. Hence, *(i)-(ii)-(iii)* is not equivalent to the usual definition of asymptotic stability of the infinite-dimensional system (4.32).

4.2.2 Back to the illustrative example

In this section, we illustrate the use of infinite-dimensional embeddings in the context of output feedback stabilization on a two-dimensional example with linear dynamics and nonlinear observation map. Let $h : \mathbb{R}^2 \to \mathbb{C}$. We consider the problem of stabilization by means of an infinite-dimensional embedding-based dynamic output feedback of the following system:

$$\begin{cases} \dot{x} = Ax + bu \\ y = h(x) \end{cases} \quad \text{with } A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{4.33}$$

Since (A, b) is stabilizable, there exists $K \in \mathbb{R}^{1\times 2}$ such that $A + bK$ is Hurwitz. Moreover, A is skew-symmetric. Hence $\kappa = |K|$ can be chosen arbitrarily small. Then, the state feedback law $\phi : x \mapsto Kx$ is such that (4.33) with $u = \phi(x)$ is globally asymptotically stable at 0. Note that (4.33) does not exactly fit the form of (4.3) since h is not necessarily radially-symmetric. Of course, our analysis is of interest only if (4.33) is not uniformly observable. In Example 4.26, we give an example of non-radially symmetric h that makes the system non-uniformly observable, and on which our (infinite-dimensional) embedding-based strategy does apply. In the following we give some sufficient conditions on h allowing the design of a stabilizing infinite-dimensional dynamic output feedback. The main result of this section, Theorem 4.39 (stated in Section 4.2.2), relies on three main hypotheses: the existence of an embedding of (4.33) into (4.32), and two observability assumptions. For each of these assumptions, we provide examples of output maps h satisfying theses hypotheses.

Unitary representations and embeddings

In [Cel+89], the authors investigated the problem of observer design for (4.33) by using infinite-dimensional embeddings. We briefly recall their strategy, that relies on representation theory (see, *e.g.,* [Vil68, BR86]).

Definition 4.22 (Group representation)**.** Let $\mathbb{G}$ be a locally compact, separable, unimodular topological group and let X be a separable Hilbert space. A map $\rho : \mathbb{G} \to \mathscr{L}(X)$ is a *representation* of $\mathbb{G}$ in X if the following conditions are satisfied:

(i) $\rho(g_1)\rho(g_2) = \rho(g_1 g_2)$ for all $g_1, g_2 \in \mathbb{G}$.

(ii) $\rho(\mathrm{e}) = \mathrm{Id}_X$ where e is the identity element of $\mathbb{G}$.

In other words, ρ is a group morphism from $\mathbb{G}$ to $\mathscr{L}(X)$. A representation ρ is *unitary* if $\rho(g)$ is a unitary operator for all $g \in \mathbb{G}$. A representation ρ is *irreducible* if it has no proper closed invariant subspace.

A group of interest in the context of control systems is the Lie group of the system.

Definition 4.23 (Lie group of a control-affine system)**.** Consider a control-affine system (see Definition 2.1) of the form

$$\dot{x} = f(x) + \sum_{i=1}^{p} u_i g_i(x) \tag{4.34}$$

where the vector fields f and g_i, $1 \leqslant i \leqslant p$ are complete. The Lie algebra of (4.34) is generated by the family

$$\mathcal{F} = \{f + \sum_{i=1}^{p} u_i g_i,\ u \in \mathbb{R}^p\}. \tag{4.35}$$

The Lie group $\mathbb{G}$ of (4.34) is generated by the family of diffeomorphisms

$$\{e^{t_k f_k} \circ \cdots \circ e^{t_1 f_1},\ t_i \in \mathbb{R}, f_i \in \mathcal{F}, k \in \mathbb{N}\}. \tag{4.36}$$

The group $\mathbb{G}$ is a subgroup of diffeomorphisms on $\mathbb{R}^n$.

More generally, Lie groups and algebras can be defined on Killing systems, that are compete systems with finite-dimensional Lie algebra (see [Cel+89]).

The Lie group $\mathbb{G}$ of system (4.33) (the group of flows generated by the dynamical system (4.33) with constant inputs) is isomorphic to $\mathbb{R}^2 \rtimes_{\mathcal{R}} H$, where $H \simeq \{e^{tA}, t \in \mathbb{R}\} \simeq \mathbb{S}^1$ is the group of rotations (isomorphic to the unit circle), $\mathcal{R} : \mathbb{S}^1 \ni \theta \mapsto e^{\theta A}$ is an automorphism of $\mathbb{R}^2$ and $\rtimes_{\mathcal{R}}$ denotes the outer semi-direct product with respect to $\mathcal{R}$.

Indeed, let $\varphi_t(x_0, u)$ denotes the flow of (4.33) for some $t \in \mathbb{R}$, $u \in \mathbb{R}$ and $x_0 \in \mathbb{R}^n$. Then

$$\varphi_t(x_0, u) = e^{tA}x_0 + \int_0^t e^{(t-s)A} B u \mathrm{d}s$$

and

$$\begin{aligned}\varphi_{t_2}(\varphi_{t_1}(x_0, u_1), u_2) &= e^{t_2 A}\left(e^{t_1 A}x_0 + \int_0^{t_1} e^{(t_1-s)A} b u_1 \mathrm{d}s\right) + \int_0^{t_2} e^{(t_2-s)A} b u_2 \mathrm{d}s \\ &= e^{(t_2+t_1)A}x_0 + e^{t_2 A}\int_0^{t_1} e^{(t_1-s)A} b u_1 \mathrm{d}s + \int_0^{t_2} e^{(t_2-s)A} b u_2 \mathrm{d}s.\end{aligned}$$

Since the pair (A, b) is controllable, $\{\int_0^t e^{(t-s)A} b u \mathrm{d}s, t \in \mathbb{R}, u \in \mathbb{R}\}$ generates $\mathbb{R}^2$. We recognize the structure of the outer semi-direct product. Thus $\mathbb{G} \simeq \mathbb{R}^2 \rtimes_{\mathcal{R}} H$ is the group of motions of the plane. According to [Vil68, Section IV.2], its unitary irreducible representations are given by a family $(\rho_\mu)_{\mu>0}$, where for each $\mu > 0$,

$$\begin{aligned}\rho_\mu : \quad \mathbb{G} &\longrightarrow \mathscr{L}(L^2(\mathbb{S}^1, \mathbb{C})) \\ (x, \vartheta) &\longmapsto \left(\xi \in L^2(\mathbb{S}^1, \mathbb{C}) \mapsto \left(\mathbb{S}^1 \ni s \mapsto e^{i\mu(1,0)e^{sA'}x}\xi(s - \vartheta)\right)\right).\end{aligned}$$

Let $X = L^2(\mathbb{S}^1, \mathbb{C})$ be the set of real-valued square-integrable functions over $\mathbb{S}^1$. Then X is a Hilbert space endowed with the scalar product defined by $\langle \xi, \zeta \rangle_X =$

$\frac{1}{2\pi}\int_0^{2\pi}\xi(s)\bar{\zeta}(s)\mathrm{d}s$ and the induced norm $\|\cdot\|_X$. Since $\mathbb{S}^1$ is compact, the constant function $\mathbb{1}: s \mapsto 1$ lies in X. Let $\mu > 0$ to be fixed later. Set

$$\begin{aligned}\tau_\mu : \mathbb{R}^2 &\longrightarrow X\\ x &\longmapsto \rho_\mu(x,0)\mathbb{1}.\end{aligned}$$

Note that τ_μ depends on μ, but from now on we omit this dependence in the notation and write τ instead of τ_μ. Since ρ_μ is a unitary representation, $\|\tau(x)\|_X = 1$ for all $x \in \mathbb{R}^2$ and $\tau(0) = \mathbb{1}$. For all $x = (x_1, x_2) = (r\cos(\theta), r\sin(\theta))$ in $\mathbb{R}^2$, we have

$$\tau(x) : \mathbb{S}^1 \ni s \mapsto e^{i\mu(x_1\cos(s)+x_2\sin(s))} = e^{i\mu r\cos(s-\theta)}. \tag{4.37}$$

If $x, \tilde{x} \in \mathbb{R}^2$ are such that $\tau(x) = \tau(\tilde{x})$, then $(x_1 - \tilde{x}_1)\cos(s) + (x_2 - \tilde{x}_2)\sin(s) = 0$ for all $s \in \mathbb{S}^1$, hence $x = \tilde{x}$. Thus τ is injective. Let $u \in C^0(\mathbb{R}_+, \mathbb{R})$. Let x be a solution of (4.33) and set $z = \tau(x) \in C^0(\mathbb{R}_+, H^1(\mathbb{S}^1, \mathbb{C})) \cap C^1(\mathbb{R}_+, X)$. Then

$$\begin{aligned}\dot{z} &= i\mu\left(\dot{x}_1\cos(s) + \dot{x}_2\sin(s)\right)z\\ &= i\mu\left(-x_2\cos(s) + x_1\sin(s) + u\sin(s)\right)z\\ &= -\frac{\partial z}{\partial s} + iu\mu\sin(s)z\\ &= \mathcal{A}(u)z\end{aligned}$$

with $\mathcal{A}(u) = -\frac{\partial}{\partial s} + iu\sin(s)$ defined on the dense domain $\mathcal{D} = H^1(\mathbb{S}^1, \mathbb{C}) = \{f \in X \;:\; f' \in X\}$. The operator $\mathcal{A}(u)$ is the skew-adjoint generator of a strongly continuous unitary group on X for any $u \in \mathbb{R}$. In order to make τ an embedding of (4.33) into (4.23), we need the output map to be in the form $\mathfrak{y} = \mathcal{C}z$. This is where the freedom degree $\mathfrak{h}$ introduced in (4.32) may be employed. More specifically, we make the following first assumption on the observation map h.

Assumption 4.24 (Linearizable output map)**.** There exist $\mathfrak{h} : \mathbb{R}^m \to \mathbb{C}^{\mathfrak{m}}$ and $\mathcal{C} \in \mathscr{L}(X, \mathbb{C}^{\mathfrak{m}})$ such that $\mathfrak{h}(h(x)) = \mathcal{C}\tau(x)$ for all $x \in \mathbb{R}^2$ and (4.30) is satisfied.

Remark 4.25. If Assumption 4.24 is satisfied, then the embedding defined in (4.37) shows that Assumption 4.18 is satisfied. Moreover, (4.30) implies that $\mathcal{C}\varepsilon \to 0$.

Example 4.26. Denote by J_k the Bessel function of the first kind of order $k \in \mathbb{Z}$, that is,

$$J_k : \mathbb{R} \ni r \mapsto \frac{1}{2\pi}\int_0^{2\pi} e^{ir\sin(s)-iks}\mathrm{d}s \in \mathbb{R}. \tag{4.38}$$

For all $k \in \mathbb{Z}$, let

$$\begin{aligned}e_k : \mathbb{S}^1 &\longrightarrow \mathbb{C}\\ s &\longmapsto e^{iks}.\end{aligned}$$

The family $(e_k)_{k\in\mathbb{Z}}$ forms a Hilbert basis of X. In the rest of the chapter, the weak norm $\mathcal{N}$ is always defined with respect to this Hilbert basis. Then, for all $x = (r\cos(\theta), r\sin(\theta)) \in \mathbb{R}^2$ and all $k \in \mathbb{Z}$,

$$\begin{aligned}\langle\tau(x), e_k\rangle_X &= \frac{1}{2\pi}\int_0^{2\pi} e^{i\mu r\cos(s-\theta)-iks}\mathrm{d}s\\ &= \frac{1}{2\pi}e^{-ik\theta+i\frac{\pi}{2}}\int_0^{2\pi} e^{i\mu r\sin(s)-iks}\mathrm{d}s\\ &= i^k J_k(\mu r)e^{-ik\theta}.\end{aligned} \tag{4.39}$$

Since $(e_k)_{k\in\mathbb{Z}}$ is a Hilbert basis of X, any function h such that

$$\mathfrak{h}(h(r\cos(\theta), r\sin(\theta))) = \sum_{k\in\mathbb{Z}} c_k J_k(\mu r) e^{-ik\theta}$$

for some map $\mathfrak{h}$ and $(c_k)_{k\in\mathbb{Z}} \in l^2(\mathbb{Z},\mathbb{C})$ satisfies $\mathfrak{h}(h(x)) = \langle \tau(x), \mathcal{C}^*\rangle_X$ with $\mathcal{C}^* = \sum_{k\in\mathbb{Z}} c_k(-i)^k e_k$. Moreover, if $c_k \neq 0$ only for a finite number of $k \in \mathbb{Z}$, then $\mathcal{C}^* \in \mathcal{D}$. Hence, for all $x \in \mathcal{D}$ and all $u \in \mathbb{R}$,

$$\|\mathcal{C}^*\mathcal{A}(u)x\|_X \leqslant \|\mathcal{C}^*\|_X \, |\mathcal{A}(u)x| = \|\mathcal{C}^*\|_X \, |\langle x, \mathcal{A}(u)\mathcal{C}^*\rangle_X| \leqslant \|\mathcal{C}^*\|_X \, \|\mathcal{A}(u)\mathcal{C}^*\|_X \, \|x\|_X$$

since $\mathcal{A}(u)$ is skew-adjoint. Thus, Assumption 4.24 is satisfied. For example, $h(x) = J_0(\mu|x|) - 1$ (with $\mathfrak{h}(y) = y+1$), $h(x) = J_2(\mu|x|)\cos(2\theta)$ (with $\mathfrak{h}(y) = y$) and $h(x) = |x|$ (with $\mathfrak{h}(y) = J_0(\mu y)$) are suitable observations maps. In each of these cases, the constant input $u \equiv 0$ makes (4.33) unobservable. Moreover, $h(x) = J_0(\mu|x|) - 1$ and $h(x) = |x|$ are radially symmetric but $h(x) = J_2(\mu|x|)\cos(2\theta)$ is not. If $h(x) = |x|$, then (4.33) is a subcase of system (4.3).

Remark 4.27. According to the Gelfand–Raïkov theorem, the finite linear combinations of pure positive-type functions (*i.e.,* of the form $(x,\vartheta) \mapsto \langle \rho_\mu(x,\vartheta)\xi, \xi\rangle_X$, where $\mu > 0$ and $\xi \in X$) is dense for the uniform convergence on compact sets, in the continuous bounded complex-valued functions on G. Hence, the set of functions of the form $(r\cos(\theta), r\sin(\theta)) \mapsto \sum_{\ell\in I_1}\sum_{k\in I_2} c_k J_k(\mu_\ell r)e^{-ik\theta}$, where I_1 and I_2 are finite subsets of $\mathbb{Z}$, $\mu_\ell > 0$ and $c_k \in \mathbb{C}$, is dense for the uniform convergence on compact sets of $\mathbb{R}^2$, in the continuous bounded complex-valued functions on $\mathbb{R}^2$. In the examples of applications of our results, we will focus on output maps h of the form $\mathfrak{h}(h(x)) = \sum_{k\in I} c_k J_k(\mu r)e^{-ik\theta}$ for some $\mathfrak{h} : \mathbb{R}^m \to \mathbb{C}^{\mathfrak{m}}$ and some fixed $\mu > 0$.

Explicit strong left-inverse

Having in mind to use the strategy developed in the previous section, we now explicitly construct a strong left-inverse π of τ defined in (4.37) over some compact set. With Corollary 4.16, we already know that a strong left-inverse π exists. However, we would like to give an explicit expression. This can be done by employing the relationship between Bessel functions of the first kind given in (4.38) and the embedding τ, as shown in equation (4.39).

Indeed, let j_1 denote the first zero of J_1'. Then J_1 is increasing over $[0, j_1]$. Denote J_1^{-1} its inverse over $[0, j_1]$. Let $\Phi : \mathbb{C} \ni x_1 + ix_2 \mapsto (x_1, x_2) \in \mathbb{R}^2$ be the canonical bijection. Let $j \in (0, j_1)$. For all $\zeta \in \mathbb{C}$, let

$$\mathfrak{f}(\zeta) = \begin{cases} 0 & \text{if } \zeta = 0 \\ \Phi\left(\frac{i\bar{\zeta}}{\mu|\zeta|} J_1^{-1}(|\zeta|)\right) & \text{if } 0 < |\zeta| \leqslant J_1(j) \\ \Phi\left(\frac{i\zeta}{\mu|\zeta|} j_1\right) & \text{if } |\zeta| \geqslant J_1(j_1) \end{cases} \tag{4.40}$$

If $J_1(j) < |\zeta| < J_1(j_1)$, define $\mathfrak{f}(\zeta)$ such that $\mathfrak{f}$ is continuously differentiable and globally Lipschitz over $\mathbb{C}$. Denote by $\ell_{\mathfrak{f}}$ its Lipschitz constant. Let $e_1 \in X$ be defined by $e_1(s) = e^{is}$ for all $s \in \mathbb{S}^1$. Let

$$\begin{aligned} \pi : X &\longrightarrow \mathbb{R}^2 \\ \xi &\longmapsto \mathfrak{f}(\langle \xi, e_1\rangle_X) \end{aligned} \tag{4.41}$$

Lemma 4.28. *The map π is a strong left-inverse of τ over $\bar{B}_{\mathbb{R}^2}(0, \frac{j}{\mu})$.*

Proof. Set $\mathcal{K}_x = \bar{B}_{\mathbb{R}^2}(0, \frac{j}{\mu})$. According to (4.40), $\phi(\xi) \in \mathcal{K}_x$ for all $x \in \mathcal{K}_x$.

Let $x = (r\cos(\theta), r\sin(\theta))$ in $\mathcal{K}_x$. Then, with (4.39),

$$\langle \tau(x), e_1 \rangle_X = ie^{-i\theta} J_1(\mu r) \in \mathcal{B}_{\mathbb{C}}\left(0, J_1\left(j\right)\right).$$

Hence $\pi(\tau(x)) = \Phi(re^{i\theta}) = x$. Let $\xi \in X$. We have

$$|\pi(\xi) - x| = |\pi(\xi) - \pi(\tau(x))| = |\mathfrak{f}(\langle \xi, e_1 \rangle_X) - \mathfrak{f}(\langle \tau(x), e_1 \rangle_X)| \leqslant \ell_{\mathfrak{f}} |\langle \xi - \tau(x), e_1 \rangle_X|.$$

Hence π is a strong left-inverse of τ over $\mathcal{K}_x$. ■

Remark 4.29. Letting μ tends towards 0, the domain of the left-inverse tends towards $\mathbb{R}^2$, which will be of use to achieve semi-global stabilization.

Well-posedness and boundedness of trajectories

We now check the well-posedness of the closed-loop system (4.32). In a second step, since $\pi(\xi)$ is meaningful only if $|\langle \xi, e_1 \rangle_X| \leqslant J_1(j)$, we show that by selecting the (perturbation) parameter δ sufficiently small, $\hat{z}$ remains in this domain along the trajectories of the closed-loop system.

Lemma 4.30. *For all $\mu, \alpha, \delta > 0$ and all $x_0, \hat{x}_0$ in $\bar{B}_{\mathbb{R}^2}(0, \frac{j}{\mu})$, the system* (4.32) *(with π as in Lemma 4.28) admits a unique solution*

$$(x, \hat{z}) \in C^1(\mathbb{R}_+, \mathbb{R}^2) \times \left(C^0(\mathbb{R}_+, \mathcal{D}) \cap C^1(\mathbb{R}_+, X)\right)$$

such that $x(0) = x_0$ and $\hat{z}(0) = \tau(\hat{x}_0)$.

Proof. Let $\mathcal{K}_x = \bar{B}_{\mathbb{R}^2}(0, \frac{j}{\mu})$ and x_0, $\hat{x}_0$ in $\mathcal{K}_x$. Set $z_0 = \tau(x_0) \in \mathcal{D}$ and $\varepsilon_0 = \tau(\hat{x}_0) - \tau(x_0) \in \mathcal{D}$. The well-posedness of system (4.32) is equivalent to the well-posedness of the following system:

$$\begin{cases} \dot{z} = \mathcal{A}(u) z \\ \dot{\varepsilon} = (\mathcal{A}(u) - \alpha \mathcal{C}^* \mathcal{C}) \varepsilon \\ u = \phi(\pi(z + \varepsilon)) + \delta \mathcal{N}^2(z + \varepsilon - \mathbb{1}) \\ z(0) = z_0,\ \varepsilon(0) = \varepsilon_0 \end{cases} \tag{4.42}$$

where $\mathcal{A}(u) = -\frac{\partial}{\partial s} + i\mu u \sin$ and $\mathcal{C} \in \mathscr{L}(X, \mathbb{C}^{\mathfrak{m}})$. Set

$$\mathcal{A}_0 = \begin{pmatrix} -\frac{\partial}{\partial s} & 0 \\ 0 & -\frac{\partial}{\partial s} - \alpha \mathcal{C}^* \mathcal{C} \end{pmatrix},$$
$$\mathcal{F} : (z, \varepsilon) \mapsto \begin{pmatrix} i\mu \left(\phi(\pi(z + \varepsilon)) + \delta \mathcal{N}^2(z + \varepsilon - \mathbb{1})\right) \sin(\cdot) z \\ i\mu \left(\phi(\pi(z + \varepsilon)) + \delta \mathcal{N}^2(z + \varepsilon - \mathbb{1})\right) \sin(\cdot) \varepsilon \end{pmatrix}.$$

Since $\mathcal{C}$ is bounded and $\mathcal{A}_0$ is diagonal, $\mathcal{A}_0$ is the generator of a strongly continuous semigroup on X^2. Since π and $\mathcal{N}^2$ are locally Lipschitz, $\mathcal{F}$ is locally Lipschitz. Hence, according to [Seg63, Theorem 1], system (4.42) admits a unique solution $(z, \varepsilon) \in C^0([0, T], X^2)$ for some $T \in \mathbb{R}_+^* \cup \{+\infty\}$. Moreover, since $\mathcal{A}(u)$ is skew-adjoint for all $u \in \mathbb{R}$, $\|z\|_X$ is constant and $\|\varepsilon\|_X$ is non-increasing. Hence, $T = +\infty$. Since π and $\mathcal{N}^2$ are continuously Fréchet differentiable, $\mathcal{F}$ is continuously Fréchet differentiable. Thus, $(z, \varepsilon) \in C^0(\mathbb{R}_+, \mathcal{D}^2) \cap C^1(\mathbb{R}_+, X^2)$. ■

Now that the existence and uniqueness of solutions of (4.32) is proved, let us show the boundedness of trajectories.

Lemma 4.31. *For all $\mu > 0$, all $R_2 \in (0, \frac{j}{\mu})$ and all $R_1 \in (0, R_2)$, there exist $R_0 \in (0, R_1)$ and $\delta_0 > 0$ such that for all $x_0, \hat{x}_0$ in $\mathcal{B}_{\mathbb{R}^2}(0, R_0)$, all $\alpha > 0$ and all $\delta \in (0, \delta_0)$, the unique solution $(x, \hat{z}) \in C^1(\mathbb{R}_+, \mathbb{R}^2) \times (C^0(\mathbb{R}_+, \mathcal{D}) \cap C^1(\mathbb{R}_+, X))$ of (4.32) such that $x(0) = x_0$ and $\hat{z}(0) = \tau(\hat{x}_0)$ satisfies $|x(t)| < R_1$, $|\langle \hat{z}(t), e_1 \rangle_X| < J_1(\mu R_2)$ and $|\pi(\hat{z}(t))| < R_2$ for all $t \in \mathbb{R}_+$.*

Proof. Recall that $\kappa = |K|$. Denote by ℓ_π the global Lipschitz constant of π. Let $R_0 \in (0, R_1)$ and $\delta_0 > 0$ satisfying the following inequalities:

$$R_0 + M\left(2\kappa\ell_\pi\sqrt{2(1 - J_0(\mu R_0))} + 16\nu^2\delta_0\right) < R_1, \tag{4.43}$$

$$2\sqrt{2(1 - J_0(\mu R_0))} + J_1(\mu R_1) < J_1(\mu R_2). \tag{4.44}$$

This is always possible by choosing R_0 and δ_0 small enough since $J_0(0) = 1$.

Let $\delta \in (0, \delta_0)$, $x_0, \hat{x}_0 \in \mathcal{B}_{\mathbb{R}^2}(0, R_0)$, $(x, \hat{z})$ as in Lemma 4.30, $z = \tau(x)$, $\varepsilon = \hat{z} - z$ and $u = \phi(\pi(\hat{z})) + \delta\mathcal{N}^2(\hat{z} - \mathbb{1})$. Set $e = b(u - Kx)$. Then $\dot{x} = (A + bK)x + e$. According to the variation of constants formula, and since $A + bK$ is Hurwitz, we get that

$$|x(t)| \leqslant |x_0| + M \sup_{s\in[0,t]} |e(s)| \qquad \forall t \in \mathbb{R}_+, \tag{4.45}$$

for some $M > 0$. Note that

$$\begin{aligned}
\|\tau(x_0) - \mathbb{1}\|_X &= \left(\|\tau(x_0)\|_X^2 + 1 - 2\langle \tau(x_0), 1\rangle_X\right)^{\frac{1}{2}} \\
&= \sqrt{2(1 - J_0(\mu|x_0|))} \\
&\leqslant \sqrt{2(1 - J_0(\mu R_0))}.
\end{aligned}$$

Then

$$\|\varepsilon_0\|_X \leqslant \|\hat{z}_0 - \mathbb{1}\|_X + \|z_0 - \mathbb{1}\|_X \leqslant 2\sqrt{2(1 - J_0(\mu R_0))}. \tag{4.46}$$

Let $t \in [0, T]$. Then

$$|e(t)| \leqslant \kappa|\pi(\hat{z}(t)) - x(t)| + \delta\mathcal{N}^2(\hat{z}(t) - \mathbb{1}). \tag{4.47}$$

On one hand,

$$\begin{aligned}
\mathcal{N}^2(\hat{z}(t) - \mathbb{1}) &\leqslant \nu^2 \|\hat{z}(t) - \mathbb{1}\|_X^2 \\
&\leqslant \nu^2 \left(\|\varepsilon(t)\|_X + \|z(t) - \mathbb{1}\|_X\right)^2 && \text{(by triangular inequality)} \\
&\leqslant \nu^2 \left(\|\varepsilon_0\|_X + 2\right)^2 \\
& && \text{(since } \|\varepsilon\|_X \text{ is non-increasing and } \|\tau(x(t))\|_X = 1\text{)} \\
&\leqslant 16\nu^2. && \text{(since } \|z_0\|_X = \|\hat{z}_0\|_X = 1\text{)}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
|\pi(\hat{z}(t)) - x(t)| &= |\pi(\hat{z}(t)) - \pi(z(t))| \\
&\leqslant \ell_\pi \|\varepsilon(t)\|_X \leqslant \ell_\pi \|\varepsilon_0\|_X \leqslant 2\ell_\pi\sqrt{2(1 - J_0(\mu R_0))}. && \text{(by (4.46))}
\end{aligned}$$

Hence,

$$|e(t)| \leqslant 2\kappa\ell_\pi\sqrt{2(1-J_0(\mu R_0))} + 16\nu^2\delta. \tag{4.48}$$

Thus, combining (4.45) and (4.43), $|x(t)| < R_1$ for all $t \in \mathbb{R}_+$. Then, for all $t \in \mathbb{R}_+$,

$$\begin{aligned}
|\langle \hat{z}(t), e_1\rangle_X| &\leqslant |\langle \varepsilon(t), e_1\rangle_X| + |\langle z(t), e_1\rangle_X| \\
&\leqslant \|\varepsilon(t)\|_X + |\langle \tau(x(t)), e_1\rangle_X| \\
&\leqslant \|\varepsilon_0\|_X + J_1(\mu|x|) \\
&\leqslant 2\sqrt{2(1-J_0(\mu R_0))} + J_1(\mu R_1) \\
&< J_1(\mu R_2).
\end{aligned}$$

Thus, (4.44) yields $|\langle \hat{z}(t), e_1\rangle_X| < J_1(\mu R_2)$ for all $t \in \mathbb{R}_+$. Finally, since $J_1(\mu R_2) < J_1(j)$, $|\pi(\hat{z}(t))| = |\mathfrak{f}(\langle \hat{z}(t), e_1\rangle_X)| = \frac{1}{\mu}J_1^{-1}(\langle \hat{z}(t), e_1\rangle_X) \leqslant R_2$. ■

In particular, we have the following corollary, which shows that the compact set of initial conditions that ensures the boundedness of trajectories can be chosen as big as desired, as soon as μ and δ are sufficiently small.

Corollary 4.32. *For all $R_0 > 0$, there exist $\mu_0 > 0$, $\delta_0 > 0$ and $R_2 > R_1 > R_0$ such that for all $x_0, \hat{x}_0$ in $\mathcal{B}_{\mathbb{R}^2}(0, R_0)$, all $\mu \in (0, \mu_0)$, all $\alpha > 0$ and all $\delta \in (0, \delta_0)$, the unique solution $(x, \hat{z}) \in C^1(\mathbb{R}_+, \mathbb{R}^2) \times (C^0(\mathbb{R}_+, \mathcal{D}) \cap C^1(\mathbb{R}_+, X))$ of (4.32) such that $x(0) = x_0$ and $\hat{z}(0) = \tau(\hat{x}_0)$ satisfies $|x(t)| < R_1$, $|\langle \hat{z}(t), e_1\rangle_X| < J_1(\mu R_2)$ and $|\pi(\hat{z}(t))| < R_2$ for all $t \in \mathbb{R}_+$.*

Proof. Let $\beta_2 > \beta_1 > 1$ to be fixed later, and let $R_1 = \beta_1 R_0$ and $R_2 = \beta_2 R_0$. Then there exist $\mu_0, \delta_0 > 0$ small enough such that (4.43) holds for all $\mu \in (0, \mu_0)$. Recall the following asymptotic expansions of the Bessel functions of the first kind at 0:

$$J_0(r) = 1 - \frac{r^2}{4} + o(r^2), \qquad J_1(r) = \frac{r}{2} + o(r).$$

Then for all $\mu > 0$,

$$2\sqrt{2(1-J_0(\mu R_0))} + J_1(\mu R_1) = \mu R_0\left(\sqrt{2} + \frac{\beta_1}{2}\right) + o(\mu),\ J_1(\mu R_2) = \mu R_0\frac{\beta_2}{2} + o(\mu).$$

Hence, if $\beta_2 > 2\sqrt{2} + \beta_1$, then there exists $\mu_0 > 0$ such that (4.44) holds for all $\mu \in (0, \mu_0)$. Set $\beta_1 = 2$ and $\beta_2 = 2\sqrt{2} + 3$. Then there exist $\mu_0 > 0$ and $\delta_0 > 0$ such that $\mu_0 R_2 < j$ and (4.43) and (4.44) are satisfied for all $\mu \in (0, \mu_0)$. Reasoning as in the proof of Lemma 4.31, the result follows. ■

Observability analysis

In order to state the main result of Section 4.2.2, we need to introduce two last assumptions on the linear output map $\mathcal{C}$ obtained from the function h in Assumption 4.24. For each assumption, we give examples of output maps h satisfying these assumptions. Following Remark 4.27, we investigate the case where at least one of the components of $\mathcal{C}$ is in the linear span of a finite number of elements of the Hilbert basis. This component is used to ensure the two observability properties. The first one states that $\mathcal{C}$ distinguishes the target point in a neighborhood of it.

Assumption 4.33 (Short time 0-detectability)**.** Let $u \in C^0(\mathbb{R}_+;\mathbb{R})$ and x be a solution of (4.33) bounded by $\frac{j}{\mu}$. If there exists $\Delta > 0$ such that $u(t_n+\cdot) \underset{n\to+\infty}{\longrightarrow} 0$ in the weak-$*$ topology of $L^\infty((0,\Delta);\mathbb{R})$ and $\mathcal{C}\tau(x(t_n+t)) \underset{n\to+\infty}{\longrightarrow} \mathcal{C}\tau(0)$ for all $t \in [0,\Delta]$, then $x(t_n) \underset{n\to+\infty}{\longrightarrow} 0$.

Remark 4.34. Assumption 4.33 implies the necessary Condition 1.11 (local). Indeed, if x is a solution of (4.33) with $u = 0$ and $h(x(t)) = 0$ for all $t \geqslant 0$, then for any positive increasing sequence $(t_n)_{n\in\mathbb{N}} \to +\infty$, $u(t_n) = 0$ and $C\tau(x(t_n+t)) = \mathfrak{h}(0)$ for all $n \in \mathbb{N}$ and all $t \geqslant 0$. Hence, according to Assumption 4.33, $x(t_n) \to 0$. Thus Condition 1.11 (local) is satisfied. Moreover, if $\mathfrak{h}$ has a continuous inverse in a neighborhood of 0, then Assumption 4.33 implies the input/output-to-state stability condition (see *e.g.,* [KSW01]), which states that any solution x of (4.33) such that $u(t) \to 0$ and $y(t) \to 0$ is such that $x(t) \to 0$ as $t \to +\infty$. This condition has proved to be of interest in the context of output feedback stabilization.

Example 4.35. Since $\mathcal{C} \in \mathscr{L}(X, \mathbb{C}^{\mathfrak{m}})$, it can be seen as a $\mathfrak{m}$-tuple of linear forms on X. Let $\zeta \in X \setminus \{0\}$ be such that $\langle \cdot, \zeta\rangle_X$ is one of these linear forms, that is, $\mathcal{C} = (\langle \cdot, \zeta\rangle_X, \dots)$. Let $(c_k)_{k\in\mathbb{Z}}$ be such that $\zeta = \sum_{k\in I} c_k e_k$, where $0 \in I \subset \mathbb{Z}$. If I is finite (see Remark 4.27), then Assumption 4.33 is satisfied.

Indeed if $\mathcal{C}\tau(x(t_n+t)) \underset{n\to+\infty}{\longrightarrow} \mathcal{C}\tau(0)$, then $\sum_{k\in I} c_k J_k(\mu r(t_n+t)) e^{-ik\theta(t_n+t)} \underset{n\to+\infty}{\longrightarrow} c_0$ (see (4.39)), where $x = (r\cos(\theta), r\sin(\theta))$. Let Δ be as in the assumption. Then, according to Duhamel's formula, for all $t \in [0,\Delta]$,

$$x(t_n+t) - e^{tA}x(t_n) \underset{n\to+\infty}{\longrightarrow} 0,$$

i.e.,

$$r(t_n+t) - r(t_n) \underset{n\to+\infty}{\longrightarrow} 0 \text{ and } e^{i\theta(t_n+t)} - e^{i(\theta(t_n)+t)} \underset{n\to+\infty}{\longrightarrow} 0,$$

Hence, $\sum_{k\in I} c_k J_k(\mu r(t_n)) e^{-ik\theta(t_n)} e^{-ikt} \underset{n\to+\infty}{\longrightarrow} c_0$ for all $t \in [0,\Delta]$. Since I is finite, this limit implies that $c_0 J_0(\mu r(t_n)) \to c_0$ and $c_k J_k(\mu r(t_n)) \to 0$ for $k \in I \setminus \{0\}$ as n goes to $+\infty$. Denote by j_0 the first zero of J_0. Then $J_k(r) \neq 0$ for any $r \in (-j_0, j_0) \setminus \{0\}$ and any $k \in \mathbb{Z}$. Since for some $k \in I$, $c_k \neq 0$, we have $J_k(\mu r(t_n)) \to J_k(0)$, hence $x(t_n) \to 0$ since $|r(t_n)| < \frac{j}{\mu}$, $j < j_1 < j_0$.

Moreover, if there exist $k_1, k_2 \in \mathbb{Z}$ with $|k_1| \neq |k_2|$, $c_{k_1} \neq 0$ and $c_{k_2} \neq 0$, then $j_0 = +\infty$ is a suitable choice due to the Bourgets's hypothesis, proved by Siegel in [Sie14].

The second hypothesis is that the unobservable input $u \equiv 0$ is isolated from other singular inputs of the infinite-dimensional system. Let us recall the usual definition of approximate observability of (4.23) (see, *e.g.,* [TW09]).

Definition 4.36 (Approximate observability (see Definition 5.22))**.** System (4.23) is said to be *approximately observable* in some time $T > 0$ for some input $u \in C^1(\mathbb{R}_+, \mathbb{R})$ if and only if

$$(\forall t \in [0,T],\ \mathcal{C}\mathbb{T}_t(z_0,u) = 0) \Longrightarrow z_0 = 0. \tag{4.49}$$

Since (4.23) is a linear system, Definition 4.36 coincides with Definition 1.15 in the finite-dimensional context. Moreover, setting $A(t) := A(u(t))$, Definition 4.36 is equivalent to Definition 5.22.

Assumption 4.37 (Isolated observability singularity)**.** Let u be a bounded input in $[-u_{\max}, u_{\max}]$ where $u_{\max} = \kappa\frac{j}{\mu} + 16\nu^2\delta$. If $u \not\equiv 0$, then u makes (4.23) approximately observable in some time $T > 0$.

Example 4.38. In [Cel+89, Example 1], the authors investigated the observability of constants inputs of (4.23) in the case where $\mu = 1$ and $\mathcal{C} = \langle \cdot, \mathbb{1} \rangle_X$ (*i.e.,* $\mathfrak{h} \circ h(x) = J_0(|x|)$, see Remark 4.26). Using a similar method, we prove that if $\mathcal{C} = (\langle \cdot, \zeta \rangle_X, \dots)$ for some $\zeta = \sum_{k\in I} c_k e_k$ in $X \setminus \{0\}$, where $I \subset \mathbb{Z}$ is finite, if $\mu u_{\max} < j_0$ for some $j_0 > 0$, then Assumption 4.37 holds at least for constant inputs bounded by $u_{\max}$. Other investigations should be carried out to deal with non constant inputs.

Note that it is always possible to make $\mu u_{\max} < j_0$ by choosing κj and $\mu\delta$ small enough. Moreover, the considered set of such maps $\mathcal{C}$ is sufficient to approximate any output map h, as explained in Remark 4.27.

Let $z_0 \in X$, $u \in \mathbb{R}\setminus\{0\}$ and $z(t) = \mathbb{T}_t(z_0, u)$ be the unique corresponding solution of (4.23). We have

$$\begin{aligned}
\langle z(t), \zeta \rangle_X &= \frac{1}{2\pi}\int_0^{2\pi} e^{-i\mu u \int_0^t \sin(s-\sigma)\mathrm{d}\sigma} z_0(s-t) \sum_{k\in I} \bar{c}_k e^{-iks}\mathrm{d}s \\
&\qquad\qquad \text{(by the method of characteristics)} \\
&= \frac{1}{2\pi}\int_0^{2\pi} \left(\sum_{k\in I} \bar{c}_k e^{-i\mu u\cos(s)-iks}\right) e^{i\mu u\cos(s-t)} z_0(s-t)\mathrm{d}s \\
&= (\psi * \Psi_0)(t)
\end{aligned}$$

where $*$ denotes the convolution product over X, $\psi : s \mapsto \sum_{k\in I} \bar{c}_k e^{-i\mu u\cos(s)-iks}$ and $\Psi_0 : s \mapsto e^{i\mu u\cos(s)} z_0(s)$. Hence, according to Parseval's theorem,

$$\frac{1}{2\pi}\int_0^{2\pi} |\langle z(t), \zeta\rangle_X|^2 \mathrm{d}t = \|\psi * \Psi_0\|_X^2 = \left\|\hat{\psi}\cdot\hat{\Psi}_0\right\|_{\hat{X}}^2 = \sum_{\ell\in\mathbb{Z}} |\langle \psi, e_\ell\rangle_X|^2 |\langle \Psi_0, e_\ell\rangle_X|^2.$$

where $\hat{\psi}$ (resp. $\hat{\Psi}_0$) denotes the Fourier series coefficients of ψ (resp Ψ_0) in $X = L^2(\mathbb{S}^1, \mathbb{C}) \subset L^1(\mathbb{S}^1, \mathbb{C})$ and $\hat{X} = l^2(\mathbb{Z}, \mathbb{C})$. Hence, it is sufficient to show that there exists $j_0 > 0$ such that, if $\mu u < j_0$, then $\langle \psi, e_\ell\rangle_X \neq 0$ for all $\ell \in \mathbb{Z}$. Indeed, it yields that if $\mathcal{C}z(t) = 0$ for all $t \in [0, 2\pi]$, then $\Psi_0 = 0$, *i.e.,* $z_0 = 0$, and thus u makes (4.23) approximately observable in time 2π.

Note that

$$\langle \psi, e_\ell\rangle_X = \frac{1}{2\pi}\int_0^{2\pi} \sum_{k\in I} \bar{c}_k e^{-i\mu u\cos(s)-i(k+\ell)s}\mathrm{d}s = \sum_{k\in I} \bar{c}_k i^k J_{k+\ell}(\mu u). \qquad \text{(by (4.39))}$$

Set $d_k = \bar{c}_k i^k$ and $F_\ell(r) = \sum_{k\in I} d_k J_{k+\ell}(r)$ for all $r \in \mathbb{R}$. Since F_ℓ is analytic for each $\ell \in \mathbb{Z}$, its zeros are isolated. Hence, for all $L > 0$, there exists $j_0 > 0$ such that, if $|\ell| < L$, then $F_\ell(r) \neq 0$ for all $r \in (-j_0, j_0) \setminus \{0\}$. Now, let $k_{\min} = \min\{k \in I \ : \ d_k \neq 0\}$ and let us prove that there exists $j_0 > 0$ such that $F_\ell(r) \neq 0$ for all $r \in (-j_0, j_0)\setminus\{0\}$ and all $\ell \geqslant -k_{\min}$. (One can reason similarly for $\ell \leqslant \max\{k \in I \ : \ d_k \neq 0\}$). We have $F_\ell(r) = d_{k_{\min}} J_{k_{\min}+\ell}(r)\left(1 + \sum_{k\in I} \frac{d_k}{d_{k_{\min}}} \frac{J_{k+\ell}(r)}{J_{k_{\min}+\ell}(r)}\right)$. According to [Neu04], $|J_{k+\ell}(r)| \leqslant \frac{1}{(k+\ell)!}\left(\frac{|r|}{2}\right)^{k+\ell}$ for all $r \in \mathbb{R}$. Moreover, according to [Laf86], if $|r| \leqslant 1$, then

$$|J_{k_{\min}+\ell}(r)| \geqslant |r|^{k_{\min}+\ell} J_{k_{\min}+\ell}(1) \geqslant \frac{|r|^{k_{\min}+\ell}}{(k_{\min}+\ell)!2^{k_{\min}+\ell}}\left(1 - \frac{1}{2(k_{\min}+\ell+1)}\right).$$

Hence

$$|F_\ell(r)| \geqslant |d_{k_{\min}}|\,|J_{k_{\min}+\ell}(r)|\left(1-\sum_{k\in I}\frac{|d_k|}{|d_{k_{\min}}|}\frac{|J_{k+\ell}(r)|}{|J_{k_{\min}+\ell}(r)|}\right)$$
$$\geqslant |d_{k_{\min}}|\,|J_{k_{\min}+\ell}(r)|\left(1-2\sum_{k\in I}\frac{|d_k|}{|d_{k_{\min}}|}\left(\frac{|r|}{2}\right)^{k-k_{\min}}\right).$$

Hence, there exits $j_0 > 0$ such that, if $0 < |r| < j_0$, $|F_\ell(r)| \geqslant \frac{|d_{k_{\min}}|}{2}\,|J_{k_{\min}+\ell}(r)|$ for all $\ell \in \mathbb{Z}$. Choosing $j_0 \leqslant \min\{r > 0 \ : \ J_0(r) = 0\}$, one has $J_{k_{\min}+\ell}(r) \neq 0$ for all $\ell \in \mathbb{Z}$, hence $F_\ell(r) \neq 0$.

In particular, if $\zeta = e_k$ for some $k \in I$, then $j_0 = \min\{r > 0 \ : \ J_0(r) = 0\}$ is a suitable choice. Indeed, $J_k(r) \neq 0$ for all $r \in (-j_0, j_0) \setminus \{0\}$ and all $k \in \mathbb{Z}$. Hence, if $\mu u_{\max} < j_0$, then u makes (4.23) approximately observable in time 2π.

Moreover, if $\mathcal{C} = (\langle \cdot, e_{k_1}\rangle_X, \langle \cdot, e_{k_2}\rangle_X, \dots)$ with $|k_1| \neq |k_2|$, then $j_0 = +\infty$ is a suitable choice due to the Bourgets's hypothesis, proved by Siegel in [Sie14].

We are now in position to state the main result of Section 4.2.2.

Theorem 4.39. *Let $\mathcal{K}_x \subset \mathbb{R}^2$ be a compact set. Let $R_0 > 0$ be such that $\mathcal{K}_x \subset B_{\mathbb{R}^2}(0, R_0)$. Let $\mu_0 > 0$ and $\delta_0 > 0$ be as in Corollary 4.32. Suppose that there exists $\mu \in (0, \mu_0)$ and $\delta \in (0, \delta_0)$ such that Assumptions 4.24, 4.33 and 4.37 are satisfied.*

Then system (4.33) *is stabilizable over $\mathcal{K}_x$ by means of an infinite-dimensional embedding-based dynamic output feedback. Moreover, the closed-loop system is explicitly given by* (4.32) *for any $\alpha > 0$ and with τ as in* (4.37) *and π as in* (4.41).

According to Examples 4.26, 4.35 and 4.38, we obtain the following corollary.

Corollary 4.40. *If $\mathfrak{h}(h(r\cos(\theta), r\sin(\theta))) = \sum_{k\in I} c_k J_k(\mu r)e^{-ik\theta}$ for some map $\mathfrak{h} : \mathbb{R}^m \to \mathbb{C}$, $\mu > 0$, $(c_k)_{k\in I} \in \mathbb{C}^I$ and $I \subset \mathbb{Z}$ finite, then Assumptions 4.24 and 4.33 are satisfied, and Assumption 4.37 is satisfied at least for constant inputs.*

Example 4.41. In this example, we assume that Assumption 4.37 is satisfied. If it is not the case, then similar results may be obtained by keeping the input constant on regular time intervals, namely, $u(t) = u(t_k)$ for all $t \in [t_k, t_{k+1})$, $t_{k+1} - t_k = \delta_t > 0$, $k \in \mathbb{N}$. Indeed, the boundedness of trajectories is still ensured, and Assumption 4.37 is only required for constant inputs.

As as application of Corollary 4.40, we provide the following examples of output maps for which our embedding based strategy allows to conclude to the stabilizability.

- If $h(x) = |x|$, then system (4.33) is semi-globally stabilizable by means of an infinite-dimensional embedding-based dynamic output feedback.

  Let $\mathcal{K}_x \subset \mathbb{R}^2$ be a compact set. Let $R_0 > 0$ be such that $\mathcal{K}_x \subset B_{\mathbb{R}^2}(0, R_0)$. Let $\mu_0 > 0$ and $\delta_0 > 0$ be as in Corollary 4.32. According to Examples 4.26 and 4.35, Assumptions 4.24 and 4.33 are satisfied for any $\mu \in (0, \mu_0)$ by considering $\mathfrak{h} : y \mapsto J_0(\mu y)$. Moreover, by choosing $\kappa j < j_0$ and $\delta < \frac{j_0 - \kappa j}{16\nu^2\mu}$, Assumption 4.37 is also satisfied according to Example 4.38. Hence, Theorem 4.39 does apply on $\mathcal{K}_x$.

- Naturally, this last example was treated via a finite-dimensional strategy in Section 4.1. Furthermore, for radially symmetric output maps, one can devise a strategy where $|x|$ is extracted (at least locally around the target by inversion) and apply the same finite-dimensional method. However, this is impossible if the output is not radially symmetric. For instance, if $h(r\cos(\theta), r\sin(\theta)) = J_2(\mu r)\cos(2\theta)$ for some $\mu > 0$, then system (4.33) is stabilizable over $\bar{B}(0, \frac{j}{\mu})$ for all $j \in (0, j_1)$ by means of an infinite-dimensional embedding-based dynamic output feedback. To our knowledge there does not exist any strategy that achieves the same result with a finite-dimensional time-independent approach.

The two following sections are devoted to the proof of Theorem 4.39. Let $\mathcal{K}_x$ be a compact subset of $\mathbb{R}^2$. Since Lemma 4.30 implies the statement *(i)* of Definition 4.20, it remains to show *(ii)* and *(iii)*. Let $R_0 > 0$ be such that $\mathcal{K}_x \subset \mathcal{B}_{\mathbb{R}^2}(0, R_0)$, $\mu \in (0, \mu_0)$ and $\delta \in (0, \delta_0)$ be as in Corollary 4.32, $\alpha > 0$, τ be as in (4.37) and π be as in (4.41). Let x_0 and $\hat{x}_0$ be in $\mathcal{K}_x$, $(x, \hat{z})$ be the corresponding solution of (4.23), $z = \tau(x)$, $\varepsilon = \hat{z} - z$ and $u = \phi(\pi(\hat{z})) + \delta\mathcal{N}^2(\hat{z} - \mathbb{1})$. Remark that

$$\begin{aligned}\|\varepsilon\|_X - \ell_\tau|x| &\leqslant \|\hat{z} - \mathbb{1}\|_X + \|z - \mathbb{1}\|_X - \ell_\tau|x| \\ &= \|\hat{z} - \mathbb{1}\|_X + \|\tau(x) - \tau(0)\|_X - \ell_\tau|x| \\ &\leqslant \|\hat{z} - \mathbb{1}\|_X\end{aligned}$$

and

$$\|\hat{z} - \mathbb{1}\|_X \leqslant \|\varepsilon\|_X + \|z - \mathbb{1}\|_X \leqslant \|\varepsilon\|_X + \ell_\tau|x|$$

where ℓ_τ is the Lipschitz constant of τ over $\mathcal{K}_x$. Hence proving statement *(ii)* of Definition 4.20 reduces to prove

(ii') For all $R_x, R_\varepsilon > 0$, there exist $r_x, r_\varepsilon > 0$ such that for all $(x_0, \hat{z}_0) \in \mathcal{K}_x \times \tau(\mathcal{K}_w)$, if $|x_0| < r_x$ and $\|\varepsilon_0\|_X < r_\varepsilon$, then $|x(t)| < R_x$ and $\|\varepsilon(t)\|_X < R_\varepsilon$ for all $t \geqslant 0$.

Since τ is continuous, if $x \to 0$, then $\tau(x) \to \mathbb{1}$, and, a fortiori, $\tau(x) \overset{w}{\rightharpoonup} \mathbb{1}$. Hence proving statement *(iii)* of Definition 4.20 reduces to prove

(iii') $x(t) \to 0$ and $\varepsilon(t) \overset{w}{\rightharpoonup} 0$ as t goes to infinity.

We prove *(ii')* in Section 4.2.2 and *(iii')* in Section 4.2.2.

Stability

Let $R_x, R_\varepsilon > 0$. We seek $r_x, r_\varepsilon > 0$ such that for all $(x_0, \hat{z}_0) \in \mathcal{K}_x \times \tau(\mathcal{K}_w)$, if $|x_0| < r_x$ and $\|\varepsilon_0\|_X < r_\varepsilon$, then $|x(t)| < R_x$ and $\|\varepsilon(t)\|_X < R_\varepsilon$ for all $t \geqslant 0$. Since $\|\varepsilon\|_X$ is non-increasing, choose $r_\varepsilon \leqslant R_\varepsilon$. Recall that x satisfies the following dynamics:

$$\dot{x} = (A + bK)x + bK(\pi(\hat{z}) - x) + \delta\mathcal{N}^2(\hat{z} - \mathbb{1})b.$$

Moreover,

$$|\pi(\hat{z}) - x| \leqslant \ell_\pi \|\varepsilon\|_X \leqslant \ell_\pi r_\varepsilon$$

where ℓ_π is the global Lipschitz constant of π and

$$\mathcal{N}(\hat{z}-\mathbb{1}) \leqslant \mathcal{N}(\varepsilon)+\mathcal{N}(z-\mathbb{1}) \leqslant \nu\left\|\varepsilon\right\|_X + \nu\left\|z-\mathbb{1}\right\|_X \leqslant \nu r_\varepsilon + \nu\ell_\tau|x|$$

where ℓ_τ is the Lipschitz constant of τ over $\mathcal{K}_x$. Since $A+bK$ is Hurwitz, there exists $P\in\mathbb{R}^{2\times 2}$ positive definite such that $P(A+bK)+(A+bK)'P<-2\mathrm{Id}_{\mathbb{R}^2}$. Denote by $\sigma_{\min}$ (resp. $\sigma_{\max}$) the smallest (resp. largest) eigenvalue of P. Then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x'Px &\leqslant -2|x|^2+2|x||Pb|\kappa|\pi(\hat{z})-x|+2|x||Pb|\delta\mathcal{N}^2(\hat{z}-\mathbb{1})\\ &\leqslant -2|x|^2+2\kappa|Pb|\ell_\pi r_\varepsilon|x|+4|Pb|\delta\nu^2(r_\varepsilon^2+\ell_\tau^2|x|^2)|x|.\end{aligned}$$

Set $r_x=\min\left(\frac{R_x}{2},\sqrt{\frac{\sigma_{\min}}{\sigma_{\max}}},\frac{1}{4|Pb|\delta\nu^2\ell_\tau^2}\right)$ and $r_\varepsilon=\min\left(R_\varepsilon,\frac{r_x}{8\kappa|Pb|\ell_\pi},\frac{\sqrt{r_x}}{4\nu\sqrt{\delta|Pb|}}\right)$. If $|x(t)|=r_x$ for some $t\in\mathbb{R}_+$, then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x'(t)Px(t) &\leqslant (-2+4|Pb|\delta\nu^2\ell_\tau^2 r_x)r_x^2+2\kappa|Pb|\ell_\pi r_\varepsilon r_x+4|Pb|\delta\nu^2 r_\varepsilon^2 r_x\\ &\leqslant -r_x^2+\frac{1}{4}r_x^2+\frac{1}{4}r_x^2\\ &<0.\end{aligned}$$

Hence, for all $t\in\mathbb{R}_+$, $|x(t)|\leqslant\sqrt{\frac{\sigma_{\max}}{\sigma_{\min}}}r_x<R_x$ and $|\varepsilon(t)|<r_\varepsilon\leqslant R_\varepsilon$.

Attractivity

Step 1: Show that $\varepsilon \overset{w}{\rightharpoonup} 0$. Let Ω be the set of limit points of $(\varepsilon(t))_{t\in\mathbb{R}_+}$ for the weak topology of X, that is, the set of points $\varepsilon^\star\in X$ such that there exists an increasing sequence $(t_n)_{n\in\mathbb{N}}$ such that $\varepsilon(t_n)\overset{w}{\rightharpoonup}\varepsilon^\star$ as $n\to+\infty$. According to (4.29), ε is bounded. Hence, by Kakutani's theorem, Ω is not empty. It remains to show that $\Omega=\{0\}$. Let $\varepsilon^\star\in\Omega$ and an increasing sequence $(t_n)_{n\in\mathbb{N}}$ such that $\varepsilon(t_n)\overset{w}{\rightharpoonup}\varepsilon^\star$ as $n\to+\infty$. Combining (4.29) and (4.28), $\hat{z}$ is also bounded. Then, after passing to a subsequence, we may assume that $\hat{z}(t_n)$ converges weakly to some $\hat{z}^\star\in X$. According to Corollary 4.32, $|u|$ is bounded by $\kappa\frac{j}{\mu}+16\nu^2\delta$. Again, after passing to a subsequence, $(u(\cdot+t_n))_{n\in\mathbb{N}}$ tends towards some $u^\star$ in the weak-$*$ topology of L^∞. If $u^\star\neq 0$, then it makes system (4.23) approximately observable. Hence, by [Cel+89, Theorem 7, Step 4] (see also Theorem 5.32), $\varepsilon^\star=0$. Now, let $\Delta>0$ be such that

$$\int_{t_n}^{t_n+\Delta} u(t)\psi(t-t_n)\mathrm{d}t \underset{n\to+\infty}{\longrightarrow} 0 \tag{4.50}$$

for all $\psi\in C^\infty((0,\Delta);\mathbb{R})$. Using the method of characteristics, one can show that for all $t,t'\in\mathbb{R}_+$ and almost all $s\in\mathbb{S}^1$,

$$z(t+t',s)=\mathcal{I}(t+t',t,s)z(t,s-t'). \tag{4.51}$$

where $\mathcal{I}(t+t',t,s)=e^{-i\mu\int_t^{t+t'}u(\sigma)\sin(s-\sigma)\mathrm{d}\sigma}$. Then, according to Duhamel's formula,

$$\hat{z}(t+t',s)=\mathcal{I}(t+t',t,s)\hat{z}(t,s-t')-\alpha\int_t^{t+t'}\mathcal{I}(t+t',\sigma,s)\Big(\left(\mathcal{C}^*\mathcal{C}\varepsilon(\sigma)\right)(s-t')\Big)\mathrm{d}\sigma. \tag{4.52}$$

Let $t' \in [0, \Delta]$. By (4.50), $\mathcal{I}(t_n + t', t_n, s) \to 1$ as $n \to +\infty$, uniformly in $s \in \mathbb{S}^1$. Then

$$\begin{aligned}\left\|\hat{z}(t_n + t', \cdot) - \hat{z}(t_n, \cdot - t')\right\|_X \leqslant & \sup_{s\in\mathbb{S}^1} |\mathcal{I}(t_n + t', t_n, s) - 1| \left\|\hat{z}(t_n)\right\|_X \\ & + \alpha \sup_{\sigma\in[t_n, t_n+t'],\, s\in\mathbb{S}^1} |\mathcal{I}(t_n + t', \sigma, s)| \left\| \int_{t_n}^{t_n+t'} \mathcal{C}^*\mathcal{C}\varepsilon(\sigma)\mathrm{d}\sigma \right\|_X \end{aligned} \tag{4.53}$$

tends towards 0 as n goes to $+\infty$ since $\hat{z}$ and ε are bounded, $t \mapsto |\mathcal{C}\varepsilon(t)|$ is integrable over $\mathbb{R}_+$ (see (4.29)) and $t' \leqslant \Delta$. Hence

$$\langle \hat{z}(t_n + t', \cdot), e_1\rangle_X \to \langle \hat{z}^\star(\cdot - t'), e_1\rangle_X = e^{it'} \langle \hat{z}^\star, e_1\rangle_X \tag{4.54}$$

and

$$\mathcal{N}^2(\hat{z}(t_n + t', \cdot) - \mathbb{1}) \to \mathcal{N}^2(\hat{z}^\star(\cdot - t') - \mathbb{1}) = \mathcal{N}^2(\hat{z}^\star - \mathbb{1}) := \mathcal{N}^2_\infty. \tag{4.55}$$

as n goes to $+\infty$. Passing to the limit in the expression of $u(t_n)$, we get the existence of $\mathcal{N}^2_\infty \in \mathbb{R}_+$ such that $\mathcal{N}^2(\hat{z}(t_n) - \mathbb{1}) \to \mathcal{N}^2_\infty$ and

$$\frac{1}{\Delta}\int_{t_n}^{t_n+\Delta} K\mathfrak{f}(e^{it}\langle \hat{z}^\star, e_1\rangle_X)\mathrm{d}t \underset{n\to+\infty}{\longrightarrow} -\delta\mathcal{N}^2_\infty. \tag{4.56}$$

For all $t \in \mathbb{R}$ and all $\zeta \in \mathcal{B}_{\mathbb{C}}(0, J_1(j))$, we have by (4.40), $K\mathfrak{f}(e^{it}\zeta) = K\mathfrak{R}(t)\mathfrak{f}(\zeta)$ where $\mathfrak{R}(t) = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix}$. Thus $\mathcal{N}^2_\infty = 0$, *i.e.* $\hat{z}(t_n) \overset{w}{\rightharpoonup} \mathbb{1}$. Combining it with (4.53), we have $\hat{z}(t_n + t') \overset{w}{\rightharpoonup} \mathbb{1}$ as n goes to $+\infty$. uniformly in $t' \in [0, \Delta]$. In particular, $\mathcal{C}\hat{z}(t_n + t') \to \mathcal{C}\tau(0)$. Since $\mathcal{C}\varepsilon \to 0$ by (4.29), we obtain $\mathcal{C}\tau(x(t_n + t')) \to \mathcal{C}\tau(0)$. Hence, by Assumption 4.33, $x(t_n) \to 0$, *i.e.*, $z(t_n) \to \mathbb{1}$. Thus $\varepsilon(t_n) \overset{w}{\rightharpoonup} 0$, *i.e.*, $\varepsilon^\star = 0$.

Step 2: Show that $x \to 0$. Recall that x satisfies the following dynamics:

$$\dot{x} = (A + bK)x + bK(\pi(\hat{z}) - x) + \delta\mathcal{N}^2(\hat{z} - \mathbb{1})b.$$

Since $A + bK$ is Hurwitz, there exists $P \in \mathbb{R}^{2\times 2}$ positive definite such that $P(A + bK) + (A + bK)'P < -2\mathrm{Id}_{\mathbb{R}^2}$. Set $V : \mathbb{R}^2 \ni x \mapsto x'Px$. Then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V(x) &\leqslant -2|x|^2 + 2|x||Pb|\kappa|\pi(\hat{z}) - x| + 2|x||Pb|\delta\mathcal{N}^2(\hat{z} - \mathbb{1}) \\ &\leqslant -2|x|^2 + 2|Pb|\frac{j}{\mu}\left(\kappa|\pi(\hat{z}) - x| + \delta\mathcal{N}^2(\hat{z} - \mathbb{1})\right).\end{aligned}$$

We have

$$\mathcal{N}(\hat{z} - \mathbb{1}) \leqslant \mathcal{N}(\varepsilon) + \mathcal{N}(z - \mathbb{1}) \leqslant \mathcal{N}(\varepsilon) + \nu\left\|z - \mathbb{1}\right\|_X \leqslant \mathcal{N}(\varepsilon) + \nu\ell_\tau|x|$$

where ℓ_τ is the Lipschitz constant of τ over $\mathcal{K}_x$. Hence, if $\delta \leqslant \frac{\mu}{4|Pb|j\nu^2\ell_\tau^2}$ (which we can assume without loss of generality by replacing δ_0 by $\min(\delta_0, \frac{\mu}{4|Pb|j\nu^2\ell_\tau^2})$, since diminishing δ), then

$$\frac{\mathrm{d}}{\mathrm{d}t}V(x) \leqslant -|x|^2 + 2|Pb|\frac{j}{\mu}\left(\kappa|\pi(\hat{z}) - x| + 2\delta\mathcal{N}^2(\varepsilon)\right).$$

Recall that $|x|$ and $|\pi(\hat{z})|$ are bounded by $\frac{j}{\mu}$. Moreover, $\mathcal{N}(\varepsilon(t)) \to 0$ as $t \to +\infty$ by Step 1, and $\pi(\hat{z}) - x \to 0$ as $t \to +\infty$ since π is a strong left-inverse of τ (see Corollary 4.16).

For all $r > 0$, set $D(r) = \{x \in \mathbb{R}^2 \ : \ V(x) \leqslant r\}$. In order to prove that $x \to 0$, we show that for all $r > 0$, there exists $T(r) \geqslant 0$ such that $x(t) \in D(r)$ for all $t \geqslant T(r)$. If $r > 0$ is such that $\bar{B}_{\mathbb{R}^2}(0, \frac{j}{\mu}) \subset D(r)$ then $T(r) = 0$ satisfies the statement. Let $0 < r < R$ be such that $\bar{B}_{\mathbb{R}^2}(0, \frac{j}{\mu}) \not\subset D(r)$ and $\bar{B}_{\mathbb{R}^2}(0, \frac{j}{\mu}) \subset D(R)$. Since $\mathcal{N}(\varepsilon(t)) \to 0$ and $\pi(\hat{z}(t)) - x(t) \to 0$, there exists $T_1(r) > 0$ such that for all $t \geqslant T_1(r)$, if $x(t) \notin D(r)$, then $\frac{\mathrm{d}}{\mathrm{d}t}V(x) < -\bar{m}$, for some $\bar{m} > 0$. First, this implies that if $x(t) \in D(r)$ for some $t \geqslant T_1(r)$, then $x(s) \in D(r)$ for all $s \geqslant t$. Second, for all $t \geqslant 0$,

$$\begin{aligned} V(x(T_1(r) + t)) &= V(x(T_1(r))) + \int_0^t \frac{\mathrm{d}}{\mathrm{d}t} V(x(T_1(r) + \tau))\mathrm{d}\tau \\ &\leqslant R - \bar{m}t \qquad \text{while } x(T_1(r) + t) \notin D(r). \end{aligned}$$

Set $T_2(r) = \frac{R-r}{m}$ and $T(r) = T_1(r) + T_2(r)$. Then for all $t \geqslant T(r)$, $x(t) \in D(r)$, which concludes the proof.

Part II

Infinite-dimensional observers

Chapter 5

Asymptotic Luenberger observers

La Nature est un temple où de vivants piliers
Laissent parfois sortir de confuses paroles ;

C. Baudelaire, *Les fleurs du mal*, "Correspondances"

Abstract. *In this chapter, we address the online state estimation problem for infinite-dimensional linear time-varying systems from the measurement of a linear output. We investigate the convergence properties of Luenberger observers for such systems. We recall some fundamental notions on evolution systems and different notions of observability. While strong exponential convergence generally holds for exactly observable systems, much less is known for approximate observability-like hypotheses on which we focus. Under a weak detectability assumption, we show that the observer estimates the so-called observable subspace of the system, at least in the weak topology of the state space. Additional conditions on the system are required to show strong convergence.*

Contents

Introduction

To analyze, monitor or control physical or biological phenomena, the first step is to provide a mathematical modeling in the form of mathematical equations that describe the evolution of the system variables. Some of these variables are accessible through measurement and others are not. A problem in control engineering is that of designing algorithms to provide real time estimates of the unmeasured data from the others, asymptotically converging to the actual data. These estimation algorithms are called state *observers* and can be found in many devices. The theory of linear autonomous finite-dimensional observers initiated by the seminal papers [Lue64, Lue71] of D. Luenberger is now well-understood. But the observer design problem remains an important challenge for the modern control community in the context of non-linear and/or infinite-dimensional systems, and the extension of the works of D. Luenberger to these systems is still an active research area.

In this chapter, we focus on infinite-dimensional time-varying linear systems with (potentially) infinite-dimensional measured output. First, in Section 5.1, we recall some notions of semigroup theory and evolution systems (mainly based on [Paz83]) to fix the functional setting of our analysis and ensure the well-posedness of the system. Then, the observer system is introduced in Section 5.2. It is based on an extension of the usual finite-dimensional asymptotic Luenberger observer, already used in [Sle72, Sle74, Cel+89, XLG95, Liu97] and in infinite-dimensional control theory in its dual form[1]. Two notions of observability (*approximate* and *exact*) are considered in Section 5.3. Except it is extended to the time-varying context, the terminology is borrowed from [TW09]. A vast literature focuses on autonomous exactly observable systems, for which strong exponential convergence of the observer generally holds (see [Liu97, Theorem 2.3], that summarizes and extends previously known results). Less is known for time-varying weakly observable systems, or for systems with non-full observable subspace. Section 5.4 contains the main results of the chapter. We show, by extending a result of [Cel+89], that weak convergence of the observer on the observable subspace can be expected, under a weak detectability assumption. Moreover, with additional hypotheses on the system, one can prove strong convergence by using very different tools inspired by [Hai14]. Main results are proved in Section 5.5.

5.1 Infinite-dimensional linear systems

Let X and Y be two Banach spaces. We consider time-varying linear systems of the form

$$\begin{cases} \dot{z} = A(t)z, & t \in \mathbb{R}_+ \\ y = Cz. \end{cases} \tag{5.1}$$

where z lying in X is the state of the system, y lying in Y is the output, $A(t) : \mathcal{D} \to X$ are linear operators defined on the same dense subspace $\mathcal{D} \subset X$ for all $t \in \mathbb{R}_+$ and $C : X \to Y$ is a linear operator. The state and output spaces X and Y may be infinite-dimensional. Before addressing the problem of observer design for such

[1] By *dual form*, we refer to the usual duality existing between observation and control problems for linear systems.

systems, we ensure the well-posedness of (5.1) by recalling some results of semigroup theory and on evolution equations.

5.1.1 Strongly continuous semigroups

First, let us consider the autonomous context. System (5.1) is said to be *autonomous* if there exists an operator $A : \mathcal{D} \to X$ such that $A(t) = A$ for all $t \in \mathbb{R}_+$. In this context, we rely on the theory of semigroups of linear operators.

Definition 5.1 (Strongly continuous semigroup)**.** A one-parameter family of operators $(\mathbb{T}(t))_{t\in\mathbb{R}_+}$ in $\mathscr{L}(X)$ is a *strongly continuous semigroup* on X if it satisfies the following properties:

(Semigroup property) $\mathbb{T}(0) = \mathrm{Id}_X$ and $\mathbb{T}(t+s) = \mathbb{T}(t)\mathbb{T}(s)$ for all $t, s \in \mathbb{R}_+$,

(Strong continuity) $\lim\limits_{\substack{t\to 0\\ t>0}} \mathbb{T}(t)z = z$ for all $z \in X$.

Definition 5.2 (Infinitesimal generator)**.** Let $\mathbb{T}$ be a strongly continuous semigroup on X. The linear operator $A : \mathcal{D} \to X$ defined by

$$\mathcal{D} = \left\{ z \in X \;\middle|\; \lim_{\substack{t\to 0\\ t>0}} \frac{\mathbb{T}(t)z - z}{t} \text{ exists} \right\},$$
$$Az = \lim_{\substack{t\to 0\\ t>0}} \frac{\mathbb{T}(t)z - z}{t}, \qquad \forall z \in \mathcal{D}.$$

is called the (infinitesimal) *generator* of $\mathbb{T}$.

Remark 5.3. If X is finite-dimensional, then every linear operator $A : X \to X$ is the generator of a strongly continuous semigroup $\mathbb{T}$ on X given by $\mathbb{T} : t \mapsto e^{tA}$.

Strongly continuous semigroups satisfy the following growth bound property.

Proposition 5.4 (Growth bound, [TW09, Proposition 2.1.2])**.** *Let $\mathbb{T}$ be a strongly continuous semigroup on X and*

$$\omega_0(\mathbb{T}) = \inf_{t\in\mathbb{R}_+^*} \frac{1}{t} \ln \|\mathbb{T}(t)\|_{\mathscr{L}(X)} \in \mathbb{R} \cup \{-\infty\} \tag{5.2}$$

Then, for all $\omega > \omega_0(\mathbb{T})$, there exists $M \in [1, +\infty)$ such that

$$\|\mathbb{T}(t)\|_{\mathscr{L}(X)} \leqslant M e^{\omega t}, \qquad \forall t \in \mathbb{R}_+. \tag{5.3}$$

Moreover, $\omega_0(\mathbb{T}) = \lim_{t\to+\infty} \frac{1}{t} \ln \|\mathbb{T}(t)\|_{\mathscr{L}(X)}$, and the flow

$$\begin{aligned} \varphi : \mathbb{R}_+ \times X &\longrightarrow X \\ (t, z) &\longmapsto \mathbb{T}(t)z \end{aligned}$$

is continuous.

Let $\rho(A) = \{\lambda \in \mathbb{C} \,:\, (\lambda \mathrm{Id}_X - A) \text{ is invertible and has bounded inverse}\}$ denote the resolvent set of A. Generators of strongly continuous semigroups are characterized by the Hille-Yosida theorem.

Theorem 5.5 (Hille-Yosida, see *e.g.,* [Paz83, Chapter 1, Theorem 3.1])**.** *Let $\mathcal{D}$ be a linear subspace of X and $A : \mathcal{D} \to X$ be a linear operator. Then A is the generator of a strongly continuous semigroup $\mathbb{T}$ satisfying* (5.3) *for some $M \geqslant 1$ and $\omega \in \mathbb{R}$ if and only if:*

- *A is closed and $\mathcal{D}$ is dense in X,*
- *$(\omega, +\infty) \subset \rho(A)$ and*

$$\left\|(\lambda \mathrm{Id}_X - A)^{-n}\right\|_{\mathscr{L}(X)} \leqslant \frac{M}{(\lambda - \omega)^n} \tag{5.4}$$

for all $\lambda > \omega$ and all positive integers n.

5.1.2 Evolution systems

The notion of strongly continuous semigroups is not sufficient to deal with time-varying systems. One needs to consider evolution systems. Adopt the convention that $[0, T] = \mathbb{R}_+$ if $T = +\infty$.

Definition 5.6 (Evolution systems)**.** Let $T \in \mathbb{R} \cup \{+\infty\}$. A two-parameter family $(\mathbb{T}(t, s))_{0 \leqslant s \leqslant t \leqslant T}$ of operators in $\mathscr{L}(X)$ is an *evolution system* on X over $[0, T]$ if it satisfies the following properties:

(Evolution property) $\mathbb{T}(t, t) = \mathrm{Id}_X$, $\mathbb{T}(t, s)\mathbb{T}(s, \tau) = \mathbb{T}(t, \tau)$ for $0 \leqslant \tau \leqslant s \leqslant t \leqslant T$,

(Strong continuity) $\lim\limits_{\substack{t \to s \\ t > s}} \mathbb{T}(t, s)z = \lim\limits_{\substack{s \to t \\ s < t}} \mathbb{T}(t, s)z = z$ for all $z \in X$.

Remark 5.7. If $(\mathbb{T}(t))_{t \in \mathbb{R}_+}$ is a strongly continuous semigroup on X, then $(\mathbb{T}(t - s))_{0 \leqslant s \leqslant t}$ is an evolution system on X over $\mathbb{R}_+$.

Before defining the notion of infinitesimal generator for evolution systems, let us define the notion of stability of a family of linear operators.

Definition 5.8 (Stable family)**.** Let $T \in \mathbb{R} \cup \{+\infty\}$. A family $(A(t))_{t \in [0,T]}$ of generators of strongly continuous semigroups on X is called *stable* if there exist $M \in [1, +\infty)$ and $\omega \in \mathbb{R}$ such that $(\omega, +\infty) \subset \rho(A(t))$ for all $t \in [0, T]$ and

$$\left\| \prod_{j=1}^{n} (\lambda \mathrm{Id}_X - A(t_j))^{-1} \right\|_{\mathscr{L}(X)} \leqslant \frac{M}{(\lambda - \omega)^n} \tag{5.5}$$

for all $\lambda > \omega$, all positive integers n and all non-decreasing sequences $(t_j)_{1 \leqslant j \leqslant n}$ in $[0, T]$.

Remark 5.9. According to the Hille-Yosida theorem, if there exists a constant $\omega \in \mathbb{R}$ such that $A(t)$ is the generator of a strongly continuous semigroup with growth bounds $M = 1$ and ω for all $t \in [0, T]$, then $(A(t))_{t \in [0,T]}$ is a stable family.

Stable families are robust to bounded perturbations. This property guarantees the well-posedness of the observer system of (5.1), defined in the next section.

Theorem 5.10 (Bounded perturbations, [Paz83, Chapter 3, Theorem 1.1])**.** *Let $(A(t))_{t\in[0,T]}$ be a stable family of generators on X. For all $B \in \mathscr{L}(X,Y)$, $(A(t)+B)_{t\in[0,T]}$ is a stable family of generators on X.*

Theorem 5.11 ([Paz83, Chapter 5, Theorem 4.8])**.** *Let $T \in \mathbb{R}\cup\{+\infty\}$ and $\mathcal{D}$ be a linear subspace of X. For all $t \in [0,T]$, let $A(t) : \mathcal{D} \to X$ be the generator of a strongly continuous semigroups on X. Assume the following hypotheses:*

- *$(A(t))_{t\in[0,T]}$ is a stable family for some constants $M \geqslant 1$ and $\omega \in \mathbb{R}$,*
- *$z \mapsto A(t)z$ is continuously differentiable in X for all $z \in \mathcal{D}$.*

Then there exists a unique evolution system $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t\leqslant T}$ on X over $[0,T]$ satisfying:

- *$\|\mathbb{T}(t,s)\|_{\mathscr{L}(X)} \leqslant Me^{\omega(t-s)}$ for all $0 \leqslant s \leqslant t \leqslant T$,*
- *$\lim\limits_{\substack{t\to s\\ t>s}} \frac{\mathbb{T}(t,s)z-z}{t} = A(s)z$ for all $z \in X$ and all $0 \leqslant s \leqslant T$,*
- *$\lim\limits_{\tau\to s} \frac{\mathbb{T}(t,\tau)z-\mathbb{T}(t,s)z}{t} = -\mathbb{T}(t,s)A(s)z$ for all $z \in X$ and all $0 \leqslant s \leqslant t \leqslant T$,*
- *$\mathbb{T}(t,s)\mathcal{D} \subset \mathcal{D}$ for all $0 \leqslant s \leqslant t \leqslant T$,*
- *$z \mapsto \mathbb{T}(t,s)z$ is continuous in $\mathcal{D}$ endowed with the graph norm $\|\cdot\|_{\mathcal{D}}^2 = \|\cdot\|_X^2 + \|A(0)\cdot\|_X^2$ for all $0 \leqslant s \leqslant t \leqslant T$.*

Definition 5.12 (Infinitesimal generator)**.** Under the assumptions of Theorem 5.11, the family $(A(t))_{t\in[0,T]}$ is called the (infinitesimal) *generator* of the evolution system $\mathbb{T}$.

Remark 5.13 (Hyperbolic context)**.** Assumptions of Theorem 5.11 are referred by [Paz83] as the *hyperbolic* case, by opposition to the *parabolic* case. Each case provides different assumptions ensuring the existence and uniqueness of an evolution system associated to a given family of operators. In [Paz83], another definition is given to take into account parabolic evolution systems.

Remark 5.14 (Autonomous context)**.** In the autonomous context, it is clear that an evolution system $\mathbb{T}$ generated by an operator A satisfies $\mathbb{T}(t,s) = \mathbb{T}(t-s,0)$ for all $t \geqslant s \geqslant 0$. By abuse of notation, the strongly continuous semigroup generated by A is also denoted by $\mathbb{T}$, so that $\mathbb{T}(t) = \mathbb{T}(t,0)$ for all $t \in \mathbb{R}_+$.

We conclude this section by ensuring the well-posedness of (5.1).

Theorem 5.15 (Well-posedness, [Paz83, Chapter 5])**.** *Let $T \in \mathbb{R}\cup\{+\infty\}$ and $z_0 \in X$. Consider the abstract Cauchy problem*

$$\begin{cases} \dot{z} = A(t)z, & \forall t \in [0,T], \\ z(0) = z_0. \end{cases} \tag{5.6}$$

If $(A(t))_{t\in\mathbb{R}_+}$ is the generator of an evolution system $\mathbb{T}$ on X over $[0,T]$, then (5.6) admits a unique solution $z \in C^0([0,T];X)$, which satisfies $z(t) = \mathbb{T}(t,0)z_0$ for all $t \in [0,T]$. Moreover, if $z_0 \in \mathcal{D}$, then $z \in C^0([0,T];\mathcal{D}) \cap C^1([0,T];X)$.

5.2 Luenberger observer

We address the problem of observer design for the observed system (5.1). Assume that $(A(t))_{t\in\mathbb{R}_+}$ is the generator of an evolution system $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t}$ on X over $\mathbb{R}_+$. Let $z_0 \in X$ and denote by (z,y) the unique solution of

$$\begin{cases} \dot{z} = A(t)z \\ z(0) = z_0 \end{cases}, \qquad y = Cz. \tag{5.7}$$

The goal is to find a new dynamical system fed by the output that asymptotically learns the state from the output dynamics. This issue was raised by D. Luenberger in his seminal paper [Lue64] in the context of finite-dimensional autonomous linear systems. In [Sle72, Sle74], J. Slemrod investigates the dual problem of stabilization in infinite-dimensional Hilbert spaces. We follow this path and introduce the usual infinite-dimensional version of the Luenberger observer.

From now on, assume that X and Y are real Hilbert spaces. All the results can easily be adapted to complex Hilbert spaces, but we prefer to restrict ourselves to real ones to simplify the presentation. Assume that C is a bounded linear operator, *i.e.,* $C \in \mathscr{L}(X,Y)$. We identify X and Y with their dual spaces via the canonical isometry, so that the adjoint of C, denoted by C^*, lies in $\mathscr{L}(Y,X)$.

Let $r > 0$ and $\hat{z}_0 \in X$. Consider the following Luenberger-like observer:

$$\begin{cases} \dot{\hat{z}} = A(t)\hat{z} - rC^*(C\hat{z} - y), \\ \hat{z}(0) = \hat{z}_0. \end{cases} \tag{5.8}$$

The parameter r is called the observer gain. Set $\varepsilon = \hat{z} - z$ and $\varepsilon_0 = \hat{z}_0 - z_0$. From now on, $\hat{z}$ represents the state estimation made by the observer system and ε the error between this estimation and the actual state of the system. Then $\hat{z}$ satisfies (5.8) if and only if ε satisfies

$$\begin{cases} \dot{\varepsilon} = (A(t) - rC^*C)\varepsilon, \\ \varepsilon(0) = \varepsilon_0. \end{cases} \tag{5.9}$$

Since $C \in \mathscr{L}(X,Y)$, Theorem 5.10 claims that $(A(t) - rC^*C)_{t\geqslant 0}$ is the generator of an evolution system on X over $\mathbb{R}_+$ denoted by $(\mathbb{S}(t,s))_{0\leqslant s\leqslant t}$. Hence, by Theorem 5.15, systems (5.8) and (5.9) have respectively a unique solution $\hat{z}$ and ε in $C^0([0,+\infty);X)$. Moreover, $\hat{z}(t) = \mathbb{T}(t,0)z_0 + \mathbb{S}(t,0)\varepsilon_0$ and $\varepsilon(t) = \mathbb{S}(t,0)\varepsilon_0$ for all $t \in [0,+\infty)$. If $(\hat{z}_0, \varepsilon_0) \in \mathcal{D}^2$, then $\hat{z}, \varepsilon \in C^0([0,+\infty);\mathcal{D}) \cap C^1([0,+\infty);X)$.

We are interested in the convergence properties of the state estimation $\hat{z}$ to the actual state z, *i.e.,* of the estimation error ε to 0. For any closed linear subspace $\mathcal{O}$ of X, let us denote by $\Pi_{\mathcal{O}} \in \mathscr{L}(X)$ the orthogonal projection onto $\mathcal{O}$.

Definition 5.16 (Asymptotic observer)**.** For any closed linear subspace $\mathcal{O}$ of X, (5.8) is said to be a strong (resp. weak) asymptotic $\mathcal{O}$-observer of (5.7) if and only if $\Pi_{\mathcal{O}}\mathbb{S}(t,0)\varepsilon_0 \to 0$ (resp. $\Pi_{\mathcal{O}}\mathbb{S}(t,0)\varepsilon_0 \overset{w}{\to} 0$) as $t \to +\infty$ for all $\varepsilon_0 \in X$. An X-observer is shortly called an observer.

5.3 Observability Gramian

A crucial operator to consider in order to investigate the convergence properties of a Luenberger-like observer is the so-called observability Gramian.

Definition 5.17 (Observability Gramian)**.** For all $t_0, \tau \in \mathbb{R}_+$, let us define

$$\begin{aligned} W(t_0,\tau) : X &\longrightarrow X \\ z_0 &\longmapsto \int_{t_0}^{t_0+\tau} \mathbb{T}(t,t_0)^* C^* C \mathbb{T}(t,t_0) z_0 \mathrm{d}t \end{aligned}$$

the *observability Gramian* of the pair $(\mathbb{T}, C)$.

The operator $W(t_0,\tau)$ is a bounded self-adjoint endomorphism of X, that characterizes the observability properties of (5.7). Let $M \in [1,+\infty)$ and $\omega \in \mathbb{R}$ be growth bounds of $\mathbb{T}$, *i.e.,* such that

$$\|\mathbb{T}(t,s)\|_{\mathscr{L}(X)} \leqslant M e^{\omega(t-s)}, \qquad \forall 0 \leqslant s \leqslant t, \tag{5.10}$$

Then W is continuous in $\mathscr{L}(X)$ with respect to (t_0,t) and we have $\|W(t_0,\tau)\|_{\mathscr{L}(X)} \leqslant \left(M e^{\omega\tau} \|C\|_{\mathscr{L}(X,Y)}\right)^2$.

Remark 5.18. In the autonomous context, $W(t_0,\tau) = W(0,\tau)$ for all $t_0, \tau \in \mathbb{R}_+$. Then, by abuse of notation, we shall write $W(\tau) := W(0,\tau)$.

Definition 5.19 (Observable subspace)**.** For all $\tau \in \mathbb{R}_+$, let

$$\mathcal{O}_\tau = (\ker W(0,\tau))^\perp . \tag{5.11}$$

be the *observable subspace* at time τ of the pair $(\mathbb{T}, C)$. Moreover, let

$$\mathcal{O} = \overline{\bigcup_{\tau>0} \mathcal{O}_\tau}. \tag{5.12}$$

be the *observable subspace* of the pair $(\mathbb{T}, C)$.

The sequence $(\mathcal{O}_\tau)_{\tau>0}$ is a non-decreasing sequence of closed linear subspaces. Hence, $\mathcal{O} = \overline{\lim_{\tau\to+\infty} \mathcal{O}_\tau}$, and it may be seen as the observable subspace in infinite time of the pair $(\mathbb{T}, C)$.

Remark 5.20 (Finite-dimensional autonomous context)**.** When (5.7) is autonomous and X and Y are finite-dimensional, we recover the usual definition (based on the observability matrix), properties (independent of observation time) and characterization (by the Hautus test) of the observable subspace:

$$\forall \tau \geqslant 0, \quad \mathcal{O}_\tau = \mathcal{O} = \left(\bigcap_{k=0}^{\dim X - 1} \ker C A^k \right)^\perp = \left(\operatorname{span} \bigcup_{\lambda \in \sigma(A)} \ker C \cap \ker(A - \lambda \mathrm{Id}) \right)^\perp .$$

For infinite-dimensional systems, there are several observability concepts that are not equivalent (see, *e.g.,* [TW09, Chapter 6] in the autonomous context), contrary to the case of finite-dimensional systems. In particular, one can distinguish the two following main concepts.

Definition 5.21 (Exact observability)**.** The pair $((A(t))_{t\in\mathbb{R}_+}, C)$ is said to be exactly observable on $(t_0, t_0+\tau) \subset \mathbb{R}_+$ if there exists $\delta > 0$ such that

$$\langle W(t_0,\tau) z_0, z_0 \rangle_X \geqslant \delta \|z_0\|_X^2 , \qquad \forall z_0 \in X. \tag{5.13}$$

Definition 5.22 (Approximate observability)**.** The pair $((A(t))_{t\in\mathbb{R}_+}, C)$ is said to be approximately observable on $(t_0, t_0+\tau) \subset \mathbb{R}_+$ if $W(t_0,\tau)$ is injective.

Clearly, the exact observability of a pair on some time interval implies its approximate observability, and the concepts are equivalent in finite-dimension. The approximate observability on $(0,\tau)$ is equivalent to the fact that $\mathcal{O}_\tau$, the observable subspace in time τ of $(\mathbb{T}, C)$, is equal to the whole state space X.

5.4 Observer convergence

Exact observability has already been deeply investigated, in particular in the autonomous context (see, *e.g.,* [Sle72,Sle74,Rus78,Liu97,Urq05]). Under this assumption, and if A is skew-adjoint, then (5.8) is a strong *exponential* asymptotic observer. More precisely, we have the following result.

Theorem 5.23 (Corollary of [Liu97, Theorem 2.3])**.** *Assume that* (5.7) *is autonomous and A is skew-adjoint (i.e., $A^* = -A$). The pair (A, C) is exactly observable on some finite time interval if and only if, for every positive-definite self-adjoint operator $S \in \mathscr{L}(Y)$, $A - C^*SC$ is the generator of an exponentially stable strongly continuous semigroup $\mathbb{S}$ on X, that is, $\omega_0(\mathbb{S}) < 0$.*

Actually, [Liu97, Theorem 2.3] is more general and summarize previously known results (see [Sle72, Sle74, Rus78]). Let us also state the following important result on strongly continuous groups (see Definition 6.1.1).

Theorem 5.24 (Corollary of [Urq05])**.** *Assume that* (5.7) *is autonomous and A is the generator of a strongly continuous group $\mathbb{T}$ on X. If $(-A, C)$ is exactly observable on some finite time interval, then for all $\omega > 0$ there exists a coercive self-adjoint operator $S \in \mathscr{L}(X)$ such that $A - S^{-1}C^*C$ is the generator of an exponentially stable strongly continuous semigroup $\mathbb{S}$ on X such that $\omega_0(\mathbb{S}) \leqslant \omega$.*

Similarly, strong exponential observers can be obtained for *uniformly* exactly observable systems, *i.e.,* if $((A(t))_{t\in[0,T]}, C)$ is exactly observable on $(t_0, t_0 + \tau)$ for some fixed constants $\delta > 0$ and $\tau > 0$ for all $t_0 \in \mathbb{R}_+$ (see Remark 5.42). On the contrary, when approximate observability holds, one can rather expect weak asymptotic observers, as in [Cel+89, XLG95].

Theorem 5.25 ([Cel+89, Theorem 7])**.** *Assume that Y is one-dimensional (i.e., C is a bounded linear form) and*

$$A(t) = A_0 + \sum_{i=1}^{p} u_i(t)A_i \tag{5.14}$$

where each $A_i : \mathcal{D} \to X$ is a skew-adjoint operator and $u_i : \mathbb{R}_+ \to \mathbb{R}$ is bounded. Assume that there exists an increasing positive sequence $(t_n)_{n\geqslant 0} \to +\infty$ such that

(i) the sequence $(t_{n+1} - t_n)_{n\in\mathbb{N}}$ is bounded,

(ii) for $1 \leqslant i \leqslant p$, $u_i(t_n + \cdot) \to u_{i,\infty}$ in the weak-topology for some bounded $u_{i,\infty} : \mathbb{R}_+ \to \mathbb{R}$ as $n \to +\infty$,*

(iii) the pair (A_∞, C) is approximately observable on some finite time-interval, where $A_\infty = A_0 + \sum_{i=1}^{p} u_{i,\infty}(t)A_i$.

Then (5.8) *is a weak asymptotic observer of* (5.7) *for all $r > 0$.*

The inputs $(u_i)_{1\leqslant i\leqslant p}$ satisfying the assumptions of Theorem 5.25 are called *regularly persistent* in [Cel+89]. Persistency refers to *(ii)-(iii)*, *i.e.,* convergence of a subsequence towards an input $(u_{i,\infty})_{1\leqslant i\leqslant p}$ (called *universal*) making the system approximately observable, while regularity refers to *(i)*, *i.e.,* boundedness of the

subsequence's times steps. Following the path of [Cel+89], we consider only approximate observability-like assumptions. We aim to relax two of the hypotheses of [Cel+89]: the particular form (5.14) of the generator and the approximate observability hypothesis. Doing so, we obtain observers converging on some subspaces of X. Moreover, with additional properties on $\mathbb{S}$, we obtain strong (non-exponential) observers by adapting a result of [Hai14] originally used for back and forth observers (see Chapter 6). All the results rely on the following weak detectability hypothesis.

5.4.1 Weak detectability

Definition 5.26 (Weak detectability)**.** Let $T \in \mathbb{R}_+ \cup \{+\infty\}$. Then $((A(t))_{t\in[0,T]}, C)$ is said to be *μ-weakly detectable* for some $\mu \geqslant 0$ if for all $t \in [0,T]$,

$$\langle A(t)z, z\rangle_X \leqslant \mu \|Cz\|_Y^2, \qquad \forall z \in \mathcal{D}. \tag{5.15}$$

Let us investigate the notion of weak detectability in the following remarks.

Remark 5.27. A pair $((A(t))_{t\geqslant 0}, C)$ is usually said to be *detectable* if for all pairs of trajectories (z_1, z_2) of (5.7), if $Cz_1(t) = Cz_2(t)$ for all $t \geqslant 0$, then $(z_1(t) - z_2(t)) \to 0$ as $t \to +\infty$. This definition is equivalent to the usual definition of detectability in finite-dimension. However, several definitions may be chosen in infinite-dimension, that are all equivalent in finite-dimension. In this remark, we show how (5.15) may be seen as a weak detectability hypothesis (although it is not implied by the finite-dimensional notion). Let $((A(t))_{t\in[0,T]}, C)$ be μ-weakly detectable for some $\mu \geqslant 0$. Then Lemma 5.39, that is proved in Section 5.5.1, states that $\mathbb{S}$ is a contraction evolution system, *i.e.,* $\|\mathbb{S}(t,s)\|_{\mathscr{L}(X)} \leqslant 1$ for $0 \leqslant s \leqslant t \leqslant T$, provided that r is selected greater that μ. Consider (z_1, z_2) two trajectories of (5.7) such that $Cz_1(t) = Cz_2(t)$ for all $t \in [0,T]$. Then z_1 and z_2 are also trajectories of (5.8), and $z_1 - z_2$ is a trajectory of (5.9). Therefore, for all $0 \leqslant s \leqslant t \leqslant T$,

$$\|z_1(t) - z_2(t)\|_X = \|\mathbb{S}(t,s)(z_1(s) - z_2(s))\|_X \leqslant \|z_1(s) - z_2(s)\|_X .$$

Hence, $[0,T] \ni t \mapsto \|z_1(t) - z_2(t)\|_X$ is non-increasing. Thus, while detectability means that indistinguishable trajectories converges one to the other, weak detectability rather means that the distance between two indistinguishable trajectories is non-increasing. However, note that detectability does not imply weak detectability.

Remark 5.28. When stating that a pair $((A(t))_{t\in[0,T]}, C)$ is μ-weakly detectable, we actually state that the pair is *uniformly* weakly detectable, in the sense that the detectability constant μ is independent of the time $t \in [0,T]$. Therefore, this assumption is stronger than the weak detectability of each pair $(A(t), C)$ for $t \in [0,T]$. If $T < +\infty$ or $t \mapsto A(t)$ is periodic, then the two statements are equivalent, due to the continuity of $[0,T] \ni t \mapsto A(t)z$ for all $z \in \mathcal{D}$.

Remark 5.29. If $A(t)$ is a *dissipative* operator for all $t \in [0,T]$, that is,

$$\langle A(t)z, z\rangle_X \leqslant 0, \qquad \forall t \in [0,T], \tag{5.16}$$

then the pair $((A(t))_{t\in[0,T]}, C)$ is 0-weakly detectable for any output operator $C \in \mathscr{L}(X,Y)$. This assumption is the one usually made in the literature to prove the weak convergence of a Luenberger-like observer in infinite-dimension (see [Sle74, Cel+89, XLG95]). Therefore, the weak detectability hypothesis may be seen as a weakening of the dissipativity hypothesis, relying on the output operator.

Remark 5.30. If there exist a bounded self-adjoint $P \in \mathscr{L}(X)$ and two constants $p > 0$ and $\mu \geqslant 0$ such that

$$\langle Px, x\rangle_X \geqslant p \|x\|_X^2, \quad \langle PA(t)x, x\rangle_X \leqslant \mu \|Cx\|_Y^2, \quad \forall x \in \mathcal{D},\ \forall t \in [0,T], \tag{5.17}$$

then the pair $((A(t))_{t\in[0,T]}, C)$ is μ-weakly detectable provided one endows the space X with the inner product $\langle P\cdot, \cdot\rangle_X$. Note that in this case the operator C^* is the adjoint of $C \in \mathscr{L}(X,Y)$ with respect to this new inner product, *i.e.,* $\langle C\cdot, \cdot\rangle_Y = \langle P\cdot, C^*\cdot\rangle_X$. If the coercivity assumption $\langle Px, x\rangle_X \geqslant p\|x\|_X^2$ is not satisfied, but P is still positive-definite, then it is still possible to apply the result in the Hilbert space X endowed with the new inner product $\langle P\cdot, \cdot\rangle_X$, but this topology is coarser than the topology associated to $\langle \cdot, \cdot\rangle_X$. Actually, if X is finite-dimensional, the existence of P (which is then a positive-definite matrix) such that (5.17) holds is a necessary condition for the existence of an asymptotic observer.

Remark 5.31. The parameter $r > 0$ is the observer gain. If $A(t)$ is a *dissipative* operator for all $t \in [0,T]$, then the convergence results hold for all gains $r > 0$. Otherwise, the gain must be chosen large enough in order to deal with the lack of dissipativity, which is replaced by weak detectability. Obviously, if a pair is μ-weakly detectable for some $\mu \geqslant 0$, then it is also λ-weakly detectable for all $\lambda \geqslant \mu$. This class of observers is called *observers with infinite gain margin* since r can be chosen as large as requested.

In the two following sections, we state the main results of this chapter.

5.4.2 Weak asymptotic observer

Theorem 5.32. *Assume that $((A(t))_{t\geqslant 0}, C)$ is μ-weakly detectable and $r > \mu$. Assume that there exist an increasing positive sequence $(t_n)_{n\geqslant 0} \to +\infty$ and an evolution system $(\mathbb{T}_\infty(t,s))_{0\leqslant s\leqslant t}$ on X such that for all $\tau \geqslant 0$,*

$$\|\mathbb{T}(t_n + t, t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \to 0 \text{ as } n \to +\infty \text{ uniformly in } t \in [0,\tau]. \tag{5.18}$$

Let $\mathcal{O}$ be the observable subspace of the pair $(\mathbb{T}_\infty, C)$. Then for all $\varepsilon_0 \in X$,

$$\Pi_{\mathcal{O}}\mathbb{S}(t_n, 0)\varepsilon_0 \xrightarrow[n\to+\infty]{w} 0. \tag{5.19}$$

Moreover, if $(t_{n+1} - t_n)_{n\geqslant 0}$ is bounded and $\mathcal{O} = X$, then (5.8) is a weak asymptotic observer of (5.7).

The proof of Theorem 5.32 is given in Section 5.5.1 and follows the steps of [Cel+89]. Note that Theorem 5.25 is a direct corollary of Theorem 5.32. In the autonomous context, every increasing positive sequence $(t_n)_{n\geqslant 0} \to +\infty$ is such that $\mathbb{T}(t_n + t, t_n) = \mathbb{T}(t)$ for all $t \geqslant 0$. Hence (5.19) holds for all such sequences $(t_n)_{n\geqslant 0}$ and with $\mathcal{O}$ the observable subspace of $(\mathbb{T}, C)$. This leads to the following corollary.

Corollary 5.33. *Suppose that (5.7) is autonomous, (A, C) is μ-weakly detectable and $r > \mu$. Let $\mathcal{O}$ be the observable subspace of $(\mathbb{T}, C)$. Then, (5.8) is a weak asymptotic $\mathcal{O}$-observer of (5.7).*

Remark 5.34. One of the assumptions is the existence of an increasing positive sequence $(t_n)_{n\geqslant 0} \to +\infty$ and an evolution system $(\mathbb{T}_\infty(t,s))_{0\leqslant s\leqslant t}$ on X such that $\|\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$ uniformly in $t \in [0,\tau]$ for all $\tau \geqslant 0$. Checking this hypothesis may be a difficult task in general. However, [IK02, Theorem 10.2] states sufficient conditions depending only on the family of generators $(A(t))_{t\geqslant 0}$ for the existence of such a sequence. In Section 6.4, we show how to check this property on a time-varying one-dimensional transport equation with periodic boundary conditions.

5.4.3 Strong asymptotic observer

With additional hypotheses on $\mathbb{S}$, we obtain the strong convergence of the observer.

Theorem 5.35. *Assume that there exists $\tau > 0$ such that $t \mapsto A(t)$ is τ-periodic. Let $\mathcal{O}_\tau$ be the observable subspace at time τ of the pair $(\mathbb{T}, C)$.*

(i) Suppose that $((A(t))_{t\geqslant 0}, C)$ is μ-weakly detectable and $r > \mu$. Assume that $\mathbb{S}(\tau,0)$ is normal and bounded from below. If $\mathcal{O}_\tau = X$, then (5.8) is a strong asymptotic observer of (5.7).

(ii) Suppose that $A(t)$ is skew-adjoint for all $t \in \mathbb{R}_+$ and $\mathbb{S}(\tau,0)$ is normal. If $\mathbb{T}(t,0)\mathcal{O}_\tau \subset \mathcal{O}_\tau$ and $\mathbb{T}(t,0)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$ for all $t \in [0,\tau]$, then (5.8) is a strong asymptotic $\mathcal{O}_\tau$-observer of (5.7) for all $r > 0$.

The proof of Theorem 5.35 is given in Section 5.5.2 and is an adaptation of [Hai14, Theorem 1.1.2] to the asymptotic time-varying context.

5.5 Proofs of the results

This section is devoted to the proofs of the results stated in Section 5.4. Throughout the section, $(A(t))_{t\in\mathbb{R}_+}$ is the generator of an evolution system $\mathbb{T}$ on X over $\mathbb{R}_+$ (in the sense of Definition 5.12), $C \in \mathscr{L}(X,Y)$ and $\mathbb{S}$ the evolution system generated by $(A(t) - rC^*C)_{t\in\mathbb{R}_+}$ (see Section 5.2). The following important remark allows us to reformulate the weak convergence results.

Remark 5.36. For any closed linear subspace $\mathcal{O}$ of X and any sequence $(x_n)_{n\geqslant 0}$ in X, recall that $\Pi_\mathcal{O} x_n \overset{w}{\to} 0$ as $n \to +\infty$ if and only if, for all $\psi \in X$, $\langle \Pi_\mathcal{O} x_n, \psi\rangle_X \to 0$. As an orthogonal projection, $\Pi_\mathcal{O}$ is a self-adjoint operator, *i.e.*, $\Pi_\mathcal{O} = \Pi_\mathcal{O}^*$, and $\operatorname{Im}\Pi_\mathcal{O} = \mathcal{O}$. Hence, $\Pi_\mathcal{O} x_n \overset{w}{\to} 0$ as $n \to +\infty$ if and only if $\langle x_n, \psi\rangle_X \to 0$ for all $\psi \in \mathcal{O}$.

All the weak convergence results are proved in the following in accordance with this remark. For example, to prove that (5.8) is a weak asymptotic $\mathcal{O}$-observer, we prove that $\langle \mathbb{S}(t,0)\varepsilon_0, \psi\rangle_X \to 0$ as $t \to +\infty$ for all $\varepsilon_0 \in X$ and all $\psi \in \mathcal{O}$.

Lemma 5.37. *Let $(L_n)_{n\in\mathbb{N}}$ be a bounded sequence of operators in $\mathscr{L}(X)$, i.e., such that $\sup_{n\in\mathbb{N}} \|L_n\|_{\mathscr{L}(X)} \leqslant M_L$ for some $M_L > 0$. Let $U, V \subset X$.*

(i) If

$$L_n\varepsilon_0 \underset{n\to+\infty}{\longrightarrow} 0, \qquad \forall \varepsilon_0 \in U$$

then

$$L_n\varepsilon_0 \underset{n\to+\infty}{\longrightarrow} 0, \qquad \forall \varepsilon_0 \in \overline{U}.$$

(ii) If

$$\langle L_n\varepsilon_0, \psi\rangle_X \underset{n\to+\infty}{\longrightarrow} 0, \qquad \forall \varepsilon_0 \in U, \quad \forall \psi \in V,$$

then

$$\langle L_n\varepsilon_0, \psi\rangle_X \underset{n\to+\infty}{\longrightarrow} 0, \qquad \forall \varepsilon_0 \in \overline{U}, \quad \forall \psi \in \overline{V}.$$

Proof of (i). Let M_L be a bound of the sequence $(L_n)_{n\in\mathbb{N}}$ in $\mathscr{L}(X)$. Let $\varepsilon_0 \in \overline{U}$ and $\eta > 0$. Then there exists $\tilde{\varepsilon}_0 \in U$ such that $\|\varepsilon_0 - \tilde{\varepsilon}_0\|_X \leqslant \eta$. Moreover, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, $\|L_n\tilde{\varepsilon}_0\|_X \leqslant \eta$. Then, for all $n \geqslant N$,

$$\|L_n\varepsilon_0\|_X \leqslant \|L_n\tilde{\varepsilon}_0\|_X + M_L\,\|\tilde{\varepsilon}_0 - \varepsilon_0\|_X \leqslant (1 + M_L)\eta$$

since $\|L_n\|_{\mathscr{L}(X)} \leqslant M_L$. Hence $L_n\varepsilon_0 \to 0$ as $n \to +\infty$. ■

Proof of (ii). Let $\varepsilon_0 \in \overline{U}$, $\psi \in \overline{V}$ and $\eta > 0$. Then there exist $\tilde{\varepsilon}_0 \in U$ and $\tilde{\psi} \in V$ such that $\|\varepsilon_0 - \tilde{\varepsilon}_0\|_X \leqslant \eta$ and $\left\|\psi - \tilde{\psi}\right\|_X \leqslant \eta$. Moreover, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, $\left|\left\langle L_n\tilde{\varepsilon}_0, \tilde{\psi}\right\rangle_X\right| \leqslant \eta$. Then, for all $n \geqslant N$,

$$\begin{aligned}
|\langle L_n\varepsilon_0, \psi\rangle_X| &\leqslant \left|\left\langle L_n\tilde{\varepsilon}_0, \tilde{\psi}\right\rangle_X\right| + \left|\left\langle L_n(\varepsilon_0 - \tilde{\varepsilon}_0), \tilde{\psi}\right\rangle_X\right| \\
&\quad + \left|\left\langle L_n\tilde{\varepsilon}_0, \psi - \tilde{\psi}\right\rangle_X\right| + \left|\left\langle L_n(\varepsilon_0 - \tilde{\varepsilon}_0), \psi - \tilde{\psi}\right\rangle_X\right| \\
&\leqslant \left(1 + M_L\left\|\tilde{\psi}\right\|_X + M_L\,\|\tilde{\varepsilon}_0\|_X + M_L\eta\right)\eta
\end{aligned}$$

since $\|L_n\|_{\mathscr{L}(X)} \leqslant M_L$. Hence $\langle L_n\varepsilon_0, \psi\rangle_X \to 0$ as $n \to +\infty$. ■

Remark 5.38. An operator $L \in \mathscr{L}(X)$ is said to be a contraction if $\|L\|_{\mathscr{L}(X)} \leqslant 1$. If $(L_n)_{n\in\mathbb{N}}$ is a sequence of contractions in $\mathscr{L}(X)$, then it is uniformly bounded by 1, hence Lemma 5.37 does apply. In the following sections, we use Lemma 5.37 only on sequences of contractions.

5.5.1 Proof of Theorem 5.32

The proof relies on the two following lemmas. The first one shows how the weak detectability is used in the proof, while the second one states a continuity property of the observability Gramian. We adapt the steps of the proof of [Cel+89, Theorem 7].

Lemma 5.39. *If $((A(t))_{t\geqslant 0}, C)$ is μ-weakly detectable and $r > \mu$, then $\mathbb{S}$ is a contraction evolution system, that is,*

$$\|\mathbb{S}(t,s)\|_{\mathscr{L}(X)} \leqslant 1, \qquad \forall t \geqslant s \geqslant 0. \tag{5.20}$$

Proof. Since $\mathcal{D}$ is dense in X, it is sufficient to show that

$$\|\mathbb{S}(t,t_0)\varepsilon_0\|_X \leqslant \|\varepsilon_0\|_X \tag{5.21}$$

for all $\varepsilon_0 \in \mathcal{D}$ and all $t \geqslant t_0 \geqslant 0$. Let $t_0 \geqslant 0$, $\varepsilon_0 \in \mathcal{D}$ and set $\varepsilon(t) = \mathbb{S}(t,t_0)\varepsilon_0$ for all $t \geqslant t_0$. Then $\varepsilon \in C^1([0,+\infty),X)$ and for all $t \geqslant t_0$,

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\varepsilon(t)\|_X^2 &= \langle \varepsilon(t), \dot{\varepsilon}(t)\rangle_X \\ &= \langle \varepsilon(t), A(t)\varepsilon(t)\rangle_X - r\,\langle \varepsilon(t), C^*C\varepsilon(t)\rangle_X \\ &\leqslant -(r-\mu)\,\|C\varepsilon(t)\|_Y^2 \quad \text{(since } ((A(t))_{t\geqslant 0}, C) \text{ is } \mu\text{-weakly detectable)} \\ &\leqslant 0\end{aligned} \tag{5.22}$$

since $r > \mu$. Hence $[t_0,+\infty) \ni t \mapsto \|\varepsilon(t)\|_X^2$ is non increasing, which yields (5.21) since $\varepsilon(t_0) = \varepsilon_0$. ■

Remark 5.40. Thanks to Lemma 5.39, we known, without using any observability hypothesis, that the observer error is non-increasing. This is a crucial aspect of the proof. Roughly speaking, even if the system has very poor observability properties on some bounded time intervals, the observer accuracy will not be affected, and will simply wait for forthcoming enriched observability properties.

Lemma 5.41. *If there exist an increasing positive sequence $(t_n)_{n\geqslant 0} \to +\infty$ and an evolution system $(\mathbb{T}_\infty(t,s))_{0\leqslant s\leqslant t}$ on X such that $\|\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$ for all $t \geqslant 0$, then $\|W(t_n,\tau) - W_\infty(0,\tau)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$.*

Proof. For all $z_0 \in X$,

$$\begin{aligned}&\|(W(t_n,\tau) - W_\infty(0,\tau))z_0\|_X \\ &\quad\leqslant \int_0^\tau \|\mathbb{T}(t_n+t,t_n)^*C^*C\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)^*C^*C\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)}\,\|z_0\|_X\,\mathrm{d}t \\ &\quad\leqslant \tau\,\|z_0\|_X \sup_{t\in[0,\tau]} \|\mathbb{T}(t_n+t,t_n)^*C^*C\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)^*C^*C\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)}.\end{aligned}$$

For all $t \in [0,\tau]$,

$$\begin{aligned}&\|\mathbb{T}(t_n+t,t_n)^*C^*C\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)^*C^*C\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \\ &\qquad\leqslant \|(\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0))^*\|_{\mathscr{L}(X)}\|C^*C\mathbb{T}(t_n+t,t_n)\|_{\mathscr{L}(X)} \\ &\qquad\quad + \|\mathbb{T}_\infty(t,0)^*C^*C\|_{\mathscr{L}(X)}\|\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \\ &\qquad\leqslant \|\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)}\|C\|_{\mathscr{L}(X,Y)}^2\Big(\|\mathbb{T}(t_n+t,t_n)\|_{\mathscr{L}(X)} \\ &\qquad\qquad + \|\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)}\Big)\end{aligned}$$

Recall that $\|\mathbb{T}(t_n+t,t_n)\|_{\mathscr{L}(X)} \leqslant Me^{\omega t}$ by (5.10) and that (5.18) implies $\|\mathbb{T}(t_n+t,t_n)\|_{\mathscr{L}(X)} \to \|\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)}$ as $n \to +\infty$. Hence, we also have $\|\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \leqslant Me^{\omega t}$. Thus,

$$\begin{aligned}&\|\mathbb{T}(t_n+t,t_n)^*C^*C\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)^*C^*C\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \\ &\qquad\leqslant 2\|C\|_{\mathscr{L}(X,Y)}^2 Me^{\omega t}\|\mathbb{T}(t_n+t,t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)}.\end{aligned}$$

Hence, according to (5.18), $\|W(t_n,\tau) - W_\infty(0,\tau)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$. ■

With these lemmas in mind, we are now able to prove the main Theorem 5.32.

Proof of Theorem 5.32. According to Lemma 5.39, $\mathbb{S}$ is a contraction evolution system. Hence, applying Lemma 5.37 (ii) with $L_n = \mathbb{S}(t_n, 0)$ for $n \in \mathbb{N}$, it is sufficient to show (5.19) for all $\psi \in \cup_{\tau \geqslant 0} (\ker W_\infty(0,\tau))^\perp$ and all $\varepsilon_0 \in \mathcal{D}$ since $\mathcal{D}$ is dense is X. Let $\varepsilon_0 \in \mathcal{D}$ and set $\varepsilon(t) = \mathbb{S}(t,0)\varepsilon_0$ for all $t \geqslant 0$. Since $\mathbb{S}$ is a contraction, $\|\varepsilon\|_X$ is non-increasing and whence converges to a finite limit. Equation (5.22) yields for all $t_0, \tau \geqslant 0$,

$$\int_{t_0}^{t_0+\tau} \|C\varepsilon(t)\|_Y^2 \,\mathrm{d}t \leqslant \frac{1}{2(r-\mu)} \left(\|\varepsilon(t_0)\|_X^2 - \|\varepsilon(t_0+\tau)\|_X^2 \right). \tag{5.23}$$

Hence,

$$\int_{t_0}^{t_0+\tau} \|C\varepsilon(t)\|_Y^2 \,\mathrm{d}t \underset{t_0 \to +\infty}{\longrightarrow} 0. \tag{5.24}$$

According to the Duhamel's formula, for all $t \geqslant t_0 \geqslant 0$,

$$\varepsilon(t) = \mathbb{T}(t,t_0)\varepsilon(t_0) - r \int_{t_0}^{t} \mathbb{T}(t,s) C^* C \varepsilon(s) \mathrm{d}s. \tag{5.25}$$

Then

$$\begin{aligned} W(t_0,\tau)\varepsilon(t_0) &= \int_{t_0}^{t_0+\tau} \mathbb{T}(t,t_0)^* C^* C \mathbb{T}(t,t_0)\varepsilon(t_0) \mathrm{d}t \\ &= \int_{t_0}^{t_0+\tau} \mathbb{T}(t,t_0)^* C^* C \varepsilon(t) \mathrm{d}t \\ &\quad + r \int_{t_0}^{t_0+\tau} \mathbb{T}(t,t_0)^* C^* C \int_{t_0}^{t} \mathbb{T}(t,s) C^* C \varepsilon(s) \mathrm{d}s \mathrm{d}t. \end{aligned}$$

By (5.10) and because C is bounded, we have

$$\begin{aligned} \|W(t_0,\tau)\varepsilon(t_0)\|_X &\leqslant M e^{\omega\tau} \|C\|_{\mathscr{L}(X,Y)} \int_{t_0}^{t_0+\tau} \|C\varepsilon(t)\|_Y \,\mathrm{d}t \\ &\quad + r\tau M^2 e^{2\omega\tau} \|C\|_{\mathscr{L}(X,Y)}^3 \int_{t_0}^{t_0+\tau} \|C\varepsilon(t)\|_Y \,\mathrm{d}t. \end{aligned}$$

Hence

$$W(t_0,\tau)\varepsilon(t_0) \underset{t_0 \to +\infty}{\longrightarrow} 0, \qquad \forall \tau \geqslant 0. \tag{5.26}$$

Remark 5.42. From (5.26), we see that a uniform exact observability assumption would imply strong convergence of ε towards 0. Indeed, if $\langle W(t_0,\tau)\varepsilon(t_0), \varepsilon(t_0)\rangle_X \geqslant \delta \|\varepsilon(t_0)\|_X^2$ for some $\tau > 0$ and $\delta > 0$ uniformly in $t_0 \in \mathbb{R}_+$, then $\|\varepsilon(t_0)\|_X^2 \to 0$ as $t_0 \to +\infty$. Moreover, the speed of convergence is exponential, with arbitrary decay rate by tuning the observer gain r. Indeed, consider $\tilde{\varepsilon} = e^{\lambda t}\varepsilon$ for some $\lambda > 0$. Then $\dot{\tilde{\varepsilon}} = (A - rC^*C + \lambda \mathrm{Id}_X)\tilde{\varepsilon}$. Hence, computing as in (5.22), we get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} \|\tilde{\varepsilon}(t)\|_X^2 \leqslant -(r-\mu) \|C\tilde{\varepsilon}(t)\|_Y^2 + \lambda \|\tilde{\varepsilon}(t)\|_X^2$$

Integrating on $[t_0, t_0+\tau]$ and using the uniform exact observability assumption, we have

$$\frac{1}{2} \left(\|\tilde{\varepsilon}(t_0+\tau)\|_X^2 - \|\tilde{\varepsilon}(t_0)\|_X^2 \right) \leqslant -(r-\mu)\delta \|\tilde{\varepsilon}(t_0)\|_X^2 + \lambda e^{2\lambda\tau} \|\tilde{\varepsilon}(t_0)\|_X^2$$

Hence, if $(r-\mu)\delta \geqslant \lambda e^{2\lambda\tau}$, $\|\tilde{\varepsilon}\|_X \leqslant M$ for some constant M, *i.e.,* $\|\varepsilon(t)\|_X \leqslant M e^{-\lambda t}$.

Now, we go back to the proof of Theorem 5.32. Let $(t_n)_{n\geqslant 0}$ and $(\mathbb{T}_\infty(t,s))_{0\leqslant s\leqslant t}$ be as in the hypotheses of Theorem 5.32. Let Ω be the set of limit points of $(\varepsilon(t_n))_{n\geqslant 0}$ for the weak topology of X, that is, the set of points $\xi \in X$ such that there exists a subsequence $(n_k)_{k\geqslant 0}$ such that $\varepsilon(t_{n_k}) \overset{w}{\rightharpoonup} \xi$ as $k \to +\infty$. Since ε is bounded in X (because $\mathbb{S}$ is a contraction), by Kakutani's theorem (see, *e.g.,* [Bre11, Theorem 3.17]), the set $\{\varepsilon(t_n), n \in \mathbb{N}\}$ is relatively weakly compact in X. Hence Ω is not empty. Let $\xi \in \Omega$ and $(\varepsilon(t_{n_k}))_{k\geqslant 0}$ be a subsequence converging weakly to ξ. Then, according to (5.26) and Lemma 5.41,

$$\begin{aligned}\left\|W_\infty(0,\tau)\varepsilon(t_{n_k})\right\|_X &\leqslant \left\|W(t_{n_k},\tau)\varepsilon(t_{n_k})\right\|_X \\ &\quad + \left\|W_\infty(0,\tau) - W(t_{n_k},\tau)\right\|_{\mathscr{L}(X)} \left\|\varepsilon_0\right\|_X \\ &\underset{k\to+\infty}{\longrightarrow} 0.\end{aligned}$$

Hence $\xi \in \ker W_\infty(0,\tau)$. Thus $\Omega \subset \ker W_\infty(0,\tau)$. Let $\psi \in X$. By definition of Ω, and since ε is bounded, for all $\eta > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, there exists $\xi_n \in \Omega$ such that

$$\left|\left\langle \varepsilon(t_n) - \xi_n, \psi\right\rangle_X\right| \leqslant \eta.$$

Then, if $\psi \in (\ker W_\infty(0,\tau))^\perp$, $\left\langle \xi_n, \psi\right\rangle_X = 0$ which yields

$$\left|\left\langle \varepsilon(t_n), \psi\right\rangle_X\right| \leqslant \left|\left\langle \varepsilon(t_n) - \xi_n, \psi\right\rangle_X\right| + \left|\left\langle \xi_n, \psi\right\rangle_X\right| \leqslant \eta.$$

Since this result holds for all $\tau \geqslant 0$,

$$\left\langle \varepsilon(t_n), \psi\right\rangle_X \underset{n\to+\infty}{\overset{w}{\longrightarrow}} 0, \qquad \forall \psi \in \bigcup_{\tau\geqslant 0} (\ker W_\infty(0,\tau))^\perp.$$

This conclude the proof of the first part of Theorem 5.32.

Now, assume that $((t_{n+1}-t_n))_{n\geqslant 0}$ is bounded and $\mathcal{O} = X$. It is sufficient to prove that for all increasing positive sequences $(\tau_k)_{k\geqslant 0} \to +\infty$, $\varepsilon(\tau_k) \overset{w}{\rightharpoonup} 0$ as $k \to +\infty$. For all $k \in \mathbb{N}$, let $n_k \in \mathbb{N}$ be such that $t_{n_k} \leqslant \tau_k < t_{n_k+1}$. Then $s_k = \tau_k - t_{n_k}$ is a non-negative bounded sequence. Hence, up to an extraction of $(t_n)_{n\geqslant 0}$, it is now sufficient to prove that $\varepsilon(t_n + s_n) \overset{w}{\rightharpoonup} 0$ as $n \to +\infty$ for all non-negative bounded sequences $(s_n)_{n\geqslant 0}$. Set $\bar{s} = \sup_{n\in\mathbb{N}} s_n$. For all $\psi \in X$,

$$\begin{aligned}\left|\left\langle \varepsilon(t_n + s_n), \psi\right\rangle_X\right| &\leqslant \left|\left\langle \mathbb{T}_\infty(s_n,0)\varepsilon(t_n), \psi\right\rangle_X\right| \\ &\quad + \left\|(\mathbb{T}(t_n+s_n,t_n) - \mathbb{T}_\infty(s_n,0))\right\|_{\mathscr{L}(X)} \left\|\varepsilon_0\right\|_X \left\|\psi\right\|_X \\ &\quad + \left\|\varepsilon(t_n+s_n) - \mathbb{T}(t_n+s_n,t_n)\varepsilon(t_n)\right\|_X \left\|\psi\right\|_X.\end{aligned}$$

By (5.18), and because $(s_n)_{n\geqslant 0}$ is bounded, it follows that

$$\left\|(\mathbb{T}(t_n+s_n,t_n) - \mathbb{T}_\infty(s_n,0))\right\|_{\mathscr{L}(X)} \underset{n\to+\infty}{\longrightarrow} 0.$$

Using (5.10), (5.25) and the Cauchy-Schwarz inequality

$$\begin{aligned}\left\|\varepsilon(t_n+s_n) - \mathbb{T}(t_n+s_n,t_n)\varepsilon(t_n)\right\|_X &\leqslant rMe^{\omega\bar{s}}\|C\|_{\mathscr{L}(X,Y)} \int_{t_n}^{t_n+\bar{s}} \left\|C\varepsilon(t)\right\|_Y \mathrm{d}t \\ &\underset{n\to+\infty}{\longrightarrow} 0.\end{aligned}$$

Hence, it remains to prove that $\mathbb{T}_\infty(s_n,0)\varepsilon(t_n) \overset{w}{\rightharpoonup} 0$ as $n \to +\infty$. For all $t \geqslant 0$, (5.10) and (5.18) yield $\|\mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \leqslant Me^{\omega t}$, and thus for $\psi \in X$,

$$|\langle \mathbb{T}_\infty(s_n,0)\varepsilon(t_n),\psi\rangle_X| \leqslant Me^{\omega \bar{s}} \|\varepsilon_0\|_X \|\psi\|_X .$$

Let $\ell \in \mathbb{R}$ and $(n_k)_{k\geqslant 0}$ a subsequence such that $\left|\langle \mathbb{T}_\infty(s_{n_k},0)\varepsilon(t_{n_k}),\psi\rangle_X\right| \to \ell$ as $k \to +\infty$. We now show that $\ell = 0$ to end the proof. Since $(s_n)_{n\geqslant 0}$ is bounded and $s \mapsto \mathbb{T}_\infty(s,0)^*\psi$ is continuous in the strong topology of X, $(\mathbb{T}_\infty(s_{n_k},0)^*\psi)_{k\geqslant 0}$ converges strongly up to a new extraction of $(s_{n_k})_{k\geqslant 0}$ to some $\xi \in X$. Then, for all $k \in \mathbb{N}$,

$$\begin{aligned}\left|\langle \mathbb{T}_\infty(s_{n_k},0)\varepsilon(t_{n_k}),\psi\rangle_X\right| &= \left|\langle \varepsilon(t_{n_k}),\mathbb{T}_\infty(s_{n_k},0)^*\psi\rangle_X\right| \\ &\leqslant \left|\langle \varepsilon(t_{n_k}),\xi\rangle_X\right| + \|\mathbb{T}_\infty(s_{n_k},0)^*\psi - \xi\|_X \|\varepsilon_0\|_X \\ &\underset{k\to+\infty}{\longrightarrow} 0.\end{aligned}$$

Thus $\ell = 0$. ■

Remark 5.43. One of the steps of the proof of Theorem 5.32 (see Appendix 5.5.1) is to show that for all $\varepsilon_0 \in \mathcal{D}$, $\varepsilon : t \mapsto \mathbb{S}(t,0)\varepsilon_0$ satisfies

$$\int_{t_0}^{t_0+\tau} \|C\varepsilon(t)\|_Y^2 \,\mathrm{d}t \underset{t_0\to+\infty}{\longrightarrow} 0, \qquad \forall \tau \geqslant 0. \tag{5.27}$$

This does not yields *a priori* that $C\varepsilon(t) \to 0$ as t goes to infinity. However, if there exists a positive constant $\alpha > 0$ such that for all $x \in \mathcal{D}$ and all $t \geqslant 0$,

$$\|C^*CA(t)x\|_X \leqslant \alpha \|x\|_X , \tag{5.28}$$

then $C\varepsilon(t) \underset{t\to+\infty}{\longrightarrow} 0$. Indeed, (5.23) will yield

$$\int_0^{+\infty} \|C\varepsilon(t)\|_Y^2 \,\mathrm{d}t < +\infty. \tag{5.29}$$

Moreover, for all $t \geqslant 0$,

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|C\varepsilon(t)\|_Y^2 &= \langle C\varepsilon(t), C\dot{\varepsilon}(t)\rangle_Y \\ &= \langle C\varepsilon(t), CA(t)\varepsilon(t)\rangle_Y - r\langle C\varepsilon(t), CC^*C\varepsilon(t)\rangle_Y \\ &= \langle \varepsilon(t), C^*CA(t)\varepsilon(t)\rangle_X - r\|C^*C\varepsilon(t)\|_X^2 \\ &\leqslant \alpha\|\varepsilon_0\|_X^2\end{aligned}$$

since $\mathbb{S}(t,0)$ is proved to be a contraction in Lemma 5.39. Thus, $\|C\varepsilon\|_Y^2$ is an integrable positive function, with bounded derivative. Hence, according to Barbalat's lemma, $\|C\varepsilon(t)\|_Y^2 \to 0$ as $t \to +\infty$.

5.5.2 Proof of Theorem 5.35

Let us first state two important lemmas. They imply that the dynamics of the error system (5.9) may be decomposed on the two subspaces $\mathcal{O}_\tau$ and $\mathcal{O}_\tau^\perp$. Therefore, the initial estimation of the unobservable part of the system $\Pi_{\mathcal{O}_\tau^\perp}\hat{z}_0$ does not affect the

reconstruction of the observable part $\Pi_{\mathcal{O}_\tau} z(t)$ at all. In Statement *(i)*, the hypothesis $\mathcal{O}_\tau = X$ holds, so that these two lemmas are useless. On contrary, they are used to prove Statement *(ii)*.

The proofs partly rely on the theory of bi-directional evolution systems, only introduced in the next chapter. We refer the reader to Section 6.1.2 for the basics notions needed in this part of the thesis.

Lemma 5.44. *Assume that $A(t)$ is skew-adjoint for all $t \in \mathbb{R}_+$. Let $\mathcal{O}_\tau$ be the observable subspace at time τ of the pair $(\mathbb{T}, C)$. Set $L = \mathbb{S}(\tau,0)^*\mathbb{S}(\tau,0)$. Then $L\mathcal{O}_\tau \subset \mathcal{O}_\tau$ and $L\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$.*

Proof. According to [DK74, Chapter 3, Lemma 1.1], since $A(t)$ is skew-adjoint for all $t \in \mathbb{R}$, it is the generator of a unitary bi-directional evolution system, still denoted by $\mathbb{T}$. In particular, for all $t \geqslant s \geqslant t_0 \in \mathbb{R}$, $\mathbb{T}(t,s)^*\mathbb{T}(t,t_0) = \mathbb{T}(s,t_0)$.

Let $\varepsilon_0 \in \mathcal{D} \cap \mathcal{O}_\tau$. For all $\psi_0 \in \mathcal{D} \cap \mathcal{O}_\tau^\perp = \mathcal{D} \cap \ker W(0,\tau)$, the Duhamel's formula (5.25) yields

$$\begin{aligned}\langle L\varepsilon_0, \psi_0\rangle_X &= \langle \mathbb{S}(\tau,0)\varepsilon_0, \mathbb{S}(\tau,0)\psi_0\rangle_X \\ &= \langle \varepsilon_0, \mathbb{T}(\tau,0)^*\mathbb{S}(\tau,0)\psi_0\rangle_X - r\int_0^\tau \langle C\mathbb{S}(s,0)\varepsilon_0, C\mathbb{T}(\tau,s)^*\mathbb{S}(\tau,0)\psi_0\rangle_X \,\mathrm{d}s.\end{aligned} \tag{5.30}$$

Since $\psi_0 \in \ker W(0,\tau)$, $C\mathbb{T}(t,0)\psi_0 = 0$ for all $t \in [0,\tau]$. Set $\psi(t) = \mathbb{T}(t,0)\psi_0$ and $\bar\psi(t) = \mathbb{S}(t,0)(-\psi_0)$. Then $\psi + \bar\psi$ is the unique solution of (5.8) starting from $0 \in \mathcal{D}$ and with $y(t) = 0$ for all $t \in [0,\tau]$. Hence, $\psi + \bar\psi = 0$ on $[0,\tau]$, *i.e.,* $\mathbb{S}(t,0)\psi_0 = \mathbb{T}(t,0)\psi_0$ for all $t \in [0,\tau]$. Then, (5.30) yields

$$\begin{aligned}\langle L\varepsilon_0, \psi_0\rangle_X &= \langle \varepsilon_0, \mathbb{T}(\tau,0)^*\mathbb{T}(\tau,0)\psi_0\rangle_X - r\int_0^\tau \langle C\mathbb{T}(s,0)\varepsilon_0, C\mathbb{T}(\tau,s)^*\mathbb{T}(\tau,0)\psi_0\rangle_X \,\mathrm{d}s. \\ &= \langle \varepsilon_0, \psi_0\rangle_X - r\int_0^\tau \langle C\mathbb{T}(s,0)\varepsilon_0, C\mathbb{T}(s,0)\psi_0\rangle_X \,\mathrm{d}s. \\ &= 0.\end{aligned}$$

Thus, since $\mathcal{D}$ is dense in X, $L\varepsilon_0 \in \mathcal{O}_\tau$ for all $\varepsilon_0 \in \mathcal{O}_\tau$. Now, let $\varepsilon_0 \in \mathcal{O}_\tau^\perp$ and $\psi_0 \in \mathcal{O}_\tau$. Since L is self-adjoint, $\langle L\varepsilon_0, \psi_0\rangle_X = \langle \varepsilon_0, L\psi_0\rangle_X = 0$ from above. Hence, $L\varepsilon_0 \in \mathcal{O}_\tau^\perp$. ∎

Lemma 5.45. *Assume that $A(t)$ is skew-adjoint for all $t \in \mathbb{R}_+$. Let $\mathcal{O}_\tau$ be the observable subspace at time τ of the pair $(\mathbb{T}, C)$. If $\mathbb{T}(t,0)\mathcal{O}_\tau \subset \mathcal{O}_\tau$ and $\mathbb{T}(t,0)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$ for all $t \in [0,\tau]$, then $\mathbb{S}(t,0)\mathcal{O}_\tau \subset \mathcal{O}_\tau$ and $\mathbb{S}(t,0)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$ for all $t \in [0,\tau]$.*

Proof. As in Lemma 5.44, $(A(t))_{t\geqslant 0}$ generates a unitary bi-directional evolution system $\mathbb{T}$. Hence, $\mathbb{T}(t,0)\mathcal{O}_\tau \subset \mathcal{O}_\tau$ if and only if $\mathbb{T}(0,t)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$ and $\mathbb{T}(t,0)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$ if and only if $\mathbb{T}(0,t)\mathcal{O}_\tau \subset \mathcal{O}_\tau$. Assume that all these inclusions hold. Let $t \in \mathbb{R}_+$ and $\varepsilon_0 \in \mathcal{O}_\tau$. For all $\psi_0 \in \mathcal{O}_\tau^\perp$, the Duhamel's formula (5.25) yields

$$\begin{aligned}\langle \mathbb{S}(t,0)\varepsilon_0, \psi_0\rangle_X &= \langle \mathbb{T}(t,0)\varepsilon_0, \psi_0\rangle_X - r\int_0^t \langle C\mathbb{S}(s,0)\varepsilon_0, C\mathbb{T}(t,s)^*\psi_0\rangle_X \,\mathrm{d}s \\ &= \langle \mathbb{T}(t,0)\varepsilon_0, \psi_0\rangle_X - r\int_0^t \langle C\mathbb{S}(s,0)\varepsilon_0, C\mathbb{T}(s,0)\mathbb{T}(0,t)\psi_0\rangle_X \,\mathrm{d}s.\end{aligned}$$

Since $\mathbb{T}(t,0)\mathcal{O}_\tau \subset \mathcal{O}_\tau$ and $\mathbb{T}(0,t)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$, it holds that $\langle \mathbb{T}(t,0)\varepsilon_0, \psi_0\rangle_X = 0$ and $C\mathbb{T}(s,0)\mathbb{T}(0,t)\psi_0 = 0$ for all $s \in [0,\tau]$, respectively. Hence, $\mathbb{S}(t,0)\mathcal{O}_\tau \subset \mathcal{O}_\tau$. Similarly, if $\varepsilon_0 \in \mathcal{O}_\tau^\perp$, then for all $\psi_0 \in \mathcal{O}_\tau$, the Duhamel's formula (5.25) yields

$$\begin{aligned}\langle \mathbb{S}(t,0)\varepsilon_0, \psi_0\rangle_X &= \langle \varepsilon_0, \mathbb{S}(0,t)\psi_0\rangle_X \\ &= \langle \varepsilon_0, \mathbb{T}(0,t)\psi_0\rangle_X + r\int_0^t \langle C\mathbb{T}(0,s)^*\varepsilon_0, C\mathbb{S}(s,0)\psi_0\rangle_X \,\mathrm{d}s \\ &= \langle \varepsilon_0, \mathbb{T}(0,t)\psi_0\rangle_X + r\int_0^t \langle C\mathbb{T}(s,0)\varepsilon_0, C\mathbb{S}(s,0)\psi_0\rangle_X \,\mathrm{d}s.\end{aligned}$$

Since $\mathbb{T}(0,t)\mathcal{O}_\tau \subset \mathcal{O}_\tau$ and $\mathbb{T}(t,0)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$, it holds that $\langle \varepsilon_0, \mathbb{T}(0,t)\psi_0\rangle_X = 0$ and $C\mathbb{T}(s,0)\varepsilon_0 = 0$ for all $s \in [0,\tau]$, respectively. Hence, $\mathbb{S}(t,0)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$. ■

With these lemmas in mind, we are now able to prove the main Theorem 5.35.

Proof of Theorem 5.35. Let $\tau > 0$ be as in the assumptions of the theorem, and set $L = \mathbb{S}(\tau,0)^*\mathbb{S}(\tau,0)$. If the hypotheses of Statement *(i)* are satisfied, then the conclusions of Lemmas 5.44 and 5.45 hold (since $\mathcal{O}_\tau = X$). Otherwise, if the hypotheses of Statement *(ii)* are satisfied, then the hypotheses and conclusions of Lemmas 5.44 and 5.45 hold. Assume for a moment that $A(t)$ is skew-adjoint for all $t \in \mathbb{R}_+$. Then $((A(t))_{t\geqslant 0}, C)$ is 0-weakly dissipative (see Remark 5.29) and $(\mathbb{T}(t,s))_{t,s\geqslant 0}$ is a unitary bi-directional evolution system (see [DK74, Chapter 3, Lemma 1.1]). Hence, applying [Paz83, Chapter 5, Theorem 2.3] to $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t\leqslant \tau}$ and $(\mathbb{T}(\tau-t,\tau-s))_{0\leqslant s\leqslant t\leqslant\tau}$ perturbed with the bounded operators $-rC^*C$ and rC^*C respectively, we obtain that $(\mathbb{S}(t,s))_{0\leqslant s\leqslant t\leqslant\tau}$ and $(\mathbb{S}(\tau-t,\tau-s))_{0\leqslant s\leqslant t\leqslant\tau}$ are two evolution systems. Moreover, the condition $\mathbb{S}(s,t)\mathbb{S}(t,s) = \mathrm{Id}_X$ for all $t,s \in \mathbb{R}_+$ is also satisfied, due to the uniqueness of solutions of (5.9). Hence $(\mathbb{S}(t,s))_{0\leqslant s\leqslant t\leqslant\tau}$ is actually a bi-directional evolution system, that can be naturally extended on $\mathbb{R}_+$.

Thus, both Statements *(i)* and *(ii)* are implied by the following:

(iii) Suppose that $((A(t))_{t\geqslant 0}, C)$ is μ-weakly detectable, $\mathbb{S}(\tau,0)$ is bounded from below and normal. Assume that the conclusions of Lemmas 5.44 and 5.45 are satisfied. Then (5.8) *is a strong asymptotic $\mathcal{O}_\tau$-observer of* (5.7) *for all $r > \mu$.*

Suppose that the assumptions of *(iii)* hold. We aim to show that $\Pi_{\mathcal{O}_\tau}\mathbb{S}(t,0)\varepsilon_0 \to 0$ as $t \to +\infty$ for all $\varepsilon_0 \in X$. The floor function is denoted by $\lfloor\cdot\rfloor$. For all $t \in \mathbb{R}_+$ and all $\varepsilon_0 \in X$,

$$\begin{aligned}\|\Pi_{\mathcal{O}_\tau}\mathbb{S}(t,0)\,\varepsilon_0\|_X &= \left\|\Pi_{\mathcal{O}_\tau}\mathbb{S}\left(t, \left\lfloor\frac{t}{\tau}\right\rfloor\tau\right)\mathbb{S}\left(\left\lfloor\frac{t}{\tau}\right\rfloor\tau, 0\right)\varepsilon_0\right\|_X \\ &= \left\|\Pi_{\mathcal{O}_\tau}\mathbb{S}\left(t - \left\lfloor\frac{t}{\tau}\right\rfloor\tau, 0\right)\mathbb{S}(\tau,0)^{\lfloor\frac{t}{\tau}\rfloor}\varepsilon_0\right\|_X \\ &\qquad\qquad \text{(since } t \mapsto A(t) \text{ is } \tau\text{-periodic)} \\ &= \left\|\mathbb{S}\left(t - \left\lfloor\frac{t}{\tau}\right\rfloor\tau, 0\right)\mathbb{S}(\tau,0)^{\lfloor\frac{t}{\tau}\rfloor}\Pi_{\mathcal{O}_\tau}\varepsilon_0\right\|_X \\ &\qquad\qquad \text{(by the conclusion of Lemma 5.45)} \\ &\leqslant \left\|\mathbb{S}\left(t - \left\lfloor\frac{t}{\tau}\right\rfloor\tau, 0\right)\right\|_X \left\|\mathbb{S}(\tau,0)^{\lfloor\frac{t}{\tau}\rfloor}\Pi_{\mathcal{O}_\tau}\varepsilon_0\right\|_X \\ &\leqslant \left\|\mathbb{S}(\tau,0)^{\lfloor\frac{t}{\tau}\rfloor}\Pi_{\mathcal{O}_\tau}\varepsilon_0\right\|_X. \qquad \text{(according to Lemma 5.39)}\end{aligned}$$

Moreover, for all $n \in \mathbb{N}$, $\langle L^n \varepsilon_0, \varepsilon_0 \rangle_X = \|\mathbb{S}(\tau,0)^n \varepsilon_0\|_X^2$ since $\mathbb{S}(\tau,0)$ is normal. Thus, applying Lemma 5.37 (i), it remains to prove that for all $\varepsilon_0 \in \mathcal{D} \cap \mathcal{O}_\tau$, $L^n \varepsilon_0 \to 0$ as $n \to +\infty$ since $\mathcal{D}$ is dense in X and L^n is a contraction for all $n \in \mathbb{N}$.

The proof is an adaptation of the strategy developed in [Hai14, Theorem 1.1.2]. First, we investigate the properties of L. It is self-adjoint positive-definite since $\mathbb{S}(\tau,0)$ is bounded from below. Let $\varepsilon_0 \in \mathcal{D} \cap \mathcal{O}_\tau$. The hypotheses of Lemma 5.39 hold. Hence, $\mathbb{S}$ is a contraction evolution system, and (5.22) yields

$$\langle L\varepsilon_0, \varepsilon_0 \rangle_X = \|\mathbb{S}(\tau,0)\varepsilon_0\|_X^2 \leqslant \|\varepsilon_0\|_X^2 - 2(r-\mu) \int_0^\tau \|C\mathbb{S}(t,0)\varepsilon_0\|_Y^2 \, \mathrm{d}t. \tag{5.31}$$

Denote by $L^{\frac{1}{2}}$ the square root of L. Then

$$\begin{aligned} \|L\varepsilon_0\|_X^2 &= \left\langle LL^{\frac{1}{2}}\varepsilon_0, L^{\frac{1}{2}}\varepsilon_0 \right\rangle_X \\ &\leqslant \left\|L^{\frac{1}{2}}\varepsilon_0\right\|_X^2 - 2(r-\mu) \int_0^\tau \left\|C\mathbb{S}(t,0)L^{\frac{1}{2}}\varepsilon_0\right\|_Y^2 \mathrm{d}t \\ &\leqslant \langle L\varepsilon_0, \varepsilon_0 \rangle_X \\ &\leqslant \|\varepsilon_0\|_X^2 - 2(r-\mu) \int_0^\tau \|C\mathbb{S}(t,0)\varepsilon_0\|_Y^2 \, \mathrm{d}t. \end{aligned}$$

If $\|L\varepsilon_0\|_X = \|\varepsilon_0\|_X$, then $C\mathbb{S}(t,0)\varepsilon_0 = 0$ for all $t \in [0,\tau]$. Hence, according to the Duhamel's formula (5.25), $\mathbb{S}(t,0)\varepsilon_0 = \mathbb{T}(t,0)\varepsilon_0$ for all $t \in [0,\tau]$. Then $W(0,\tau)\varepsilon_0 = 0$, *i.e.,* $\varepsilon_0 \in \mathcal{O}_\tau \cap \mathcal{O}_\tau^\perp = \{0\}$.

Thus, $\|L\varepsilon_0\|_X < \|\varepsilon_0\|_X$ if $\varepsilon_0 \neq 0$. Moreover, (5.22) yields for all $\varepsilon_0 \in X$ and all $n \in \mathbb{N}$

$$\begin{aligned} \left\langle L^{n+1}\varepsilon_0, \varepsilon_0 \right\rangle_X - \langle L^n \varepsilon_0, \varepsilon_0 \rangle_X &= \|\mathbb{S}((n+1)\tau,0)\varepsilon_0\|_X^2 - \|\mathbb{S}(n\tau,0)\varepsilon_0\|_X^2 \\ &\leqslant -2(r-\mu) \int_{n\tau}^{(n+1)\tau} \|C\mathbb{S}(t,0)\varepsilon_0\|_Y^2 \, \mathrm{d}t \\ &\leqslant 0. \end{aligned}$$

Then $(L^n)_{n\geqslant 0}$ is a non-increasing sequence of bounded self-adjoint positive-definite operators on the vector space $\mathcal{O}_\tau$ (by the invariance property). Hence, according to [TW09, Lemma 12.3.2], there exists a bounded self-adjoint positive-definite operator $L_\infty \in \mathscr{L}(\mathcal{O}_\tau)$ such that $L_\infty \leqslant L^n$ for all $n \in \mathbb{N}$ and $L^n\varepsilon_0 \to L_\infty\varepsilon_0$ as $n \to +\infty$ for all $\varepsilon_0 \in \mathcal{O}_\tau$. It remains to prove that $L_\infty = 0$.

For all $x_1, x_2 \in \mathcal{O}_\tau$ and all $n \in \mathbb{N}$,

$$\begin{aligned} \langle L_\infty x_1, L_\infty x_2 \rangle_X = \langle L_\infty x_1, (L_\infty - L^n)x_2 \rangle_X + \langle (L_\infty - L^n)x_1, L^n x_2 \rangle_X \\ + \langle L^n x_1, L^n x_2 \rangle_X . \end{aligned}$$

Since L is self-adjoint,

$$\langle L^n x_1, L^n x_2 \rangle_X = \left\langle L^{2n} x_1, x_2 \right\rangle_X \underset{n\to+\infty}{\longrightarrow} \langle L_\infty x_1, x_2 \rangle_X .$$

Hence $L_\infty^2 = L_\infty$. Moreover, for all $\varepsilon_0 \in \mathcal{O}_\tau \setminus \{0\}$,

$$\|L_\infty\varepsilon_0\|_X^2 = \left\langle L_\infty^2\varepsilon_0, \varepsilon_0 \right\rangle_X = \langle L_\infty\varepsilon_0, \varepsilon_0 \rangle_X \leqslant \left\langle L^2\varepsilon_0, \varepsilon_0 \right\rangle_X = \|L\varepsilon_0\|_X^2 < \|\varepsilon_0\|_X^2 .$$

Hence $\|L_\infty\varepsilon_0\|_X^2 = \|L_\infty^2\varepsilon_0\|_X^2 < \|L_\infty\varepsilon_0\|_X^2$ if $L_\infty\varepsilon_0 \neq 0$. Thus $L_\infty\varepsilon_0 = 0$ for all $\varepsilon_0 \in \mathcal{O}_\tau$, which ends the proof. ■

Chapter 6

Back and Forth Nudging

Comme de longs échos qui de loin se confondent
Dans une ténébreuse et profonde unité

C. Baudelaire, *Les fleurs du mal*, "Correspondances"

Abstract. *In this chapter, we consider the offline estimation problem of the initial data for infinite-dimensional linear time-varying system from the measurement of a linear output over a finite-time interval. We use the so-called Back and Forth Nudging (BFN) algorithm, based on iterations of forward and backward asymptotic observers learning the state from its output. We recall notions on bi-directional evolution systems, and rely on the Luenberger observers involved in Chapter 5. As in the asymptotic context, we prove under a weak detectability assumption that the observer estimates the so-called observable subspace of the system, at least in the weak topology of the state space. Additional conditions on the system are required to show strong convergence. We illustrate the results on a transport equation.*

Contents

Introduction

When only part of the state of an infinite-dimensional system is measured on some finite time interval, an important inverse problem is the one of estimating the initial state by looking at the corresponding output on the time interval. This problem arises for example in oceanography and meteorology (see *data assimilation* problems in [AB05, AB08, Aur09]) and in process control (see Chapter 7). Although various inverse problems techniques may be applied, the Back and Forth Nudging (BFN) algorithm (also called time reversal based algorithm in [IRT11]) proves to be of deep interest, because of its strong use of the system dynamics. It relies on the theory of Luenberger observers (see Chapter 5). But since the observation time is finite, asymptotic observers must be adapted. For systems admitting both a forward and a backward evolution (see Section 6.1), it is possible to emulate a backward dynamics, hence a backward observer. Then, the main idea is to use iteratively forward and backward observers, working on the same bounded time interval, and using the same measurement. After each iteration, the final estimation of the state made by the observer is used as the initial condition of the next iteration. This methodology leads to the back and forth observer described in Section 6.2.

In the autonomous context, strong convergence results have already been obtained for both exactly (see [RTW10, IRT11]) and approximately (see [HR11a, HR11b, Hai14]) observable systems. Adapting the asymptotic results obtained in Chapter 5, we show under less restrictive hypotheses the weak convergence of the BFN algorithm, and extend the results of [Hai14] to the time-varying context. Main results are stated in Section 6.3 and proved in Section 6.5.

6.1 Backward and forward systems

Let X and Y be two Banach spaces. We consider time-varying linear systems of the form

$$\begin{cases} \dot{z} = A(t)z, & t \in [0,T] \\ y = Cz. \end{cases} \tag{6.1}$$

where $T \in \mathbb{R}_+$, z lying in X is the state of the system, y lying in Y is the output, $A(t) : \mathcal{D} \to X$ are linear operators defined on the same dense subspace $\mathcal{D} \subset X$ for all $t \in [0,T]$ and $C : X \to Y$ is a linear operator. Contrarily to the previous chapter, we consider the system over a bounded time interval, *i.e.,* $T < +\infty$.

Before addressing the problem of back and forth observer design for such systems, we ensure the well-posedness of (6.1) by recalling some results on bi-directional evolution equations.

6.1.1 Strongly continuous groups

Definition 6.1 (Strongly continuous group)**.** A one-parameter family $(\mathbb{T}(t))_{t\in\mathbb{R}}$ of operators in $\mathscr{L}(X)$ is a *strongly continuous group* on X if it satisfies the following properties:

(Group property) $\mathbb{T}(0) = \mathrm{Id}_X$ and $\mathbb{T}(t+s) = \mathbb{T}(t)\mathbb{T}(s)$ for all $t, s \in \mathbb{R}$,

(Strong continuity) $\lim_{t\to 0} \mathbb{T}(t)z = z$ for all $z \in X$.

Moreover, if $\mathbb{T}(t)$ is unitary for all $t \in \mathbb{R}$, then $\mathbb{T}$ is said to be unitary.

Infinitesimal generators of strongly continuous groups are defined in the same way as for semigroups. Stone's theorem characterizes the generators of unitary groups.

Theorem 6.2 (Stone, see, *e.g.,* [TW09, Theorem 3.8.6])**.** *An operator $A : \mathcal{D} \to X$ is skew-adjoint (i.e., $A^* = -A$) if and only if A is the generator of a unitary group on X.*

In order to build this observer, we need to assume that the family $(A(t))_{t\in[0,T]}$ is the generator of a *bi-directional* evolution system on X denoted by $(\mathbb{T}(t,s))_{0\leqslant s,t\leqslant T}$.

6.1.2 Bi-directional evolution systems

Definition 6.3 (Bi-directional evolution systems)**.** Let $T \in \mathbb{R} \cup \{+\infty\}$. A two-parameter family $(\mathbb{T}(t,s))_{t,s\in[0,T}$ of operators in $\mathscr{L}(X)$ is a *bi-directional evolution system* on X over $[0,T]$ if it satisfies the following properties:

(Evolution property) $\mathbb{T}(t,t) = \mathrm{Id}_X$, $\mathbb{T}(t,s)\mathbb{T}(s,\tau) = \mathbb{T}(t,\tau)$ for $\tau, s, t \in [0,T]$,

(Strong continuity) $\lim\limits_{t\to s} \mathbb{T}(t,s)z = z$ for all $s \in [0,T]$ and all $z \in X$.

Moreover, if $\mathbb{T}(t,s)$ is unitary for all $s,t \in [0,T]$, then $\mathbb{T}$ is said to be unitary.

In particular, if $\mathbb{T}$ is a bi-directional evolution system, then it consists of invertible operators due to the evolution property: $\mathbb{T}(t,s) = \mathbb{T}(s,t)^{-1}$. Conversely, we have the following characterization:

Theorem 6.4 ([NN02, Lemma 4.3])**.** *An evolution system $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t\leqslant T}$ on X over $[0,T]$ consists of invertible operators if and only if there exists a family of invertible bounded operators $(U(t))_{[0,T]}$ in $\mathscr{L}(X)$ such that*

$$\mathbb{T}(t,s) = U(t)U(s)^{-1} \tag{6.2}$$

Moreover, by setting $\mathbb{T}(s,t) = \mathbb{T}(t,s)^{-1}$ for $s < t$, the evolution system is extended to a bi-directional evolution system $(\mathbb{T}(t,s))_{t,s\in[0,T]}$.

Using Definition 5.6, one can check that a family $(\mathbb{T}(t,s))_{0\leqslant s,t\leqslant T}$ of bounded linear operators on X is a bi-directional evolution system if and only if:

(a) $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t\leqslant T}$ is an evolution systems on X,

(b) $(\mathbb{T}(T-t,T-s))_{0\leqslant s\leqslant t\leqslant T}$ is an evolution system on X,

(c) for all $t,s \in [0,T]$, $\mathbb{T}(s,t)\mathbb{T}(t,s) = \mathrm{Id}_X$.

This characterization leads us to the notion of infinitesimal generators of bi-directional evolution systems.

Definition 6.5 (Bi-directional generators)**.** A family of operators $(A(t))_{t\in[0,T]}$ is said to be the generator of a bi-directional evolution system $\mathbb{T}$ on X over $[0,T]$ if and only if it is the generator of an evolution system $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t\leqslant T}$, $(-A(T-t))_{t\in[0,T]}$ is the generator of an evolution system $(\mathbb{T}(T-t,T-s))_{0\leqslant s\leqslant t\leqslant T}$ and condition (c) is satisfied.

In particular, if $(A(t))_{t\in[0,T]}$ is the generator of a bi-directional evolution system $\mathbb{T}$ on X over $[0,T]$, then

$$\|\mathbb{T}(t,s)\|_{\mathscr{L}(X)} \leqslant Me^{\omega(t-s)}, \qquad \forall s,t\in[0,T]. \tag{6.3}$$

Hence, the flow

$$\begin{aligned}\varphi : [0,T]^2 \times X &\longrightarrow X\\ (t,s,z) &\longmapsto \mathbb{T}(t,s)z\end{aligned}$$

is continuous. Indeed, if $(t_n,s_n,z_n)_{n\in\mathbb{N}} \to (t,s,z) \in [0,T]^2\times X$ in the product topology, then

$$\begin{aligned}&\|\varphi(t_n,s_n,z_n)-\varphi(t,s,z)\|_X \qquad (6.4)\\ &\leqslant \|\mathbb{T}(t_n,s_n)(z_n-z)\|_X + \|(\mathbb{T}(t_n,s_n)-\mathbb{T}(t,s))\,z\|_X \quad \text{(by triangular inequality)}\\ &\leqslant Me^{\omega(t_n-s_n)}\,\|z_n-z\|_X + \|(\mathbb{T}(t_n,s_n)-\mathbb{T}(t,s))\,z\|_X \qquad \text{(by (6.3))}\\ &\underset{n\to+\infty}{\longrightarrow} 0. \qquad \text{(by strong continuity of } \mathbb{T})\end{aligned}$$

For skew-adjoint families of operators, we have the following result which follows directly from Theorem 5.11 and Remark 5.9.

Theorem 6.6 (Unitary bi-directional evolution systems)**.** *Let $T\in\mathbb{R}_+\cup\{+\infty\}$ and $\mathcal{D}$ be a subspace of X. If $A(t):\mathcal{D}\to X$ is a linear skew-adjoint operator for all $t\in[0,T]$ and $z\mapsto A(t)z$ is continuously differentiable in X for all $z\in\mathcal{D}$, then $(A(t))_{t\in[0,T]}$ is the generator of a unitary bi-directional evolution system $\mathbb{T}$ over X on $[0,T]$.*

The well-posedness follows from Theorem 5.15.

Theorem 6.7 (Well-posedness)**.** *Let $T\in\mathbb{R}_+\cup\{+\infty\}$, $t_0\in[0,T]$ and $z_0\in X$. Consider the abstract Cauchy problem*

$$\begin{cases}\dot{z} = A(t)z, & \forall t\in[0,T],\\ z(t_0)=z_0.\end{cases} \tag{6.5}$$

If $(A(t))_{t\in\mathbb{R}_+}$ is the generator of a bi-directional evolution system $\mathbb{T}$ on X over $[0,T]$, then (6.5) admits a unique solution $z\in C^0([0,T];X)$, which satisfies $z(t)=\mathbb{T}(t,t_0)z_0$ for all $t\in[0,T]$. Moreover, $z\in C^0([0,T];\mathcal{D})\cap C^1([0,T];X)$ if $z_0\in\mathcal{D}$.

6.2 Back and forth observer

We address the problem of reconstructing the state of (5.7) from the knowledge of its output on a bounded time interval $[0,T]$, $T\in\mathbb{R}_+$. Assume that $(A(t))_{t\in[0,T]}$ is the generator of a bi-directional evolution system on X over $[0,T]$. To achieve this new state estimation problem, we iteratively use forward and backward Luenberger observers. This methodology is called the Back and Forth Nudging in [AB05, AB08, AN12], or the time reversal based algorithm in [IRT11].

Let $\hat{z}_0 \in X$. For every $n \in \mathbb{N}$, we consider the following dynamical systems defined on $[0,T]$ as in [RTW10] by

$$\begin{cases} \dot{\hat{z}}^{2n} = A(t)\hat{z}^{2n} - rC^*(C\hat{z}^{2n} - y) \\ \hat{z}^{2n}(0) = \begin{cases} \hat{z}^{2n-1}(0) & \text{if } n \geqslant 1 \\ \hat{z}_0 & \text{otherwise.} \end{cases} \end{cases} \tag{6.6}$$

$$\begin{cases} \dot{\hat{z}}^{2n+1} = A(t)\hat{z}^{2n+1} + rC^*(C\hat{z}^{2n+1} - y) \\ \hat{z}^{2n+1}(T) = \hat{z}^{2n}(T). \end{cases} \tag{6.7}$$

Remark 6.8. System (6.6) is the usual asymptotic Luenberger observer of (5.7) (see (5.8)), whereas system (6.7) may be seen as an asymptotic Luenberger observer of (5.7) in reversed time. Indeed, $\hat{z}^{2n+1}(t)$ satisfies (6.6) if and only if $\hat{z}_r^{2n+1}(t) := \hat{z}^{2n+1}(T-t)$ satisfies

$$\begin{cases} \dot{\hat{z}}_r^{2n+1} = -A(T-t)\hat{z}_r^{2n+1} - rC^*(C\hat{z}_r^{2n+1} - y(T-t)) \\ \hat{z}_r^{2n+1}(0) = \hat{z}^{2n}(T). \end{cases}$$

Therefore, the coupled system (6.6)-(6.7) where $n \in \mathbb{N}$ is an iteration of successive Luenberger observers forward and backward in time. The final value of the estimation obtained after an iteration is used as the initial condition of the next iteration.

Let $\varepsilon_0 = \hat{z}_0 - z_0$ and $\varepsilon^n = \hat{z}^n - z$ for all $n \in \mathbb{N}$. Then $\hat{z}^{2n}$ and $\hat{z}^{2n+1}$ satisfy respectively (6.6) and (6.7) if and only if ε^{2n} and ε^{2n+1} satisfy

$$\begin{cases} \dot{\varepsilon}^{2n} = (A(t) - rC^*C)\varepsilon^{2n} \\ \varepsilon^{2n}(0) = \begin{cases} \varepsilon^{2n-1}(0) & \text{if } n \geqslant 1 \\ \varepsilon_0 & \text{otherwise.} \end{cases} \end{cases} \tag{6.8}$$

$$\begin{cases} \dot{\varepsilon}^{2n+1} = (A(t) + rC^*C)\varepsilon^{2n+1} \\ \varepsilon^{2n+1}(T) = \varepsilon^{2n}(T). \end{cases} \tag{6.9}$$

Since $C \in \mathscr{L}(X,Y)$, Theorem 5.10 claims that both $(A(t) - rC^*C)_{t\in[0,T]}$ and $(A(t) + rC^*C)_{t\in[0,T]}$ are stable families of generators of strongly continuous semigroups that generate bi-directional evolution systems on X denoted respectively by $(\mathbb{S}_+(t,s))_{0\leqslant s,t\leqslant T}$ and $(\mathbb{S}_-(t,s))_{0\leqslant s,t\leqslant T}$. Then, for all $n \in \mathbb{N}$, (6.6), (6.7), (6.8) and (6.9) have respectively a unique solution $\hat{z}^{2n}$, $\hat{z}^{2n+1}$, ε^{2n} and ε^{2n+1} in $C^0([0,T];X)$. Moreover, $\hat{z}^{2n}(t) = \mathbb{T}(t,0)z_0 + \mathbb{S}_+(t,0)\varepsilon^{2n}(0)$, $\hat{z}^{2n+1}(t) = \mathbb{T}(t,T)z(T) + \mathbb{S}_-(t,T)\varepsilon^{2n+1}(T)$, $\varepsilon^{2n}(t) = \mathbb{S}_+(t,0)\varepsilon^{2n}(0)$ and $\varepsilon^{2n+1}(t) = \mathbb{S}_-(t,T)\varepsilon^{2n+1}(T)$ for all $t \in [0,T]$. In particular, note that

$$\varepsilon^{2n}(0) = (\mathbb{S}_-(0,T)\mathbb{S}_+(T,0))^n \varepsilon_0. \tag{6.10}$$

If $(\hat{z}_0, \varepsilon_0) \in \mathcal{D}^2$, then $\hat{z}^n, \varepsilon^n \in C^0([0,T];\mathcal{D}) \cap C^1([0,T];X)$ for all $n \in \mathbb{N}$.

We are interested in the convergence properties of the initial state estimation $\hat{z}^{2n}(0)$ to the actual state $z(0)$, *i.e.,* of the estimation error $\varepsilon^{2n}(0)$ to 0, as n goes to infinity. Recall that for any closed linear subspace $\mathcal{O}$ of X, $\Pi_{\mathcal{O}} \in \mathscr{L}(X)$ denotes the orthogonal projection onto $\mathcal{O}$.

Definition 6.9 (Back and forth observer)**.** For any closed linear subspace $\mathcal{O}$ of X, the system (6.6)-(6.7) is said to be a strong (resp. weak) back and forth $\mathcal{O}$-observer of (5.7) if and only if $\Pi_{\mathcal{O}}\varepsilon^{2n}(0) \to 0$ (resp. $\Pi_{\mathcal{O}}\varepsilon^{2n}(0) \overset{w}{\to} 0$) as $n \to +\infty$ for all $\varepsilon_0 \in X$. An X-observer is shortly called an observer.

6.3 Back and forth convergence

Back and forth observer convergence has been investigated mainly in the autonomous exactly observable context, in which strong exponential convergence with arbitrary decay rate can be proved. This is the framework adopted by the authors of [IRT11] and [RTW10]. In [Hai14], the author remained in the autonomous context, but removed the exact observability assumption, and obtained strong convergence on the observable subspace for skew-adjoint operators A. His main result is the following.

Theorem 6.10 ([Hai14, Theorem 1.1])**.** *Assume that* (5.7) *is autonomous and A is skew-adjoint. Let $T \in \mathbb{R}_+$ and $\mathcal{O}_T$ be the observable subspace at time T of the pair $(\mathbb{T}, C)$.*

(i) The sequence $\left(\left\|(\mathrm{Id}_X - \Pi_{\mathcal{O}_T})\,\varepsilon^{(2n)}(0)\right\|_X\right)_{n\in\mathbb{N}}$ *is constant.*

(ii) The sequence $\left(\left\|\Pi_{\mathcal{O}_T}\varepsilon^{(2n)}(0)\right\|_X\right)_{n\in\mathbb{N}}$ *is decreasing and tends towards* 0*.*

(iii) The two following propositions are equivalent:

- *There exists $\delta > 0$ such that*

$$\langle W(t_0,\tau)z_0, z_0\rangle_X \geqslant \delta \|z_0\|_X^2, \qquad \forall z_0 \in \mathcal{O}_T. \tag{6.11}$$

- *There exists $\gamma \in (0,1)$ such that*

$$\left\|\Pi_{\mathcal{O}_T}\varepsilon^{(2n)}(0)\right\|_X \leqslant \gamma^n \|\Pi_{\mathcal{O}_T}\varepsilon_0\|_X, \qquad \forall \varepsilon_0 \in X. \tag{6.12}$$

Note that (6.11) is an exact observability-like property, holding only on the observable subspace $\mathcal{O}_T$. In particular, Theorem 6.10 *(iii)* implies that if A is skew-adjoint and (A, C) is exactly observable, then system (6.6)-(6.7) is an *exponential* strong observer of (5.7) with decay rate γ.

In [IRT11], the authors showed that instead of considering backward observers as usual observers acting on the system in reversed time, it is possible to consider other *time reversal operators*, for example for the Schrödinger or wave equations. But as in [RTW10], only exact observability hypotheses are considered, leading to exponential convergence of the observer (as in Theorem 6.10 *(iii)*).

In this chapter, we focus on time-varying systems and approximate observability-like hypotheses. In particular, we will adapt the tools of Theorem 5.32 to the back and forth context, and extend the results of [Hai14] in the time-varying context. As in the asymptotic context (see Chapter 5), only weak back and forth observers, converging on the observable subspace $\mathcal{O}_T$, are obtained in general. Additional properties on $\mathbb{S}$ are required for strong convergence to hold. As in the autonomous context, strong exponential convergence can be obtained for exactly observable systems on $[0,T]$ (see Remark 5.42). We rely on a weak detectability hypothesis on both $((A(t))_{t\in[0,T]}, C)$ and $((-A(t))_{t\in[0,T]}, C)$, which is equivalent to

$$|\langle A(t)x, x\rangle_X| \leqslant \mu \|Cx\|_Y^2, \qquad \forall x \in \mathcal{D}. \tag{6.13}$$

Remark 6.11. The considered inner product on X is the same for both the forward and the backward observer. If one must change the inner product with a self-adjoint operator P as in Remark 5.30, then this change must be done for both observers. In [HPR14], the authors proved in the autonomous finite-dimensional context the existence of such a common operator P for both A and $-A$. In the autonomous infinite-dimensional context, a similar result can be obtained if the pair $(-A, C)$ is exactly observable in some finite time. Indeed, according to Theorem 5.24, there exists $S \in \mathscr{L}(X)$ coercive self-adjoint such that $A - S^{-1}C^*C$ is the generator of an exponentially stable strongly continuous semigroup $\mathbb{U}$. Define the infinite-time observability Gramian $P = \int_0^{+\infty} \mathbb{U}(t)^*C^*C\mathbb{U}(t)$, which is well defined since $\mathbb{U}$ is exponentially stable. If $(A - S^{-1}C^*C, C)$ is exactly observable, then P is coercive. According to [Phó91, Corollary 8], P is the unique solution of $P(A - S^{-1}C^*C) + (A - S^{-1}C^*C)^*P = -C^*C$. Hence

$$\left|\left\langle P(A - S^{-1}C^*C)x, x\right\rangle_X\right| = \|Cx\|_Y^2, \qquad \forall x \in \mathcal{D}. \tag{6.14}$$

Remark that for non-homogeneous systems of the form $\dot{x} = Ax + u(t)$, it is still possible to apply the forward and backward observers described in this chapter by setting $\hat{x} = A\hat{x} + u(t) - rC^*C\varepsilon$. Then, apply the back and forth observer on $\dot{x} = Ax = A_S x + u(t)$ with $A_S = (A - S^{-1}C^*C)$ and $u = S^{-1}C^*Cx$. Doing so, the inner product defined by the operator P is such that (A_S, C) and $(-A_S, C)$ are μ-weakly detectable with $\mu = 1$.

In the two following sections, we state the main results of this chapter. All the remarks made on the results of Chapter 5 are also valid for these results.

6.3.1 Weak back and forth observer

Theorem 6.12. *Assume that $(\mathbb{T}(t,s))_{0\leqslant s,t\leqslant T}$ is a bi-directional evolution system. Suppose that both $((A(t))_{t\in[0,T]}, C)$ and $((-A(t))_{t\in[0,T]}, C)$ are μ-weakly detectable and $r > \mu$. Let $\mathcal{O}_T$ be the observable subspace at time T of the pair $(\mathbb{T}, C)$. Then, the system* (6.6)-(6.7) *is a weak back and forth $\mathcal{O}_T$-observer of* (5.7).

The proof of Theorem 6.12 is given in Section 6.5.1. Under additional assumptions on the system, strong convergence of the observer holds.

6.3.2 Strong back and forth observer

Theorem 6.13. *Assume that $(\mathbb{T}(t,s))_{0\leqslant s,t\leqslant T}$ is a bi-directional evolution system. Let $\mathcal{O}_T$ be the observable subspace at time T of the pair $(\mathbb{T}, C)$. Suppose that both $((A(t))_{t\in[0,T]}, C)$ and $((-A(t))_{t\in[0,T]}, C)$ are μ-weakly detectable and $r > \mu$. Assume that $\mathbb{S}_-(0,T) = \mathbb{S}_+(T,0)^*$. If $\mathcal{O}_T = X$, then the system* (6.6)-(6.7) *is a strong back and forth observer of* (5.7).

The proof of Theorem 6.13 given in Section 6.5.2 is an adaptation of [Hai14, Theorem 1.1.2] to the time-varying context.

6.4 Application to a transport equation

Consider a one-dimensional time-varying transport equation with periodic boundary conditions. More precisely, let $x_1 > x_0 \geqslant 0$ and $X = L^2((x_0, x_1); \mathbb{R})$ the set of real-valued square-integrable functions over (x_0, x_1), endowed with the inner product $\langle f, g \rangle_X = \int_{x_0}^{x_1} fg$ for all $f, g \in X$. Let $\mathcal{D} = \{\psi \in X \mid \psi(x_0) = \psi(x_1), \psi' \in X\}$ and $G \in C^1([0,T]; \mathbb{R})$. For all $t \geqslant 0$, let

$$\begin{aligned} A(t) : \mathcal{D} &\longrightarrow X \\ \psi &\longmapsto -G(t)\frac{\mathrm{d}\psi}{\mathrm{d}x}. \end{aligned}$$

Then $A(t)$ is a skew-adjoint operator for all $t \geqslant 0$. Hence $(A(t))_{t\geqslant 0}$ is a stable family of generators of strongly continuous groups that share the same domain $\mathcal{D}$. Moreover $t \mapsto A(t)f$ is continuously differentiable for all $f \in \mathcal{D}$ since G is of class C^1. Then Theorem 6.6 ensures that $(A(t))_{t\in[0,T]}$ is the generator of a unique bi-directional unitary (*i.e.,* forward and backward contraction) evolution system on X denoted by $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t}$. Moreover, $\mathbb{T}(t,s)$ is defined for all $t \geqslant s \geqslant 0$ and all $z_0 \in X$ by

$$(\mathbb{T}(t,s)z_0)(x) = z_0(v(x,t,s)), \tag{6.15}$$

where

$$v(x,t,s) = x_0 + \left(\left(x - x_0 - \int_s^t G(\tau)\mathrm{d}\tau\right) \bmod (x_1 - x_0)\right) \tag{6.16}$$

for almost all $x \in (x_0, x_1)$.

Hence, for all real Hilbert spaces Y and all output operators $C \in \mathscr{L}(X, Y)$, the pair $((A(t))_{t\in[0,T]}, C)$ is 0-weakly detectable, as well as the pair $((-A(t))_{t\in[0,T]}, C)$. Consequently, the transport equation with periodic boundary conditions is a good candidate to apply the observer methodology developed in Chapters 5 and 6. Moreover, in the asymptotic context, we have the following proposition, which is useful to apply Theorem 5.32.

Proposition 6.14. *Assume that $T = +\infty$ and that both G and its derivative G' are bounded. If there exist $G_\infty \in C^1(\mathbb{R}_+, \mathbb{R})$ and an increasing positive sequence $(t_n)_{n\geqslant 0} \to +\infty$ such that $G(t_n + t) \to G_\infty(t)$ as $n \to +\infty$ for all $t \geqslant 0$, then $\|\mathbb{T}(t_n + t, t_n) - \mathbb{T}_\infty(t, 0)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$ uniformly in $t \in [0, \tau]$ for all $\tau \geqslant 0$, where $\mathbb{T}_\infty$ is the evolution system generated by $\left(-G_\infty(t)\frac{\mathrm{d}}{\mathrm{d}x}\right)_{t\geqslant 0}$.*

In particular, note that if G is periodic, then G and G' are bounded and there exist a *bounded* sequence $(t_n)_{n\geqslant 0}$ and a constant $G_\infty > 0$ such that $\|\mathbb{T}(t_n + t, t_n) - \mathbb{T}_\infty(t)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$ uniformly in $t \in [0, \tau]$ for all $\tau \geqslant 0$, where $\mathbb{T}_\infty$ is the strongly continuous semigroup generated by $-G_\infty\frac{\mathrm{d}}{\mathrm{d}x} : \mathcal{D} \to X$.

Proof of Proposition 6.14. It is a direct application of [IK02, Theorem 10.2.b]. The consistency condition (C) of [IK02] is satisfied since for all $z_0 \in \mathcal{D}$,

$$A(t_n + t)z_0 = -G(t_n + t)\frac{\mathrm{d}z_0}{\mathrm{d}x} \underset{n\to+\infty}{\longrightarrow} -G_\infty(t)\frac{\mathrm{d}z_0}{\mathrm{d}x} \tag{6.17}$$

Moreover, $(\|A(t_n + t)z_0\|_X)_{n\geqslant 0}$ is bounded by $\sup_{\mathbb{R}_+} |G| \left\|\frac{\mathrm{d}z_0}{\mathrm{d}x}\right\|_X$ for all $t \geqslant 0$ and all $z_0 \in \mathcal{D}$. For all $\eta t_{\max}, z_2 \in \mathcal{D}$, all $n \in \mathbb{N}$ and all $t, \tau \geqslant 0$, we have the following

inequalities:

$$\begin{aligned}
|\langle A(t_n+t+\tau)z_1 - A(t_n+t)z_2,\ z_1-z_2\rangle_X| & \\
&\leqslant |\langle (A(t_n+t+\tau)-A(t_n+t))z_1, z_1-z_2\rangle_X| \\
&\quad + |\langle A(t_n+t)(z_1-z_2), z_1-z_2\rangle_X| \\
&\leqslant |G(t_n+t+\tau)-G(t_n+t)| \left\|\frac{\mathrm{d}z_1}{\mathrm{d}x}\right\|_X \|z_1-z_2\|_X \\
&\qquad\qquad \text{(since } A(t_n+t) \text{ is skew-adjoint)} \\
&\leqslant \sup_{\mathbb{R}_+} |G'|\,\tau \left\|\frac{\mathrm{d}z_1}{\mathrm{d}x}\right\|_X \|z_1-z_2\|_X . \qquad (6.18)
\end{aligned}$$

Hence, the condition (E2u) of [IK02] is also satisfied. Therefore, all the hypotheses of [IK02, Theorem 10.2.b] are met, which ends the proof. ■

In the following sections, the form of the output operator is investigated.

6.4.1 Geometric conditions on the output operator

If the kernel of the output operator $C \in \mathscr{L}(X,Y)$ satisfies some geometric conditions, then the kernel of the observability Gramian of the system may be linked to the kernel of C. Indeed, assume that there exists a set $U \subset [x_0, x_1]$ such that

$$\ker C = \{\psi \in X \mid \psi|_U = 0\}, \qquad (6.19)$$

where $f|_U$ denotes the restriction of f to U. Then $z_0 \in \ker W(t_0,\tau)$ for some $t_0, \tau \geqslant 0$ if and only if $(\mathbb{T}(s,t_0)z_0)\,|_U = 0$ for almost all $s \in (t_0, t_0+\tau)$, *i.e.*, $z_0(v(x,s,t_0)) = 0$ for almost all $s \in (t_0, t_0+\tau)$ and almost all $x \in U$. Hence

$$\ker W(t_0,\tau) = \{\psi \in X \mid \psi|_{U_{\max}} = 0\} \qquad (6.20)$$

where $U_{\max} = \{v(x,s,t_0), x \in U, s \in [t_0, t_0+\tau]\}$. Moreover, note that

$$\ker W(t_0,\tau)^\perp = \left\{\psi \in X \;\middle|\; \psi|_{[x_0,x_1]\setminus U_{\max}} = 0\right\}. \qquad (6.21)$$

This leads to the following result. Roughly speaking, it states that if the observation time τ is sufficiently large for all the data to pass through the observation window $[x_{\min}, x_{\max}]$, then the observable part of the state is actually the full state.

Proposition 6.15. *Assume that* $\ker C \subset \left\{\psi \in X \;\middle|\; \psi|_{[x_{\min},x_{\max}]} = 0\right\}$ *for some interval* $[x_{\min}, x_{\max}] \subset [x_0, x_1]$. *If*

$$\left|\int_{t_0}^{t_0+\tau} G(t)\mathrm{d}t\right| \geqslant (x_1-x_0) - (x_{\max}-x_{\min}), \qquad (6.22)$$

for some $t_0, \tau \geqslant 0$, *then* $\ker W(t_0,\tau) = \{0\}$.

Proof. According to (6.20), it is sufficient to prove that $U_{\max} = [x_0, x_1]$ when $U = [x_{\min}, x_{\max}]$. Clearly, $U \subset U_{\max}$.

Now, let $x \in [x_0, x_1] \setminus U$. If $\int_{t_0}^{t_0+\tau} G(t)\mathrm{d}t \geqslant 0$, set $d = (x_{\min} - x) \bmod (x_1 - x_0)$. Then $d \leqslant (x_1-x_0) - (x_{\max}-x_{\min})$. Hence, according to the intermediate value theorem, there exists $s \in [t_0, t_0+\tau]$ such that $\int_{t_0}^{s} G(t)\mathrm{d}t = d$. Using (6.16), we obtain

$x = v(x_{\min}, s, t_0)$. Otherwise, $\int_{t_0}^{t_0+\tau} G(t)\mathrm{d}t \leqslant 0$. Set $d = (x - x_{\max}) \bmod (x_1 - x_0)$. Similarly, $d \leqslant (x_1 - x_0) - (x_{\max} - x_{\min})$. Hence, according to the intermediate value theorem, there exists $s \in [t_0, t_0 + \tau]$ such that $\int_{t_0}^{s} G(t)\mathrm{d}t = -d$. Using (6.16), we obtain $x = v(x_{\max}, s, t_0)$. Thus, in both cases, there exists $\tilde{x} \in U$ and $s \in [t_0, t_0 + \tau]$ such that $x = v(\tilde{x}, s, t_0)$.

■

6.4.2 Integral output operator with bounded kernel

Assume that $C \in \mathscr{L}(X, Y)$ is an integral output operator with bounded kernel, that is, there exists $k \in L^\infty((x_0, x_1); Y)$ (*i.e.,* with $\operatorname{ess\,sup}_{x\in(x_0,x_1)} \|k(x)\|_Y < +\infty$) such that[1]

$$C\psi = \int_{x_0}^{x_1} k(x)\psi(x)\mathrm{d}x \tag{6.23}$$

for all $\psi \in X$. Then, there is no time interval $(t_0, t_0 + \tau) \subset \mathbb{R}_+$ such that the pair $((A(t))_{t\geqslant 0}, C)$ is exactly observable on $(t_0, t_0 + \tau)$.

Proposition 6.16. *If $C \in \mathscr{L}(X, Y)$ satisfies (6.23) for some $k \in L^\infty((x_0, x_1); Y)$, then for all $t_0, \tau \geqslant 0$ and all $\delta > 0$, there exists $z_0 \in X$ such that*

$$\langle W(t_0, \tau)z_0, z_0\rangle_X \leqslant \delta \|z_0\|_X^2 . \tag{6.24}$$

Hence, for such output operators, the convergence of an observer must rely on weaker observability assumptions, such as the approximate observability. In the application of the results to a crystallization process (see Chapter 7), the reader will find that C is precisely an integral output operator with bounded kernel.

Proof of Proposition 6.16. Let $t_0, \tau \geqslant 0$, $z_0 \in X$ and $z(t) = \mathbb{T}(t_0 + t, t_0)z_0$ for all $t \geqslant t_0$. Since (x_0, x_1) is bounded, any $\psi \in L^2((x_0, x_1); \mathbb{R})$ is also integrable. Set $\|\psi\|_{L^1((x_0,x_1);\mathbb{R})} = \int_{x_0}^{x_1} |\psi(x)|\, dx$. Then

$$\begin{aligned}
\langle W(t_0, \tau)z_0, z_0\rangle_X &= \int_{t_0}^{t_0+\tau} \|Cz(t)\|_Y^2 \,\mathrm{d}t \\
&\leqslant \int_{t_0}^{t_0+\tau} \left(\int_{x_0}^{x_1} \|k(x)z(t,x)\|_Y \,\mathrm{d}x\right)^2 \mathrm{d}t \qquad \text{(Bochner inequality)} \\
&\leqslant \int_{t_0}^{t_0+\tau} \left(\int_{x_0}^{x_1} \|k(x)\|_Y \,|z(t,x)|\,\mathrm{d}x\right)^2 \mathrm{d}t \\
&\leqslant \|k\|^2_{L^\infty((x_0,x_1);Y)} \int_{t_0}^{t_0+\tau} \left(\int_{x_0}^{x_1} |z(t,x)|\, dx\right)^2 \mathrm{d}t \\
&\leqslant \tau\|k\|^2_{L^\infty((x_0,x_1);Y)} \sup_{t\in[t_0,t_0+\tau]} \|z(t)\|^2_{L^1((x_0,x_1);\mathbb{R})}.
\end{aligned}$$

Moreover, by the usual transport properties of v, we get for all $t \in [t_0, t_0 + \tau]$ that

$$\|z(t)\|^2_{L^1((x_0,x_1);\mathbb{R})} = \|z_0(v(t, t_0, \cdot))\|^2_{L^1((x_0,x_1);\mathbb{R})} = \|z_0\|^2_{L^1((x_0,x_1);\mathbb{R})}.$$

Hence

$$\langle W(t_0, \tau)z_0, z_0\rangle_X \leqslant \tau\|k\|_{L^\infty((x_0,x_1);Y)}\|z_0\|^2_{L^1((x_0,x_1);\mathbb{R})}.$$

The result follows from the fact that the norms $\|\cdot\|_{L^1((x_0,x_1);\mathbb{R})}$ and $\|\cdot\|_{L^2((x_0,x_1);\mathbb{R})}$ are not equivalent. ■

[1] C is well-defined because $[x_0, x_1] \ni x \mapsto k(x)f(x)$ is Bochner integrable, since $x \mapsto \|k(x)\|_Y$ is bounded and $x \mapsto f(x)$ is integrable (since (x_0, x_1) has finite length and $f \in L^2((x_0, x_1); \mathbb{R})$).

Remark 6.17. According to Remark 5.43, the boundedness of the operator C^*CA from $(\mathcal{D}, \|\cdot\|_X)$ to $(X, \|\cdot\|_X)$ is an interesting property for the convergence to 0 of the correction term $C\varepsilon$ of the observers. If we ask more regularity to the solutions of the transport equation, then the integral output operators in the form of (6.23) satisfy this assumption. Indeed, assume (in this remark *only*) that $X = \{\psi \in L^2(x_0, x_1; \mathbb{R}) \ : \ f' \in L^2(x_0, x_1; \mathbb{R})\}$ endowed with the inner product $\langle f, g\rangle_X = \int_{x_0}^{x_1}(fg + f'g')$ and $\mathcal{D}_{\text{new}} = \{\psi \in X \ : \ \psi(x_1) = \psi(x_1), \psi'(x_1) = \psi'(x_1), f'' \in L^2(x_0, x_1; \mathbb{R})\}$. Then, for all $z_0 \in \mathcal{D}_{\text{new}}$,

$$\begin{aligned}\|CAz_0\|_Y^2 &\leqslant \left(\int_{x_0}^{x_1} \left\|k(x)\frac{\mathrm{d}z_0}{\mathrm{d}x}(x)\right\|_Y dx\right)^2 \\ &\leqslant \|k\|_{L^\infty((x_0,x_1),Y)} \left(\int_{x_0}^{x_1} \left|\frac{\mathrm{d}z_0}{\mathrm{d}x}(x)\right| dx\right)^2 \\ &\leqslant \|k\|_{L^\infty((x_0,x_1),Y)}(x_1 - x_0)\,\|z_0\|_X^2\end{aligned}$$

by the Cauchy-Schwarz inequality. Thus, $C^*CA \in \mathscr{L}((\mathcal{D}_{\text{new}}, \|\cdot\|_X), (X, \|\cdot\|_X))$ since C is bounded.

6.5 Proofs of the results

This section is devoted to the proofs of the results stated in Section 6.3. Throughout the section, $(A(t))_{t\in\mathbb{R}_+}$ is the generator of a bi-directional evolution system $\mathbb{T}$ on X over $[0, T]$ (in the sense of Definition 6.5) for some $T \in \mathbb{R}_+$, $C \in \mathscr{L}(X, Y)$ and $\mathbb{S}_+$ and $\mathbb{S}_-$ are bi-directional evolution systems generated by $(A(t) - rC^*C)_{t\in\mathbb{R}_+}$ and $(A(t) + rC^*C)_{t\in\mathbb{R}_+}$, respectively (see Section 6.2).

6.5.1 Proof of Theorem 6.12

We adapt the proof of Theorem 5.32 to the BFN algorithm (see Section 5.5.1). The lemmas involved and steps of the proof are very similar.

Lemma 6.18. *If $((A(t))_{t\in[0,T]}, C)$ and $((-A(t))_{t\in[0,T]}, C)$ are μ-weakly detectable and $r > \mu$, then $\mathbb{S}_+$ (resp. $\mathbb{S}_-$) is a forward (resp. backward) contraction bi-directional evolution system, that is,*

$$\|\mathbb{S}_+(t, s)\|_{\mathscr{L}(X)} \leqslant 1 \quad \textit{and} \quad \|\mathbb{S}_-(s, t)\|_{\mathscr{L}(X)} \leqslant 1, \qquad \forall t \geqslant s \geqslant 0. \tag{6.25}$$

Proof. Since $\mathcal{D}$ is dense in X, it is sufficient to show that

$$\|\mathbb{S}_+(t, t_0)\varepsilon_0\|_X \leqslant \|\varepsilon_0\|_X \quad \text{and} \quad \|\mathbb{S}_-(t, t_0)\varepsilon_0\|_X \geqslant \|\varepsilon_0\|_X \tag{6.26}$$

for all $\varepsilon_0 \in \mathcal{D}$ and all $t \geqslant t_0 \geqslant 0$. Let $t_0 \geqslant 0$, $\varepsilon_0 \in \mathcal{D}$ and set $\varepsilon_+(t) = \mathbb{S}_+(t, t_0)\varepsilon_0$ and $\varepsilon_-(t) = \mathbb{S}_-(t, t_0)\varepsilon_0$ for all $t \geqslant t_0$. Then $\varepsilon^i \in C^1([0, +\infty), X)$ for $i \in \{0, 1\}$ and for all $t \geqslant t_0$,

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\varepsilon_+(t)\|_X^2 &= \langle \varepsilon_+(t), \dot{\varepsilon}_+(t)\rangle_X \\ &= \langle \varepsilon_+(t), A(t)\varepsilon_+(t)\rangle_X - r\,\langle \varepsilon_+(t), C^*C\varepsilon_+(t)\rangle_X \\ &\leqslant -(r - \mu)\,\|C\varepsilon_+(t)\|_Y^2 \quad \text{(since } ((A(t))_{t\geqslant 0}, C) \text{ is } \mu\text{-weakly detectable)} \qquad (6.27) \\ &\leqslant 0\end{aligned}$$

and

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\varepsilon_-(t)\right\|_X^2 &= \left\langle \varepsilon_-(t), \dot{\varepsilon}_-(t)\right\rangle_X \\ &= \left\langle \varepsilon_-(t), A(t)\varepsilon_-(t)\right\rangle_X + r\left\langle \varepsilon_-(t), C^*C\varepsilon_-(t)\right\rangle_X \\ &\geqslant (r-\mu)\left\|C\varepsilon_-(t)\right\|_Y^2 \quad \text{(since } ((-A(t))_{t\geqslant 0}, C) \text{ is } \mu\text{-weakly detectable)} \\ &\geqslant 0\end{aligned} \tag{6.28}$$

since $r > \mu$. Hence $[t_0, +\infty) \ni t \mapsto \left\|\varepsilon_+(t)\right\|_X^2$ is non-increasing and $[t_0, +\infty) \ni t \mapsto \left\|\varepsilon_-(t)\right\|_X^2$ is non-decreasing, which yields (5.21) since $\varepsilon_+(t_0) = \varepsilon_-(t_0) = \varepsilon_0$. ■

We are now able to prove the main Theorem 6.12.

Proof of Theorem 6.12. According to Lemma 6.18, $\mathbb{S}_+$ (resp. $\mathbb{S}_-$) is a forward (resp. backward) contraction bi-directional evolution system. Let $L = \mathbb{S}_-(0,T)\mathbb{S}_+(T,0) \in \mathscr{L}(X)$. Then L^n is a contraction for all $n \in \mathbb{N}$. Hence, applying Lemma 5.37 (ii), it is sufficient to show that $\left\langle L^n\varepsilon_0, \psi\right\rangle_X \to 0$ as $n \to +\infty$ for all $\psi \in \cup_{\tau\geqslant 0}(\ker W(0,T))^\perp$ and all $\varepsilon_0 \in \mathcal{D}$ since $\mathcal{D}$ is dense is X. Let $\varepsilon_0 \in \mathcal{D}$ and set $\varepsilon^{2n}(t) = \mathbb{S}_+(t,0)L^n\varepsilon_0$ for all $t \geqslant 0$ and all $n \in \mathbb{N}$. Since L is a contraction, $\left\|\varepsilon^{2n}(0)\right\|_X$ is non-increasing and thus has a finite limit as n goes to infinity. Moreover,

$$\begin{aligned}\left\|\varepsilon^{2n}(T)\right\|_X = \left\|\mathbb{S}_+(T,0)L^n\varepsilon_0\right\|_X &= \left\|\mathbb{S}_-(T,0)L^{n+1}\varepsilon_0\right\|_X \\ &= \left\|\mathbb{S}_-(T,0)\varepsilon^{2(n+1)}(0)\right\|_X \geqslant \left\|\varepsilon^{2(n+1)}(0)\right\|_X.\end{aligned}$$

Then (6.27) yields for all $n \in \mathbb{N}$

$$\begin{aligned}\int_0^T \left\|C\varepsilon^{2n}(t)\right\|_Y^2 \mathrm{d}t &\leqslant \frac{1}{2(r-\mu)}\left(\left\|\varepsilon^{2n}(0)\right\|_X^2 - \left\|\varepsilon^{2n}(T)\right\|_X^2\right) \\ &\leqslant \frac{1}{2(r-\mu)}\left(\left\|\varepsilon^{2n}(0)\right\|_X^2 - \left\|\varepsilon^{2(n+1)}(0)\right\|_X^2\right).\end{aligned}$$

Hence,

$$\int_0^T \left\|C\varepsilon^{2n}(t)\right\|_Y^2 \mathrm{d}t \underset{n\to+\infty}{\longrightarrow} 0. \tag{6.29}$$

According to the Duhamel's formula, for all $n \in \mathbb{N}$,

$$\varepsilon^{2n}(t) = \mathbb{T}(t,0)\varepsilon^{2n}(0) - r\int_0^t \mathbb{T}(t,s)C^*C\varepsilon^{2n}(s)\mathrm{d}s. \tag{6.30}$$

Then

$$\begin{aligned}W(0,T)\varepsilon^{2n}(0) &= \int_0^T \mathbb{T}(t,0)^*C^*C\mathbb{T}(t,0)\varepsilon^{2n}(0)\mathrm{d}t \\ &= \int_0^T \mathbb{T}(t,0)^*C^*C\varepsilon^{2n}(t)\mathrm{d}t \\ &\quad + r\int_0^T \mathbb{T}(t,0)^*C^*C\int_0^t \mathbb{T}(t,s)C^*C\varepsilon^{2n}(s)\mathrm{d}s\mathrm{d}t.\end{aligned}$$

According to (6.3) and because C is bounded,

$$\begin{aligned}\left\|W(0,T)\varepsilon^{2n}(0)\right\|_X &\leqslant Me^{\omega T}\|C\|_{\mathscr{L}(X,Y)}\int_0^T \left\|C\varepsilon^{2n}(t)\right\|_Y \mathrm{d}t \\ &\quad + rTM^2e^{2\omega T}\|C\|^3_{\mathscr{L}(X,Y)}\int_0^T \left\|C\varepsilon^{2n}(t)\right\|_Y \mathrm{d}t.\end{aligned}$$

Hence $W(0,T)\varepsilon^{2n}(0) \to 0$ as $n \to +\infty$.

Remark 6.19. Reasoning as in Remark 5.42, if the pair $((A(t))_{t\in[0,T]}, C)$ is exactly observable, then (6.6)-(7.62) is a strong exponential observer with arbitrary decay rate by tuning the observer gain r.

Now, let Ω be the set of limit points of $(\varepsilon^{2n}(0))_{n\geqslant 0}$ for the weak topology of X, that is, the set of points $\xi \in X$ such that there exists a subsequence $(n_k)_{k\geqslant 0}$ such that $\varepsilon^{2n_k}(0) \overset{w}{\rightharpoonup} \xi$ as $k \to +\infty$. Since $(\varepsilon^{2n}(0))_{n\geqslant 0}$ is bounded in X (because L is a contraction), by Kakutani's theorem (see, *e.g.,* [Bre11, Theorem 3.17]), the set $\{\varepsilon^{2n}(0), n \in \mathbb{N}\}$ is relatively weakly compact in X. Hence Ω is not empty. Let $\xi \in \Omega$ and $(\varepsilon^{2n_k}(0))_{k\geqslant 0}$ be a subsequence converging weakly to ξ. Then $W(0,T)\xi = 0$ by uniqueness of the weak limit. Thus $\Omega \subset \ker W(0,T)$. Let $\psi \in X$. By definition of Ω, and since $(\varepsilon^{2n}(0))_{n\geqslant 0}$ is bounded, for all $\eta > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, there exits $\xi_n \in \Omega$ such that

$$|\left\langle \varepsilon^{2n}(0) - \xi_n, \psi \right\rangle_X| \leqslant \eta.$$

Then, if $\psi \in (\ker W(0,T))^\perp$, $\langle \xi_n, \psi\rangle_X = 0$ which yields

$$\left|\left\langle \varepsilon^{2n}(0), \psi \right\rangle_X\right| \leqslant \left|\left\langle \varepsilon^{2n}(0) - \xi_n, \psi \right\rangle_X\right| + |\langle \xi_n, \psi\rangle_X| \leqslant \eta,$$

i.e.,

$$\left\langle \varepsilon^{2n}(0), \psi \right\rangle_X \xrightarrow[n\to+\infty]{w} 0, \qquad \forall \psi \in \bigcup_{\tau\geqslant 0} (\ker W(0,T))^\perp.$$

This ends the proof of Theorem 6.12. ■

6.5.2 Proof of Theorem 6.13

Proof of Theorem 6.13. Assume that $T < +\infty$ and $(\mathbb{T}(t,s))_{0\leqslant s,t\leqslant T}$ is a bi-directional evolution system. Suppose that $((A(t))_{t\in[0,T]}, C)$ and $((-A(t))_{t\in[0,T]}, C)$ are μ-weakly detectable and $r > \mu$. Assume also that $\mathcal{O}_T = X$ and $\mathbb{S}_-(0,T) = \mathbb{S}_+(T,0)^*$. We follow the same strategy as in the proof of Theorem 5.35 (see Section 5.5.2).

Let $L = \mathbb{S}_-(0,T)\mathbb{S}_+(T,0) = \mathbb{S}_+(T,0)^*\mathbb{S}_+(T,0)$ (as in the proof of Theorem 6.12, Section 6.5.1). Then, it is sufficient to prove that for all $\varepsilon_0 \in \mathcal{O}_\tau$, $L^n\varepsilon_0 \to 0$ as $n \to +\infty$. The operator L is self-adjoint positive-definite since $\mathbb{S}_+(\tau,0)$ is bounded from below (since $\mathbb{S}_+$ is bi-directional). Let $\varepsilon_0 \in X$. The hypotheses of Lemma 6.18 hold. Hence, L is a contraction and (6.27) yields

$$\langle L\varepsilon_0, \varepsilon_0\rangle_X = \|\mathbb{S}_+(T,0)\varepsilon_0\|_X^2 \leqslant \|\varepsilon_0\|_X^2 - 2(r-\mu)\int_0^\tau \|C\mathbb{S}_+(t,0)\varepsilon_0\|_Y^2 \,\mathrm{d}t. \tag{6.31}$$

From there, the proof is identical to the proof of Theorem 5.35, from equation (5.31) to the end, by replacing τ by T, $\mathbb{S}$ by $\mathbb{S}_+$ and $\mathcal{O}_\tau$ by X. Hence, $L^n\varepsilon_0 \to 0$ as $n \to \infty$, which ends the proof of Theorem 6.13. ■

Chapter 7

Observers for a crystallization process

Among all of the mathematical disciplines the theory of differential equations is the most important... It furnishes the explanation of all those elementary manifestations of nature which involve time.

S. Lie

Abstract. *During a batch crystallization process, the Particle Size Distribution (PSD) is of major importance. However, measuring the PSD is difficult, and a popular approach is to estimate the PSD from other measurements. In this chapter we focus on three measures: temperature, solute concentration, and Chord Length Distribution (CLD). After modeling the process and the sensors, we propose different strategies of estimation. The first one is a direct approach based on a Tikhonov regularization procedure using the CLD but not relying on the dynamical model of the PSD. The second one is a Kazantzis-Kravaris/Luenberger (KKL) observer using only temperature and solute concentration as measurements. The last one is an infinite-dimensional Luenberger observer using the CLD based on the theory established in Chapters 5 and 6, still effective when polymorphism occurs.*

Contents

Figures and Tables

Introduction

Crystallization is one of the oldest and major processes used in industry (chemical, pharmaceutical, food, *etc.*) to produce, purify or separate solid compounds or products [Bis13]. This unit operation aims to produce solid crystals with well defined specifications including (among others) the Particle Size Distribution (PSD) which is of critical importance. At the industrial scale, the PSD is neither well controlled nor monitored during the crystallization process and a grinding step is usually performed before delivering the final product. Measuring a PSD remains a challenging problem, tackled by modern Process Analytical Technologies (PATs) with various measures and approaches, such as image processing [Pre+10, Gao+18], dynamical observers and moments based methods [Mes+11,Ucc11,Vis12,Leb+15,Gru17,PÖ17] Some PATs, such as the Focused Beam Reflectance Measurement (FBRM) or the BlazeMetrics® technologies, give access to the Chord Length Distribution (CLD) [LCK98, WHM05, Agi+15, PR16].

In this chapter, we aim to address the problem of reconstructing the PSD from thee measurements: temperature, solute concentration and CLD. In Section 7.1, a model of the process is derived from a population balance equation, as well as a model of the measurements, depending on the shape of the crystals. A direct

reconstruction of the PSD from the measurements, based on inverse problems techniques, is investigated in Section 7.3. But this approach makes no use of the system dynamics, and fails in many situations. On the contrary, in Section 7.4, we build two state observers using the measurements. We consider both the online and the offline estimation problems seen in Chapters 5 and 6, respectively. In particular, we apply the results of Chapter 6 to prove the convergence of a back and forth observer reconstruction the PSD from the knowledge of the CLD over a finite time interval.

7.1 Modeling the batch crystallization process

7.1.1 Population balance in the single-shape case

In a first step a batch crystallization process is modeled in the case where the size of the crystals is described by a single scalar parameter r, and all crystals have the same shape. Typically, crystals are spherical and r represents their radius. We denote by $\psi(t,\cdot)$ the PSD at time t in the reactor, so that $\int_{r_1}^{r_2}\psi(t,r)\mathrm{d}x$ is the number of crystals in the reactor at time t having a radius r between r_1 and r_2. Let $r_{\max}$ be a maximal radius that no crystals of any shape can reach during the process (such as the size of the reactor):

$$\psi(t,r_{\max})=0, \qquad \forall t\in[0,t_{\max}]. \tag{7.1}$$

We assume that all crystals appear at the same minimal radius $r_{\min}>0$, and denote by $u(t)$ the appearance of particles of size $r_{\min}$ at time t:

$$\psi(t,r_{\min})=u(t), \qquad \forall t\in[0,t_{\max}]. \tag{7.2}$$

The function u is linked to the nucleation rate R and the growth rate G by the following formula:

$$u(t)=\frac{R_n(t)}{G(t)}. \tag{7.3}$$

Note however that in our approaches for PSD estimation, we do not need to know precisely this expression. We do not use any model of u, and assume this quantity to be unknown. The growth rate is supposed to be positive at any time. Moreover, considering the McCabe assumption, G is independent of the crystals size. The population balance leads to (see *e.g.*, [MEH01, Mul01])

$$\frac{\partial\psi}{\partial t}(t,r)+G(t)\frac{\partial\psi}{\partial r}(t,r)=0, \tag{7.4}$$

i.e., a one-dimensional time-varying transport equation. Finally, assume that seed particles with PSD ψ_0 may lie in the reactor at time $t=0$:

$$\psi(0,r)=\psi_0(r), \qquad \forall r\in[r_{\min},r_{\max}]. \tag{7.5}$$

To summarize, the evolution of the PSD through the process follows the set of partial differential equations (PDEs)

$$\begin{cases} \frac{\partial\psi}{\partial t}(t,r)+G(t)\frac{\partial\psi}{\partial r}(t,r)=0 & \forall t\in(0,t_{\max}),\forall r\in(r_{\min},r_{\max}) \\ \psi(0,r)=\psi_0(r) & \forall r\in[r_{\min},r_{\max}] \\ \psi(t,r_{\min})=u(t) & \forall t\in[0,t_{\max}] \end{cases} \tag{7.6}$$

with the additional boundary condition (7.1).

7.1.2 Well-posedness

The well-posedness is ensured by the following result.

Theorem 7.1. *If G is positive and continuous, $\psi_0 \in L^2((r_{\min}, r_{\max}); \mathbb{R})$ and $u \in L^2((0, t_{\max}); \mathbb{R})$, then (7.6) admits a unique solution $\psi \in C^0([0, t_{\max}]; L^2((r_0, r_1); \mathbb{R}))$.*

Moreover, for all $t \in [0, t_{\max}]$ and almost all $r \in [r_{\min}, r_{\max}]$,

$$\psi(t,r) = \begin{cases} \psi_0(r - \mathfrak{G}(t)) & \text{if } r - r_{\min} \geqslant \mathfrak{G}(t) \\ u \circ \mathfrak{G}^{-1}(\mathfrak{G}(t) - r + r_{\min}) & \text{else.} \end{cases} \tag{7.7}$$

where $\mathfrak{G} : [0, t_{\max}] \ni t \mapsto \int_0^t G(\tau)\mathrm{d}\tau$.

Moreover, if $\psi_0 \in H^1((r_{\min}, r_{\max}); \mathbb{R})$, $u \in H^1((0, t_{\max}); \mathbb{R})$ and $u(0) = \psi_0(r_{\min})$, then

$$\psi \in C^0([0, t_{\max}]; H^1(r_{\min}, r_{\max})) \cap C^1([0, t_{\max}]; L^2(r_{\min}, r_{\max})).$$

The proof of this theorem can be found in [Cor07, Theorem 2.4] in the case $G = 1$, and can be easily adapted by means of a time reparametrization (set $\frac{\mathrm{d}t_{\text{new}}}{\mathrm{d}t} = G(t)$). Alternatively, a proof based on evolution systems theory (see Section 5.1.2) is given in Section 7.4.2. It is worth noticing that this theorem does not take into account condition (7.1). However, the following proposition holds.

Proposition 7.2. *Assume that the hypotheses of Theorem 7.1 are satisfied. Assume that $\psi_0(r) = 0$ for all $r \in [\bar{r}, r_{\max}]$ for some $\bar{r} \in [r_{\min}, r_{\max})$. If*

$$\bar{r} + \mathfrak{G}(t_{\max}) < r_{\max}, \tag{7.8}$$

then $\psi(t, r) = 0$ for all $t \in [0, t_{\max}]$ and all $r \in [\bar{r} + \mathfrak{G}(t_{\max}), r_{\max}]$.

Proof. Let $t \in [0, t_{\max}]$ and $r \in [\bar{r} + \mathfrak{G}(t_{\max}), r_{\max}]$. Then

$$r - \mathfrak{G}(t_{\max}) \geqslant \bar{r} \geqslant r_{\min}.$$

Consequently, according to (7.7), $\psi(t, r) = \psi_0(r - \mathfrak{G}(t)) = 0$. ■

Hence, one must choose $r_{\max}$ large enough so that the particles do not reach the size $r_{\max}$ in time $t_{\max}$. In the rest of the chapter, we always assume that (7.1) is satisfied.

7.1.3 Multi-shape case

Polymorphism is a common phenomenon that may occur during crystallization: crystals may have several metastable shapes. We assume that only a finite number N of shapes may appear during the process, and that the size of a crystal having the shape $i \in \{1, \dots, N\}$ is still described by a single parameter r. Denoting by ψ_i the PSD associated to each shape $i \in \{1, \dots, N\}$ and reasoning as in the single-shape case, we get the following set of partial differential equations (PDEs)

$$\forall i \in \{1, \dots, N\}, \begin{cases} \frac{\partial \psi_i}{\partial t}(t,r) + G_i(t)\frac{\partial \psi_i}{\partial r}(t,r) = 0 & \forall t \in (0, t_{\max}), \forall r \in (r_{\min}, r_{\max}) \\ \psi_i(0,r) = \psi_{0,i}(r) & \forall r \in [r_{\min}, r_{\max}] \\ \psi_i(t, r_{\min}) = u_i(t) & \forall t \in [0, t_{\max}] \end{cases} \tag{7.9}$$

with the additional boundary condition

$$\psi_i(t, r_{\max}) = 0, \qquad \forall t \in [0, t_{\max}],\ \forall i \in \{1, \dots, N\}. \tag{7.10}$$

Note that each shape i has a specific growth rate G_i, nucleation u_i and initial condition $\psi_{0,i}$. Since PSDs of different shapes do not interact with each other, Theorem 7.1 still ensures the well-posedness of (7.9).

7.2 Modeling the measurements

7.2.1 Solute concentration and temperature

First, we consider the case where the measured outputs are the temperature and the solute concentration (denoted by $C_c(t)$). We restrict ourselves to the single-shape case. These two measurements allow to obtain online estimation of the growth rate (*i.e.* $G(t)$) and the third moment of the PSD (denoted by $y(t)$).

Estimation of G

The knowledge of the temperature and the solute concentration allows to obtain an approximation of the growth rate G. Indeed, following [Ucc11], a model of G can be given for all times $t \in [0, t_{\max}]$ by

$$G(t) = k_g \frac{C_c(t) - C^*(t)}{C^*(t)} \tag{7.11}$$

where

- k_g is a known growth rate parameter (in $\mathrm{m.s}^{-1}$),
- $C^*(t)$ is the solubility at time t (in kg of solute per kg of solvent),
- $C_c(t)$ is the solute concentration at time t (in kg of solute per kg of solvent).

Since $C^*(t)$ depends on the temperature at time t, the growth rate G of the crystals can be estimated online with the available sensors. Other model expressions of G are available in the literature, for more details one may refer to [MEH01, Mul01]. The two dynamical observers developed in Section 7.4 will use the knowledge of G in the observer design.

Estimation of the third moment of the PSD

It is possible to link the solute concentration with the PSD. Indeed, for each $t \in \mathbb{R}_+$, let $C_s(t)$ (in kg of solid per kg of solvent) be the solid concentration in the reactor at time t, in other words, the ratio between the total crystals mass in the reactor at time t and the solvent mass. Let ρ_s (in $\mathrm{kg.m}^{-3}$) be the density of the solute in solid phase and M_e the solvent mass (in kg). It yields:

$$C_s(t) = \frac{\rho_s}{M_e} V_s(t)$$

where $V_s(t)$ is the volume (in m^3) occupied by the crystals at time t. Then the volume of a crystal with size r (in m) is simply $V = k_v r^3$ where k_v is a volumetric shape factor (see *e.g.* [HK64, RL88]). For example, $k_v = \frac{4\pi}{3}$ for spherical crystals. The total volume of the crystals is then

$$V_s(t) = k_v \int_{r_{\min}}^{r_{\max}} \psi(t,r) r^3 \mathrm{d}r$$

in the single-shape case. Hence, the solid concentration in the reactor can be expressed as follows.

$$\forall t \in [0, t_{\max}], \quad C_s(t) = \frac{\rho_s k_v}{M_e} \int_{r_{\min}}^{r_{\max}} \psi(t,r) r^3 \mathrm{d}r. \tag{7.12}$$

Assume moreover that ρ_s is a known parameter. This implies that we can associate to system (7.6) the measurement y defined as

$$\forall t \in [0, t_{\max}], \quad y(t) = \int_{r_{\min}}^{r_{\max}} \psi(t,r) r^3 \mathrm{d}r, \tag{7.13}$$

that is the third moment of $\psi(t,\cdot)$. The purpose of Section 7.4.1 is to propose an observer to solve the problem of online estimation of ψ from the knowledge of y and G.

Remark 7.3 (Multi-shape case)**.** In the multi-shape case, we would obtain the following measurement:

$$\forall t \in [0, t_{\max}], \quad C_s(t) = \sum_{i=1}^{N} \frac{\rho_{s_i} k_{v_i}}{M_e} \int_{r_{\min}}^{r_{\max}} \psi(t,r) r^3 \mathrm{d}r. \tag{7.14}$$

where k_{v_i} and ρ_{s_i} are respectively the volumetric shape factor and the density in solid phase associated to the shape i. But we will not use this measurement in the multi-shape case.

7.2.2 Chord Length Distribution

The FBRM and BlazeMetrics® technologies are *in situ* sensors measuring data online during a crystallization process. The probe is equipped with a laser beam in rotation that scans across the particles. While the beam hit a particle, light is backscattered to the probe. The sensor counts the number of distinct light pulses and their duration. For each pulse, a length on a particle (*i.e.,* a chord length) can be determined, since the rotation speed of the beam is known and the speed of the particle is supposed to be insignificant. Hence, one can deduce the Chord Length Distribution (CLD) of the particles. The reader may refer to [BG99, SLB99, LW05] for more details about this technology, and how it is linked to the CLD. Using a CLD to recover the corresponding PSD is a major current issue in process engineering.

Understanding the PSD-to-CLD relation is an essential step in recovering the desired PSD when using the above-mentioned technologies. Naturally, this relation is heavily influenced by the shape of the particles. In [Hob+91, BG99, Lan+01], the authors considered spherical particles. It often occurs in crystallization processes that particles cannot be assumed to have such symmetries. In [Agi+15] for

instance, needle-shaped particles were modeled as cylinders. In this thesis, we consider crystals whose shape can be approximated by a spheroid (also called ellipsoid of revolution). A spheroid is a surface of revolution, obtained as the rotation of an ellipse along one of its two principal axes. In particular, spheres are spheroids. These shapes have the advantage of allowing to model both spheres and elongated needle-shaped particles with only one shape tuning parameter. In that respect, we gather different shapes under the same mathematical umbrella while retaining many computational properties of the spherical model. Note that, unlike [LW05] who considered two-dimensional ellipses, we consider proper three-dimensional spheroids that can be measured by the probe in any possible orientation. Spheroids were already considered in [Kel84], but the experimental assumptions lead to differing probabilistic models and distributions.

From spheroid geometry to chord length

When scanning across some particles, the sensor measures chords on the projection of the particle on the plane that is orthogonal to the probe's laser beam. Hence, two sources of hazards must be considered to model the random choice of the chords measured by the sensor:

- choice of orientation of the spheroid with respect to the probe;
- choice of the chord on the projection of the spheroid with selected orientation.

Step 1: Choosing an orientation. A spheroid of radius r in elementary orientation can be represented as the set of points $(x, y, z) \in \mathbb{R}^3$ such that

$$\begin{pmatrix} x & y & z \end{pmatrix} D \begin{pmatrix} x \\ y \\ z \end{pmatrix} \leqslant r^2 \quad \text{with} \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\eta^2} \end{pmatrix}. \tag{7.15}$$

The parameter η is the ratio of the diameter of the spheroid along the axis of rotation by the diameter perpendicular to this axis. It characterizes the eccentricity of the spheroid. The spheroid is said to be prolate if $\eta > 1$ and oblate if $\eta < 1$. When $\eta = 1$, the particle is a sphere. The volume of such a particle is given by $\frac{4\pi}{3}\eta r^3$.

Without loss of generality, we assume that the probe's laser beam is parallel to the z-axis. The solid can be oriented in any direction in space. Since the solid is a spheroid, it has an axis of symmetry and any orientation is equivalent to picking a point on the sphere in 3d space, corresponding, for instance, to the position of the north pole of the spheroid (see Figure 7.1). For this reason, we obtain an orientation following spherical coordinates. Hence a sequence of two rotations of the elementary spheroid (7.15) allows to choose any possible orientation.

- First, we rotate the space around the y-axis with an angle $\theta \in [0, \pi]$, leading to a change of coordinates of the matrix

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

- Second, we rotate the space around the z-axis with an angle $\phi \in [0, 2\pi]$, leading to a change of coordinates of the matrix

$$\rho_z(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The change of coordinates $(x, y, z)^\top \mapsto \rho_z(\phi)\rho_y(\theta)(x, y, z)^\top$ has the effect of mapping the point $(0, 0, 1)$ to any point on the sphere. Furthermore, it is an isometry. If $(\phi, \theta) \in [0, 2\pi] \times [0, \pi]$ is picked according to the probability measure $\mathrm{d}\mu = \frac{\sin\theta}{4\pi}\mathrm{d}\phi\mathrm{d}\theta$, this equals to *uniformly* picking a random orientation for the spheroid (that is, the measure μ gives a uniform probability of picking a point on the sphere). Then, the change of coordinates implies that the rotated spheroid has equation

$$\begin{pmatrix} x & y & z \end{pmatrix} A \begin{pmatrix} x \\ y \\ z \end{pmatrix} \leqslant r^2 \quad \text{with} \quad A = \rho_z(\phi)\rho_y(\theta)\, D\, \rho_y(-\theta)\rho_z(-\phi). \tag{7.16}$$

In Figure 7.1, the elementary spheroid (7.15) (on the left) is rotated by $\rho_z(\phi)\rho_y(\theta)$ to obtain the rotated spheroid (7.16) (on the right).

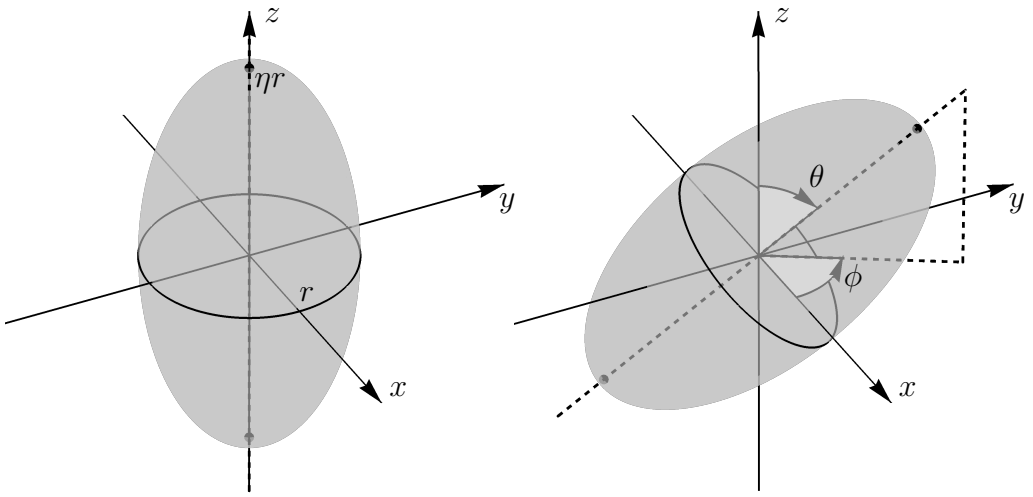


Figure 7.1 – *On the left:* elementary spheroid of parameter η and radius r (equation (7.15)). *On the right:* rotation of the elementary spheroid with angles ϕ, θ (equation (7.16))

Step 2: Projecting the spheroid on the (x, y)-plane. Given an arbitrary orientation of the particle in space, the sensor measure (assumed parallel to the z-axis) is the same as the one given by the ellipse obtained by projection of the solid on the (x, y)-plane. Hence, the next step is to transfer the geometry of the 3d spheroid onto its shadow in the (x, y)-plane. The shell of the spheroid is given by $(x, y, z)A(x, y, z)^\top = r^2$ for some $(\phi, \theta) \in [0, 2\pi] \times [0, \pi]$. For completeness sake, the full expression of matrix A is the following:

$$\begin{pmatrix} \bar{s}\cos^2\phi + \sin^2\phi & -\bar{\eta}\sin^2\theta\sin 2\phi & -\bar{\eta}\sin 2\theta\cos\phi \\ -\bar{\eta}\sin^2\theta\sin 2\phi & \bar{s}\sin^2\phi + \cos^2\phi & -\bar{\eta}\sin 2\theta\sin\phi \\ -\bar{\eta}\sin 2\theta\cos\phi & -\bar{\eta}\sin 2\theta\sin\phi & \frac{1}{\eta^2}\cos^2\theta + \sin^2\theta \end{pmatrix} \quad \text{with} \quad \begin{array}{c} \bar{\eta} = \frac{\eta^2-1}{2\eta^2}, \\ \bar{s} = \frac{\sin^2\theta + \eta^2\cos^2\theta}{\eta^2}. \end{array}$$

If we are looking at points that appear at the edge of the shadow of the spheroid, it is clear that these must be such that the tangent plane to the spheroid at that point is vertical (see Figure 7.2).

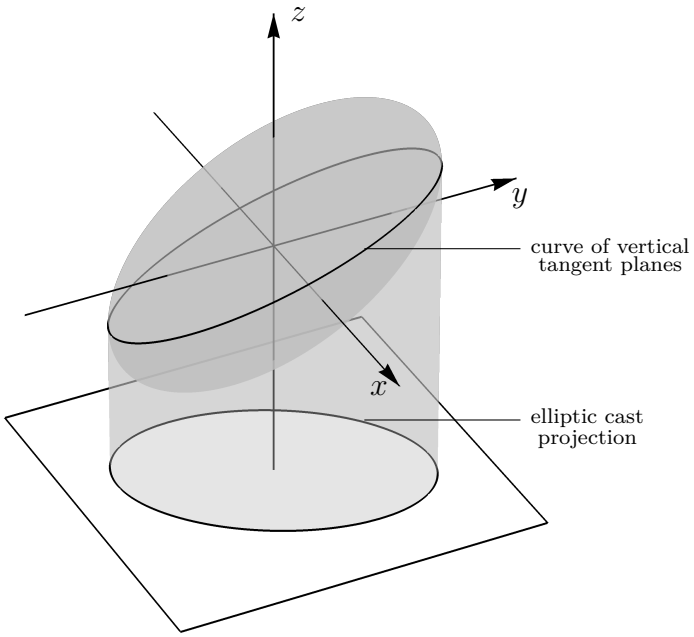


Figure 7.2 – Projection of a spheroid on the (x, y)-plane.

Since the spheroid is given by an implicit definition of the form $g(x, y, z) = r^2$, the tangent plane to the spheroid at a point (x, y, z) is actually the plane that is orthogonal to $\nabla g(x, y, z)$, the gradient of g at (x, y, z). Hence, to find points (x, y) in the plane that lie at the border of the shadow cast by the spheroid, we solve

$$g(x, y, z) = r^2, \qquad (\nabla g(x, y, z))^\top \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0.$$

In the case of a spheroid, $g(x, y, z) = (x, y, z)A(x, y, z)^\top$, hence $\nabla g(x, y, z) = A \cdot (x, y, z)^\top$. In other words, we solve

$$\begin{pmatrix} x & y & z \end{pmatrix} A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = r^2, \qquad \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0.$$

In the (x, y)-plane, solutions to this pair of equations are points of the planar ellipse

$$\alpha x^2 + \beta y^2 + \gamma xy = r^2, \tag{7.17}$$

with

$$\alpha = \frac{\cos^2 \phi}{\cos^2 \theta + \eta^2 \sin^2 \theta} + \sin^2 \phi, \tag{7.18}$$

$$\beta = \frac{\sin^2 \phi}{\cos^2 \theta + \eta^2 \sin^2 \theta} + \cos^2 \phi, \tag{7.19}$$

$$\gamma = -\frac{(\eta^2 - 1) \sin^2 \theta \sin 2\phi}{\cos^2 \theta + \eta^2 \sin^2 \theta}. \tag{7.20}$$

Naturally, $\cos^2\theta+\eta^2\sin^2\theta > 0$ for all $\eta > 0$ and $\theta \in [0,\pi]$. In conclusion, the shadow cast by the spheroid has the shape of an ellipse of orientation and eccentricity determined by the quantities α, β, γ, themselves functions of ϕ, θ and η. When necessary, we write α_η to underline the η-dependence.

Step 3: Choosing a chord on the projection. Since we considered all the possible orientations of the spheroid in space, we can consider with no loss of generality that the probe's laser cut the two-dimensional projection (7.17) at constant y. Hence, the length of a chord on (7.17) at some constant $y \in \mathbb{R}$ is the distance between the two x-solutions, if they exist, of

$$\alpha x^2 + \gamma y x + \beta y^2 - r^2 = 0. \tag{7.21}$$

Let $\Delta = \gamma^2 y^2 - 4\alpha(\beta y^2 - r^2)$ be the discriminant of the quadratic and let

$$y_{\max} = \frac{2\sqrt{\alpha} r}{\sqrt{4\alpha\beta - \gamma^2}}. \tag{7.22}$$

(Let us specify that $4\alpha\beta - \gamma^2 = \frac{8}{1+\eta^2+(\eta^2-1)\cos 2\theta} > 0$ for all $\eta > 0$ and all $\theta \in [0,\pi]$.) If $|y| \leqslant y_{\max}$, then $\Delta \geqslant 0$ and the length of the chord cutting (7.17) at y is given by

$$\ell = \frac{\sqrt{\Delta}}{\alpha}. \tag{7.23}$$

Otherwise, if $|y| > y_{\max}$, *i.e.*, $\Delta < 0$, then no chord cuts the ellipse at y. Hence, the maximum chord length is $\frac{2r}{\sqrt{\alpha}}$, reached at $y = 0$. For all $\ell \in [0, \frac{2r}{\sqrt{\alpha}}]$, let y_ℓ be such that the chord length ℓ is reached at $y = y_\ell$, so that y_ℓ is implicitly defined by (7.23):

$$y_\ell = \frac{\sqrt{4\alpha r^2 - \alpha^2\ell^2}}{\sqrt{4\alpha\beta - \gamma^2}}. \tag{7.24}$$

If $\ell > \frac{2r}{\sqrt{\alpha}}$, adopt the convention $y_\ell = 0$. Doing so, y_ℓ is a continuous function of ℓ. These notations are summarized in Figure 7.3.

To conclude, for a given spheroid of radius r and ratio η with orientation (ϕ, θ) in space, the chord length ℓ is measured by the sensor when cutting the projection of the particle on the (x, y)-axis at constant $y = y_\ell$.

From spheroid distribution to cumulative CLD

Let us denote by $\psi(r)$ a PSD of spheroids of parameter η (dimensionless) and radius r between $r_{\min}$ and $r_{\max}$, generating a CLD measured by the sensor in a batch reactor. For $r_1 < r_2$, the integral $\int_{r_1}^{r_2} \psi(r)\mathrm{d}r$ represents the number of particles with radius r between r_1 and r_2 per unit of volume. The corresponding CLD is denoted by $q(\ell)$. Note that the largest possible chord of a spheroid of radius r is the diameter of the spheroid, namely, $\ell_{\max} = 2r_{\max}\max(\eta, 1)$. Hence $0 \leqslant \ell \leqslant \ell_{\max}$. Then $\int_{\ell_1}^{\ell_2} q(\ell)\mathrm{d}\ell$ represents the number of chords with length ℓ between ℓ_1 and ℓ_2 measured by the sensor per unit of volume. The cumulative CLD is denoted by $Q(\ell) = \int_0^\ell q(l)\mathrm{d}l$. Then, the normalized functions $\bar{\psi}(r) = \frac{1}{\int_{r_{\min}}^{r_{\max}} \psi(\rho)\mathrm{d}\rho}\psi(r)$ and $\bar{q}(\ell) = \frac{1}{Q(\ell_{\max})}q(\ell)$ are probability density functions and $\bar{Q}(\ell) = \frac{1}{Q(\ell_{\max})}Q(\ell)$ is a cumulative distribution function (dimensionless).

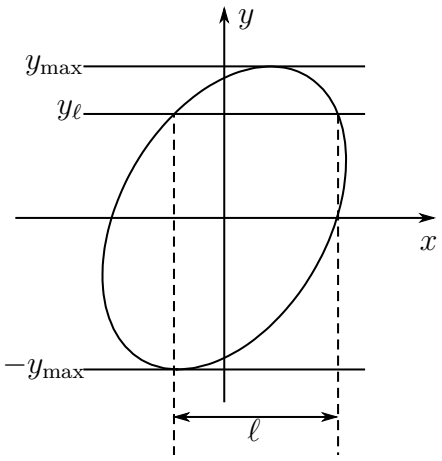


Figure 7.3 – Length ℓ of an horizontal chord on an ellipse at $y = y_\ell \in [-y_{\max}, y_{\max}]$.

Let R be a random variable representing the radius of a particle, and L be a random variable representing a measured chord length. By law of total expectation,

$$\bar{Q}(\ell) := \int_0^\ell \bar{q}(l)\mathrm{d}l = \mathbb{P}(L < \ell) = \mathbb{E}(\mathbb{1}_{L\leqslant\ell}) = \mathbb{E}(\mathbb{E}(\mathbb{1}_{L\leqslant\ell}|R)) = \int_{r_{\min}}^{r_{\max}} k(\ell, r)\bar{\psi}(r)\mathrm{d}r. \tag{7.25}$$

where

$$k(\ell, r) = \mathbb{P}(L < \ell \ : \ R = r)$$

encodes the probability of measuring a chord length less than ℓ assuming a particle of radius r crosses the sensor. Hence

$$Q(\ell) = \kappa \int_{r_{\min}}^{r_{\max}} k(\ell, r)\psi(r)\mathrm{d}r \tag{7.26}$$

where

$$\kappa = \frac{Q(\ell_{\max})}{\int_{r_{\min}}^{r_{\max}} \psi(r)\mathrm{d}r}$$

is the ratio between the number of particles and the number of chords measured by the sensor, which depends on the experimental conditions.

For a given radius r, and a given orientation of the particle, encoded by (ϕ, θ), the chord length is measured according to the situation described in the previous section. That is, the chord length corresponds to a chord length at constant y for an ellipse in the (x, y)-plane (of shape determined by r, ϕ, θ and η). Then, $L < \ell$ is achieved if the horizontal chord has ordinate y belonging to the set

$$(-y_{\max}, -y_\ell) \cup (y_\ell, y_{\max}) \tag{7.27}$$

where $y_{\max}$ is as in (7.22) and y_ℓ as in (7.24). Since $\ell < \frac{2r}{\sqrt{\alpha}}$ with α as in (7.18), the probability that $L < \frac{2r}{\sqrt{\alpha}}$ is full. Hence the probability that the measured chord length L is less than ℓ is given by

$$\frac{2(y_{\max} - y_\ell)}{2y_{\max}} = 1 - \sqrt{1 - \left(\frac{\ell}{2r}\right)^2 \alpha},$$

which means that the ordinate of the chord length is chosen uniformly in the set (7.27).

Uniformly choosing an orientation of the spheroid means that the angles (ϕ, θ) are picked in $[0, 2\pi] \times [0, \pi]$ according to the probability measure $\mathrm{d}\mu = \frac{\sin\theta}{4\pi}\mathrm{d}\phi\mathrm{d}\theta$. Then, by the law of total expectation,

$$k(\ell, r) = 1 - \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \sqrt{1 - \left(\frac{\ell}{2r}\right)^2 \alpha_\eta(\phi, \theta)}\frac{\sin\theta}{4\pi}\mathrm{d}\theta\mathrm{d}\phi, \tag{7.28}$$

with

$$\alpha_\eta(\phi, \theta) = \frac{\cos^2\phi}{\cos^2\theta + \eta^2\sin^2\theta} + \sin^2\phi. \tag{7.29}$$

Remark 7.4 (Spheres)**.** For spherical crystals (*i.e.,* when $\eta = 1$), (7.29) yields $\alpha_1 = 1$. Hence, (7.28) has the simpler expression

$$k(\ell, r) = 1 - \sqrt{1 - \left(\frac{\ell}{2r}\right)^2} \tag{7.30}$$

Combining the expression of k with (7.26), we get a function that maps a PSD of spheroids to the corresponding cumulative CLD up to the constant κ. In particular, if $\bar{\psi}$ is a Dirac distribution at some fixed radius r (which means that all particles have the same radius r), then (7.25) yields $\bar{Q}(\ell) = k(\ell, r)$. In Figure 7.4, we plot $\bar{Q}(\ell)$ for a Dirac distribution of particles at $r = 1\text{mm}$, and three different values of η. This emphasizes the influence of the shape parameter on the CLD.

Remark 7.5 (Multi-shape case)**.** In the multi-shape case (see Section 7.1.3), the CLD data collected by the sensor is the sum of the CLDs associated to each PSD. More precisely, with the notations of (7.26), the measured cumulative CLD Q satisfies

$$Q(\ell) = \sum_{i=1}^{N} \kappa_i \int_{r_{\min}}^{r_{\max}} k_i(\ell, r)\psi_i(r)\mathrm{d}r, \tag{7.31}$$

where k_i is the kernel defined in (7.28) with $\eta = \eta_i$ and

$$\kappa_i = \frac{\int_0^{\ell_{\max}} q(\ell)\mathrm{d}\ell}{\int_{r_{\min}}^{r_{\max}} \psi_i(r)\mathrm{d}r}.$$

7.3 Direct approach

With the measurements modeled in Section 7.2, is it possible to reconstruct directly a PSD, *i.e.,* without using the dynamical model established in Section 7.1? Clearly, the measurement of the third moment of the PSD is insufficient: it is easy to construct two distributions having the same third moment and yet being very different in L^2. However, there is much more information in the CLD: it is an infinite-dimensional measurement. In this section, we show that it is possible, in the single-shape case and for spheroid particles, to reconstruct the PSD from the CLD, up to a multiplicative factor (that can be determined with a measure of the solute concentration). In particular, since the dynamical model is not used, this approach may be used for any suspended particles whose chords are measured by a sensor, and not only for

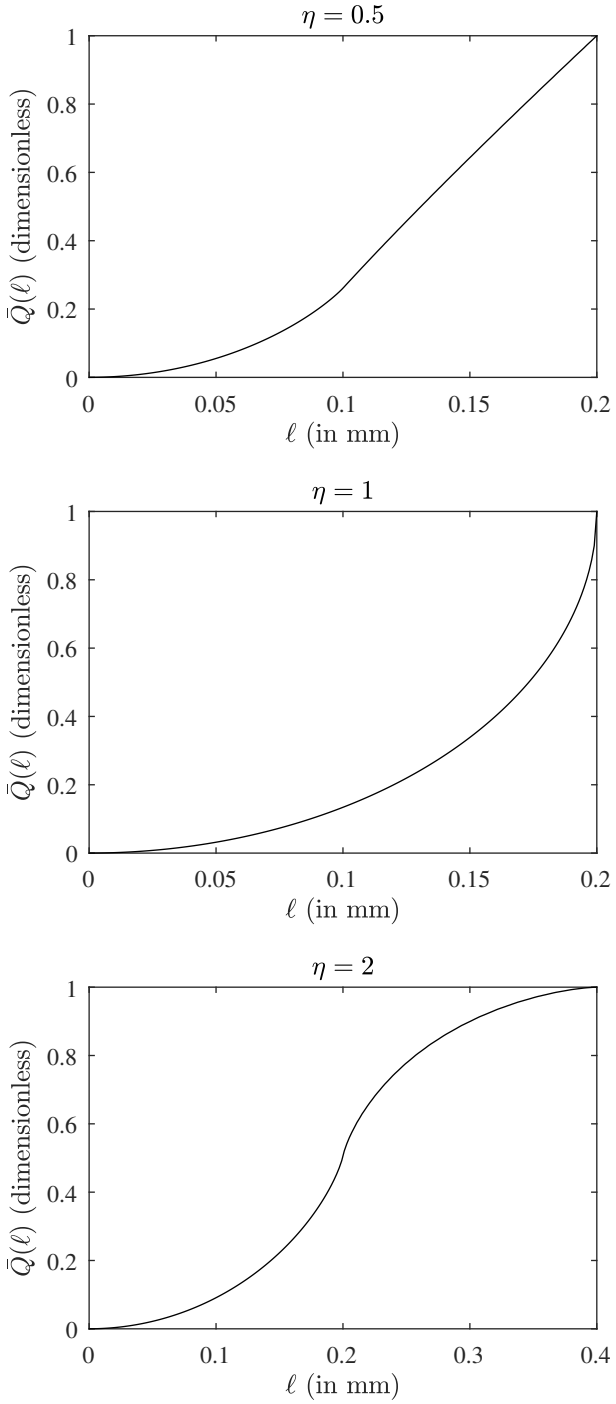


Figure 7.4 – Normalized CLD $\bar{Q}$ associated to a Dirac distribution of spheroids at $r = 1$mm for $\eta = 0.5$, 1, 2.

crystals. We use a Tikhonov regularization procedure. We prove the injectivity of the PSD-to-CLD map, and illustrate the strategy on numerical simulations.

Consider a PSD of spheroids sharing the same shape parameter η. According to (7.26), it is possible to compute the corresponding cumulative CLD up to the knowledge of the parameter κ. Conversely, for a given CLD, is it possible to estimate

the corresponding PSD? This question is a crucial issue in process control. Indeed, PATs like the FBRM sensor are able to measure the CLD online, for example during a crystallization process. But the main distribution to be known, and governing the physico-chemical properties of solids, is the PSD. In this section, we propose a two-steps procedure to recover the PSD from two measures: the CLD, and the solid concentration in the reactor.

1. First, using the knowledge of the CLD and (7.25), we estimate the normalized PSD $\bar{\psi}$.
2. Second, using the CLD and the solid concentration, we estimate the number of particles per unit of volume $\int_{r_{\min}}^{r_{\max}} \psi(r)\mathrm{d}r$.

Combining these steps with the relation

$$\psi(r) = \bar{\psi}(r) \int_{r_{\min}}^{r_{\max}} \psi(\rho)\mathrm{d}\rho, \qquad \forall r \in [r_{\min}, r_{\max}], \tag{7.32}$$

we obtain an estimation of the PSD ψ. Sometimes, the knowledge of the number of particles is not to be determined: only the "shape" of the PSD is of interest. In this case, only the first step needs to be applied. In numerical simulations, we focus on this first step.

7.3.1 Estimation of $\bar{\psi}$ with a Tikhonov regularization procedure

Let $X = L^2((r_{\min}, r_{\max}); \mathbb{R})$ be the set of real square integrable functions over $(r_{\min}, r_{\max})$, and $Y = L^2((0, \ell_{\max}); \mathbb{R})$ with $\ell_{\max} = 2r_{\max}\max(\eta, 1)$. Then a (normalized) PSD may be viewed as an element of X, while a (normalized) CLD is an element of Y. Let us define the following map:

$$\begin{aligned} \mathcal{K} : X &\longrightarrow Y \\ \bar{\psi} &\longmapsto \left(\ell \mapsto \textstyle\int_{r_{\min}}^{r_{\max}} k(\ell, r)\bar{\psi}(r)\mathrm{d}r\right) \end{aligned}$$

Equation (7.25) may be rewritten as

$$\mathcal{K}\bar{\psi} = \bar{Q}. \tag{7.33}$$

For a given CLD q, it is easy to compute the cumulative normalized CLD $\bar{Q}$. Then, reconstructing $\bar{\psi}$ from $\bar{Q}$ is solving the inverse problem (7.33) with unknown $\bar{\psi}$ in $L^2((r_{\min}, r_{\max}); \mathbb{R})$. However, this problem admits a solution only if $\bar{Q}$ lies in the image of $\mathcal{K}$, denoted by $\operatorname{Im}\mathcal{K} = \{\mathcal{K}\bar{\psi}, \bar{\psi} \in X\}$. Due to measurements noise on $\bar{Q}$, this condition is generally not satisfied. To overcome this problem, we reformulate (7.33) as a minimization problem:

$$\textit{Find } \bar{\psi} \in X \textit{ minimizing } \|\mathcal{K}\bar{\psi} - \bar{Q}\|_Y^2. \tag{7.34}$$

where $\|\cdot\|_Y$ denotes the L^2-norm, that is,

$$\|\mathcal{K}\bar{\psi} - \bar{Q}\|_Y^2 = \int_0^{\ell_{\max}} |(\mathcal{K}\bar{\psi})(\ell) - \bar{Q}(\ell)|^2 \mathrm{d}\ell \tag{7.35}$$

Denoting by $\operatorname{argmin}_{\bar{\psi}\in X} \|\mathcal{K}\bar{\psi} - \bar{Q}\|^2$ the set of solutions of (7.34), the following facts hold (see, *e.g.*, [IJ14]):

- If $\mathcal{K}$ is injective, then (7.34) has at most one solution.
- If $\bar{Q} \in \operatorname{Im}\mathcal{K} \oplus (\operatorname{Im}\mathcal{K})^{\perp}$, then the set $\operatorname{argmin}_{\bar{\psi}\in X} \|\mathcal{K}\bar{\psi} - \bar{Q}\|^2$ is closed, convex and non-empty (in particular (7.34) admits at least one solution).
- If $\mathcal{K}$ is injective and admits a left inverse denoted by $\mathcal{K}^{-1}$, then the unique solution of (7.34) is $\bar{\psi} = \mathcal{K}^{-1}\bar{Q}$.

The direct sum $\operatorname{Im}\mathcal{K} \oplus (\operatorname{Im}\mathcal{K})^{\perp}$ being dense in Y, we assume in the following that $\bar{Q}$ lies in this set. The direct approach developed in this section is justified by the following theorem.

Theorem 7.6. *The operator $\mathcal{K}$ is injective.*

Proof. Let $\psi \in L^2((r_{\min}, r_{\max}); \mathbb{R})$ such that $\mathcal{K}\psi = 0$. Then, for almost every $\ell \in (0, \ell_{\max})$, we have:

$$\begin{aligned} 0 &= \int_{r_{\min}}^{r_{\max}} k(\ell, r)\psi(r)\mathrm{d}r \\ &= \int_{r_{\min}}^{r_{\max}} \psi(r)\mathrm{d}r - \int_{r_{\min}}^{r_{\max}} \psi(r) \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \sqrt{1 - \left(\frac{\ell}{2r}\right)^2 \alpha_\eta(\phi,\theta)}\frac{\sin\theta}{4\pi}\mathrm{d}\theta\mathrm{d}\phi\mathrm{d}r. \end{aligned} \tag{7.36}$$

Let us consider the sequence $(\mathcal{K}\psi)^{(2n)}(0)$. It can be computed using differentiation of the parameter integral. The function $\ell \mapsto \sqrt{1 - \left(\frac{\ell}{2r}\right)^2 \alpha}$ is analytic at 0, and we have (from the series expansion of $\sqrt{1-\ell^2}$) that

$$\left.\frac{\mathrm{d}^{2n}}{\mathrm{d}\ell^{2n}}\sqrt{1 - \left(\frac{\ell}{2r}\right)^2 \alpha}\right|_{\ell=0} = \frac{(2n)!}{(n!)^2(1-2n)4^{2n}}\frac{\alpha^n}{r^{2n}}$$

Hence, for $n \geqslant 1$,

$$(\mathcal{K}\psi)^{(2n)}(0) = \frac{(2n)!}{(n!)^2(1-2n)4^{2n}} \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \alpha_\eta^n(\phi,\theta)\frac{\sin\theta}{4\pi}\mathrm{d}\theta\mathrm{d}\phi \int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n}}\mathrm{d}r.$$

Let us denote

$$a_n(\eta) = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \alpha_\eta^n(\phi,\theta)\frac{\sin\theta}{4\pi}\mathrm{d}\theta\mathrm{d}\phi, \qquad b_n = \frac{(2n)!}{(n!)^2(1-2n)4^{2n}}$$

so that

$$(\mathcal{K}\psi)^{(2n)}(0) = a_n(\eta)b_n \int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n}}\mathrm{d}r.$$

Since $\mathcal{K}\psi$ is constantly equal to 0, $\mathcal{K}\psi^{(2n)}(0) = 0$ for all $n \in \mathbb{N}^*$. Since $a_n(\eta)b_n > 0$ for all $\eta > 0$ and all $n \in \mathbb{N}^*$, having $(\mathcal{K}\psi)^{(2n)}(0) = 0$ for all $n \in \mathbb{N}^*$ implies that

$$\int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n}}\mathrm{d}r = 0 \qquad \forall n \in \mathbb{N}^*.$$

Let $n \in \mathbb{N}^*$. Then,

$$\begin{aligned} 0 &= \int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n}} \mathrm{d}x \\ &= \int_{\frac{1}{r_{\max}}}^{\frac{1}{r_{\min}}} \psi\left(\frac{1}{\tilde{r}}\right) \tilde{r}^{2n-2} \mathrm{d}\tilde{r}. \qquad (\tilde{r} = \frac{1}{r}) \end{aligned}$$

Set $\tilde{\psi} : [\frac{1}{r_{\max}}, \frac{1}{r_{\min}}] \ni \tilde{r} \longmapsto \psi(\frac{1}{\tilde{r}})$. Then,

$$\begin{aligned} 0 &= \int_{\frac{1}{r_{\max}}}^{\frac{1}{r_{\min}}} \tilde{\psi}(\tilde{r}) \tilde{r}^{2n-2} \mathrm{d}\tilde{r} \\ &= \frac{1}{2} \int_{\frac{1}{r_{\max}^2}}^{\frac{1}{r_{\min}^2}} \frac{\tilde{\psi}(\sqrt{\bar{r}})}{\sqrt{\bar{r}}} \bar{r}^{n-1} \mathrm{d}\bar{r}. \qquad (\bar{r} = \tilde{r}^2) \end{aligned} \tag{7.37}$$

Set $\bar{\psi} : [\frac{1}{r_{\max}^2}, \frac{1}{r_{\min}^2}] \ni \bar{r} \longmapsto \frac{\tilde{\psi}(\sqrt{\bar{r}})}{\sqrt{\bar{r}}}$. Then we have

$$0 = \int_{\frac{1}{r_{\max}^2}}^{\frac{1}{r_{\min}^2}} \bar{\psi}(x) \bar{r}^{n-1} \mathrm{d}\bar{r}. \tag{7.38}$$

Since the family $(r \mapsto r^n)_{n \geqslant 0}$ is a total family in $L^2\left(\left(\frac{1}{r_{\max}^2}, \frac{1}{r_{\min}^2}\right); \mathbb{R}\right)$ from the Weierstrass approximation theorem, $\bar{\psi} = 0$. Hence $\psi(r) = 0$ for all $r \in [r_{\min}, r_{\max}]$. ■

Therefore, the problem (7.34) admits exactly one solution. However, numerically computing this solution remains challenging, because the problem is still ill-posed. Indeed, the operator $\mathcal{K}$ is compact, as an integral operator with square-integrable kernel. Hence, its left-inverse cannot be continuous, which implies that any small measurement noise on $\bar{Q}$ leads to a major perturbation of the estimated normalized PSD $\bar{\psi}$. To tackle this issue, a typical approach is the Tikhonov regularization procedure.

Proposition 7.7 (see, *e.g.*, [IJ14])**.** *For any $\delta > 0$, the minimization problem*

$$\textit{Find } \bar{\psi} \in X \textit{ minimizing } \|\mathcal{K}\bar{\psi} - \bar{Q}\|_Y^2 + \delta \|\bar{\psi}\|_X^2. \tag{7.39}$$

admits a unique solution, which depends continuously on $\bar{Q}$.

The Tikhonov regularization consists in replacing the ill-posed problem (7.34) by the well-posed (7.39). The parameter δ is called the *regularization parameter*. Letting δ tend towards zero, we recover the original problem (7.34). As δ tends towards infinity, the solution of (7.39) tends towards zero. The choice of δ is a trade-off: the regularized problem must be sufficiently close to the original problem (δ sufficiently small) to have a similar solution, but not too close to remain robust to measurement noise (δ sufficiently large). It must be experimentally selected. One can interpret δ as a confidence measure: the more uncertain the sensor is, the larger δ should be.

This regularization procedure is justified by the following theoretical result, that describes what happens when δ goes to zero.

Theorem 7.8 ([Ker16, Theorem 6.1]). *Let $\mathcal{T}$ in $\mathscr{L}(X,Y)$ and $z \in \operatorname{Im}\mathcal{T}$. Let $\psi_0 \in X$ and ψ the solution of (7.34) closest to ψ_0. Let $(Q_n)_{n\in\mathcal{N}}$ be a sequence in Y converging to Q. Let $\varepsilon_n = \|Q_n - Q\|_Y$. Let $(\delta_n)_{n\in\mathcal{N}}$ be a sequence of regularization parameters converging to zero. For any $n \in \mathcal{N}$, let ψ_n be the solution of the problem (7.39) associated to Q_n and δ_n. Then,*

- $\|\mathcal{T}\psi_n - Q_n\|_Y \underset{n\to+\infty}{\longrightarrow} 0$;
- *if* $\dfrac{\varepsilon_n}{\delta_n} \underset{n\to+\infty}{\longrightarrow} 0$, *then* $\|\mathcal{T}\psi_n - Q_n\|_Y = \mathcal{O}(\varepsilon_n)$ *and* $\psi_n \underset{n\to+\infty}{\longrightarrow} \psi$;
- *if* $\dfrac{\varepsilon_n}{\delta_n^2} \underset{n\to+\infty}{\longrightarrow} 0$, $\psi \in (\operatorname{Im}\mathcal{T})^*$, *then* $\|\mathcal{T}\psi_n - Q_n\|_Y = \mathcal{O}(\varepsilon_n^2)$, $\|\psi_n - \psi\|_X = \mathcal{O}(\varepsilon_n)$.

Finally, since $\bar{\psi}$ is known to be a probability density function, one can constrain the minimization problem:

$$\textit{Find } \bar{\psi} \in X \textit{ minimizing } \|\mathcal{K}\bar{\psi} - \bar{Q}\|_Y^2 + \delta\|\bar{\psi}\|_X^2 \textit{ subject to } \bar{\psi} \geqslant 0. \tag{7.40}$$

Denoting by $\bar{\psi}$ the solution of this latter problem, we now aim to find the PSD ψ.

7.3.2 Estimation of the number of particles

In this section, we propose to estimate $\int_{r_{\min}}^{r_{\max}} \psi(r)\mathrm{d}r$ by using a measurement of the solid concentration C_s (in kg of solid per kg of solvent). As in Section 7.2.1, let ρ_s (in kg.m^{-3}) be the density of the solute in solid phase, M_e be the solvent mass (in kg), and V_s (in m^3) be the volume occupied by the particles in the reactor. Then $C_s = \frac{\rho_s}{M_e} V_s$ and

$$V_s = \frac{4\pi}{3}\eta \int_{r_{\min}}^{r_{\max}} \psi(r) r^3 \mathrm{d}r = \frac{4\pi}{3}\eta \int_{r_{\min}}^{r_{\max}} \bar{\psi}(r) r^3 \mathrm{d}r \int_{r_{\min}}^{r_{\max}} \psi(r)\mathrm{d}r. \tag{7.41}$$

since the volumetric shape factor k_v of a spheroid with parameter η is $k_v = \frac{4\pi}{3}\eta$. Using the estimation of $\bar{\psi}$ obtained in the previous step, we get

$$\int_{r_{\min}}^{r_{\max}} \psi(r)\mathrm{d}r = \frac{3}{4\pi\eta}\frac{M_e}{\rho_s}\frac{C_s}{\int_{r_{\min}}^{r_{\max}} \bar{\psi}(r) r^3 \mathrm{d}r}. \tag{7.42}$$

Thus, if ρ_s, M_e and η are known, and $\bar{\psi}$ is estimated in the previous step, it is possible to estimate the number of particles per unit of volume with a measurement of the solid concentration.

7.3.3 Numerical simulations

For simulations, we consider a bi-modal normalized PSD $\bar{\psi}(r)$ of spheroid particles with shape parameter $\eta = 2$ and radius r between $r_{\min} = 1.0 \times 10^{-4}$m and $r_{\max} = 3.0 \times 10^{-4}$m, attaining its maximum at $r = 1.5 \times 10^{-4}$m and $r = 2.5 \times 10^{-4}$m. More precisely, we choose

$$\bar{\psi}(r) = \frac{e^{-30(r-1.5\times10^{-4})^2} + e^{-30(r-2.5\times10^{-4})^2}}{\int_{1\times10^{-4}}^{3\times10^{-4}} e^{-30(\rho-1.5\times10^{-4})^2} + e^{-30(\rho-2.5\times10^{-4})^2}\mathrm{d}\rho}. \tag{7.43}$$

The corresponding CLD $\bar{q}$ satisfies $\bar{Q} = \mathcal{K}\bar{\psi}$, where $\bar{Q}$ is the cumulative CLD. The chord lengths ℓ lie in $[0, \ell_{\max}]$, with $\ell_{\max} = 2r_{\max}\eta = 12$mm. We add a zero mean Gaussian noise to q with variance deviation of 2% of the maximum of q. Then, we apply the Tikhonov regularization procedure to estimate ψ from the noised CLD q. Intervals $[r_{\min}, r_{\max}]$ and $[0, \ell_{\max}]$ are discretized with 200 equally spaced points. We use three different values of the regularization parameter $\delta(= 10^{-5}, 10^{-3}, 10^{-1})$. We plot the results in Figure 7.5. For all the considered values of δ, the bi-modality of the PSD is recovered by the estimation. However, when $\delta = 10^{-5}$, the regularization parameter is too small. The discontinuity issues of the non-regularized problem (7.33) still appear. On the contrary, $\delta = 10^{-1}$ is too large. The regularized problem is too far from the original minimization problem and some information on the amplitude of the PSD is lost. With $\delta = 10^{-3}$, we recover a satisfying estimation of the original PSD by balancing these two effects.

7.3.4 Conclusion

In this section, we have shown that a direct approach allows to reconstruct the unknown PSD from its CLD when crystals are spheroids sharing the same shape factor η. The method relies on a regularization method to inverse the PSD-to-CLD relation, and can be used online or offline. However, this strategy has several drawbacks:

- If the only accessible measures are the temperature and the solute concentration, the method does not apply. In Section 7.4.1, we will show that an observer using these measures (and not the CLD) may be designed and is able to reconstruct partially the PSD on numerical simulations.

- It does not use the dynamical model of the batch crystallization process introduced in Section 7.1. Several improvements can be made to take into account this additional knowledge. First, the resolution of the regularized problem (7.40) is usually done with iterative algorithms, such as interior-point methods (see [BV04, Chapter 11]). Hence, when solving online the minimization problem, an important improvement in the computational cost can be made by choosing as the initial guess a shifted version of the previous estimation (because the actual solution satisfies a transport equation). Second, if the method is applied offline (after measuring the CLD over a finite time-interval), the minimization problem (7.40) can be reformulated to take into account the fact that the solution satisfies a transport equation:

$$\begin{aligned} &\textit{Find } \bar{\psi}_0 \in X, u \in L^2(0, t_{\max}) \textit{ minimizing} \\ &\int_0^{t_{\max}} \|\mathcal{K}\bar{\psi}(t) - \bar{Q}(t)\|_Y^2 + \delta\|\bar{\psi}(t)\|_X^2 \mathrm{d}t \textit{ subject to } \bar{\psi}(t) \geqslant 0, \end{aligned}$$

  where $\psi(t)$ is the solution of (7.6) associated to ψ_0 and u and $Q(t)$ is the corresponding cumulative CLD.

- The method is unable to deal with crystals having several shape factors or separated in different clusters (multi-shape case). Indeed, while the single-shape operator $\mathcal{K}$ is injective (see Theorem 7.6), the corresponding multi-shape

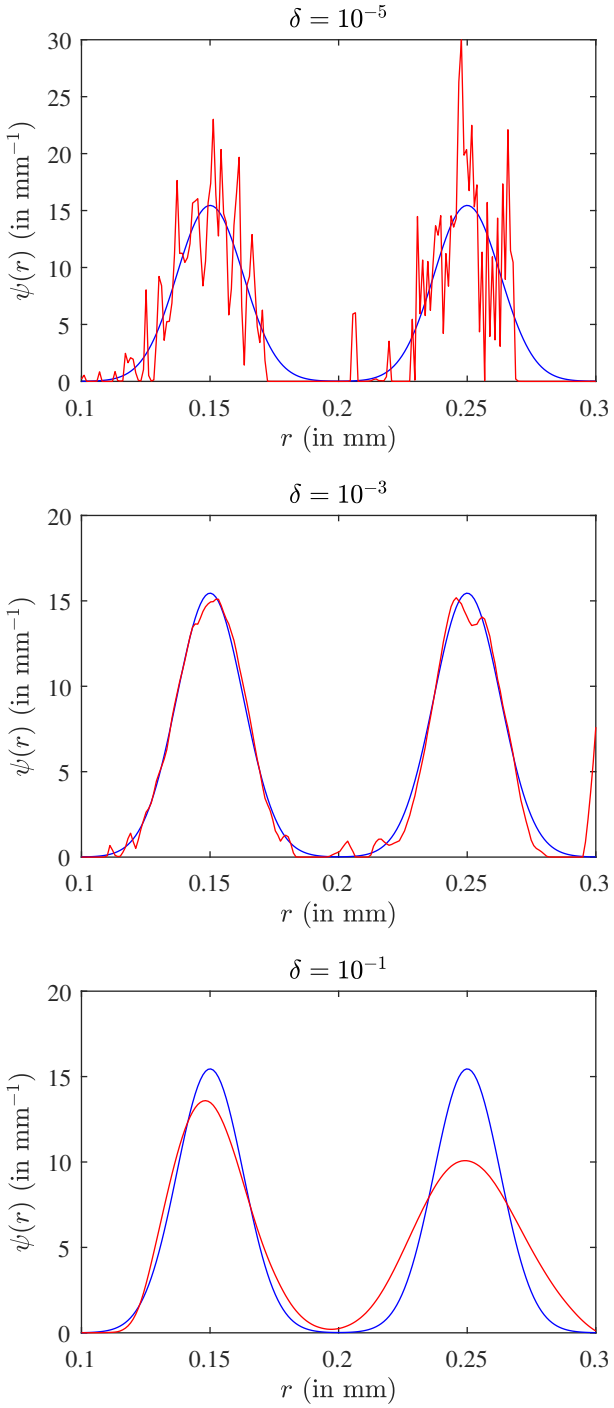


Figure 7.5 – Estimation of the PSD by the Tikhonov regularization method. In blue: the PSD given by (7.43). In red: the PSD estimated by the Tikhonov regularization method for $\delta = 10^{-5}, 10^{-3}, 10^{-1}$.

operator (see Remark 7.5), still denoted by $\mathcal{K}$, defined by

$$\begin{aligned} \mathcal{K}: \quad & X^N \longrightarrow Y \\ & (\bar{\psi}_i)_{1\leqslant i\leqslant N} \longmapsto \left(\ell \mapsto \sum_{i=1}^N \int_{r_{\min}}^{r_{\max}} k_i(\ell, r)\bar{\psi}_i(r)\mathrm{d}r\right) \end{aligned}$$

may not be injective. Indeed, the different PSDs are intertwined in the CLD.

In particular, in the case where $\eta_i = \eta_j$ for some $i \neq j$, there is no way to differentiate the part of the CLD due to ψ_i and the part due to ψ_j. Therefore, applying the Tikhonov regularization procedure in this case is not a convenient approach. In Section 7.4.2, we design an observer able to estimate each PSD from the common CLD.

7.4 Observer approach

7.4.1 KKL observer with measured solute concentration

In this section, we consider that crystals are all spheroids of the same shape (single-shape case) and we have access to the measurements of the temperature and solute concentration (but not CLD). According to Section 7.2.1, these measurements allow to access to the growth rate and the third moment of the PSD. The observation problem we intend to solve is the following.

Problem 7.9. From the knowledge of the output function y given by (7.13) and the growth rate G, give an online estimation of the PSD ψ solution of (7.6).

Observability analysis

In this section, we study how the third moment y may help us to estimate the PSD. First, we have the following result.

Proposition 7.10. *Let $\tau \in (0, t_{\max}]$. Assume that there exists $\mu > 0$ such that $G(t) \geqslant \mu$ for all $t \in [0, t_{\max}]$. Then for all $y \in C^0(0,\tau)$, there exists at most one function $u \in H^4(0,\tau)$ such that the solution ψ of (7.6) given by u and $\psi_0 = 0$ satisfies*

$$y(t) = \int_{r_{\min}}^{r_{\max}} \psi(t,r) r^3 \mathrm{d}r, \qquad \forall t \in [0,\tau].$$

In other words, Proposition 7.10 states that the map $u \mapsto y$ is injective, where y denotes the third moment of the solution of (7.6) with null initial condition. Its proof is given below. Hence, one can hope that our method may reconstruct ψ from y, at least when the initial condition is zero (*i.e.* there is no crystals at the beginning of the process).

However, one can wonder what happens if the initial condition is not zero. Can we still reconstruct the state from the measurement of its third moment and the knowledge of its dynamics? In other words, is the map $\psi \mapsto y$ injective? If yes, then one can hope that our algorithm is robust, so that the estimation of the state converges to the actual PSD. Unfortunately, the answer is no. Indeed, we have the following proposition which is a slight modification of [Vis12, Theorem 3.2.3] that we state in our own context only. Its proof is given in Appendix B.1.

Proposition 7.11. *Let $\tau \in (0, t_{\max}]$. Assume that there exists $\mu > 0$ such that $G(t) \geqslant \mu$ for all $t \in [0, t_{\max}]$. There exist infinitely many solutions of (7.6) with different initial conditions and boundary conditions that have the same third moment $y \in C^0(0,\tau)$.*

We shall say that system (7.6) with measurement (7.13) is not observable. Thus, we cannot guarantee that our estimation of the PSD converges to the actual PSD. Despite this fact, our methodology should be able to reconstruct partially the actual PSD. Indeed, the linear function that maps the PSD to its third moment has rank 4. The image of this state-output mapping is the observable part of the system (see Section 5.3).

Kazantzis-Kravaris/Luenberger observers

In the seminal paper [Lue64], D. Luenberger proposed an original observer design, based on a two steps procedure. The approach is somehow different from the way finite-dimensional Luenberger observers are introduced nowadays (which follows D. Luenberger second paper on observers [Lue71]). More recently, this two-steps strategy has been employed in [KK98] to design local observers for finite-dimensional nonlinear dynamics, leading to the so-called Kazantzis-Kravaris/Luenberger (KKL) observers. The two steps suggested in [Lue64] are as follows: *(i)* estimate a function of the state; *(ii)* invert the function of the previous step. In the context of nonlinear systems, this path has been explored by the authors of [AP06, Ber+17, BA19]. The results have been extended to discrete-time systems in [BAS19]. A reproduction of this last article is presented in Appendix C, in order to recall the results existing for finite-dimensional systems. In the present section, we aim to apply this strategy to Problem 7.15, hence to extend the two-steps strategy to the infinite-dimensional context. Note that, due to the lack of observability of the system (see Proposition 7.11), no proof of convergence of the observer will be obtained. However, as illustrated later on numerical simulations, the observer seems to reconstruct at least partially the state. The use of KKL observers for infinite-dimensional systems is an active research area, at the heart of the project ANR ODISSE (ANR-19-CE48-0004-01).

Step 1: reconstruction of a function of the state. In [AP06] and [BA19], the authors show that it is always possible to exponentially estimate a function of the state of a nonlinear dynamical system that will carry enough information about the state to estimate it completely in Step 2. In order to do so in the infinite-dimensional context, we introduce an auxiliary dynamical system fed by the measured output such that its solutions provide an estimation of this function of the state.

Let us first consider an abstract context in which the state space X is a Hilbert space and the Cauchy problem is

$$\dot{\psi} = A\psi, \quad \psi(0) = \psi_0, \tag{7.44}$$

where $A : \mathcal{D} \subset X \to X$ is a linear operator which is the generator of a strongly continuous semigroup (see Section 5.1.1) denoted by $(\mathbb{T}(t))_{t\in\mathbb{R}_+}$ in $\mathscr{L}(X)$ and $\psi_0 \in \mathcal{D}$. Let $\rho(A) = \{\lambda \in \mathbb{C} \ : \ (A - \lambda \mathrm{Id})^{-1} \in \mathscr{L}(X)\}$ denote the resolvent set of A. Moreover, consider a bounded output operator

$$y = C\psi, \tag{7.45}$$

where $C \in \mathscr{L}(X, \mathbb{R})$ is bounded linear form. Following the KKL methodology, we obtain the following proposition.

Proposition 7.12. *For all λ in $\rho(A)\cap\mathbb{R}^*_-$, let $\mathcal{T}_\lambda$ in $\mathscr{L}(X,\mathbb{R})$ be the operator defined as*

$$\mathcal{T}_\lambda : X \ni \psi \mapsto C(A-\lambda\mathrm{Id})^{-1}\psi \in \mathbb{R}\ .$$

Then, the dynamical system

$$\dot{z}_\lambda = \lambda z_\lambda + y, \tag{7.46}$$

is an exponential observer for $\mathcal{T}_\lambda\psi$. More precisely, for all (ψ_0, z_0) in $\mathcal{D}\times\mathbb{R}$, it yields for all $t \geqslant 0$

$$\mathcal{T}_\lambda(\mathbb{T}(t)\psi_0) - z_\lambda(t) = \exp(\lambda t)\left(\mathcal{T}_\lambda(\psi_0) - z_0\right). \tag{7.47}$$

where $z_\lambda : \mathbb{R}_+ \to \mathbb{R}$ is the solution of system (7.46) when y is given by (7.45) and initiated from z_0.

Proof. Let ψ_0 be in $\mathcal{D}$. Equations (7.44)–(7.46) yield

$$\begin{aligned}\frac{d}{dt}\Big(\mathcal{T}_\lambda(\mathbb{T}(t)\psi_0) - z_\lambda(t)\Big) &= \mathcal{T}_\lambda(A\mathbb{T}(t)\psi_0) - \lambda z_\lambda(t) - C\mathbb{T}(t)\psi_0\\ &= \mathcal{T}_\lambda(A-\lambda\mathrm{Id})\mathbb{T}(t)\psi_0 + \lambda\Big(\mathcal{T}_\lambda(\mathbb{T}(t)\psi_0) - z_\lambda(t)\Big) - C\mathbb{T}(t)\psi_0\\ &= \lambda(\mathcal{T}_\lambda(\mathbb{T}(t)\psi_0) - z_\lambda(t)),\end{aligned}$$

where the last equality follows since $\mathcal{T}_\lambda(A-\lambda\mathrm{Id}) = C$. Hence, (7.47) follows by integrating in time the former equation. Keeping in mind that λ is negative in Proposition 7.12, (7.47) implies

$$\lim_{t\to+\infty} |\mathcal{T}_\lambda(\mathbb{T}(t)\psi_0) - z_\lambda(t)| = 0. \tag{7.48}$$

This ends the proof. ■

Remark 7.13. The operator $\mathcal{T}_\lambda$ is solution to the Sylvester equation :

$$A\mathcal{T}_\lambda = \lambda\mathcal{T}_\lambda + C. \tag{7.49}$$

We recognize here the algebraic equation which was already given in Luenberger seminal paper [Lue64] and which becomes a nonlinear partial differential equation in [AP06].

Step 2: reconstruction of the entire state of the system. According to step 1, we can easily estimate $\mathcal{T}_\lambda\psi$ for all λ in $\rho(A)\cap\mathbb{R}^*_-$ via the observer system (7.46). The idea of the KKL observer methodology is to consider the mapping $\mathcal{T} : X \mapsto \mathbb{R}^p$ given by $\psi \mapsto (\mathcal{T}_{\lambda_1}\psi, \ldots, \mathcal{T}_{\lambda_p}\psi)$ which will be exponentially estimated along the trajectory of (7.44) via a bench of observers of the form (7.46). To solve the estimation problem, the question is to solve the inverse problem

$$\mathcal{T}\hat{\psi} = z \tag{7.50}$$

with the unknown $\hat{\psi}$ in X. To do so, we apply the Tikhonov regularization procedure introduced in Section 7.3.1. Hence, the observer $\hat{\psi}$ is obtained as the solution of the minimization problem

$$\textit{Find } \hat{\psi}(t) \in X \textit{ minimizing } \|\mathcal{T}\hat{\psi}(t) - z(t)\|_Y^2 + \delta\|\hat{\psi}(t)\|_X^2 \textit{ subject to } \hat{\psi} \geqslant 0. \tag{7.51}$$

for some $\delta > 0$, where $z(t)$ is given by (7.46).

Application to the batch crystallization process.

To apply the KKL strategy to the process under consideration, we consider λ a negative real number and the dynamical system

$$\dot{z} = \lambda z + y \ . \tag{7.52}$$

We must find a mapping $\mathcal{T}_\lambda$ which is estimated by this dynamical equation. Let $X = L^2(r_{\min}, r_{\max})$ We have the following proposition.

Proposition 7.14. *Let $\mathcal{T}_\lambda : C^1([0,t_{\max}];X) \mapsto C^1([0,t_{\max}];\mathbb{R})$ be the operator defined by*

$$\mathcal{T}_\lambda(\psi) : t \mapsto \int_{r_{\min}}^{r_{\max}} a(t,r)\psi(t,r)\mathrm{d}r \tag{7.53}$$

where a is the unique solution of

$$\begin{cases} \frac{\partial a}{\partial t}(t,r) + G(t)\frac{\partial a}{\partial r}(t,r) = \lambda a(t,r) + r^3 & \forall t \in (0,t_{\max}),\ \forall x \in (r_{\min}, r_{\max}) \\ a(0,r) = 0 & \forall r \in [r_{\min}, r_{\max}] \\ a(t,r_{\min}) = 0 & \forall t \in [0,t_{\max}]. \end{cases} \tag{7.54}$$

Then, if ψ is a solution of (7.6) satisfying (7.1), z is a solution of (7.52) and y is given by (7.13), we have for all $t \in [0,t_{\max}]$:

$$\mathcal{T}_\lambda(\psi)(t) - z(t) = \exp(\lambda t)(\mathcal{T}_\lambda(\psi)(0) - z_0). \tag{7.55}$$

Proof. Using (7.6) and an integration by parts yields

$$\begin{aligned} \frac{d}{dt}(\mathcal{T}_\lambda(\psi)(t) - z(t)) &= \int_{r_{\min}}^{r_{\max}} \partial_t a(t,x)\psi(t,r)\mathrm{d}r - \int_{r_{\min}}^{r_{\max}} G(t)a(t,r)\partial_x\psi(t,r)\mathrm{d}r \\ &\quad - \lambda z(t) - \int_{r_{\min}}^{r_{\max}} r^3\psi(t,r)\mathrm{d}r \\ &= \int_{r_{\min}}^{r_{\max}} \partial_t a(t,x)\psi(t,r)\mathrm{d}r + \int_{r_{\min}}^{r_{\max}} G(t)\partial_x a(t,r)\psi(t,r)\mathrm{d}r \\ &\quad - G(t)\Big[a(t,r)\psi(t,r)\Big]_{r_{\min}}^{r_{\max}} - \lambda z(t) - \int_{r_{\min}}^{r_{\max}} x^3\psi(t,r)\mathrm{d}r. \end{aligned}$$

Hence, with (7.54) and also the boundary condition in (7.6) and (7.1), this implies

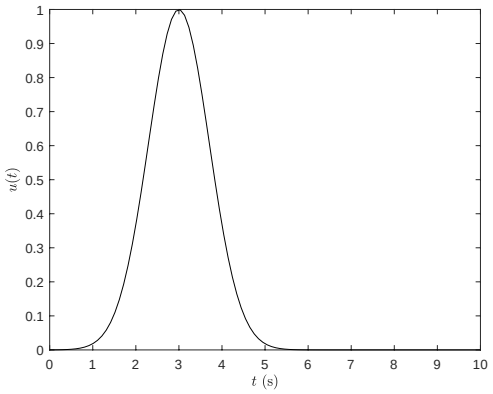
$$\frac{d}{dt}(\mathcal{T}_\lambda(\psi)(t) - z(t)) = \lambda(\mathcal{T}_\lambda(\psi)(t) - z(t)).$$

By integrating in time the former equation, we obtain (7.55). ■

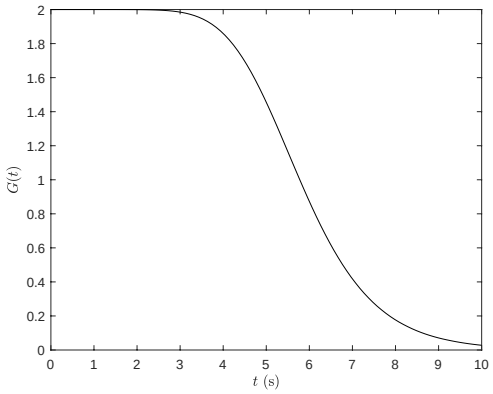
Consequently, for each $\lambda < 0$ we exponentially estimate the functional $\mathcal{T}_\lambda\psi(t)$. It is interesting to remark that no information on the nucleation rate is needed to obtain this estimation.

At any fixed time t, the operator $X \ni \psi \mapsto \int_{r_{\min}}^{r_{\max}} a(t,r)\psi(r)\mathrm{d}r \in \mathbb{R}$ is an integral operator, hence is not continuously invertible, which justifies the use of a Tikhonov regularization procedure as in Section 7.3.1. To summarize, the observer $\hat{\psi}$ is then given by

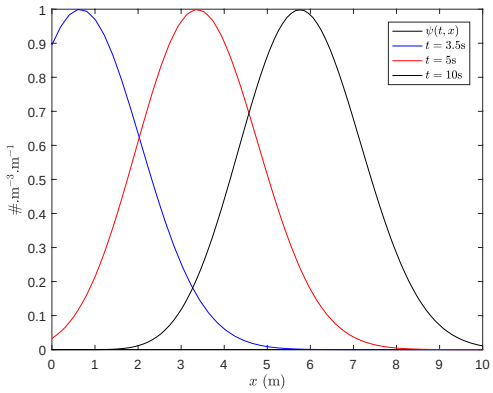
$$\begin{cases} \dot{z} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix} z + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} y \\ \hat{\psi}(t) = \mathrm{argmin}_{\tilde{\psi}\in X}\left\{\|\mathcal{T}(\tilde{\psi})(t) - z(t)\|^2 + \delta\|\tilde{\psi}\|^2\right\}, \quad \delta > 0 \\ \mathcal{T} = \left(\mathcal{T}_{\lambda_1}, \ldots, \mathcal{T}_{\lambda_p}\right) \end{cases} \tag{7.56}$$



(a) Boundary condition u due to the nucleation.



(b) Growth rate G, given by (7.11).



(c) Simulation of the PSD ψ.

Figure 7.6 – Numerical simulation of the batch crystallization process with $(r_{\min}, r_{\max}) = (0, 10)$, $t_{\max} = 10$ and $N_x = N_t = 100$.

Numerical simulations

In this section numerical simulations are carried out. Let $(\mathbb{r}_j)_{1\leqslant j\leqslant N_x}$ be a uniform discretization of the space interval $(r_{\min}, r_{\max})$ with space step Δx and $(\mathbb{t}_k)_{1\leqslant k\leqslant N_t}$ be a uniform discretization of the time interval $(0, t_{\max})$ with time step Δt. We fix $N_x = N_t = 100$. Let $(\lambda_j)_{1\leqslant i\leqslant p}$ be the considered negative values of λ. An approximation of $(\mathcal{T}_{\lambda_i}\psi)(\mathbb{t}_k)$ is given by $\Delta x \sum_{j=1}^{N_x} a_{i,j,k}\psi_{j,k}$ where $a_{i,j,k}$ is an approximation of $a_{\lambda_i}(\mathbb{t}_k, \mathbb{r}_j)$ (solution of (7.54)) and $\psi_{j,k}$ an approximation of $\psi(\mathbb{t}_k, \mathbb{r}_j)$. The transport equation which describes the crystallization process is simulated via the method of characteristics.

We consider system (7.6) with G as in (7.11) with a null initial condition $z_0 = 0$ and a boundary condition similar to a truncated normal distribution reaching its maximum at $t = 3$ s and with a compact support $[0, 6]$ (see Figure 7.6a). The unique solution of this system is drawn in Figure 7.6c (solid line), and the corresponding growth rate is drawn in Figure 7.6b.

Step 1: reconstruction of a function of the state. Following the methodology developed in Section 7.4.1, we first try to estimate the function $\mathcal{T}_\lambda(z)$ of the state via the dynamical system (7.52) for some fixed negative values of λ. All along

the simulation of (7.6), we compute y and estimate the solution of (7.54) via the method of characteristics. We integrate the solution of (7.52) with the first order Euler's method. Then we plot the evolution of the relative error between z and $\mathcal{T}_\lambda(z)$ in Figure 7.7 for some values of λ. One can check that the error goes to zero as $t \to +\infty$. Moreover, the bigger is $|\lambda|$, the faster is the convergence. This is due to the exponential convergence of $z - \mathcal{T}_\lambda(z)$ to zero given by (7.47). Hence, we are able to approximate any function $\mathcal{T}_\lambda(z)$ of the state. Now, we can move to the second part of the methodology of Section 7.4.1.

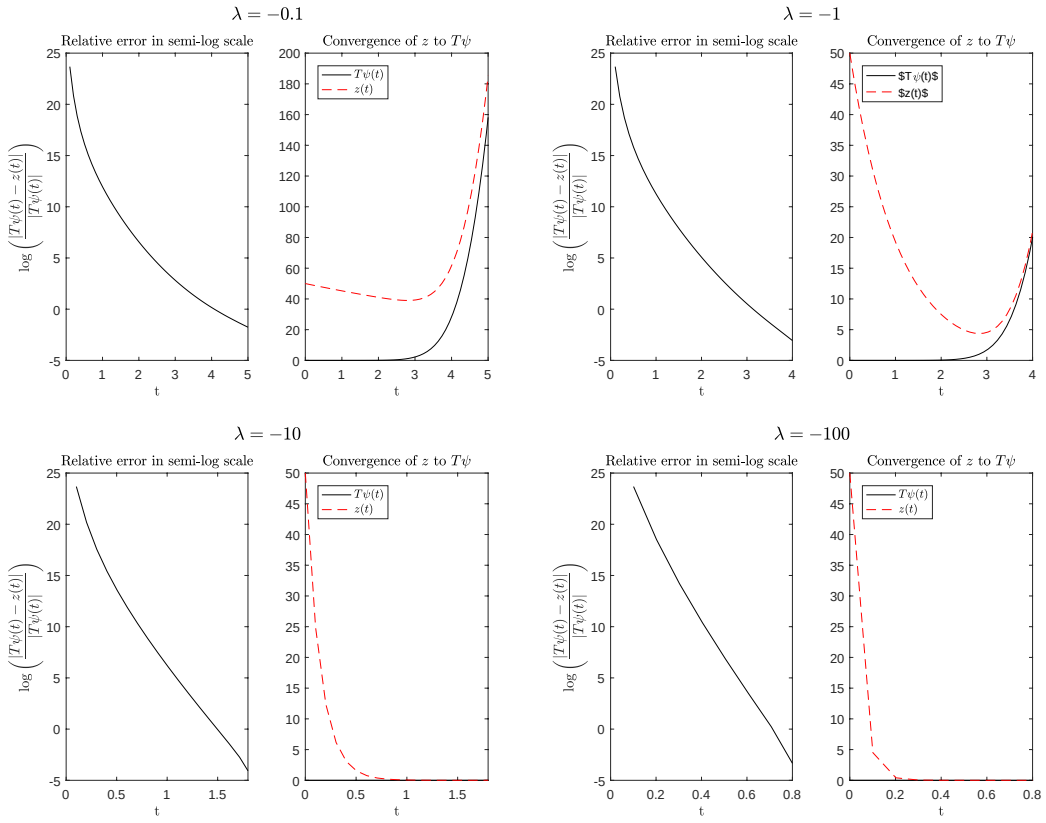


Figure 7.7 – Convergence of $\mathcal{T}_\lambda(z) - z$ to zero for different values of λ. We choose $z(0) \neq 0$ arbitrarily. The bigger is $|\lambda|$, the faster is the convergence. By means of a linear regression, one can estimate the convergence rate of the relative error to zero: $\mathcal{O}(e^{-7.4t})$ if $\lambda = -0,1$, $\mathcal{O}(e^{-8.2t})$ if $\lambda = -1$, $\mathcal{O}(e^{-14.2t})$ if $\lambda = -10$, $\mathcal{O}(e^{-32.9t})$ if $\lambda = -100$.

Step 2: reconstruction of the entire state of the system Following Step 1, we estimate simultaneously numerous functions $\mathcal{T}_{\lambda_i}(z)$ which correspond to different values $\lambda_i < 0$. These estimations are denoted z_{λ_i}. The aim of this section is to estimate the state ψ from the knowledge of $(z_{\lambda_i})_{1\leqslant i\leqslant p}$. Then, we choose a regularization parameter $\delta > 0$ and solve the discrete version of the quadratic minimization problem (7.51) at each time step, that is for each time $\mathbb{t}_k$, find $\left(\psi_{j,k}\right)_{1\leqslant j\leqslant N_x}$ minimizing

$$\left\|\Delta x\left(a_{i,j,k}\right)_{1\leqslant i\leqslant p,1\leqslant j\leqslant N_x} \cdot \left(\psi_{j,k}\right)_{1\leqslant j\leqslant N_x} - \left(z_{\lambda_i}(\mathbb{t}_k)\right)_{1\leqslant i\leqslant p}\right\|^2 + \delta \left\|\left(\psi_{j,k}\right)_{1\leqslant j\leqslant N_x}\right\|^2. \tag{7.57}$$

This is a quadratic minimization problem, which we solve via an interior-point method (see *e.g.* [BV04, Chapter III.11]. We need to fix an initial condition to apply this algorithm. Following a continuation method, we choose as an initial condition at time $\mathbb{t}_k$ the minimum value obtained at time $\mathbb{t}_{k-1}$, transported during a time Δt at speed $G(\mathbb{t}_{k-1})$. The choice of parameters p, $\lambda_1, \dots, \lambda_p$ and δ and their influence are investigated in the paragraphs below.

- **Choice of p and $(\lambda_i)_{1\leqslant i\leqslant p}$.**

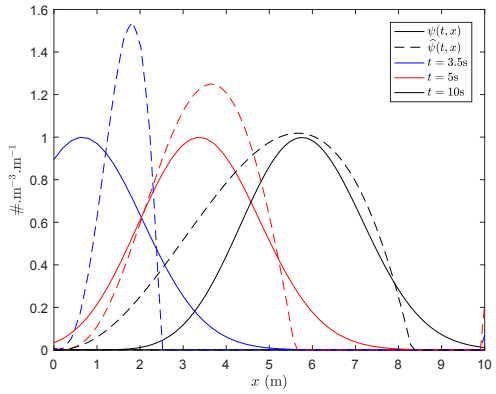
  Note that the matrix $\left(a_{i,j,k}\right)_{1\leqslant i\leqslant p, 1\leqslant j\leqslant N_x}$ may be injective only if $p \geqslant N_x$, that is if the discretization in λ is thinner than in r. Therefore, we fix $p = 2N_x = 200$. Moreover, even if the matrix $(a_{i,j,k})_{i,j}$ is injective, a regularization method is needed to left-inverse it. Indeed, for all $t \in (t_0, t_1)$, the operator

  $$L^2(r_{\min}, r_{\max}) \ni \psi \mapsto \left(\lambda \mapsto \int_{r_{\min}}^{r_{\max}} a_\lambda(t, r)\psi(r)\mathrm{d}r\right) \in L^2(\lambda_{\min}, \lambda_{\max})$$

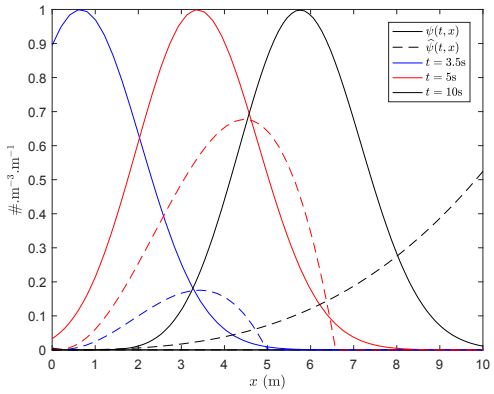
  is compact (as an integral operator). Hence, even if it is injective, its inverse is not continuous. The matrix $(a_{i,j,k})_{i,j}$ is a discretization of this operator. Then, the more the discretization is thinner, the more it is ill-conditioned. This emphasizes the necessity of using a regularization method. In Figure 7.8, we plot the estimation of the PSD for different values of (λ_i). For large values of $|\lambda|$, z converges quickly to $\mathcal{T}\psi$. However, it appears that functions a_λ carry less information for large values of $|\lambda|$, so that the map $\mathcal{T}$ is more difficult to inverse. This explains Figure 7.8b, on which the estimation $\hat{\psi}$ is worst than on Figure 7.8c. On the contrary, for small values of $|\lambda|$, it seems that functions a_λ carry more information, since the estimation $\hat{\psi}$ is similar on Figure 7.8a and Figure 7.8c at $t = 10$ s. However, we also see a peaking phenomenon (for $t \leqslant 5$ s on Figure 7.8a), due to the fact that z is slower to converge to $\mathcal{T}\psi$ than for large values of $|\lambda|$. Thus, one must find a compromise for the choice of (λ_i): take large values for fast convergence and avoiding peaking, and small values for efficient estimation.

- **Choice of the regularization parameter δ.**

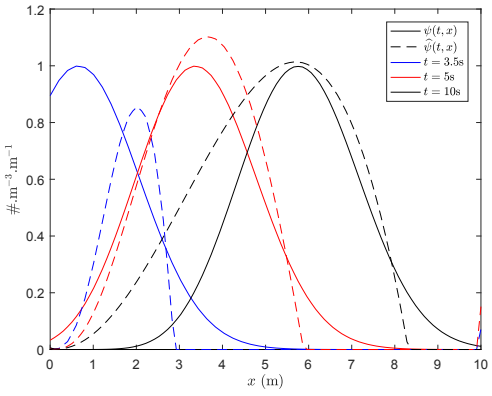
  The regularization parameter δ must be chosen numerically, in order to find a compromise between the minimization of the norm of the state, and the minimization of the gap $\mathcal{T}\psi - z$. This compromise can be interpreted as a measurement reliability. Indeed, if the measurement has a small uncertainty, then we choose a small δ. On the contrary, if the measurement is highly uncertain, then we fix a large value of δ in order to regularize the solution. In Figure 7.9, we plot the actual PSD z and its estimation $\hat{z}$ at different times, for different values of δ, and with or without measurement noise. Measurement noise is fixed at 2% of the maximal value of the output on the time interval. For small values of δ and/or with measurement noise, we see that a peaking phenomenon appear: this is due to a lack of regularization of the solution. On the contrary, if δ is too large, then the minimization of the norm of the state takes too much importance in the minimization problem, and $\hat{\psi}$ is too attenuated.



(a) $(\lambda_i) \subset [-10, -1]$



(b) $(\lambda_i) \subset [-100, -10]$



(c) $(\lambda_i) \subset [-100, -1]$

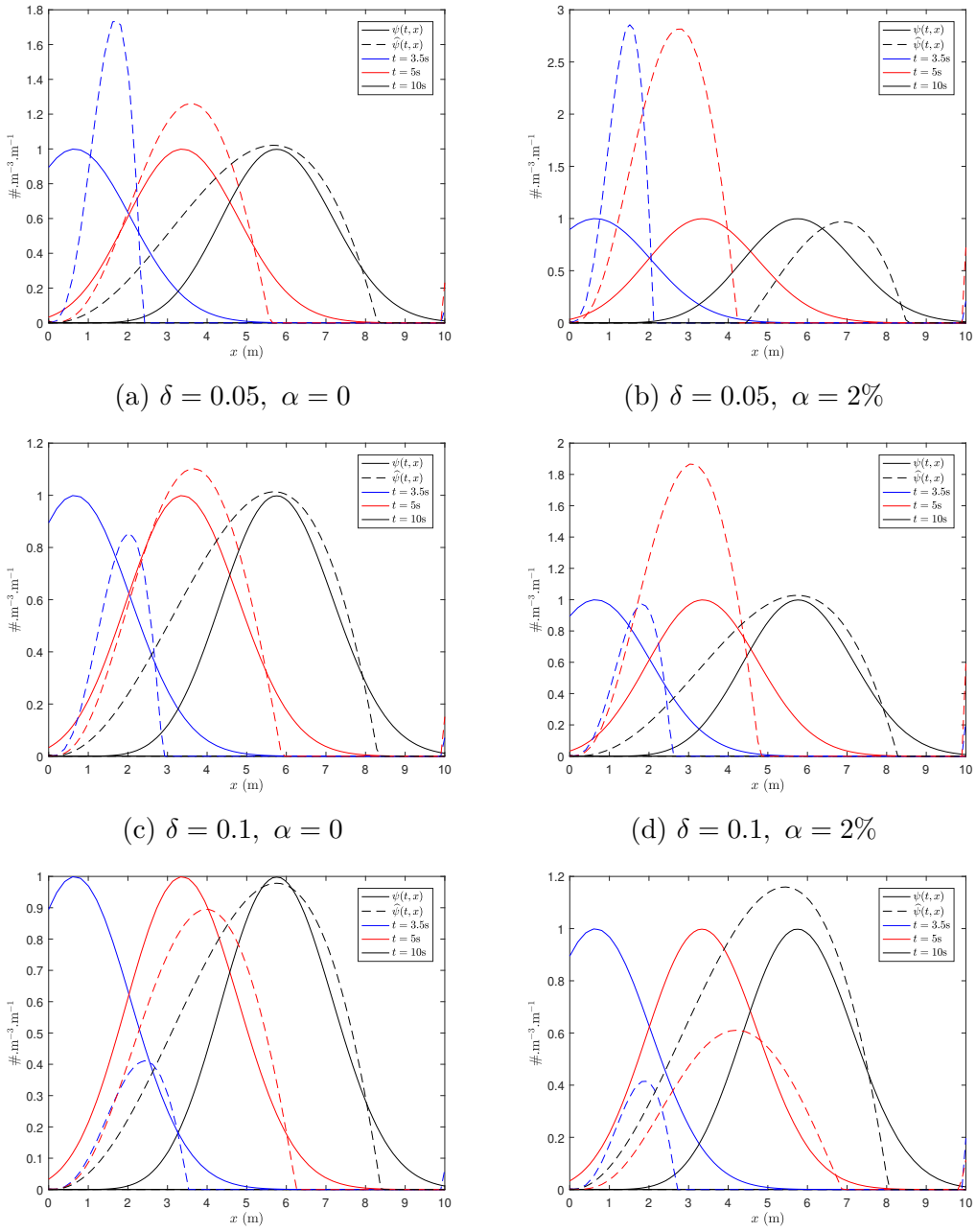
Figure 7.8 – Influence of $(\lambda_i)_{1 \leqslant i \leqslant p}$ on the reconstruction of the PSD.

Conclusion

The use of the KKL methodology for infinite-dimensional systems is promising, and the numerical results suggest that the knowledge of the temperature solute concentration allow to obtain a suitable online approximation of the PSD for well-chosen parameters (see, *e.g.,* Figure 7.8c). However, the system lacks of observability (see Proposition 7.11), and new measures need to be considered, such as the CLD. In particular, the multi-shape case must be investigated.

7.4.2 Luenberger observer with measured CLD

In this section, we consider that crystals may have several shapes during the process (see Section 7.1.3), and we have access to the measurement of the CLD over a finite time interval. The growth rate of each shape is supposed to be known, and we try to estimate the PSD associated to each shape, up to a constant multiplicative factor. In the single-shape case, this problem can be solved by the direct approach with a Tikhonov regularization procedure (see Section 7.3.1). However, this approach does not use the dynamical system (7.6) and cannot be applied in the multi-shape case (see Section 7.3.4). The estimation problem of this section can be reformulated in the following manner.



(a) $\delta = 0.05,\ \alpha = 0$ (b) $\delta = 0.05,\ \alpha = 2\%$

(c) $\delta = 0.1,\ \alpha = 0$ (d) $\delta = 0.1,\ \alpha = 2\%$

(e) $\delta = 0.2,\ \alpha = 0$ (f) $\delta = 0.2,\ \alpha = 2\%$

Figure 7.9 – Influence of the regularization parameter δ and measurement noise α on the reconstruction of the PSD

Problem 7.15. From the knowledge of the CLD over a finite time interval and the growth rate associated to each shape, give an estimation of the corresponding PSD solution of (7.9).

Modeling the batch crystallization process as a one-dimensional time-varying transport equation with periodic boundary conditions

Let $\psi_{0,i} \in L^2((r_{\min}, r_{\max}); \mathbb{R})$ and $u_i \in L^2((0, t_{\max}); \mathbb{R})$ be the initial condition and boundary condition of (7.9). Since u_i is not supposed to be measured, it is part of

the unknown data to be reconstructed. Define $\psi_i(t,r)$ for $r_{\min} - \int_t^{t+t_{\max}} G_i(s)\mathrm{d}s \leqslant r \leqslant r_{\min}$ in the following manner:

$$\psi_i(t,r) = u_i(t+\tau) \text{ with } \tau \geqslant 0 \text{ such that } \int_t^{t+\tau} G_i(s)\mathrm{d}s = r_{\min} - r. \tag{7.58}$$

Roughly speaking, $\psi_i(t,r)$ for $r < r_{\min}$ represents crystals that did not yet appear at time t, but will appear later at some time $t+\tau$. If $t+\tau > t_{\max}$, set $\psi_i(t,r) = 0$. Combining all the PSDs in a unique vector $\psi = (\psi_i)_{1\leqslant i\leqslant N}$, $G(t) = \mathrm{diag}((G_i(t))_{1\leqslant i\leqslant N})$ and $\psi_0(r) = (\psi_{0,i}(r))_{1\leqslant i\leqslant N}$, system (7.9) can be rewritten as

$$\begin{cases} \frac{\partial\psi}{\partial t}(t,r) + G(t)\frac{\partial\psi}{\partial r}(t,r) = 0 & \forall t \in (0,t_{\max}), \forall r \in (r_0,r_1) \\ \psi(0,r) = \psi_0(r) & \forall r \in [r_0,r_1] \end{cases} \tag{7.59}$$

where $r_0 = r_{\min} - \max_{1\leqslant i\leqslant N} \int_0^{t_{\max}} G_i(s)\mathrm{d}s$ and $r_1 = r_{\max}$ and with periodic boundary conditions $\psi(t,r_{\min}) = \psi(t,r_{\max})$ (since the right boundary term does not influence $\psi(t,r)$ for $r > r_{\min}$ and $t \leqslant t_{\max}$). Then, any solution ψ of (7.59) is such that $\psi(t,r)$ is the corresponding solution of (7.9) when restricted to $t \in [0,t_{\max}]$ and $r \in [r_{\min}, r_{\max}]$.

Proposition 7.16 (Well-posedness)**.** *If G_i is positive and continuously differentiable, $\psi_{0,i} \in L^2((r_{\min},r_{\max});\mathbb{R})$ and $u_i \in L^2((0,t_{\max});\mathbb{R})$ for all $i \in \{1,\dots,N\}$, then system* (7.59) *admits a unique solution $\psi \in C^0((0,t_{\max}); L^2((r_0,r_1);\mathbb{R})^N)$.*

Proof. The proof relies on the theory of linear evolution systems (see *e.g.* [Paz83]). Let $X = L^2((r_0,r_1);\mathbb{R})$ and $\mathcal{D} = \{\psi \in X \ : \ \psi' \in X, \psi(r_0) = \psi(r_1)\}$. The operator $-G(t)\frac{\partial}{\partial r} : \mathcal{D}^N \to X^N$ is linear, unbounded, and skew-adjoint for all $t \in [0,t_{\max}]$. Since G is C^1, $t \mapsto -G(t)\frac{\partial\psi}{\partial r}$ is continuously differentiable for all $\psi \in \mathcal{D}^N$. Hence, according to Theorem 6.6, it is the generator of a bidirectional evolution system on X^N. In particular, (7.59) admits a unique solution $\psi \in C^0((0,t_{\max}); X^N)$ for each $\psi_0 \in X^N$. ∎

Application of Chapters 5 and 6 to the batch crystallization process

Now, the crystallization process has been reformulated as a one-dimensional time-varying transport equation with periodic boundary conditions. Hence, we can apply the results obtained in Chapter 6, and more precisely in Section 6.4. Recall that in the multi-shape case, the measured cumulative CLD Q is given by (7.31). Abusing notations, let us replace $\kappa_i\psi_i$ by ψ_i, which satisfies the same PDE (7.9). Suppose that κ_i is independent of time for all i, that is, the ratio between the number of particles and the number of chords seen by the sensor is constant. Our goal is to estimate $\psi(t,r) = (\psi_i(t,r))_{1\leqslant i\leqslant N}$ from the knowledge of the cumulative CLD $Q(t,\ell)$ over the time interval $[0,t_{\max}]$, given by

$$Q(t,\ell) = \sum_{i=1}^{N} \int_{r_{\min}}^{r_{\max}} k_i(\ell,r)\psi_i(t,r)\mathrm{d}r. \tag{7.60}$$

Let $X = L^2((r_0,r_1);\mathbb{R})$, $\ell_{\max} = 2r_{\max}\max_{1\leqslant i\leqslant N}(\eta_i)$ and $Y = L^2((0,\ell_{\max});\mathbb{R})$, so that $Q(t,\cdot) \in Y$ for all $t \in [0,t_{\max}]$. Define the operator

$$\begin{aligned} \mathcal{K} : X^N &\longrightarrow Y \\ \psi &\longmapsto \left(\ell \mapsto \textstyle\sum_{i=1}^N \int_{r_{\min}}^{r_{\max}} k_i(\ell,r)\psi_i(r)\mathrm{d}r\right). \end{aligned}$$

Its adjoint operator is

$$\begin{aligned} \mathcal{K}^* : Y &\longrightarrow X^N \\ Q &\longmapsto \left(r \mapsto \int_0^{\ell_{\max}} k_i(\ell, r) Q(\ell) \mathrm{d}\ell\right)_{1 \leqslant i \leqslant N} \end{aligned}$$

with $k_i(\ell, r) = 0$ for $r \notin [r_{\min}, r_{\max}]$ or $\ell \notin [0, 2r_{\max}\eta_i]$. In our context, the back and forth observer system (7.61)-(7.62) can be written as

$$\begin{cases} \dfrac{\partial \hat{\psi}^{2n}}{\partial t}(t,r) = -G(t)\dfrac{\partial \hat{\psi}^{2n}}{\partial r}(t,r) - \mu \mathcal{K}^*(\mathcal{K}\hat{\psi}^{2n}(t,\cdot) - \bar{Q}(t,\cdot)) \\ \hat{\psi}^{2n}(0,r) = \begin{cases} \hat{\psi}^{2n-1}(0,r) & \text{if } n \geqslant 1 \\ \hat{\psi}_0(r) & \text{otherwise} \end{cases} \end{cases} \tag{7.61}$$

$$\begin{cases} \dfrac{\partial \hat{\psi}^{2n+1}}{\partial t}(t,r) = -G(t)\dfrac{\partial \hat{\psi}^{2n+1}}{\partial r}(t,r) + \mu \mathcal{K}^*(\mathcal{K}\hat{\psi}^{2n+1}(t,\cdot) - \bar{Q}(t,\cdot)) \\ \hat{\psi}^{2n+1}(t_{\max}, r) = \hat{\psi}^{2n}(t_{\max}, r) \end{cases} \tag{7.62}$$

where $t \in (0, t_{\max})$, $r \in (r_0, r_1)$ and $\mu > 0$ is a degree of freedom. In this system, $\hat{\psi}^n(t,r)$ is the estimation of the actual PSD $\psi(t,r)$ obtained after n iterations of the algorithm. Note that the algorithm relies only on the knowledge of the normalized CLD $\bar{Q}(t,\ell)$ on the time interval $[0, t_{\max}]$. The following result ensures the convergence of $\hat{\psi}^n$ to ψ up to an observability condition.

Theorem 7.17. *Assume that for all $\psi_0 \in X$, the following implication is satisfied:*

$$(\forall t \in [0, t_{\max}],\ \mathcal{K}\psi(t,\cdot) = 0) \quad \Longrightarrow \quad \psi_0 = 0, \tag{7.63}$$

where ψ denotes the solution of (7.59) with initial condition ψ_0. Then, for all $\mu > 0$, all $t \in [0, t_{\max}]$ and almost all $r \in [r_0, r_1]$,

$$\hat{\psi}^n(t,r) \underset{n \to +\infty}{\longrightarrow} \psi(t,r). \tag{7.64}$$

Proof. This result is an application of Theorem 6.13. Let $X = L^2((r_0, r_1); \mathbb{R})$ and $\mathcal{D} = \{\psi \in X \ : \ \psi' \in X, \psi(r_0) = \psi(r_1)\}$. As in Proposition 7.16, $-G(t)\frac{\partial}{\partial r} : \mathcal{D}^N \to X^N$ is skew-adjoint for all $t \in [0, t_{\max}]$. Moreover, (7.63) states that (7.59) with output $\mathcal{K}\psi$ is observable, that is, its observable subspace is X. Hence, all the hypotheses of Theorem 6.13 are satisfied, so that the BFN algorithm converges to the actual state of the system as the number of iterations goes to infinity. ■

Condition (7.63) is a *weak observability* condition, and can be reformulated in the following way. If two initial conditions ψ_0 and $\tilde{\psi}_0$ (*i.e.*, (u_i), $(\psi_{0,i})$, $(\tilde{u}_i)$, $(\tilde{\psi}_{0,i})$, $1 \leqslant i \leqslant N$) are such that the corresponding cumulative CLDs Q and $\tilde{Q}$ are the same on the whole time interval $[0, t_{\max}]$, then $\psi_0 = \tilde{\psi}_0$, which implies that the two PSDs are also the same on $[0, t_{\max}]$. Indeed, by taking the difference $\psi - \tilde{\psi}$, we recover (7.63). Hence, the main question to investigate is now: when does the observability condition (7.63) holds?

Due to the injectivity of the operator $\mathcal{K}$ in the single-shape case, the observability condition (7.63) is satisfied when $N = 1$.

Theorem 7.18. *If $N = 1$ (single-shape case), then for all $\mu > 0$, all $t \in [0, t_{\max}]$ and almost all $r \in [r_0, r_1]$,*

$$\hat{\psi}^n(t,r) \underset{n \to +\infty}{\longrightarrow} \psi(t,r). \tag{7.65}$$

In some crystallization processes, there are two shapes of crystals of the same species appearing simultaneously in the reactor due to polymorphism. Frequently, one of these shapes is almost spherical, and the other is very elongated (see Figure 7.10 and the experiments of [Gao+18] for example). Hence, according to Theorem 7.17, the BFN algorithm is able to estimate the actual PSD of each shape from the knowledge of the CLD during the process.

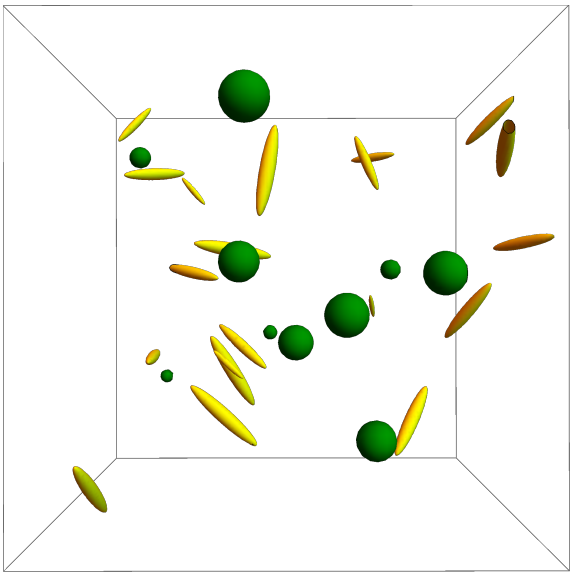


Figure 7.10 – Simulated suspension of ideal particles of two shapes, distributed in size: spheres ($\eta = 1$, in green) and prolate spheroids ($\eta = 6$, in yellow), in a cubic volume.

Theorem 7.19. *Consider two clusters of crystals* ($N = 2$) *with shapes* $\eta_1 = 1$ *and* $\eta_2 > 1$. *Assume that their growth rate have constant ratio* $\frac{g_1}{g_2}$, *i.e.,* $g_2 G_1(t) = g_1 G_2(t)$ *for all* $t \in [0, t_{\max}]$. *Then for all* $\psi_0 \in H^2(r_0, r_1)$ *satisfying the boundary condition* (7.10),

$$(\forall t \in [0, t_{\max}],\ \mathcal{K}\psi(t, \cdot) = 0) \quad \Longrightarrow \quad \psi_0 = 0, \tag{7.66}$$

Theorem (7.19) is proved in Appendix B.2.

Remark 7.20. The time $t_{\max} > 0$ is not necessarily the duration of the full process, it can theoretically be chosen as small as desired. This property is called "small time" observability. Even if the knowledge of the CLD at a fixed time t is not sufficient to estimate the corresponding PSD, measuring the CLD on a small time interval $[t, t + dt]$ on which the process occurs is sufficient to estimate the PSD on this same interval. Moreover, this property can be used to build an "almost" online observer in the sense that the BFN algorithm can be launched online on small time intervals during the process.

Numerical simulations

For the numerical simulations, we consider the set of parameters given in Table 7.11. Simulations of (7.59) and (7.61)-(7.62) are performed with forward/backward finite differences, with spacing $dr = \frac{1}{100}$ for ψ_1 with growth rate G_1 and $dr = \frac{1}{50}$ for ψ_2

$r_{\min} = 1.0 \times 10^{-4}$m	$r_{\max} = 2.0 \times 10^{-4}$m	$t_{\max} = 1$h	$N = 2$
$G_1 = 1.0 \times 10^{-4}$m.h^{-1}	$G_2 = 2.0 \times 10^{-4}$m.h^{-1}	$\eta_1 = 1$	$\eta_2 = 2$

Table 7.11 – Parameters of the numerical simulation of the BFN algorithm.

with growth rate G_2. We fix $\psi_1 = \psi_2 = 0$ at the initial time $t = 0$, and choose the nucleation rates u_1 and u_2 such that, at time $t = 1$h, we have (see blue line on Figure 7.12)

$$\psi_1(t_{\max}, r) = \psi_2(t_{\max}, r) = \frac{e^{-30(r-1.5\times10^{-4})^2}}{\int_{1\times10^{-4}}^{2\times10^{-4}} e^{-30(\rho-1.5\times10^{-4})^2}\mathrm{d}\rho}. \tag{7.67}$$

The BFN algorithm is initialized at $\hat{\psi}_1 = \hat{\psi}_2 = 0$. On Figure 7.12, we plot the estimations $\hat{\psi}_1$ and $\hat{\psi}_2$ obtained by BFN after $2n = 20$ and 100 iterations. After 20 iterations, the shape of the two PSDs is already well estimated. After 100 iterations, the estimation of ψ_2 is far more accurate. The error between the actual PSD and the estimation made by BFN, given by

$$\|\varepsilon^{2n}(t)\|_{L^2}^2 = \int_{1\times10^{-4}}^{2\times10^{-4}} \left(\psi_1(t,r) - \hat{\psi}_1^{2n}(t,r)\right)^2 + \left(\psi_2(t,r) - \hat{\psi}_2^{2n}(t,r)\right)^2 \mathrm{d}r, \tag{7.68}$$

is plotted in Figure 7.13. Applying a linear regression for $2n \geqslant 30$, the rate of convergence is estimated as $\|\varepsilon^{2n}(t)\|_{L^2} \approx 0.156 \times 0.986^n$.

Conclusion

In this section, we have shown how the results of Chapter 6 can be applied to the batch crystallization process under consideration. When measuring the CLD, the BFN algorithm is able to reconstruct the PSD at least in two different cases: *(i)* the single-shape case, where crystals are spheres; *(ii)* the two-shapes case, where crystals are either spheres, or prolate spheroids having a common shape factor. However, the convergence speed has not been investigated, and seems to be slow on numerical simulations. Theoretically, convergence holds in the strong topology (see Theorem (6.13)) but is not exponential since the system is not exactly observable (see Proposition 6.16). A spectral analysis of the family of operators $-G(t) - \mu\mathcal{K}^*\mathcal{K}$ should be carried out.

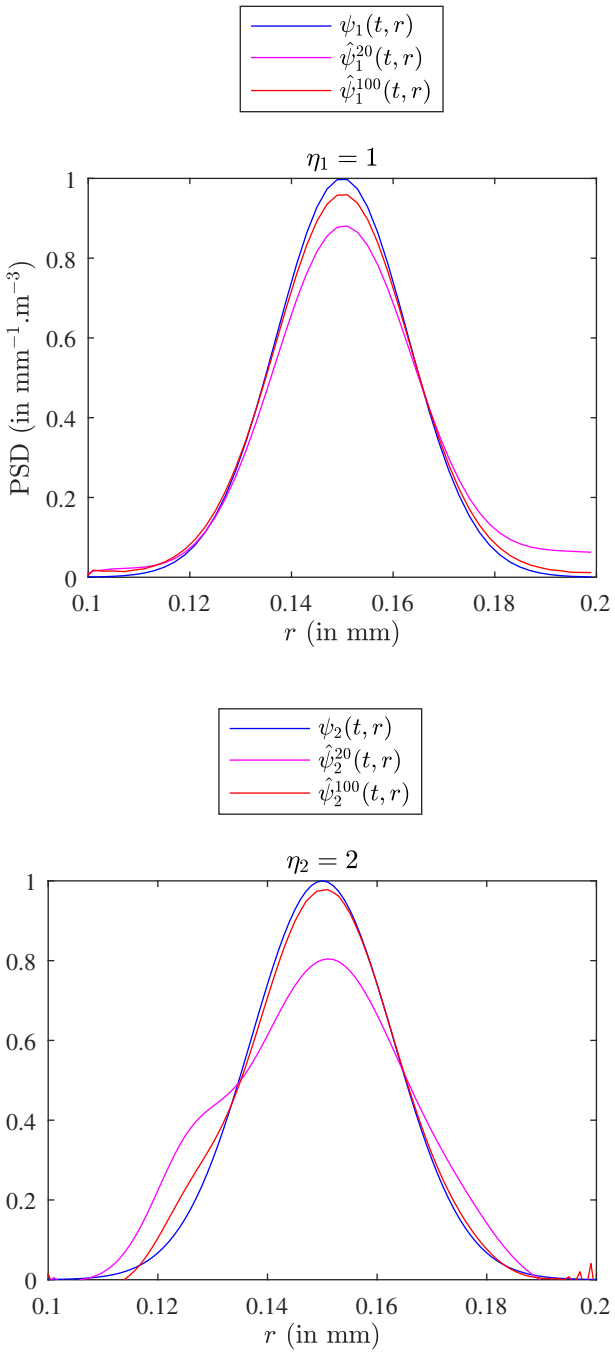


Figure 7.12 – PSDs ψ_1 and ψ_2 at time $t = 1$h and their estimation obtained by the BFN algorithm after $2n = 20$ and 100 iterations.

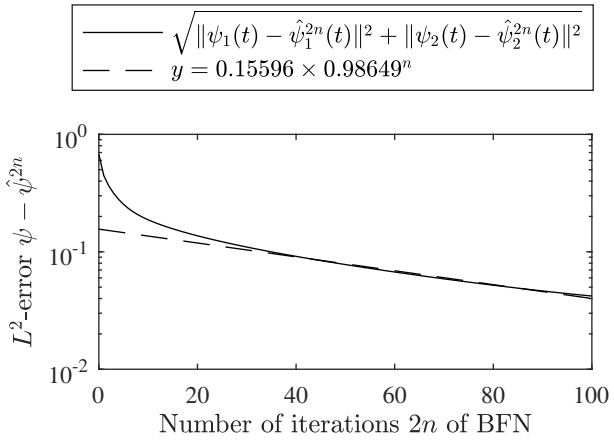


Figure 7.13 – Evolution of the absolute error between the actual PSDs ψ_1 and ψ_2 at time $t = 1$h and the estimations $\hat{\psi}_1^{2n}$ and $\hat{\psi}_2^{2n}$ obtained by (7.61)-(7.62) through iterations of the BFN algorithm.

Conclusion and perspectives

In the first part of this thesis, the problem of dynamic output feedback stabilization was investigated. Chapter 1 served as an introduction to this issue. Several necessary conditions were given, and the notion of uniform observability was recalled. A vast literature is devoted to the semi-global stabilization of uniformly observable systems by means of separation principles. However, it is not generic for nonlinear systems to be uniformly observable when the dimension of the output is less or equal than the dimension of the input. This motivates the emphasis made in the following chapters on non-uniformly observable systems.

Chapter 2 investigated SISO bilinear systems. Generically, there exist singular inputs making these systems unobservable, but the value of the input at the target point is not singular. To tackle the problem of semi-global dynamic output feedback stabilization for such systems, we set up a separation principle: the existence of a stabilizing state feedback is assumed, an observer is designed, and these two elements are combined in closed-loop. A major hurdle to overcome in order to prove the effectiveness of this method is that the inputs generated in closed-loop may be singular. In that case, observer convergence cannot be guaranteed, which in turn prevents stabilization. To manage this issue, we proposed to perturb the stabilizing state feedback to get new observability properties. We proved, under genericity assumptions on the system and for arbitrary compact sets of initial conditions, the existence of a dense open set of perturbations guaranteeing that the perturbed feedback laws were still stabilizing and preventing the closed-loop inputs to render the system unobservable. These results rely on transversality theory. We applied them to Luenberger and Kalman observers.

Thanks to this approach, if the trajectories of a system are *a priori* bounded, then semi-global output feedback stabilization can be achieved. However, proving the boundedness of trajectories remains an open problem. For uniformly observable systems, this is usually done by selecting a sufficiently large observer gain, which allows to tune the observer's convergence rate. For non-uniformly observable systems, nothing guarantees that increasing the observer gain actually increases the observer's convergence rate. Indeed, this would require that the inputs generated in closed-loop not only avoid those making the system unobservable, but also remain sufficiently far from these singular inputs. This issue should be addressed in future works. Combined with the results of Chapter 2, this would allow to set up a generic separation principle for SISO bilinear systems.

Chapter 3 dealt with state-affine dissipative systems. For these systems, a Luenberger observer with non-increasing error can be designed. This is a key tool to tackle observability singularities in the context of output feedback stabilization. Indeed, even if the input of the closed-loop system occasionally makes the system unobservable, this does not affect the observer performance. Thanks to this prop-

erty, we proved that 0-detectability is a necessary and sufficient condition to achieve a separation principle for dissipative systems. Moreover, no perturbation strategy is needed. We illustrated the results on the Ćuk converter and a heat exchanger. In the future, investigating how this result could be extended to infinite-dimensional systems would be interesting. Indeed, after choosing a suitable functional setting, the non-increase of the observer error should be preserved. However, local asymptotic stabilization (using the center manifold theorem) and ω-limit set techniques (using compactness conditions) should be carefully revised.

In Chapter 4, new guidelines for output feedback stabilization at an unobservable target point were introduced. Combining the insights suggested by the previous chapters, we illustrated two main ideas: well chosen perturbations of a stabilizing state feedback can yield new observability properties of the closed-loop system, and embedding into bilinear systems admitting observers with dissipative error systems allows to mitigate the observability issues. In a first part, an illustrative example of linear conservative system with nonlinear output was considered. The specific form of the (quadratic) output allowed to embed the system into a bilinear one, for which a Luenberger observer with dissipative error system could be designed. Using this new dissipativity property in conjunction with a feedback perturbation, semi-global dynamic output feedback stabilization was achieved. Since the embedding strategy highly depended on the relation between the system dynamics and the output, a natural question to ask is: under which conditions can a nonlinear observation-control system be embedded into a system for which an observer with dissipative error can be designed? Then, semi-global stabilization at an unobservable target point could be investigated for this class of systems.

Moreover, these embedding strategies could be used to design time-varying dynamic output feedbacks. Indeed, being able to guarantee that the error between the state of the system and the estimation made by the observer is non-increasing is an important tool for the design of switched feedback laws, where the input is alternatively chosen to estimate or stabilize the state. This would help to coalesce the work of [Cor94a] and [ST03], and to get semi-global stabilization at an unobservable target. This is the direction taken by [BS21a].

In the second part of Chapter 4, we proposed to embed the original system into a unitary bilinear infinite-dimensional one. The strategy relies on unitary representation theory. An infinite-dimensional Luenberger observer was designed on the embedded system, under the assumption that the output has been linearized by the embedding. The infinite-dimensional system being unitary, the observer error system was dissipative. Finally, a perturbation of the feedback law was used to close the loop. Under assumptions of short-time 0-detectability and isolated observability singularity, we proved the asymptotic stability of the closed-loop system combining the finite-dimensional system and the infinite-dimensional observer. This strategy allowed to consider more general output maps, the counterpart being the infinite-dimensionality of the closed-loop system. Extending this approach to a wider class of nonlinear systems is an important question that will be raised in future works.

Beyond the specific embedding technique introduced in this thesis, let us stress that topological obstructions to output feedback stabilization are lifted when infinite-dimensional observers are considered. For example, the obstruction brought up in [Cor94a] regarding the stabilizability of $\dot{x} = u$, $y = x^2$ vanishes if one extends the usual definition of dynamic output feedback stabilizability by allowing infinite-

dimensional states fed by the output. Finding an example of system that is not stabilizable by means of a finite-dimensional dynamic output feedback but that is stabilizable by means of an infinite-dimensional one would fully justify the need of such embeddings. Finally, the use of infinite-dimensional observers naturally led us to the following chapters.

In the second part of this thesis, we addressed the problem of observer design for infinite-dimensional time-varying linear systems. The convergence properties of infinite-dimensional Luenberger observers were investigated in Chapter 5. We focused on approximate observability-like hypotheses, and relied on a weak detectability assumption. This assumption guaranteed that the distance between two trajectories sharing the same output was non-increasing. Since the system was time-varying, we studied the convergence of the observer for time sequences for which the evolution system converged and used (almost) periodicity assumptions. We showed that the observer estimated the so-called observable subspace of the system, at least in the weak topology of the state space. Strong convergence was obtained under additional hypotheses on the system.

Similar techniques were used in Chapter 6 to tackle the problem of offline estimation of the initial data of the system from the knowledge of the output over a finite time interval. The system was assumed to be bi-directional, and the BFN algorithm was used. It is based on iterations of forward and backward observers. After each iteration, the final estimation of the state obtained by the observer is chosen as the initial condition of the next observer. Using the tools developed in Chapter 5, we gave sufficient conditions for the observer convergence in the weak and strong topologies.

While the BFN algorithm is originally meant to be used on finite time intervals, one can imagine to build an online asymptotic observer based on its paradigm. Indeed, instead of using an asymptotic observer to estimate the state of a system from the online measurement of the output, one could use a BFN observer on the moving horizon of past times. This would require to record the values of the output, and to emulate an accelerated observer dynamics reusing these values several times. Then, two important questions naturally arise: is this observer still converging? If yes, can we compare its convergence speed with that of the usual asymptotic observer? These questions could be investigated in future works.

In Chapter 7, various observer problems were considered for a model of crystallization processes. The evolution of the PSD (infinite-dimensional state of the system) was modeled, as well as several measurements. In particular, a model of the CLD was derived for spheroid crystals. Three techniques of PSD estimation were proposed. The first one was a direct approach based on a Tikhonov regularization method, that recovers the PSD from the CLD without using the system dynamics. The main benefit of this method was that it is easily implementable in practice, since it does not rely on any evolution model of the process. However, it was unable to deal with the case of crystals having different shapes, and the lack of dynamical model limits its performances. The second one was an infinite-dimensional extension of the KKL observer, based on temperature and solute concentration. These measurements are not sufficient to fully characterized the PSD. Still, the method could be used to obtain some preliminary information on the PSD. The last one was based on the infinite-dimensional Luenberger observers investigated in Chapters 5 and 6. It was the most promising method, and was able to deal with the multi-shape

case for some specific combinations of spheroid crystals. Extending the results to more general shapes or combinations of shapes, as it is done in [BS21b], would be interesting. For each method, numerical simulations have been carried out, and they should be evaluated on experiments in a future research project.

Appendix A

Appendix of Part I

A.1 Proof of Lemma 2.30

We prove the first part of the statement by induction on i. For $i = 1$, one easily checks that

$$Q_1^{1+k} = B, \qquad \forall k \in \mathbb{N}. \tag{A.1}$$

Assuming the desired property for i, we prove the existence of $R_{i+1}^0, \dots, R_{i+1}^i \in \mathrm{End}(\mathbb{R}^n)[X_0, \dots X_i]$ such that

$$Q_{i+1}^{i+1+k} = \sum_{j=0}^{i} k^j R_{i+1}^j, \qquad \forall k \geqslant 0.$$

Using the definition of $Q_{i+1}^{i+1+\ell}$ and the recurrence relation (2.21) yields

$$Q_{i+1}^{i+1+\ell} = \Psi(Q_i^{i+\ell}) + Q_{i+1}^{i+\ell}, \qquad \forall \ell \geqslant 1. \tag{A.2}$$

Consequently, for all $k \geqslant 0$,

$$\begin{aligned}
Q_{i+1}^{i+1+k} &= \sum_{\ell=1}^{k} \left(Q_{i+1}^{i+1+\ell} - Q_{i+1}^{i+\ell} \right) + Q_{i+1}^{i+1} \\
&= \sum_{\ell=1}^{k} \left(\Psi(Q_i^{i+\ell}) \right) + Q_{i+1}^{i+1} && \text{(by (A.2))} \\
&= \sum_{\ell=1}^{k} \left(\sum_{j=0}^{i-1} \ell^j \Psi(R_i^j) \right) + Q_{i+1}^{i+1} && \text{(by induction hypothesis)} \\
&= \sum_{j=0}^{i-1} \left(\sum_{\ell=1}^{k} \ell^j \right) \Psi(R_i^j) + Q_{i+1}^{i+1} \\
&= \sum_{j=0}^{i-1} S^j(k) \Psi(R_i^j) + Q_{i+1}^{i+1}, \qquad \text{with } S^j(k) = \sum_{\ell=1}^{k} \ell^j.
\end{aligned}$$

Note that $Q_{i+1}^{i+1}, \Psi(R_i^j) \in \mathrm{End}(\mathbb{R}^n)[X_0, \dots, X_i]$ for all $j \in \{0, \dots, i-1\}$ ($Q_{i+1}^{i+1} = \partial P_{i+1}/\partial X_0$). Moreover, according to Faulhaber's formula, we have

$$S^j(k) = \frac{k^{j+1}}{j+1} + T^j(k), \qquad \forall j, k \in \mathbb{N},$$

where $T^j(k)$ is a polynomial in the variable k of degree j with no constant term. Consequently,

$$\begin{aligned} Q_{i+1}^{i+1+k} &= \frac{k^i}{i}\Psi(R_i^{i-1}) + \left(T^{i-1}(k)\Psi(R_i^{i-1}) + \sum_{j=0}^{i-2} S^j(k)\Psi(R_i^j)\right) + Q_{i+1}^{i+1} \\ &= k^i R_{i+1}^i + \sum_{j=1}^{i-1} k^j R_{i+1}^j + R_{i+1}^0 \\ &= \sum_{j=0}^{i} k^j R_{i+1}^j, \end{aligned}$$

with $R_{i+1}^i = \Psi(R_i^{i-1})/i$, $R_{i+1}^0 = Q_{i+1}^{i+1}$ and $R_{i+1}^j \in \mathrm{End}(\mathbb{R}^n)[X_0, \dots, X_i]$ for all $j \in \{0, \dots, i\}$.

The second part of the statement easily follows by induction. Indeed,

$$BP_0 = Q_1^1 = \sum_{j=0}^{0} 0^j R_1^j = R_1^0,$$

and

$$R_{i+1}^i = \frac{\Psi(R_i^{i-1})}{i} = \frac{1}{i}\Psi\left(\frac{1}{(i-1)!}BP_{i-1}\right) = \frac{1}{i!}B\Psi(P_{i-1}) = \frac{1}{i!}BP_i.$$

The statement follows.

A.2 Proof of Lemma 4.9

Let us compute the determinant of $\mathcal{Q}$.

$$\det \mathcal{Q} = \begin{vmatrix} K & \delta & 0 \\ KA & 0 & -\delta\alpha \\ \vdots & \vdots & \vdots \\ KA^{n+1} & 0 & \delta(-\alpha)^{n+1} \end{vmatrix} = (-1)^{n+1}\delta^2\alpha \begin{vmatrix} KA & 1 \\ \vdots & \vdots \\ KA^{n+1} & (-\alpha)^n \end{vmatrix} = -\delta^2\alpha \sum_{k=0}^{n} \alpha^k Q(k)$$

where

$$Q(k) = \begin{vmatrix} \tilde{K}A^0 \\ \vdots \\ \tilde{K}A^{k-1} \\ \tilde{K}A^{k+1} \\ \vdots \\ \tilde{K}A^n \end{vmatrix}, \qquad \tilde{K} = KA, \qquad k \in \{0, \dots, n\}.$$

Let $P(X) = \sum_{k=0}^{n} c_k X^k$ be the characteristic polynomial of A. Since A is skew-symmetric and invertible, it holds that n is even, P is minimal for A, positive on $\mathbb{R}$, $c_n = 1$. Then,

$$A^n = -\sum_{k=0}^{n-1} c_k A^k.$$

Let Δ be the determinant of the Kalman observability matrix of $(\tilde{K}, A)$. Since (K, A) is observable and A is invertible, $\Delta \neq 0$. Then for $k < n$,

$$Q(k) = \begin{vmatrix} \tilde{K}A^0 \\ \vdots \\ \tilde{K}A^{k-1} \\ \tilde{K}A^{k+1} \\ \vdots \\ \sum_{i=0}^{n-1} c_i \tilde{K}A^i. \end{vmatrix} = \begin{vmatrix} \tilde{K}A^0 \\ \vdots \\ \tilde{K}A^{k-1} \\ \tilde{K}A^{k+1} \\ \vdots \\ -c_k \tilde{K}A^k. \end{vmatrix} = -c_k(-1)^{n-k}\Delta.$$

The case $k = n$ simply yields $Q(n) = \Delta$. Then

$$\det \mathcal{Q} = \delta^2 \alpha \Delta \sum_{k=0}^{n} c_k(-1)^k \alpha^k = \delta^2 \alpha \Delta P(-\alpha).$$

Since P is positive on $\mathbb{R}$, $\det \mathcal{Q} \neq 0$ as soon as $\alpha > 0$.

Appendix B

Appendix of Part II

B.1 Proof of Propositions 7.10 and 7.11.

Assume that $G \geqslant \mu > 0$. Let $\tau \in (0, t_{\max}]$ and $u \in H^4(0,\tau)$. Let ψ be the solution of (7.6) with initial condition ψ_0 and boundary condition u. We introduce a time reparametrization $\tilde{t} = \int_0^t G(s)\mathrm{d}s$, which is well defined since $G \geqslant \mu$. Let $\tilde{\psi}$, $\tilde{u}$ and $\tilde{y}$ be such that $\tilde{\psi}(\tilde{t}) = \psi(t)$, $\tilde{u}(\tilde{t}) = u(t)$ and $\tilde{y}(\tilde{t}) = y(t)$ for all $t \in [0, t_{\max}]$. Then

$$\begin{cases} \partial_{\tilde{t}}\tilde{\psi}(\tilde{t},r) = -\partial_r\tilde{\psi}(\tilde{t},r) \\ \tilde{\psi}(0,r) = \psi_0(r) \\ \tilde{\psi}(\tilde{t},r_{\min}) = \tilde{u}(\tilde{t}) \end{cases} \tag{B.1}$$

and $\tilde{y}(\tilde{t}) = \int_{r_{\min}}^{r_{\max}} \tilde{\psi}(\tilde{t},r) r^3 \mathrm{d}r$. Since the observability properties are not affected by the time reparametrization, one can investigate observability properties of the system (B.1) instead of (7.6). Therefore, one can assume without loss of generality that $G = 1$ in the rest of the proof. Since $u \in H^4(0,\tau)$, we have $y \in C^4(0,\tau)$. Equation (7.1) and system (7.6) yield

$$\begin{aligned} y' &= 3\int_{r_{\min}}^{r_{\max}} r^2\psi(\cdot,r)\mathrm{d}r - \left[r^3\psi(\cdot,r)\right]_{r_{\min}}^{r_{\max}} \\ &= 3\int_{r_{\min}}^{r_{\max}} r^2\psi(\cdot,r)\mathrm{d}r + r_{\min}^3 u, \qquad \text{(B.2)} \\ y^{(2)} &= 6\int_{r_{\min}}^{r_{\max}} r\psi(\cdot,r)\mathrm{d}r + 3\left[r^2\psi(\cdot,r)\right]_{r_{\min}}^{r_{\max}} + r_{\min}^3 u' \\ &= 6\int_{r_{\min}}^{r_{\max}} r\psi(\cdot,r)\mathrm{d}r + 3r_{\min}^2 u + r_{\min}^3 u', \qquad \text{(B.3)} \\ y^{(3)} &= 6\int_{r_{\min}}^{r_{\max}} \psi(\cdot,r)\mathrm{d}r - 6\left[r\psi(\cdot,r)\right]_{r_{\min}}^{r_{\max}} + 3\, r_{\min}^2 u' + r_{\min}^3 u^{(2)} \\ &= 6\int_{r_{\min}}^{r_{\max}} \psi(\cdot,r)\mathrm{d}r + 6r_{\min}u + 3r_{\min}^2 u' + r_{\min}^3 u^{(2)}, \qquad \text{(B.4)} \\ y^{(4)} &= -6\left[\psi(\cdot,r)\right]_{r_{\min}}^{r_{\max}} + 6r_{\min}u' + 3r_{\min}^2 u^{(2)}(t) + r_{\min}^3 u^{(3)} \\ &= 6u + 6r_{\min}u' + 3r_{\min}^2 u^{(2)} + r_{\min}^3 u^{(3)}. \qquad \text{(B.5)} \end{aligned}$$

End of the proof of Proposition 7.10. By hypothesis, $\psi_0 = 0$. Consequently, Equa-

tions (B.2)–(B.4) yield

$$\begin{cases} y'(0) = r_{\min}^3 u(0) \\ y^{(2)}(0) = 3r_{\min}^2 u(0) + r_{\min}^3 u'(0) \\ y^{(3)}(0) = 6r_{\min} u(0) + 3r_{\min}^2 u'(0) + r_{\min}^3 u^{(2)}(0), \end{cases}$$

which is a triangular system with non vanishing diagonal since $r_{\min} > 0$. Hence $u(0)$, $u'(0)$ and $u^{(2)}(0)$ are determined by y. Moreover, on $[0, \tau]$, u satisfies Equation (B.5) which is a 3rd order ordinary differential equation. Hence, according to the Cauchy-Lipschitz theorem, there exits a unique solution u to this problem. Thus y determines u uniquely, that $u \mapsto y$ is injective. ■

End of the proof of Proposition 7.11. Substituting the boundary condition in equation (B.5) with $u = 0$ yields $y^{(4)} = 0$ identically on $[0, \tau]$. Hence y is a polynomial function of degree less or equal than 3. Thus the linear function that maps any solution of (7.6) with null boundary condition to its third moment has rank 4. Since ψ lies in an infinite dimensional vector space, we get by the rank-nullity theorem that its kernel is non-trivial, *i.e.* the state-output map $\psi \mapsto y$ is not injective, and the system has a 4-dimensional observable part. ■

Note that Proposition 7.11 relies deeply on Hypothesis (7.1). Hence the non-injectivity of the measurement is due to the fact that the system is observed on a too small time interval. If the system was observed on $[0, +\infty)$, then one could show with similar argues an injectivity result.

B.2 Proof of Theorem 7.19

Since $t_{\max}$ can be chosen as small as desired (see remark 7.20), we actually show that if $\psi_0 \neq 0$, the set of times $t \in [0, t_{\max}]$ such that $C\psi(t) \neq 0$ is dense in $[0, t_{\max}]$. Moreover, $C\psi = C_1\psi_1 + C_2\psi_2$, where

$$\begin{aligned} \mathcal{K}_i : L^2((r_{\min}, r_{\max}); \mathbb{R}) &\longrightarrow L^2((0, \ell_{\max}); \mathbb{R}) \\ \psi &\longmapsto \left(\ell \mapsto \textstyle\int_{r_{\min}}^{r_{\max}} k_i(\ell, r)\psi(r)\mathrm{d}r\right) \end{aligned}$$

Hence, proving (7.66) is equivalent to proving that if $\psi_1(t, \cdot) = C_2\psi_2(t, \cdot)$ for all $t \in [0, t_{\max}]$, then $\psi_1(t, r) = \psi_2(t, r) = 0$ for all $t \in [0, t_{\max}]$ and all $r \in [r_{\min}, r_{\max}]$. The proof relies on properties of the successive derivatives of $C_i\psi_i$.

Let $\mathcal{F} : L^2((r_{\min}, r_{\max}); \mathbb{R}) \to \mathbb{R}^{\mathbb{N}^*}$ be the linear map such that

$$\mathcal{F}_n(\psi) = (\mathcal{F}(\psi))_n = \int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n}}\mathrm{d}r, \qquad \forall \psi \in L^2((r_{\min}, r_{\max}); \mathbb{R}), n \in \mathbb{N}^*.$$

For all $\eta > 0$, recall the definitions of Section 7.2.2:

$$a_n(\eta) = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \alpha_\eta^n(\phi, \theta) \frac{\sin\theta}{4\pi} \mathrm{d}\theta\mathrm{d}\phi, \qquad b_n = \frac{(2n)!}{(n!)^2(1-2n)4^{2n}}.$$

Let $\mathcal{A}(\eta)$ and $\mathcal{B}$ be the linear endomorphisms on $\mathbb{R}^{\mathbb{N}^*}$ such that, for any $(u_n)_{n\in\mathbb{N}^*}$

$$(\mathcal{A}(\eta)u)_n = a_n(\eta)u_n, \qquad (\mathcal{B}u)_n = b_n u_n.$$

Then

$$\left((\mathcal{K}_i\psi)^{(2n)}(0)\right)_{n\in\mathbb{N}^*} = \mathcal{B}\mathcal{A}(\eta_i)\mathcal{F}\psi.$$

Lemma B.1 (Asymptotic properties of (a_n))**.** *The sequence* $(a_n(\eta))_{n\in\mathbb{N}}$ *is such that*

$$\lim_{n\to\infty}\frac{a_{n+1}(\eta)}{a_n(\eta)}=\begin{cases}1 & \text{if } \eta\geqslant 1,\\ \dfrac{1}{\eta^2} & \text{if } \eta<1.\end{cases}$$

Furthermore, $a_n(\eta)>\sqrt{\pi/n}$ *and if* $\eta>1$ *then* $a_n(\eta)\to 0$.

Proof. Recall that $\alpha_\eta(\phi,\theta)=\dfrac{\cos^2\phi}{\cos^2\theta+\eta^2\sin^2\theta}+\sin^2\phi$. Then

$$\|\alpha_\eta\|_\infty=\max_{\substack{\phi\in[0,2\pi]\\ \theta\in[0,\pi]}}\alpha_\eta(\phi,\theta)=\begin{cases}1 & \text{if } \eta\geqslant 1,\\ \dfrac{1}{\eta^2} & \text{if } \eta<1.\end{cases}$$

Recall that $\frac{\sin\theta}{4\pi}$ is the density of a probability measure μ on $(\phi,\theta)\in[0,2\pi]\times[0,\pi]$. If we denote by $\mathbb{E}_\mu$ the expected value with respect to μ, we obtain $a_n(\eta)=\mathbb{E}_\mu\left(\alpha_\eta^n\right)$. Then

$$a_{n+1}(\eta)=\mathbb{E}_\mu\left(\alpha_\eta^{n+1}\right)\leqslant\|\alpha_\eta\|_\infty\mathbb{E}_\mu\left(\alpha_\eta^n\right)=\|\alpha\|_\infty a_n(\eta).$$

On the other hand, $a_{n+1}(\eta)=\mathbb{E}_\mu\left(\left(\alpha_\eta^n\right)^{\frac{n+1}{n}}\right)$. Notice that the function $x\mapsto x^{\frac{n+1}{n}}=x^{1+\frac{1}{n}}$ is convex. Hence, Jensen's inequality implies

$$a_{n+1}(\eta)=\mathbb{E}_\mu\left(\left(\alpha_\eta^n\right)^{\frac{n+1}{n}}\right)\geqslant\left(\mathbb{E}_\mu\left(\alpha_\eta^n\right)\right)^{1+\frac{1}{n}}=(a_n(\eta))^{1+\frac{1}{n}}$$

Thus $(a_n(\eta))^{\frac{1}{n}}\leqslant\frac{a_{n+1}(\eta)}{a_n(\eta)}\leqslant\|\alpha_\eta\|_\infty$. Since μ is a probability measure,

$$(a_n(\eta))^{\frac{1}{n}}=\left(\mathbb{E}_\mu\left(\alpha_\eta^n\right)\right)^{\frac{1}{n}}=\|\alpha_\eta\|_{L^n(\mu)}\xrightarrow[n\to\infty]{}\|\alpha_\eta\|_{L^\infty(\mu)}=\|\alpha_\eta\|_\infty,$$

which concludes the proof of the first stated limit.

Regarding the supplementary asymptotic information, we first have naturally

$$a_n(\eta)\geqslant\int_0^{2\pi}\sin^{2n}\phi\mathrm{d}\phi=2\pi\frac{(2n)!}{2^{2n}(n!)^2}\sim 2\sqrt{\frac{\pi}{n}}.$$

The last limit stated is a consequence of Lebesgue's dominated convergence theorem, since $\alpha_\eta^n(\phi,\theta)\xrightarrow[n\to\infty]{}0$ for all (θ,ϕ) such that $\phi\neq k\pi$, $k\in\mathbb{Z}$, (in which case $\alpha_\eta^n(\phi,\theta)=1$), and $0\leqslant\alpha_\eta^n(\phi,\theta)\leqslant 1$. ∎

Regarding the map $\mathcal{F}$, we have the following lemma.

Lemma B.2. *Let* ψ *be continuous and such that* $\psi(r_{\min})\neq 0$. *Then*

$$\int_{r_{\min}}^{r_{\max}}\frac{\psi(r)}{r^{2n}}\mathrm{d}r\sim\frac{\psi(r_{\min})}{2nr_{\min}^{2n-1}}.$$

Proof. Without loss of generality, we can assume that $\psi(r_{\min})>0$. Let $\mu\in(0,1)$, then by continuity of ψ, there exists $R\in(r_{\min},r_{\max}]$ such that for all $r\in[r_{\min},R)$, $\psi(r)\in(\psi(r_{\min})(1-\mu),\psi(r_{\min})(1+\mu))$. Then

$$\int_{r_{\min}}^{r_{\max}}\frac{\psi(r)}{r^{2n}}\mathrm{d}r=\int_{r_{\min}}^{R}\frac{\psi(r)}{r^{2n}}\mathrm{d}r+\int_{R}^{r_{\max}}\frac{\psi(r)}{r^{2n}}\mathrm{d}r$$

$$\left|\int_R^{r_{\max}} \frac{\psi(r)}{r^{2n}}\mathrm{d}r\right| \leqslant \frac{1}{2n-1}\frac{\|\psi\|_\infty}{R^{2n-1}}.$$

On the other hand,

$$\left|\int_{r_{\min}}^{R} \frac{\psi(r)}{r^{2n}}\mathrm{d}r - \frac{\psi(r_{\min})}{2nr_{\min}^{2n-1}}\right| \leqslant \frac{\psi(r_{\min})}{2n(2n-1)r_{\min}^{2n-1}} + \frac{\mu\psi(r_{\min})}{(2n-1)r_{\min}^{2n-1}} + \frac{\psi(r_{\min})(1+\mu)}{(2n-1)R^{2n-1}}.$$

As a consequence,

$$\begin{aligned}\left|\frac{2nr_{\min}^{2n+1}}{\psi(r_{\min})}\int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n}}\mathrm{d}r - 1\right| \leqslant & \mu\frac{2n}{2n-1} + \frac{1}{2n-1} + \frac{(1+\mu)2n}{(2n-1)}\left(\frac{r_{\min}}{R}\right)^{2n-1} \\ & + \frac{2n}{2n-1}\frac{\|\psi\|_\infty}{\psi(r_{\min})}\left(\frac{r_{\min}}{r_{\max}}\right)^{2n-1}.\end{aligned}$$

The right-hand side has limit μ for any μ (independently of the value of R, which is always larger than $r_{\min}$), hence the left hand side has limit 0. ■

By integration by parts, we can obtain a corollary.

Corollary B.3. *Let ψ be continuously differentiable and such that $\psi(r_{\min}) = 0$, $\psi(r_{\max}) = 0$ and $\psi'(r_{\min}) \neq 0$. Then*

$$\int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n}}\mathrm{d}r \sim \frac{\psi'(r_{\min})}{4n^2 r_{\min}^{2n-1}}.$$

These last two results allow to prove the following statement.

Proposition B.4. *There are no solutions ψ_1, ψ_2 to $\mathcal{F}(\psi_1) = \mathcal{A}(\eta)\mathcal{F}(\psi_2)$ (with $\psi_i(r_{\max}) = 0$) such that $\psi_1(r_{\min}), \partial_r\psi_1(r_{\min}), \psi_2(r_{\min}), \partial_r\psi_2(r_{\min})$ are not all equal to 0.*

Proof. According to Lemmas B.1-B.2 and Corollary B.3, if $\psi_1(r_{\min}) \neq 0$, then $r_{\min}^{2n-1}\mathcal{F}_n(\psi_i)$ converges to 0 slower than $r_{\min}^{2n-1}a_n(\eta)\mathcal{F}_n(\psi_2)$, since $a_n(\eta) \to 0$. On the other hand, if $\psi_1(r_{\min}) = 0$ then having $\psi_2(r_{\min}) \neq 0$ implies that $a_n(\eta)r_{\min}^{2n-1}\mathcal{F}_n(\psi_2)$ now converges slower than $r_{\min}^{2n-1}\mathcal{F}_n(\psi_1)$ since $a_n(\eta) \geqslant 2\sqrt{\pi/n}$. Hence this implies that we must also have $\psi_2(r_{\min}) = 0$. The same argument repeated on the derivatives yields the statement. ■

In this first case, observability is proved by a sort of injectivity argument, the images of $\mathcal{K}_1$ and $\mathcal{K}_2$ are such that their intersection cannot be reached through functions ψ that do not vanish at $r_{\min}$.

Proposition B.5. *Assume $\eta_1 = 1$ and $\eta_2 = \eta > 1$. If ψ_1, ψ_2 are two non-zero solutions of the transport equation such that for some $\tau \in [0, t_{\max}]$, $\psi_1(\tau), \psi_2(\tau) \in H_0^2(r_{\min}, r_{\max})$, then there exists no $\varepsilon > 0$ such that*

$$\mathcal{K}_1(\psi_1(t)) = \mathcal{K}_2(\psi_2(t)) \qquad \forall t \in (\tau-\varepsilon, \tau+\varepsilon) \cap [0, t_{\max}].$$

Proof. By iterated integration by parts, for any $\psi \in H_0^2(r_{\min}, r_{\max})$

$$\int_{r_{\min}}^{r_{\max}} \frac{\psi''(r)}{r^{2n}}\mathrm{d}r = \left[\frac{\psi'(r)}{r^{2n}}\right]_{r_{\min}}^{r_{\max}} - \left[2n\frac{\psi(r)}{r^{2n+1}}\right]_{r_{\min}}^{r_{\max}} + 2n(2n+1)\int_{r_{\min}}^{r_{\max}} \frac{\psi(r)}{r^{2n+2}}\mathrm{d}r.$$

Hence for both ψ_i, $i \in \{1,2\}$, at $t=\tau$,

$$\mathcal{F}_n(\psi_i''(\tau)) = 2n(2n+1)\mathcal{F}_{n+1}(\psi_i(\tau)).$$

We prove the result by contradiction. Assume there exists $\varepsilon > 0$ such that

$$\mathcal{K}_1(\psi_1(t)) = \mathcal{K}_2(\psi_2(t)), \qquad \forall t \in (\tau-\varepsilon, \tau+\varepsilon) \cap [0, t_{\max}],$$

implies that

$$\mathcal{BA}(\eta_1)\mathcal{F}\psi_1(t) = \mathcal{BA}(\eta_2)\mathcal{F}\psi_2(t), \qquad \forall t \in (\tau-\varepsilon, \tau+\varepsilon) \cap [0, t_{\max}], \tag{B.6}$$

and, term-wise,

$$\mathcal{F}_n(\psi_2(t)) = \frac{a_n(\eta_1)}{a_n(\eta_2)}\mathcal{F}_n(\psi_1(t)), \qquad \forall t \in (\tau-\varepsilon, \tau+\varepsilon) \cap [0, t_{\max}], \forall n \in \mathbb{N}^*.$$

On the other hand, equation (B.6) can be differentiated with respect to time. With

$$\frac{g_i}{G_i(t)}\frac{\partial}{\partial t}\frac{g_i}{G_i(t)}\frac{\partial}{\partial t}\psi_i(t,r) = g_i^2\frac{\partial^2\psi_i}{\partial r^2}(t,r) \qquad \forall t \in [0, t_{\max}]$$

hence, by the assumption that $g_1/G_1(t) = g_2/G_2(t)$,

$$g_1^2\mathcal{K}_1\left(\frac{\partial^2\psi_1}{\partial r^2}(t)\right) = g_2^2\mathcal{K}_2\left(\frac{\partial^2\psi_2}{\partial r^2}(t)\right), \qquad \forall t \in (\tau-\varepsilon, \tau+\varepsilon) \cap [0, t_{\max}].$$

Likewise, this implies

$$g_1^2\mathcal{BA}(\eta_1)\mathcal{F}\frac{\partial^2\psi_1}{\partial r^2}(t) = g_2^2\mathcal{BA}(\eta_2)\mathcal{F}\frac{\partial^2\psi_2}{\partial r^2}(t), \qquad \forall t \in (\tau-\varepsilon, \tau+\varepsilon) \cap [0, t_{\max}], \tag{B.7}$$

and, term-wise,

$$\mathcal{F}_n\left(\frac{\partial^2\psi_2}{\partial r^2}(t)\right) = \frac{g_1^2}{g_2^2}\frac{a_n(\eta_1)}{a_n(\eta_2)}\mathcal{F}_n\left(\frac{\partial^2\psi_2}{\partial r^2}(t)\right), \qquad \forall t \in (\tau-\varepsilon, \tau+\varepsilon) \cap [0, t_{\max}], \forall n \in \mathbb{N}^*.$$

Since equations (B.6)-(B.7) hold, we have both

$$\begin{aligned} \mathcal{F}_n(\psi_1''(\tau)) &= 2n(2n+1)\mathcal{F}_{n+1}(\psi_1(\tau)), \\ \frac{g_1^2}{g_2^2}\frac{a_n(\eta_1)}{a_{n+1}(\eta_1)}\frac{a_{n+1}(\eta_2)}{a_n(\eta_2)}\mathcal{F}_n(\psi_1''(\tau)) &= 2n(2n+1)\mathcal{F}_{n+1}(\psi_1(\tau)). \end{aligned}$$

If there isn't an infinity of non-zero terms in $\mathcal{F}\psi_1$, the function ψ_1 must be equal to zero since the family $(r \mapsto 1/r^{2n})_{n\in\mathbb{N}^*}$ is total. Assuming $\psi_1 \neq 0$, then there is an infinity of non zero terms and, up to an extraction $(n_k)_{k\in\mathbb{N}^*}$ such that $\mathcal{F}_{n_k}(\psi_1) \neq 0$ for all $k \in \mathbb{N}^*$, and

$$\frac{g_1^2}{g_2^2}\frac{a_{n_k}(\eta_1)}{a_{n_k+1}(\eta_1)}\frac{a_{n_k+1}(\eta_2)}{a_{n_k}(\eta_2)} = 1. \tag{B.8}$$

If $\eta \geqslant 1$, $\frac{a_n(\eta)}{a_{n+1}(\eta)} \to 1$, hence (B.8) is is leading to an incoherent limit except in the case $g_1^2 = g_2^2$. However, since $a_n(1) = 1$ and $\frac{a_{n+1}(\eta_2)}{a_n(\eta_2)} > 1$, (B.8) cannot be satisfied termwise if $g_1^2 = g_2^2$. ■

Proposition B.6. *Assume $\eta_1 = 1$ and $\eta_2 = \eta > 1$. Let ψ_1, ψ_2 be two non-zero $H^2(r_{\min}, r_{\max})$ solutions of their respective transport equations such that*

$$\psi_i(r_{\max}, t) = 0, \qquad \forall t \in [0, t_{\max}], i \in \{1, 2\}.$$

Then the set of times $t \in [0, t_{\max}]$ such that

$$\mathcal{K}_1(\psi_1(t)) \neq \mathcal{K}_2(\psi_2(t))$$

is dense in $[0, t_{\max}]$.

Proof. Pick a time $t \in [0, t_{\max}]$. On the one hand, if $\psi_i(r_{\min}, t) \neq 0$, or $\partial_r \psi_i(r_{\min}, t) \neq 0$ for $i = 1$ or $i = 2$, then Proposition B.4 applies to prove that $\mathcal{K}_1(\psi_1(t)) \neq \mathcal{K}_2(\psi_2(t))$. On the other hand, if $\psi_1(t), \psi_2(t) \in H_0^2(r_{\min}, r_{\max})$, then Proposition B.5 applies to prove that if t is such that $\mathcal{K}_1(\psi_1(t)) = \mathcal{K}_2(\psi_2(t))$ then any open interval containing t must also contain a time t' for which $\mathcal{K}_1(\psi_1(t')) \neq \mathcal{K}_2(\psi_2(t'))$. This proves the statement. ■

Proposition B.6 implies Theorem 7.19, which concludes the observability analysis.

Appendix C

Article on discrete KKL observers

Luenberger observers for discrete-time nonlinear systems

Lucas Brivadis, Vincent Andrieu and Ulysse Serres

The authors are with Univ. Lyon, Université Claude Bernard Lyon 1, CNRS, LAGEPP UMR 5007, 43 bd du 11 novembre 1918, F-69100 Villeurbanne, France (e-mail: lucas.brivadis@gmail.com, vincent.andrieu@gmail.com, ulysse.serres@gmail.com)

February 5, 2020

Abstract

In this paper, we consider the problem of designing an asymptotic observer for a nonlinear dynamical system in discrete-time following Luenberger's original idea. This approach is a two-step design procedure. In a first step, the problem is to estimate a function of the state. The state estimation is obtained by inverting this mapping. Similarly to the continuous-time context, we show that the first step is always possible provided a linear and stable discrete-time system fed by the output is introduced. Based on a weak observability assumption, it is shown that picking the dimension of the stable auxiliary system sufficiently large, the estimated function of the state is invertible. This approach is illustrated on linear systems with polynomial output. The link with the Luenberger observer obtained in the continuous-time case is also investigated.

1 Introduction

1.1 Context

The design of observers for nonlinear discrete-time systems remains a challenging and open problem despite a burgeoning literature. Since no universal method exists, several approaches have been developed. Most of them have first been developed for continuous-time systems, and then extended to the discrete case. Some of them, such as the well-known *extended Kalman filter* ([4, 11]), provide only a local convergence of the observer, and are based on a linearization of the system. Others (as [5] or [7]) consist in applying an invertible change of coordinates that transforms the original system in an other form for which it is much more easier to design an observer. Still others deal with Lipschitz nonlinear systems ([13, 12], among others), that occur frequently in practice, and are based on linear matrix inequalities that provide Lyapunov functions for the error system.

A completely different idea is to try to reproduce the Luenberger's initial methodology originally developed for linear continuous-time system in [9], which differs from what is now usually called *Luenberger observer*. This path has been mapped in the case of discrete-time systems by N. Kazantzis and C. Kravaris in [8]. It consists to estimate first a function of the state, thanks to a linear stable system fed by the output, and then to inverse this mapping. However, strong assumptions such as analyticity of the system and observablity of the linearized system are required, and the invertibility of the function is obtained only locally.

In the following, we relax those assumptions following the strategy developed in the continuous case in [2] and later in [1] and [3]. We require the system to be time reversible, and

replace the observability hypothesis of the linearized system by a backward distinguishability hypothesis on the nonlinear system itself. In so doing, we obtain the existence and the injectivity (not only locally) of such a function of the state.

This paper is organized as follows. In the next part of the introduction (Section 1.2), we state our problem in a more precise way and introduce some notations and definitions. We also prove a first result that guarantees the existence of an observer as soon as there exists a continuous uniformly injective map satisfying some functional equation. Our main results can be found in Section 2. We state sufficient conditions for the existence, injectivity and also unicity of such a map. We provide in Section 3 some examples and applications of those results. we examine linear systems with polynomial output and also discrete-time systems that approximate continuous-time systems.

Throughout the paper, we denote by $|\cdot|$ the usual Euclidean norm and by $\|\cdot\|$ the induced matrix norm.

1.2 Problem statement

We consider the discrete-time system

$$x_{k+1} = f(x_k), \qquad y_k = h(x_k), \tag{1}$$

with state $x \in \mathbb{R}^n$, output $y \in \mathbb{R}^p$ and suitable functions f and h. In this paper, we deal with the problem of existence of an observer for system (1). We denote $X_k(x_0) = f^k(x_0)$ the value at time k of the unique solution of system (1) initialized at $x_0 \in \mathbb{R}^n$, and $Y_k(x_0) = h(X_k(x_0))$ the corresponding output. Let $\mathcal{X}_0 \subset \mathcal{X} \subset \mathbb{R}^n$ such that for all initial condition $x_0 \in \mathcal{X}_0$ and all $k \in \mathbb{N} \cup \{0\}$, $X_k(x_0) \in \mathcal{X}$.

Definition 1. *Let m be a positive integer, $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ and $\psi : \mathbb{R}^m \to \mathbb{R}^n$. The discrete-time dynamical system given by*

$$\xi_{k+1} = \varphi(\xi_k, y_k), \qquad \hat{x}_k = \psi(\xi_k), \tag{2}$$

is called an observer for (1) *if and only if, for all $(x_0, \xi_0) \in \mathcal{X}_0 \times \mathbb{R}^m$, the solution of the coupled system* (1)-(2)*, denoted by $(X_k(x_0), \hat{X}_k(x_0, \xi_0))_{k \geqslant 0}$, satisfies*

$$\lim_{k \to +\infty} \left| X_k(x_0) - \hat{X}_k(x_0, \xi_0) \right| = 0. \tag{3}$$

Note that, even if $\hat{X}_k$ seems to depend directly of x_0, it is actually not the case. As (2) says, $\hat{X}_k$ depends only of the measurements $Y_0(x_0), Y_1(x_0), \ldots, Y_{k-1}(x_0)$ through the dynamic of $(\xi_k)_{k \geqslant 0}$.

We follow the Luenberger-like methodology in order to design an observer for system (1). Let m be a positive integer. First, we try to transform (1) into

$$\xi_{k+1} = A\xi_k + By_k. \tag{4}$$

with $A \in \mathbb{R}^{m \times m}$ a matrix with spectral radius $\rho(A) < 1$ and $B \in \mathbb{R}^{m \times p}$. In order to do this, we look for a continuous map $T : \mathcal{X} \to \mathbb{R}^m$ such that, for any $x_0 \in \mathcal{X}_0$ and any $k \in \mathbb{N} \cup \{0\}$,

$$T(X_{k+1}(x_0)) = AT(X_k(x_0)) + BY_k(x_0). \tag{5}$$

Let $\Xi_k(x_0, \xi_0)$ denote the value at time k of the unique solution of system (4) with initial condition $\xi_0 \in \mathbb{R}^m$ and measurements $y_k = Y_k(x_0)$. Note that, for any $(x_0, \xi_0) \in \mathcal{X}_0 \times \mathbb{R}^m$,

$$\Xi_{k+1}(x_0, \xi_0) - T(X_{k+1}(x_0)) = A(\Xi_k(x_0, \xi_0) - T(X_k(x_0))) \tag{6}$$

and since $\rho(A) < 1$, $\Xi_k(x_0, \xi_0) - T(X_k(x_0))$ converges geometrically towards zero. Hence, implementing system (4), one can deduce an approximation of $T(x_k)$ as k goes to infinity. Then, if T is injective, one can estimate the state of system (1). More precisely, we have the following theorem.

Theorem 1. *Let m be a positive integer, $A \in \mathbb{R}^{m\times m}$ such that $\rho(A) < 1$ and $B \in \mathbb{R}^{m\times p}$. Let $T : \mathcal{X} \to \mathbb{R}^m$ be a continuous map. Assume the following:*

1. *For all $x \in \mathcal{X}$, T satisfies*
$$T(f(x)) = AT(x) + Bh(x). \tag{7}$$
2. *T is uniformly injective, that is, there exists α a class $\mathcal{K}^\infty$ function such that for all $(x_1, x_2) \in \mathcal{X}^2$,*
$$|x_1 - x_2| \leqslant \alpha(|T(x_1) - T(x_2)|. \tag{8}$$

Then there exists a map $T^* : \mathbb{R}^m \to \mathbb{R}^n$ such that $(\hat{X}_k)_{k\geqslant 0}$ defined by $\hat{X}_k(x_0, \xi_0) = T^*(\Xi_k(x_0, \xi_0))$ for all $(x_0, \xi_0) \in \mathcal{X}_0 \times \mathbb{R}^m$ is the solution of an observer for (1)*.*

Proof. Clearly, (7) implies that (5) is satisfied for all $x_0 \in \mathcal{X}_0$ and all $k \in \mathbb{N} \cup \{0\}$. Let $(x_0, \xi_0) \in \mathcal{X}_0 \times \mathbb{R}^m$. Since $\rho(A) < 1$, it follows from (6) that

$$\lim_{k\to+\infty} \Xi_k(x_0, \xi_0) - T(X_k(\xi_0)) = 0. \tag{9}$$

From the uniform injectivity of T, there exists a pseudo-inverse $T^{-1} : T(\mathcal{X}) \to \mathbb{R}^n$ such that for all x in $\mathcal{X}$ $T^{-1}(T(x)) = x$ and for all $(\xi_1, \xi_2) \in T(\mathcal{X})^2$,

$$|T^{-1}(\xi_1) - T^{-1}(\xi_2)| \leqslant \alpha(|\xi_1 - \xi_2|). \tag{10}$$

According to [10, Theorem 2], there exists a function $T^* : \mathbb{R}^m \to \mathbb{R}^n$, that is an extension to $\mathbb{R}^m$ of T^{-1}, satisfying (10) for all $(\xi_1, \xi_2) \in (\mathbb{R}^m)^2$. Hence,

$$|T^*(\xi) - x| \leqslant \alpha(|\xi - T(x)|) \quad \forall \xi \in \mathbb{R}^m,\ \forall x \in \mathcal{X}. \tag{11}$$

Thus $|T^*(\Xi_k(x_0, \xi_0)) - X_k(\xi_0)| \to 0$ as k goes to infinity. Setting $\varphi : (\xi, y) \in \mathbb{R}^n \times \mathbb{R}^p \mapsto A\xi + By$ and $\psi = T^*$, it follows from the Definition 1 that $(\hat{X}_k)_{k\geqslant 0}$ defined by $\hat{X}_k(x_0, \xi_0) = T^*(\Xi_k(x_0, \xi_0))$ is the solution of an observer for (1). □

Then it is sufficient to prove the existence of a uniformly injective continuous map $T : \mathcal{X} \mapsto \mathbb{R}^m$ satisfying (7) for some positive integer m in order to design an observer for (1). In the next section, we state sufficient conditions for the existence, injectivity, and also unicity of a continuous map T solution of (7).

Remark 1. *Note that if $\mathcal{X}$ is a compact subset of $\mathbb{R}^n$, then every continuous injective map $T : \mathcal{X} \to \mathbb{R}^m$ is also uniformly injective in the sense of* (8)*. In the following, we are interested in the injectivity of T. If uniform injectivity is required (for example to apply Theorem 1), then one must either assume $\mathcal{X}$ compact or prove the uniform injectivity by other means.*

2 Results and comments

2.1 Existence of the transformation

First, we are interested in the existence of a map T satisfying (7). In [2], V. Andrieu and L. Praly have proved the existence of a so-called Kazantzis–Kravaris/Luenberger observer for continuous-time systems of the form

$$\dot{x} = f(x), \qquad y = h(x). \tag{12}$$

We follow the same methodology and adapt it in the discrete case. We need to make some assumptions on the system.

Assumption 1. *f is invertible and f^{-1} and h are continuous.*

Assumption 2. *There exist four non-negative constants C_1, C_2, C_1' and C_2' such that, for all $x \in \mathbb{R}^n$,*

$$|x| \leqslant C_1 + C_2|f(x)|, \qquad |h(x)| \leqslant C_1' + C_2'|x|. \tag{13}$$

Remark 2. *Note that Assumptions 1 and 2 are satisfied in particular if f is invertible and both f^{-1} and h are globally Lipschitz. We will use this remark in the next section about the injectivity of T.*

For all non-negative integer i, we denote $\circ$ the composition operator and

$$f^i = \underbrace{f \circ f \circ \cdots \circ f}_{i \text{ times}}, \qquad f^{-i} = (f^{-1})^i.$$

Theorem 2. *Let m be a positive integer, $A \in \mathbb{R}^{m\times m}$ a normal matrix such that $\rho(A) < \min\{1, 1/C_2\}$ and $B \in \mathbb{R}^{m\times p}$. Assume that Assumptions 1 and 2 are satisfied. For all $x \in \mathcal{X}$, set*

$$T(x) = \sum_{i=0}^{+\infty} A^i B h(f^{-(i+1)}(x)). \tag{14}$$

Then $T : \mathcal{X} \to \mathbb{R}^m$ is well defined, continuous, and satisfies (7).

Proof. For all $x \in \mathcal{X}$ and all non-negative integer i, let $a_i(x) = A^i B h(f^{-(i+1)}(x))$. According to Assumption 1, each a_i is continuous on $\mathcal{X}$. Note that, since A is normal, $\rho(A) = \|A\|$. Then, according to Assumption 2, we have for all $x \in \mathcal{X}$

$$|a_i(x)| \leqslant \rho(A)^i \|B\| \left(C_1' + C_2' \left(C_2^{i+1}|x| + C_1 \sum_{j=0}^{i} C_2^j \right) \right). \tag{15}$$

Since $\rho(A) < 1$ and $\rho(A)C_2 < 1$ the Lebesgue dominated convergence theorem applied on any compact set implies that (14) defines a continuous function. Moreover, for any $x \in \mathcal{X}$,

$$\begin{aligned} T(f(x)) &= \sum_{i=0}^{+\infty} A^i B h(f^{-(i+1)}(f(x))) \\ &= A \sum_{i=0}^{+\infty} A^{i-1} B h(f^{-i}(x)) \\ &= A \sum_{i=0}^{+\infty} A^i B h(f^{-(i+1)}(x)) + Bh(x) \\ &= AT(x) + Bh(x), \end{aligned}$$

which shows that T satisfies (7). □

2.2 Injectivity with backward distinguishability

In order to obtain that T defined by (14) is injective, we introduce the following *backward distinguishability* assumption on the system.

Assumption 3. *For all $(x_1, x_2) \in \mathcal{X}^2$, if $x_1 \neq x_2$, then there exists a positive integer i such that $h(f^{-i}(x_1)) \neq h(f^{-i}(x_2))$.*

We also need stronger hypothesis on the system than in the previous section.

Assumption 4. *f is invertible and f^{-1} and h are of class C^1 and globally Lipschitz.*

According to the Remark 2, if Assumption 4 holds, then Assumptions 1 and 2 are satisfied. We denote by I_k the identity $k \times k$ matrix, by $\otimes$ the Kronecker product and by A^* the conjugate transpose matrix of A.

Theorem 3. *Let Assumptions 3 and 4 hold. Let $m = (n+1)p$ and $B = (1, \dots, 1)^* \otimes I_p \in \mathbb{C}^{m \times p}$. Let $C_2 = \sup\{|(f^{-1})'(x)|, x \in \mathcal{X}\}$ and $\mathcal{D}$ be the open disc of $\mathbb{C}$ of radius $\min\{1, 1/C_2\}$. Then there exists a subset $\mathcal{R} \subset \mathcal{D}^{n+1}$ of zero Lebesgue measure in $\mathbb{C}^{n+1}$ such that, for any $(\lambda_1, \dots \lambda_{n+1}) \in \mathcal{D}^{n+1} \setminus \mathcal{R}$, the matrix $A = \operatorname{diag}(\lambda_1, \dots, \lambda_{n+1}) \otimes I_p \in \mathbb{C}^{m \times m}$ is such that the map $T : \mathcal{X} \to \mathbb{C}^m$ defined by (14) is well-defined, of class C^1 and one-to-one.*

Proof. Let $(\lambda_1, \dots \lambda_{n+1}) \in \mathcal{D}^{n+1}$ and $A = \operatorname{diag}(\lambda_1, \dots, \lambda_{n+1}) \otimes I_p \in \mathbb{C}^{m \times m}$. Let $T : \mathcal{X} \to \mathbb{C}^m$ be defined as in (14). For all $\lambda \in \mathcal{D}$, let

$$T_\lambda(x) = \sum_{i=0}^{+\infty} \lambda^i h(f^{-(i+1)}(x)), \quad \forall x \in \mathcal{X}. \tag{16}$$

Let $a_i(x) = \lambda^i h(f^{-(i+1)}(x))$ for all $x \in \mathcal{X}$. Then each a_i is of class C^1 on $\mathcal{X}$ by Assumption 4, and we have the following domination:

$$\left|a_i'(x)\right| \leqslant \lambda^i C_2' C_2^{i+1}$$

with $C_2' = \sup\{|h'(x)|, x \in \mathcal{X}\}$. Moreover, $\lambda C_2 < 1$. So the Lebesgue dominated convergence theorem implies that for each $\lambda \in \mathcal{D}$, $T_\lambda : \mathcal{X} \to \mathbb{C}^p$ is well-defined and of class C^1. Considering the structure of A and B, remark that up to a permutation of coordinates we have

$$T(x) = \left(T_{\lambda_1}(x), \dots, T_{\lambda_{n+1}}(x)\right)^*$$

It is sufficient to prove that $T : \mathcal{X} \to \mathbb{C}^p$ is one-to-one for almost all $(\lambda_1, \dots, \lambda_{n+1}) \in \mathcal{D}^{n+1}$.

In order to do this, we need the following lemma, established by L. Praly and V. Andrieu in [2, Lemma 1], which is a modified version of [6, Lemma 3.2] due to J.-M. Coron.

Lemma 1. *Let $\mathcal{D}$ and Γ be open subsets of $\mathbb{C}$ and $\mathbb{R}^{2n}$, respectively. Let $g : \Gamma \times \mathcal{D} \to \mathbb{C}^p$ be a function which is holomorphic in λ for each $\underline{x} \in \Gamma$ and C^1 in $\underline{x}$ for each $\lambda \in \mathcal{D}$. If for each $\underline{x} \in \Gamma$, the function $\lambda \in \mathcal{D} \mapsto g(\underline{x}, \lambda)$ is not constantly zero, then the set*

$$\mathcal{R} = \bigcup_{\underline{x} \in \Gamma} \left\{ (\lambda_1, \dots, \lambda_{n+1}) \in \mathcal{D}^{n+1} \;\middle|\; \forall i \in \{1, \dots, n+1\},\ g(\underline{x}, \lambda_i) = 0 \right\} \tag{17}$$

has zero Lebesgue measure in $\mathbb{C}^{n+1}$.

We apply this lemma to $\Gamma = \{(x_1, x_2) \in \mathcal{X}^2 \mid x_1 \neq x_2\}$ and $g = \Delta T$ defined as follows:

$$\Delta T : (x_1, x_2, \lambda) \in \mathcal{X}^2 \times \mathcal{D} \mapsto T_\lambda(x_1) - T_\lambda(x_2) \tag{18}$$

Clearly, $\Delta T(x_1, x_2, \cdot)$ is holomorphic on $\mathcal{D}$ for each $(x_1, x_2) \in \mathcal{X}^2$ and $\Delta T(\cdot, \lambda)$ is of class C^1 on $\mathcal{X}^2$ for each $\lambda \in \mathcal{D}$. Fix $(x_1, x_2) \in \Gamma$. Now, we prove that $\Delta T(x_1, x_2, \cdot)$ is not identically zero on $\mathcal{D}$. Assume the contrary. By unicity of the power series expansion, we get that for all positive integer i,

$$h(f^{-i}(x_1)) = h(f^{-i}(x_2)) \tag{19}$$

According to the *backward distinguishability* Assumption 3, it implies that $x_1 = x_2$ which is contradictory with the fact that $(x_1, x_2) \in \Gamma$. Hence, $\Delta T(x_1, x_2, \cdot)$ is not identically zero on $\mathcal{D}$.

Since $\mathcal{D}$ is a convex subset of $\mathbb{C}$ and $\Delta T(x_1, x_2, \cdot)$ is holomorphic, its zero are isolated and with finite multiplicity. Hence the hypotheses of Lemma 1 are satisfied. Thus, $\mathcal{R} \subset \mathcal{D}^{n+1}$ has zero Lebesgue measure and for all $(\lambda_1, \ldots, \lambda_{n+1}) \in \mathcal{D}^{n+1} \setminus \mathcal{R}$, T is injective by definition of ΔT. □

Remark 3. *The function T and the matrices A and B defined Theorem 3 take complex values while previous Theorems 1 and 2 remain in the real frame. However, one can choose two different ways to bridge this gap.*

- *State Theorems 1 and 2 in the complex frame. The proofs remain identical. One should simply change the domains and codomains of f and h.*
- *Instead of considering $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n+1}) \otimes I_p \in \mathbb{C}^{m \times m}$ and $B = (1, \ldots, 1)^* \otimes I_p \in \mathbb{C}^{m \times p}$, one should either consider $\tilde{A} = \mathrm{diag}(\Lambda_1, \ldots, \Lambda_{n+1}) \otimes I_p \in \mathbb{R}^{2m \times 2m}$ and $\tilde{B} = (\mathbb{I}, \ldots, \mathbb{I})^* \otimes I_p \in \mathbb{R}^{2m \times p}$, where*

$$\Lambda_i = \begin{pmatrix} \Re(\lambda_i) & -\Im(\lambda_i) \\ \Im(\lambda_i) & \Re(\lambda_i) \end{pmatrix}, \qquad \mathbb{I} = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

  Then for all real sequence of measurements $(y_k)_{k \geqslant 0}$ the solutions of $\tilde{\xi}_{k+1} = \tilde{A}\tilde{\xi}_k + \tilde{B}y_k$ contain the real and imaginary parts of the solutions of $\xi_{k+1} = A\xi_k + By_k$.

2.3 Unicity

One can also wonder in which cases does the unicity of T satisfying (7) holds. More than a theoretical question, this fact may be useful in practice in order to obtain the injectivity of T. Most of the time, the function T given by (14) is difficult to compute. Since the matrix A has spectral radius strictly inferior to 1, an approximation of T is given by

$$T_N(x) = \sum_{i=0}^{N} A^i B h(f^{-(i+1)}(x)), \quad \forall x \in \mathcal{X}. \tag{20}$$

for all $N \geqslant 0$. Then $|T(x) - T_n(x)| \to 0$ as $N \to +\infty$. However, if f and h have more properties (for example if f is linear and h is polynomial, see Section 3.1), there may exist another solution $\tilde{T}$ of (7) much more easier to compute than T. Then, the question of the injectivity of that new $\tilde{T}$ remains open *a priori*. But if (7) has a unique solution for A and B complex matrices chosen has in Theorem 3, then $T = \tilde{T}$ and hence $\tilde{T}$ is injective. Now, we state our unicity theorem.

Theorem 4. *Let m be a positive integer, $A \in \mathbb{R}^{m\times m}$ such that $\rho(A) < 1$ and $B \in \mathbb{R}^{m\times p}$. Let Assumption 1 hold and make the following backward stability hypothesis on $\mathcal{X}$:*

$$\forall x \in \mathcal{X},\ \forall i \geqslant 1, \quad f^{-i}(x) \in \mathcal{X}. \tag{21}$$

Assume also that $\mathcal{X}$ is compact. Then there exists one and only one continuous function $T : \mathcal{X} \to \mathbb{R}^m$ that satisfy (7) for all $x \in \mathcal{X}$.

Proof. First, we prove that the continuous solution of (7) is unique. Let $T_1,\ T_2 : \mathcal{X} \to \mathbb{R}^m$ be two continuous solutions of (7). Let $x \in \mathcal{X}$. Then for all $i \in \mathbb{N} \cup \{0\}$,

$$\begin{aligned} T_1(x) - T_2(x) &= (T_1 - T_2)(f^i(f^{-i}(x))) \\ &= A^i(T_1 - T_2)(f^{-i}(x)). \qquad \text{(from (7))} \end{aligned}$$

Since $\mathcal{X}$ is compact, satisfy (21) and T_1 and T_2 are continuous, there exists a constant $K > 0$ such that $|(T_1 - T_2)(f^{-i}(x))| \leqslant K$ for all $i \in \mathbb{N} \cup \{0\}$. Since moreover $\rho(A) < 1$, $A^i(T_1 - T_2)(f^{-i}(x)) \to 0$ as $i \to +\infty$. Thus $T_1(x) - T_2(x) = 0$.

The existence of a continuous T satisfying (7) follows from the Theorem 2 and from the fact that Assumption 2 can be replaced in its proof by the fact that $\mathcal{X}$ is compact and backward stable[1]. Indeed, the series (14) still defines a continuous function since the domination

$$|a_i(x)| \leqslant \rho(A)^i ||B|| \sup_{\tilde{x}\in\mathcal{X}} h(\tilde{x}) \tag{22}$$

holds for all $x \in \mathcal{X}$ and can replace (15). Then one may apply the Lebesgue dominated convergence on $\mathcal{X}$. □

To conclude this section, recall that we have now at our disposal three theorems that ensures under different conditions on (1) the existence, unicity and injectivity of a continuous map T satisfying (7). In the next section, we illustrate on examples how to use those tools. In particular, we study systems with linear dynamics and polynomial output, and emphasize the link between the Luenberger observers developed in [2] for continuous-time systems and the discrete-time observers developed in this paper for theirs first-order approximations.

3 Examples

3.1 Linear dynamics with polynomial output

We consider first the system with linear dynamic and polynomial output of degree d

$$x_{k+1} = Fx_k, \quad y_k = HP_d(x) \tag{23}$$

with $P_d : \mathbb{R}^n \to \mathbb{R}^{k_d}$ a vector containing the k_d possible monomials with degree less or equal than d, $F \in \mathbb{R}^{n\times n}$ and $H \in \mathbb{R}^{p\times k_d}$. Then we have the following proposition.

Proposition 1. *Let m be a positive integer and $B \in \mathbb{R}^{m\times p}$. There exists a subset $\mathcal{S}$ of zero Lebesgue measure in $\mathbb{R}^{m\times m}$ such that for all $A \in \mathbb{R}^{m\times m}\backslash\mathcal{S}$, there exists a function $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ of the form*

$$T(x) = MP_d(x), \qquad \forall x \in \mathbb{R}^n \tag{24}$$

for some $M \in \mathbb{R}^{m\times k_d}$, that satisfies (7) for any $x \in \mathbb{R}^n$.

[1] Similarly, using the same trick, one can easily show that the hypothesis of *globally Lipschitz* in Assumption 4 can be replaced in the proof of Theorem 3 by the fact that $\mathcal{X}$ is compact and backward stable.

Proof. First, note that since $P_d(Fx)$ is a vector containing polynomials of x with degree inferior to d, there exists a matrix $D \in \mathbb{R}^{k_d \times k_d}$ such that

$$P_d(Fx) = DP_d(x), \qquad \forall x \in \mathbb{R}^n. \tag{25}$$

Since the set of eigenvalues of D is finite, the spectra of D and $-A$ are disjoint for almost all $A \in \mathbb{R}^{m\times m}$ *i.e.* there exists a subset $\mathcal{S} \subset \mathbb{R}^{m\times m}$ of zero Lebesgue measure such that the spectra of D and $-A$ are disjoint for all $A \in \mathbb{R}^{m\times m} \setminus \mathcal{S}$. For such matrices A the Sylvester equation

$$MD = AM + BH \tag{26}$$

has a unique solution $M \in \mathbb{R}^{m\times k_d}$. Set T as in (24). It remains to check that (7) is satisfied for $f = F$ and $h = HP_d$. For all $x \in \mathbb{R}^n$,

$$\begin{aligned} T(Fx) &= MP_d(Fx) && \text{(from (24))} \\ &= MDP_d(x) && \text{(from (25))} \\ &= AMP_d(x) + BHP_d(x) && \text{(from (26))} \\ &= AT(x) + BHP_d(x). \end{aligned}$$

□

Remark 4. *Note that the result is still true if A and B are complex matrices. Then T takes complex values. The proof remains identical.*

Remark 5. *Choose a set $\mathcal{X}_0 \subset \mathbb{R}^n$ of initial condition and let $\mathcal{X}$ be as usual such that $X_k(x_0) \in \mathcal{X}$ for all $x_0 \in \mathcal{X}_0$ and all $k \in \mathbb{N} \cup \{0\}$. Note that if F is invertible and if $\mathcal{X}$ is compact and backward stable, then the assumptions of the Theorem 4 hold. Assume also that Assumptions 3 and 4 hold and apply Theorem 3 with $m = (n+1)p$. Then, for almost all $(\lambda_1, \dots, \lambda_{n+1}) \in \mathbb{C}^{n+1}$, and for complex matrices A and B as in Theorem 3, we have*

$$T(x) = MP_d(x) = \sum_{i=0}^{+\infty} A^i Bh(f^{-(i+1)}(x)) \tag{27}$$

for all $x \in \mathcal{X}$. In particular, T defined by (24) is injective.

3.2 Link with the continuous Luenberger observer

In this section, we are interested in the link between the continuous Luenberger observer developed in [2] for system (12) and the discrete observer developed in the previous sections for a discrete-time version of (12).

3.2.1 Continuous-time system

We consider the following example with linear dynamic and polynomial output:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_1 \end{cases}, \quad y = x_1^2 - x_2^2 + x_1 + x_2. \tag{28}$$

It can be shown that this system is weakly differentially observable[2] of order 4 on $\mathbb{R}^2$ in the sense of [3, Definition 1]. Following [3], we seek $T_\lambda : \mathbb{R} \to \mathbb{R}^n$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t} T_\lambda(x) = \lambda T_\lambda(x) + y \tag{29}$$

[2]First, the map $(x_1^2 - x_2^2, x_1 + x_2) \mapsto (y, \ddot{y})$ is injective. Similarly, $(x_1 x_2, x_1 - x_2) \mapsto (\dot{y}, \dddot{y})$ is also injective. Combining those results, we get that $(x_1, x_2) \mapsto (y, \dot{y}, \ddot{y}, \dddot{y})$ is injective.

for some $\lambda < 0$. Since (28) has linear dynamic and polynomial output of degree 2, one can look for T of the form

$$T_\lambda(x) = x^* \begin{pmatrix} a & c/2 \\ c/2 & b \end{pmatrix} x + \begin{pmatrix} d & e \end{pmatrix} x \tag{30}$$

for some $(a, b, c, d, e) \in \mathbb{R}^5$. Then (29) holds if and only if

$$\begin{aligned} -c = \lambda a + 1, \quad c = \lambda b - 1, \quad 2(a-b) = \lambda c, \\ -e = \lambda d + 1, \quad d = \lambda e + 1. \end{aligned} \tag{31}$$

The only solution of this equation is

$$\begin{aligned} a = -\frac{\lambda}{4+\lambda^2}, \quad b = \frac{\lambda}{4+\lambda^2}, \quad c = -\frac{4}{4+\lambda^2}, \\ d = \frac{1-\lambda}{1+\lambda^2}, \quad e = -\frac{1+\lambda}{1+\lambda^2}. \end{aligned} \tag{32}$$

Since T_λ is stationary, one could believe that this function provide an observer that could be efficient even for a numerical approximation of (28). However, as we will see in the following, it is not the case: for a given discrete approximation of (28), it is better to design an observer based on the discrete-time system rather than to use the one given by T_λ.

3.2.2 Associated first-order discrete-time system

For some discretization parameter $dt > 0$, the associated first-order approximation[3] of (28) is

$$\begin{cases} x_1(k+1) = x_1(k) + dt x_2(k) \\ x_2(k+1) = x_2(k) - dt x_1(k) \\ y_k = x_1(k)^2 - x_2(k)^2 + x_1(k) + x_2(k) \end{cases}. \tag{33}$$

We seek a function $T^d_\lambda : \mathbb{R} \to \mathbb{R}^n$ satisfying a first-order approximation of (29) given by the Euler explicit method:

$$T^d_\lambda(x(k+1)) = (1 + \lambda dt) T^d_\lambda(x(k)) + dt y_k. \tag{34}$$

Since $\lambda < 0$, it is sufficient to choose $\lambda dt > -2$ to have $-1 < 1 + \lambda dt < 1$. Now, we seek T^d_λ of the form

$$T^d_\lambda(x) = x^* \begin{pmatrix} a' & c'/2 \\ c'/2 & b' \end{pmatrix} x + \begin{pmatrix} d' & e' \end{pmatrix} x \tag{35}$$

for some $(a', b', c', d', e') \in \mathbb{R}^5$. Then (34) holds if and only if (d', e') satisfy the same equation that (d, e) in (31) and (a', b', c') satisfy

$$\begin{cases} -c' + b' dt = \lambda a' + 1, \\ c' + a' dt = \lambda b' - 1, \\ 2(a' - b') - c' dt = \lambda c'. \end{cases} \tag{36}$$

Remark that this equation is the same than (32) when $dt = 0$. This is coherent with the fact that (33) is a discretization of (28). Then, the only solution of (36) is such that $(d', e') = (d, e)$

[3] Since (28) is weakly differentially observable, it can be shown that (33) is backward distinguishable as soon as dt is small enough.

for all $dt > 0$ and (a', b', c') converges to (a, b, c) as dt goes to 0:

$$\begin{cases} a' = -\dfrac{\lambda + dt}{4 + (\lambda + dt)^2}, \\ b' = \quad \dfrac{\lambda + dt}{4 + (\lambda + dt)^2}, \\ c' = -\dfrac{4}{4 + (\lambda + dt)^2}. \end{cases} \tag{37}$$

For $dt > 0$, the discrete observer given by T_λ^d is therefore different from the continuous observer given by T_λ, even if their difference goes to 0 as dt goes to 0.

3.2.3 Comparison of the observers

Consider a numerical simulation of the continuous-time system (28) obtained by the Euler explicit first-order method, which corresponds to the discrete-time system (33). Then the map T_λ^d given by (34) is much more adapted to the design of a numerically efficient observer than the function T_λ given by (29) that has been designed for (28). More generally, in order to implement an observer for a continuous-time varying system, it is better to develop a discrete-time observer based on the numerical approximation of the system, rather than a continuous-time observer based on the original system itself.

In order to highlight numerically this fact, we simulate the system (28) thanks to (33) and compare the accuracy of two observers: one based on functions of the form T_λ^d, and another based on functions of the form T_λ. To obtain the observers, we fix $dt > 0$ and three arbitrary values $\lambda_i < 0$ satisfying $\lambda_i dt > -2$ and use the fact that

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ a_1 & c_1 & d_1 & e_1 \\ a_2 & c_2 & d_2 & e_2 \\ a_3 & c_3 & d_3 & e_3 \end{pmatrix} \begin{pmatrix} x_1^2 - x_2^2 \\ x_1 x_2 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y \\ T_1(x) \\ T_2(x) \\ T_3(x) \end{pmatrix}. \tag{38}$$

where (a_i, c_i, d_i, e_i) is given by (32) (resp. (37)) with $\lambda = \lambda_i$ and $T_i = T_{\lambda_i}$ (resp. $T_i = T_{\lambda_i}^d$). Fix the following parameters and initial conditions:

$$dt = 0.01,\ x(0) = (1, 0),\ \lambda_i = -10 \times i,\ \xi^i(0) = 0. \tag{39}$$

Then the 4×4 matrix defined in (38) is invertible. Hence one can reconstruct an approximation $(\hat{x}_1, \hat{x}_2)$ of the state (x_1, x_2) from the measurement y and approximations of $T_i(x)$ given by the dynamic $\xi_{k+1}^i = (1 + \lambda_i dt)\xi_k^i + dt y_k$.

On Fig. 1, we plot on a semi-log scale the evolution of the absolute error $\varepsilon_k = |x_k - \hat{x}_k|$ between the state and its observer for $k \in \{0, \ldots, 500\}$ (*i.e.* $t \in [0, 5]$) for the observer based on functions T_{λ_i} designed for the original continuous-time system. Similarly, we make on Fig. 2 the same plot but for the observer based on functions $T_{\lambda_i}^d$ designed for the discrete-time system. We clearly see that the observer based on $T_{\lambda_i}^d$ is much more efficient than the one based on T_{λ_i}. On one hand, using $T_{\lambda_i}^d$, the error go to zero until it achieve 10^{-12}, which is close to the machine epsilon ($\approx 10^{-16}$). Moreover, the state observer seems to converge exponentially to the state, with a rate $r \approx -4.58$ (estimation based on a linear regression made on $[0.5, 3]$). On the other hand, with T_{λ_i}, the observer does not converge to the state: it keeps an absolute error oscillating around 10^{-2}. This phenomenon is due to the fact that the trajectory of (33) is not invariant for this observer: even if it is well initialized (*i.e.* $x(0) = \hat{x}(0)$), the observer will oscillate around the state.

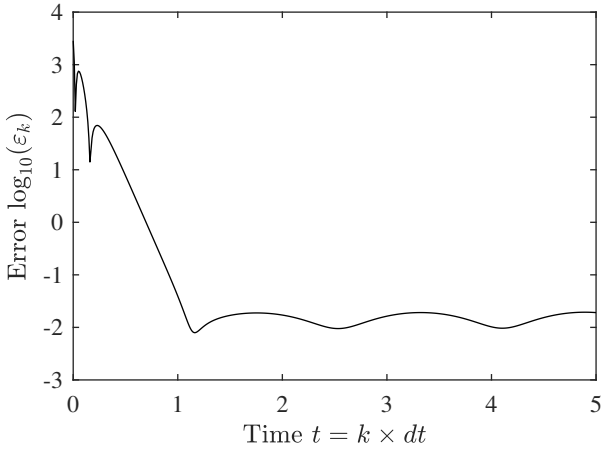


Figure 1: Evolution of the error between the state and the observer based on T_{λ_i} in semi-log scale

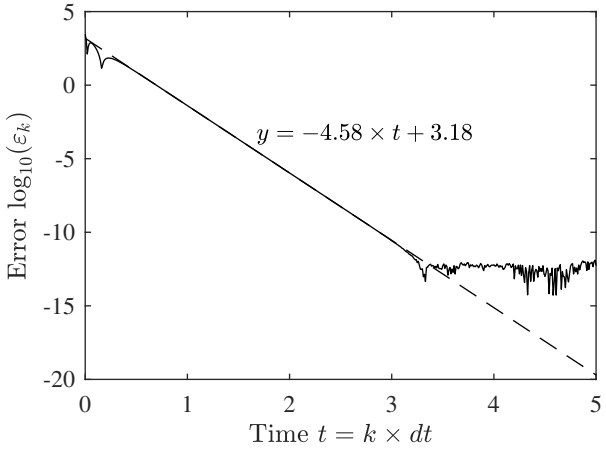


Figure 2: Evolution of the error between the state and the observer based on $T^d_{\lambda_i}$ in semi-log scale

4 Conclusion

We have shown how the initial Luenberger methodology can be applied to nonlinear discrete-time systems. It is based on the existence of a map satisfying some functional equation linked to the system, that transform the original system into a linear asymptotically stable one fed by the output. As soon as this map is uniformly injective, it allows us to estimate the state of the nonlinear system by simulating an autonomous system fed by the output and inverting this map. We stated sufficient conditions for the existence of such a map. In particular, we need the system to be reversible in time. Under a backward distinguishability hypothesis, we also proved that this map is injective.

References

[1] V. Andrieu. Convergence speed of nonlinear luenberger observers. *SIAM Journal on Control and Optimization*, 52(5):2831–2856, 2014.

[2] Vincent Andrieu and Laurent Praly. On the existence of a kazantzis–kravaris/luenberger observer. *SIAM J. Control and Optimization*, 45:432–456, 02 2006.

[3] Pauline Bernard and Vincent Andrieu. Luenberger observers for nonautonomous nonlinear systems. *IEEE Transactions on Automatic Control*, PP:1–1, 09 2018.

[4] M. Boutayeb and D. Aubry. A strong tracking extended kalman observer for nonlinear discrete-time systems. *IEEE Transactions on Automatic Control*, 44(8):1550–1556, Aug 1999.

[5] C. Califano, S. Monaco, and D. Normand-Cyrot. On the observer design in discrete-time. *Systems & Control Letters*, 49(4):255 – 265, 2003.

[6] Jean-Michel Coron. On the stabilization of controllable and observable systems by an output feedback law. *Mathematics of Control, Signals and Systems*, 7(3):187–216, 1994.

[7] Henri Huijberts. On existence of extended observers for nonlinear discrete-time systems. *Lecture Notes in Control and Information Sciences*, 244, 03 1999.

[8] Nikolaos Kazantzis and Costas Kravaris. Discrete-time nonlinear observer design using functional equations. *Systems & Control Letters*, 42(2):81 – 94, 2001.

[9] D. G. Luenberger. Observing the state of a linear system. *IEEE Transactions on Military Electronics*, 8(2):74–80, April 1964.

[10] E. J. McShane. Extension of range of functions. *Bull. Amer. Math. Soc.*, 40(12):837–842, 12 1934.

[11] K Reif and R Unbehauen. The extended kalman filter as an exponential observer for nonlinear systems. *Signal Processing, IEEE Transactions on*, 47:2324 – 2328, 09 1999.

[12] A. Zemouche and M. Boutayeb. Observer design for lipschitz nonlinear systems: The discrete-time case. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 53(8):777–781, Aug 2006.

[13] A Zemouche and M Boutayeb. Observers design for discrete-time lipschitz nonlinear systems. state of the art and new results. *Proceedings of the IEEE Conference on Decision and Control*, pages 4780–4785, 12 2012.

Appendix D

Article on weakly contractive systems

From local to global asymptotic stabilizability for weakly contractive control systems

L. Brivadis[1], L. Sacchelli[2], V. Andrieu[1], J.-P. Gauthier[3], and U. Serres[1]

[1]Univ. Lyon, Université Claude Bernard Lyon 1, CNRS, LAGEPP UMR 5007, 43 bd du 11 novembre 1918, F-69100 Villeurbanne, France
[2]Department of Mathematics, Lehigh University, Bethlehem, PA, USA
[3]Université de Toulon, Aix Marseille Univ, CNRS, LIS, France

Tuesday 25th August, 2020

Abstract

A nonlinear control system is said to be weakly contractive in the control if the flow that it generates is non-expanding (in the sense that the distance between two trajectories is a non-increasing function of time) for some fixed Riemannian metric independent of the control. We prove in this paper that for such systems, local asymptotic stabilizability implies global asymptotic stabilizability by means of a dynamic state feedback. We link this result and the so-called Jurdjevic and Quinn approach.

Keywords: Nonlinear control systems, Feedback stabilization, Asymptotic stability.

1 Main result

1.1 Statement of the result

Consider the following nonlinear continuous-time control system:

$$\dot{x} = f(x,u) = f_u(x), \quad f(0,0) = 0, \tag{1}$$

where x lives in $\mathbb{R}^n$ and u is the control input taking values in an open subset $\mathcal{U}$ of $\mathbb{R}^m$ containing zero. We assume that $f_u \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ for all $u \in \mathcal{U}$, $\frac{\partial f}{\partial x} \in C^0(\mathbb{R}^n \times \mathcal{U}, \mathbb{R}^n)$ and $f(x,\cdot)$ is locally Lipschitz for all $x \in \mathbb{R}^n$.

Definition 1 (Static stabilizability)**.** *System* (1) *is said to be locally (resp. globally) asymptotically stabilizable by a static state feedback if there exists a locally Lipschitz mapping $\lambda : \mathbb{R}^n \to \mathcal{U}$ such that*

$$\dot{x} = f(x, \lambda(x)) \tag{2}$$

is locally (resp. globally) asymptotically stable at the origin.

Local asymptotic stabilizability is usually obtained by investigating first order or homogeneous approximations of the dynamical system around the origin. Yet obtaining global stabilizability from local stabilizability is not an easy task and may fail in general. For autonomous systems, the famous Markus-Yamabe conjecture emphasizes this fact (see, *e.g.* [3, 11, 14, 15]).

However, there are classes of system for which we know how to bridge the gap between local and global asymptotic stabilizability. This is obviously the case if the feedback law λ is such that $x \mapsto f(x, \lambda(x))$ is a linear vector field. More generally, it still holds for homogeneous systems admitting a homogeneous feedback law (see e.g. [8, 16]). Note also that it is shown in [6] that when the locally stabilizing state feedback fails to share the same homogeneity property than the vector field, global (or semi-global) property can still be achieved by a dynamic state feedback.

Definition 2 (Dynamic stabilizability)**.** *System* (1) *is said to be locally (resp. globally) asymptotically stabilizable by a dynamic state feedback if there exist a positive numbber* m*, a map* $\hat{f} : \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{U} \to \mathbb{R}^m$ *such that* $\hat{f}(\cdot, \cdot, u) \in C^1(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^m)$ *for all* $u \in \mathcal{U}$*,* $\frac{\partial \hat{f}}{\partial(x,\hat{x})} \in C^0(\mathbb{R}^n \times \mathbb{R}^n \times \mathcal{U}, \mathbb{R}^n)$ *and* $\hat{f}(x, \hat{x}, \cdot)$ *is locally Lipschitz for all* $(x, \hat{x}) \in \mathbb{R}^n \times \mathbb{R}^m$ *and a locally Lipschitz mapping* $\lambda : \mathbb{R}^m \to \mathcal{U}$ *such that*

$$\dot{x} = f(x, \lambda(\hat{x})), \quad \dot{\hat{x}} = \hat{f}(x, \hat{x}, \lambda(\hat{x})) \tag{3}$$

is locally (resp. globally) asymptotically stable at the origin.

In this paper, we give another class of dynamical systems which share the same property that static local asymptotic stabilizability implies dynamic global asymptotic stabilizability: namely, weakly contractive control systems.

Definition 3 (Weakly contractive)**.** *Let* g *be a* C^1 *Riemannian metric on* $\mathbb{R}^n$*. System* (1) *is said to be weakly contractive with respect to* g *if*

$$\forall u \in \mathcal{U}, \quad L_{f_u} g \leqslant 0, \tag{4}$$

where $L_{f_u} g$ *denotes the Lie derivative*[1] *of the metric* g *with respect to the vector field* f_u*.*

A vector field F over $\mathbb{R}^n$ is usually said to be contractive with respect to a metric g if $L_F g$ is negative. Here we insist on the fact that the vector fields f_u are only *weakly* contractive with respect to the metric g, in the sense that $L_{f_u} g$ is only non-positive.

For all point $x \in \mathbb{R}^n$, and all pair of vectors $(\varphi, \psi) \in \mathbb{R}^n \times \mathbb{R}^n$, denote by $\langle \varphi, \psi \rangle_{g(x)} = g(x)(\varphi, \psi)$ the inner product between the two vectors φ and ψ at the point x for the metric g, and set $|\varphi|_{g(x)} = \sqrt{\langle \varphi, \varphi \rangle_{g(x)}}$.

Recall that associated to the metric g we can define a distance d_g between a pair of points of $\mathbb{R}^n$ in the following way. The length of any piecewise C^1 path $\gamma : [s_1, s_2] \to \mathbb{R}^n$ between two arbitrary points $x_1 = \gamma(s_1)$ and $x_2 = \gamma(s_2)$ in $\mathbb{R}^n$ is defined as:

$$\ell(\gamma) = \int_{s_1}^{s_2} |\gamma'(s)|_{g(\gamma(s))} \mathrm{d}s \tag{5}$$

The distance $d_g(x_1, x_2)$ is defined as the infimum of the length over all such paths. We denote d_g^2 the square of the distance function.

For all point $(x, \hat{x}) \in \mathbb{R}^n \times \mathbb{R}^n$, we denote (*if it exists*) $\nabla_{g(\hat{x})} d_g^2(x, \hat{x})$ the gradient of the function $\hat{x} \mapsto d_g^2(x, \hat{x})$ at the point $\hat{x}$ for the metric g. Fix $x \in \mathbb{R}^n$. Then $\nabla_{g(\hat{x})} d_g^2(x, \hat{x})$ is well-defined at each $\hat{x} \in \mathbb{R}^n$ if and only if, for all $\hat{x} \in \mathbb{R}^n$, there exists a unique length-minimizing

[1] For all point $x \in \mathbb{R}^n$, $L_{f_u} g(x)$ satisfies for all pair of vectors $(\varphi, \psi) \in \mathbb{R}^n \times \mathbb{R}^n$, $L_{f_u} g(x)(\varphi, \psi) = \frac{\partial}{\partial x} \left(\langle \varphi, \psi \rangle_{g(x)} \right)(x)[f_u(x)] + 2\langle \varphi, \frac{\partial f_u}{\partial x}(x)[\psi] \rangle_{g(x)}$. Then, $L_{f_u} g \leqslant 0$ if and only if $L_{f_u} g(x)(\varphi, \varphi) \leqslant 0$ for all point $x \in \mathbb{R}^n$ and all vector $\varphi \in \mathbb{R}^n$.

curve γ joining x to $\hat{x}$, *i.e.* such that $\ell(\gamma) = d_g(x, \hat{x})$. Equivalently, the Riemannian exponential map at the point $\hat{x}$ (denoted by $\exp_{\hat{x}}$) is invertible[2] and we have

$$\nabla_{g(\hat{x})} d_g^2(x, \hat{x}) = -2 \exp_{\hat{x}}^{-1}(x)$$

for all $\hat{x} \in \mathbb{R}^n$, which yields

$$\nabla_{g(\hat{x})} d_g^2(x, \hat{x}) = 0 \quad \text{if and only if} \quad x = \hat{x}. \tag{6}$$

Also, by definition of the Riemannian gradient, for all vectors $\varphi \in \mathbb{R}^n$,

$$\left\langle \nabla_{g(\hat{x})} d_g^2(x, \hat{x}), \varphi \right\rangle_{g(\hat{x})} = \frac{\partial d_g^2}{\partial \hat{x}}(x, \hat{x})[\varphi]. \tag{7}$$

Assume that f is C^1. If (1) is a weakly contractive vector field, then for all C^1 control $u : \mathbb{R}_+ \to \mathcal{U}$ the time-varying vector field f_u generates a non-expanding flow in the sense that, if x_1 and x_2 satisfy $\dot{x}_i = f_u(x_i)$ for $i \in \{1, 2\}$, then the distance $d_g(x_1, x_2)$ between the two trajectories is a non-increasing function of time. We give in appendix a short proof of this well-known statement to be self-contained.

The following theorem is the main result of the paper.

Theorem 4. *Let g be a C^2 complete Riemannian metric on $\mathbb{R}^n$ such that d_g^2 is a C^2 function. Assume that* (1) *is weakly contractive with respect to g, and $f \in C^1(\mathbb{R}^n \times \mathcal{U}, \mathbb{R}^n)$. If* (1) *is locally asymptotically stabilizable by a static state feedback $\lambda \in C^1(\mathbb{R}^n, \mathcal{U})$, then it is also globally asymptotically stabilizable by a dynamic state feedback given by*

$$\dot{x} = f(x, \lambda(\hat{x})), \quad \dot{\hat{x}} = f(\hat{x}, \lambda(\hat{x})) + k(x, \hat{x}) \tag{8}$$

where $\hat{x} \in \mathbb{R}^n$ and

$$k(x, \hat{x}) = -\alpha(x, \hat{x}) \nabla_{g(\hat{x})} d_g^2(x, \hat{x})$$

for some positive locally Lipschitz map α.

1.2 Discussion on the result

The idea of the proof is somehow counter-intuitive. Indeed, the feedback depends only on $\hat{x}$. By selecting α sufficiently small, we make sure that $\hat{x}$ remains in the basin of attraction of the origin for the vector field associated to the state feedback. On the other hand, the correction terms k acting on $\dot{\hat{x}}$ forces x to converge to $\hat{x}$, which implies that x goes to zero.

An interesting aspect of our approach is that no structural constraints is imposed on the local asymptotic stabilizer. This one can be designed for qualitative purposes and can be for instance bounded or optimal as long as this one ensures a local asymptotic stability property. This technique offers another approach to solve the global asymptotic stabilization with local optimal behavior as for instance studied in [2] or [5]. The main difference with these studies being that the local optimal behavior is reproduced asymptotically in time (as x converges to $\hat{x}$).

To construct the feedback law one needs to compute $\nabla_{g(\hat{x})} d_g^2(x, \hat{x})$ which may be difficult to obtain analytically in general (except in some simple cases, *e.g.*, if the metric is constant). Some ways of constructing similar correction terms may be obtained following observer designs based on Riemannian approaches as in [1, 17]. In particular in [17, Lemma 3.6], the authors

[2] see e.g. [4, Chap. 7, Theorem 3.1] for sufficient geometric conditions.

introduced a "distance-like" function δ, that is of crucial importance in the construction of the correction term.

1.3 Proof

Let λ be a C^1 locally asymptotically stabilizing feedback law. Let $\mathcal{D}$ be the basin of attraction of the origin for the vector field $x \mapsto f(x, \lambda(x))$, which is a non-empty open subset of $\mathbb{R}^n$. According to the converse Lyapunov theorem [18, Theorem 1] (based on the previous works of [9, 10, 12]), there exists a proper function $V \in C^\infty(\mathcal{D}, \mathbb{R}_+)$ such that $V(0) = 0$ and

$$\frac{\partial V}{\partial x}(x) f(x, \lambda(x)) \leqslant -V(x), \quad \forall x \in \mathcal{D} \ . \tag{9}$$

For all $r > 0$, set $D(r) = \{x \in \mathbb{R}^n \mid V(x) \leqslant r\}$ which is a compact subset of $\mathcal{D}$. Let $\alpha : \mathbb{R}^n \times \mathcal{D} \to \mathbb{R}_+$ be the positive and locally Lipschitz function given by

$$\alpha(x, \hat{x}) = \frac{\max\{V(\hat{x}), 1\}}{2\left(1 + \left|\frac{\partial V}{\partial x}(\hat{x})\right|\right)\left(1 + \left|\nabla_{g(\hat{x})} d_g^2(x, \hat{x})\right|\right)}. \tag{10}$$

It yields

$$|k(x, \hat{x})| \leqslant \frac{\max\{V(\hat{x}), 1\}}{2\left(1 + \left|\frac{\partial V}{\partial x}(\hat{x})\right|\right)}, \quad \forall (x, \hat{x}) \in \mathbb{R}^n \times \mathcal{D}. \tag{11}$$

We prove Theorem 4 in three steps.

Step 1 : the $\hat{x}$-component of semi-trajectories of (8) remain in a compact subset of $\mathcal{D}$. For all $(x, \hat{x}) \in \mathbb{R}^n \times \mathcal{D}$, it follows from (9) and (11) that

$$\begin{aligned}\frac{\partial V}{\partial x}(\hat{x})[f(\hat{x}, \lambda(\hat{x})) + k(x, \hat{x})] &\leqslant -V(\hat{x}) + \left|\frac{\partial V}{\partial x}(\hat{x})\right| \frac{\max\{V(\hat{x}), 1\}}{2\left(1 + \left|\frac{\partial V}{\partial x}(\hat{x})\right|\right)} \\ &\leqslant -V(\hat{x}) + \frac{1}{2}\max\{V(\hat{x}), 1\}.\end{aligned}$$

Hence, if $\hat{x} \in \mathcal{D} \setminus D(1)$,

$$\frac{\partial V}{\partial x}(\hat{x})(f(\hat{x}, \lambda(\hat{x})) + k(x, \hat{x})) \leqslant -\frac{1}{2} V(\hat{x}) \ . \tag{12}$$

For all initial conditions $(x_0, \hat{x}_0) \in \mathbb{R}^n \times \mathcal{D}$, the solution $(x, \hat{x})$ of the closed-loop system (8) satisfies

$$V(\hat{x}(t)) \leqslant \max\{V(\hat{x}_0), 1\},$$

for all $t \geqslant 0$, in the time domain of existence of the solution. In other words, $\hat{x}(t) \in D(1) \cup D(V(\hat{x}_0))$ which is a compact subset of $\mathcal{D}$.

Step 2 : the distance between $\hat{x}$ and x is non-increasing and has limit zero. System (8) can be rewritten as

$$\begin{bmatrix}\dot{x} \\ \dot{\hat{x}}\end{bmatrix} = F(x, \hat{x}) + K(x, \hat{x}) \tag{13}$$

by setting $F(x, \hat{x}) = \begin{bmatrix} f(x, \lambda(\hat{x})) \\ f(\hat{x}, \lambda(\hat{x}))\end{bmatrix}$ and $K(x, \hat{x}) = \begin{bmatrix} 0 \\ -\alpha(x, \hat{x}) \nabla_{g(\hat{x})} d_g^2(x, \hat{x})\end{bmatrix}$.

Since (1) is weakly contractive with respect to g, the result proved in appendix applied to the control $u = \lambda(\hat{x})$ shows that

$$L_F d_g^2(x, \hat{x}) \leqslant 0.$$

Thus, by (7),

$$L_{F+K} d_g^2(x,\hat{x}) \leqslant -\alpha(x,\hat{x}) \left| \nabla_{g(\hat{x})} d_g^2(x,\hat{x}) \right|^2_{g(\hat{x})} . \tag{14}$$

Hence, for all $(x_0,\hat{x}_0) \in \mathbb{R}^n \times \mathcal{D}$, $t \mapsto d_g(x(t),\hat{x}(t))$ is non-increasing and for all $t \geqslant 0$ on the time domain of existence of the solution we have

$$(x(t),\hat{x}(t)) \in \Gamma(x_0,\hat{x}_0),$$

where

$$\Gamma(x_0,\hat{x}_0) = \Big\{ (\xi,\hat{\xi}) \in \mathbb{R}^n \times \mathcal{D} \mid \hat{\xi} \in D(1) \cup D(V(\hat{x}_0)), \\ d_g(\xi,\hat{\xi}) \leqslant d_g(x_0,\hat{x}_0) \Big\}.$$

The set $\Gamma(x_0,\hat{x}_0)$ is closed and bounded, and g is a complete metric. Hence, according to the Hopf-Rinow theorem, $\Gamma(x_0,\hat{x}_0)$ is compact. Therefore, $(x,\hat{x})$ remains in a compact subset of $\mathbb{R}^n \times \mathcal{D}$. Thus, solutions of (8) are complete in positive time.

Given $(x_0,\hat{x}_0) \in \mathbb{R}^n \times \mathcal{D}$, let $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ be the function defined by

$$\kappa(s) = \min_{(\xi,\hat{\xi}) \in \Gamma(x_0,\hat{x}_0) | d_g(\xi,\hat{\xi}) = s} \alpha(\xi,\hat{\xi}) \left| \nabla_{g(\hat{\xi})} d_g^2(\xi,\hat{\xi}) \right|^2_{g(\hat{\xi})} .$$

Note that if $x_0 \neq \hat{x}_0$, then, for all $s > 0$, $\kappa(s) > 0$ since α takes positive values and (6) holds. Hence, (14) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t} d_g^2(x(t),\hat{x}(t)) \leqslant -\kappa(d_g^2(x(t),\hat{x}(t))) \ , \ \forall t \geqslant 0. \tag{15}$$

Thus $\lim_{t\to+\infty} d_g(x(t),\hat{x}(t)) = 0$.

Step 3 : attractivity and local asymptotic stability of the origin. Given $(x_0,\hat{x}_0)$ in $\mathbb{R}^n \times \mathcal{D}$, let $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ be the function defined by

$$\mu(s) = \max_{(\xi,\hat{\xi}) \in \Gamma(x_0,\hat{x}_0) | d_g(\xi,\hat{\xi}) \leqslant s} \left| \frac{\partial V}{\partial x}(\hat{\xi}) k(\xi,\hat{\xi}) \right| .$$

Then μ is non-decreasing, continuous and $\mu(0) = 0$. Moreover, the solution $(x,\hat{x})$ of (8) initialized at $(x_0,\hat{x}_0) \in \mathbb{R}^n \times \mathcal{D}$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} V(\hat{x}(t)) \leqslant -V(\hat{x}(t)) + \mu(d_g(x(t),\hat{x}(t)). \tag{16}$$

From this inequality and Step 2 we conclude that $\lim_{t\to+\infty}(x(t),\hat{x}(t)) = (0,0)$.

Inequalities (14) and (16) being true for all solutions starting in $\Gamma(x_0,\hat{x}_0)$, this implies also stability of $(0,0)$.

2 Link with Jurdjevic and Quinn approach

2.1 Jurdjevic and Quinn result

The next result follows from the work of Jurdjevic and Quinn in [7]. The version that we state here is a direct corollary of [13, Theorem II.1]

Theorem 5 (Jurdjevic and Quinn approach)**.** *Consider the control system*

$$\dot{x} = a(x) + b(x,u)u, \tag{17}$$

with a and b two C^1 functions. Assume that there exists a C^1 positive definite proper function $V : \mathbb{R}^n \mapsto \mathbb{R}_+$ such that

$$L_a V \leqslant 0.$$

If the only solution of the system

$$\dot{x} = a(x), \quad L_{b(\cdot,0)}V(x) = 0, \quad L_a V(x) = 0 \tag{18}$$

is $x \equiv 0$, then (17) is globally asymptotically stabilizable by a static state feedback.

In the context of weakly contractive control systems, the Jurdjevic and Quinn approach leads to the following corollary.

Corollary 6. *Let g be a complete Riemannian metric on $\mathbb{R}^n$. Assume that (1) is weakly contractive with respect to g and that $f \in C^2(\mathbb{R}^n \times \mathcal{U}, \mathbb{R}^n)$. If the only solution of the system*

$$\dot{x} = f(x,0), \ \left(L_{b(\cdot,0)} d_g^2(\cdot,0)\right)(x) = 0, \ L_{f_0} d_g^2(x,0) = 0$$

where $b(\cdot,0) = \frac{\partial f}{\partial u}(\cdot,0)$ is $x \equiv 0$, then (1) is globally asymptotically stabilizable by a static state feedback.

To prove this corollary, it is sufficient to apply Theorem 5 with $V : x \mapsto d_g(x,0)^2$, $a : x \mapsto f(x,0)$ and $b : (x,u) \mapsto \int_0^1 \frac{\partial f}{\partial u}(x,su)ds$.

2.2 Link with our result

Note that the Jurdjevic-Quinn approach guarantees the existence of a *static* state feedback, contrarily to our main Theorem 4 which build a *dynamic* state feedback. However, the feedback obtained by their approach is implicit, while our dynamic state feedback is explicitly given by (8).

Moreover, our feedback law differs strongly with the one given in Jurdjevic-Quinn approach. Indeed, in their approach the feedback is designed small enough to make sure that it acts in a good direction related to the Lyapunov function. In our framework, this is no more a *small feedback approach* but more a *small correction term for an observer approach.*

Let us consider the particular case in which

$$f(x,u) = Ax + b(x)u. \tag{19}$$

where $A \in \mathbb{R}^{n\times n}$ and $b \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Then (19) is weakly contractive with respect to some constant metric g if and only if $L_A g \leqslant 0$ and $L_b g = 0$[3]. Moreover, the pair $(A, b(0))$ is controllable if and only if (19) is locally asymptotically stabilizable by a static feedback. Then, if all these hypotheses hold, a dynamic globally stabilizing state feedback is given by Theorem 4.

We can also show under the same hypotheses that the Jurdjevic and Quinn approach can be applied. Indeed, the system in Corollary 6 is equivalent to

$$\dot{x} = Ax, \quad \left(L_b d_g^2(\cdot,0)\right)(x) = 0, \quad L_A d_g^2(x,0) = 0 \tag{20}$$

which implies that $x \equiv 0$ when the pair $(A, b(0))$ is controllable. Then, according to Corollary 6, (19) is globally asymptotically stabilizable by a static state feedback. However, it is not clear in general that both contexts are equivalent. Finding an example of nonlinear system satisfying the hypotheses of Theorem 4 but not the hypotheses of Corollary 6 remains an open question.

[3]It is easy to check that it is the case if and only if $b(x) = b(0) + Jx$ with $L_J g = 0$.

3 Appendix on weakly contractive vector fields

For all $u : \mathbb{R}_+ \to \mathcal{U}$ and all $x \in \mathbb{R}^n$, denote by $t \mapsto X_u(x,t)$ the solution of (1) with initial condition x. Let $u : \mathbb{R}_+ \to \mathcal{U}$ be such that X_u is well-defined and C^2 on $\mathbb{R}^n \times \mathbb{R}_+$. Let $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$ and $\gamma : [s_1, s_2] \to \mathbb{R}^n$ be a C^2 path between the points $x_1 = \gamma(s_1)$ and $x_2 = \gamma(s_2)$. For all $(s,t) \in [s_1, s_2] \times \mathbb{R}_+$, set $\Gamma(s,t) = X_u(\gamma(s), t)$ and $\rho(s,t) = \left|\frac{\partial \Gamma}{\partial s}(s,t)\right|^2_{g(\Gamma(s,t))}$. Then ρ is C^1 and

$$\frac{\partial \rho}{\partial t}(s,t) = L_{f_u} g(\Gamma(s,t)) \left(\frac{\partial \Gamma}{\partial s}(s,t), \frac{\partial \Gamma}{\partial s}(s,t)\right) \leqslant 0,$$

which yields

$$\begin{aligned}\frac{\mathrm{d}\ell(\Gamma(\cdot,t))}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{s_1}^{s_2} \sqrt{\rho(s,t)}\mathrm{d}s \\ &= \int_{s_1}^{s_2} \frac{1}{2\sqrt{\rho(s,t)}} \frac{\partial \rho}{\partial t}(s,t)\mathrm{d}s \\ &\leqslant 0.\end{aligned}$$

Hence $d_g(X_u(x_1,t), X_u(x_2,t)) \leqslant \ell(\Gamma(\cdot,t)) \leqslant \ell(\gamma)$. Choosing a sequence of paths $(\gamma_n)_{n\in\mathbb{N}}$ such that $\ell(\gamma_n) \to d_g(x_1,x_2)$ and passing to the limit we get

$$d_g(X_u(x_1,t), X_u(x_2,t)) \leqslant d_g(x_1,x_2).$$

Since this inequality is true for any control input u, $t \mapsto d_g(X_u(x_1,t), X_u(x_2,t))$ is non-increasing for all control u and all points x_1, x_2.

References

[1] N. Aghannan and P. Rouchon. An intrinsic observer for a class of Lagrangian systems. *IEEE Trans. Automat. Control*, 48(6):936–945, 2003.

[2] S. Benachour, H. Stein Shiromoto, and V. Andrieu. Locally optimal controllers and globally inverse optimal controllers. *Automatica J. IFAC*, 50(11):2918–2923, 2014.

[3] J. Bernat and J. Llibre. Counterexample to Kalman and Markus-Yamabe conjectures in dimension larger than 3. *Dynam. Contin. Discrete Impuls. Systems*, 2(3):337–379, 1996.

[4] M. P. do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[5] K. Ezal, Z. Pan, and P. V. Kokotović. Locally optimal and robust backstepping design. *IEEE Trans. Automat. Control*, 45(2):260–271, 2000.

[6] H. Hammouri and J.-C. Marquès. Two global stabilizability results for homogeneous systems from a local stabilizability assumption. *IEEE Trans. Automat. Control*, 54(9):2239–2244, 2009.

[7] V. Jurdjevic and J. P. Quinn. Controllability and stability. *J. Differential Equations*, 28(3):381–389, 1978.

[8] M. Kawski. Homogeneous stabilizing feedback laws. *Control Theory Adv. Tech.*, 6(4):497–516, 1990.

[9] J. Kurzweil. On the inversion of Lyapunov's second theorem on stability of motion. *Czechoslovak Math. J.*, 06(2):217–219, 1956.

[10] J. Kurzweil. On the inversion of Lyapunov's second theorem on stability of motion. *Amer. Math. Soc. Transl.*, 2(24):19–77, 1963.

[11] L. Markus and H. Yamabe. Global stability criteria for differential systems. *Osaka Math. J.*, 12:305–317, 1960.

[12] J. L. Massera. Contributions to stability theory. *Ann. of Math. (2)*, 64:182–206, 1956.

[13] F. Mazenc and L. Praly. Adding integrations, saturated controls, and stabilization for feedforward systems. *IEEE Trans. Automat. Control*, 41(11):1559–1578, 1996.

[14] G. Meisters. A biography of the Markus-Yamabe conjecture. *Aspects of mathematics*, 1996.

[15] C. Olech. On the global stability of an autonomous system on the plane. *Contributions to Differential Equations*, 1:389–400, 1963.

[16] L. Rosier. Homogeneous Lyapunov function for homogeneous continuous vector fields. *Systems Control Lett.*, 19(6):467–473, 1992.

[17] R. G. Sanfelice and L. Praly. Convergence of nonlinear observers on $\mathbb{R}^n$ with a Riemannian metric (Part I). *IEEE Trans. Automat. Control*, 57(7):1709–1722, 2012.

[18] A. R. Teel and L. Praly. A smooth Lyapunov function from a class-$\mathcal{KL}$ estimate involving two positive semidefinite functions. *ESAIM Control Optim. Calc. Var.*, 5:313–367, 2000.

Résumé détaillé

Contexte

Cette thèse s'articule autour de deux thèmes différents mais liés. Dans une première partie, nous nous intéressons au problème de stabilisation par bouclage de sortie dynamique. Lorsque seulement une partie de l'état d'un système dynamique est connue, un bouclage d'état stabilisant ne peut pas être implémenté. Dès lors, une stratégie possible pour stabiliser l'état sur un point cible consiste à concevoir un observateur, afin d'estimer asymptotiquement l'état en filtrant la sortie au cours du temps, et à utiliser comme contrôleur la loi de commande stabilisante appliquée à l'observateur. Cette approche est connue pour être efficace sur les systèmes uniformément observables, c'est-à-dire observables pour toute entrée. Cependant, les systèmes non-linéaires ne sont génériquement pas uniformément observables lorsque la dimension de la sortie est inférieure ou égale à celle de l'entrée. Ainsi, en présence de singularités d'observabilité, de nouvelles techniques restent à développer.

Dans une seconde partie, nous traitons du problème de synthèse d'observateur pour les systèmes linéaires temps-variant de dimension infinie. L'objectif est de concevoir un système dynamique capable d'estimer l'état du système de départ à partir d'une mesure et de sa dynamique. La notion d'observabilité peut se généraliser de plusieurs façons en dimension infinie. En particulier, on distingue les hypothèses d'observabilité exacte et approchée. Alors qu'une convergence exponentielle des observateurs de Luenberger peut généralement être montrée sous des hypothèses d'observabilité exacte, les résultats portant sur des hypothèses d'observabilité approchée, auxquelles nous nous intéressons, sont plus rares. Ces observateurs peuvent également être utilisés dans le contexte de la reconstitution de la condition initiale d'un système. La procédure, appelée Back and Forth Nudging (BFN), est alors basée sur des itérations successives d'observateurs en temps positifs et en temps rétrograde. Ces méthodes peuvent être appliquées à un procédé de cristallisation par lots, dans lequel l'état à estimer est la distribution en taille des particules (PSD).

Principales contributions

Dans le Chapitre 1, nous formulons le problème de stabilisation par bouclage de sortie dynamique, et énonçons des conditions nécessaires. Des résultats de la littérature existante sont rappelés. Nous distinguons deux grandes classes de systèmes non-uniformément observables, selon que leur cible correspond à une commande observable ou non. Dans le Chapitre 2, nous nous intéressons aux systèmes observables à la cible. La difficulté réside dans l'existence d'entrées rendant le système inobservable que la boucle de rétroaction peut produire au cours de la stabilisation. Éviter ces entrées serait un premier pas vers la réalisation d'un principe de séparation

générique. Notre principale contribution à ce problème est énoncée dans [Bri+21b].

Contribution 1. Pour les systèmes bilinéaires possédant une entrée et une seule sortie (SISO), des perturbations génériques de la loi de commande garantissent que les entrées produites par la boucle de rétroaction rendent le système observable.

Dans le Chapitre 3, nous soulignons l'utilité de propriétés de dissipation dans le contexte de la stabilisation par bouclage de sortie. Deux trajectoires d'un même système dissipatif ne s'éloignent pas l'une de l'autre au cours du temps. Cela permet de construire des observateurs de Luenberger à système d'erreur dissipatif, c'est-à-dire dont l'erreur est décroissante, indépendamment des propriétés d'observabilité. Nos résultats sont énoncés dans [Sac+20], et un corollaire concernant la stabilisation par bouclage d'état est démontré dans [Bri+21c].

Contribution 2. Pour les systèmes dissipatifs affines en l'état stabilisables par bouclage d'état, la 0-détectabilité est une condition est nécessaire et suffisante à l'existence d'un bouclage de sortie dynamique globalement stabilisant.

Les systèmes inobservables à la cible sont étudiés dans le Chapitre 4. Nous rassemblons les idées développées dans les chapitres précédents afin de tracer des lignes directrices en vue de la résolution du problème, et illustrons ces principes sur des exemples de systèmes à dynamique linéaire et sortie non-linéaire. Afin de tirer parti de propriétés de dissipation sur une classe plus large de système, nous proposons une stratégie basée sur l'utilisation de plongements : l'observateur est alors synthétisé sur le système plongé, qui a été conçu pour une admettre un système d'erreur dissipatif. Ce travail a abouti au manuscrit [Bri+20b].

Contribution 3. Sur des exemples de systèmes non-linéaires, nous illustrons trois grands principes pour la stabilisation par bouclage de sortie dynamique à une cible inobservable :

- des perturbations additives de la loi de commande engendrent de nouvelles propriétés d'observabilité, sans compromettre le processus de stabilisation ;
- les observateurs à système d'erreur dissipatifs sont robustes aux singularités d'observabilité ;
- des plongements dans des systèmes de dimension finie ou infinie permettent de concevoir des observateurs de Luenberger avec des systèmes d'erreur dissipatifs.

Dans le Chapitre 4, nous utilisons des observateurs de Luenberger de dimension infinie, sous des hypothèses d'observabilité approchée. Cela nous conduit naturellement à la seconde partie de cette thèse. Nos principaux résultats théoriques dans cette partie sont énoncés dans les Chapitres 5 et 6.

Contribution 4. Sous une hypothèse dite de détectabilité faible, les observateurs de Luenberger de dimension infinie convergent vers la partie observable de l'état dans la topologie faible de l'espace d'état.

Contribution 5. Les résultats de convergence de la Contribution 4 s'adaptent au contexte BFN.

Ces travaux ont abouti à la publication [Bri+21a]. Durant cette thèse, le problème de synthèse d'observateurs de dimension infinie a été considéré sur un procédé de cristallisation par lots. L'état à estimer, à partir de différentes mesures, est alors la distribution en taille des particules (PSD). Nous proposons trois stratégies d'estimation, qui sont énoncées dans le Chapitre 7 et ont conduit aux articles [Bri+20a, BS20, Bri+21a].

Contribution 6. Dans le contexte d'un procédé de cristallisation par lots, nous proposons plusieurs stratégies de construction de la PSD :

- une approche directe, basée sur une méthode de régularisation de Tikhonov, utilisant la mesure de la distribution en taille des cordes (CLD) ;
- un observateur de Kazantzis-Kravaris/Luenberger (KKL), utilisant la mesure de la température et de la concentration en soluté ;
- un observateur de Luenberger de dimension infinie, basé sur les Contributions 4 et 5, utilisant la mesure de la CLD.

Chaque chapitre débute avec un résumé indépendant.

Publications

Au cours de cette thèse, les articles suivants ont été publiés ou proposés à la publication. Les articles [MBA20a] et [MBA20b] traitent du problème de régulation de la sortie pour des systèmes couplés EDO/EDP. Ce sujet n'est pas abordé dans cette thèse, mais est lié aux hypothèses d'observabilité approchées étudiées dans le Chapitre 5. Les récents articles [BS21a] et [BS21b] ne sont pas discutés dans la thèse.

- Articles de journaux :

  1. L. Brivadis, V. Andrieu, É. Chabanon, É. Gagnière, N. Lebaz et U. Serres. "New dynamical observer for a batch crystallization process based on solute concentration". *Journal of Process Control* 87 (2020), p. 17-26. ISSN : 0959-1524. DOI : 10.1016/j.jprocont.2019.12.012
  2. L. Brivadis, L. Sacchelli, V. Andrieu, J.-P. Gauthier et U. Serres. "From local to global asymptotic stabilizability for weakly contractive control systems". *Automatica J. IFAC* 124 (2021). ISSN : 0005-1098. DOI : 10.1016/j.automatica.2020.109308.
  3. L. Brivadis, V. Andrieu, U. Serres et J.-P. Gauthier. "Luenberger observers for infinite-dimensional systems, Back and Forth Nudging, and application to a crystallization process". *SIAM Journal on Control and Optimization* 59.2 (2021), p. 857-886. DOI : 10.1137/20M1329020.
  4. L. Brivadis, J.-P. Gauthier, L. Sacchelli et U. Serres. "Avoiding observability singularities in output feedback bilinear systems". *SIAM Journal on Control and Optimization* 59.3 (2021), p. 1759-1780. DOI : 10.1137/19M1272925.
  5. S. Marx, L. Brivadis et D. Astolfi. "Forwarding techniques for the global stabilization of dissipative infinite-dimensional systems coupled with an ODE". Submitted to *Mathematics of Control, Signals, and Systems*. Under review. Sept. 2020.

6. L. Brivadis, J.-P. Gauthier, L. Sacchelli et U. Serres. “New perspectives on output feedback stabilization at an unobservable target”. Submitted to *ESAIM. Control, Optimisation and Calculus of Variations*. Under review. Nov. 2020.
7. L. Brivadis et L. Sacchelli. “New inversion methods for the single/multi-shape CLD-to-PSD problem with spheroid particles”. Submitted to *Journal of Process Control*. Under review. Déc. 2020.

- Articles de conférences :

1. L. Brivadis, V. Andrieu et U. Serres. “Luenberger observers for discrete-time nonlinear systems”. In : *2019 IEEE 58th Conference on Decision and Control (CDC)*. 2019, p. 3435-3440. doi : 10.1109/CDC40024.2019.9029220.
2. S. Marx, L. Brivadis et D. Astolfi. “Forwarding design for stabilization of a coupled transport equation-ODE with a cone-bounded input nonlinearity”. In : *2020 59th IEEE Conference on Decision and Control (CDC)*. 2020, p. 640-645. doi : 10.1109/CDC42340.2020.9304178.
3. L. Sacchelli, L. Brivadis, V. Andrieu, U. Serres et J.-P. Gauthier. “Dynamic output feedback stabilization of non-uniformly observable dissipative systems”. *IFAC-PapersOnLine* 53.2 (2020). 21th IFAC World Congress, p. 4923-4928. issn : 2405-8963. doi : https://doi.org/10.1016/j.ifacol.2020.12.1071.
4. L. Brivadis et L. Sacchelli. “A switching technique for output feedback stabilization at an unobservable target”. Submitted to *2021 60th IEEE Conference on Decision and Control (CDC)*. Under review. Mar. 2021.
5. L. Brivadis et L. Sacchelli. “Approximate observability and back and forth observer of a PDE model of crystallisation process”. Submitted to *2021 60th IEEE Conference on Decision and Control (CDC)*. Under review. Mar. 2021.

Chapitre 1 : Problématique

Résumé. *Ce chapitre est une introduction à la problématique de stabilisation par bouclage de sortie dynamique. Les concepts essentiels sont définis, et des conditions nécessaires sont énoncées. Des principes de séparations usuels sur les systèmes uniformément observables sont rappelés, et des stratégies existantes sur les systèmes non-uniformément observables sont introduites.*

Introduction

Stabiliser l'état d'un système dynamique sur un point cible est un problème classique en théorie du contrôle. Cependant, dans beaucoup de problèmes physiques, seulement une partie de l'état, appelée la sortie, est connue. Dès lors, une stabilisation par bouclage d'entrée ne peut pas être implémentée. Seule la sortie, et l'état d'un système dynamique gouverné par la sortie, peuvent être utilisés pour stabiliser l'état du système de départ. Ce problème, connu sous le nome de *stabilisation par bouclage de sortie dynamique*, a été largement étudié (voir, par exemple, [GB81,EK92,GK92, KE93,Cor94a,TP94,JG95,TP95,AK99,MPI07,AP09]). Lorsqu'une loi de commande stabilisante par bouclage d'état peut être synthétisée, une stratégie couramment employée pour résoudre le problème du bouclage de sortie consiste à synthétiser un observateur, c'est-à-dire un système dynamique guidé par la sortie estimant l'état du système original au cours du temps, et à appliquer la loi de commande stabilisante à l'estimation obtenue par l'observateur. Cette stratégie connue pour être efficace sur les systèmes uniformément observables depuis les travaux [TP94, TP95] et [JG95]. L'*observabilité* d'un système de contrôle pour une entrée fixée qualifie la capacité à distinguer l'état à partir de la connaissance de la sortie. Elle caractérise le fait que deux trajectoires du système peuvent être différenciées par leur sortie respective sur un intervalle de temps donné. Cette notion cruciale constitue un champ d'étude à part entière (voir, par exemple, [GK01, AP09, Ber+17, Ber19]). Un système est uniformément observable dès lors qu'il est observable pour toute entrée. Cependant, comme montré dans [GK01], l'uniforme observabilité n'est pas une propriété générique sur les systèmes de contrôle lorsque la dimension de l'entrée égale ou excède celle de la sortie. Il existe alors des entrées singulières rendant le système inobservable, et le bouclage de sortie peut engendrer de telles entrées. Cela contrarie la stabilisation par bouclage de sortie dynamique, qui demeure un problème ouvert lorsque de telles entrées existent. La première partie de cette thèse est consacrée à l'étude de cette problématique.

On distinguera deux grandes classes de systèmes, selon que la valeur de la loi de commande (stabilisante par bouclage d'état) correspond ou non à une entrée constante rendant le système observable. Le Chapitre 2 est dédié au premier cas, et le Chapitre 4 au second. Le Chapitre 3 présente un résultat intermédiaire reliant ces deux parties.

Définition

Soient $n, m, p \in \mathbb{N}$, $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ et $h : \mathbb{R}^n \to \mathbb{R}^m$. Pour tout $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$, considérons le système suivant :

$$\begin{cases} \dot{x} = f(x, u) \\ y = h(x) \end{cases} \tag{1.1}$$

où x est l'état du système, u est le contrôle (ou entrée) et y est l'observation (ou sortie, ou mesure).

La première partie de la thèse est consacrée au problème de stabilisation par bouclage de sortie dynamique de (1.1). L'objectif est d'utiliser la mesure de y au cours du temps pour stabiliser, en agissant sur le contrôle u, l'état x à un point cible $x^\star \in \mathbb{R}^n$. Une condition nécessaire est l'existence de $u^\star \in \mathbb{R}^p$ thel que $x^\star$ est un point d'équilibre de $\dot{x} = f(x, u^\star)$, *i.e.,* $f(x^\star, u^\star) = 0$. Quitte à changer les coordonnées du système, on supposera sans perte de généralité $(x^\star, u^\star) = (0, 0)$ et $h(0) = 0$. Afin de garantir le caractère bien posé du problème de Cauchy associé au système en boucle ouverte (1.1), supposons f continue et uniformément localement lipschitzienne par rapport à x. D'après le théorème de Cauchy-Lipschitz, pour tout $u \in C^0(\mathbb{R}_+, \mathbb{R}^n)$ et chaque $x_0 \in \mathbb{R}^n$, il existe une unique solution maximale $\varphi_t(x_0, u)$ définie pour $t \in [0, T_u(x_0))$ telle que $\varphi_0(x_0, u) = x_0$ et $\frac{\partial \varphi_t(x_0,u)}{\partial t} = f(\varphi_t(x_0, u), u(t))$. L'application φ est continue et appelée le flot de (1.1).

Définition 1.1 (Stabilisabilité par bouclage de sortie dynamique)**.** Le système (1.1) est dit *localement* (resp. *globalement*) *stabilisable par bouclage de sortie dynamique* si et seulement si la propriété suivante est vérifiée.

Il existe deux fonctions continues $\nu : \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^q$ et $\varpi : \mathbb{R}^q \times \mathbb{R}^m \to \mathbb{R}^p$ pour un certain $q \in \mathbb{N}$ tel que $(0, 0) \in \mathbb{R}^n \times \mathbb{R}^q$ est un point d'équilibre localement (resp. globalement) asymptotiquement stable du système suivant :

$$\begin{cases} \dot{x} = f(x, u) \\ y = h(x) \end{cases}, \qquad \begin{cases} \dot{w} = \nu(w, u, y) \\ u = \varpi(w, y). \end{cases} \tag{1.2}$$

De plus, si pour tout compact $\mathcal{K}_x \subset \mathbb{R}^n$, il existe deux fonctions continues $\nu : \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^q$ et $\varpi : \mathbb{R}^q \times \mathbb{R}^m \to \mathbb{R}^p$ pour un certain $q \in \mathbb{N}$, et un compact $\mathcal{K}_w \subset \mathbb{R}^q$ tel que $(0, 0) \in \mathbb{R}^n \times \mathbb{R}^q$ est un point d'équilibre localement asymptotiquement stable de (1.2) avec un bassin d'attraction contenant $\mathcal{K}_x \times \mathcal{K}_w$, alors (1.1) est dit *semi-globalement stabilisable par bouclage de sortie dynamique.*

Conditions nécessaires

Le problème de stabilisation par bouclage d'état dynamique est équivalent au problème de stabilisation par bouclage de sortie dynamique lorsque $h(x) = x$. Par conséquent, il est une condition nécessaire au problème de stabilisation par bouclage de sortie dynamique. Dès lors, on peut se demander si la stabilisation par bouclage d'état *statique* est une condition nécessaire au bouclage de sortie dynamique. Le lemme [AP09, Lemma 1, (1)] répond par la positive dans le cas où une fonction de sélection suffisamment lisse existe. Par conséquent, dans le cadre de la stabilisation par bouclage de sortie dynamique, on supposera la condition suivante vérifiée.

Condition 1.2 (Stabilisabilité par bouclage d'état statique)**.** — local, semi-global, global] Le ststème (1.1) est localement (resp. semi-globalement, globalement) stabilisable par bouclage d'état statique.

Par ailleurs, nous démontrons que les deux conditions suivantes (inspirées par [Cor94a] et [Son81]) sont nécessaires au problème de stabilisation par bouclage de sortie dynamique.

Condition 1.3 (0-détectabilité — local, global)**.** Soit $\mathcal{X}_0 = \{x_0 \in \mathbb{R}^n \ : \ \forall t \in [0, T_0(x_0)), h(\varphi_t(x_0, 0)) = 0\}$. Le point $0 \in \mathcal{X}_0$ est un point d'équilibre localement (resp. globalement) asymptotiquement stable du champ de vecteur $\mathcal{X}_0 \ni x \mapsto f(x, 0)$.

Condition 1.4 (Indistinguabilité $\implies$ stabilisabilité simultanée — local, global)**.** Pour tout x_0, $\tilde{x}_0$ dans un voisinage de $0 \in \mathbb{R}^n$ (resp. pour tout x_0, $\tilde{x}_0$ dans $\mathbb{R}^n$), si pour tout $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ tel que $T_u(x_0) = +\infty$, $h(\varphi_t(x_0, u)) = h(\varphi_t(\tilde{x}_0, u))$ pour tout $t \in [0, T_u(\tilde{x}_0))$, alors il existe $v \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ tel que $\varphi_t(x_0, v)$ et $\varphi_t(\tilde{x}_0, v)$ sont bien définis pour tout $t \in \mathbb{R}_+$ et tendent vers 0 quand t tend vers l'infini.

Principe de séparation, observabilité uniforme

Dans cette thèse, nous nous concentrons sur des approches de stabilisation indirectes au sens de [AP09]. Notre objectif est de construire un observateur $\hat{x}$ de l'état x, basé sur la mesure y, puis d'appliquer le contrôle $u = \phi(\hat{x})$, où ϕ est une loi de commande stabilisante par bouclage d'état. Pour les systèmes linéaires, il est suffisant de synthétiser « séparément » l'observateur et la loi de commande. Ce principe est connu comme le *principe de séparation*. Cela n'est plus vrai pour les systèmes non-linéaires. D'une part, l'existence d'un observateur convergeant vers l'état et d'une loi de commande stabilisante n'est pas suffisante pour garantir la stabilisation par bouclage de sortie dynamique. D'autre part, même lorsque cela est possible, certains paramètres de l'observateur doivent dépendre de la loi de commande. Pour cette raison, les principes de séparation pour les systèmes non-linéaires sont également appelés commandes par bouclage de sortie basées-observateur.

Une notion cruciale pour garantir l'efficacité des principes de séparation usuels est l'observabilité.

Définition 1.5 (Observabilité)**.** Le système (1.1) est dit *observable* en temps T pour une entrée $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$ si et seulement si, pour toutes conditions initiales $x_0 \neq \tilde{x}_0 \in \mathbb{R}^n$, l'ensemble

$$\left\{\forall t \in [0, \min(T, T_u(x_0), T_u(\tilde{x}_0))), \ h(\varphi_t(x_0, u)) \neq h(\varphi_t(\tilde{x}_0, u))\right\} \tag{1.3}$$

est de mesure non-nulle. Si (1.1) est observable en tout temps $T > 0$ pour toute entrée u, alors il est dit *uniformément observable* en temps petit.

Les notions (plus fortes) d'observabilité uniforme complète et d'observabilité différentielle forte permettent de mettre en place des principes de séparation non-linéaires, comme montré dans [TP94, TP95] et [JG95], respectivement. Lorsque la dimension de la sortie est supérieure à celle de l'entre, *i.e.,* $m > P$, l'observabilité différentielle forte est une propriété générique des systèmes non-linéraires (voir [GK01, Chapitre 4, Théorème 2.2]). Cependant, cela n'est plus le cas dès lors que $m \leqslant p$. La question de la stabilisation par bouclage de sortie dynamique reste donc largement ouverte pour ces systèmes, dit non-uniformément observables.

Systèmes non-uniformément observables

On distinguera deux classes de systèmes on-uniformément observables, selon que leur point cible correspond à une entrée observable ou non.

Définition 1.6 (Observabilité à la cible)**.** Le système (1.1) est *observable à la cible* en un temps $T > 0$ s'il est observable en temps T pour l'entrée constante $u \equiv 0$. Sinon, (1.1) est *inobservable à la cible*.

L'entrée $u \equiv 0$ correspond à la valeur du contrôle au point cible de la boucle fermée (1.2). Si la cible est observable, alors lorsque l'état tend vers la cible, le contrôle utilisé dans la boucle fermée tend vers un contrôle rendant le système observable. Par conséquent, les singularités d'observabilité ne peuvent qu'être traversées durant le régime transitoire. La principale difficulté consiste alors à éviter ces entrées inobservables. Si la cible est inobservable, alors la singularité d'observabilité est en quelque sorte inévitable. En effet, si la stratégie de stabilisation porte ses fruits, le contrôle tend à rendre le système de moins en moins observable. Par conséquent, démontrer la convergence de l'observateur lorsque l'état est proche de la cible s'avère difficile.

Dans la littérature scientifique existante, une attention moindre a été portée aux systèmes non-uniformément observables, mais de récents travaux illustrant l'apparition de singularités d'observabilité dans des systèmes d'ingénierie variés (voir [HPR14, Com+16, Fla19, Sur+19, Aja+20, RD20, Sur+20, AGS21]) ont conduit à un renouveau de l'intérêt suscité par cette question. Une technique populaire, en particulier dans le cadre des systèmes inobservables à la cible, consiste à modifier le contrôle afin d'engendrer de nouvelles propriétés d'observabilité sans pour autant entraver la stabilisation. Cette approche a été inspirée par [Cor94a], dans lequel la stabilisation locale par bouclage de sortie dynamique *temps-variant périodique* est démontrée, sous une condition d'observabilité à zéro de Lie. Le résultat est basé sur une stratégie de stabilisation en deux étapes :

(Phase d'observation) Sur un intervalle de temps $[0, T]$, le système est « excité » par une entrée temps-variante rendant le système observable, mais s'annulant lorsque l'état atteint l'équilibre. L'observateur converge en temps fini vers l'état.

(Phase de stabilisation) Sur un intervalle de temps $[T, 2T]$, l'état est stabilisé en temps fini grâce à l'estimation exacte obtenue par l'observateur.

Une méthode similaire est employée dans [ST03], dans laquelle une stabilisation *pratique* semi-globale est obtenue. Plus récemment, des auteurs ont proposé d'accomplir les phases d'observation et de stabilisation à partir du même contrôle, choisi pour une être une légère modification du contrôle usuel $u = \phi(\hat{x})$. Le signal est choisi suffisamment proche de la commande originale, afin de ne pas contrarier le processus de stabilisation, mais la modulation engendrée permet d'améliorer les propriétés d'observabilité. C'est par exemple le cas de la méthode dite des « mesures virtuelles » proposée dans [Com+16, Sur+19, Sur+20]. C'est également le cas de la perturbation autonome de la loi de commande utilisée dans [LSG17] ;

Inspirés par les travaux de [LSG17], nous utilisons cette stratégie de perturbation de la loi de commande pour stabiliser des systèmes non-uniformément observables à des cibles observables (Chapitre 2) ou inobservable (Chapitre 4).

Chapitre 2 : Cible observable

Résumé. *Nous considérons le problème de la stabilisation par bouclage de sortie dynamique à une cible inobservable sur les systèmes bilinéaires SISO. La difficulté réside dans l'existence d'entrées rendant le système inobservable. Durant le processus de stabilisation, le contrôle généré par le système en boucle fermée peut être l'une de ces entrées, et donc empêcher la convergence de l'observateur. Pour aborder ce problème, nous proposons une stratégie de perturbation autonome de la loi de commande. La loi de commande perturbée conserve son caractère stabilisant par bouclage d'état, mais empêche le contrôle de la boucle fermée de rendre le système inobservable. Nous démontrons, sous des hypothèses génériques sur le système, l'existence d'un ensemble ouvert dense de telles perturbations. Nous appliquons ces résultats aux observateurs de Luenberger et de Kalman. Nous expliquons comment cette stratégie pourrait ouvrir la voie à la synthèse d'un principe de séparation générique sur les systèmes bilinéaires SISO.*

Introduction

Dans ce chapitre, nous nous restreignons à l'étude des systèmes bilinéaires à une seule entrée et une seule sortie (SISO) avec observation linéaire qui sont stabilisables par bouclage d'état statique à un point cible que, sans perte de généralité, nous supposons être l'origine. Nous supposons également que le système est observable à la cible, c'est-à-dire que le contrôle constant obtenu par évaluation de la loi de commande à la cible n'est pas singulier. Cette classe de système est un choix d'étude naturel pour deux raisons. D'abord, l'observabilité uniforme demeure une hypothèse non générique lorsque la dimension de l'entrée est supérieure ou égale à celle de la sortie. Ensuite, d'après un résultat de [FK83], tout système affine en le contrôle ayant un espace d'observation de dimension finie se plonge dans un système bilinéaire.

L'existence d'entrées rendant le système inobservable rend le problème de stabilisation par bouclage de sortie difficile, et aucune stratégie générale n'existe, même si la cible est observable. Le principal obstacle est que le contrôle généré par le système en boucle fermée peut être l'une de ces entrées singulières. Dans [LSG17], les auteurs proposent une stratégie de perturbation de la loi de commande afin d'éviter ce phénomène. Inspirée par ces travaux, une question peut se poser : « Peut-on assurer que seules des entrées observables seront produites par le bouclage de sortie dynamique obtenu en combinant un observateur et une loi de commande stabilisante ? » Cette question s'inscrit dans le cadre plus général de la synthèse d'un principe de séparation en présence d'observabilité de singularités. On ne peut espérer que toutes les lois de commandes stabilisantes par bouclage d'état satisfont cette propriété. En revanche, nous montrons que pour toute loi de commande, il existe de petites perturbations additives de la loi qui satisfont cette propriété d'observabilité tout en conservant le caractère stabilisant de la loi de commande. La théorie de la transversalité est utilisée pour démontrer l'existence d'un ouvert dense de telles perturbations. Ainsi, presque toutes les lois de commandes stabilisantes par retour d'état garantissent l'observabilité du système en boucle fermée. Concevoir un principe de séparation générique sur les systèmes bilinéaires SISO est au-delà du cadre de cette étude, mais les résultats obtenus peuvent ouvrir la voie à ce travail. Des hypothèses sur l'observateur utilisé sont requises pour démontrer nos résultats. Nous

vérifions ces conditions sur les observateurs de Luenberger et de Kalman.

Problématique

Soient $n \in \mathbb{N}$, $A, B \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{1\times n}$, $b \in \mathbb{R}^n$ et $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$. Posons $A_u = A + uB$ et considérons le système suivant :

$$\begin{cases} \dot{x} = A_u x + bu \\ y = Cx. \end{cases} \tag{2.1}$$

Notons $\mathbb{S}^{n-1}$ la sphère unité de $\mathbb{R}^n$. Du fait de la structure bilinéaire de (2.1), l'observabilité se caractérise de la façon suivante.

Proposition 2.1. *Le système* (2.1) *est observable en temps T pour le contrôle u si et seulement si pour tout $\omega_0 \in \mathbb{S}^{n-1}$, la solution de $\dot{\omega} = A_{u(t)}\omega$ initialisée à ω_0 vérifie $C\omega|_{[0,T]} \not\equiv 0$.*

La stabilisation par bouclage d'état des systèmes bilinéaires est un problème important pour lequel des stratégies variées ont été développées (voir, par exemple, [Qui80, Gut81, Bac90, BB91, CV00]). Dans ce chapitre, nous supposons l'existence d'une loi de commande stabilisante lisse. Soit $\lambda \in C^\infty(\mathbb{R}^n, \mathbb{R})$ tel que 0 est un point d'équilibre localement asymptotiquement stable du champ de vecteur $x \mapsto A_{\lambda(x)}x + b\lambda(x)$ sur un certain bassin d'attraction $\mathcal{D}(\lambda)$. Quitte à remplacer A par $A + \lambda(0)B$, nous supposons sans perte de généralité que $\lambda(0) = 0$.

Fixons maintenant la structure de l'observateur. Soit $\mathcal{S}_n \subset \mathbb{R}^{n\times n}$ la sous-variété des matrices symétriques définies positives et soit $L : \mathcal{S}_n \to \mathbb{R}^{n\times 1}$. Pour tout $u \in \mathbb{R}$, soit $f(\cdot, u)$ un champ de vecteur sur $\mathcal{S}_n$. En notant $\varepsilon = \hat{x} - x$ l'erreur d'estimation, nous introduisons l'observateur suivant, dépendant de la paire (f, L) :

$$\begin{cases} \dot{\hat{x}} = A_u\hat{x} + bu - L(\xi)C\varepsilon \\ \dot{\varepsilon} = (A_u - L(\xi)C)\,\varepsilon \\ \dot{\xi} = f(\xi, u). \end{cases} \tag{2.2}$$

Remarquons que les observateurs de Luenberger et Kalman s'écrivent sous cette forme en posant $f(\xi, u) = 0$ (observateur de Luenberger) ou $f_Q^{\text{Kalman}}(\xi, u) = \xi A_u' + A_u\xi + Q - \xi C'C\xi$ pour un certain $Q \in \mathcal{S}_n$ (observateur de Kalman), et, dans les deux cas, $L(\xi) = \xi C'$.

Enfin, nous considérons des perturbations δ de la loi de commande λ afin de garantir l'observabilité du système en boucle fermée. Le système en boucle fermée s'écrit donc :

$$\begin{cases} \dot{\hat{x}} = A_{(\lambda+\delta)(\hat{x})}\hat{x} + b(\lambda+\delta)(\hat{x}) - L(\xi)C\varepsilon \\ \dot{\varepsilon} = \left(A_{(\lambda+\delta)(\hat{x})} - L(\xi)C\right)\varepsilon \\ \dot{\xi} = f(\xi, (\lambda+\delta)(\hat{x})) \\ \dot{\omega} = A_{(\lambda+\delta)(\hat{x})}\omega. \end{cases} \tag{2.3}$$

Soit $\mathcal{K} = \mathcal{K}_x \times \mathcal{K}_\varepsilon \times \mathcal{K}_\xi$ un sous-ensemble compact semi-algébrique de $\mathcal{D}(\lambda) \times \mathbb{R}^n \times \mathcal{S}_n$. Cet ensemble sera celui des conditions initiales de (2.2). Pour tout $R > 0$, posons

$$\mathcal{V}_R = \{\delta \in C^\infty(\mathbb{R}^n, \mathbb{R}) \ : \ \forall x \in B(0, R), \quad \delta(x) = 0\}.$$

Afin d'établir nos résultats d'observabilité, nous faisons les hypothèses suivantes sur l'observateur (f, L) :

(FC) (Complétude en temps positif.) Pour tout $u \in C^\infty(\mathbb{R}_+, \mathbb{R})$, le champ de vecteur temps-variant $f(\cdot, u)$ est complet en temps positif. De plus, pour tout $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathcal{K} \times \mathbb{S}^{n-1}$ et tout $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$ borné sur $\mathcal{D}(\lambda)$, le système en boucle fermée (2.3) a une unique solution $(\hat{x}, \varepsilon, \xi, \omega) \in C^\infty(\mathbb{R}_+, \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n \times \mathbb{R}^n)$ définie sur $[0, +\infty)$.

(NFOT) (Aucune trajectoire plate de l'observateur.) Pour tout $R > 0$, il existe $\eta > 0$ tel que pour tout $\delta \in \mathcal{V}_R$ vérifiant $\sup\{|\delta(x)| \ : \ x \in \mathcal{K}_x\} < \eta$, et pour tout $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0) \in \mathcal{K} \times \mathbb{S}^{n-1}$ tel que $(\hat{x}_0, \varepsilon_0) \neq (0, 0)$, il existe un entier positif k tel que la solution de (2.3) initialisée à $(\hat{x}_0, \varepsilon_0, \xi_0, \omega_0)$ vérifie $\hat{x}^{(k)}(0) \neq 0$.

En particulier, nous montrons que les observateurs de Luenberger et Kalman vérifient ces deux hypothèses, de sorte que nos résultats s'appliquent à ces observateurs. Pour tout $k \in \mathbb{N}$, $K \subset \mathbb{R}^n$ et $\delta \in C^\infty(\mathbb{R}^n, \mathbb{R})$, définissons

$$\|\delta\|_{k,K} = \sup\left\{ \left| \frac{\partial^\ell \delta}{\partial x_{i_1} \cdots \partial x_{i_\ell}}(x) \right| \ : \ 0 \leqslant \ell \leqslant k, \quad 1 \leqslant i_1 \leqslant \cdots \leqslant i_\ell \leqslant n, \quad x \in K \right\}.$$

Pour tout $k \in \mathbb{N}$, tout compact $K \subset \mathbb{R}^n$ et tout $\eta > 0$, $k \in \mathbb{N}$, posons

$$\mathscr{N}(k, K, \eta) = \left\{ \delta \in C^\infty(\mathbb{R}^n, \mathbb{R}) \ : \ \|\delta\|_{k,K} < \eta \right\}.$$

Le problème auquel nous nous attaquons est le suivant.

Problem D.1. Soit $T > 0$. Sous des hypothèses génériques sur (A, B, C), existe-t-il $R, \eta > 0$, un entier positif k et un ensemble résiduel $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ tels que la propriété suivante soit vérifiée : pour tout $\delta \in \mathscr{O} \cap \mathcal{V}_R$ et toute condition initiale $(\hat{x}_0, \varepsilon_0, \xi_0) \in \mathcal{K}$, le système (2.1) est observable en temps T pour le contrôle $u = (\lambda + \delta) \circ \hat{x}$, où $\hat{x}$ suit la dynamique de (2.3) avec la conditions initiale $(\hat{x}_0, \varepsilon_0, \xi_0)$ et la perturbation δ ?

Principaux résultats d'observabilité

Nous énonçons d'abord notre résultat principal, qui traite de l'observabilité de (2.3). Sa preuve est la partie la plus technique du chapitre, et repose essentiellement sur des outils de la théorie de la transversalité.

Théorème 2.2. *Supposons que les paires (C, A) et (C, B) sont observables. Supposons de plus que $0 \notin \mathcal{K}_x$. Alors il existe $\eta > 0$, un entier positif k et un ouvert dense (dans la topologie C^∞ de Whitney) $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ tels que la solution de (2.3) avec $\delta \in \mathscr{O}$ partant d'une condition initiale $(\hat{x}(0), \varepsilon(0), \xi(0), \omega(0)) \in \mathcal{K} \times \mathbb{S}^{n-1}$ vérifie*

$$\exists k_0 \in \{0, \ldots, k\} \quad : \quad \left. \frac{\mathrm{d}^{k_0}}{\mathrm{d}t^{k_0}} \right|_{t=0} C\omega(t) \neq 0. \tag{2.4}$$

La propriété (2.4) est plus forte que l'observabilité de (2.3) en tout temps positif. L'hypothèse $0 \notin \mathcal{K}_x$ est supprimée dans le corollaire Corollaire 2.3, en affaiblissant légèrement le résultat d'observabilité.

Corollaire 2.3. *Supposons que les paires (C, A) et (C, B) sont observables. Supposons de plus que 0 est dans l'intérieur de $\mathcal{K}_x$. Soit $T > 0$. Alors il existe $\eta > 0$, un entier positif k et un ouvert dense (dans la topologie C^∞ de Whitney) $\mathscr{O} \subset \mathscr{N}(k, \mathcal{K}_x, \eta)$ tels que la solution de (2.3) avec $\delta \in \mathscr{O}$ partant d'une condition initiale $(\hat{x}(0), \varepsilon(0), \xi(0), \omega(0)) \in \mathcal{K} \times \mathbb{S}^{n-1}$ vérifie*

$$\exists t \in [0, T] \quad : \quad C\omega(t) \neq 0,$$

c'est-à-dire que le système (2.1) est observable en temps T pour le contrôle $u = (\lambda + \delta) \circ \hat{x}$, où $\hat{x}$ suit la dynamique (2.3) avec la condition initiale $(\hat{x}_0, \varepsilon_0, \xi_0)$ et la perturbation δ.

Ce résultat implique un propriété d'observabilité générique directement sur la loi de commande stabilisante λ.

Corollaire 2.4. *Supposons que les paires (C, A) et (C, B) sont observables. Supposons de plus que 0 est dans l'intérieur de $\mathcal{K}_x$. Notons Λ l'ensemble des lois de commande $\lambda \in C^\infty(\mathbb{R}^n, \mathbb{R})$ telles que 0 est un point d'équilibre localement asymptotiquement stable du champ de vecteur $x \mapsto A_{\lambda(x)}x + b\lambda(x)$. Soit $T > 0$ et $\Lambda_T \subset \Lambda$ l'ensemble des lois de commande $\lambda \in \Lambda$ telles que (2.1) est observable en temps T pour le contrôle $u = \lambda \circ \hat{x}$,où $\hat{x}$ suit la dynamique (2.3) avec la condition initiale $(\hat{x}_0, \varepsilon_0, \xi_0)$ et la perturbation nulle $\delta \equiv 0$. Alors Λ_T est un sous-ensemble ouvert dense de Λ.*

Ce dernier corollaire est un pas important dans la direction d'un principe de séparation générique pour les systèmes bilinéaires SISO. En effet, il énonce que si un système est stabilisable par bouclage d'état statique, alors génériquement sur la loi de commande et sur le système, les entrées produites par la boucle fermée rendent le système observable.

Comme $\mathcal{V}_R$ n'est pas un ouvert de la topologie C^∞ de Whitney, l'ensemble $\mathscr{O}$ défini dans le Corollaire 2.3 n'est pas, lui non plus, un ouvert de cette topologie. En revanche, c'est bien un ouvert de la topologie induite sur $\mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_R$. De plus, l'ensemble des matrices $(A, B, C) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{1\times n}$ telles que (C, A) et (C, B) sont deux paires observables est ouvert et dense. Ainsi, cette hypothèse est générique. Contrairement à la stratégie mise en place dans [LSG17] ou dans le Chapitre 4 sur des exemples spécifiques, les résultats de ce chapitre ne proposent pas une synthèse explicite de la perturbation δ, mais montrent plutôt que pour presque tout système bilinéaire SISO, presque toute perturbation $\delta \in \mathscr{N}(k, \mathcal{K}_x, \eta) \cap \mathcal{V}_R$ permet de rendre le système observable.

Enfin, le théorème suivant montre que les observateurs usuels de Luenberger et de Kalman vérifient les hypothèses (FC) et (NFOT). Ainsi, nos résultats peuvent être appliqués sur ces observateurs.

Théorème 2.5. *Supposons que (C, A) est observable et λ est borné sur $\mathcal{D}(\lambda)$. Soit $Q \in \mathcal{S}_n$. Pour tout $\xi \in \mathcal{S}_n$ et tout $u \in \mathbb{R}$, considérons les observateurs suivants :*

$$f^{\text{Luenberger}}(\xi, u) = 0 \qquad \text{(observateur de Luenberger)}$$
$$f_Q^{\text{Kalman}}(\xi, u) = \xi A_u' + A_u\xi + Q - \xi C'C\xi \qquad \text{(observateur de Kalman)}$$

et $L(\xi) = \xi C'$. Alors le système en boucle fermée (2.3) donné par (f, L) satisfait les hypothèses (FC) et (NFOT) pour tout $f \in \{f^{\text{Luenberger}}, f_Q^{\text{Kalman}}\}$.

Chapitre 3 : Systèmes dissipatifs

Résumé. *La distance entre deux trajectoires d'un même système dissipatif affine en l'état est toujours décroissante. Dans ce chapitre, nous montrons que cette propriété se révèle être un outil puissant dans le contexte de la stabilisation par bouclage de sortie. Contrairement au chapitre précédent, nous ne supposons pas le point cible observable, et nous ne faisons pas usage de stratégies de perturbation de la loi de commande. Nous montrons que la condition de* 0*-détectabilité est nécessaire et suffisante à la mise en place d'un principe de séparation sur les systèmes dissipatifs. La preuve est basée sur un observateur de Luenberger à petit gain. Les résultats sont appliqués à un convertisseur Ćuk et un échangeur de chaleur. Des simulations numériques sont proposées.*

Introduction

Dans le Chapitre 2, nous avons montré sur les systèmes bilinéaires SISO que l'observabilité à la cible est une condition suffisante pour mettre en place une stratégie de perturbation de la loi de commande garantissant l'observabilité du système en boucle fermée. Toutefois, cette stratégie ne permet pas, à elle seule, de réaliser la stabilisation par bouclage de sortie dynamique. La principale difficulté réside dans la potentielle non-bornitude des trajectoires du système en boucle fermée, que l'ajustement du gain de l'observateur ne permet pas de résoudre. Pour lutter contre ce phénomène, le cas des systèmes à trajectoires bornées pourrait être étudié. En particulier, si le système bilinéaire est de la forme $\dot{x} = (A+uB)x$ avec $A+uB$ un matrice dissipative pour tout $u \in \mathbb{R}$, alors toutes les trajectoires du système sont bornées. Ainsi, on peut appliquer la stratégie de perturbation du Chapitre 2, et montrer que la stabilisation par bouclage de sortie dynamique est possible. Cependant, l'objectif de ce chapitre est de montrer que pour de tels systèmes, la stratégie de perturbation de la loi de commande est superflue. En effet, une condition nécessaire et suffisante à la stabilisation par bouclage de sortie dynamique des systèmes dissipatifs est la 0-détectabilité, qui est plus faible que l'observabilité à la cible.

Problématique

Définition 3.1 (Systèmes affines en l'état)**.** Un système de contrôle est dit *affine en l'état* s'il est de la forme

$$\dot{x} = A(u)x + b(u) \tag{3.1}$$

où $x \in \mathbb{R}^n$ est l'état du système, $u \in C^0(\mathbb{R}_+, \mathcal{U})$ est l'entrée, $\mathcal{U} \subset \mathbb{R}^p$ est l'ensemble des contrôles admissibles et $A : \mathcal{U} \to \mathbb{R}^{n\times n}$ et $b : \mathcal{U} \to \mathbb{R}^n$ sont des fonctions continues.

En particulier, les systèmes bilinéaires considérés dans le Chapitre 2 sont affines en l'état. Notons que dans ce chapitre, le système n'est plus supposé SISO.

Définition 3.2 (Systèmes dissipatifs)**.** Le système affine en l'état (3.1) est dit *dissipatif* sur un ensemble de contrôles admissible $\mathcal{U} \subset \mathbb{R}^p$ s'il existe une matrice symétrique définie-positive $P \in \mathbb{R}^{n\times n}$ telle que pour tout $u \in \mathcal{U}$,

$$PA(u) + A(u)'P \leqslant 0. \tag{3.2}$$

De nombreux systèmes physiques satisfont cette propriété de dissipativité. C'est par exemple le case des systèmes entrée-état-sortie port-Hamiltoniens (voir, par exemple, [SJ+14]). La clé de cette propriété réside dans la proposition suivante, qui énonce que la distance (dans la métrique associée à P) entre deux trajectoires subissant le même contrôle est décroissante.

Proposition 3.3. *Soient x_1 et x_2 deux solutions du système dissipatif* (3.1) *partageant la même entrée $u : \mathbb{R}_+ \to \mathcal{U}$. Alors $t \mapsto (x_1(t) - x_2(t))'P(x_1(t) - x_2(t))$ est décroissant.*

L'objectif de ce chapitre est de montrer l'intérêt de la dissipativité dans le contexte de la stabilisation par bouclage de sortie.

Principaux résultats sur les systèmes dissipatifs

Soit $n, m, p \in \mathbb{N}$. Soient $A : \mathbb{R}^p \to \mathbb{R}^{n\times n}$ et $B : \mathbb{R}^p \to \mathbb{R}^n$ deux fonctions localement lipschitziennes, et $C \in \mathbb{R}^{m\times n}$. Pour tout $u \in C^0(\mathbb{R}_+, \mathbb{R}^p)$, considérons le système suivant :

$$\begin{cases} \dot{x} = A(u)x + B(u) \\ y = Cx \end{cases} \tag{3.3}$$

où x est l'état du système, u est l'entrée et y est la sortie.

Théorème 3.4. *Supposons que* (3.3) *soit* 0*-détectable et localement stabilisable par bouclage d'état statique. Soit $\mathcal{D}(\lambda)$ le bassin d'attraction d'une loi de commande stabilisante par bouclage d'état λ. Supposons de plus que λ est localement lipschitzienne. Si* (3.3) *est dissipatif sur un l'ensemble admissible $\mathcal{U} = \lambda(\mathcal{D}(\lambda))$, alors il est globalement stabilisable par bouclage de sortie dynamique.*

De plus, le bouclage de sortie est donnée par l'observateur de Luenberger à gain dynamique suivant :

$$\begin{cases} \dot{x} = A(\lambda(x))x + B(\lambda(x)) \\ \dot{\hat{x}} = A(\lambda(\hat{x}))\hat{x} + B(\lambda(\hat{x})) - \alpha(\hat{x}, C\varepsilon)P^{-1}C'C\varepsilon \end{cases} \tag{3.4}$$

avec $\hat{x}(0) \in \mathcal{D}(\lambda)$ et α une certaine fonction localement lipschitzienne.

En contraignant l'observateur de Luenberger à garder un gain constant, il est toujours possible d'obtenir un résultat semi-global.

Théorème 3.5. *Supposons que* (3.3) *soit* 0*-détectable et localement stabilisable par bouclage d'état statique. Soit $\mathcal{D}(\lambda)$ le bassin d'attraction d'une loi de commande stabilisante par bouclage d'état λ. Supposons de plus que λ est localement lipschitzienne. Si* (3.3) *est dissipatif sur un l'ensemble admissible $\mathcal{U} = \lambda(\mathcal{D}(\lambda))$, alors pour tout compact $\mathcal{K}_x \times \mathcal{K}_{\hat{x}} \subset \mathbb{R}^n \times \mathcal{D}(\lambda)$, il existe $\alpha_0 > 0$ tel que pour tout $\alpha \in (0, \alpha_0)$, $(0, 0)$ est un point d'équilibre localement asymptotiquement stable de*

$$\begin{cases} \dot{x} = A(\lambda(x))x + B(\lambda(x)) \\ \dot{\hat{x}} = A(\lambda(\hat{x}))\hat{x} + B(\lambda(\hat{x})) - \alpha P^{-1}C'C\varepsilon \end{cases} \tag{3.5}$$

avec un bassin d'attraction contenant $\mathcal{K}_x \times \mathcal{K}_{\hat{x}}$.

Exemples d'applications

Nous proposons deux applications du Théorème 3.5 (stabilisation semi-globale). Remarquons d'abord que si $A(u) = (J(u) - R(u))\mathcal{H}$, avec $\mathcal{H}$ une matrice symétrique définie-positive et $R(u)$ (resp. $J(u)$) une matrice symétrique semi-définie-positive (resp. antisymétrique), Best linéaire et $C = B'\mathcal{H}$, alors (3.3) est un système entrée-état-sortie port-Hamiltoniens (voir [SJ+14]). Dans ce cas, un bouclage de sortie statique stabilisant peut être implémenté en définissant $u = -ky$ pour $k > 0$. Cependant, pour un système ayant la même dynamique, mais une sortie différente (*i.e.* $C \neq B'\mathcal{H}$), notre résultat propose une stratégie de stabilisation par bouclage de sortie globale ou semi-globale dès lors que la paire $(C, A(0))$ est détectable. Les deux exemples traités sont de cette forme. Soit $x^\star$ le point cible visé par le système, et $u^\star$ la valeur du contrôle à la cible. Après un changement de coordonnées, les systèmes considérés s'écrivent sous la forme suivante.

Exemple 3.6 (Convertisseur Ćuk)**.**

$$A(\bar{u}) = \begin{pmatrix} 0 & -(1-u^\star-\bar{u}) & 0 & 0 \\ 1-u^\star-\bar{u} & 0 & u^\star+\bar{u} & 0 \\ 0 & -u^\star-\bar{u} & 0 & -1 \\ 0 & 0 & 1 & -\frac{1}{R} \end{pmatrix} P,$$

$$B(\bar{u}) = \bar{u}b \text{ avec } b = \begin{pmatrix} C_2 x_2^\star \\ L_3 x_3^\star - L_1 x_1^\star \\ -C_2 x_2^\star \\ 0 \end{pmatrix} \text{ et } C = \begin{pmatrix} 0,\ 1,\ 0,\ 0 \end{pmatrix}.$$

Exemple 3.7 (Échangeur de chaleur)**.**

$$A(\bar{u}) = \begin{pmatrix} -k\mathrm{Id}_3 + \gamma_1(u^\star+\bar{u})J & k\mathrm{Id}_3 \\ k\mathrm{Id}_3 & -k\mathrm{Id}_3 + \gamma_2 J' \end{pmatrix},\ B(\bar{u}) = \bar{u}b$$

$$\text{avec } b = \begin{pmatrix} E - \gamma_1 x_1^\star,\ \gamma_1(x_1^\star - x_2^\star),\ \gamma_1(x_2^\star - x_3^\star),\ 0,\ 0,\ 0 \end{pmatrix}',$$

et $C = \begin{pmatrix} 0,\ 0,\ 0,\ 1,\ 0,\ 0 \end{pmatrix}$ où Id_3 est la matrice identité 3×3, k, γ_1, γ_2, E sont des constantes physiques positives, et

$$J = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Nous montrons que ces systèmes ne sont pas uniformément observables, mais que les hypothèses des résultats principaux de ce chapitre sont vérifiées. Ces systèmes sont dissipatifs, stabilisables par bouclage d'état statique, et 0-détectables. Dès lors, nous appliquons le Théorème 3.5 pour garantir la stabilisation semi-globale vie l'observateur de Luenberger donné par (3.5), où α est une constante positive suffisamment petite. Des simulations numériques sont proposées, et les résultats sont comparés pour différentes valeurs de α.

Chapitre 4 : Cible inobservable

Résumé. *Nous considérons le problème de la stabilisation par bouclage de sortie dynamique à une cible inobservable. La difficulté réside dans l'apparente contradiction entre l'objectif à réaliser et les propriétés du système. En effet, le système tend à être de moins en moins observable à mesure que l'état se rapproche de la cible. Nous illustrons deux idées principales : des perturbations bien choisies de la loi de commande stabilisante par retour d'état peuvent engendrer de nouvelles propriétés d'observabilité du système en boucle fermée, et des plongements dans des systèmes admettant des observateurs à système d'erreur dissipatif permettent de compenser les singularités d'observabilité. Nous les appliquons à des systèmes à dynamique linéaire et sortie non-linéaire, et proposons une stratégie* ad hoc *de plongement en dimension finie. Plus généralement, nous proposons une nouvelle stratégie basée sur des plongements unitaires en dimension infinie. Pour cela, nous étendons la définition usuelle de la stabilisation par bouclage de sortie dynamique afin d'autoriser l'usage d'observateurs de dimension infinie. En particulier, nous tirerons parti d'observateurs de Luenberger de dimension infinie étudiés dans le Chapitre 5. Nous montrons comment cette technique, basée sur la théorie des représentations, peut s'avérer utile dans le cadre de la stabilisation par bouclage de sortie à une cible inobservable.*

Introduction

Stabiliser l'état d'un système à un point cible inobservable est un problème survenant dans des systèmes d'ingénierie pratiques, pour lesquels des stratégies originales ont été explorées ces dernières années [HPR14, Com+16, Fla19, Sur+19, Aja+20, RD20, Sur+20, AGS21], conduisant à un regain d'intérêt pour le sujet. La difficulté réside dans l'apparente contradiction existant entre les objectifs d'estimation et de stabilisation : à mesure que l'état du système approche la cible, les propriétés d'observabilité s'amenuisent, ce qui détériore l'estimation de l'état par l'observateur, et donc empêche potentiellement la stabilisation.

Des méthodes générales, basées sur des lois de commande temps-variantes, ont été développées pour traiter les entrées singulières. Mentionnons l'article [Cor94a] dans lequel J.-M. Coron propose une méthode de stabilisation locale avec un bouclage de sortie dynamique temps-variant périodique, sous une hypothèse d'observabilité à zéro de Lie. Une stratégie similaire, dite "sample-and-hold", a été mise en place dans [ST03] pour parvenir à la stabilisation pratique semi-globale. De plus, des méthodes de perturbation de l'entrée avec des signaux haute-fréquences [Com+16, Sur+19, Sur+20] ou des bruits stochastiques [Fla19] ont été étudiées et permettent d'améliorer les propriétés d'observabilité du système sans contrarier l'objectif de stabilisation. Dans [LSG17], les auteurs introduisent une perturbation explicite de la loi de commande sur un exemple spécifique de système bilinéaire issu du contrôle quantique. Cette idée nous a guidé dans le Chapitre 2 pour trouver des perturbations génériques dans le cas des systèmes observables à la cible. Cette stratégie à l'avantage de conserver le caractère autonome du système en boucle fermée, ce qui est intéressant pour certaines applications en ingénierie. Dans ce chapitre, nous utilisons à nouveau de telles perturbations autonomes pour engendrer une meilleure observabilité du système. Mais contrairement au Chapitre 2, elles sont synthétisées de manière explicite.

Un autre outil qui s'avère important en théorie de la stabilisation est l'utilisation de systèmes ayant des flots non-expansifs, tels que les systèmes dissipatifs étudiés dans le Chapitre 3 et les systèmes faiblement contractant de l'Appendice D. En effet, dans [LSG17], la stratégie de perturbation de la loi de commande était utilisée en conjonction avec la propriété de contraction du système de contrôle pour démontrer la stabilisation par bouclage de sortie à la cible inobservable. La propriété principale utilisée est que l'erreur de l'observateur d'état est décroissante quelque soit l'observabilité du système. Par conséquent, l'estimation de l'état ne se détériore pas à mesure qu'il s'approche de la cible.

Dans ce chapitre, nous rassemblons les idées développées dans les Chapitres 2 et 3 afin de tracer des lignes directrices en vue de la résolution du problème de la stabilisation à une cible inobservable. Nous considérons essentiellement des systèmes à dynamique linéaire conservative et observation non-linéaire. Un exemple élémentaire de J.-M. Coron dans [Cor94a] illustre la raison pour laquelle certains systèmes ne peuvent être stabilisés par bouclage de sortie dynamique autonome, même localement. En revanche, nous montrons comment des exemples similaires, partageant les mêmes propriétés d'inobservabilité, peuvent toutefois être stabilisés.

Afin de bénéficier des propriétés des systèmes dissipatifs, nous nous intéressons à des techniques de plongement. Dans [Cel+89], les auteurs proposent une méthode de synthèse d'observateur basée sur des plongements unitaires en dimension infinie. Nous réutilisons cette approche dans le contexte de la stabilisation par bouclage de sortie, ce qui nous amène à coupler le système non-linéaire de départ de dimension finie avec un observateur dissipatif de dimension infinie. De façon intéressante, ajouter un état de dimension infinie permet d'abolir les obstructions topologiques identifiées dans[Cor94a]. Nous illustrons la stratégie sur un exemple qui, nous l'espérons, pourra ouvrir la voie à des résultats plus généraux à l'avenir.

Un exemple éclairant en dimension finie

Considérons le cas où le système (1.1) est SISO et f est une application linéaire, de sortie qu'il se réécrit sous la forme :

$$\begin{cases} \dot{x} = Ax + bu, \\ y = h(x). \end{cases} \tag{4.1}$$

où $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^{n\times 1}$ et $h : \mathbb{R}^n \to \mathbb{R}$. Si h est non-linéaire et n'est pas une transformation inversible d'une application linéaire, alors la théorie usuelle des systèmes linéaires ne s'applique pas. La Condition 1.2 se réduit à la stabilizability de la paire (A, b). Si elle est vérifiée, alors (4.1) est globalement asymtptoqieuement stabilisable par une bouclage d'état statique linéaire.

Dans [Cor94a], J.-M. Coron propose l'exemple unidimensionnel suivant :

$$\dot{x} = u, \qquad y = x^2. \tag{4.2}$$

il montre que (4.2) n'est pas localement stabilisable par bouclage de sortie dynamique, à moins de considérer des lois de commande temps-variantes. La difficulté de ce système vient de l'inobservabilité du point cible 0. En effet, (4.2) n'est pas observable pour l'entrée constante $u \equiv 0$, puisque les conditions initiales x_0, $-x_0$ $\in \mathbb{R}$ sont indistinguables.

En particulier, le système n'est pas uniformément observable, et les résultats de [TP94, JG95, TP95] ne s'appliquent donc pas. Pour résoudre ce problème, [Cor94a] considère des lois de commande par bouclage de sortie temps-variantes, et démontrer ainsi la stabilisabilité locale. Ce système peut également être stabilisé par les méthodes dites "dead-beat" ou "sample-and-hold" (voir [NS98] et [ST03], respectivement).

Une généralisation possible de (4.2) en dimension plus élevée est :

$$\begin{cases} \dot{x} = Ax + bu, \\ y = h(x) \end{cases} \tag{4.3}$$

pour une matrice antisymétrique A et une fontion d'observation radialement symétrique[1] h. À nouveau, l'entrée constante $u \equiv 0$ rend le système inobservable en tout temps $T > 0$ puisque deux conditions initiales x_0, $\tilde{x}_0$ dans $\mathbb{R}^n$ telles que $|x_0| = |\tilde{x}_0|$ sont indistinguables par la mesure. La Condition 1.2 (global) se réduit à la stabilisabilité de la paire (A, b) et la Condition 1.3 (global) est toujours satisfaite. En adaptant le résultat de [Cor94a], on obtient également la condition nécessaire suivante, due à une obstruction topologique.

Théorème 4.1. *Si* (4.3) *est localement stabilisable par retour de sortie dynamique, alors A est inversible.*

En particulier, le théorème spectral induit le corollaire suivant.

Corollaire 4.2. *Si n est impair et A antisymétrique, alors* (4.3) *n'est pas localement stabilisable par retour de sortie dynamique.*

L'un des principaux résultat de ce chapitre est la réciproque de ce théorème dans le cas où $h(x) = \frac{1}{2}|x|^2$. La preuve se base sur les grands principes donnés en introduction : nous plongeons le système dans un système bilinéaire admettant un observateur à système d'erreur dissipatif et ajoutons une perturbation de la loi de commande.

Théorème 4.3. *Si A est antisymétrique et inversible et (A, b) est stabilisable, alors*

$$\begin{cases} \dot{x} = Ax + bu, \\ y = h(x) = \dfrac{1}{2}|x|^2. \end{cases} \tag{4.3'}$$

est semi-globalement stabilisable par retour de sortie dynamique.

Nous proposons une simulation numérique du bouclage de sortie dynamique synthétisé dans la preuve de ce théorème.

Une nouvelle perspective en dimension infinie

Inspirés par la stratégie de plongement mise en place sur l'exemple (4.3'), nous proposons une méthode plus générale basée sur les mêmes outils. Malheureusement, le plongement utilisé apparaît comme trop dépendant du lien entre le système et

[1] Quitte à changer de produit scalaire, on peut également considérer le cas où $PA + AP = 0$ pour une matrice symétrique définie-positive $P \in \mathbb{R}^{n\times n}$ et h tel que $(x_1' P x_1 = x_2' P x_2) \Rightarrow (h(x_1) = h(x_2))$.

sa fonction d'observation, et donc difficilement généralisable. De nouvelles techniques doivent être explorées. Dans [Cel+89], les auteurs utilisent des plongements, construits à base de représentations unitaires de groupes, dans des systèmes bilinéaires dans le contexte de la synthèse d'observateurs. L'approche est bien plus générale que celle proposée dans l'exemple (4.3'), mais le prix à payer est que l'observateur obtenu peut être de dimension infinie. Nous réutilisons ces observateurs avec pour objectif la stabilisation à une cible inobservable. Nous montrons d'abord quelques résultats généraux, avant de considérer à nouveau des systèmes à dynamique linéaire et sortie non-linéaire. Les outils utilisés relient la théorie des observateurs de Luenberger de dimension infinie étudiés dans la Partie II et le problème de stabilisation par bouclage de sortie dynamique de la Partie I.

L'objectif est de plonger le système de départ (1.1) dans un système de la forme

$$\begin{cases} \dot{z} = A(u(t))z \\ \mathfrak{y} = \mathcal{C}z. \end{cases} \tag{4.4}$$

où $A(u) : \mathcal{D} \to X$ est un opérateur anti-adjoint défini sur le domaine $\mathcal{D}$ dense dans le Hilbert X générant un système d'évolution bidirectionnel unitaire $(\mathbb{T}_t(\cdot, u))_{t\in\mathbb{R}_+}$ et $C \in \mathscr{L}(X, \mathbb{C}^{\mathfrak{m}})$ pour un certain entier positif $\mathfrak{m}$. On définit donc la notion de plongement de la façon suivante.

Définition 4.4 (Plongement)**.** Une application injective $\tau : \mathbb{R}^n \mapsto X$ est un plongement de (1.1) dans le système unitaire (4.4) s'il existe une application $\mathfrak{h} : \mathbb{R}^m \to \mathbb{C}^{\mathfrak{m}}$ telle que le diagramme suivant soit commutatif pour tout $t \in \mathbb{R}_+$ et tout $u \in C^1(\mathbb{R}_+, \mathbb{R}^p)$:

$$\begin{array}{ccccccc} \mathbb{R}^n & \xrightarrow{\varphi_t(\cdot,u)} & \mathbb{R}^n & \xrightarrow{h} & \mathbb{R}^m & \xrightarrow{\mathfrak{h}} & \mathbb{C}^{\mathfrak{m}} \\ {\scriptstyle\tau}\downarrow & & \downarrow{\scriptstyle\tau} & & & \nearrow_{\mathcal{C}} & \\ X & \xrightarrow[\mathbb{T}_t(\tau(\cdot),u)]{} & X & & & & \end{array} \tag{4.5}$$

i.e., , pour tout $x_0 \in \mathbb{R}^n$, $\tau(\varphi_t(x_0, u)) = \mathbb{T}_t(\tau(x_0), u)$ et $\mathfrak{h}(h(x_0)) = \mathcal{C}\tau(x_0)$.

Si un tel plongement existe, on construit ensuite :

- un observateur de Luenberger de dimension infinie de (4.4) dont l'erreur est décroissante ;
- un inverse à gauche de τ étendu sur X ;
- une perturbation de la loi de commande améliorant l'observabilité du système en boucle fermée.

Nous montrons la stabilisation semi-globale par bouclage de sortie de (4.3') pour une classe de fonctions de sortie h satisfaisant des hypothèses d'observabilité du système de départ et du système plongé, et garantissant l'existence du plongement. En particulier, nous montrons que les fonctions h telles que

$$\mathfrak{h}(h(r\cos(\theta), r\sin(\theta))) = \sum_{k\in I} c_k J_k(\mu r) e^{-ik\theta}$$

pour un certain $\mathfrak{h}$, où J_k est la k-ième fonction de Bessel de première espèce, satisfont les hypothèses énoncées.

Chapitre 5 : Observateurs asymptotiques de Luenberger

Résumé. *Dans ce chapitre, nous considérons le problème d'estimation en ligne de l'état d'un système dynamique linéaire temps-variant de dimension infinie à partir de la mesure d'une sortie linéaire. Nous analysons les propriétés de convergence des observateurs de Luenberger pour ces systèmes. Nous rappelons certaines notions fondamentales sur les systèmes d'évolution et plusieurs notions d'observabilité. Tandis que la convergence forte exponentielle de l'observateur peut être espérée pour les systèmes exactement observables, les résultats portant sur des hypothèses d'observabilité approchée, sur lesquelles nous nous focalisons, sont plus rares. Sous une hypothèse de détectabilité faible, nous démontrons que l'observateur permet d'estimer la partie dite observable du système, au moins dans la topologie faible de l'espace d'état. Des hypothèses supplémentaires sur le système d'erreur permettent de démontrer la convergence forte. Nous illustrons les résultats sur une équation de transport.*

Introduction

Pour analyser, surveiller ou contrôler un système physique ou biologique, la première étape est d'établir un modèle mathématique décrivant l'évolution des différentes variables au cours du temps. Certaines de ces variables sont accessibles par des mesures ; d'autres ne le sont pas. L'un des problèmes de la théorie du contrôle est la synthèse d'algorithmes capables d'estimer en temps réel les variables non-mesurées à partir des autres, et convergeant asymptotiquement vers les données réelles. Ces algorithmes d'estimations sont appelés *observateurs*, et sont couramment utilisés dans de nombreux domaines. La théorie des observateurs linéaires de dimension finie, initiée par les travaux [Lue64, Lue71] de D. Luenberger, est maintenant bien connue. Mais la synthèse d'observateur pour des systèmes non-linéaires et/ou de dimension infinie reste un enjeu majeur de la théorie du contrôle, et l'extension des travaux de D. Luenberger à ces systèmes est un domaine actif de recherche.

Dans ce chapitre, nous considérons des systèmes de dimension infinie linéaires temps-variants, ayant une mesure de dimension (potentiellement) infinie. Nous rappelons d'abord quelques notions élémentaires sur la théorie des semi-groupes et les systèmes d'évolution (principalement issus de [Paz83]) et énonçons le cadre fonctionnel de notre étude. Nous définissons les observateurs de Luenberger considérés, basés sur une extension usuelle des observateurs de diension finie ([Sle72, Sle74, Cel+89, XLG95, Liu97]). Les notions d'observabilité approchée et exacte sont définies, et coïncident avec [TW09] dans le contexte autonome. Une large littérature est consacrée aux systèmes autonomes exactement observables, pour lesquels la convergence forte exponentielle de l'observateur se vérifie en général (voir par exemple [Liu97, Theorem 2.3], qui harmonise et étend des résultats précédemment connus). Les systèmes temps-variant approximativement observables, voire à sous-espace observable non-plein, sont quant à eux moins étudiés. Nous montrons, en étendant un résultat de [Cel+89], la convergence faible de l'observateur sur la partie observable de l'état, sous une hypothèse de détectabilité faible. De plus, sous des hypothèses supplémentaires sur le système d'erreur, la convergence forte peut être montrée en utilisant des outils différents inspirés par [Hai14].

Observateur de Luenberger

Soient X et Y deux espaces de Hilbert réels. Nous considérons le système :

$$\begin{cases} \dot{z} = A(t)z, & t \in \mathbb{R}_+ \\ y = Cz. \end{cases} \tag{5.1}$$

où z dans X est l'état du système, y dans Y est la sortie, $C \in \mathscr{L}(X,Y)$ est un opérateur linéaire borné et $(A(t))_{t\in\mathbb{R}_+}$ est une famille d'opérateurs non-bornés de domaine $\mathcal{D}$ dense dans X et à valeur dans X générant un système d'évolution $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t}$ sur X sur $\mathbb{R}_+$. Nous considérons le problème d'estimation de l'état z de (5.1) à partir de la mesure y. Soit $z_0 \in X$. Notons (z,y) l'unique solution correspondante de (5.1) Nous cherchons à synthétiser un nouveau système dynamique apprenant z à partir de y et sa dynamique. Nous proposons l'observateur de Luenberger usuellement utilisé en dimension infinie. Soit $r > 0$ et $\hat{z}_0 \in X$. L'observateur est donnée par :

$$\begin{cases} \dot{\hat{z}} = A(t)\hat{z} - rC^*(C\hat{z} - y), \\ \hat{z}(0) = \hat{z}_0. \end{cases} \tag{5.2}$$

Le paramètre r est le gain de l'observateur. Posons $\varepsilon = \hat{z} - z$ et $\varepsilon_0 = \hat{z}_0 - z_0$. La variable $\hat{z}$ représente l'estimation de l'état faite par l'observateur, et ε l'écart entre cette estimation et l'état réel, de sorte qu'il vérifie

$$\begin{cases} \dot{\varepsilon} = (A(t) - rC^*C)\varepsilon, \\ \varepsilon(0) = \varepsilon_0. \end{cases} \tag{5.3}$$

Notons $(\mathbb{S}(t,s))_{0\leqslant s\leqslant t}$ le système d'évolution généré par la famille d'opérateurs $(A(t) - rC^*C)_{t\geqslant 0}$. Les solutions $\hat{z}$ et ε dans $C^0([0,+\infty);X)$ de (5.2) et (5.3) satisfont $\hat{z}(t) = \mathbb{T}(t,0)z_0 + \mathbb{S}(t,0)\varepsilon_0$ and $\varepsilon(t) = \mathbb{S}(t,0)\varepsilon_0$ pour tout $t \in [0,+\infty)$.

Nous nous intéressons aux propriétés de convergence de l'estimation $\hat{z}$ vers l'état z, *i.e.,* de l'erreur ε vers 0. Pour tout sous-espace vectoriel fermé $\mathcal{O}$ de X, on note $\Pi_{\mathcal{O}} \in \mathscr{L}(X)$ la projection orthogonale sur $\mathcal{O}$.

Définition 5.1 (Observateur asymptotique)**.** Pour tout sous-espace vectoriel fermé $\mathcal{O}$ de X, (5.2) est un $\mathcal{O}$-observateur asymptotique fort (resp. faible) de (5.1) si et seulement si $\Pi_{\mathcal{O}}\mathbb{S}(t,0)\varepsilon_0 \to 0$ (resp. $\Pi_{\mathcal{O}}\mathbb{S}(t,0)\varepsilon_0 \overset{w}{\to} 0$) quand $t \to +\infty$ pour tout $\varepsilon_0 \in X$. Un X-observer est simplement appelé un observateur.

Grammien d'observabilité

Le Grammien d'observabilité est un opérateur dont il est crucial d'analyser les propriétés pour étudier la convergence des observateurs de Luenberger.

Définition 5.2 (Grammien d'observabilité)**.** Pour tout $t_0, \tau \in \mathbb{R}_+$, définissons

$$\begin{aligned} W(t_0,\tau) : X &\longrightarrow X \\ z_0 &\longmapsto \int_{t_0}^{t_0+\tau} \mathbb{T}(t,t_0)^* C^* C \mathbb{T}(t,t_0) z_0 \mathrm{d}t, \end{aligned}$$

le *Grammien d'observabilité* de la paire $(\mathbb{T}, C)$.

L'opérateur $W(t_0, \tau)$ est un endomorphisme borné auto-adjoint de X, qui caractérise les proprités d'observabilité de (5.1). Dans le contexte autonome, $W(t_0, \tau) = W(0, \tau)$ pour tout $t_0, \tau \in \mathbb{R}_+$. On notera alors $W(\tau) := W(0, \tau)$.

Définition 5.3 (Sous-espace observable)**.** pour tout $\tau \in \mathbb{R}_+$, soit

$$\mathcal{O}_\tau = (\ker W(0, \tau))^\perp . \tag{5.4}$$

le *sous-espace observable* au temps τ de la paire $(\mathbb{T}, C)$. De plus, soit

$$\mathcal{O} = \overline{\bigcup_{\tau>0} \mathcal{O}_\tau}. \tag{5.5}$$

le *sous-espace observable* de la paire $(\mathbb{T}, C)$.

La suite $(\mathcal{O}_\tau)_{\tau>0}$ est une suite décroissante de sous espaces fermés. Donc $\mathcal{O} = \overline{\lim_{\tau\to+\infty} \mathcal{O}_\tau}$ peut être interprété comme le *sous-espace observable* en temps infini de la paire $(\mathbb{T}, C)$.

Lorsque (5.1) est autonome et que X et Y sont de dimension finie, on retrouve la définition usuelle (basé sur la matrice d'observabilité), les propriétés, et la caractérisation par le test d'Hautus, de l'espace observable :

$$\forall \tau \geqslant 0, \quad \mathcal{O}_\tau = \mathcal{O} = \left(\bigcap_{k=0}^{\dim X - 1} \ker CA^k \right)^\perp = \left(\operatorname{span} \bigcup_{\lambda \in \sigma(A)} \ker C \cap \ker(A - \lambda \mathrm{Id}) \right)^\perp .$$

En dimension infinie, plusieurs concepts d'observabilité coexistent (voir notamment [TW09, Chapter 6]). En particulier, on distingue les deux notions suivantes.

Définition 5.4 (Observabilité exacte)**.** La paire $((A(t))_{t\in[0,T]}, C)$ est dite exactement observable sur l'intervalle de temps $(t_0, t_0 + \tau) \subset [0, T]$ s'il existe $\delta > 0$ tel que

$$\langle W(t_0, \tau) z_0, z_0 \rangle_X \geqslant \delta \left\| z_0 \right\|_X^2 , \qquad \forall z_0 \in X. \tag{5.6}$$

Définition 5.5 (Observabilité approchée)**.** La paire $((A(t))_{t\in[0,T]}, C)$ est dite approximativement observable sur l'intervalle de temps $(t_0, t_0 + \tau) \subset [0, T]$ si $W(t_0, \tau)$ est injectif.

Clairement, l'observabilité exacte implique l'observabilité approchée, elles sont équivalentes en dimension finie. De plus, l'observabilité approchée sur $(0, \tau)$ est équivalente à $\mathcal{O}_\tau = X$.

Convergence de l'observateur

Nos résultats se concentrent sur des hypothèses d'observabilité approchée, et reposent sur l'hypothèse suivante.

Définition 5.6 (Détectabilité faible)**.** Soit $T \in \mathbb{R}_+ \cup \{+\infty\}$. Alors $((A(t))_{t\in[0,T]}, C)$ est dit *μ-faiblement détectable* pour $\mu \geqslant 0$ si pour tout $t \in [0, T]$,

$$\langle A(t) z, z \rangle_X \leqslant \mu \left\| Cz \right\|_Y^2 , \qquad \forall z \in \mathcal{D}. \tag{5.7}$$

On obtient sous cette condition les deux résultats principaux de ce chapitre.

Théorème 5.7. *Supposons que $((A(t))_{t\geqslant 0}, C)$ est μ-faiblement détectable et $r > \mu$. Supposons qu'il existe une suite positive strictement croissante $(t_n)_{n\geqslant 0} \to +\infty$ et un système d'évolution $(\mathbb{T}_\infty(t,s))_{0\leqslant s\leqslant t}$ sur X tel que pour tout $\tau \geqslant 0$,*

$$\|\mathbb{T}(t_n + t, t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \to 0 \text{ quand } n \to +\infty \text{ uniformément en } t \in [0,\tau], \tag{5.8}$$

Soit $\mathcal{O}$ le sous espace observable de la paire $(\mathbb{T}_\infty, C)$. Alors pour tout $\varepsilon_0 \in X$,

$$\Pi_{\mathcal{O}}\mathbb{S}(t_n,0)\varepsilon_0 \xrightarrow[n\to+\infty]{w} 0. \tag{5.9}$$

De plus, si $(t_{n+1} - t_n)_{n\geqslant 0}$ est borné et $\mathcal{O} = X$, alors (5.2) *est un observateur asymptotique faible de* (5.1).

Dans le contexte autonome, toute suite $(t_n)_{n\geqslant 0} \to +\infty$ est telle que $\mathbb{T}(t_n+t, t_n) = \mathbb{T}(t)$ pour tout $t \geqslant 0$. Donc (5.9) est vérifiée pour toute suite $(t_n)_{n\geqslant 0}$ avec $\mathcal{O}$ le sous-espace observable de $(\mathbb{T}, C)$. Cette remarque conduit au corollaire suivant

Corollaire 5.8. *Supposons que* (5.1) *est autonome, (A, C) est μ-faiblement détectable et $r > \mu$. Soit $\mathcal{O}$ le sous-espace observable de $(\mathbb{T}, C)$. Alors,* (5.2) *est un $\mathcal{O}$-observateur asymptotique faible de* (5.1).

Sous une condition supplémentaire sur le système, la convergence forte de l'observateur est obtenue.

Théorème 5.9. *Supposons qu'il existe $\tau > 0$ tel que $t \mapsto A(t)$ est τ-périodique. Soit $\mathcal{O}_\tau$ la partie observable en temps τ de la paire $(\mathbb{T}, C)$.*

(i) Supposons que $((A(t))_{t\geqslant 0}, C)$ est μ-faiblement détectable et $r > \mu$. Supposons que $\mathbb{S}(\tau, 0)$ est normal et intérieurement borné. Si $\mathcal{O}_\tau = X$, alors (5.2) *est un observateur asymptotique fort de* (5.1).

(ii) Supposons que $A(t)$ est anti-adjoint pour tout $t \in \mathbb{R}_+$ et que $\mathbb{S}(\tau, 0)$ est normal. Si $\mathbb{T}(t,0)\mathcal{O}_\tau \subset \mathcal{O}_\tau$ et $\mathbb{T}(t,0)\mathcal{O}_\tau^\perp \subset \mathcal{O}_\tau^\perp$ pour tout $t \in [0,\tau]$, alors (5.2) *est un $\mathcal{O}_\tau$-observateur asymptotique fort de* (5.1) *pour tout $r > 0$.*

Chapitre 6 : Back and Forth Nudging

Résumé. *Dans ce chapitre, nous considérons le problème d'estimation hors-ligne de la condition initiale d'une système dynamique linéaire temps-variant de dimension infinie à partir de la mesure d'une sortie linéaire sur un intervalle de temps fini. Nous utilisons l'algorithme Back and Forth Nudging (BFN), qui se base sur des itérations d'observateurs asymptotiques en temps positif et en temps négatif pour apprendre l'état à partir de la sortie. Nous rappelons quelques notions sur les systèmes d'évolution bidirectionnels, et ré-employons les observateurs de Luenberger du Chapitre 6. Comme dans le contexte asymptotique, nous démontrons sous une hypothèse de détectabilité faible, nous démontrons que l'observateur permet d'estimer la partie dite observable du système, au moins dans la topologie faible de l'espace d'état. Des hypothèses supplémentaires sur le système d'erreur permettent de démontrer la convergence forte.*

Introduction

Lorsque seule une partie de l'état d'un système dynamique de dimension infinie est mesurée sur un intervalle de temps fini, une question importante est celle de l'estimation de la condition initiale de l'état à partir de la connaissance de la mesure sur l'intervalle de temps. Ce problème apparaît par exemple en océanographie ou en météorologie (voir les problèmes d'estimation de donnée de [AB05, AB08, Aur09]) ainsi qu'en génie des procédés (voir le Chapitre 7). Bien que de nombreuses techniques issues des problèmes inverses puissent être appliquées, l'algorithme Back and Forth Nudging (BFN) (également appelé algorithme par inversion temporelle dans [IRT11]) se montre particulièrement efficace dans ce contexte, du fait de sa forte utilisation de la dynamique du système. Il repose sur la théorie des observateurs de Luenberger en dimension infinie (voir le Chapitre 7). Mais puisque le temps d'observation est fini, les observateurs asymptotiques doivent être adaptés. Pour les systèmes admettant à la fois une évolution en temps positif et en temps négatif, il est possible de simuler la dynamique du système en temps rétrograde, et donc de synthétiser un observateur rétrograde. Dès lors, l'idée est d'utiliser itérativement des observateurs en sens positif et en sens négatif, travaillant sur le même intervalle borné, et utilisant la même mesure. Après chaque itération, l'estimation finale de l'état obtenue par l'observateur est utilisée comme condition initiale de l'itération suivante de l'observateur. Cette méthodologie conduit à l'observateur dit « back and forth » décrit dans ce chapitre.

Dans le contexte autonome, des résultats de convergence forte ont été obtenus, à la fois pour les systèmes exactement [RTW10,IRT11] et approximativement [HR11a, HR11b, Hai14] observables. En adaptant les résultats asymptotiques obtenus dans le Chapitre 5, nous montrons sous des hypothèses moins restrictives la convergence faible de l'algorithme BFN, et étendons les résultats de [Hai14] au contexte temps-variant.

Observateur « back and forth »

Nous considérons le problème d'estimation de la condition initiale z_0 de (5.1) à partir de la mesure de y sur l'intervalle de temps fini $[0, T]$. On suppose que $(A(t))_{t\in[0,T]}$ est

le générateur d'un système d'évolution bidirectionnel sur X sur $[0,T]$. La méthode proposée est celle du BFN, basées sur les travaux de [AB05, AB08, IRT11, AN12].

Soit $\hat{z}_0 \in X$. Pour chaque $n \in \mathbb{N}$,on considère le système dynamique sur $[0,T]$ défini comme dans [RTW10] par

$$\begin{cases} \dot{\hat{z}}^{2n} = A(t)\hat{z}^{2n} - rC^*(C\hat{z}^{2n} - y) \\ \hat{z}^{2n}(0) = \begin{cases} \hat{z}^{2n-1}(0) & \text{si } n \geqslant 1 \\ \hat{z}_0 & \text{sinon.} \end{cases} \end{cases} \tag{6.1}$$

$$\begin{cases} \dot{\hat{z}}^{2n+1} = A(t)\hat{z}^{2n+1} + rC^*(C\hat{z}^{2n+1} - y) \\ \hat{z}^{2n+1}(T) = \hat{z}^{2n}(T). \end{cases} \tag{6.2}$$

Le système (6.1) est l'observateur asymptotique usuel de (5.1) (voir (5.2)), tandis que (6.2) peut être vu comme un observateur de Luenberger de (5.1) en temps rétrograde. En effet, $\hat{z}^{2n+1}(t)$ vérifie (6.1) si et seulement si $\hat{z}_r^{2n+1}(t) := \hat{z}^{2n+1}(T-t)$ vérifie

$$\begin{cases} \dot{\hat{z}}_r^{2n+1} = -A(T-t)\hat{z}_r^{2n+1} - rC^*(C\hat{z}_r^{2n+1} - y(T-t)) \\ \hat{z}_r^{2n+1}(0) = \hat{z}^{2n}(T). \end{cases}$$

Ainsi, le système couplé (6.1)-(6.2) avec $n \in \mathbb{N}$ est un itération d'observateurs en temps positifs et négatifs. La valeur finale de l'estimation obtenue après une itération sert de condition initiale à l'itération suivante.

Soit $\varepsilon_0 = \hat{z}_0 - z_0$ et $\varepsilon^n = \hat{z}^n - z$ pour tout $n \in \mathbb{N}$. Alors $\hat{z}^{2n}$ et $\hat{z}^{2n+1}$ satisfont respectivement (6.1) et (6.2) si et seulement si ε^{2n} et ε^{2n+1} sont solutions de

$$\begin{cases} \dot{\varepsilon}^{2n} = (A(t) - rC^*C)\varepsilon^{2n} \\ \varepsilon^{2n}(0) = \begin{cases} \varepsilon^{2n-1}(0) & \text{si } n \geqslant 1 \\ \varepsilon_0 & \text{sinon.} \end{cases} \end{cases} \tag{6.3}$$

$$\begin{cases} \dot{\varepsilon}^{2n+1} = (A(t) + rC^*C)\varepsilon^{2n+1} \\ \varepsilon^{2n+1}(T) = \varepsilon^{2n}(T). \end{cases} \tag{6.4}$$

Notons $(\mathbb{S}_+(t,s))_{0\leqslant s,t\leqslant T}$ et $(\mathbb{S}_-(t,s))_{0\leqslant s,t\leqslant T}$ les systèmes d'évolution bidirectionnels engendrés respectivement par $(A(t) - rC^*C)_{t\in[0,T]}$ et $(A(t) + rC^*C)_{t\in[0,T]}$. Les solutions $\hat{z}^{2n}$, $\hat{z}^{2n+1}$, ε^{2n} et ε^{2n+1} dans $C^0([0,T];X)$ de (6.1), (6.2), (6.3) et (6.4) vérifient $\hat{z}^{2n}(t) = \mathbb{T}(t,0)z_0 + \mathbb{S}_+(t,0)\varepsilon^{2n}(0)$, $\hat{z}^{2n+1}(t) = \mathbb{T}(t,T)z(T) + \mathbb{S}_-(t,T)\varepsilon^{2n+1}(T)$, $\varepsilon^{2n}(t) = \mathbb{S}_+(t,0)\varepsilon^{2n}(0)$ et $\varepsilon^{2n+1}(t) = \mathbb{S}_-(t,T)\varepsilon^{2n+1}(T)$ pour tout $t \in [0,T]$. En particulier,

$$\varepsilon^{2n}(0) = (\mathbb{S}_-(0,T)\mathbb{S}_+(T,0))^n \varepsilon_0. \tag{6.5}$$

Nous nous intéressons aux propriétés de convergence de l'estimation $\hat{z}^{2n}(0)$ vers la condition initiale réelle $z(0)$,*i.e.,* de l'erreur $\varepsilon^{2n}(0)$ vers 0 quand n tend vers l'infini.

Rappelons que pour tout sous-espace vectoriel fermé $\mathcal{O}$ de X, on note $\Pi_{\mathcal{O}} \in \mathscr{L}(X)$ la projection orthogonale sur $\mathcal{O}$.

Définition 6.1 (Observateur « back and forth »)**.** Pour tout sous-espace vectoriel fermé $\mathcal{O}$ de X, le système (6.1)-(6.2) est un $\mathcal{O}$-observateur « back and forth » fort (resp. faible) de (5.1) si et seulement si $\Pi_{\mathcal{O}}\varepsilon^{2n}(0) \to 0$ (resp. $\Pi_{\mathcal{O}}\varepsilon^{2n}(0) \overset{w}{\to} 0$) quand $n \to +\infty$ pour tout $\varepsilon_0 \in X$. Un X-observateur est simplement appelé un observateur.

Convergence de l'observateur

Nos résultats reposent sur la même hypothèse de détectabilité faible que dans le chapitre précédent, et utilisent les mêmes techniques de preuve.

Théorème 6.2. *Supposons que $(\mathbb{T}(t,s))_{0\leqslant s,t\leqslant T}$ est un système d'évolution bidirectionnel. Supposons que $((A(t))_{t\in[0,T]}, C)$ et $((-A(t))_{t\in[0,T]}, C)$ sont μ-faiblement détectable et $r > \mu$. Soit $\mathcal{O}_T$ le sous-espace observable au temps T de la paire $(\mathbb{T}, C)$. Alors le système* (6.1)-(6.2) *est un $\mathcal{O}_T$-observateur faible « back and forth » de* (5.1).

Sous une condition supplémentaire sur le système, la convergence forte de l'observateur est obtenue.

Théorème 6.3. *Supposons que $(\mathbb{T}(t,s))_{0\leqslant s,t\leqslant T}$ est un système d'évolution bidirectionnel. Soit $\mathcal{O}_T$ le sous-espace observable au temps T de la paire $(\mathbb{T}, C)$. Supposons que $((A(t))_{t\in[0,T]}, C)$ et $((-A(t))_{t\in[0,T]}, C)$ sont μ-faiblement détectable et $r > \mu$. Supposons de plus que $\mathbb{S}_-(0,T) = \mathbb{S}_+(T,0)^*$. Si $\mathcal{O}_T = X$, alors le système* (6.1)-(6.2) *est un observateur « back and forth » fort de* (5.1).

Application à une équation de transport

Nous appliquons nos résultats à une équation de transport temps-variante unidimensionnelle avec des conditions aux bords périodiques. Plus précisément, soient $x_1 > x_0 \geqslant 0$ et $X = L^2((x_0, x_1); \mathbb{R})$ l'ensemble des fonctions à valeurs réelles de carré intégrable sur (x_0, x_1), muni du produit scalaire $\langle f, g\rangle_X = \int_{x_0}^{x_1} fg$, pour $f, g \in X$. Soit $\mathcal{D} = \{\psi \in X \mid \psi(x_0) = \psi(x_1), \psi' \in X\}$ et $G \in C^1([0,T]; \mathbb{R})$. Pour tout $t \geqslant 0$, soit

$$\begin{aligned} A(t) : \mathcal{D} &\longrightarrow X \\ \psi &\longmapsto -G(t)\frac{\mathrm{d}\psi}{\mathrm{d}x}. \end{aligned}$$

Alors $A(t)$ est anti-adjoint pour tout $t \geqslant 0$ et $(A(t))_{t\geqslant 0}$ engendre un système d'évolution unitaire sur X noté $(\mathbb{T}(t,s))_{0\leqslant s\leqslant t}$ vérifiant

$$(\mathbb{T}(t,s)z_0)(x) = z_0(v(x,t,s)), \tag{6.6}$$

où

$$v(x,t,s) = x_0 + \left(\left(x - x_0 - \int_s^t G(\tau)\mathrm{d}\tau\right) \bmod (x_1 - x_0)\right) \tag{6.7}$$

pour presque tout $x \in (x_0, x_1)$.

Ainsi, quelque soit $T \in \mathbb{R}_+ \cup \{+\infty\}$, l'espace de Hilbert Y et l'opérateur de sortie $C \in \mathscr{L}(X, Y)$, les paires $((A(t))_{t\in[0,T]}, C)$ et $((-A(t))_{t\in[0,T]}, C)$ sont 0-faiblement détectables. L'équation de transport $\dot{z} = A(t)z$ est donc une bonne candidate pour appliquer la théorie développée dans les Chapitres 5 et 6. De plus, la proposition suivante est utile pour vérifier les hypothèses du théorème 5.7.

Proposition 6.4. *Supposons que G et G' sont bornés. S'il existe $G_\infty \in C^1(\mathbb{R}_+, \mathbb{R})$ et une suite positive strictement croissante $(t_n)_{n\geqslant 0} \to +\infty$ telle que $G(t_n + t) \to G_\infty(t)$ quand $n \to +\infty$ pour tout $t \geqslant 0$, alors $\|\mathbb{T}(t_n+t, t_n) - \mathbb{T}_\infty(t,0)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$ uniformément en $t \in [0,\tau]$ pour tout $\tau \geqslant 0$, où $\mathbb{T}_\infty$ est le système d'évolution engendré par $\left(-G_\infty(t)\frac{\mathrm{d}}{\mathrm{d}x}\right)_{t\geqslant 0}$.*

En particulier, si G est périodique, alors G et G' sont bornés et il existe une suite bornée $(t_n)_{n\geqslant 0}$ et une constante $G_\infty > 0$ tels que $\|\mathbb{T}(t_n + t, t_n) - \mathbb{T}_\infty(t, 0)\|_{\mathscr{L}(X)} \to 0$ as $n \to +\infty$ uniformément en $t \in [0, \tau]$ pour tout $\tau \geqslant 0$.

Nous analysons également différents opérateurs de sortie $C \in \mathscr{L}(X, Y)$. D'abord, si le noyau de C satisfait une certaine condition géométrique, alors le noyau du Grammien d'observabilité est directement lié à celui de C. En effet, supposons qu'il existe $U \subset [x_0, x_1]$ tel que

$$\ker C = \{\psi \in X \mid \psi|_U = 0\}, \tag{6.8}$$

où $f|_U$ est la restriction de f à U.

Alors $z_0 \in \ker W(t_0, \tau)$ pour $t_0, \tau \geqslant 0$ si et seulement si

$(\mathbb{T}(s, t_0)z_0)\,|_U = 0$ pour presque tout $s \in (t_0, t_0 + \tau)$, *i.e.,* $z_0(v(x, s, t_0)) = 0$ pour presque tout $s \in (t_0, t_0 + \tau)$ et $x \in U$. Donc

$$\ker W(t_0, \tau) = \{\psi \in X \mid \psi|_{U_{\max}} = 0\} \tag{6.9}$$

avec $U_{\max} = \{v(x, s, t_0), x \in U, s \in [t_0, t_0 + \tau]\}$. De plus,

$$\ker W(t_0, \tau)^\perp = \left\{\psi \in X \;\middle|\; \psi|_{[x_0,x_1]\setminus U_{\max}} = 0\right\}. \tag{6.10}$$

Cette remarque conduit à la proposition suivante, qui affirme que si τ est suffisamment grand pour faire parcourir à toute la donnée initiale la fenêtre d'observation $[x_{\min}, x_{\max}]$, alors le système est approximativement observable en temps τ.

Proposition 6.5. *Supposons que* $\ker C \subset \left\{\psi \in X \;\middle|\; \psi|_{[x_{\min},x_{\max}]} = 0\right\}$ *pour un certain intervalle* $[x_{\min}, x_{\max}] \subset [x_0, x_1]$. *S'il existe* $t_0, \tau \geqslant 0$, *tels que*

$$\left|\int_{t_0}^{t_0+\tau} G(t)\mathrm{d}t\right| \geqslant (x_1 - x_0) - (x_{\max} - x_{\min}), \tag{6.11}$$

alors $\ker W(t_0, \tau) = \{0\}$.

Supposons maintenant que $C \in \mathscr{L}(X, Y)$ est un opérateur intégral à noyau borné, c'est-à-dire qu'il existe $k \in L^\infty((x_0, x_1); Y)$ tel que

$$C\psi = \int_{x_0}^{x_1} k(x)\psi(x)\mathrm{d}x \tag{6.12}$$

pour tout $\psi \in X$. Alors il n'existe aucun intervalle $(t_0, t_0 + \tau) \subset \mathbb{R}_+$ tel que que la paire $((A(t))_{t\geqslant 0}, C)$ est exactement observable sur $(t_0, t_0 + \tau)$.

Proposition 6.6. *Si* $C \in \mathscr{L}(X, Y)$ *est de la forme* (6.12) *pour un certain* $k \in L^\infty((x_0, x_1); Y)$, *alors pour tout* $t_0, \tau \geqslant 0$ *et tout* $\delta > 0$, *il existe* $z_0 \in X$ *tel que*

$$\langle W(t_0, \tau)z_0, z_0\rangle_X \leqslant \delta \|z_0\|_X^2. \tag{6.13}$$

Ainsi, pour de tels opérateurs, la convergence d'un observateur doit reposer sur des hypothèses d'observabilité plus faible, comme l'observabilité approchée. Dans l'application de ces résultats à un procédé de cristallisation (voir Chapitre 7), le lecteur remarquera que C est précisément un opérateur intégral à noyau borné.

Chapitre 7 : Observateurs et procédés de cristallisation

Résumé. *Durant un procédé de cristallisation par lots, la distribution en taille des particules (PSD) est d'une importance capitale. Cependant, mesurer la PSD est difficile, et une approche populaire consiste à l'estimer à partir d'autres mesures. Dans ce chapitre, nous en considérons principalement trois : la température, la concentration en soluté, et le distribution en taille des cordes (CLD). Après avoir modélisé le procédé et les capteurs physiques, nous proposons différentes stratégies d'estimation. D'abord, une approche directe basée sur une procédure de régularisation de Tikhonov utilisant la CLD, mais indépendante du modèle dynamique de la PSD. Ensuite, un observateur de Kazantzis-Kravaris/Luenberger (KKL) utilisant uniquement comme mesures la température et la concentration en soluté. Enfin, un observateur de Luenberger de dimension infinie utilisant la CLD basé sur la théorie développée dans les Chapitres 5 et 6, également efficace lorsque des phénomènes de polymorphisme ont lieu au cours du procédé.*

Introduction

La cristallisation est l'un des procédés les plus anciens utilisés dans l'industrie (chimique, pharmaceutique, agro-alimentaire, *etc.*) pour produire, purifier ou séparer des produits ou composés solides [Bis13]. Cette opération unitaire a pour objectif de produire des cristaux solides aux spécifications précises incluant (parmi d'autres) la distribution en taille des cristaux (PSD), qui est d'une importance critique. À l'échelle industrielle, la PSD n'est difficilement contrôlable au cours du procédé de cristallisation, et une étape de broyage/tamisage est généralement requise avant l'obtention du produit final. Les technologies d'analyse des procédés (PATs) modernes proposent des mesures et des techniques variées pour reconstruire le PSD, telles que l'analyse d'image [Pre+10, Gao+18], des observateurs dynamiques et des méthodes basées sur la mesure des moments [Mes+11, Ucc11, Vis12, Leb+15, Gru17, PÖ17]. Certaines PATs, telles que la "Focused Beam Reflectance Measurement" FBRM ou la technologie BlazeMetrics®, donnent accès à la distribution en taille des cordes CLD [LCK98, WHM05, Agi+15, PR16].

Dans ce chapitre, nous considérons le problème de reconstruction de la PSD à partir de trois mesures : la température, la concentration en soluté, et la CLD. Nous proposons d'abord un modèle dynamique du procédé à partir d'un bilan de population, ainsi qu'un modèle des mesures, qui dépendent de la forme des cristaux formés. Une reconstruction directe de la PSD à partir des mesures, basées sur des méthodes de problèmes inverses, est envisagée dans un premier temps. Mais cette approche n'utilise pas la connaissance de la dynamique du système, et échoue dans de nombreuses situations. Nous développons donc dans un second temps deux observateurs d'états. Les problèmes d'estimation en ligne ou hors-ligne sont tous les deux considérés. En particulier, nous utilisons la théorie développée dans les Chapitres 5 et 6 pour montrer la convergence d'un observateur estimant la PSD à partir de la mesure de la CLD.

Modélisation du procédé

Dans un premier temps, nous établissons une modélisation du procédé de cristallisation par lots dans le cas de cristaux dont la taille est décrite par un paramètre scalaire r et ayant tous la même forme (par exemple, des cristaux sphériques de rayon r). Notons $\psi(t,\cdot)$ la PSD au temps t dans le réacteur, de sorte que $\int_{r_1}^{r_2}\psi(t,r)\mathrm{d}x$ est le nombre de cristaux dans le réacteur au temps t ayant un rayon r entre r_1 et r_2. Soit $r_{\max}$ un rayon maximal que les cristaux ne peuvent atteindre au cours du procédé (par exemple la taille du réacteur) :

$$\psi(t,r_{\max})=0, \qquad \forall t\in[0,t_{\max}]. \tag{7.1}$$

Supposons que tous les cristaux se forment à la même taille minimale $r_{\min}>0$, et notons $u(t)$ l'apparition des cristaux de taille $r_{\min}$ au temps t :

$$\psi(t,r_{\min})=u(t), \qquad \forall t\in[0,t_{\max}]. \tag{7.2}$$

La fonction u est liée au taux de germination R et au taux de croissance G par la relation :

$$u(t)=\frac{R_n(t)}{G(t)}. \tag{7.3}$$

Notons cependant que nos approches d'estimation de la PSD ne reposent aucunement sur cette expression. Nous n'utilisons aucun modèle de u, et supposons cette quantité inconnue.

Le taux de croissance est supposé positif à tout instant. De plus, sous l'hypothèse de McCabe, G est indépendant de la taille des cristaux. Le bilan de population conduit finalement à (voir [MEH01, Mul01]) :

$$\frac{\partial\psi}{\partial t}(t,r)+G(t)\frac{\partial\psi}{\partial r}(t,r)=0, \tag{7.4}$$

i.e., une équation de transport unidimensionnelle temps-variante. Enfin, nous supposons que des particules de germe de PSD ψ_0 peuvent se trouver dans le réacteur à $t=0$:

$$\psi(0,r)=\psi_0(r), \qquad \forall r\in[r_{\min},r_{\max}]. \tag{7.5}$$

Pour résumer, l'évolution de la PSD au cours du procédé suit l'EDP

$$\begin{cases} \frac{\partial\psi}{\partial t}(t,r)+G(t)\frac{\partial\psi}{\partial r}(t,r)=0 & \forall t\in(0,t_{\max}),\forall r\in(r_{\min},r_{\max}) \\ \psi(0,r)=\psi_0(r) & \forall r\in[r_{\min},r_{\max}] \\ \psi(t,r_{\min})=u(t) & \forall t\in[0,t_{\max}] \end{cases} \tag{7.6}$$

avec la condition au bord additionnelle (7.1).

On assure le caractère bien posé du problème avec le théorème suivant.

Théorème 7.1. *Si G est une fonction continue à valeurs strictement positive, $\psi_0\in L^2((r_{\min},r_{\max});\mathbb{R})$ et $u\in L^2((0,t_{\max});\mathbb{R})$, alors* (7.6) *admet une unique solution $\psi\in C^0([0,t_{\max}];L^2((r_0,r_1);\mathbb{R}))$.*

De plus, pour tout $t\in[0,t_{\max}]$ et presque tout $r\in[r_{\min},r_{\max}]$,

$$\psi(t,r)=\begin{cases} \psi_0(r-\mathfrak{G}(t)) & \text{si } r-r_{\min}\geqslant\mathfrak{G}(t) \\ u\circ\mathfrak{G}^{-1}(\mathfrak{G}(t)-r+r_{\min}) & \text{sinon.} \end{cases} \tag{7.7}$$

où $\mathfrak{G} : [0, t_{\max}] \ni t \mapsto \int_0^t G(\tau)\mathrm{d}\tau$.

De plus, si $\psi_0 \in H^1((r_{\min}, r_{\max}); \mathbb{R})$, $u \in H^1((0, t_{\max}); \mathbb{R})$ *et* $u(0) = \psi_0(r_{\min})$, *alors*

$$\psi \in C^0([0, t_{\max}]; H^1(r_{\min}, r_{\max})) \cap C^1([0, t_{\max}]; L^2(r_{\min}, r_{\max})).$$

Au cours d'un procédé de cristallisation, le polymorphisme est un phénomène courant : des cristaux peuvent avoir différentes formes stables ou métastables. Nous supposons qu'un nombre fini N de telles formes peuvent coexister dans le réacteur, et que la taille d'un cristal de forme $i \in \{1, \dots, N\}$ est toujours décrite par un unique scalaire r. En notant ψ_i la PSD associée à chaque forme i, et en raisonnant comme précédemment, on obtient le modèle d'évolution suivant, où les différentes forment n'interagissent pas entre elles :

$$\forall i \in \{1, \dots, N\}, \begin{cases} \frac{\partial \psi_i}{\partial t}(t, r) + G_i(t)\frac{\partial \psi_i}{\partial r}(t, r) = 0 & \forall t \in (0, t_{\max}), \forall r \in (r_{\min}, r_{\max}) \\ \psi_i(0, r) = \psi_{0,i}(r) & \forall r \in [r_{\min}, r_{\max}] \\ \psi_i(t, r_{\min}) = u_i(t) & \forall t \in [0, t_{\max}] \end{cases} \tag{7.8}$$

Modélisation des mesures

Une mesure de la température et de la concentration en soluté permettent de connaître le taux de croissance $G(t)$ des cristaux et le troisième moment de la PSD donné par

$$y(t) = \int_{r_{\min}}^{r_{\max}} \psi(t, r) r^3 \mathrm{d}r. \tag{7.9}$$

Nous construirons un observateur de KKL basé sur ces deux mesures. La troisième mesure considérée dans ce chapitre est la CLD. Les technologies FBRM et BlazeMertics® sont des sondes *in situ* qui mesurent des données au cours du procédé. La sonde est équipée d'un laser en rotation qui balaye les particules. Lorsque le rayon frappe un cristal, de la lumière est rétro-diffusée en direction de la sonde. Un capteur compte le nombre et la durée des impulsions lumineuses reçues. À chaque impulsion correspond une longueur sur la particule, c'est-à-dire une longueur de corde, qui peut être estimée puisque la vitesse de rotation du laser est connue. Ainsi, on en déduit la distribution en taille des cordes (CLD). Le lecteur trouvera les détails de cette technologie et son lien avec la CLD dans [BG99, SLB99, LW05]. Estimer la PSD à partir de la CLD est un enjeu moderne en génie des procédés. Dans un premier temps, la relation existant entre ces deux données, qui dépend fortement de la forme des particules, doit être modélisée. Dans [Hob+91, BG99, Lan+01], les auteurs considèrent des particules sphériques. Mais les cristaux ont rarement des formes présentant une telle symétrie. Dans [Agi+15] par exemple, les cristaux sont modélisés par des cylindres allongés. Dans cette thèse, nous proposons d'approcher la forme des cristaux par des sphéroïdes (également appelés ellipsoïdes de révolution). Un sphéroïde est une surface de révolution, obtenue par rotation d'une ellipse autour de l'un de ses deux axes principaux. En particulier, les sphères sont des sphéroïdes. Ces formes ont l'avantage de permettre de modéliser à la fois des sphères et des particules allongées en forme d'aiguilles (fréquentes en cristallisation).

Lorsque le laser de la sonde balaye les particules, le capteur mesure des longueurs de cordes sur la projection de la particule dans le plan orthogonal au laser. Ainsi, deux sources de hasard sont considérées pour modéliser la mesure :

- le choix de l'orientation du sphéroïde par rapport à la sonde ;
- le choix de la corde mesurée sur la projection du sphéroïde dans l'orientation sélectionnée.

Notons η le rapport entre le rapport entre le diamètre du sphéroïde le long de son axe de rotation par le diamètre perpendiculaire à cet axe. Ce paramètre caractérise l'excentricité du sphéroïde. Il est dit allongé si $\eta > 1$, aplati si $\eta < 1$. Le sphéroïde est un sphère lorsque $\eta = 1$.

Soit ψ la PSD associée à une famille de sphéroïdes de paramètre η et de rayon r compris entre $r_{\min}$ et $r_{\max}$, et q la CLD associée. Notons que la plus grande corde mesurable est $\ell_{\max} = 2r_{\max}\max(\eta, 1)$. Ainsi, pour $0 \leqslant \ell \leqslant \ell_{\max}$, $\int_{\ell_1}^{\ell_2}$ représente le nombre de cordes de longueur comprise entre ℓ_1 et ℓ_2 mesurées par la sonde. La CLD cumulée est notée $Q(\ell) = \int_0^\ell q(l)\mathrm{d}l$. Enfin, on définit les fonctions normalisées $\bar{\psi}(r) = \frac{1}{\int_{r_{\min}}^{r_{\max}} \psi(\rho)\mathrm{d}\rho}\psi(r)$ et $\bar{q}(\ell) = \frac{1}{Q(\ell_{\max})}q(\ell)$ qui sont des fonctions de densité et $\bar{Q}(\ell) = \frac{1}{Q(\ell_{\max})}Q(\ell)$ qui est une fonction de répartition.

Par le théorème de l'espérance totale et une modélisation aléatoire uniforme, on obtient la relation suivante entre la PSD et le CLD cumulée :

$$Q(\ell) = \kappa \int_{r_{\min}}^{r_{\max}} k(\ell, r)\psi(r)\mathrm{d}r \tag{7.10}$$

avec

$$\kappa = \frac{Q(\ell_{\max})}{\int_{r_{\min}}^{r_{\max}} \psi(r)\mathrm{d}r},$$

$$k(\ell, r) = 1 - \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \sqrt{1 - \left(\frac{\ell}{2r}\right)^2 \alpha_\eta(\phi, \theta)}\frac{\sin\theta}{4\pi}\mathrm{d}\theta\mathrm{d}\phi, \tag{7.11}$$

et

$$\alpha_\eta(\phi, \theta) = \frac{\cos^2\phi}{\cos^2\theta + \eta^2\sin^2\theta} + \sin^2\phi. \tag{7.12}$$

Approche directe

Nous montrons dans cette section qu'il est possible de reconstruire la PSD à partir de la CLD, à un facteur multiplication près (qui peut être déterminée par la mesure de la concentration en soluté), dans le cas de cristaux sphéroïdaux d'une seule forme. Aucun modèle dynamique n'est utilisé : l'approche est directement basée sur une méthode d'inversion de l'expression (7.10).

Soient $X = L^2((r_{\min}, r_{\max}); \mathbb{R})$ et $Y = L^2((0, \ell_{\max}); \mathbb{R})$, $\ell_{\max} = 2r_{\max}\max(\eta, 1)$. Une PSD (normalisée) peut être vu comme un élément de X, tandis qu'une CLD (normalisée) est un élément de Y. En définissant l'opérateur

$$\begin{aligned} \mathcal{K} : X &\longrightarrow Y \\ \bar{\psi} &\longmapsto \left(\ell \mapsto \textstyle\int_{r_{\min}}^{r_{\max}} k(\ell, r)\bar{\psi}(r)\mathrm{d}r\right), \end{aligned}$$

la relation (7.10) se réécrit

$$\mathcal{K}\bar{\psi} = \bar{Q}. \tag{7.13}$$

Il s'agit dès lors de proposer une méthode d'inversion de $\mathcal{K}$. Nous démontrons le théorème suivant.

Théorème 7.2. *L'opérateur $\mathcal{K}$ est injectif.*

Par conséquent, il est théoriquement possible de reconstruire la PSD à partir de la CLD. Reste à proposer une méthode suffisamment robuste aux erreurs de mesures. En effet, l'opérateur $\mathcal{K}$ est compact. Par conséquent, son inverse ne peut être continu. Nous optons donc pour une méthode de régularisation de Tikhonov. Pour un paramètre de régularisation $\delta > 0$, nous cherchons à résoudre le problème de minimisation suivant :

$$\textit{Trouver } \bar{\psi} \in X \textit{ minimisant } \|\mathcal{K}\bar{\psi} - \bar{Q}\|_Y^2 + \delta\|\bar{\psi}\|_X^2 \textit{ tel que } \bar{\psi} \geqslant 0. \tag{7.14}$$

Lorsque δ tend vers 0, on retrouve le problème d'inversion de départ. Au contraire, quand δ tend vers l'infini, la solution de (7.14) tend vers 0. Le choix de δ est donc un compromis : le problème régularisé doit être suffisamment proche du problème de départ (δ assez petit) pour avoir une solution proche, mais en rester assez éloigné pour garantir la robustesse au bruit de mesure (δ assez grand). Il est sélectionné expérimentalement, à la façon d'un indice de confiance dans la mesure réalisé.

Observateur de KKL

Dans cette section, nous proposons de construire un observateur estimant en ligne la PSD au cours du procédé en utilisant le modèle dynamique du système et la mesure de la température et de la concentration en soluté. Nous proposons une méthode en deux étapes, basée sur l'approche des observateurs de KKL, usuellement utilisés sur les systèmes non-linéaires de dimension finie.

Dans un premier temps, nous cherchons à reconstruire des fonctions $\mathcal{T}_\lambda\psi$ de l'état ψ à estimer.

Proposition 7.3. *Soit $\mathcal{T}_\lambda : C^1([0, t_{\max}]; X) \mapsto C^1([0, t_{\max}]; \mathbb{R})$ l'opérateur défini par*

$$\mathcal{T}_\lambda(\psi) : t \mapsto \int_{r_{\min}}^{r_{\max}} a(t, r)\psi(t, r)\mathrm{d}r \tag{7.15}$$

où a est l'unique solution de

$$\begin{cases} \frac{\partial a}{\partial t}(t, r) + G(t)\frac{\partial a}{\partial r}(t, r) = \lambda a(t, r) + r^3 & \forall t \in (0, t_{\max}),\ \forall x \in (r_{\min}, r_{\max}) \\ a(0, r) = 0 & \forall r \in [r_{\min}, r_{\max}] \\ a(t, r_{\min}) = 0 & \forall t \in [0, t_{\max}]. \end{cases} \tag{7.16}$$

Si ψ est une solution de (7.6) vérifiant (7.1) et z est une solution de

$$\dot{z} = \lambda z + y\ . \tag{7.17}$$

où y est donné par (7.9), alors on a pour tout $t \in [0, t_{\max}]$:

$$\mathcal{T}_\lambda(\psi)(t) - z(t) = \exp(\lambda t)(\mathcal{T}_\lambda(\psi)(0) - z_0). \tag{7.18}$$

Par conséquent, pour chaque $\lambda < 0$, il est possible d'estimer $\mathcal{T}_\lambda\psi$ exponentiellement en simulant le système dynamique (7.17).

Reste ensuite à accomplir la seconde étape de la stratégie KKL : en estimant un nombre suffisant de fonctions $\mathcal{T}_\lambda\psi$, est-il possible d'estimer l'état complet ψ ? Pour répondre à cette question, nous utilisons une fois de plus la méthode de régularisation

de Tikhonov, car les opérateurs $\mathcal{T}_\lambda$ sont à noyau. On obtient finalement la procédure d'estimation suivante :

$$\begin{cases} \dot{z} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix} z + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} y \\ \hat{\psi}(t) = \operatorname{argmin}_{\tilde{\psi} \in X} \left\{ \|\mathcal{T}(\tilde{\psi})(t) - z(t)\|^2 + \delta \|\tilde{\psi}\|^2 \right\}, \quad \delta > 0 \\ \mathcal{T} = \left(\mathcal{T}_{\lambda_1}, \dots, \mathcal{T}_{\lambda_p} \right) \end{cases} \tag{7.19}$$

Observateur de Luenberger

Dans cette dernière section, nous considérons que les cristaux peuvent prendre plusieurs formes sphéroïdales au cours du procédé, et que nous mesurons la CLD sur un intervalle de temps fini. Le taux de croissance de chaque famille de cristaux est supposé connu. Nous cherchons à estimer la PSD associée à chaque forme, à une constante multiplicative près. Comme nous l'avons précédemment, la procédure de régularisation de Tikhonov est efficace dans le cas où les cristaux ont tous la même forme. Mais cette approche ne se généralise pas lorsqu'il y a plusieurs formes. En effet, la CLD est commune à tous les cristaux, et s'exprime sous la forme

$$Q(t, \ell) = \sum_{i=1}^{N} \int_{r_{\min}}^{r_{\max}} k_i(\ell, r) \psi_i(t, r) \mathrm{d}r. \tag{7.20}$$

Définissons l'opérateur

$$\begin{aligned} \mathcal{K} : X^N &\longrightarrow Y \\ \psi &\longmapsto \left(\ell \mapsto \textstyle\sum_{i=1}^{N} \int_{r_{\min}}^{r_{\max}} k_i(\ell, r) \psi_i(r) \mathrm{d}r \right) \end{aligned}$$

et son adjoint

$$\begin{aligned} \mathcal{K}^* : Y &\longrightarrow X^N \\ Q &\longmapsto \left(r \mapsto \textstyle\int_0^{\ell_{\max}} k_i(\ell, r) Q(\ell) \mathrm{d}\ell \right)_{1 \leqslant i \leqslant N} \end{aligned}$$

où $k_i(\ell, r) = 0$ pour $r \notin [r_{\min}, r_{\max}]$ ou $\ell \notin [0, 2 r_{\max} \eta_i]$. L'opérateur $\mathcal{K}$ n'étant pas nécessairement injectif, nous proposons d'utiliser, la dynamique du système et de construire un observateur « back and forth » étudié dans le Chapitre 6. Dans notre contexte, (7.21)-(7.22) s'exprime sous la forme

$$\begin{cases} \dfrac{\partial \hat{\psi}^{2n}}{\partial t}(t, r) = -G(t) \dfrac{\partial \hat{\psi}^{2n}}{\partial r}(t, r) - \mu \mathcal{K}^*(\mathcal{K} \hat{\psi}^{2n}(t, \cdot) - \bar{Q}(t, \cdot)) \\ \hat{\psi}^{2n}(0, r) = \begin{cases} \hat{\psi}^{2n-1}(0, r) & \text{si } n \geqslant 1 \\ \hat{\psi}_0(r) & \text{sinon} \end{cases} \end{cases} \tag{7.21}$$

$$\begin{cases} \dfrac{\partial \hat{\psi}^{2n+1}}{\partial t}(t, r) = -G(t) \dfrac{\partial \hat{\psi}^{2n+1}}{\partial r}(t, r) + \mu \mathcal{K}^*(\mathcal{K} \hat{\psi}^{2n+1}(t, \cdot) - \bar{Q}(t, \cdot)) \\ \hat{\psi}^{2n+1}(t_{\max}, r) = \hat{\psi}^{2n}(t_{\max}, r) \end{cases} \tag{7.22}$$

Sous une hypothèse d'observabilité approchée, les résultats du Chapitre 6 garantissent la convergence de l'observateur. Nous démontrons cette observabilité dans le cas où deux familles de cristaux coexistent dans le réacteur : des cristaux sphériques et des cristaux en forme de sphéroïdes allongés.

Bibliography

[AR67] R. Abraham and J. Robbin. *Transversal mappings and flows*. An appendix by Al Kelley. W. A. Benjamin, Inc., New York-Amsterdam, 1967, pp. x+161.

[Agi+15] O. S. Agimelen, P. Hamilton, I. Haley, A. Nordon, M. Vasile, J. Sefcik, and A. J. Mulholland. "Estimation of particle size distribution and aspect ratio of non-spherical particles from chord length distribution". *Chemical Engineering Science* 123 (2015), pp. 629–640. ISSN: 0009-2509. DOI: 10.1016/j.ces.2014.11.014.

[Aja+20] A. Ajami, M. Brouche, J. .-P. Gauthier, and L. Sacchelli. "Output stabilization of military UAV in the unobservable case". In: *2020 IEEE Aerospace Conference*. 2020, pp. 1–6. DOI: 10.1109/AERO47225.2020.9172770.

[AGS21] A. Ajami, J.-P. Gauthier, and L. Sacchelli. "Dynamic output stabilization of control systems: An unobservable kinematic drone model". *Automatica J. IFAC* 125 (2021), p. 109383. ISSN: 0005-1098. DOI: https://doi.org/10.1016/j.automatica.2020.109383.

[AP09] V. Andrieu and L. Praly. "A unifying point of view on output feedback designs for global asymptotic stabilization". *Automatica J. IFAC* 45.8 (2009), pp. 1789–1798. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2009.04.015.

[AP08] V. Andrieu and L. Praly. "Global asymptotic stabilization for nonminimum phase nonlinear systems admitting a strict normal form". *IEEE Transactions on Automatic Control* 53.5 (2008), pp. 1120–1132. DOI: 10.1109/TAC.2008.923657.

[And05] V. Andrieu. "Bouclage de sortie et observateur". Thèse de doctorat dirigée par Praly, Laurent Mathématiques et automatique Paris, ENMP 2005. PhD thesis. 2005, 1 vol. (231 p.)

[AP06] V. Andrieu and L. Praly. "On the existence of a Kazantzis–Kravaris/Luenberger observer". *SIAM Journal on Control and Optimization* 45 (Feb. 2006), pp. 432–456. DOI: 10.1137/040617066.

[AK01] M. Arcak and P. Kokotovic. "Observer-based control of systems with slope-restricted nonlinearities". *IEEE Transactions on Automatic Control* 46.7 (2001), pp. 1146–1150. DOI: 10.1109/9.935073.

[AK99] A. N. Atassi and H. K. Khalil. "A separation principle for the stabilization of a class of nonlinear systems". *IEEE Transactions on Automatic Control* 44.9 (1999), pp. 1672–1687. ISSN: 0018-9286. DOI: 10.1109/9.788534.

[Aur09] D. Auroux. "The back and forth nudging algorithm applied to a shallow water model, comparison and hybridization with the 4D-VAR". *International Journal for Numerical Methods in Fluids* 61.8 (2009), pp. 911–929. ISSN: 0271-2091. DOI: 10.1002/fld.1980.

[AB08] D. Auroux and J. Blum. "A nudging-based data assimilation method: the Back and Forth Nudging (BFN) algorithm". *Nonlinear Processes in Geophysics* 15.2 (2008), pp. 305–319. DOI: 10.5194/npg-15-305-2008.

[AB05] D. Auroux and J. Blum. "Back and forth nudging algorithm for data assimilation problems". *Comptes Rendus Mathématique. Académie des Sciences. Paris* 340.12 (2005), pp. 873–878. ISSN: 1631-073X. DOI: 10.1016/j.crma.2005.05.006.

[AN12] D. Auroux and M. Nodet. "The back and forth nudging algorithm for data assimilation problems: theoretical results on transport equations". *ESAIM. Control, Optimisation and Calculus of Variations* 18.2 (2012), pp. 318–342. ISSN: 1292-8119. DOI: 10.1051/cocv/2011004.

[Bac90] A. Bacciotti. "Constant feedback stabilizability of bilinear systems". In: *Realization and modelling in system theory: Proceedings of the International Symposium MTNS-89, Volume I*. Ed. by M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran. Boston, MA: Birkhäuser Boston, 1990, pp. 357–367. ISBN: 978-1-4612-3462-3. DOI: 10.1007/978-1-4612-3462-3_40.

[BB91] A. Bacciotti and P. Boieri. "A characterization of single-input planar bilinear systems which admit a smooth stabilizer". *Systems & Control Letters* 16.2 (1991), pp. 139–144. ISSN: 0167-6911. DOI: 10.1016/0167-6911(91)90008-3.

[BG99] P. Barrett and B. Glennon. "In-line FBRM monitoring of particle size in dilute agitated suspensions". *Particle & Particle Systems Characterization* 16.5 (1999), pp. 207–211. DOI: 10.1002/(SICI)1521-4117(199910)16:5<207::AID-PPSC207>3.0.CO;2-U.

[BR86] A. O. Barut and R. Raczka. *Theory of group representations and applications*. Second. World Scientific Publishing Co., Singapore, 1986, pp. xx+717. ISBN: 9971-50-217-8. DOI: 10.1142/0352.

[Ber19] P. Bernard. *Observer design for nonlinear systems*. Vol. 479. Lecture Notes in Control and Information Sciences. Springer, Cham, 2019, pp. xi+187. ISBN: 978-3-030-11146-5. DOI: 10.1007/978-3-030-11146-5.

[BA19] P. Bernard and V. Andrieu. "Luenberger Observers for Nonautonomous Nonlinear Systems". *IEEE Transactions on Automatic Control* 69 (2019), pp. 270–281. DOI: 10.1109/TAC.2018.2872202.

[Ber+17] P. Bernard, L. Praly, V. Andrieu, and H. Hammouri. "On the triangular canonical form for uniformly observable controlled systems". *Automatica J. IFAC* 85 (2017), pp. 293–300. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2017.07.034.

[Bes07] G. Besançon. *Nonlinear observers and applications*. Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 2007. ISBN: 9783540735038.

[BBH96] G. Besançon, G. Bornard, and H. Hammouri. "Observer synthesis for a class of nonlinear control systems". *European Journal of Control* 2.3 (1996), pp. 176–192. ISSN: 0947-3580. DOI: 10.1016/S0947-3580(96)70043-2.

[Bis13] B. Biscans. "Cristallisation en solution - Procédés et types d'appareils". *Techniques de l'ingénieur. Génie des procédés* J2788 v2 (2013), pp. 1–25.

[BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. New York, NY, USA: Cambridge University Press, 2004. ISBN: 0521833787.

[Bre11] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011, pp. xiv+599. ISBN: 978-0-387-70913-0.

[BAS19] L. Brivadis, V. Andrieu, and U. Serres. "Luenberger observers for discrete-time nonlinear systems". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. 2019, pp. 3435–3440. DOI: 10.1109/CDC40024.2019.9029220.

[Bri+20a] L. Brivadis, V. Andrieu, É. Chabanon, É. Gagnière, N. Lebaz, and U. Serres. "New dynamical observer for a batch crystallization process based on solute concentration". *Journal of Process Control* 87 (2020), pp. 17–26. ISSN: 0959-1524. DOI: 10.1016/j.jprocont.2019.12.012.

[Bri+21a] L. Brivadis, V. Andrieu, U. Serres, and J.-P. Gauthier. "Luenberger observers for infinite-dimensional systems, Back and Forth Nudging, and application to a crystallization process". *SIAM Journal on Control and Optimization* 59.2 (2021), pp. 857–886. DOI: 10.1137/20M1329020.

[Bri+20b] L. Brivadis, J.-P. Gauthier, L. Sacchelli, and U. Serres. "New perspectives on output feedback stabilization at an unobservable target". Submitted to *ESAIM. Control, Optimisation and Calculus of Variations.* Under review. Nov. 2020.

[Bri+21b] L. Brivadis, J.-P. Gauthier, L. Sacchelli, and U. Serres. "Avoiding observability singularities in output feedback bilinear systems". *SIAM Journal on Control and Optimization* 59.3 (2021), pp. 1759–1780. DOI: 10.1137/19M1272925.

[BS20] L. Brivadis and L. Sacchelli. "New inversion methods for the single/multi-shape CLD-to-PSD problem with spheroid particles". Submitted to *Journal of Process Control*. Under review. Dec. 2020.

[BS21a] L. Brivadis and L. Sacchelli. "A switching technique for output feedback stabilization at an unobservable target". Submitted to *2021 60th IEEE Conference on Decision and Control (CDC)*. Under review. Mar. 2021.

[BS21b] L. Brivadis and L. Sacchelli. “Approximate observability and back and forth observer of a PDE model of crystallisation process”. Submitted to *2021 60th IEEE Conference on Decision and Control (CDC)*. Under review. Mar. 2021.

[Bri+21c] L. Brivadis, L. Sacchelli, V. Andrieu, J.-P. Gauthier, and U. Serres. “From local to global asymptotic stabilizability for weakly contractive control systems”. *Automatica J. IFAC* 124 (2021). ISSN: 0005-1098. DOI: 10.1016/j.automatica.2020.109308.

[Cel+89] F. Celle, J.-P. Gauthier, D. Kazakos, and G. Sallet. “Synthesis of nonlinear observers: a harmonic-analysis approach”. *Mathematical Systems Theory*. 22.4 (1989), pp. 291–322. ISSN: 0025-5661. DOI: 10.1007/BF02088304.

[CV00] O. Chabour and J. Vivalda. “Remark on local and global stabilization of homogeneous bilinear systems”. *Systems & Control Letters* 41.2 (2000), pp. 141–143. ISSN: 0167-6911. DOI: 10.1016/S0167-6911(00)00045-1.

[Com+16] P. Combes, A. K. Jebai, F. Malrait, P. Martin, and P. Rouchon. “Adding virtual measurements by signal injection”. In: *2016 American Control Conference (ACC)*. 2016, pp. 999–1005. DOI: 10.1109/ACC.2016.7525045.

[Cor07] J. Coron. *Control and Nonlinearity*. Mathematical surveys and monographs. American Mathematical Society, 2007. ISBN: 9780821836682.

[Cor94a] J.-M. Coron. “On the stabilization of controllable and observable systems by an output feedback law”. *Mathematics of Control, Signals, and Systems* 7.3 (1994), pp. 187–216. ISSN: 0932-4194. DOI: 10.1007/BF01212269.

[Cor94b] J.-M. Coron. “Relations entre commandabilité et stabilisations non linéaires”. In: *Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. XI (Paris, 1989–1991)*. Vol. 299. Pitman Research Notes In Mathematics Series. Harlow: Longman Scientific & Technical, 1994, pp. 68–86.

[DK74] J. L. Dalec’kii and M. G. Krein. *Stability of solutions of differential equations in Banach space*. Translated from the Russian by S. Smith, Translations of Mathematical Monographs, Vol. 43. American Mathematical Society, Providence, R.I., 1974, pp. vi+386.

[Dri98] L. P. D. v. d. Dries. *Tame topology and O-minimal structures*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1998. DOI: 10.1017/CBO9780511525919.

[EN00] K.-J. Engel and R. Nagel. *One-parameter semigroups for linear evolution equations*. Vol. 194. Graduate Texts in Mathematics. With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt. Springer-Verlag, New York, 2000, pp. xxii+586. ISBN: 0-387-98463-1.

[EK92] F. Esfandiari and H. K. Khalil. “Output feedback stabilization of fully linearizable systems”. *International Journal of Control* 56.5 (1992), pp. 1007–1037. ISSN: 0020-7179. DOI: 10.1080/00207179208934355.

[Fla19] E. Flayac. “Coupled methods of nonlinear estimation and control applicable to terrain-aided navigation”. Thèse de doctorat dirigée par Jean, Frédéric Mathématiques appliquées. PhD thesis. Université Paris-Saclay (ComUE), 2019.

[FK83] M. Fliess and I. Kupka. “A finiteness criterion for nonlinear input-output differential systems”. *SIAM Journal on Control and Optimization* 21.5 (1983), pp. 721–728. ISSN: 0363-0129. DOI: 10.1137/0321044.

[Gao+18] Z. Gao, Y. Wu, Y. Bao, J. Gong, J. Wang, and S. Rohani. “Image analysis for in-line measurement of multidimensional size, shape, and polymorphic transformation of l-glutamic acid using deep learning-based image segmentation and classification”. *Crystal Growth & Design* 18.8 (Aug. 2018), pp. 4275–4281. DOI: 10.1021/acs.cgd.8b00883.

[GB81] J.-P. Gauthier and G. Bornard. “Observability for any $u(t)$ of a class of nonlinear systems”. *IEEE Transactions on Automatic Control* 26.4 (1981), pp. 922–926. ISSN: 0018-9286. DOI: 10.1109/TAC.1981.1102743.

[GK92] J.-P. Gauthier and I. Kupka. “A separation principle for bilinear systems with dissipative drift”. *IEEE Transactions on Automatic Control* 37.12 (1992), pp. 1970–1974. ISSN: 0018-9286. DOI: 10.1109/9.182484.

[GK01] J.-P. Gauthier and I. Kupka. *Deterministic observation theory and applications*. Cambridge University Press, Cambridge, 2001, pp. x+226. ISBN: 0-521-80593-7. DOI: 10.1017/CBO9780511546648.

[GG74] M. Golubitsky and V. Guillemin. *Stable mappings and their singularities*. Graduate texts in mathematics. Springer, 1974.

[GM88] M. Goresky and R. MacPherson. *Stratified Morse theory*. Vol. 14. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Springer-Verlag, Berlin, 1988, pp. xiv+272. ISBN: 3-540-17300-5. DOI: 10.1007/978-3-642-71714-7.

[Gru17] F. Gruy. “Chord Length Distribution: relationship between distribution moments and Minkowski functionals”. Submitted to *Journal of Process Control*. Under review. 2017.

[Gut81] P.-O. Gutman. “Stabilizing controllers for bilinear systems”. *IEEE Transactions on Automatic Control* 26.4 (1981), pp. 917–922. DOI: 10.1109/TAC.1981.1102742.

[Hai14] G. Haine. “Recovering the observable part of the initial data of an infinite-dimensional linear system with skew-adjoint generator”. *Mathematics of Control, Signals, and Systems* 26.3 (2014), pp. 435–462. ISSN: 0932-4194. DOI: 10.1007/s00498-014-0124-z.

[HR11a] G. Haine and K. Ramdani. “Observateurs itératifs en horizon fini. Application à la reconstruction de données initiales pour des EDP d’évolution”. *Journal Européen des Systèmes Automatisés (JESA)* 45 (Dec. 2011), pp. 715–724. DOI: 10.3166/jesa.45.715-724.

[HR11b] G. Haine and K. Ramdani. "Reconstructing initial data using observers: Error analysis of the semi-discrete and fully discrete approximations". *Numerische Mathematik* 120 (July 2011), pp. 307–343. DOI: 10.1007/s00211-011-0408-x.

[Hd90] H. Hammouri and J. de Leon Morales. "Observer synthesis for state-affine systems". In: *29th IEEE Conference on Decision and Control*. 1990, 784–785 vol.2. DOI: 10.1109/CDC.1990.203695.

[Hir94] M. W. Hirsch. *Differential topology*. Vol. 33. Graduate Texts in Mathematics. Corrected reprint of the 1976 original. Springer-Verlag, New York, 1994, pp. x+222. ISBN: 0-387-90148-5.

[HPR14] T.-B. Hoang, W. Pasillas-Lépine, and W. Respondek. "A switching observer for systems with linearizable error dynamics via singular time-scaling". In: *MTNS 2014*. Groningen, Netherlands, July 2014.

[Hob+91] E. F. Hobbel, R. Davies, F. W. Rennie, T. Allen, L. E. Butler, E. R. Waters, J. T. Smith, and R. W. Sylvester. "Modern methods of on-line size analysis for particulate process streams". *Particle & Particle Systems Characterization* 8.1-4 (1991), pp. 29–34. DOI: 10.1002/ppsc.19910080107.

[HK64] H. Hulburt and S. Katz. "Some problems in particle technology: A statistical mechanical formulation". *Chemical Engineering Science* 19.8 (1964), pp. 555–574.

[IJ14] K. Ito and B. Jin. *Inverse Problems*. Vol. 22. Series on Applied Mathematics. World Scientific, Oct. 2014, p. 332. DOI: 10.1142/9120.

[IK02] K. Ito and F. Kappel. *Evolution equations and approximations*. Vol. 61. Series on Advances in Mathematics for Applied Sciences. World Scientific Publishing Co., Inc., River Edge, NJ, 2002, pp. xiv+498. ISBN: 981-238-026-4. DOI: 10.1142/9789812777294.

[IRT11] K. Ito, K. Ramdani, and M. Tucsnak. "A time reversal based algorithm for solving initial data inverse problems". *Discrete and Continuous Dynamical Systems* 4.3 (2011), pp. 641–652. ISSN: 1937-1632. DOI: 10.3934/dcdss.2011.4.641.

[JG95] P. Jouan and J.-P. Gauthier. "Finite singularities of nonlinear systems. Output stabilization, observability and observers". In: *Proceedings of 1995 34th IEEE Conference on Decision and Control*. Vol. 4. 1995, 3295–3299 vol.4. DOI: 10.1109/CDC.1995.478688.

[JQ78] V. Jurdjevic and J. Quinn. "Controllability and stability". *Journal of Differential Equations* 28.3 (1978), pp. 381–389. ISSN: 0022-0396. DOI: 10.1016/0022-0396(78)90135-3.

[Jur96] V. Jurdjevic. *Geometric Control Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1996. DOI: 10.1017/CBO9780511530036.

[KK98] N. Kazantzis and C. Kravaris. "Nonlinear observer design using Lyapunov's auxiliary theorem". *Systems & Control Letters* 34.5 (1998), pp. 241–247. ISSN: 0167-6911. DOI: 10.1016/S0167-6911(98)00017-6.

[Kel84] A. M. Kellerer. “Chord-length distributions and related quantities for spheroids”. *Radiation Research* 98.3 (1984), pp. 425–437. ISSN: 00337587, 19385404.

[Ker16] M. Kern. *Méthodes numériques pour les problèmes inverses*. Collection Mathématiques et statisitques. ISTE Éditions, Mar. 2016, p. 222.

[KE93] H. K. Khalil and F. Esfandiari. “Semiglobal stabilization of a class of nonlinear systems using output feedback”. *IEEE Transactions on Automatic Control* 38.9 (Sept. 1993), pp. 1412–1415. ISSN: 0018-9286. DOI: 10.1109/9.237658.

[KZ84] M. A. Krasnosel’skii and P. P. Zabreiko. *Geometrical methods of nonlinear analysis*. Vol. 263. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Translated from the Russian by Christian C. Fenske. Springer-Verlag, Berlin, 1984, pp. xix+409. ISBN: 3-540-12945-6. DOI: 10.1007/978-3-642-69409-7.

[KSW01] M. Krichman, E. D. Sontag, and Y. Wang. “Input-Output-to-State Stability”. *SIAM Journal on Control and Optimization* 39.6 (2001), pp. 1874–1928. DOI: 10.1137/S0363012999365352.

[Laf86] A. Laforgia. “Inequalities for Bessel functions”. *Journal of Computational and Applied Mathematics* 15.1 (1986), pp. 75–81. ISSN: 0377-0427. DOI: 10.1016/0377-0427(86)90239-6.

[LSG17] M. Lagache, U. Serres, and J. Gauthier. “Exact output stabilization at unobservable points: Analysis via an example”. In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. Dec. 2017, pp. 6744–6749. DOI: 10.1109/CDC.2017.8264676.

[Lan+01] P. Langston, A. Burbidge, T. Jones, and M. Simmons. “Particle and droplet size analysis from chord measurements using Bayes’ theorem”. *Powder Technology* 116.1 (2001), pp. 33–42. ISSN: 0032-5910. DOI: 10.1016/S0032-5910(00)00359-4.

[Leb+15] N. Lebaz, A. Cockx, M. Spérandio, and J. Morchain. “Reconstruction of a distribution from a finite number of its moments: A comparative study in the case of depolymerization process”. *Computers & Chemical Engineering* 84 (Sept. 2015). DOI: 10.1016/j.compchemeng.2015.09.008.

[LW05] M. Li and D. Wilkinson. “Determination of non-spherical particle size distribution from chord length measurements. Part 1: Theoretical analysis”. *Chemical Engineering Science* 60.12 (2005), pp. 3251–3265. ISSN: 0009-2509. DOI: 10.1016/j.ces.2005.01.008.

[Liu97] K. Liu. “Locally distributed control and damping for the conservative systems”. *SIAM Journal on Control and Optimization* 35.5 (1997), pp. 1574–1590. ISSN: 0363-0129. DOI: 10.1137/S0363012995284928.

[LCK98] W. Liu, N. N. Clark, and A. I. Karamavruç. "Relationship between bubble size distributions and chord-length distribution in heterogeneously bubbling systems". *Chemical Engineering Science* 53.6 (1998), pp. 1267–1276. ISSN: 0009-2509. DOI: https://doi.org/10.1016/S0009-2509(97)00426-0.

[Lue71] D. Luenberger. "An introduction to observers". *IEEE Transactions on Automatic Control* 16.6 (1971), pp. 596–602. DOI: 10.1109/TAC.1971.1099826.

[Lue64] D. G. Luenberger. "Observing the state of a linear system". *IEEE Transactions on Military Electronics* 8.2 (Apr. 1964), pp. 74–80. ISSN: 0536-1559. DOI: 10.1109/TME.1964.4323124.

[MPI07] L. Marconi, L. Praly, and A. Isidori. "Output stabilization via nonlinear Luenberger observers". *SIAM Journal on Control and Optimization* 45.6 (2007), pp. 2277–2298. ISSN: 0363-0129. DOI: 10.1137/050642344.

[MBA20a] S. Marx, L. Brivadis, and D. Astolfi. "Forwarding design for stabilization of a coupled transport equation-ODE with a cone-bounded input nonlinearity". In: *2020 59th IEEE Conference on Decision and Control (CDC)*. 2020, pp. 640–645. DOI: 10.1109/CDC42340.2020.9304178.

[MBA20b] S. Marx, L. Brivadis, and D. Astolfi. "Forwarding techniques for the global stabilization of dissipative infinite-dimensional systems coupled with an ODE". Submitted to *Mathematics of Control, Signals, and Systems*. Under review. Sept. 2020.

[MP93] F. Mazenc and L. Praly. "Global stabilization for nonlinear systems". Preprint. CAS, ENSMP. Jan. 1993.

[MEH01] A. Mersmann, A. Eble, and C. Heyer. "Crystal growth". In: *Crystallization Technology Handbook*. Ed. by A. Mersmann. Marcel Dekker Inc., 2001, pp. 48–111. ISBN: 0824705289.

[Mes+11] A. Mesbah, A. E. Huesman, H. J. Kramer, and P. M. Van den Hof. "A comparison of nonlinear observers for output feedback model-based control of seeded batch crystallization processes". *Journal of Process Control* 21.4 (2011), pp. 652–666. ISSN: 0959-1524. DOI: 10.1016/j.jprocont.2010.11.013.

[MK80] R. R. Mohler and W. J. Kolodziej. "An overview of bilinear system theory and applications". *IEEE Transactions on Systems, Man, and Cybernetics* 10.10 (1980), pp. 683–688. DOI: 10.1109/TSMC.1980.4308378.

[Mul01] J. Mullin. *Crystallization*. 4th ed. Elsevier, 2001. ISBN: 9780080530116.

[NN02] R. Nagel and G. Nickel. "Well-posedness for nonautonomous abstract Cauchy problems". In: *Evolution Equations, Semigroups and Functional Analysis: In Memory of Brunello Terreni*. Basel: Birkhäuser Basel, 2002, pp. 279–293. ISBN: 978-3-0348-8221-7. DOI: 10.1007/978-3-0348-8221-7_15.

[Nar66] R. Narasimhan. *Introduction to the theory of analytic spaces*. Lecture Notes in Mathematics, No. 25. Springer-Verlag, Berlin-New York, 1966, pp. iii+143.

[NS98] D. Nešić and E. Sontag. “Input-to-state stabilization of linear systems with positive outputs”. *Systems & Control Letters* 35.4 (1998), pp. 245–255. ISSN: 0167-6911. DOI: 10.1016/S0167-6911(98)00060-7.

[Neu04] E. Neuman. “Inequalities involving Bessel functions of the first kind”. *JIPAM. Journal of Inequalities in Pure and Applied Mathematics* 5.4 (2004), Article 94, 4. ISSN: 1443-5756.

[PR16] A. V. Pandit and V. V. Ranade. “Chord length distribution to particle size distribution”. *AIChE Journal* 62.12 (2016), pp. 4215–4228.

[Paz83] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*. Vol. 44. Applied Mathematical Sciences. Springer-Verlag, New York, 1983, pp. viii+279. ISBN: 0-387-90845-5. DOI: 10.1007/978-1-4612-5561-1.

[Phó91] V. Q. Phóng. “The operator equation $AX - XB = C$ with unbounded operators A and B and related abstract Cauchy problems”. *Mathematische Zeitschrift* 208.1 (1991), pp. 567–588. DOI: 10.1007/BF02571546.

[PQ05] J. Polendo and C. Qian. “A generalized framework for global output feedback stabilization of genuinely nonlinear systems”. In: *Proceedings of the 44th IEEE Conference on Decision and Control*. 2005, pp. 2646–2651.

[PÖ17] M. Porru and L. Özkan. “Monitoring of batch industrial crystallization with growth, nucleation, and agglomeration. Part 2: Structure design for state estimation with secondary measurements”. *Industrial & engineering chemistry research* 56.34 (2017), pp. 9578–9592. DOI: 10.1021/acs.iecr.7b00243.

[Pre+10] B. Presles, J. Debayle, G. Fevotte, and J.-C. Pinoli. “Novel image analysis method for in situ monitoring the particle size distribution of batch crystallization processes”. *Journal of Electronic Imaging* 19.3 (2010), pp. 1–7. DOI: 10.1117/1.3462800.

[PP04] C. Prieur and L. Praly. “A tentative direct Lyapunov design of output feedbacks”. *IFAC Proceedings Volumes* 37.13 (2004). 6th IFAC Symposium on Nonlinear Control Systems 2004 (NOLCOS 2004), Stuttgart, Germany, 1-3 September, 2004, pp. 847–852. ISSN: 1474-6670. DOI: 10.1016/S1474-6670(17)31331-9.

[Qui80] J. P. Quinn. “Stabilization of bilinear systems by quadratic feedback controls”. *Journal of Mathematical Analysis and Applications* 75.1 (1980), pp. 66–80. ISSN: 0022-247X. DOI: https://doi.org/10.1016/0022-247X(80)90306-6.

[RTW10] K. Ramdani, M. Tucsnak, and G. Weiss. “Recovering and initial state of an infinite-dimensional system using observers”. *Automatica J. IFAC* 46.10 (2010), pp. 1616–1625. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2010.06.032.

[RL88] A. D. Randolph and M. A. Larson. *Theory of particulate processes*. 2nd ed. Academic Press, 1988. ISBN: 9780125796521.

[RD20] A. Rapaport and D. Dochain. "A robust asymptotic observer for systems that converge to unobservable states—a batch reactor case study". *IEEE Transactions on Automatic Control* 65.6 (2020), pp. 2693–2699. DOI: 10.1109/TAC.2019.2940870.

[ROE01] H. Rodriguez, R. Ortega, and G. Escobar. "A new family of energy-based non-linear controllers for switched power converters". In: *ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 01TH8570)*. Vol. 2. 2001, pp. 723–727.

[Rus78] D. L. Russell. "Controllability and stabilizability theory for linear partial differential equations: recent progress and open questions". *SIAM Review* 20.4 (1978), pp. 639–739. DOI: 10.1137/1020095.

[Sac+20] L. Sacchelli, L. Brivadis, V. Andrieu, U. Serres, and J.-P. Gauthier. "Dynamic output feedback stabilization of non-uniformly observable dissipative systems". *IFAC-PapersOnLine* 53.2 (2020). 21th IFAC World Congress, pp. 4923–4928. ISSN: 2405-8963. DOI: https://doi.org/10.1016/j.ifacol.2020.12.1071.

[SJ+14] A. van der Schaft, D. Jeltsema, et al. "Port-Hamiltonian systems theory: An introductory overview". *Foundations and Trends® in Systems and Control* 1.2-3 (2014), pp. 173–378.

[Seg63] I. Segal. "Non-linear semi-groups". *Annals of Mathematics. Second Series* 78 (1963), pp. 339–364. ISSN: 0003-486X. DOI: 10.2307/1970347.

[ST03] H. Shim and A. Teel. "Asymptotic controllability and observability imply semiglobal practical asymptotic stabilizability by sampled-data output feedback". *Automatica J. IFAC* 39.3 (2003), pp. 441–454. ISSN: 0005-1098. DOI: 10.1016/S0005-1098(02)00278-9.

[Sie14] C. L. Siegel. "Über einige Anwendungen diophantischer Approximationen [reprint of Abhandlungen der Preußischen Akademie der Wissenschaften. Physikalisch-mathematische Klasse 1929, Nr. 1]". In: *On some applications of Diophantine approximations*. Vol. 2. Quad./Monogr. Ed. Norm., Pisa, 2014, pp. 81–138.

[SLB99] M. Simmons, P. Langston, and A. Burbidge. "Particle and droplet size analysis from chord distributions". *Powder Technology* 102.1 (1999), pp. 75–83. ISSN: 0032-5910. DOI: 10.1016/S0032-5910(98)00197-1.

[Sle72] M. Slemrod. "The linear stabilization problem in Hilbert space". *Journal of Functional Analysis* 11.3 (1972), pp. 334–345. ISSN: 0022-1236. DOI: https://doi.org/10.1016/0022-1236(72)90073-0.

[Sle74] M. Slemrod. "A note on complete controllability and stabilizability for linear control systems in Hilbert space". *SIAM Journal on Control* 12.3 (1974), pp. 500–508. DOI: 10.1137/0312038.

[Son81] E. D. Sontag. "Conditions for abstract nonlinear regulation". *Information and Control* 51.2 (1981), pp. 105–127. ISSN: 0019-9958. DOI: 10.1016/S0019-9958(81)90217-5.

[Sur+19] D. Surroop, P. Combes, P. Martin, and P. Rouchon. "Third-order virtual measurements with signal injection". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. 2019, pp. 642–647. DOI: 10.1109/CDC40024.2019.9029751.

[Sur+20] D. Surroop, P. Combes, P. Martin, and P. Rouchon. "Adding virtual measurements by PWM-induced signal injection". In: *2020 American Control Conference (ACC)*. 2020, pp. 2692–2698. DOI: 10.23919/ACC45564.2020.9147700.

[TP94] A. Teel and L. Praly. "Global stabilizability and observability imply semi-global stabilizability by output feedback". *Systems & Control Letters* 22.5 (1994), pp. 313–325. ISSN: 0167-6911. DOI: 10.1016/0167-6911(94)90029-9.

[TP95] A. Teel and L. Praly. "Tools for semiglobal stabilization by partial state and output feedback". *SIAM Journal on Control and Optimization* 33.5 (1995), pp. 1443–1488. ISSN: 0363-0129. DOI: 10.1137/S0363012992241430.

[TP00] A. R. Teel and L. Praly. "A smooth Lyapunov function from a class-$\mathcal{KL}$ estimate involving two positive semidefinite functions". *ESAIM. Control, Optimisation and Calculus of Variations* 5 (2000), pp. 313–367. ISSN: 1292-8119. DOI: 10.1051/cocv:2000113.

[TW09] M. Tucsnak and G. Weiss. *Observation and control for operator semigroups*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Verlag, Basel, 2009, pp. xii+483. ISBN: 978-3-7643-8993-2. DOI: 10.1007/978-3-7643-8994-9.

[Ucc11] B. Uccheddu. "Observer for a batch crystallization process". Theses. Université Claude Bernard - Lyon I, July 2011.

[Urq05] J. M. Urquiza. "Rapid exponential feedback stabilization with unbounded control operators". *SIAM Journal on Control and Optimization* 43.6 (2005), pp. 2233–2244. ISSN: 0363-0129. DOI: 10.1137/S0363012901388452.

[Vil68] N. J. Vilenkin. *Special functions and the theory of group representations*. Translated from the Russian by V. N. Singh. Translations of Mathematical Monographs, Vol. 22. American Mathematical Society, Providence, R. I., 1968, pp. x+613.

[Vis12] J. A. W. Vissers. "Model-based estimation and control methods for batch cooling crystallizers". PhD thesis. Technische Universiteit Eindhoven, 2012.

[WHM05] J. Worlitschek, T. Hocker, and M. Mazzotti. "Restoration of PSD from Chord Length Distribution Data using the Method of Projections onto Convex Sets". *Particle & Particle Systems Characterization* 22.2 (2005), pp. 81–98. DOI: https://doi.org/10.1002/ppsc.200400872.

[XLG95] C.-Z. Xu, P. Ligarius, and J.-P. Gauthier. "An observer for infinite-dimensional dissipative bilinear systems". *Computers & Mathematics with Applications*. 29.7 (1995), pp. 13–21. ISSN: 0898-1221. DOI: 10.1016/0898-1221(95)00014-P.

[Zit+20] B. Zitte, B. Hamroun, D. Astolfi, and F. Couenne. "Robust control of a class of bilinear systems by forwarding: application to counter current heat exchanger". *IFAC-PapersOnLine* 53.2 (2020). 21th IFAC World Congress, pp. 11515–11520. ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2020.12.603.