LNCIS
LECTURE NOTES IN CONTROL AND INFORMATION SCIENCES
328

Antonio Loría
Françoise Lamnabhi-Lagarrigue
Elena Panteley (Eds.)

Advanced Topics in Control Systems Theory

Lecture Notes from FAP 2005

Springer

Lecture Notes in Control and Information Sciences 328

Editors: M. Thoma · M. Morari

Antonio Loría · Françoise Lamnabhi-Lagarrigue
Elena Panteley (Eds.)

Advanced Topics in Control Systems Theory

Lecture Notes from FAP 2005

With 33 Figures

Springer

Editors

Dr Antonio Loría, PhD
Dr Françoise Lamnabhi-Lagarrigue, PhD
Dr Elena Panteley, PhD

Laboratoire des Signaux et Systèmes
Centre National de la Recherche Scientifique (CNRS)
SUPELEC
3 rue Joliot Curie
91192 Gif-sur-Yvette
France

British Library Cataloguing in Publication Data
Advanced topics in control systems theory : lecture notes from FAP 2005. -
(Lecture notes in control and information sciences ; 328)
1. Automatic control - Congresses 2. Automatic control - Mathematical models - Congresses
3. Control theory - Congresses 4. Systems engineering - Congresses
I.Loria,A. (Antonio) II.Lamnabhi-Lagarrigue, F. (Francoise), 1953-
III.Panteley, E. (Elena) IV.FAP 2005
(Graduate school . 2005 : Paris)
629.8'312

ISBN-13 9781846283130
ISBN-10 1846283132

Library of Congress Control Number: 2005938387

Lecture Notes in Control and Information Sciences ISSN 0170-8643
ISBN-10: 1-84628-313-2 Printed on acid-free paper
ISBN-13: 978-1-84628-313-0

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Typesetting: Data conversion by authors.
Final processing by PTP-Berlin Protago-TEX-Production GmbH, Germany
Cover-Design: design & production GmbH, Heidelberg

Printed in Germany

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

To our lovely daughters,
AL & EP.

Preface

Advanced Topics in Control Systems Theory is a byproduct of the European school "Formation en Automatique de Paris 2005" (Paris Graduate School on Automatic Control). The school has taken place in spring every year since 2003 and is open to PhD students in control theory throughout Europe. In 2005, the school benefited of the valuable participation of 23 European renowned control researchers and more than 80 European PhD students. While the program consisted of the modules listed below, the contents of the present monograph collects selected notes provided by the lecturers and is by no means exhaustive.

Program of FAP 2005:

P1 Control theory of linear and nonlinear distributed systems
Y. Chitour, E. Trélat

P2 Nonsmooth Analysis and Control Theory
F. Clarke

P3 Efficient methods for linear control and estimation: an algebraic approach
H. Bourles, M. Fliess

P4 Nonlinear optimal control
B. Bonnard

P5 Sampled-data control systems
A. Astolfi, D. Shona-Laila

P6 Nonlinear adaptive control with applications
A. Astolfi, D. Karagianis, R. Ortega

P8 Tools for analysis and control of time-varying systems
A. Loria, E. Panteley

P9 Control of oscillating mechanical systems, synchronization and chaos
J. Levine, H. Nijmeijer

P10 Stability and control of time-delay systems
S. Niculescu, Y. Chitour

P11 On observer design for nonlinear systems
G. Besancon, E. Busvelle

P12 Hybrid systems modeling and control in automotive applications
K.H. Johansson, A. Balluchi, W. Pasillas

P13 Algebraic analysis of multidimensional control systems
J.-F. Pommaret

P14 Geometry of static and dynamic feedback
W. Respondek

As for previous FAP schools each module was taught over 21hrs within one week. Therefore, the contents of the present monograph may be used in support to either a one-term general advanced course on nonlinear control theory, thereby devoting a few lectures to each topic, or it may be used in support to more focused intensive courses at graduate level. The academic requirement for the class student or the reader in general is a basic knowledge on control theory (linear and non linear).

Advanced Topics in Control Systems Theory also constitutes an ideal start for researchers in control theory who wish to broaden their general culture or to get involved in fields different to their expertise, while avoiding a thorough book-keeping. Indeed, the monograph presents in a concise but pedagogical manner diverse aspects of modern control and dynamic systems theory: optimal control, output feedback control, infinite-dimensional systems, systems with delays, sampled-data systems and stability theory. In particular, these lecture notes are based on the material taught in modules P1, P3, P4, P5, P8, P10, and P11.

This is the second of a series of yearly volumes, which shall prevail beyond the lectures taught in class during each FAP season (spring). Further information on FAP, in particular, on the scientific program for the subsequent years is updated on `www.lss.supelec.fr/~loria/FAP2005/Program/` approximately during fall preceeding a FAP season..

FAP is organized within the context of the European teaching network "Control Training Site" sponsored by the European Community through the Marie Curie program. The editors of the present text greatefully acknowledge such sponsorship. We also take this oportunity to acknowledge the French national center for scientific research (C.N.R.S.) which provides us with a working environment and ressources probably unparalleled in the world.

Gif sur Yvette, France.
October 2005

Antonio Loría,
Françoise Lamnabhi-Lagarrigue,
Elena Panteley.

Contents

5 Controllability of Partial Differential Equations

6 Stability, Told by Its Developers

List of Contributors

A. Astolfi
Department of Electrical and
Electronic Engineering, Imperial
College,
Exhibition Road, London SW7
2BT, UK.
a.astolfi@imperial.ac.uk

G. Besançon
Laboratoire d'Automatique de
Grenoble,
Ecole Nationale Supérieure des
Ingénieurs Electriciens de Grenoble
B.P. 46, 38402 St. martin d'Hères,
Cedex, France.
Gildas.Besancon@inpg.fr

B. Bonnard
Institut de Mathématiques de
Bourgogne, UMR CNRS 5584,
B.P 47870, 21078 Dijon Cedex,
France.
bbonnard@u-bourgogne.fr

H. Bourlès
SATIE, ENS de Cachan et CNAM,
61 Ave du Président Wilson,
94230 Cachan, France.
henri.bourles@satie.ens_cachan.fr

J.-B. Caillau
ENSEEIHT-IRIT, UMR CNRS
5505, 2 rue Camichel,
31071 Toulouse, France.
caillau@n7.fr

Y. Chitour
Laboratoire des Signaux et
Systèmes,
Université Paris-Sud, C.N.R.S.,
Supélec,
3, Rue Joliot Curie, 91192
Gif-sur-Yvette, France.
chitour@lss.supelec.fr

K. Gu
Department of Mechanical and
Industrial Engineering,
Southern Illinois University
Edwardsville, Edwardsville,
Illinois 62026-1805, USA
kgu@siue.edu

A. Loría
C.N.R.S., Laboratoire des Signaux
et Systèmes, Supélec,
3, Rue Joliot Curie, 91192
Gif-sur-Yvette, France.
loria@lss.supelec.fr

E. Panteley
C.N.R.S., Laboratoire des Signaux
et Systèmes, Supélec,
3, Rue Joliot Curie, 91192
Gif-sur-Yvette, France.
panteley@lss.supelec.fr

D. Nešić
Department of Electrical and
Electronic Engineering,
The University of Melbourne,
Parkville, VIC 3001, Australia
d.nesic@ee.unimelb.edu.au

S. Niculescu
C.N.R.S., HEUDIASYC,
Centre de Recherches de Royallieu
BP 20529 - 60205 Compiègne
CEDEX, France
niculescu@hds.utc.fr

D. Shona-Laila
Department of Electrical and
Electronic Engineering, Imperial
College,
Exhibition Road, London SW7
2BT, UK.
d.laila@imperial.ac.uk

A. Teel
Department of Electrical and
Computer Engineering, University
of California, Santa Barbara, CA
93106, USA. teel@ece.ucsb.edu

E. Trélat
Laboratory AN-EDP, CNRS UMR
8628,
Université Paris-Sud
Orsay, France.
emmanuel.trelat@math.u-psud.fr

L. Zaccarian
Dipartimento di Informatica,
Sistemi e Produzione,
University of Rome, Tor Vergata,
00133 Rome, Italy
zack@disp.uniroma2.it

1

Introduction to Nonlinear Optimal Control

Bernard Bonnard[1] and Jean-Baptiste Caillau[2]

[1] Université de Bourgogne, Bâtiment Sciences Mirande, BP 47870, F-21078 Dijon.
E-mail: Bernard.Bonnard@u-bourgogne.fr

[2] ENSEEIHT-IRIT (UMR CNRS 5505), 2 Rue Camichel, F-31071 Toulouse.
E-mail: caillau@n7.fr.

The maximum principle is presented in the weak and general forms. The standard proofs are detailed, and the connection with the shooting method for numerical resolution is made. A brief introduction to the micro-local analysis of extremals is also provided. Regarding second-order conditions, small time-optimality is addressed by means of high order generalized variations. As for local optimality of extremals, the conjugate point theory is introduced both for regular problems and for minimum time singular single input affine control systems. The analysis is applied to the minimum time control of the Kepler equation, and the numerical simulations for the corresponding orbit transfer problems are given. In the case of state constrained optimal control problems, necessary conditions are stated for boundary arcs. The junction and reflection conditions are derived in the Riemannian case.

1.1 Introduction

The objective of this article is to present available techniques to analyze optimal control problems of systems governed by ordinary differential equations. Coupled with numerical methods, they provide tools to solve practical problems. This will be illustrated by the minimum time transfer between Keplerian orbits.

The material is organized as follows. Section 1.2 is devoted to the standard maximum principle who was formulated and proved by Pontryagin and his collaborators in 1956. We follow in the presentation the line of the discovery, see [9]. First of all, we give the weak version, assuming the control domain open. Then we formulate and prove the general theorem along the lines of [15].

Work supported in part by the French Space Agency (contract 02/CNES/0257/00) and done in the framework of the HYCON Network of Excellence, contract number FP6-IST-511368.

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 1–60, 2006.

The maximum principle is a necessary optimality result and further conditions are usually required to select minimizers. The aim of Section 1.3 is to present the recent techniques developed to achieve this task. They use the second-order variation along a reference extremal solution of the maximum principle and are directly applicable when the control domain is open. The problem is to test the sign of this second variation. This is done in two steps. First, we must check optimality for small time. To this end, we use special variations and make direct evaluations of the accessibility set, especially using the Baker-Campbell-Hausdorff formula. This approach has provided a generalization of the maximum principle called the high order maximum principle, first obtained by Krener [13]. This result can be applied in the so-called singular case where the standard maximum principle is not able to distinguish minima from maxima. A consequence of this generalization is to get second-order computable conditions in the singular case: generalized Legendre and Goh conditions. The second step, which does not concern small time, is the concept of conjugate point: the problem is to compute in the $\mathcal{C}^1$ topology the first time when a reference trajectory loses local optimality. We present an algorithm to compute this time in the smooth case. This computation is based on the concept of Lagrangian singularity related to the second-order derivative. We give the elements of symplectic geometry necessary to the understanding. One practical motivation for the discovery of the maximum principle was coming from the space engineering. In Section 1.4 we present applications of the afore-mentioned techniques to investigate the minimum time transfer of a spaceship between Keplerian orbits. They are combined with geometrical analysis and numerical simulations so as to compute the optimal solution. The final section deals with the necessary conditions for state constrained problems. The presentation is geometric, is the spirit of Gamkrelidze approach [18]. The conditions, due to Weierstraß, are proved in the planar case.

1.2 Optimal Control and Maximum Principle

1.2.1 Preliminaries

In this section, we consider a system written in local coordinates as

$$\dot{x} = f(x, u)$$

where, for each time t, $x(t)$ is in $\mathbf{R}^n$, $u(t)$ in $U \subset \mathbf{R}^m$, and where (x, u) represents a trajectory-control pair defined on an interval $[0, T]$. We denote by $\mathcal{U}$ the class of admissible controls. To each trajectory we assign a cost of the form

$$c(x, u) = \int_0^T f^0(x, u)dt$$

where T can be fixed or not. The optimal control problem is to minimize this cost functional among all trajectories of the system satisfying prescribed boundary conditions of the form

$$x(0) \in M_0, \; x(T) \in M_1.$$

Our system can be extended to a state-cost system according to

$$\dot{x}^0 = f^0(x, u) \tag{1.1}$$

$$\dot{x} = f(x, u) \tag{1.2}$$

which we will also write, with $\widetilde{x} = (x^0, x)$ and $x^0(0) = 0$,

$$\dot{\widetilde{x}} = \widetilde{f}(\widetilde{x}, u).$$

In order to define the necessary optimality conditions, our problem has to be tamed in the following way. For each admissible control u, the corresponding solution $\widetilde{x}(t, \widetilde{x}_0, u)$ starting at time $t = 0$ from $\widetilde{x}_0 = (0, x_0)$ has to be uniquely defined on a maximal interval and has to be an absolutely continuous solution of the system (1.1)-(1.2) almost everywhere. Moreover, the differential of this solution with respect to the initial condition has to be defined, absolutely continuous and solution of the linear differential system

$$\frac{d}{dt}\frac{\partial \widetilde{x}}{\partial \widetilde{x}_0} = \frac{\partial \widetilde{f}}{\partial \widetilde{x}}(\widetilde{x}(t, \widetilde{x}_0, u))\frac{\partial \widetilde{x}}{\partial \widetilde{x}_0}$$

called the *variational system*. Those basic existence, uniqueness and regularity results are standard under the following assumptions.

(i) The set of admissible controls if the set of locally bounded mappings defined on the real line.

(ii) The function $\widetilde{f}$ and its partial derivative $\partial \widetilde{f}/\partial \widetilde{x}$ are continuous.

(iii)The prescribed boundary manifolds are regular submanifolds of $\mathbf{R}^n$.

The approach of our work is geometric and the important concept is the *accessibility set* attached to the system $\dot{x} = f(x, u)$ defined by

$$\mathcal{A}_{x_0,T} = \{x(T, x_0, u), \; u \in \mathcal{U}\}$$

when the initial condition is x_0 and the final time T. Observe that if (x, u) is optimal, the extremity $\widetilde{x}(T, \widetilde{x}_0, u)$ of the extended trajectory must clearly belong to the boundary of the accessibility set of the extended system. The maximum principle is a necessary condition for $\widetilde{x}(T, \widetilde{x}_0, u)$ to belong to $\partial \mathcal{A}_{\widetilde{x}_0,T}$.

1.2.2 The Weak Maximum Principle

We assume that f is smooth and that the set of admissible controls is the set of locally bounded mappings taking values in U, an *open* subset of $\mathbf{R}^m$. If we introduce the *endpoint mapping*, x_0 and T being fixed,

$$E_{x_0,T} \ : \ u \in \mathcal{U} \mapsto x(T, x_0, u)$$

then the accessibility set is the image of the mapping. Since the final time is fixed, the set $\mathcal{U}$ is endowed with the $\mathrm{L}^\infty([0,T])$-norm topology:

$$\|u\| = \text{Ess Sup}_{t\in[0,T]}|u(t)|$$

where $|.|$ is any equivalent norm on $\mathbf{R}^n$.

First and Second Variation

It can be easily proved that the endpoint mapping is $\mathcal{C}^\infty$ for the L^∞ topology and that the first and second variations are computed in the following way. Fix $x(0) = x_0$ and denote by (x, u) the reference solution defined on $[0, T]$. Let $x + \delta x$ be the solution starting from x_0 and generated by $u + \delta u$ where δu is an L^∞ variation. Since f is smooth we can write:

$$\begin{aligned} f(x+\delta x, u+\delta u) &= f(x,u) + \frac{\partial f}{\partial x}(x,u)\delta x + \frac{\partial f}{\partial u}(x,u)\delta u \\ &+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x,u)(\delta x,\delta x) + \frac{\partial^2 f}{\partial x\partial u}(x,u)(\delta x,\delta u) + \frac{1}{2}\frac{\partial^2 f}{\partial u^2}(x,u)(\delta u,\delta u) + \cdots \end{aligned}$$

Writing that $x + \delta x$ is solution, we have

$$\dot{x} + \delta\dot{x} = f(x+\delta x, u+\delta u)$$

and we can decompose δx as $\delta_1 x + \delta_2 x + \cdots$ where $\delta_1 x$ is linear in u and $\delta_2 x$ quadratic. By identification,

$$\delta_1\dot{x} = \frac{\partial f}{\partial x}(x(t),u(t))\delta_1 x + \frac{\partial f}{\partial u}(x(t),u(t))\delta u(t)$$

that is $\delta_1 x$ is solution of the system linearized along the reference trajectory and

$$\begin{aligned} \delta_2\dot{x} = &\frac{\partial f}{\partial x}(x(t),u(t))\delta_2 x + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x(t),u(t))(\delta_1 x(t),\delta_1 x(t)) \\ &+\frac{\partial^2 f}{\partial x\partial u}(x(t),u(t))(\delta_1 x(t),\delta u(t)) + \frac{1}{2}\frac{\partial^2 f}{\partial u^2}(x(t),u(t))(\delta u(t),\delta u(t)) \end{aligned} \tag{1.3}$$

with $\delta_1 x(0) = \delta_2 x(0) = 0$ since $\delta x(0) = 0$. Let $A(t)$ be the matrix $\partial f/\partial x$ along $(x(t), u(t))$ and let Φ be the matrix valued fundamental solution of

$$\dot{\Phi} = A(t)\Phi$$

with $\Phi(0) = I$. We observe that the first and second variations can be computed using the standard formula to integrate linear differential equations. In particular, by setting $B(t) = \partial f/\partial u$ along $(x(t), u(t))$, the Fréchet derivative is:

$$\delta_1 x(T) = \Phi(T)\int_0^T \Phi^{-1}(s)B(s)\delta u(s)ds. \tag{1.4}$$

Statement and Proof of the Weak Maximum Principle

Let (x, u) be the reference trajectory defined on $[0, T]$, T fixed. Assume that $x(T)$ belongs to the boundary of the accessibility set. Then, from the open mapping theorem, the control has to be a singularity of the endpoint mapping and we must have

$$\text{rank } E'_{x_0,T}(u) < n$$

where $E'_{x_0,T}(u)$ is the Fréchet derivative at u computed according to (1.4),

$$E'_{x_0,T}(u) = \delta_1 x(T).$$

To get the weak maximum principle, we take a nonzero covector $\bar{p}$ orthogonal to the image of $E'_{x_0,T}(u)$ and we set:

$$p(t) = \bar{p}\Phi(T)\Phi^{-1}(t).$$

Hence, p is solution of the *adjoint equation*

$$\dot{p} = -p\frac{\partial f}{\partial x}(x(t), u(t))$$

and, by construction,

$$\int_0^T p(t)B(t)\delta u(t)dt = 0$$

for all variations in $\mathrm{L}^\infty([0, T])$. As a result, $p(t)B(t)$ is zero almost everywhere and we have proved the following proposition.

Proposition 1.1. *Let (x, u) be a trajectory defined on $[0, T]$ such that $x(T)$ belongs the boundary of $\mathcal{A}_{x_0,T}$, the control set being open in $\mathbf{R}^m$. There exists an absolutely continuous nonvanishing covector function p defined on $[0, T]$ such that the triple (x, p, u) is almost everywhere solution of*

$$\dot{x} = \frac{\partial H}{\partial p}(x, p, u), \ \dot{p} = -\frac{\partial H}{\partial x}(x, p, u)$$
$$\frac{\partial H}{\partial u}(x, p, u) = 0$$

where $H(x, p, u) = \langle p, f(x, u)\rangle$ is the Hamiltonian *of the system.*

The covector function p is called the *adjoint state*. In particular, if (x, u) is time minimizing, $x(T)$ belongs to the boundary of the accessibility set and satisfies the previous necessary conditions.

1.2.3 The Maximization Condition

Actually, the second-order variation can be used so as to derive more conditions for time-optimality as explained in [9]. Let us denote by Π the image of the Fréchet derivative of the endpoint mapping at u. As previously noticed, if the reference trajectory is optimal, the hyperplane Π is at least of codimension one. Consider now the generic case where Π is of codimension exactly one, and where the reference trajectory is differentiable at T and intersects Π transversely. The adjoint vector at T is orthogonal to Π and thus uniquely defined up to a scalar. Morevor, since the trajectory is transverse to Π at T, we can use the normalization

$$p(T)f(x(T), u(T)) > 0.$$

We introduce the *intrinsic second-order derivative* which is defined as the restriction of the second variation to the kernel K of $E'_{x_0,T}(u)$ projected to $\Pi^\perp$. It is given by

$$\delta u \in K \mapsto p(T)\delta_2 x(T)$$

with $\delta_2 x$ computed by means of (1.3). If u is time-optimal, we must have (see Fig. 1.1):

$$p(T)\delta_2 x(T) \leq 0, \;\; \delta u \in K.$$

Expliciting $\delta_2 x(T)$, one gets the additional standard *Legendre-Clebsch* condition,

$$\frac{\partial^2 H}{\partial u^2} \leq 0$$

and finally obtains the (local) maximization condition : almost everywhere,

$$H(x(t), p(t), u(t)) = \max_{v \in V_t} H(x(t), p(t), v)$$

with, for each t, V_t a neighbourhood of $u(t)$.

1.2.4 Maximum Principle, Fixed Time

Statement

Consider a system $\dot{x} = f(x, u)$ with, as before, f and $\partial f/\partial x$ continuous functions on an open subset of $\mathbf{R}^{n+m}$. The set of admissible controls $\mathcal{U}$ is again the set of locally bounded functions taking values in a fixed subset U of $\mathbf{R}^m$, and such that the responses starting at $t = 0$ from x_0 are defined on the whole interval $[0, T]$, T fixed. Let (x, u) be a reference trajectory such that the endpoint $x(T)$ belongs to the boundary of the accessibility set. Then, there exists a non-trivial covector absolutely continuous function p such that the triple (x, p, u) is almost everywhere solution of the equations

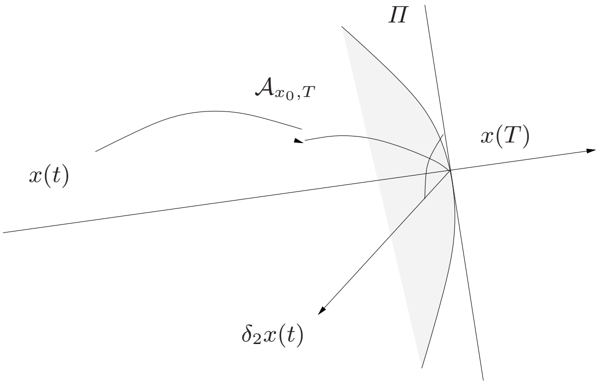

Fig. 1.1. Legendre-Clebsch condition: non-positivity of the intrinsic second-order derivative

$$\dot{x} = \frac{\partial H}{\partial p}(x,p,u),\ \dot{p} = -\frac{\partial H}{\partial x}(x,p,u) \tag{1.5}$$

where $H(x,p,u) = \langle p, f(x,u)\rangle$ is the Hamiltonian. Moreover, the maximization condition holds almost everywhere along the extremal triple,

$$H(x(t),p(t),u(t)) = M(x(t),p(t))$$

where $M(x,p) = \max_{v\in U} H(x,p,v)$, and $t \mapsto M(x(t),u(t))$ is constant on $[0,T]$.

The Proof of the Maximum Principle

Needle variations. The basic concept needed to prove the maximum principle is the concept of *needle variation*. Indeed, because the control domain is arbitrary, standard L^∞ variations of the reference control used when U is open have to be replaced by L^1 elementary ones of the form:

$$u_{\pi_1}(t,\varepsilon) = \begin{cases} u_1 \text{ on } [t_1 - \varepsilon l_1, t_1] \\ u(t) \text{ everywhere else on } [0,T] \end{cases}$$

where the needle variation is the triple $\pi_1 = (t_1, l_1, u_1)$, $0 < t_1 < T$, $l_1 \geq 0$, u_1 in U. For $\varepsilon > 0$ small enough, the perturbed control is a well defined admissible control with response $x_{\pi_1}(t,\varepsilon)$ starting from x_0. Clearly, $x_{\pi_1}(t,\varepsilon)$ tends to $x(t)$ uniformly on $[0,T]$ when ε tends to 0, and is continuous with respect to (π_1, t, ε). To get differentiability with respect to ε, we require that t_1 be a Lebesgue point so that

$$\int_{t_1-\varepsilon}^{t_1} f(x(t),u(t))dt = f(x(t_1),u(t_1)) + o(\varepsilon).$$

From standard integration theory, the subset $\mathcal{L}$ of Lebesgue points has full measure on $[0,T]$. If π_1 is such a needle variation, then the corresponding response defines a curve at $x(t_1)$, $\alpha(\varepsilon) = x_{\pi_1}(t_1, \varepsilon)$ whose tangent vector is

$$\dot\alpha(0) = l_1(f(x(t_1), u_1) - f(x(t_1), u(t_1))).$$

This comes from the estimate

$$\begin{aligned} x_{\pi_1}(t_1,\varepsilon) &= x(t_1 - l_1\varepsilon) + \int_{t_1 - l_1\varepsilon}^{t_1} f(x_{\pi_1}(t,\varepsilon), u_1)dt \qquad (1.6)\\ &= x(t_1) - \varepsilon l_1 f(x(t_1), u(t_1)) + \varepsilon l_1 f(x(t_1), u_1) + o(\varepsilon). \qquad (1.7)\end{aligned}$$

This tangent vector is called the *elementary perturbation vector* associated to the needle variation and is denoted v_{π_1}.

Remark 1.1. If t_1 is a Lebesgue point, for any positive η, from the definition one can find another Lebesgue point t such that $|t - t_1| \leq \eta$ and $|f(x(t), u(t)) - f(x(t_1), u(t_1))| \leq \eta$.

Parallel displacements along the trajectory. We first recall a standard but crucial result. Let $\dot x = X(x)$ be a smooth differential equation, and let $\varphi_t = \exp tX$ define the local one parameter group. If $\alpha(\varepsilon)$ is a smooth curve at x_0, then $(t,\varepsilon) \mapsto \beta(t,\varepsilon) = \varphi_t(\alpha(\varepsilon))$ is a smooth two-dimensional surface. Let $x(t) = \exp tX(x_0)$ be the reference curve and Φ_t be the matrix valued solution of the variational equation

$$\delta\dot x = \frac{\partial X}{\partial x}(x(t))\delta x \qquad (1.8)$$

with $\Phi_0 = I$. Then, for each fixed t, the derivative at 0 of the curve $\varepsilon \mapsto \beta(t,\varepsilon)$ is the so-called *parallel displacement* $w(t)$ given by

$$w(t) = \Phi_t v$$

where $v = \dot\alpha(0)$. Moreover, if X is analytic and if $\dot\alpha(0) = Y(x_0)$ with Y another analytic vector field, w can be computed using the ad-formula

$$w = \sum_k \frac{t^k}{k!}\mathrm{ad}^k X \cdot Y(x(t))$$

where

$$[X,Y] = \frac{\partial X}{\partial x}(x)Y(x) - \frac{\partial Y}{\partial x}(x)X(x) \qquad (1.9)$$

is the Lie bracket[3] and $\mathrm{ad}X$ is the linear operator $\mathrm{ad}X \cdot Y = [X,Y]$. Extending this result to the time depending case, we can transport an elementary

[3] Beware of the sign, here, opposite to some classical texts, *e.g.* [11].

pertubation vector v_{π_1} at $x(t_1)$ along the reference trajectory. The variational equation is

$$\dot{v} = \frac{\partial f}{\partial x}(x(t), u(t))v$$

and if p is solution of the adjoint equation then, by construction,

$$p(t)v(t) = \text{constant}.$$

We note Φ_{t,t_1} the fundamental matrix solution of the variational equation (1.8) with initial condition $\Phi_{t_1,t_1} = I$. From our previous analysis, we know that if v_{π_1} is the elementary perturbation vector, tangent to the curve $\alpha(\varepsilon)$, then for $t \geq t_1$, $\Phi_{t,t_1} v_{\pi_1}$ is the tangent vector to a curve $\beta(t, \varepsilon)$, image of α by the flow (see Fig. 1.2).

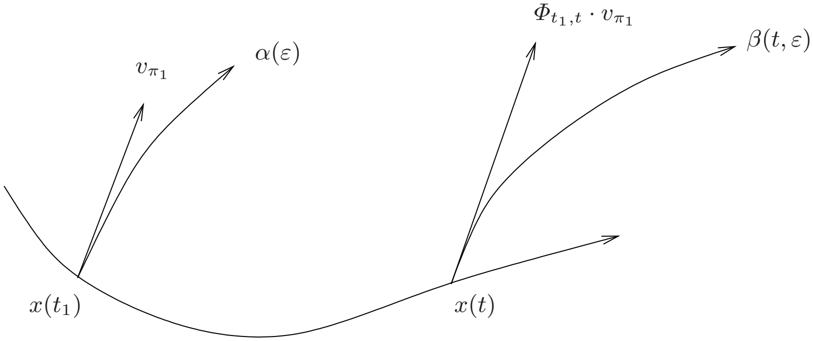

Fig. 1.2. Transport of the elementary perturbation vector by the flow

Definition 1.1. *The* first tangent perturbation cone *or* first Pontryagin cone K_t *at any time* $0 < t \leq T$ *is the smallest convex cone*[4] *in the tangent space at* $x(t)$ *that contains all parallel displacements of all elementary perturbation vectors at Lebesgue points on* $]0, t]$,

$$K_t = \text{cone}(\{\Phi_{t,t_1} v_{\pi_1},\ \pi_1 = (t_1, l_1, u_1) \in \mathcal{L} \times \mathbf{R}_+^* \times U,\ 0 < t_1 \leq t\}).$$

Let now $\pi_i = (t_i, l_i, u_i)$, $i = 1, \dots, k$, be needle variations with distinct times t_i. Let $\pi = (\pi_1, \dots, \pi_k)$ be the *complex variation* associated to the perturbed control

$$u_\pi(t, \varepsilon) = \begin{cases} u_i \text{ on } [t_i - \varepsilon l_i, t_i] \\ u(t) \text{ everywhere else on } [0, T] \end{cases}$$

which is well defined for ε small enough because the t_i are distinct. Clearly, the estimate (1.7) can be extended to complex variations as follows.

[4] For a subset A of $\mathbf{R}^n$, the smallest convex cone containing A is $\text{cone}\, A = \{\sum_{i=1}^{k} \lambda_k x_k,\ k \geq 1, \lambda_1 \geq 0, \dots, \lambda_k \geq 0\}$.

Lemma 1.1. *Let $v_i = \Phi_{t,t_i} v_{\pi_i}$ be parallel displacements of elementary perturbation vectors defined by needle variations $\pi_i = (t_i, l_i, u_i)$ with distinct times t_i, $i = 1, \ldots, k$. Then, the convex combination $\lambda_1 v_1 + \cdots + \lambda v_k$, $\lambda_i \geq 0$ and $\sum_{i=1}^{k} \lambda_i = 1$, is tangent to $x_\pi(t, \varepsilon)$, the response to the perturbed control $u_\pi(t, \varepsilon)$ where π is the complex variation $((t_1, \lambda_1 l_1, u_1), \ldots, (t_k, \lambda_k l_k, u_k))$:*

$$x_\pi(t, \varepsilon) = x(t) + \varepsilon(\lambda_1 v_1 + \cdots + \lambda v_k) + o(\varepsilon).$$

Fundamental lemma. In order to prove the maximum principle, we need a technical lemma which is a consequence of the following byproduct of the Brouwer fixed point theorem [15].

Proposition 1.2. *Let f be a continuous mapping from the closed unit ball B of $\mathbf{R}^n$ into $\mathbf{R}^n$. Let $0 < \varepsilon < 1$ be such that, for all x in the unit sphere S,*

$$|f(x) - x| \leq \varepsilon.$$

Then, $f(B)$ contains the open ball of radius $1 - \varepsilon$ centered at the origin.

Lemma 1.2. *Let v be a nonzero vector interior to K_t, then $x(t) + \varepsilon v$ lies interior to the accessibility set $\mathcal{A}_{x_0,t}$ for all small enough and positive ε.*

Proof. Let v be nonzero and interior to K_t. There are independent parallel displacements $v_1, \ldots, v_n$ in K_t such that v is interior to the convex set generated by $v_1, \ldots, v_n$. Let Π be the hyperplane defined by these vectors. Since v is interior, any point y in the interior of the cone generated by $v_1, \ldots, v_n$ can be written $y = x(t) + \varepsilon(v + r)$ with $\varepsilon > 0$, and r in a suitable open subset of the $n - 1$ dimensional vector space parallel to Π (see Fig. 1.3). For such an r, there are nonnegative scalars $\lambda_1, \ldots, \lambda_n$, $\sum_{i=1}^{n} \lambda_i = 1$, such that $v + r = \lambda_1 v_1 + \cdots + \lambda_n v_n$. Besides, there are needle variations $\pi_i = (t_i, l_i, u_i)$ such that $v_i = \Phi_{t,t_i} v_{\pi_i}$, $i = 1, \ldots, n$, and one can assume all Lebesgue points t_i distinct (see remark 1.1). Hence, for ε small enough it is possible to define the perturbed control u_r associated to the complex variation $(t_i, \lambda_i l_i, u_i)_i$. If x_r denotes the corresponding response, Lemma 1.1 asserts that

$$\begin{aligned} x_r(t, \varepsilon) &= x(t) + \varepsilon(\lambda_1 v_1 + \cdots + \lambda_n v_n) + o(\varepsilon) \\ &= x(t) + \varepsilon(v + r) + o(\varepsilon). \end{aligned}$$

Let then define the continuous mapping $g : (\varepsilon, r) \mapsto x_r(t, \varepsilon)$ into the accessibility set $\mathcal{A}_{x_0,T}$. In coordinates (ε, r), $g(\varepsilon, r) = (\varepsilon + o(\varepsilon), r + o(1))$. As a result, $|g(\varepsilon, r) - (\varepsilon, r)|$ tends to zero when ε does so and, by Proposition 1.2, one can find ε_0, positive and small enough, such that the image by g of $[0, \varepsilon_0] \times \{|r| \leq \varepsilon_0\}$ (with g continuously extended at $\varepsilon = 0$ according to $g(0, r) = (0, r)$) contains $]0, \varepsilon_0[\times \{|r| < \varepsilon_0\}$. Therefore, $\mathcal{A}_{x_0,t}$ is a neighbourhood of $x(t) + \varepsilon v$ for $0 < \varepsilon < \varepsilon_0$, hence the result. ■

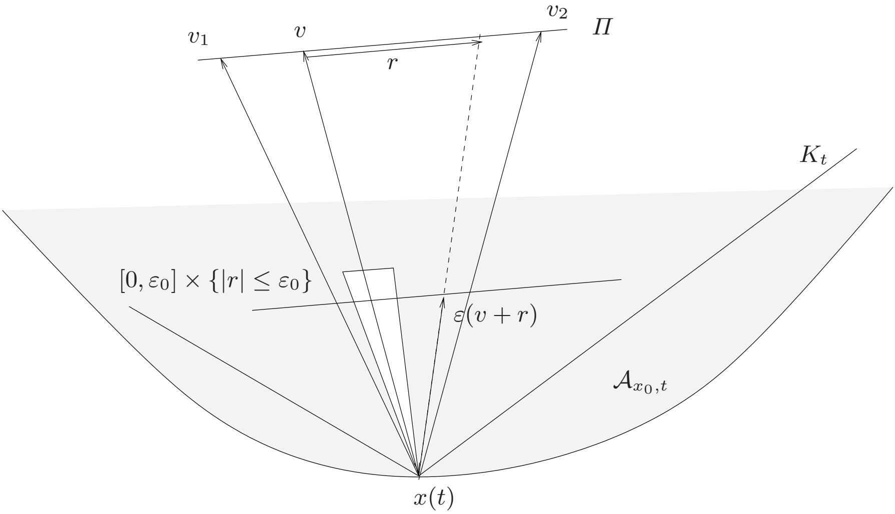

Fig. 1.3. Conical neighbourhood of vector v in the accessibility set

End of the proof. To finish the proof of the maximum principle, we just use a geometric separation argument. Indeed, if $x(T)$ belongs to the boundary of $\mathcal{A}_{x_0,T}$, then there exists a sequence of points x_n not belonging to the interior of the accessibility set, converging to $x(T)$ and such that, up to a subsequence, the unit vectors $(x_n - x(T))/|x_n - x(T)|$ have a limit v when n tends to $+\infty$. This vector v is not interior to K_T otherwise, from the fundamental Lemma 1.2, $x(T) + \varepsilon v$ would be interior to $\mathcal{A}_{x_0,T}$ for any small and positive ε, and so would be x_n for n big enough. The convex cone K_T is thus included in a half-space defined by a separating hyperplane Π. Let $\bar{p}$ be the unit normal to Π oriented outwards K_T, and let us denote p the solution of the adjoint equation

$$\dot{p} = -p\frac{\partial f}{\partial x}(x(t), u(t))$$

satisfying $p(T) = \bar{p}$. Then, the maximization condition must hold almost everywhere. Indeed, let t_1 in $]0, T[$ be a Lebesgue point, and let u_1 be in U. The elementary perturbation vector $v_{\pi_1} = f(x(t_1), u_1) - f(x(t_1), u(t_1))$ associated to $\pi_1 = (t_1, 1, u_1)$ is in K_{t_1}, so $v = \Phi_{T,t_1} v_{\pi_1}$ is in K_T and

$$\langle p(t_1), v_{\pi_1} \rangle = \langle \bar{p}, v \rangle \leq 0$$

that is $H(x(t_1), p(t_1), u_1) \leq H(x(t_1), p(t_1), u(t_1))$. Accordingly, the Hamiltonian is maximized at t_1, $H(x(t_1), p(t_1), u(t_1)) = \max_{v \in U} H(x(t_1), p(t_1), v)$, and the conclusion proceeds from the fact that the set of Lebesgue points has full measure. Standard arguments allow to prove that $t \mapsto M(x(t), p(t)) =$

$\max_{v \in U} H(x(t), p(t), u)$ is absolutely continuous with zero derivative almost everywhere : hence M is constant along (x, p).

Application to Time-optimal Control

We can apply our result to the time-optimal control problem. Indeed, assume that the reference control is time-optimal on $[0, T]$. Then, for each t in $]0, T]$, the point $x(t)$ is in $\partial A(x_0, T)$ so that $\dot{x}(t) = f(x(t), u(t))$ cannot be interior to the first order cone $K(t)$. Indeed, from the fundamental lemma, $x(t+\varepsilon)$ would be in $\mathcal{A}_{x_0,t}$ for $\varepsilon > 0$ small enough, otherwise, contradicting optimality. Hence, we have the additional condition $\langle p(t), f(x(t), u(t))\rangle \geq 0$ and the reduced Hamiltonian is constant and positive.

1.2.5 Maximum Principle, General Case

We formulate the result which can be used to analyze general finite dimensional optimal control problems. We consider a system $\dot{x} = f(x, u)$ written in local coordinates x in $\mathbf{R}^n$, where the set $\mathcal{U}$ of admissible controls is the set of locally bounded functions valued in a fixed control domain $U \subset \mathbf{R}^m$. Let M_0 and M_1 be the regular submanifolds defining the boundary conditions, and let

$$c(x, u) = \int_0^T f^0(x, u)dt$$

be the cost functional assigned to an admissible control and its response x assumed to be defined on $[0, T]$, T *free*. As before, $\widetilde{x} = (x^0, x)$ is the cost extended state and $\widetilde{f}$ the extended dynamics. We assume that $\widetilde{f}$ satisfies the previous regularity assumptions, namely that it is continuous on $\mathbf{R}^{1+n+m}$, together with its partial derivative $\partial\widetilde{f}/\partial\widetilde{x}$. Let

$$\widetilde{H}(\widetilde{x}, \widetilde{p}, u) = p^0 f^0(x, u) + \langle p, f(x, u)\rangle$$

be the extended Hamiltonian and

$$\widetilde{M}(\widetilde{x}, \widetilde{p}) = \max_{v \in U} H(\widetilde{x}, \widetilde{p}, v).$$

Theorem 1.1. *If u is an optimal control on $[0, T]$ then there exists an absolutely continuous extended adjoint covector function $\widetilde{p} = (p^0, p)$, nonzero on $[0, T]$ and such that the following equations are satisfied almost everywhere by the triple $(\widetilde{x}, \widetilde{p}, u)$:*

$$\dot{\widetilde{x}} = \frac{\partial \widetilde{H}}{\partial \widetilde{p}}(\widetilde{x}, \widetilde{p}, u), \ \dot{\widetilde{p}} = -\frac{\partial \widetilde{H}}{\partial \widetilde{x}}(\widetilde{x}, \widetilde{p}, u)$$
$$\widetilde{H}(\widetilde{x}, \widetilde{p}, u) = \widetilde{M}(\widetilde{x}, \widetilde{p}).$$

Moreover, $\widetilde{M}(\widetilde{x},\widetilde{p})=0$ *on* $[0,T]$ *and* p^0 *is constant and non-positive. Eventually,* p *can be selected at the extremities so as to satisfy the transversality conditions*

$$p(0)\perp T_{x(0)}M_0,\ p(T)\perp T_{x(T)}M_1.$$

Proof. We use the necessary conditions for the fixed time case and we extend the cone with additional directions. Indeed, since u is optimal, the endpoint $(x^0(T),x(T))$ of the extended system belongs to the boundary of the extended accessibility set. So there exists a non-trivial augmented adjoint covector $\widetilde{p}=(p^0,p)$ such that, almost everywhere,

$$\widetilde{H}(\widetilde{x},\widetilde{p},u)=\widetilde{M}(\widetilde{x},\widetilde{p})$$

and the maximized Hamiltonian $\widetilde{M}$ is constant along $(\widetilde{x},\widetilde{p})$ on $[0,T]$. In order to extend the first tangent cone $\widetilde{K}_t$ of the extended system, we proceed as follows. Since the time is not fixed, by making time variations $t+\varepsilon\delta t$ of Lebesgue points we can add to $\widetilde{K}_t$ the two vectors $v_\pm=\pm\widetilde{f}(\widetilde{x}(t),u(t))$. The two manifolds M_0, M_1 are embedded into $\mathbf{R}^{1+n}$ by taking $\widetilde{M}_0=(0,M_0)$ and $\widetilde{M}_1=(0,M_1)$, with respective tangent bundles $T\widetilde{M}_0$, $T\widetilde{M}_1$. Since the initial condition is relaxed to $\widetilde{M}_0$, we can add to $\widetilde{K}_t$ the parallel displacements in the tangent space to $\widetilde{M}_0$. Hence, the *second tangent perturbation cone* $\widetilde{K}'_t$, $0<t\leq T$, is defined as the convex cone generated by the vectors:

(i) $\widetilde{\Phi}(t,0)w,\ w\in T_{\widetilde{x}(0)}\widetilde{M}_0$.

(ii) $\widetilde{\Phi}(t,t_1)(\widetilde{f}(\widetilde{x}(t_1),v)-\widetilde{f}(\widetilde{x}(t_1),u(t_1)))$ where v is in U and $t_1\leq t$ is a Lebesgue point.

(iii) $\pm\widetilde{\Phi}(t,t_1)\widetilde{f}(\widetilde{x}(t_1),u(t_1))$ with $t_1\leq t$ a Lebesgue point.

According to Lemma 1.2, for any vector w interior to $\widetilde{K}'_t$ there exists $\lambda>0$ and a conic neighbourhood of λw included in the accessibility set $\widetilde{\mathcal{A}}_{x_0}=\cup_{t>0}\widetilde{\mathcal{A}}_{x_0,t}$. In particular, since $\widetilde{x}(T)$ is optimal, the vector $(-1,0)$ of $\mathbf{R}^{1+n}$ does not belong the interior of $\widetilde{K}'_t$, otherwise we could find an admissible control minimizing the cost even more. In order to obtain the transversality condition at the endpoint, we introduce the cone T_1 at $\widetilde{x}(T)$ which is generated by $T_{\widetilde{x}(T)}\widetilde{M}_1$ and the downward vector $(-1,0)$. The second perturbation cone $\widetilde{K}'_t$ and T_1 are separated by an hyperplane Π. Here, we can take a normal vector $\bar{\widetilde{p}}=(p^0,\bar{p})$ at $\widetilde{x}(T)$ with $p^0\leq 0$ and

$$\bar{\widetilde{p}}\widetilde{K}'_t\leq 0,\ \bar{\widetilde{p}}T_1\geq 0.$$

The corresponding solution $\widetilde{p}$ of the adjoint system with $\tilde{p}(T)=\bar{\widetilde{p}}$ satisfies the maximum principle, including the required transversality conditions. ■

Remark 1.2. The case where the time is fixed can be reduced to our previous case by introducing the time as a new space variable. The result is the same, the maximized Hamiltonian still being constant along $(\widetilde{x}, \widetilde{p})$ but not necessarily zero anymore. The non-autonomous case can be similarly analyzed.

Definition 1.2. *We call* extremal *any triple* (x, p, u) *solution of the Hamiltonian system and verifying the maximization condition. An extremal also satisfying the transversality conditions is called a* BC-extremal.

1.2.6 Maximum Principle and Shooting Problem

Consider any optimal control problem. It is well posed if there exists an optimal solution. This can be checked by applying the standard Filippov theorem (see [15], p. 259). We assume that there is a solution satisfying the maximum principle. If we denote by $M_i^\perp$, $i = 1, 2$, the cotangent lifts

$$M_i^\perp = \{(x, p) \in T^*M_i \mid x \in M_i,\ p \perp T_xM_i\}$$

we can define the *shooting mapping*

$$S\ :\ (x_0, p_0) \in M_0^\perp \mapsto (x(T), p(T)) \in M_1^\perp. \tag{1.10}$$

An important remark is that the Hamiltonian is linear with respect to the adjoint covector p in order that p has to be taken in the projective space $\mathbf{P}^{n-1} \subset \mathbf{R}^n$. With this normalization, the number of equations is equal to the number of variables. For instance, if we consider the time-optimal control problem with fixed extremities x_0, x_1, the shooting problem is to find a time T and an initial adjoint covector p_0 in $\mathbf{P}^{n-1}$ such that

$$x(T, u, p_0) - x_1 = 0$$

where u is computed by means of the maximization condition, and where $x(., u, p_0)$ is the solution of the Hamiltonian system (1.5) with $x(0) = x_0$ and $p(0) = p_0$.

1.2.7 Introduction to the Micro-analysis of the Extremal Solutions

Consider first the case where the control domain U is open. The maximization condition gives us the conditions

$$\frac{\partial \widetilde{H}}{\partial u} = 0,\ \frac{\partial^2 \widetilde{H}}{\partial u^2} \leq 0$$

and the *regular case* occurs when the strict Legendre-Clebsch condition is satisfied:

$$\frac{\partial^2 \widetilde{H}}{\partial u^2} < 0.$$

In this case, applying the implicit function theorem to solve $\partial\widetilde{H}/\partial u = 0$ leads to compute the reference control as a smooth *dynamic feedback*,

$$(x, p) \mapsto u(x, p). \tag{1.11}$$

By plugging (1.11) into $\widetilde{H}$, we define a true Hamiltonian function. However, $\partial\widetilde{H}/\partial u = 0$ is in general a nonlinear equation with several zeros associated to various local maxima $u_i(x, p)$ of $\widetilde{H}$. The master Hamiltonian thus defines several Hamiltonian functions H_i among which an absolute maximum must be chosen. Memory of all those Hamiltonians must be kept since, along a reference extremal, bifurcations between different local maxima may occur to provide the global maximum. This key phenomenon is crucial in the analysis of the extremal solutions, see for instance the pioneering article [8] where the problem is addressed in the framework of calculus of variations with a non-convex one-dimensional Lagrangian function.

1.2.8 Affine Control Systems

In many applications the control system is

$$\dot{x} = F_0(x) + \sum_{i=1}^{m} u_i F_i(x)$$

and, for the time-optimal control problem, the *reduced Hamiltonian* is considered:

$$H = H_0 + \sum_{i=1}^{m} u_i H_i$$

where $H_i = \langle p, F_i \rangle$, $i = 1, \ldots, m$ are the Hamiltonian lifts of the vector fields. In this case, the Hamiltonian is affine in the control and the problem is singular in the sense that

$$\frac{\partial^2 H}{\partial u^2} = 0.$$

Hence, the Legendre-Clebsch condition cannot be used to separate maxima from minima if the extremal solution is interior to the control domain and higher order conditions are required.

1.3 More Second-order Conditions

1.3.1 High-order Maximum Principle

Consider first the time-optimal control problem for a single input affine control system

$$\dot{x} = F_0(x) + uF_1(x)$$

where $|u| \leq 1$. According to the maximum principle, the extremals are solutions of

$$\dot{x} = \frac{\partial H}{\partial p}(x, p, u), \; \dot{p} = -\frac{\partial H}{\partial x}(x, p, u)$$

together with the maximization condition

$$H(x, p, u) = \max_{|v| \leq 1} H(x, p, v)$$

where $H(z, u) = H_0(z) + uH_1(z)$, $H_i = \langle p, F_i \rangle$ for $i = 0, 1$, and $z = (x, p)$.

Definition 1.3. *Let (z, u) be a reference extremal defined on $[0, T]$. It is called* regular *if $u(t) = \text{sign } H_1(z(t))$ almost everywhere on $[0, T]$, and* singular *if $H_1(z(t)) = 0$ for all t in $[0, T]$.*

More general extremals are concatenation of regular and singular subarcs. We begin by computing the singular controls defined by the constraint $H_1(z) = 0$.

Computation of Singular Extremals

The weak and the general maximum principle lead to the same equation, $H_1(z) = 0$. To compute the corresponding trajectories, we differentiate with respect to time and use the Hamiltonian formalism. Differentiating twice with respect to t in $[0, T]$, we get:

$$\{H_1, H_0\}(z(t)) = 0$$
$$\{\{H_1, H_0\}, H_0\}(z(t)) + u(t)\{\{H_1, H_0\}, H_1\}(z(t)) = 0$$

with the Poisson brackets given by $\{H_X, H_Y\} = \langle p, [X, Y] \rangle$ and the Lie bracket defined as in (1.9).

Definition 1.4. *A singular extremal z is said to be of* minimal order *if, everywhere on $[0, T]$,*

$$\{\{H_1, H_0\}, H_1\}(z(t)) \neq 0.$$

Proposition 1.3. *If z is a singular arc of minimal order, the corresponding singular control is the dynamic feedback*

$$u_s(t) = -\frac{\{\{H_1, H_0\}, H_0\}}{\{\{H_1, H_0\}, H_1\}}(z(t))$$

and the extremal curve is smooth and solution of

$$\dot{z} = \overrightarrow{H}_s(z)$$

contained in $H_1 = \{H_1, H_0\} = 0$ and $H_s = H_0 + u_s H_1$.

A Standard Normalization

Hence a singular arc is in general smooth. Take such an arc, $t \mapsto x(t)$. Restricting if necessary its domain of definition, we can assume that it is one-to-one. Hence, it can be identified locally with the curve $\gamma \ : \ t \mapsto (t, 0, \ldots, 0)$ and is the response of to a smooth control denoted u_γ. The control can be normalized to zero by the feedback $v = u - u_\gamma$. Then, differentiating as before $H_1(z) = 0$, one gets that, everywhere

$$\langle p(t), \mathrm{ad}^k F_0 \cdot F_1(\gamma(t)) \rangle = 0, \ k \geq 0.$$

We proved the following.

Proposition 1.4. *Let z be a smooth singular extremal on $[0, T]$, corresponding to a singular control identified to zero. The maximum principle is equivalent to*

$$\mathrm{ad}^k H_0 \cdot H_1(z(t)) = 0, \ k \geq 0 \tag{1.12}$$

everywhere on $[0, T]$.

In (1.12), $\mathrm{ad}^k H_0 \cdot H_1$ denotes the k-th Poisson bracket of H_1 with H_0. This is clearly equivalent to the following lemma.

Lemma 1.3. *Let x be a trajectory defined on $[0, T]$ and associated to the zero control. Assume that, for each t, $V_1(t) = \mathrm{Span}\{\mathrm{ad}^k F_0 \cdot F_1(x(t)), \ k \geq 0\}$ is of maximum rank n. Then, for each $0 \leq t_0 < t_1 \leq T$, the linearized system along x restricted to $[t_0, t_1]$ is controllable and $x(t_1)$ belongs to the interior of the accessibility set $\mathcal{A}_{x(t_0), t_1 - t_0}$.*

This gives a simple interpretation of the maximum principle for single input affine control systems in the open control case.

The Analytic Case

Consider now the case where F_0 and F_1 are real analytic vector fields and let z be the reference extremal defined on $[0, T]$ associated to a control normalized to zero. As before, the maximum principle is subsumed by (1.12) and $V_1(T)$ is the image of the Fréchet derivative of the endpoint mapping which coincides with the first order Pontryagin cone constructed in the proof of the principle. Indeed, if $v_{\pi_1}(t)$ is an elementary perturbation vector with $\pi_1 = (t_1, 1, u_1)$, one has

$$\begin{aligned} v_{\pi_1}(t) &= (F_0 + u_1 F)(x(t_1)) - F_0(x(t_1)) \\ &= u_1 F_1(x(t_1)) \end{aligned}$$

and we can take $u_1 = \pm 1$. The parallel transport can be evaluated using the ad-formula

$$(\exp tF_0)'(F_1(x)) = \sum_k \frac{t^k}{k!} \mathrm{ad}^k F_0 \cdot F_1(x(t))$$

where $x(t) = (\exp tF_0)(x_0)$. Special variations of the reference zero control can be applied to generate the Lie brackets $\mathrm{ad}^k F_0 \cdot F_1(x(t))$. We present a computation based on the Baker-Campbell-Hausdorff formula.

Generalized Variations

To simplify, we restrict ourselves to the $\mathcal{C}^\omega$–real analytic case: $\dot{x} = F_0(x) + uF_1(x)$ where $\gamma(t) = (\exp tF_0)(x_0)$ is the reference singular trajectory associated to the control normalized to zero and defined on $[0, T]$. We assume that $|u| \leq 1$. A *positive rational polynomial* is a function of the form

$$\sum_{i=1}^{p} c_i t^{q_i}, \ c_i \geq 0, \ q_i \in \mathbf{Q}.$$

A vector W belongs to the generalized Pontryagin cone $\mathcal{E}^+$ if there exist positive rational polynomials $r_1, \sigma_1, \ldots, r_{2k}, \sigma_{2k}$ associated to a perturbation π such that:

$$\begin{aligned} \alpha_\pi(y, \varepsilon) &= \exp(\varepsilon W + o(\varepsilon))(y) \\ &= (\exp \sigma_{2k}(\varepsilon) F_0)(\exp r_{2k}(\varepsilon)(F_0 - F_1)) \\ &\quad (\exp \sigma_{2k-1}(\varepsilon) F_0)(\exp r_{2k-1}(\varepsilon)(F_0 + F_1)) \\ &\cdots (\exp \sigma_1(\varepsilon) F_0)(\exp r_1(\varepsilon)(F_0 + F_1)) \\ &\quad (\exp -(\Sigma_{i=1}^{2k} \sigma_k(\varepsilon) + r_k(\varepsilon)) F_0)(y) \end{aligned}$$

where $y = \exp TF_0(x)$. By construction, for $\varepsilon > 0$ small enough $\alpha_\pi(y, \varepsilon)$ is in $\mathcal{A}(x_0, T)$ and W is the right derivative of α at 0. Moreover, from the Baker-Campbell-Hausdorff formula, the derivative belongs to the Lie algebra generated by F_0 and F_1. As for the maximum principle, a crucial property is to have convexity. To prove this property we proceed as follows. Let $\pi_1 = (r_{i,1}, \sigma_{i,1})_i$ and $\pi_2 = (r_{i,2}, \sigma_{i,2})_i$ be two perturbations with respective tangent vectors W_1 and W_2. The composition of π_1 and π_2 is defined as:

$$(\alpha_{\pi_2}(\varepsilon))(\exp \mu(\varepsilon) F_0)(\alpha_{\pi_1}(\varepsilon))(\exp -\mu(\varepsilon) F_0)$$

where $\mu(\varepsilon) = \sum_i \sigma_{i,2}(\varepsilon) + r_{i,2}(\varepsilon)$. From the ad-formula, W_1 is a tangent vector to $(\exp \mu(\varepsilon) F_0)(\alpha_{\pi_1}(\varepsilon))(\exp -\mu(\varepsilon) F_0)$. Therefore, using the Baker-Campbell-Hausdorff formula $\exp X \exp Y = \exp(X + Y + \cdots)$ we have the lemma hereafter.

Lemma 1.4. *The sum $W_1 + W_2$ is the tangent vector corresponding to the composition of π_1 and π_2. In particular, $\mathcal{E}^+$ is a convex cone.*

Next, we prove the following additional result.

Lemma 1.5. *If $\pm W$ is in $\mathcal{E}^+$, then $\pm \mathrm{ad}^k F_0 \cdot W$ is in $\mathcal{E}^+$ as well for $k \geq 0$.*

Proof. We prove the result by recurrence on k. Let $\alpha_{\pi_\pm}(\varepsilon)$ be admissible variations with respective tangent vectors $\pm W$:

$$\alpha_{\pi_\pm}(\varepsilon) = \exp(\pm\varepsilon W + o(\varepsilon^p))$$

where $p > 1$ is a rational number. Let q in $\mathbf{Q}$ be such that $0 < q < 1$ and $pq > 1$, then:

$$\begin{aligned}&(\alpha_{\pi_+}(\varepsilon^q))(\exp \varepsilon^{1-q} F_0)(\alpha_{\pi_-}(\varepsilon^q))(\exp -\varepsilon^{1-q} F_0) = \\ &\quad (\exp(\varepsilon^q W + o(\varepsilon pq)))(\exp \varepsilon^{1-q} F_0)(\exp(-\varepsilon^q W + o(\varepsilon pq)))(\exp -\varepsilon^{1-q} F_0)\end{aligned}$$

which, because of the Baker-Campbell-Hausdorff formula, is equal to

$$\exp(W(\varepsilon^q - \varepsilon^q) + F_0(\varepsilon^{1-q} - \varepsilon^{1-q}) + \varepsilon[W, F_0] + o(\varepsilon)).$$

Thus, $[W, F_0]$ belongs to $\mathcal{E}^+$. ■

In particular, using the previous variations we can recover the conditions from the maximum principle. They concern only the linearized system. An important second-order condition is given by the result hereafter.

Proposition 1.5. *The Lie bracket $[F_1, [F_1, F_0]]$ belongs to $\mathcal{E}^+$.*

Proof. Applying the Baker-Campbell-Hausdorff formula, we get:

$$\begin{aligned}&(\exp \varepsilon^{1/3}(F_0 - F_1))(\exp 2\varepsilon^{1/3}(F_0 + F_1))(\exp \varepsilon^{1/3}(F_0 - F_1))(\exp -4\varepsilon^{1/3} F_0) \\ &\quad = \exp(2\varepsilon/3 \ \mathrm{ad}^2 F_1 \cdot F_0 - 2\varepsilon \mathrm{ad}^2 F_0 \cdot F_1 + o(\varepsilon)).\end{aligned}$$

Hence the vector $\frac{2}{3}\mathrm{ad}^2 F_1 \cdot F_0 - 2\mathrm{ad}^2 F_0 \cdot F_1$ belongs to $\mathcal{E}^+$. Since $\mathcal{E}^+$ is a convex cone containing $\pm \mathrm{ad}^2 F_0 \cdot F_1$, this proves the result. ■

As in the maximum principle, $\mathcal{E}^+$ provides an approximating cone of $\mathcal{A}_{x_0,T}$ and we obtain the following result.

Proposition 1.6. *Let x be a time-optimal trajectory defined on $[0, T]$ and associated to a control normalized to zero. Then, there exists p such that the extremal $z = (x, p)$ satisfies everywhere the conditions:*

(i) $\dot{z} = \overrightarrow{H}_0(z)$, *Hamiltonian system defined by* H_0.

(ii)$\mathrm{ad}^k H_0 \cdot H_1(z(t)) = 0, \ k \geq 0.$

(iii)$H_0(z(t)) \geq 0.$

(iv)$\{H_1, \{H_1, H_0\}\}(z(t)) \geq 0.$

Definition 1.5. *The condition (iv) is called the generalized Legendre-Clebsch condition.*

Application and Geometric Interpretation

Assume that the vector field F_1 is transverse to the trajectory. Then, we can find local coordinates in which $F_1 = \partial/\partial x_n$ so that the system is written

$$\dot{x}' = F(x', x_n),\ \dot{x}_n = F_{0,n}(x) + u$$

where $x' = (x_1, \ldots, x_{n-1})$. The system in x' where x_n is taken as the new control variable is called the *reduced system*. Let $H' = \langle p', F(x', x_n)\rangle$ be the corresponding reduced Hamiltonian, $p' = (p_1, \ldots, p_{n-1})$. A straightforward computation gives

$$\begin{aligned}\frac{d}{dt}\frac{\partial H}{\partial u}(z,u) &= \{H_1, H_0\}(z) = -\frac{\partial H'}{\partial x_n}(z', x_n)\\ \frac{\partial}{\partial u}\frac{d^2}{dt^2}\frac{\partial H}{\partial u}(z,u) &= \{H_1, \{H_1, H_0\}\}(z) = -\frac{\partial^2 H'}{\partial x_n^2}(z', x_n)\end{aligned}$$

along an extremal curve: the generalized Legendre-Clebsch condition is the Legendre-Clebsch condition for the reduced system.

Multi-Input Case, Goh Condition

Similarly, higher order variations can be applied in the multi-input case to obtain further necessary conditions. The most important are the so-called *Goh conditions* that we present now. Consider a system of the form

$$\dot{x} = F_0(x) + \sum_{i=1}^{m} u_i F_i(x).$$

If $z = (x, p, u)$ is a reference singular extremal defined on $[0, T]$ then, in order to be time-optimal, the following condition has to be satisfied:

$$\{H_v, H_w\}(z(t)) = 0,\ t \in [0, T]$$

for every pair of vector fields F_v and F_w in $\mathrm{Span}\{F_1, \ldots, F_m\} \cdots$

1.3.2 Intrinsic Second-order Derivative and Conjugate Times

In the previous section we have generated special variations to obtain further necessary conditions for affine control systems. They concern Lie brackets of the form $[F_1,[F_1,F_0]]$ (generalized Legendre-Clebsch condition), or $[v,w]$ with v,w in $\mathrm{Span}\{F_1,\dots,F_m\}$ (Goh condition). These brackets are related to the second-order derivative and to the necessary conditions for *small time* optimality. We introduce now a different concept related to the loss of optimality because of the *cumulated effect of time*. It is the concept of *conjugate time* associated to the spectral properties of the intrinsic second-order derivative, and to the notion of Lagrangian manifolds in symplectic geometry. We begin by presenting these geometric tools.

Symplectic Geometry and Lagrangian Manifolds

Linear symplectic manifolds and symplectic group. We recall some standard facts about symplectic geometry. Let (V,ω) be a linear symplectic space of dimension $2n$. We can choose a basis called *Darboux* or *canonical linear coordinates* such that $V \simeq \mathbf{R}^{2n}$ and $\omega(x,y) = {}^t xJy$ where

$$J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}. \tag{1.13}$$

A subspace L of V is called *isotropic* if $\omega|_L = 0$. An isotropic of maximal dimension n is called a *Lagrangian subspace*. Linear isomorphisms preserving ω are called *symplectomorphisms* and, in Darboux coordinates, they are identified with the elements of the *symplectic group* $\mathrm{Sp}(n,\mathbf{R})$ of matrices S satisfying ${}^tSJS = J$. Decomposing S into $n\times n$ blocks,

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

we obtain the relations:

$${}^tAD = {}^tBC = I,\ \ {}^tAC = {}^tCA,\ \ {}^tBD = {}^tDB.$$

The Lie algebra $\mathfrak{sp}(n,\mathbf{R})$ of $\mathrm{Sp}(n,\mathbf{R})$ is the algebra of order $2n$ matrices H such that $\exp tH$ is in $\mathrm{Sp}(n,\mathbf{R})$. These matrices are characterized by ${}^tHJ + H\,{}^tJ = 0$ and, decomposing H into blocks,

$$H = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

we obtain the equivalent definition:

$$\mathfrak{sp}(n,\mathbf{R}) = \{H = \begin{bmatrix} A & B \\ C & -{}^tA \end{bmatrix},\ B \text{ and } C \text{ symmetric}\} \cdots$$

The symplectic group acts on Lagrangian subspaces and we have the following representation of Lagrangian subspaces. Let L be a Lagrangian subspace, and let $\Pi : (x,p) \mapsto x$ be the canonical projection written in Darboux coordinates. If the restriction to L of Π is regular, L can be represented as

$$\begin{bmatrix} x \\ Cx \end{bmatrix}$$

that is as the image of $\{x\}$ by the $2n \times n$ matrix

$$\begin{bmatrix} I \\ C \end{bmatrix}$$

where C is symmetric. More generally, let L be a Lagrangian subspace represented by the $2n \times n$ matrix

$$\begin{bmatrix} A \\ B \end{bmatrix}.$$

Then, from the definition, one must have ${}^tAB - {}^tBA = 0$ and the matrix

$$\begin{bmatrix} A & -B \\ B & A \end{bmatrix}$$

is symplectic. In particular, the symplectic group acts transitively on the Lagrangian subspaces.

Symplectic and Lagrangian manifolds on the cotangent bundle. On the cotangent bundle T^*M of any smooth manifold M exists a canonical symplectic structure associated with the *Liouville form* written in coordinates as $\alpha = pdx$, where x are coordinates on M and p the dual ones. The symplectic form is defined by $\omega = d\alpha = dp \wedge dx$. We denote by Π the standard projection, $\Pi : (x,p) \in T^*M \mapsto x \in M$. Locally, we can identify M with $\mathbf{R}^n$, but globally, an important topological invariant is the space H^1 which is the quotient of the space of closed 1-forms by the space of exact 1-forms. If L is a regular submanifold of (T^*M, ω), it is called isotropic (*resp.* Lagrangian) if at each point the tangent space is isotropic (*resp.* Lagrangian). A canonical example in $\mathbf{R}^{2n}$ is constructed as follows. Let $S : x \mapsto S(x)$ be a smooth function in $\mathbf{R}^n$ and consider the graph $L = \{p = \partial S/\partial x\}$: L is a Lagrangian manifold and the projection $\Pi : L \to \mathbf{R}^n$ is regular. We generalize now the representation result of Lagrangian manifolds obtained in the linear case.

Proposition 1.7. *Let L be a Lagrangian manifold of (T^*M, ω). Then, locally, there are Darboux coordinates (x,p) together with a smooth function S of $(x_I, p_{\bar I})$ with $I = \{1, \ldots, m\}$ and $\bar I = \{m+1, \ldots, n\}$ such that L is defined by the equations*

$$p_I = \frac{\partial S}{\partial x_I}, \; x_{\bar I} = -\frac{\partial S}{\partial p_{\bar I}}\cdot$$

The mapping S is called the generating mapping of L.

Definition 1.6. *Let L be a Lagrangian submanifold of (T^*M, ω). A nonzero vector v, tangent to L at x, is called* vertical *whenever $d\Pi(x)v = 0$. The* caustic *of L is the set of points at which there exists at least one vertical tangent vector.*

Example 1.1. For any x in M, the fiber $L = T_x^*M$ is a linear Lagrangian submanifold and all tangent vectors are vertical. More generally, if M_0 is a regular submanifold of M, the submanifold $M_0^\perp$ defined by the transversality relation

$$M_0^\perp = \{(x,p) \in T^*M \mid p \perp T_xM\}$$

is a Lagrangian submanifold of M.

Hamiltonian vector fields and variational equation. In order to simplify our presentation, we use local coordinates, identifying locally M to $\mathbf{R}^n$, T^*M to $\mathbf{R}^{2n}$, and ω to the standard 2-form $dp \wedge dx$. Hence, any time dependent Hamiltonian vector field is defined by the equations

$$\dot{x} = \frac{\partial H}{\partial p}(t,z),\ \dot{p} = -\frac{\partial H}{\partial x}(t,z)$$

where $z = (x,p)$ and $H(t,z)$ is the Hamiltonian. Using J as defined by (1.13), the previous equation can be written in the compact form

$$\dot{z} = J\nabla_z H(t,z) \tag{1.14}$$

where ∇_z stands for the gradient with respect to z. When the Hamiltonian is a quadratic form

$$H(t,z) = \frac{1}{2}\,{}^t z S(t) z$$

with $S(t)$ symmetric, we get a linear Hamiltonian system

$$\dot{z} = JS(t)z = A(t)z$$

where $A(t)$ is a Hamiltonian matrix element of $\mathfrak{sp}(n, \mathbf{R})$. In order to make our geometric analysis, an important issue is the action of the group of symplectic transformations on Hamiltonian vector fields. Let $\dot{z} = J\nabla_z H(t,z)$ be a Hamiltonian vector field and consider a change of variables $z \mapsto \xi = \Phi(t,z)$. The transformation is *symplectic* if $\partial\Phi(t,z)/\partial z$ belongs to the symplectic group. Computing, one has

$$\dot{\xi} = \frac{\partial \Phi}{\partial t}(t,z) + \frac{\partial \Phi}{\partial z}(t,z)\dot{z}.$$

Since the transformation is symplectic,

$$\frac{\partial \Phi}{\partial z}(t,z) J\nabla_z H(t,z) = J\nabla_\xi \hat{H}(t,\xi)$$

where $\hat{H}(t,\xi) = H(t,z)$. Using Poincaré lemma, we can write locally

$$\frac{\partial \Phi}{\partial t}(t,z) = J\nabla_\xi R(t,\xi)$$

where R is called the *remainder function*: we have showed that any symplectic change of coordinates transforms a Hamiltonian vector field into another Hamiltonian vector field. If $z(t,t_0,z_0)$ is the solution of (1.14) starting from z_0 at t_0, then the flow $z_0 \mapsto z(t,t_0,z_0)$ is symplectic for any fixed t, t_0. Differentiating with respect to z we define the variational equation

$$\delta\dot{z} = J\nabla^2_{z^2} H(t, z(t,t_0,z_0))\delta z.$$

Symplectomorphisms induce time dependent linear symplectic isomorphisms on the corresponding variational equation. The action of the linear symplectic group on linear Hamiltonian differential equations is a standard action and numerous tensor analysis exist in the litterature. For instance, a standard result is the following.

Proposition 1.8. *Let $\dot{x} = A(t)$ be a Hamiltonian differential equation on $\mathbf{R}^{2n}$ and let $z_1,\dots,z_n$ be n independent solutions such that $\omega(z_i,z_j) = 0$. Then, a complete set of solutions can be computed by quadrature.*

Proof. Let L be the $2n \times n$ matrix whose columns are the independent solutions. By construction, it is a one parameter Lagrangian manifold and we have

$$\dot{L} = A(t)L,\ \ {}^tL(t)JL(t) = 0.$$

Since the solution are independent, the matrix ${}^tLL$ is a non singular $n \times n$ matrix. Define the $2n \times n$ matrix $L' = JL({}^tLL)^{-1}$. Hence, ${}^tL'JL = 0$ and ${}^tLJL' = -I$. Therefore, $P = (L', L)$ is a symplectic matrix and we have

$$P^{-1} = \begin{bmatrix} -{}^tLJ \\ {}^tL'J \end{bmatrix}.$$

If we make the symplectic change of coordinates $x = Py$, we get the Hamiltonian equation

$$\dot{y} = P^{-1}(AP - \dot{P})y$$

and using the notation $\dot{x} = Ax = JS$ where S is symmetric. Decomposing $y = (u,v)$, we obtain the equations

$$\begin{aligned} \dot{u} &= 0 \\ \dot{v} &= -{}^tL'(SL' + J\dot{L}')u \end{aligned}$$

and the solution can be computed by quadrature. ■

Similar tensor analysis can be developed to study the standard LQ problem.

Geometric analysis of linear quadratic problems. Consider the smooth linear system in $\mathbf{R}^n$ $\dot{x} = A(t) + B(t)u$ and the problem of minimizing a cost defined by

$$c(x,u) = \int_0^T ({}^t xW(t)x + {}^t uU(t)u)dt$$

with fixed time $T > 0$ and prescribed boundary conditions. The symmetric matrices $W(t)$ and $U(t)$ are smooth with respect to t and we assume that the strict Legendre-Clebsch condition holds for all t: $U(t) > 0$. By applying a proper feedback we can renormalize $U(t)$ to I. If we apply the maximum principle, the optimal solutions have to be found among the following extremals:

$$\dot{x} = A(t)x + B(t)U^{-1}(t)\,{}^tB(t)p \tag{1.15}$$
$$\dot{p} = {}^tW(t)x - {}^tA(t)p \tag{1.16}$$

which can be rewritten $\dot{z} = Hz$ with

$$H = \begin{bmatrix} A & C \\ D & -{}^tA \end{bmatrix}$$

and $C = B\,{}^tB$ ($U(t)$ being identified with I). We can assume B of full rank so that C is definite positive. In order to identify a *curvature*-like invariant connected to the optimality properties of the reference solution, a standard reduction is to write the equation as a second-order differential equation. Using the first equation, we write

$$p = C^{-1}(\dot{x} - Ax)$$

which, plugged into the second equation, gives after a left product by C,

$$\ddot{x} + \widetilde{A}\dot{x} + \widetilde{B}x = 0.$$

By setting $x(t) = S(t)X(t)$ where $S(t)$ is properly chosen it can be written

$$\ddot{X} + K(t)X = 0.$$

The matrix $K(t)$ is the curvature invariant of the problem, related to the distribution of conjugate points to be defined later. It is an invariant of the action of the symplectic subgroup of matrices of the form

$$P(t) = \begin{bmatrix} A(t) & 0 \\ B(t) & C(t) \end{bmatrix}$$

which preserves in fact the subspace δx because we must keep track of the state space. By counting the respective dimensions, a normal form contains $n(n+1)/2$ parameters which correspond to the symmetric tensor identified in our reduction. Another useful representation which will be used later is the

Riccati equation. Let Φ be the fundamental matrix solution of (1.15)-(1.16). Decomposing $\Phi(t)$ into $n \times n$ blocks

$$\Phi(t) = \begin{bmatrix} \Phi_1(t) & \Phi_3(t) \\ \Phi_2(t) & \Phi_4(t) \end{bmatrix}$$

we define the one parameter family of Lagrangian subspaces

$$L(t) = \begin{bmatrix} \Phi_3(t) \\ \Phi_4(t) \end{bmatrix}.$$

The projection $\Pi : (x,p) \mapsto x$ restricted to L is regular if and only if the matrix $\Phi_3(t)$ is invertible. We have

$$\begin{bmatrix} \dot{\Phi}_3 \\ \dot{\Phi}_4 \end{bmatrix} = \begin{bmatrix} A & B\,{}^tB \\ W & -\,{}^tA \end{bmatrix} \begin{bmatrix} \Phi_3 \\ \Phi_4 \end{bmatrix}.$$

In the regular case, we introduce $R(t) = \Phi_4(t)\Phi_3^{-1}(t)$ which satisfies the symmetric Riccati equation

$$\dot{R} = W - {}^tAR - RA - RB\,{}^tBR$$

whose solution is symmetric whenever $R(0)$ is symmetric.

Symplectic transformation and generating function. Let φ be a symplectomorphism. Let us prove that, locally, φ is parameterized by a *generating function*. We proceed as follows. Since the result is local, we identify the symplectic space with $\mathbf{R}^{2n}$. Let $\varphi : (x,p) \mapsto (X,P)$ be a symplectic change of coordinates. Then, the 1-form $\sigma_1 = xdp - XdP$ is closed. Assume that (p,P) define coordinates then, locally, there is a function $S_1(p,P)$ such that $\sigma_1 = dS_1$ and we get the relation

$$x = \frac{\partial S_1}{\partial p},\ X = -\frac{\partial S_1}{\partial P}$$

which defines locally the change of coordinates. We proceed similarly with the 1-forms

$$\begin{aligned} \sigma_2 &= xdp + PdX \\ \sigma_3 &= pdx - PdX \\ \sigma_4 &= pdx + XdP \end{aligned}$$

to which we associate the generating mappings S_2, S_3, S_4. In particular, each diffeomorphism $X = \varphi(x)$ can be lifted onto a symplectomorphism $\overrightarrow{\varphi}$ given by

$$X = \varphi(x),\ p = \frac{{}^t\partial\varphi}{\partial x}(x)P$$

and defined by the generated mapping $S_4(x,P) = {}^t\varphi(x)P$. The next step is to define the geometric concept of conjugate point.

Definition 1.7. *Let $\overrightarrow{H}(t,z)$ be a smooth Hamiltonian vector field whose integral curves are the extremals of an optimal control problem with fixed time T. Let $z=(x,p)$ be a reference extremal. Then the variational equation*

$$\delta\dot{z} = \frac{\partial\overrightarrow{H}}{\partial z}(t,z(t))\delta z$$

is called the Jacobi equation*. A* Jacobi field *$J=\delta z$ is a non-trivial solution of this equation. In accordance with the Lagrangian terminology (see def. 1.6), it is called* vertical *at time t if $\delta x(t)=0$, that is if $d\Pi(z(t))J(t)=0$. The time t_c is called* conjugate *if there exists a Jacobi field vertical both at $t=0$ and t_c. In this case, $x(t_c)$ is said to be conjugate to $x(0)$ along the reference solution.*

Definition 1.8. *If $z(t,t_0,z_0)$ is the integral curve of $\overrightarrow{H}(t,z)$ with initial condition z_0 at $t=0$, the* exponential mapping *at t is defined by*

$$\exp_{x_0,t} \; : \; p_0 \mapsto \Pi(z(t,x_0,p_0)).$$

The following result is a consequence of the previous analysis.

Proposition 1.9. *Let z be a reference extremal with initial condition $z_0=(x_0,p_0)$ defined on $[0,T]$. Let L_0 be the fiber $T^*_{x_0}M$ and let L be its image by the one parameter group* $\exp t\mathbf{H}$*. Then L is a one parameter family of Lagrangian submanifolds along the reference extremal curve and t_c is conjugate if and only if (L,Π) is singular at t_c, that is if p_0 is a singular point of the exponential mapping at time t_c.*

The generalization to control problems with arbitrary initial conditions is straightforward.

Definition 1.9. *Let $\overrightarrow{H}(t,z)$ be a smooth Hamiltonian vector field whose integral curves are the extremals of an optimal control problem with fixed time T and initial manifold M_0. The time t_f is a* focal time *along the BC-extremal z if there is a Jacobi field J such that $J(0)$ is in $T_{z(0)}M_0^{\perp}$ and J is vertical at t_f.*

Both concepts fit in the same geometric framework: a one parameter family of Lagrangian manifolds obtained by transporting the initial submanifold with the flow. The Jacobi fields span the tangent spaces of the Lagrangian manifolds computed along the reference extremal. They are the image of the initial tangent space by the fundamental matrix of the variational equation and conjugate or focal points are obtained using a verticality test. Curvature type invariants are related to tensor analysis of each problem. The analysis in the next paragraph shows the connection between the concept of conjugate point and the intrinsic second-order derivative. We derive $\mathcal{C}^1$-sufficient second order optimality conditions in the smooth case.

Conjugate Points of Smooth Time-optimal Control Problems

Preliminaries. We restrict our presentation to a smooth time optimal control problem $\dot{x} = f(x, u)$ where u belongs to U, assumed to be an open subset of $\mathbf{R}^m$. The Hamiltonian of the problem is

$$\widetilde{H}(x, p, u) = p^0 + H(x, p, u)$$

with $H(x, p, u) = \langle p, f(x, u)\rangle$ and $p^0 \leq 0$. The scalar p^0, dual to the cost functional $c(x, u) = 1$, can be normalized to 0 or to 1. The case $p^0 = 1$ is called the *normal case*. The maximum principle asserts that time-optimal solutions satisfy $\partial H/\partial u = 0$ and $\partial^2 H/\partial u^2 \leq 0$. In the so-called regular case, the strict Legendre-Clebsch condition holds and $\partial H/\partial u = 0$ is solved by the implicit function theorem. By plugging the dynamic feedback $\hat{u} \ : \ (x, p) \mapsto \hat{u}(x, p)$ into H, a true Hamiltonian function H_r is defined:

$$H_r(x, p) = H(x, p, \hat{u}(x, p)).$$

As usual, $t \mapsto z(t, z_0)$ is the extremal solution with initial condition $z_0 = (x_0, p_0)$. Since H is linear in p, we have a first lemma.

Lemma 1.6. *The two components of an extremal solution verify*

$$x(t, x_0, \lambda p_0) = x(t, x_0, p_0), \ \ p(t, x_0, \lambda p_0) = \lambda p(t, x_0, p_0).$$

In particular, the rank of the exponential mapping $\exp_{x_0,t}$ *at a given time* t *is at most* $(n-1)$*.*

The aim of this section is twofold. First, thanks to the concept of conjugate point, we obtain second-order necessary and sufficient conditions for optimality, the set of controls being endowed with the L^∞ topology. Then, using standard field theory, we extend those optimality results to the $\mathcal{C}^0$ topology on the set of trajectories of the system. In order to carry out a more complete analysis applicable to affine systems, we make the following prolongation. We set $\dot{u} = v$ and extend the original system to a control affine one:

$$\begin{aligned} \dot{x} &= f(x, u) \\ \dot{u} &= v. \end{aligned}$$

If we write the system $\dot{y} = F_0(y) + \sum_{i=1}^m v_i F_i(y)$ with $y = (x, u)$, the controlled distribution is flat: $[F_i, F_j] = 0$, $i, j = 1, \ldots, m$. Our analysis also applies to control affine systems whose distribution is involutive. A prototype of such systems is the single input control system of the form: $\dot{y} = F_0(y) + vF_1(y)$. Having made our prolongation, we must change the L^∞ control topology on u into the L^1 topology on $v = \dot{u}$. According to Section 1.3.1, there is a one-to-one correspondance between the extremal solutions of the original system and the affine system obtained by prolongation. As a consequence, we shall be able to translate the relevant optimality results.

Second order sufficient optimality conditions for single input affine systems. We consider a single input affine control system

$$\dot{x} = F_0(x) + uF_1(x)$$

and we assume that the control domain is $U = \mathbf{R}$. The controlled vector field F_1 is called the *cheap direction* and time-optimal curves are to be searched among concatenation of standard extremals with jumps into this cheap direction. We compute a normal form under the action of the *feedback group*. The group acts locally with the following transformations:

(i) Change of coordinates, $y = \varphi(x)$,

$$(F_0, F_1) \mapsto (\varphi_* F_0, \varphi_* F_1).$$

(ii) Feedback transformation, $v = \alpha(x) + \beta(x)u$ where β is invertible,

$$(F_0, F_1) \mapsto (F_0 + \alpha F_1, \beta F_1).$$

The following result is standard.

Proposition 1.10. *The singularities of the endpoint mapping corresponding to extremals curves of the time-optimal control problem are feedback invariant.*

Hence, we shall use the action of the feedback group to normalize our system along a reference extremal, each change of coordinates $y = \varphi(x)$ being lifted onto a symplectic diffeomorphism $\overrightarrow{\varphi}$ acting on the extremal flow.

Geometric reduction. We proceed in two steps. We first pick a reference smooth extremal trajectory γ defined on $[0, T]$. Assuming it is one-to-one, we can identify it with $t \mapsto (t, 0, \ldots, 0)$ in suitable coordinates $x = (x_1, \ldots, x_n)$. A tubular neighbourhood of γ is characterized by small x_i's for $i \geq 2$. Then we consider the Taylor expansion of the pair F_0, F_1 along γ: the *jet* of order one (*resp.* two) is the collection of all linear (*resp.* quadratic) terms. The control is also normalized to zero thanks to the feedback $v = u - u(x_1)$ (see Section 1.3.1). Besides, if F_1 is tranverse to γ, we can choose the coordinates in the neighbourhood of the curve such that F_1 is identified with $\partial/\partial x_n$. From our preliminary analysis, we know that the first order Fréchet derivative of the endpoint mapping depends only upon the jet of order one, while the second-order intrinsic derivative depends only upon the jet of order two. Furthermore, all the information about first and second variations is collected by Lie brackets within the two spaces $E_1(t) = \mathrm{Span}\{\mathrm{ad}^k F_0 \cdot F_1(\gamma(t))\}$ and $E_2(t)$ which is generated by the restriction to γ of Lie brackets with at most two occurences of F_1. The second normalization is performed choosing a reference extremal meeting the generic requirements hereafter:

(i) $E_1(t)$ is of codimension one and is generated by the first $(n-1)$ brackets, $\mathrm{ad}^k F_0 \cdot F_1(\gamma(t))$, $k = 0, \ldots, n-2$, for any t in $[0, T]$.

(ii) The Lie bracket $\mathrm{ad}^2 F_1 \cdot F_0(\gamma(t))$ is not contained in $E_1(t)$ for t in $[0, T]$.

(iii)The vector field F_0 restricted to γ is tranverse to $E_1(t)$ on $[0, T]$.

This has the following implications: first, for each $0 < t_0 < t_1 \leq T$, the singularity of the endpoint mapping at the zero control defined on $[t_0, t_1]$ is of codimension one and the image of its Fréchet derivative is $E_1(t_1)$. Secondly, the adjoint covector p is unique up to a scalar and oriented in order that $H_0 = \langle p, F_0 \rangle$ be positive. The singular trajectory which is of minimal order by virtue of requirement (ii), is said to be *hyperbolic* if $\langle p, \mathrm{ad}^2 F_1 \cdot F_0(\gamma(t)) \rangle < 0$ on $[0, T]$, *elliptic* if $\langle p, \mathrm{ad}^2 F_1 \cdot F_0(\gamma(t)) \rangle > 0$. Observe that the generalized Legendre-Clebsch condition is only satisfied in the hyperbolic case. It is now crucial to notice that since the reference curve is a one-dimensional manifold, we can normalize any independent family of Lie brackets to form a frame along it. Our assumptions allow us to pick coordinates preserving the previous normalizations and defining a moving frame defined by:

$$\mathrm{ad}^k F_0 \cdot F_1(\gamma(t)) = \frac{\partial}{\partial x_{n-k}},\ k = 0, \ldots, n-1,\ t \in [0, T].$$

Moreover, since the feedback is chosen so that u is zero along γ, we can impose the linearization condition $\mathrm{ad}^k F_0 \cdot F_1(\gamma(t)) = 0$ for $k > n-2$ and $t \in [0, T]$. These computations can be explicited. In particular, the moving frame construction amounts to a time dependent linear transformation. Having made these normalizations, we have the following.

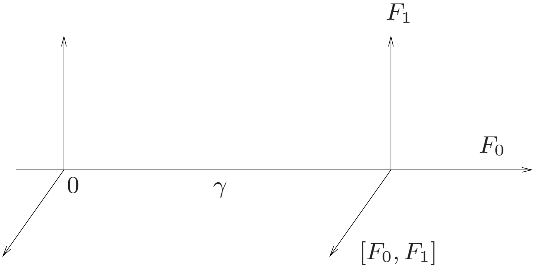

Fig. 1.4. Canonical moving frame

Proposition 1.11. *Along the reference curve, the system is feedback equivalent to the system defined by the two vector fields*

$$\begin{aligned} F_0 &= \frac{\partial}{\partial x_1} + \sum_{i=2}^{n-2} x_{i+1} \frac{\partial}{\partial x_i} + \sum_{i,j=2}^{n} a_{ij}(x_1) x_i x_j \frac{\partial}{\partial x_1} + R \\ F_1 &= \frac{\partial}{\partial x_n} \end{aligned}$$

where the remainder $R = \sum_{i=1}^{n} R_i \frac{\partial}{\partial x_i}$ *is such that the* 1*-jets of the* R_i*'s along* γ *are zero,* $i = 1, \dots, n$*, as well as the* 2*-jets for* $i \geq 2$*.*

Definition 1.10. *The truncated system*

$$\begin{aligned} F_0 &= \frac{\partial}{\partial x_1} + \sum_{i=2}^{n-2} x_{i+1} \frac{\partial}{\partial x_i} + \sum_{i,j=2}^{n} a_{ij}(x_1) x_i x_j \frac{\partial}{\partial x_1} \\ F_1 &= \frac{\partial}{\partial x_n} \end{aligned}$$

is called the approximating model *along* γ*.*

Properties of the model. In this model, we have gathered in one normal form all the information required to evaluate the endpoint mapping (and thus the accessibility set) up to second-order relevant terms. The adjoint covector is oriented by the condition $H_0 \geq 0$ and normalized to $p = (1, 0, \dots, 0)$ The linearized system along the reference trajectory is a constant linear system in Brunovsky normal form. Indeed,

$$\begin{aligned} \dot{x}_1 &= 1 + q(x_1, x_2, \dots, x_n) \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_n &= u \end{aligned}$$

with $q(x_1, x_2, \dots, x_n) = \sum_{i,j=2}^{n} a_{ij} x_i x_j$. Setting $x(t) = t + \xi(t)$ we get

$$\begin{aligned} \dot{\xi}_1 &= \sum_{i,j=2}^{n} a_{ij} x_i x_j \\ \dot{\xi}_2 &= \xi_3 \\ &\vdots \\ \dot{\xi}_n &= u \end{aligned}$$

and the system describing the evolution of ξ is the linearized system. We substitute x_1 by t in the quadratic form q. The last diagonal coefficient $a_{nn}(t)$ is

$\langle p, \partial^2 F_0/\partial x_n^2\rangle(\gamma(t))$, which is also equal to the opposite of the Poisson bracket $\{\{H_1, H_0\}, H_1\}(z(t))$ involved in the generalized Legendre-Clebsh condition: it is negative in the hyperbolic case, and positive in the elliptic one. The kernel K_t of the first order derivative is $\dot{\xi}_2 = \xi_3, \ldots, \dot{\xi}_{n-1} = \xi_n$ with the boundary conditions $\xi_2(0) = \cdots = \xi_n(0) = 0$, $\xi_2(t) = \cdots = \xi_n(t) = 0$, and the quadratic form q represents in fact the intrinsic second-order derivative defined on $[0, T]$ by the restriction to K_t of

$$Q_t(\xi) = \int_0^t \sum_{i,j=2}^{n} a_{ij}(s)\xi_i(s)\xi_j(s)ds.$$

By construction, our affine system is the prolongation of a regular system in $\mathbf{R}^{n-1}$ where the control variable is $x_n = \xi_n$.

Accessory problem and intrinsic derivative. By taking ξ_n as control variable and approximating by the model, clearly, the reference extremal curve is time-optimal on $[0, T]$ if and only if Q_t is negative for each t in $]0, T]$. This leads to consider the so-called *accessory problem*, $\varepsilon Q_t \to \min$, with $\varepsilon = -1$ in the hyperbolic case, and $\varepsilon = 1$ in the elliptic one. This is a standard problem in differential operator theory. We can rewrite the intrinsic second order derivative as

$$Q_t = \int_0^T q(y(s))ds$$

with $y = \xi_2$ and where

$$q(y(t)) = \sum_{i,j=0}^{n-2} b_{ij}(t)y^{(i)}(t)y^{(j)}(t)$$

the b_{ij} being symmetric functions. The boundary conditions on $[0, T]$ define the set $\mathcal{C}_t$ of smooth curves such that $y(0) = \cdots = y^{(n-3)}(0) = 0$, $y(t) = \cdots = y^{(n-3)}(t) = 0$. Let D be the differential operator of order $2(n-2)$ defined by

$$Dy = \frac{1}{2}\sum_{i=0}^{n-2}(-1)^i \frac{d^i}{dt^i}\frac{\partial q}{\partial y^{(i)}}(y).$$

It is the *Euler-Lagrange operator* associated to the accessory minimization problem and it can be written

$$Dy = \sum_{i,j=0}^{n-2}(-1)^j \frac{d^j}{dt^j} b_{ij}(t) \frac{d^i}{dt^i}\cdot$$

Its restriction D_t to $\mathcal{C}_t$ is a self-adjoint differential operator representing the second-order intrinsic derivative. This operator is regular since $b_{n-2,n-2}(t) = \{\{H_1, H_0\}, H_1\}(\gamma(t))$ is nonzero. The following result holds.

Lemma 1.7. *The equation $Dy = 0$ is equivalent to Jacobi equation along the reference extremal and is Euler-Lagrange equation associated to the accessory problem. If J is a Jacobi field solution of the variational equation, then J is vertical at 0 and t_c if and only if $D_{t_c} J = 0$ so that*

$$Q_{t_c}(J) = \int_0^{t_c} D_{t_c} J(s) \cdot J(s) ds = 0.$$

The spectral properties of Q_t are investigated using the classical theory on linear differential operators, see [17], and we get the next proposition.

Proposition 1.12. *For each t in $]0, T]$, there exists a sequence of eigenvectors and eigenvalues $(e_{t,\alpha}, \lambda_{t,\alpha})_{\alpha \geq 1}$ such that*

(i) The eigenvectors $e_{t,\alpha}$ belong to $\mathrm{L}^2([0,T]) \cap \mathcal{C}_t$ and $D_t e_{t,\alpha} = \lambda_{t,\alpha} e_{t,\alpha}$.

(ii) Each curve y in $\mathcal{C}_t$ can be represented by its uniformly convergent Fourier series,

$$y = \sum_{\alpha \geq 1} y_\alpha e_{t,\alpha}.$$

We order the eigenvalues increasingly, $\lambda_{t,1} \leq \lambda_{t,2} \leq \ldots$, and state the Morse result about the time evolution of the spectrum of D_t.

Proposition 1.13. *Let y be in $\mathcal{C}_t$ with Fourier series $y = \sum_{\alpha \geq 1} y_\alpha e_{t,\alpha}$. Then $Q_t(y) = \sum_{\alpha \geq 1} \lambda_{t,\alpha} y_\alpha^2$ and Q_t is positive for t small enough. The first conjugate time to 0, t_{1c}, is the smallest t such that $\lambda_{t,1} = 0$. If $t < t_{1c}$, the only minimizer of Q_t on $\mathcal{C}_t$ is $y = 0$. If $t > t_{1c}$, the infimum of Q_t is $-\infty$.*

Proof. Rather than using the standard Morse theory, we make a simple proof of the loss of optimality after the first conjugate time based on the geometric argument of the Riemannian case [7]. Indeed, let t_{1c} be the first conjugate time along the reference trajectory γ. There exists a Jacobi field vertical at 0 and t_{1c} corresponding to a variation of γ with $\delta x(0) = \delta x(t_{1c}) = 0$. Then, for $t > t_{1c}$ we can construct a broken solution with the same time duration (see Fig. 1.5). But in our regular case, an optimal solution cannot be broken. In fact, by smoothing the corner we obtain a shortest path. Since the model approximates our system up to relevant terms of order two, we conclude that optimality is lost. ■

Proposition 1.14. *Consider a single input affine control system defined by the pair F_0, F_1. Under our assumptions, a reference trajectory is time minimal (*resp. *maximal) in the hyperbolic (*resp. *elliptic) case up to the first conjugate time with respect to all trajectories with the same extremities and contained in a tubular $\mathcal{C}^1$-neighbourhood of the reference extremal.*

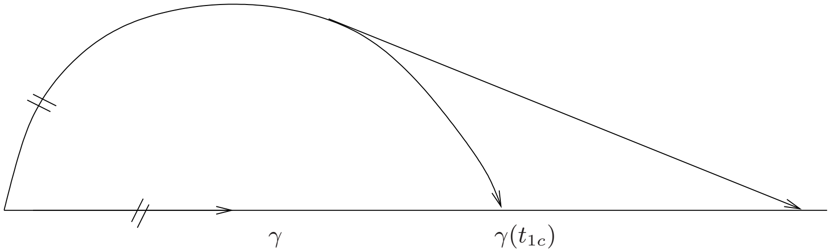

Fig. 1.5. Broken solution with same time duration and shortest path

The same optimality result holds for the restricted system, the set of controls being endowed with the L^∞-norm topology.

Computation and estimation of conjugate points. The simplest test deals with the nonlinear system $\dot{x} = f(x, u)$. We denote by $H_r(x, p)$ the smooth maximized Hamiltonian function and we restrict our equation to the level set $H_r = 1$ so as to break the symmetry due to the linearity with respect to p. This amounts to assume p in the projective space $\mathbf{P}^{n-1}$. We note $J_1, \ldots, J_{n-1}$ a basis of Jacobi fields which are vertical at 0. Let $L(t) = \mathrm{Span}\{J_1(t), \ldots, J_{n-1}(t)\}$ be the corresponding isotropic space. The numerical test for conjugate times is

$$\mathrm{rank}\ d\varPi(z(t))L(t) < n - 1 \tag{1.17}$$

which can easily be tested numerically. Moreover, since the reference trajectory is tranverse, the test is equivalent to

$$\det(d\varPi(z(t))L(t), f(x(t), u(t))) = 0. \tag{1.18}$$

In order to estimate the conjugate points we can use the curvature tensor according to the Sturm comparison theorem.

Lemma 1.8. *(Sturm) Let v be the solution of $\ddot{v} + A(t)v = 0$ with $v(0) = 0$, $\dot{v}(0) = 1$, and let w be the solution of $\ddot{w} + B(t)w = 0$ with the same initial conditions. Suppose $A(t) \leq B(t)$. If a (*resp. *b) is the first positive zero of v (*resp. *w), then $a \leq b$*

.

Proof. In accordance with initial conditions, v and w are positive for $0 < t < a$ and $0 < t < b$, respectively. Assume by contradiction that $a > b$. One has

$$0 = \int_0^b (v(\ddot{w} + Bw) - w(\ddot{v} + Av))dt \tag{1.19}$$

$$= [v\dot{w} - w\dot{v}]_0^b + \int_0^b (B - A)vwdt. \tag{1.20}$$

Since $A(t) \geq B(t)$ and since v and w are positive on $[0, b]$, then the integral in (1.20) is non-positive. Therefore, $[v\dot{w} - w\dot{v}]_0^b$ is nonnegative, that is $v(b)\dot{w}(b) \geq 0$. But $v(b) > 0$ and $\dot{w}(b) < 0$ (since w is not identically zero, w and $\dot{w}$ cannot be both zero at b), hence the contradiction. ■

Corollary 1.1. *Let $\delta\ddot{x} + K(t)\delta x = 0$ be the one-dimensional Jacobi equation in normal form. Assume $0 < K_1 \leq K(t) \leq K_2$. If t_{1c} is the first conjugate time, then t_{1c} belongs to $[\pi/\sqrt{K_2}, \pi/\sqrt{K_1}\,]$.*

In higher dimension, this result can be applied using the *sectional curvature*. In our problem, the Jacobi equation is identified with a differential operator and we can use a different normal form with less invariant terms than in the curvature because, while evaluating the intrinsic second-order derivative, we have reduced the terms by integrating by parts. The normal form is given by the proposition hereafter.

Proposition 1.15. *Any self-adjoint differential operator l with real coefficients is of even order and can be written according to*

$$l(y) = (p_0 y^{(q)})^{(q)} + (p_1 y^{(q-1)})^{(q-1)} \cdots + (p_q y).$$

Accordingly, l is defined by the $q+1$ functions of time $p_0, \ldots, p_q$.

We now strengthen our optimality results to get $\mathcal{C}^0$-sufficient optimality conditions.

Central field, Hamilton-Jacobi-Bellman equation and $\mathcal{C}^0$-sufficient optimality conditions. Let $(\bar{z}, \bar{u})$ be an extremal defined on $[0, T]$ of the time-optimal control problem of $\dot{x} = f(x, u)$, x in a manifold M, extremities being fixed. As previously, we make the following regularity assumptions:

(i) The strict Legendre-Clebsch conditions holds along the extremal,

$$\frac{\partial^2 H}{\partial u^2}(\bar{z}(t), \bar{u}(t)) < 0, \;\; t \in [0, T].$$

(ii) On each subinterval $0 < t_0 < t_1 \leq T$ the singularity is of codimension one.

(iii)We are in the normal case where $H = \langle p, f(x, u)\rangle$ is not zero and p can be chosen on the level set $\widetilde{H} = -1 + \langle p, f(x, u)\rangle = 0$.

In this case, our reference control is smooth and the reference extremal is solution of a system defined by a smooth Hamiltonian H_r. We denote by φ_t the one parameter local group

$$\varphi_t = \exp t\overrightarrow{H}_r$$

and by $L(t)$ the one parameter family of Lagrangian manifolds image of the fiber $T^*_{x_0}M$. Assume that the reference extremal curve is one-to-one and that there exists no conjugate point on $[0, T]$. Then we can imbed the reference solution into a *central field*, projection of the Lagrangian submanifolds on M.

This construction is valid in a neighbourhood of the reference curve, but it can be prolongated to a maximal open set W homeomorphic to a convex cone, each point of the domain being related to a unique point of $\Pi(L(t))$. Our aim is to prove that the reference extremal curve is optimal with respect to all trajectories with the prescribed extremities contained in this set W. The first result is the following [14].

Lemma 1.9. *(Verification lemma) Excluding $x(0)$, assume that there exists an open neighbourhood N of the reference trajectory and two smooth mappings $S : W \to \mathbf{R}$ and $\hat{u} : W \to U$ such that for each (x, u) in $W \times U$ we have the maximization condition*

$$\widetilde{H}(x, dS(x), \hat{u}(x)) \geq \widetilde{H}(x, dS(x), u)$$

and $\widetilde{H}(x, dS(x), \hat{u}(x)) = 0$. Then the reference trajectory is optimal among all the trajectories of the system with the same extremities and contained in W.

Proof . Let $0 < \bar{t}_0 \leq \bar{t}_1 \leq T$ and let (x, u) be a trajectory of the system defined on $[t_0, t_1]$, contained in W and satisfying the boundary conditions $x(t_0) = \bar{x}(\bar{t}_0)$, $x(t_1) = \bar{x}(\bar{t}_1)$. If we note $T(x, u)$ the transfer time $t_1 - t_0$, we must prove that $T(\bar{x}, \bar{u}) \leq T(x, u)$. By definition,

$$\begin{aligned} 1 &= \langle dS(x(t)), f(x(t), u(t))\rangle - \widetilde{H}(x(t), dS(x(t)), u(t)) \\ &= dS(x(t))\dot{x}(t) - \widetilde{H}(x(t), dS(x(t)), u(t)). \end{aligned}$$

Therefore we obtain

$$T(x, u) = \int_{t_0}^{t_1} dt = S(x(t_1)) - S(x(t_0)) - \int_{t_0}^{t_1} \widetilde{H}(x(t), dS(x(t)), u(t))dt.$$

Besides, since along the reference curve we have $\widetilde{H} = 0$, one has

$$T(\bar{x}, \bar{u}) = S(\bar{x}(\bar{t}_1)) - S(\bar{x}(\bar{t}_0)).$$

Because the extremities are fixed we get

$$T(x, u) - T(\bar{x}, \bar{u}) = -\int_{t_0}^{t_1} \widetilde{H}(x(t), dS(x(t)), u(t))dt$$

which is nonnegative by virtue of the maximization condition and $T(x, u) \geq T(\bar{x}, \bar{u})$. This proves the result. ■

The construction of S is equivalent to solve the standard *Hamilton-Jacobi-Bellman* equation

$$\max_{u\in U} H(x, \frac{\partial S}{\partial x}) = 1$$

the transfer times of the extremal curves being $T(x,u) = S(x(t_1)) - S(x(t_0))$ and the adjoint covector p being $\partial S/\partial x$.

Geometric construction and concluding result. Restated in symplectic formalism, the construction of the solution S of the Hamilton-Jacobi-Bellman equation is clear. The set $L = \{p = \partial S/\partial x\}$ is a Lagrangian manifold and the restriction of Π to L is a diffeomorphism. To construct L we use the central field and we take $L = (\cup_{t>0} L(t)) \cap \{\widetilde{H} = 0\}$. Indeed, for each t the manifold $L(t)$ is Lagrangian and homogeneous with respect to p which is normalized by the condition $p^0 = 1$, p^0 being the dual to the cost. The projection gives the set of points at time t from x_0. In particular, $L(t) \cap \{\widetilde{H} = 0\}$ is an isotropic manifold of dimension $n-1$. It is straightforward to prove that it defines a Lagrangian manifold when saturated by the flow.

Proposition 1.16. *Under our assumptions, the reference extremal curve is optimal with respect to all curves solution of the system with same extremities and contained in the domain covered by the central field.*

1.3.3 Examples

Sub-Riemannian Heisenberg Case

We consider a system in $\mathbf{R}^n$ of the form

$$\dot{x} = \sum_{i=1}^{m} u_i F_i(x).$$

Such systems are called symmetric and are relevant in robotics. In particular, they are connected to motion planning: given a path $\gamma \ : \ t \mapsto \gamma(t)$, t in $[0,1]$, find an admissible trajectory (x,u) of the system with same extremities. Moreover, in order to avoid obstacles, we fix around γ a tube W in which the admissible trajectories have to stay. For such problems we can also assign a cost, *e.g.* minimum time or minimum length. We shall consider here the latter case where the length is defined by a Riemannian metric. The restriction of this metric to a the distribution generated by $F_1, \ldots, F_m$ defines a sub-Riemannian problem. If we choose an orthonormal subframe in the distribution, the problem becomes $\dot{x} = \sum_{i=1}^{m} u_i F_i(x)$ with criterion

$$\int_0^T (\sum_{i=1}^{m} u_i^2)^{1/2} dt$$

and the size of the tube W can be set by the metric. We restrict the analysis to the standard Heisenberg contact case in $\mathbf{R}^3$.

Consider the following system in $\mathbf{R}^3$:

$$\dot{q} = u_1 F_1(q) + u_2 F_2(q)$$

where $F_1 = \partial/\partial x + y\partial/\partial z$, $F_2 = \partial/\partial y - x\partial/\partial z$. If we set $F_3 = \partial/\partial z$, we get $[F_1, F_2] = 2\partial/\partial z$. Moreover, all the Lie brackets of order three or more are zero. The distribution D spanned by F_1, F_2 is a *contact distribution* defined as the kernel of the 1-form $\alpha = dz + (xdy - ydx)$. A sub-Riemannian metric is associated to a metric of the form $g = a(q)dx^2 + 2b(q)dxdy + c(q)dy^2$. By choosing suitable coordinates, the smooth functions a, b and c can be normalized to $a = c = 1$ and $b = 0$. The case $g = dx^2 + dy^2$ is called the *Heisenberg case* or the *sub-Riemannian flat contact case.*

Heisenberg sub-Riemannian geometry and the Dido problem. We observe that the previous problem can be written $\dot{x} = u_1$, $\dot{y} = u_2$, $\dot{z} = \dot{x}y - \dot{y}x$, and

$$\int_0^T (\dot{x}^2 + \dot{y}^2)^{1/2} dt \to \min$$

in order that:

(i) The length of a curve $t \mapsto (x(t), y(t), z(t))$ is the length of the projection in the xy-plane.

(ii) If a curve joins (x_0, y_0, z_0) to (x_1, y_1, z_1), $z_1 - z_0$ is proportional to the area swept by the projection of the curve on the xy-plane.

Hence, our problem is dual to the *Dido problem* whose solutions are circles: find closed curves in the plane of prescribed length and maximum enclosed area.

Computation of the extremal curves. According to Maupertuis principle, minimizing the length is the same as minimizing the energy,

$$\int_0^T (u_1^2 + u_2^2) dt \to \min$$

the final time T being fixed. The Hamiltonian is

$$\widetilde{H}(x, p, u) = \sum_{i=1}^{2} (p^0 u_i^2 + u_i P_i)$$

with the Hamiltonian lifts or *Poincaré coordinates* $P_i = \langle p, F_i \rangle$, $i = 0, \ldots, 2$ and $F_0 = \partial/\partial z$. We are in the normal case and set $p^0 = -1/2$ so that, in these coordinates, the extremals are the solutions of:

$$\dot{x} = P_1, \ \dot{y} = P_2, \ \dot{z} = P_1 y - P_2 x$$

and

$$\dot{P}_1 = 2P_3P_2, \ \dot{P}_2 = -2P_3P_1, \ \dot{P}_3 = 0.$$

By setting $P_3 = \lambda/2$, we obtain the equation of a linear pendulum, $\ddot{P}_1 + \lambda^2 P_1 = 0$ and the equations are integrable by quadrature with trigonometric functions. The integration is straightforward if we observe that

$$\ddot{z} - \frac{\lambda}{2}\frac{d}{dt}(x^2 + y^2) = 0.$$

We get the following parameterization of the extremals starting from $z_0 = (0, 0, 0)$. If $\lambda = 0$, $x(t) = At\cos\varphi$, $y(t) = At\sin\varphi$, $z(t) = 0$, that is we have straight lines. If $\lambda \neq 0$, $x(t) = A/\lambda\ (\sin(\lambda t + \varphi) - \sin\varphi)$, $y(t) = A/\lambda\ (\cos(\lambda t + \varphi) - \cos\varphi)$, $z(t) = (A^2/\lambda)t - (A^2/\lambda^2)\sin\lambda t$, with $A = (P_1^2 + P_2^2)^{1/2}$ and φ is the argument of the vector $(\dot{x}, -\dot{y})$.

Conjugate points, global optimality. The computation of conjugate points by means of the previous parameterization is obvious: the extremal straight lines have no conjugate points, and the extremals which project onto circles in the xy-plane have their first conjugate points after one revolution. We note $S(a, r)$ the sphere of points at sub-Riemannian distance r of the center a. Note that the sub-Riemannian distance is continuous. It is isometric to the sphere of same radius centered at the origin, $S(0, r)$, which is a surface of revolution with respect to the z-axis and symmetric with respect to the xy-plane. Eventually, we can take a unit radius by homogeneity. Standard existence theorems tell us that the sphere is made of extremity points of minimizing extremals of unit length. As a consequence of our computations, a conjugate point is also a *cut point* where a minimizer ceases to be globally optimal. This is a degenerate situation similar to the Riemannian case on $\mathbf{S}^2$. Arguably, as the sphere is a surface of revolution with respect to the z-axis, there is a one parameter family of extremal curves intersecting exactly at the same point.

The Flat Torus

Another interesting example is the flat torus $\mathbf{T}^2$ obtained by identifying points on the opposite sides of the square $[0, 1] \times [0, 1]$. The extremals with respect to length are the images of straight lines in $\mathbf{R}^2$ through this identification. Since the problem is flat, the curvature is zero and the Jacobi equation is trivial. There is no conjugate point and optimality is related to the topology of the torus. Actually, if x_0 is chosen at the center of the square, then an extremal is optimal until it reaches the sides where it meets another minimizing curve (there are up to four such curves at a corner). To describe the underlying topology, we can choose coordinates l_1, l_2 which are angles. Given any extremal, there are infinitely many extremals with same extremities but different rotation numbers. This property is crucial to understand the orbit transfer problem.

1.4 Time-optimal Transfer Between Keplerian Orbits

An important matter in astronautics is to transfer a satellite between elliptic orbits. Optimal criterions related to this issue are the maximization of the final mass (which amounts to minimizing the fuel consumption) or the minimization of the transfer time. We shall consider this second problem for two reasons. First, recent research projects concern orbit transfer with electro-ionic propulsion for which the thrust is very low and the transfer duration very long (up to several months). Moreover, since the transfer is always with maximum thrust, the structure of the minimum time extremals is simpler than in the minimum consumption case [10] and fit in the smooth case previously analyzed.

1.4.1 Model and Basic Properties

Let m be the mass of the satellite, and let F be the thrust of the engine. The equation describing the system in Cartesian coordinates are:

$$\ddot{q} = -q\frac{\mu}{r^3} + \frac{F}{m}$$

where q is the position of the satellite measured in a fixed frame I, J, K whose origin is the Earth center, where $r = |q| = (q_1^2 + q_2^2 + q_3^2)^{1/2}$, and where μ is the gravitation constant. The free motion with $F = 0$ is *Kepler equation*. The thrust is bounded, $|F| \leq F_{\max}$, and the mass variation is described by

$$\dot{m} = -\frac{|F|}{v_e} \tag{1.21}$$

where v_e is a positive constant (see Table 1.1). Practically, the initial value m_0 of the mass is known, and m has to remain greater than the mass of the

Table 1.1. Physical constants

Variable	**Value**
μ	5165.8620912 Mm^3·h^{-2}
$1/v_e$	$1.42e-2$ Mm^{-1}·h
m_0	1500 kg
$F_{\max}$	3 N

satellite without fuel: $m \geq \chi_0$. If (1.21) is not taken into account, we have a simplified *constant mass model*. Roughly speaking, this model is sufficient for our geometric analysis, but the mass variation has to be included for numerical computation. Besides, if the thrust is maximal, maximizing the final mass reduces to minimizing the transfer time. If $q \wedge \dot{q}$ is not zero, the thrust can be decomposed in a moving frame attached to the satellite. A canonical choice consists in the *radial-orthoradial frame*: $F = u_r F_r + u_{\mathrm{or}} F_{\mathrm{or}} + u_c F_c$ with

$$F_r = \frac{q}{r}\frac{\partial}{\partial \dot{q}}, \ F_c = \frac{q \wedge \dot{q}}{|q \wedge \dot{q}|}\frac{\partial}{\partial \dot{q}}$$

and $F_{\mathrm{or}} = F_c \wedge F_r$. If $u_c = 0$, the state space is the tangent space to the osculating plane generated by q and $\dot{q}$ and we have a 2D-problem. For the sake of simplicity, we shall restrict our study to this setting which already exhibits all the relevant features of the whole system. We first recall the classical properties of the Kepler equation.

Proposition 1.17. *The two vectors below are first integrals of the Kepler equation:*

$$\begin{aligned} c &= q \wedge \dot{q} \ \textit{(angular momentum)} \\ L &= -q\frac{\mu}{r} + \dot{q} \wedge c \ \textit{(Laplace vector).} \end{aligned}$$

Moreover, the energy $H(q, \dot{q}) = 1/2\ \dot{q}^2 - \mu/r$ *is preserved and the following relations hold:*

$$L.c = 0, \ L^2 = \mu^2 + 2Hc^2.$$

Proposition 1.18. *If* $c = 0$*, the motion is on a colliding line. Otherwise, if* $L = 0$ *the motion is circular while if* $L \neq 0$ *and* $H < 0$ *the trajectory is an ellipse:*

$$r = \frac{c^2}{\mu + |L| \cos(\theta - \theta_0)}$$

where θ_0 *is the argument of the pericenter.*

Definition 1.11. *The domain* $\Sigma_e = \{H < 0, \ c \neq 0\}$ *is filled by elliptic orbits and is called the* elliptic (2D-elliptic in the planar case) domain.

Geometric coordinates. To each pair (c, L) corresponds a unique oriented ellipse. Using these coordinates, we have a natural representation of the system. Namely,

$$\begin{aligned} \dot{c} &= q \wedge \frac{F}{m} \\ \dot{L} &= F \wedge c + \dot{q} \wedge (q \wedge \frac{F}{m})\cdot \end{aligned}$$

Since Σ_e is a fiber bundle, one needs a coordinate to describe the evolution on the fiber itself. The fiber is $\mathbf{S}^1$ and the coordinate is the so-called *cumulated longitude*. A second representation is thus provided by the *orbit elements*. Restricting to the 2D case, one has $x = (P, e, l)$ with

- P, *semi-latus rectum* related to the semi-major axis by $P = a(1 - e^2)$.
- $e = (e_x, e_y)$, *eccentricity vector* oriented along L, that is along the semi-major axis, and whose norm is the eccentricity of the ellipse.
- l, *cumulated longitude* measured with respect to the q_1-axis.

Using the radial-orthoradial decomposition in which the dynamics is $\dot{x} = F_0 + u_r F_r + u_{\mathrm{or}} F_{\mathrm{or}}$, the vector fields are

$$F_0 = \sqrt{\frac{\mu}{P}} \frac{W^2}{P} \frac{\partial}{\partial l}$$

$$F_r = \sqrt{\frac{P}{\mu}} \left(+ \sin l \frac{\partial}{\partial e_x} - \cos l \frac{\partial}{\partial e_y} \right)$$

$$F_{\mathrm{or}} = \sqrt{\frac{P}{\mu}} \left(\frac{2P}{W} \frac{\partial}{\partial P} + (\cos l + \frac{e_x + \cos l}{W}) \frac{\partial}{\partial e_x} + (\sin l + \frac{e_y + \sin l}{W}) \frac{\partial}{\partial e_y} \right) .$$

According to the data provided by the French Space Agency (CNES), the problem is to transfer the system from a low eccentric initial ellipse towards the geostationnary orbit. The boundary conditions are given in Table 1.2. Though the longitude is free on the initial and terminal orbits, we set $l(0) = \pi$ for numerical issues[5].

Table 1.2. Boundary conditions

Variable	Initial cond.	Final cond.
P	11.625 Mm	42.165 Mm
e_x	0.75	0
e_y	0	0
h_x	0.0612	0
h_y	0	0
l	π rad	103 rad

1.4.2 Maximum Principle and Extremal Solutions

Definition 1.12. *We call SR-problem with drift the time-optimal problem for a system of the form*

[5] This amounts to start at the apocenter where the attraction is the weakest. The numerical integration is thus improved.

$$\dot{x} = F_0 + \sum_{i=1}^{m} u_i F_i$$

with $x \in \mathbf{R}^n$, $F_0, \ldots, F_m$ smooth vector fields, the control u in $\mathbf{R}^m$ being bounded by $\sum_{i=1}^m |u_i|^2 \leq 1$.

Let the H_i's be the usual Hamiltonian lifts $\langle p, F_i \rangle$, $i = 0, \ldots, m$, and let Σ be the switching surface $\{H_i = 0,\ i = 1, \ldots, m\}$. The maximization of the Hamiltonian $H = H_0 + \sum_{i=1}^m u_i H_i$ outside Σ implies that

$$u_i = \frac{H_i}{\sqrt{\sum_{i=1}^m H_i^2}},\ i = 1, \ldots, m. \tag{1.22}$$

Plugging (1.22) into H, one defines the Hamiltonian function

$$H_r = H_0 + \left(\sum_{i=1}^m H_i^2 \right)^{\frac{1}{2}}. \tag{1.23}$$

Definition 1.13. *The corresponding solutions are called* order zero extremals. *From the maximum principle, optimal extremals are contained in the level set $\{H_r \geq 0\}$. Those in $\{H_r = 0\}$ are* exceptional.

The following result is standard.

Proposition 1.19. *The order zero extremals are smooth responses to smooth controls on the boundary of $|u| \leq 1$. They are singularities of the endpoint mapping $E_{x_0,T} : u \mapsto x(T, x_0, u)$ for the L^∞-topology when u is restricted to the unit sphere $\mathbf{S}^{m-1}$.*

In order to construct all extremals, we must analyze the behaviour of those of order zero near the switching surface. On one hand, observe that we can connect two such arcs at a point located on Σ if we respect the Weierstraß-Erdmann conditions

$$p(t+) = p(t-),\ H_r(t+) = H_r(t-)$$

where t is the time of contact with the switching surface. Those conditions, obtained in classical calculus by means of specific variations, are contained in the maximum principle. On the other hand, singular extremals satisfy $H_i = 0$, $i = 1, \ldots, m$, and are contained in Σ. They are singularities of the endpoint mapping if u is interior to the control domain, $|u| < 1$. For the 2D orbit transfer, we restrict ourselves to the constant mass model. Since Lie brackets can easily be computed in Cartesian coordinates, the system is written $\dot{x} = F_0(x) + u_1 F_1(x) + u_2 F_2(x)$ where $x = (q, \dot{q})$ and where F_0 is the Kepler vector field with $F_1 = \partial/\partial_{\dot{q}_1}$, $F_2 = \partial/\partial_{\dot{q}_2}$. The following result is straightforward.

Lemma 1.10. *Let $\mathcal{D}$ be the controlled distribution generated by F_1 and F_2. Then $[\mathcal{D},\mathcal{D}]=0$ and at each point the rank of the span of F_1, F_2, $[F_0,F_1]$ and $[F_0,F_2]$ is four.*

This allows to make a complete classification of the extremals. Differentiating along an extremal curve, one gets

$$\dot{H}_i=\{H_i,H_0\},\ i=1,2.$$

At a switching point, $H_1=H_2=0$ but, since F_1, F_2, $[F_0,F_1]$ and $[F_0,F_2]$ form a frame, both Poisson brackets $\{H_1,H_0\}$ and $\{H_2,H_0\}$ cannot be vanish (this is the so-called *order one* case [2]). In order to understand the behaviour of extremals in the neighbourhood of such a point, we make a polar blowing up

$$H_1=\rho\cos\varphi,\ H_2=\rho\sin\varphi$$

and we get

$$\begin{aligned}\dot{\rho}&=\cos\varphi\{H_1,H_0\}+\sin\varphi\{H_2,H_0\}\\ \dot{\varphi}&=1/\rho\ \left(-\sin\varphi\{H_1,H_0\}+\cos\varphi\{H_2,H_0\}\right).\end{aligned}$$

Extremals curves crossing Σ are obtained by solving $\dot{\varphi}=0$ and the following holds.

Proposition 1.20. *There are extremals curves made of two order zero extremals concatenated and crossing Σ with a given slope. The corresponding control rotates instantaneously of an angle π at the contact with the switching surface. The resulting singularity of the extremal curve is called a Π-*singularity.

Since these singularities must be isolated, the only extremal curves for the orbit transfer are order zero extremal or finite concatenation of such arcs with Π-singularities at the junctions. Hence, numerically, we only compute order zero extremals since Π-singularities where the switching surface is crossed with a given slope are properly handled by an integrator with adaptive step length.

1.4.3 Numerical Resolution

Shooting Function

The boundary value problem defined by the maximum principle is solved by shooting. The shooting function is defined by (1.10). Namely, when the initial and final states are fixed, $x(0)=x_0$ and $x(T)=x_1$, one has

$$S \ : \ (T, p_0) \in \mathbf{R}_+^* \times \mathbf{P}^{n-1} \mapsto \exp_{x_0,T}(p_0) - x_1 \tag{1.24}$$

where the exponential mapping at time T is as before defined on $\mathbf{P}^{n-1}$ by $\exp_{x_0,T}(p_0) = \Pi(z(T, x_0, p_0))$. In the standard transfer case, the final longitude is free and the equation involving l_f has to be replaced by the transversality condition $p_l(T) = 0$, where p_l is the adjoint state to l. Practically, the n-dimensional nonlinear equation (1.24) on $\mathbf{R}\times\mathbf{P}^{n-1}$ is treated as an equation on $\mathbf{R}^{n+1}$, the initial adjoint covector p_0 being taken in $\mathbf{R}^n$ and normalized according to $p_0 \in \mathbf{S}^{n-1}$. Alternatively, in the normal case one can also prescribe the Hamiltonian level set, *e.g.* $H_r = 1$. Since equation (1.24) is to be solved by a Newton-like method, an important issue is the regularity of the shooting mapping. Here, the extremals are smooth outside Π-singularities so the analysis consists in studying the shooting mapping in the neighbourhood of such points. To this end, we use the *nilpotent model* of the 2D problem obtained in [2]. Here, the dimension is four and a nilpotent approximation with brackets of length greater than three all vanishing is

$$\begin{aligned} \dot{x}_1 &= 1 + x_3 & \dot{x}_2 &= x_2 \\ \dot{x}_3 &= u_1 & \dot{x}_4 &= u_2. \end{aligned}$$

The coupling of the system arises from the constraint on the control, $u_1^2+u_2^2 \leq 1$. Clearly enough, the extremals of such a system are given by

$$\dot{x}_3 = \frac{at+b}{\sqrt{(at+b)^2+(ct+d)^2}}$$

$$\dot{x}_4 = \frac{ct+d}{\sqrt{(at+b)^2+(ct+d)^2}}$$

where $a = -p_1(0)$, $b = p_3(0)$, $c = -p_2(0)$ and $d = p_4(0)$. The *switching function* is $t \mapsto (at + b, ct + d)$, and the set Σ of initial covectors generating switch points is stratified as follows. If a and c are both nonzero, the existence of a time such that the two components of the switching function vanish simultaneously reduces to the condition $ad - bc = 0$, so the first strata is the quadric $\Sigma_1 = \{p_1 \neq 0,\ p_2 \neq 0,\ p_1p_4 - p_2p_3 = 0\}$. If a and b are zero while c is not, there is no condition on d and, by symmetry, we also get the two disjoint unions of half planes $\Sigma_2^1 = \{p_1 = p_2 = 0,\ p_3 \neq 0\}$ and $\Sigma_2^2 = \{p_3 = p_4 = 0,\ p_1 \neq 0\}$. Eventually, note that a, b, c and d all zero is impossible since no singular control is allowed here (see Section 1.4.2). As a result, Σ is partitionned into a set of codimension one, and two sets of codimension two:

$$\Sigma = \Sigma_1 \cup (\Sigma_2^1 \cup \Sigma_2^2).$$

The fact that these subsets are of codimension greater or equal to one implies that, despite the existence of the singularities illustrated below, the numerical computation is essentially reduced to the smooth case and thus tractable. Regarding Σ_1, let a and c be nonzero reals. Let then $\delta = (b/a - d/c)/2$ be the half distance between the roots of each component of the switching function. By symmetry, we can assume that a and c are equal and positive. Up to a translation of time, integrating the nilpotent model amounts to integrate

$$\dot{\xi} = \frac{t-\delta}{\sqrt{t^2+\delta^2}}$$

as well as the symmetric term ($t+\delta$ at the numerator). For nonzero δ, one gets

$$\xi_t(\delta) = \sqrt{t^2+\delta^2} - |\delta|(\operatorname{arc\,sh}\frac{t}{\delta}+1) + \text{constant}$$

where t is a fixed positive time. We have a singularity of the kind $\delta \log|\delta|$ and the shooting mapping is continuous but not differentiable at $\delta = 0$, that is when Σ_1 is crossed. Finally, as for Σ_2^2, let c and d be zero (the case of Σ_2^1 being treated analogously), and let $\sigma = -b/a$ be the unique root of the first component of the switching function when a is not zero. Assume for instance that a is positive so that the exponential is computed by integrating

$$\dot{\xi} = \frac{t-\sigma}{|t-\sigma|}.$$

Hence, for a positive fixed t, up to a constant

$$\begin{aligned} \xi_t(\delta) &= t && \text{if } \sigma < 0 \\ &= t - 2\sigma && \text{if } 0 \le \sigma \le t \\ &= -t && \text{if } \sigma > t. \end{aligned}$$

Accordingly, the function is not differentiable at $\sigma = 0$ and $\sigma = t$, and with zero derivative outside $[0, t]$. In other words, for $p_1 > 0$, the shooting mapping is singular outside the cone $\{-p_1 t \le p_3 \le 0\}$ included in Σ_2^2, and not differentiable on its boundary.

Homotopy on the Maximal Thrust

Beyond regularity issues, a delicate task is to provide the Newton method with nice initial guesses for T and p_0. Since the convergence is only local, no matter how smooth the function may be, a relevant approach is *homotopy*. Indeed, our transfer problem is naturally embedded in a family of such problems parameterized by the maximum thrust $F_{\max}$. Moreover, one can expect that given two such thrusts $F^0_{\max}$ and $F^1_{\max}$ close enough, the associated solutions T^i, p^i_0, $i = 0, 1$ also be close. This is the basic idea of homotopy which connects the simple problem with $F^0_{\max}$ big (the bigger the thrust, the shorter the transfer time and the easier the control problem) to the more intricate one with $F^1_{\max}$ smaller. One may for instance consider the sequence of intermediate problems generated by the convex homotopy $F_{\max} = (1-\lambda)F^0_{\max} + \lambda F^1_{\max}$, where λ in $[0, 1]$ is the homotopy parameter. Then *discrete homotopy* consists in picking a finite sequence $\lambda_0 = 0 < \cdots < \lambda_k < \cdots < \lambda_N = 1$ and trying to follow the associated path of zeros: the solution at step k is supposed to be an initial guess precise enough to ensure convergence of the solver at step $k+1$. More subtile alternatives where the step on the homotopic parameter is automatically adjusted are *simplicial* or *differential homotopy* [10]. In the case of discrete homotopy—often refered to as *discrete continuation*—, the minimal regularity required to ensure the process is relevant is provided by the following proposition [5].

Proposition 1.21. *The value function $F_{\max} \mapsto T(F_{\max})$ mapping to each positive maximum thrust the corresponding minimum time is right continuous for the transfer problem (2D or 3D, constant mass or not).*

As a matter of fact, we will use a decreasing sequence of thrusts bounds $(F^k_{\max})_k$. Therefore, right continuity of the value function is enough to guarantee that $T(F^k_{\max})$ tends to $T(F_{\max})$ when the thrusts decrease to $F_{\max}$. But while mere discrete homotopy is used to initialize the search for p_0, a much more precise guess for the minimum time is available. Clearly, the value function $T(F_{\max})$ is decreasing, and one can easily prove that the product $T(F_{\max}) \cdot F_{\max}$ is bounded below. Actually, it is also bounded over (see [6]) and the conjecture is that it has a limit when $F_{\max}$ tends to 0. In practice, we use the heuristic $T(F_{\max}) \cdot F_{\max} \simeq$ constant. Figure 1.6 presents the result of such a computation for a medium thrust of 3 Newtons.

Conjugate Points

In order to deal with conjugate and not focal points, we restrict ourselves to the transfer with fixed final longitude, l^f. As a result, the initial and final state are prescribed, and we are in the situation of Section 1.3.2. The shooting mapping is exactly defined by (1.24) and an extremal is easily computed by solving the shooting problem for a fixed final longitude close to the one obtained with l_f free. So as to integrate the Jacobi equation

$$\delta \dot{z} = d\overrightarrow{H}_r(z(t)) \cdot \delta z$$

along the resulting extremal, $z(t) = z(t, t_0, x_0, \bar{p}_0)$ where $\bar{p}_0$ is a root of the shooting mapping, the standard procedure is to consider the augmented system

$$z = \overrightarrow{H}_r(z),\ \delta \dot{z} = d\overrightarrow{H}_r(z) \cdot \delta z$$

with initial condition $(x_0, \bar{p}_0)$ on z. As for $\delta z = \delta x, \delta p)$, the initial Jacobi fields must span the $(n-1)$ dimensional tangent space to $\{x_0\} \times \mathbf{S}^{n-1}$ so that

$$\delta x_i(0) = 0,\ \delta p_i(0) \perp \bar{p}(0),\ i = 1, \ldots, n-1.$$

According to (1.18), conjugate times are roots of

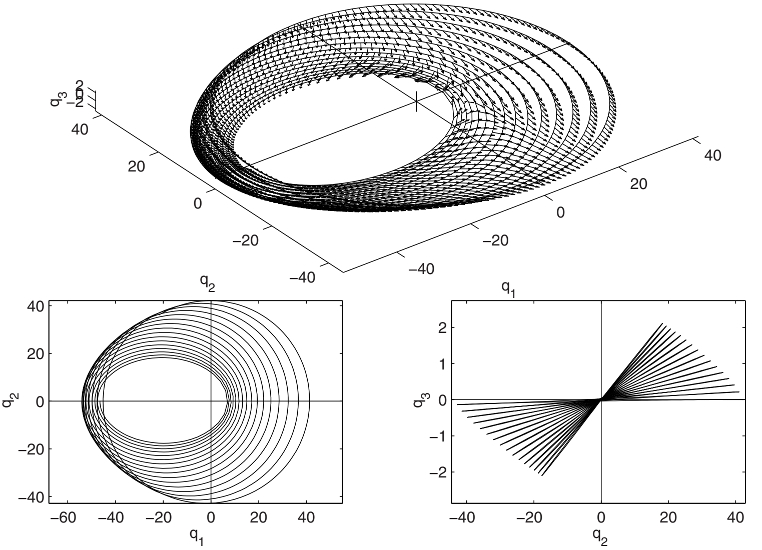

Fig. 1.6. Three dimensional transfer for 3 Newtons. The arrows indicate the action of the thrust. The main picture is 3D, the other two are projections. The duration is about twelve days.

$$\det(\delta_1 x(t), \ldots, \delta_{n-1} x(t), \dot{x}(t)) = 0.$$

This test is important since, numerically, a generic root can be detected by a change in signs. Hence, a rough estimate of conjugate times is easily obtained by dichotomy and then made more accurate, *e.g.*, by a few Newton steps. Nevertheless, it is still compulsary to refine the computation to take into account cases when the rank of the exponential mapping becomes strictly less than $n-2$ (see (1.17)). To this end, we also perform a singular value decomposition so as to check the rank of the span of $\delta x_1, \ldots, \delta x_{n-1}$ along the extremal. The counterpart is that, since the singular values are nonnegative, the detection of zeros is more intricate. An example of this kind of computation in the orbit transfer case is shown at Fig. 1.7.

1.5 Introduction to Optimal Control with State Constraints

In many applied control problems, the systems has state constraints. For instance, in the shuttle re-entry case, there is a constraint on the thermic flow in order to avoid the destruction of the ship. Whereas standard existence theorems hold, the necessary conditions become intricate. Indeed, general extremal

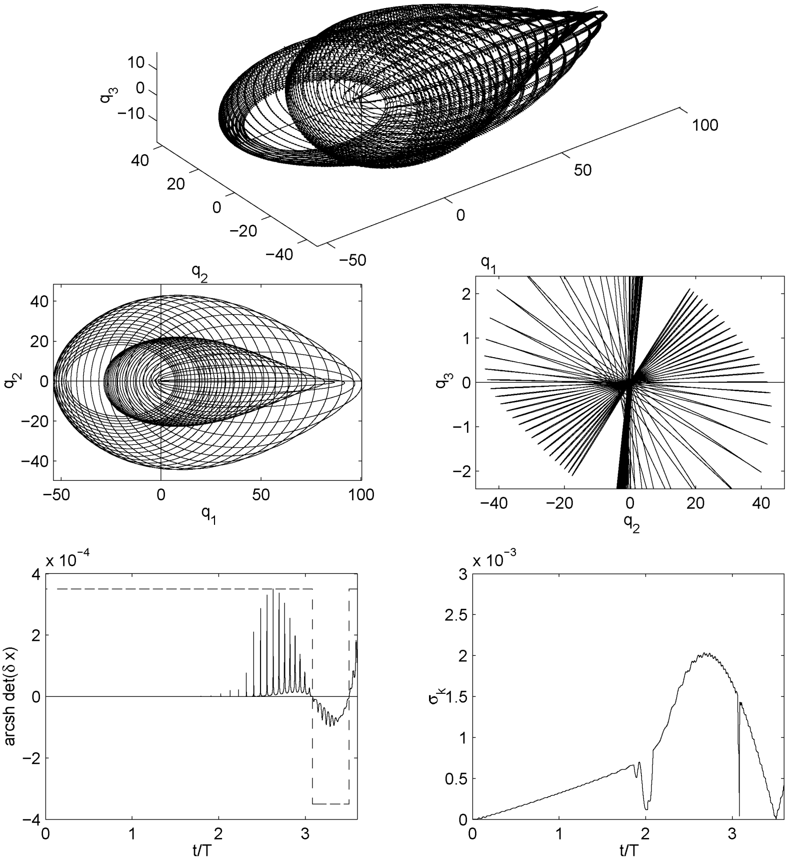

Fig. 1.7. An extremal, which is roughly the same as in Fig. 1.6 (the difference being the fixed final longitude), is extended until 3.5 times the minimum time. Bottom left, the determinant, bottom right, the smallest singular value of the Jacobi fields associated to the extremal. There, two conjugate times are detected. The optimality is lost about three times the minimum time.

curves are parameterized by measures supported by the boundary of the domain. Hence, the key point is to make a geometric analysis of the accessibility set near the constraints of order one to derive these conditions in a geometric form. The ultimate goal is to glue together extremals of the non-constrained problem with those on the boundary to provide an optimal synthesis. This kind of analysis comes from the pioneering work of Weierstraß in 1870 who solved the problem of minimizing the length of a planar curve in the presence of obstacles. The resulting necessary conditions can also be used to analyze *hybrid systems* defined by two subsystems $\dot{x} = f_1(x, u)$, $\dot{x} = f_2(x, u)$, each subsystem describing the evolution in two domains separated by a surface. In this case, Descartes-like refraction rules obtain as consequences of the maximum principle with state constraints. In the orbit transfer case, this approach allows us to take into account the eclipse phenomenon associated to electro-ionic propulsion.

1.5.1 The Geometric Framework

We consider a system of the form $\dot{x} = f(x, u)$, x in $\mathbf{R}^n$ and u in U, subset of $\mathbf{R}^m$, with a cost

$$c(x, u) = \int_0^T f^0(x, u) dt$$

in the presence of one state constraint of the form $g(x) \leq 0$, $g \; : \; \mathbf{R}^n \to \mathbf{R}$. We denote by $\widetilde{x} = (x^0, x)$ the extended state, the extended dynamics by $\widetilde{f}$, and by $\widetilde{g} = (0, g)$ the extended state constraint. In order to make a geometric analysis, we restrict our study to piecewise smooth pairs $(\widetilde{x}, \widetilde{u})$ defined on $[0, T]$. An optimal solution $\widetilde{x}$ is thus made of extremal subarcs contained in the open domain $\{g < 0\}$ where the constraint is not active and where the standard maximum principle holds, and of subarcs contained in the boundary, namely *boundary arcs*. In order to decide upon optimality, we split the problem in two.

Optimality of boundary arcs. Clearly, a boundary arc has to be optimal with respect not only to all trajectories contained in the boundary, but also to all neighbouring arcs in the open domain $\{g < 0\}$. This is illustrated by Fig. 1.8 where the hypersurface is a sphere and where the two kinds of variations are represented.

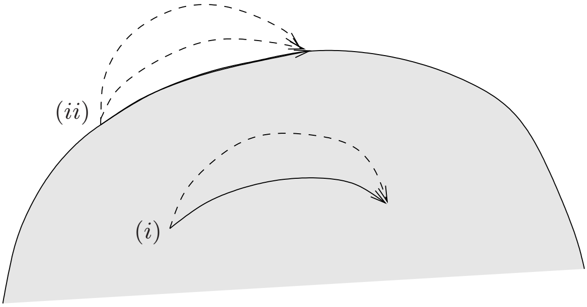

Fig. 1.8. Boundary curve (i) and neighbouring curve outside the constraint (ii)

Optimality conditions at junctions or reflections with the boundary. In this case, the matter is to glue together extremal curves of the non-constrained problem. The junction and reflection conditions were derived by Weierstraß by applying the variations of Fig. 1.9.

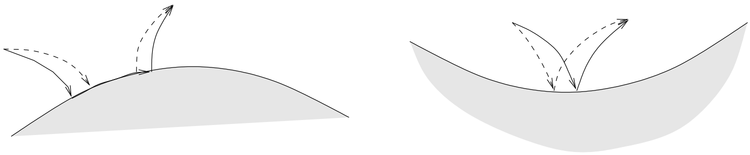

Fig. 1.9. Weierstraß variations

The two previous drawings are the keys of the geometric analysis which is organized as follows. We give first the necessary conditions of [18], illustrating them by several examples. Since the proof is technical, we present next the original proof of Weierstraß using standard calculus of variations.

1.5.2 Necessary Optimality Conditions for Boundary Arcs

Statement

For the sake of simplicity, we shall assume the control domain to be a smooth manifold with boundary defined by $q(u) \leq 0$. Differentiating the constraint along the solution, we get the Lie derivative with respect to the dynamics:

$$\begin{aligned} h(x,u) &= L_{f(x,u)}g \\ &= d/dt\,(g(x)) \\ &= (\nabla g(x)|f(x,u)). \end{aligned}$$

The crucial concept is given by the following definition.

Definition 1.14. *The pair* (x,u) *is of* order one *if*

(i) $h(x,u) = 0$

(ii)$\partial h(x,u)/\partial u \neq 0$

which corresponds to a contact of minimal order *with the boundary.*

In this case, we can define locally a system in the boundary by choosing controls such that $h(x,u)$ is zero. In order to ensure the existence of variations, we further impose the following regularity condition: if u belongs to the boundary of the control domain, $\partial h/\partial u(x,u)$ and $dq(u)/du$ are linearly independent,

$$\frac{\partial h}{\partial u} \wedge \frac{dq}{du} \neq 0.$$

Let $\widetilde{H}(\widetilde{x},\widetilde{p},u)=p^0f^0(x,u)+\langle p,f(x,u)\rangle$ be the Hamiltonian of the problem. If we maximize $\widetilde{H}$ over the set $U(x)$ defined by $h(x,u)=0$, $q(u)\leq 0$, then there are Lagrange multipliers λ and ν such that

$$\frac{\partial \widetilde{H}}{\partial u}=\lambda\frac{\partial h}{\partial u}+\nu\frac{dq}{du}\cdot \tag{1.25}$$

We can now formulate the necessary optimality conditions for boundary arcs.

Theorem 1.2. *Let (x,u) be a smooth optimal solution defined on $[0,T]$ of the problem with fixed extremities. Then, there is a continuous adjoint covector (p^0,p), nonzero, and a scalar function λ such that the following conditions are satisfied:*

$$\dot{x}=\frac{\partial \widetilde{H}}{\partial p}(\widetilde{x},\widetilde{p},u),\ \dot{p}=-\frac{\partial \widetilde{H}}{\partial x}(\widetilde{x},\widetilde{p},u)+\lambda\frac{\partial h}{\partial x}(x,u)$$
$$\widetilde{H}(\widetilde{x},\widetilde{p},u)=\max\nolimits_{v\in U(x)}\widetilde{H}(\widetilde{x},\widetilde{p},v)=0$$

where, for each t, $\lambda(t)$ is a Lagrange multiplier defined by (1.25). Moreover, p^0 is non-positive, $p(0)$ can be chosen tangent to $\{g=0\}$ and, at each derivability point of λ, the vector $\dot{\lambda}(t)\nabla g(x(t))$ is zero or pointing towards the interior of the domain.

Application in Riemannian Geometry

We consider a smooth hypersurface M defined by the equation $g(x)=0$ and imbedded in the Euclidean space $\mathbf{R}^n$. The manifold M is thus Riemannian for the induced metric. Outside M, the curves of minimum length are straight lines and we can recover the geometric properties defining the extremals from Theorem 1.2. Clearly, they have to be extremal curves for the induced Riemannian metric. Besides, the convexity properties of the surface are important as illustrated by Fig. 1.10.

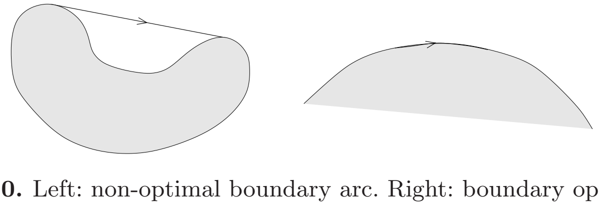

Fig. 1.10. Left: non-optimal boundary arc. Right: boundary optimal arc

The problem can be formulated as the time-optimal control problem of the system $\dot{x}=u$, x in $\mathbf{R}^n$ and u in $\mathbf{R}^n$, where the control domain is the unit sphere $\sum_{i=1}^{n} u_i^2 = 1$. One has $h(x,u) = (\nabla g(x)|u)$ and we consider an arc $t \mapsto x(t)$ such that $g(x) = 0$ and $h(x,u) = 0$. The Hamiltonian is $\widetilde{H} = p^0 + \langle p, u\rangle$ and the adjoint system satisfies

$$\dot{p} = \lambda \frac{\partial h}{\partial x} = \lambda \frac{d}{dt} \nabla g(x) \tag{1.26}$$

The cost multiplier p^0 is not zero and can be normalized to $p^0 = -1$. Hence we get $\langle p, u\rangle = 1$. Moreover, $\partial \widetilde{H}/\partial u = p = \lambda \partial h(x,u)/\partial u + 2\nu dq(u)/du$, so

$$p = \lambda \nabla g(x) + 2\nu u$$

and, multiplying by $u = \dot{x}$, we obtain

$$1 = \langle p, u\rangle = \lambda \langle \nabla g(x), u\rangle + 2\nu |u|^2.$$

Therefore, $\nu = 1/2$ and, using relation (1.26), we get:

$$\dot{\lambda} \nabla g(x) + \dot{u} = 0.$$

This relation tells us that the acceleration $\ddot{x} = \dot{u}$ is perpendicular to the tangent space of M: this is the standard characterization of the geodesic curves on the surface. Moreover,

$$\ddot{x} = -\dot{\lambda} \nabla g(x)$$

and $\ddot{x}$ is pointing outwards which is the convexity relation. Hence, we have a complete description of optimal curves on the surface thanks to Theorem 1.2. In the next paragraph, we present the junction and reflection conditions so as to provide an exhaustive portrait of optimal solutions.

1.5.3 Junction and Reflection Conditions

Statement

We shall consider an arc x defined on $[0, T]$ and meeting the boundary of the domain at a unique time $0 < \tau < T$.

Definition 1.15. *The point $x(\tau)$ is called a* junction point *if $x(t)$ is contained in the boundary for $t \geq \tau$, and a* reflection point *if the arc is contained in the interior of the domain when $t \neq \tau$.*

Let $\widetilde{p} = (p^0, p)$ be the adjoint covector associated to an optimal solution. For a junction point, the *jump condition* is

$$p(\tau+) = p(\tau-) + \mu \nabla g(x(\tau)).$$

For a reflection point, the condition is

$$p(\tau+) = p(\tau-) + \mu \nabla g(x(\tau)), \ \mu \geq 0.$$

Geometric Consequence

Consider the junction condition. Since for boundary arcs we can replace $p(\tau+)$ by $p(\tau+) + \nu \nabla g(x(\tau))$ by virtue of the tangency property to $\{g = 0\}$ of Theorem 1.2, at a junction point p can be normalized according to

$$p(\tau+) = p(\tau-).$$

Lemma 1.11. *At a junction point, the adjoint vector can be chosen continuous.*

Furthermore, using the junction condition one has

$$\langle \widetilde{p}(\tau), \widetilde{f}(\widetilde{x}(\tau), u(\tau-)) \rangle = \langle \widetilde{p}(\tau), \widetilde{f}(\widetilde{x}(\tau), u(\tau+)) \rangle = \max_{v \in U(x)} \widetilde{H}(\widetilde{x(\tau)}, \widetilde{p}(\tau), v)$$

where the maximized Hamiltonian is zero by virtue of Theorem 1.2. As a result, if the control is deduced from the maximization of the Hamiltonian is unique, it has to remain continuous when connecting the trajectory to the boundary. This is the case in the Riemannian problem.

Lemma 1.12. *In the Riemannian case, the straight lines connecting the boundary arcs are tangent to the surface at the junction points.*

1.5.4 Proof of the Necessary Conditions in the Riemannian Case

We consider the problem of minimizing in the plane

$$\int_{t_0}^{t_1} F(x, y, \dot{x}, \dot{y}) dt$$

with (x, y) is in $\mathbf{R}^2$ and where F defines a Riemannian metric. In particular, F satisfies the homogeneity relation

$$F(x, y, k\dot{x}, k\dot{y}) = kF(x, y, \dot{x}, \dot{y}), \ k > 0. \tag{1.27}$$

Though homogeneity can be relaxed by imposing a parameterization $\dot{x}^2 + \dot{y}^2 = 1$, we shall keep the problem in its general form. This will result in additional properties of the extremals. Now, if ξ and η are variations of the reference curve on the same interval $[t_0, t_1]$, the length variation is:

$$\begin{aligned} \delta l &= \int_{t_0}^{t_1} (F_x \xi + F_y \eta) + (F_{\dot{x}} \dot{\xi} + F_{\dot{y}} \dot{\eta}) dt \\ &\simeq l(x + \xi, y + \eta) - l(x, y) \end{aligned}$$

so that, integrating by parts and using zero boundary conditions $\xi(t_0) = \xi(t_1) = 0$, $\eta(t_0) = \eta(t_1) = 0$,

$$F_x - \frac{d}{dt}F_{\dot{x}} = 0, \; F_y - \frac{d}{dt}F_{\dot{y}} = 0.$$

These are the *Euler-Lagrange* equations, not independent because of homogeneity. Indeed, differentiating (1.27) with respect to k at $k = 1$ one obtains

$$\dot{x}F_{\dot{x}} + \dot{y}F_{\dot{y}} = F(x, y, \dot{x}, \dot{y}).$$

Differentiating with respect to (x, y),

$$\begin{aligned} F_x &= \dot{x}F_{\dot{x}x} + \dot{y}F_{\dot{y}x} \\ F_y &= \dot{x}F_{\dot{x}y} + \dot{y}F_{\dot{y}y} \end{aligned}$$

then with respect to $\dot{x}$, we get

$$\begin{aligned} \dot{x}F_{\dot{x}\dot{x}} + \dot{y}F_{\dot{y}\dot{x}} &= 0 \\ \dot{x}F_{\dot{x}\dot{y}} + \dot{y}F_{\dot{y}\dot{y}} &= 0 \end{aligned}$$

and the problem is not regular because the Hessian matrix of the Legendre-Clebsch condition is not invertible. For $(\dot{x}, \dot{y}) \neq (0, 0)$, there is a function F_1 defined by

$$F_{\dot{y}\dot{y}} = \dot{x}^2 F_1, \; F_{\dot{x}\dot{x}} = \dot{y}^2 F_1$$

and

$$F_{\dot{x}\dot{y}} = -\dot{x}\dot{y}F_1.$$

If we introduce

$$T = (F_{x\dot{y}} - F_{y\dot{x}}) + F_1(\dot{x}\ddot{y} - \ddot{x}\dot{y})$$

we get

$$F_x - \frac{d}{dt}F_{\dot{x}} = \dot{y}T, \; F_y - \frac{d}{dt}F_{\dot{y}} = -\dot{x}T$$

and Euler equation is equivalent to the *Weierstraß* equation, $T = 0$, for $(\dot{x}, \dot{y}) \neq (0, 0)$.

Application: Necessary Boundary Optimality Conditions

The previous formulæ for the first variation will be used to derive the necessary boundary conditions. Assume the boundary is one dimensional, and let $(\widetilde{x}, \widetilde{y})$ be a boundary arc on $[t_0, t_1]$. We introduce the variations represented by Fig. 1.11.

At each point of the boundary, we construct a vector n with length u, orthogonal to the boundary and oriented towards the interior of the domain. Namely, $n = (\xi, \eta)$ with

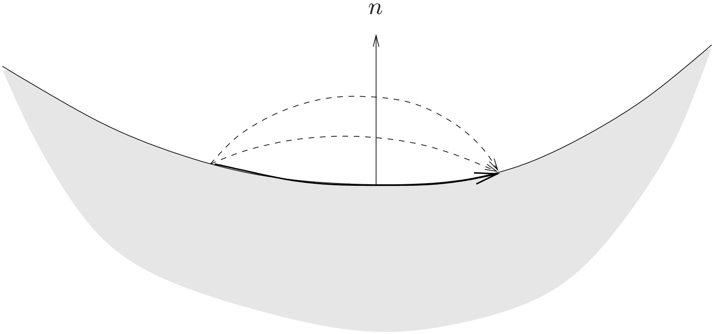

Fig. 1.11. Variation of the boundary arc

$$\xi = -\frac{u\dot{\tilde{y}}}{\sqrt{\dot{\tilde{x}}^2 + \dot{\tilde{y}}^2}}, \quad \eta = \frac{u\dot{\tilde{x}}}{\sqrt{\dot{\tilde{x}}^2 + \dot{\tilde{y}}^2}}$$

and we consider the variations of the reference curve $\tilde{x}+\xi$, $\tilde{y}+\eta$. Let $u = \varepsilon p$ for ε positive and p a nonnegative function on $[t_0, t_1]$ such that $p(t_0) = p(t_1) = 0$. The associated variation has zero boundary conditions and, from the previous computation, the length variation is

$$\begin{aligned}\delta J &= \int_{t_0}^{t_1} (F_x - \dot{F}_{\dot{x}})\xi + (F_y + \dot{F}_{\dot{y}})\eta dt \\ &= -\varepsilon \int_{t_0}^{t_1} \widetilde{T} p \sqrt{\dot{\tilde{x}}^2 + \dot{\tilde{y}}^2} dt.\end{aligned}$$

Accordingly, if the boundary arc is optimal, one must have $\widetilde{T}$ non-positive along $(\tilde{x}, \tilde{y})$. In the Riemannian case, the Legendre-Clebsch condition $F_1 > 0$ is satisfied and we ge the curvature relation between the extremal tangent to the boundary and the boundary arc itself:

$$\frac{F_{x\dot{y}} - F_{y\dot{x}}}{F_1 \left(\dot{\tilde{x}}^2 + \dot{\tilde{y}}^2\right)^{3/2}} \leq \frac{\dot{y}\ddot{x} - \ddot{y}\dot{x}}{\left(\dot{\tilde{x}}^2 + \dot{\tilde{y}}^2\right)^{3/2}}$$

which amounts to the standard convexity relation for $F = \sqrt{\dot{\tilde{x}}^2 + \dot{\tilde{y}}^2}$.

Junction Conditions with the Boundary

In this case, the variation is represented by Fig. 1.12 and concerns the entry or exit point. The geometric situation leads us to consider two central fields associated respectively to the initial and final point (labels 0 and 1).

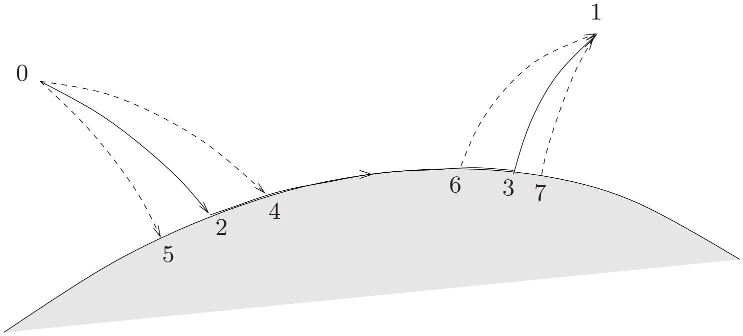

Fig. 1.12. Variations on the entry and exit junction points

In particular, consider the variation of the entry point 2 between 4 and 5. Indexing the length by the extremities of the arcs, we must estimate $l_{04} - (l_{02}+l_{24})$. This computation uses the general formula to estimate the variation of the cost between two curves. Let us denote $\gamma = (x, y)$ the extremal arc 02 defined on $[t_0, t_1]$, $\widetilde{\gamma} = (\widetilde{x}, \widetilde{y})$ the boundary arc defined on $[t_1, t_1 + h]$ $(h > 0)$, and $\gamma + \nu$ the arc 04 also defined on $[t_0, t_1]$. Since the reference curve γ is an extremal, the length variation is, up to first order,

$$\delta l = [F_{\dot{x}}\xi + F_{\dot{y}}\eta]_{t_0}^{t_1} - F(\widetilde{x}_2, \widetilde{y}_2, \dot{\widetilde{x}}_2, \dot{\widetilde{y}}_2)h$$

where $\xi(t_0) = \eta(t_0) = 0$ since the initial point is fixed, and $\xi(t_1) = h\dot{\widetilde{x}}$, $\eta(t_1) = h\dot{\widetilde{y}}$ at the junction point. Hence, the length variation is

$$\begin{aligned} \delta l &= h\left[(\dot{\widetilde{x}}F_{\dot{x}} + \dot{\widetilde{y}}F_{\dot{y}})|_{\gamma} - F|_{\widetilde{\gamma}}\right] \\ &= -hE \end{aligned}$$

where E is the *Weierstraß excess function*:

$$E(x, y, \dot{x}, \dot{y}, \dot{\widetilde{x}}, \dot{\widetilde{y}}) = F(\widetilde{x}, \widetilde{y}, \dot{\widetilde{x}}, \dot{\widetilde{y}}) - (\dot{\widetilde{x}}F_{\dot{x}}(x, y, \dot{x}, \dot{y}) + \dot{\widetilde{y}}F_{\dot{y}}(x, y, \dot{x}, \dot{y})).$$

Replacing the arc 24 by 25, we get the necessary optimality condition

$$E(x, y, \dot{x}, \dot{y}, \dot{\widetilde{x}}, \dot{\widetilde{y}}) = 0$$

at the entry point 2, $(\dot{x}, \dot{y})$ being the tangent to the reference extremal curve and $(\dot{\widetilde{x}}, \dot{\widetilde{y}})$ being the tangent to the boundary. The excess function has the following homogeneity induced by the metric:

$$E(x, y, k\dot{x}, k\dot{y}, \widetilde{k}\dot{\widetilde{x}}, \widetilde{k}\dot{\widetilde{y}}) = \widetilde{k}E(x, y, \dot{x}, \dot{y}, \dot{\widetilde{x}}, \dot{\widetilde{y}})$$

for each positive k, $\widetilde{k}$. Introducing the slopes

$$p = \frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}} = \cos\theta, \quad q = \frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}} = \sin\theta$$
$$\widetilde{p} = \frac{\dot{\widetilde{x}}}{\sqrt{\dot{\widetilde{x}}^2+\dot{\widetilde{y}}^2}} = \cos\widetilde{\theta}, \quad \widetilde{q} = \frac{\dot{\widetilde{y}}}{\sqrt{\dot{\widetilde{x}}^2+\dot{\widetilde{y}}^2}} = \sin\widetilde{\theta}$$

we get

$$E(x,y,\dot{x},\dot{y},\dot{\widetilde{x}},\dot{\widetilde{y}}) = \sqrt{\dot{\widetilde{x}}^2+\dot{\widetilde{y}}^2}\,E(x,y,p,q,\widetilde{p},\widetilde{q}).$$

In accordance with the mean value theorem, there is θ^* between θ and $\widetilde{\theta}$ such that

$$E(x,y,\cos\theta,\sin\theta,\cos\widetilde{\theta},\sin\widetilde{\theta}) = (1-\cos(\widetilde{\theta}-\theta))F_1(x,y,\cos\theta^*,\sin\theta^*).$$

In the regular case where F_1 is positive, we deduce the following.

Proposition 1.22. *In the regular case, at the entrance and exit junction points with a boundary arc, one must have $\theta = \widetilde{\theta}$: the extremal has to be tangent to the boundary.*

As a consequence, this gives the junction condition of Lemma 1.12, previously obtained as a consequence of the jump condition.

Reflection Condition on the Boundary

In this case, the variation is on the reflection point, see Fig. 1.13.

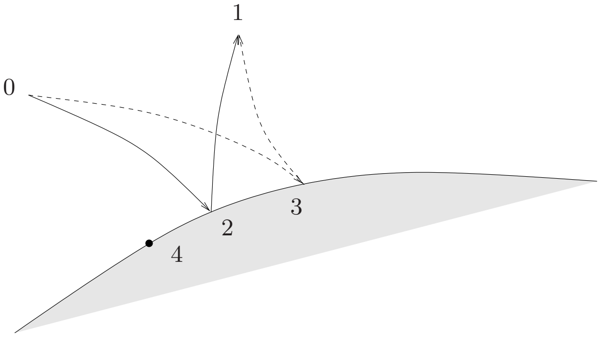

Fig. 1.13. Variations on the entry and exit junction points

The cost variation $l_{031} - l_{021}$ is evaluated by gluing together the central fields with initial point 0 and terminal point 1 along the common boundary arc 23, that is evaluating $(l_{03}-(l_{02}+l_{23}))+(l_{31}+l_{23}-l_{21})$. If h is the variation parameter on the boundary arc, we get

$$\delta l = h\left(E(x_2, y_2, \dot{x}_2^+, \dot{y}_2^+, \dot{\tilde{x}}_2, \dot{\tilde{y}}_2) - E(x_2, y_2, \dot{\bar{x}}_2, \dot{\bar{y}}_2, \dot{\tilde{x}}_2, \dot{\tilde{y}}_2)\right)$$

where $(\dot{\bar{x}}_2, \dot{\bar{y}}_2)$, $(\dot{x}_2^+, \dot{y}_2^+)$ and $(\dot{\tilde{x}}_2, \dot{\tilde{y}}_2)$ are respectively tangent to the arcs 02, 21 and 03. Hence, we get the necessary optimality condition at the reflection point in terms of the corresponding slopes:

$$E(x_2, y_2, p_2^+, q_2^+, \widetilde{p}_2, \widetilde{q}_2) = E(x_2, y_2, \bar{p}_2, \bar{q}_2, \widetilde{p}_2, \widetilde{q}_2).$$

This relation will give us the standard reflection condition if the metric is $F = \sqrt{\dot{x}^2 + \dot{y}^2}$. Indeed, $F_1(x, y, \cos\theta^*, \sin\theta^*) = 1$ and the Descartes rule is obtained:

$$\cos(\widetilde{\theta}_2 - \theta_2^-) = \cos(\widetilde{\theta}_2 - \theta_2^+).$$

The lines reflecting on the boundary must have equal angles. The same approach can be applied for the refraction rule where we glue together on the boundary two central fields with different extremal curves, see Fig. 1.14. In both cases, the rules are given by a jump condition on the adjoint state.

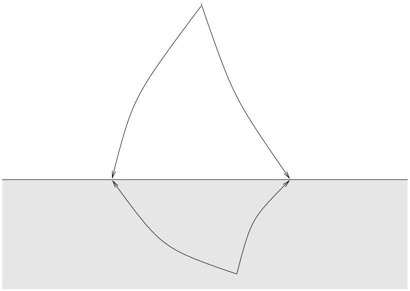

Fig. 1.14. Refraction of two central fields

Bibliographical Notes

For the maximum principle, see the introduction to the discovery in [9]. For the proof, we have followed [15]. The high order maximum principle is due to Krener [13], and the presentation using the Baker-Campbell-Hausdorff formula is inspired by [12]. Elements of symplectic geometry are borrowed from [16], and for the concept of conjugate point we use [4]. The example in sub-Riemannian geometry is excerpted from [3]. For an introduction on orbital transfer and numerical techniques, see [5, 10]. The necessary optimality conditions in the state constrained case are from [18]. For the proof in the planar case, we have followed [1].

References

1. O. Bolza (1904,). *Lectures on the calculus of variations.* Dover, New-York.
2. B. Bonnard, J.-B. Caillau, and E. Trélat (submitted,). Geometric optimal control of Keplerian orbits. *Discrete Cont. Dyn. Syst. Series B.*
3. B. Bonnard and M. Chyba (2003,). *Singular trajectories and their role in control theory.* Number 40 in Math. and Applications. Springer Verlag.
4. B. Bonnard and I. Kupka (1993,). Théorie des singularités de l'application entrée-sortie et optimalité des trajectoires singulières dans le problème du temps minimal. *Forum Mathematicum*, 5:111–159.
5. J.-B. Caillau (2000,). *Contribution à l'étude du contrôle en temps minimal des transferts orbitaux.* PhD thesis, ENSEEIHT, Institut National Polytechnique, Toulouse.
6. J.-B. Caillau, J. Gergaud, and J. Noailles (to appear in *Open Problems in Mathematical Systems and Control*, V. Blondel and A. Megretski Eds.,). *Minimum time control of the Kepler equation.* Princeton University Press.
7. M. P. Do Carmo (1993,). *Riemannian geometry.* Birkhäuser.
8. I. Ekeland (1977,). Discontinuité des champs Hamiltoniens et existence de solutions optimales en calcul des variations. *Pub. IHES*, 47:5–32.
9. R. V. Gamkrelidze (1999,). Discovery of the maximum principle. *Journal of dynamical and control systems*, 5(4):437–451.
10. J. Gergaud and T. Haberkorn (submitted,). Homotopy method for minimum consumption orbit transfer. *ESAIM COCV.*
11. C. Godbillon (1985,). *Géométrie différentielle et mécanique analytique.* Hermann, Paris.
12. H. Hermes (1978,). Lie algebras of vector fields and local approximation of attainable sets. *SIAM J. Control Optim.*, 16(5):715–727.
13. A. J. Krener (1977,). The high order maximum principle and its application to singular extremals. *SIAM J. Control Optim.*, 15(2):256–293.
14. I. Kupka. Personal communication.
15. E. B. Lee and L. Markus (1967,). *Foundations of optimal control theory.* John Wiley, New-York.
16. K. R. Meyer and G. R. Hall (1991,). *Introduction to Hamiltonian dynamical systems.* Springer Verlag.
17. M. A. Naimark (1968,). *Linear differential operators.* Frederick Ungar Publishing Co.
18. L. Pontryagin, V. Boltiansky, and R. Gamkrelidze (1974,). *Théorie mathématique des processus optimaux.* Éditions MIR, Moscow.

2

Observer Design for Nonlinear Systems

Gildas Besançon

LAG-ENSIEG, BP 46, 38402, St Martin d'Hères, France.
E-mail: Gildas.Besancon@lag.ensieg.inpg.fr

> "*Measure what is measurable and make it measurable what is not so.*"
> Galileo Galilei.

Summary: In this chapter, an overview of main observability problems and possible observer designs for nonlinear systems is proposed. In particular the observer problem and related observability conditions are first given. Then, some *basic* designs are reviewed, divided into Luenberger-like designs for so-called uniformly observable systems, and Kalman-like designs for non uniformly observable ones. Finally, two directions for more *advanced* designs - in the sense that they are based on the previously listed ones - are proposed: designs based on observer interconnections on the one hand, and designs based on system transformations on the other hand.

2.1 Introduction

When using a *system* approach in front of a problem, the issue of *observer* design arises as soon as one needs some *internal* information from *external* (directly available) measurements. In general indeed, due to sensor limitations (for cost reasons, technological constraints, *etc.*), the directly measured signals do not coincide with all signals characterizing the system behavior. Those signals of interest roughly include time-varying signals characterizing the system (*state variables*), constant ones (*parameters*), and unmeasured external ones (*disturbances*). This need for internal information can be motivated by various purposes: modelling (*identification*), monitoring (*fault detection*), or driving (*control*) the system, all these being required for keeping a system under control, as summarized by figure 2.1 hereafter. This makes the reconstruction - or observer - problem the heart of a general control problem.

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 61–89, 2006.

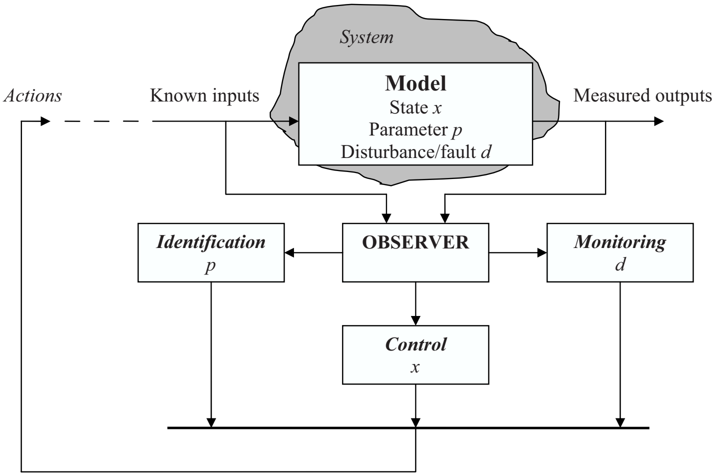

Fig. 2.1. Observer as the heart of control systems.

More formally, in short, an observer relies on a model, with on-line adaptation based on available measurements, and aiming at information reconstruction, *i.e.* it can be characterized as a model-based, measurement-based, closed-loop, information reconstructor.

Usually the model is a state-space representation, and it will be assumed here that all pieces of information to be reconstructed are born by state variables. In front of this, one can try to design an explicit differential system whose state should give an estimate of the actual state of the considered model, or just settle the problem as an optimization problem. This second case will not be considered in details here, the reader being referred to optimization tools for such an approach.

About the considered model, it can in general be either continuous-time or discrete-time, deterministic or stochastic, finite-dimensional or infinite-dimensional, smooth or "with singularities". But the presentation will be restricted here to the case of smooth, finite-dimensional, deterministic, continuous-time state-space descriptions.

In this framework, section 2.2 will be dedicated to the problem formulation together with main related definitions (observer and observability). Section 2.3 will present main available observer designs in terms of two categories: those with a constant (input-independent) correction gain, and those with a time-varying (possibly input-dependent) correction gain. Section 2.4 will then propose some ways to extend those designs to more general classes of systems, either by means of interconnexions, or by transformations. Some conclusions and further remarks will finally be given in section 2.5.

Notice that the material presented here results from some own research and viewpoint on the problem, in the continuity of [5] for instance, as well as from the quite large amount of available results, including overviews of [12] or [21] for instance.

In all the subsequent sections, the following notations/terminology will be used:

- I for the identity matrix of appropriate dimensions,
- v_i for the ith component of a vector v,
- $M = M^T > 0$ for a symmetric positive definite matrix M,
- *Stable matrix* for a matrix with all eigenvalues having strictly negative real parts.

2.2 Main Problem and Definitions

2.2.1 Problem Formulation

Model Under Consideration

Let us assume that the system under consideration can be described by a differential state-space representation of the following form:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t)) \end{aligned} \tag{2.1}$$

where x denotes the state vector, taking values in X a connected manifold of dimension n, u denotes the vector of known external inputs, taking values in some open subset U of $\mathbf{R}^m$, and y denotes the vector of measured outputs taking values in some open subset Y of $\mathbf{R}^p$.
Functions f and h will in general be assumed to be $\mathcal{C}^\infty$ w.r.t. their arguments, and input functions $u(.)$ to be locally essentially bounded and measurable functions in a set $\mathcal{U}$.
The system will be assumed to be complete.
More generally, the dynamics might explicitly depend on time via $f(x(t), u(t), t)$, while y might further directly depend on u and even t, via $h(x(t), u(t), t)$. Such an explicitly time-dependent system is usually called 'time-varying':

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), t) \\ y(t) &= h(x(t), u(t), t) \end{aligned} \tag{2.2}$$

Various particular cases arise as follows:

- Control-affine systems:

$$f(x,u) = f_0(x) + g(x)u$$

- State-affine systems[1]:

$$f(x,u) = A(u)x + B(u), \quad h(x) = Cx \ (or \ C(u)x + D(u))$$

- Linear Time-Varying (LTV) systems:

$$f(x,u,t) = A(t)x + B(t)u, \quad h(x,u,t) = C(t)x + D(t)u$$

- Linear Time-Invariant (LTI) systems:

$$f(x,u) = Ax + Bu, \quad h(x,u) = Cx + Du$$

In any case, let $\chi_u(t, x_{t_0})$ denote the solution of the state equation in (2.1) under the application of input u on $[t_0, t]$ and satisfying $\chi_u(t_0, x_{t_0}) = x_{t_0}$.

Observer Problem

Given a model (2.1), the purpose of acting on the system, or monitoring it, will in general need to know $x(t)$, while in practice one has only access to u and y. The observation problem can then be formulated as follows:

> Given a system described by a representation (2.1), find an estimate $\hat{x}(t)$ for $x(t)$ from the knowledge of $u(\tau), y(\tau)$ for $0 \leq \tau \leq t$.

Clearly this problem makes sense when one cannot invert h w.r.t. x at any time.
In front of this, one can look for a solution in terms of optimization, by looking for the best estimate $\hat{x}(0)$ of $x(0)$ which can explain the evolution $y(\tau)$ over $[0, t]$, and from this, get an estimate $\hat{x}(t)$ by integrating (2.1) from $\hat{x}(0)$ and under $u(\tau)$. In order to cope with disturbances, one should rather optimize the estimate of some initial state over a moving horizon, namely minimize some criterion of the form:

$$\int_{t-T}^{t} \|h(\chi_u(\tau, z_{t-T})) - y(\tau)\|^2 d\tau$$

w.r.t. z_{t-T} for any $t > T$, and $y(\tau)$ corresponding to the measured output over $[t-T, t]$ under the effect of the considered input u.
This is a general formulation for a solution to the problem, relying on available optimization tools and results for practical use and guarantees (see *e.g.* [40, 37,

[1] with bilinear systems as a particular case

1]): so it takes advantage of its systematic formulation, but suffers from usual drawbacks of nonlinear optimization (computational burden, local minima...). Alternatively, one can use the idea of an explicit "feedback" in estimating $x(t)$, as this is done for control purposes: more precisely, noting that if one knows the initial value $x(0)$, one can get an estimate for $x(t)$ by simply integrating (2.1) from $x(0)$, the feedback-based idea is that if $x(0)$ is unknown, one can try to correct on-line the integration $\hat{x}(t)$ of (2.1) from some erroneous $\hat{x}(0)$, according to the measurable error $h(\hat{x}(t))-y(t)$, namely to look for an estimate $\hat{x}$ of x as the solution of a system:

$$\dot{\hat{x}}(t) = f(\hat{x}(t), u(t)) + k(t, h(\hat{x}(t)) - y(t)), \ \textit{with } k(t, 0) = 0. \tag{2.3}$$

Such an auxiliary system is what will be defined as an *observer*, and the above equation is the most common form of an observer for a system (2.1) (as in the case of linear systems [31, 36]).
More generally, an observer can be defined as follows:

Definition 2.1. *Observer.*
Considering a System (2.1), an observer is given by an auxiliary system:

$$\begin{aligned} \dot{X}(t) &= F(X(t), u(t), y(t), t) \\ \hat{x}(t) &= H(X(t), u(t), y(t), t) \end{aligned} \tag{2.4}$$

such that:

(i) $\hat{x}(0) = x(0) \Rightarrow \hat{x}(t) = x(t), \quad \forall t \geq 0$*;*

(ii) $\|\hat{x}(t) - x(t)\| \to 0$ *as* $t \to \infty$*;*

If (ii) holds for any $x(0), \hat{x}(0)$*, the observer is* global*.*
If (ii) holds with exponential convergence, the observer is exponential*.*
If (ii) holds with a convergence rate which can be tuned, the observer is tunable*.*

Notice that the overview on observer design presented in the sequel will mainly be dedicated to global exponential tunable observers.
Notice also that with notations of (2.1) and (2.4), the difference $\hat{x} - x$ will be called *observer error*.
Notice finally that with the above point of view, the observation problem turns to be a problem of observer design.

2.2.2 Conditions for a Solution

The observation problem arises as soon as one does not have directly access to the state vector at each time t, but only to a function of the state corresponding to the measured output $y(t)$. Thus at a first glance, the problem will

be solvable only if $y(t)$ bears the information on the full state vector when considered over some time interval: this roughly corresponds to the notion of "*observability*".
However, when restricting the definition of an observer strictly to items (i)-(ii), one can find observers yielding solutions to the observation problem even in cases when y does not bear the full information on the state vector:
Consider for instance the simple system:

$$\dot{x} = -x + u, \;\; y = 0$$

Clearly one cannot get any information on x from y, and yet the system:

$$\dot{\hat{x}} = -\hat{x} + u$$

satisfies (i)-(ii) and yields an estimate of x, since:

$$\overset{\cdot}{\overbrace{\hat{x} - x}} = -(\hat{x} - x).$$

This corresponds to a notion of "*detectability*". Notice that in that case, however, the rate of convergence cannot be tuned. Additional remarks in that respect can be found *e.g.* in [5].
If we restrict ourselves to the case of observers in the sense of *tunable* observers, then observability becomes a necessary condition.

About "Necessary" Conditions

For a possible design of a (tunable) observer, one must be able to recover the information on the state via the output measured from the initial time, and more particularly to recover the corresponding initial value of the state. This means that observability is characterized by the fact that from an output measurement, one must be able to distinguish between various initial states, or equivalently, one cannot admit *indistinguishable* states (following [28]):

Definition 2.2. *Indistinguishability.*
A paire $(x_0, x'_0) \in \mathbf{R}^n \times \mathbf{R}^n$ is indistinguishable for a System (2.1) if:

$$\forall u \in \mathcal{U}, \; \forall t \geq 0, \; h(\chi_u(t, x_0)) = h(\chi_u(t, x'_0)).$$

A state x is indistinguishable from x_0 if the paire (x, x_0) is indistinguishable.

From this, observability can be defined:

Definition 2.3. *Observability [resp. at x_0].*
A System (2.1) is observable [resp. at x_0] if it does not admit any indistinguishable paire [resp. any state indistinguishable from x_0].

This definition is quite general (global), and even too general for practical use, since one might be mainly interested in distinguishing states from their neighbors:
Consider for instance the case of the following system:

$$\dot{x} = u, \quad y = sin(x). \tag{2.5}$$

Clearly, y cannot help distinguishing between x_0 and $x_0 + 2k\pi$, and thus the system is not observable. It is yet clear that y allows to distinguish states of $]-\frac{\pi}{2}, \frac{\pi}{2}[$.
This brings to consider a weaker notion of observability:

Definition 2.4. *Weak observability [resp. at x_0].*
A System (2.1) is weakly observable [resp. at x_0] if there exists a neighborhood U of any x [resp. of x_0] such that there is no indistinguishable state from x [resp. x_0] in U.

Notice that this does not prevent from cases where the trajectories have to go far from U before one can distinguish between two states of U.
Consider for instance the case of a system:

$$\dot{x} = u; \quad y = h(x)$$

with h a $\mathcal{C}^\infty$ function as in figure 2.2 below: clearly the system is weakly observable since any state is distinguishable from any other one by applying some nonzero input u, but distinguishing two points of $[-1, 1]$ needs to wait for y to move away from 0.

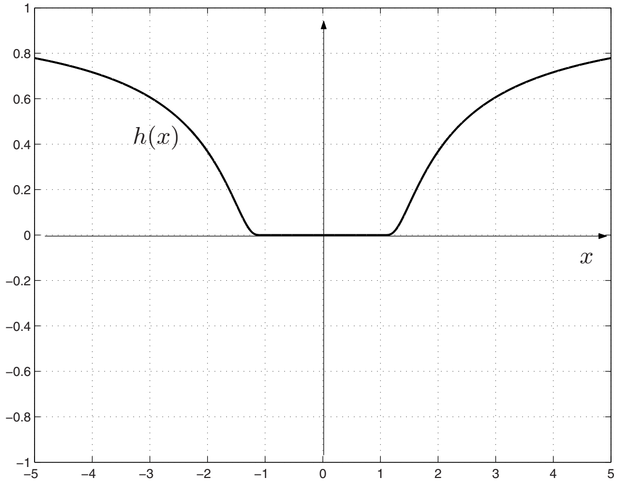

Fig. 2.2. Output function of a weakly but not locally observable system.

Hence, to prevent from this situation, an even more local definition of observability can be given:

Definition 2.5. *Local weak observability [resp. at x_0].*
A System (2.1) is locally weakly observable [resp. at x_0] if there exists a neighborhood U of any x [resp. of x_0] such that for any neighborhood V of x [resp. x_0] contained in U, there is no indistinguishable state from x [resp. x_0] in V when considering time intervals for which trajectories remain in V.

This roughly means that one can distinguish every state from its neighbors without "going too far". This notion is of more interest in practice, and also presents the advantage of admitting some 'rank condition' characterization. Such a condition relies on the notion of *observation space* roughly corresponding to the space of all observable states:

Definition 2.6. *Observation space.*
The observation space for a System (2.1) is defined as the smallest real vector space (denoted by $\mathcal{O}(h)$) of $\mathcal{C}^\infty$ functions containing the components of h and closed under Lie derivation along $f_u := f(.,u)$ for any constant $u \in$ $\mathbf{R}^m$ (namely such that for any $\varphi \in \mathcal{O}(h)$, $L_{f_u}\varphi \in \mathcal{O}(h)$, where $L_{f_u}\varphi(x) = \frac{\partial\varphi}{\partial x} f(x,u)$).

Definition 2.7. *Observability rank condition [resp. at x_0].*
A System (2.1) is said to satisfy the observability rank condition [resp. at x_0] if:

$$\forall x, \quad dimd\mathcal{O}(h)\mid_x = n \quad [\text{resp. } dimd\mathcal{O}(h)\mid_{x_0} = n]$$

where $d\mathcal{O}(h)\mid_x$ is the set of $d\varphi(x)$ with $\varphi \in \mathcal{O}(h)$.

From this we have [28]:

Theorem 2.1. *A System (2.1) satisfying the observability rank condition at x_0 is locally weakly observable at x_0.*
More generally a System (2.1) satisfying the observability rank condition is locally weakly observable.
Conversely, a System (2.1) locally weakly observable satisfies the observability rank condition in an open dense subset of X.

As an example, consider again System (2.5): for this system clearly $d\mathcal{O}(h) = span\{cos(x)dx, sin(x)dx\}$ and thus $dimd\mathcal{O}(h)\mid_{x_0} = 1$ for any x_0, namely the system satisfies the observability rank condition.
As a second example, consider a system:

$$\begin{aligned} \dot{x} &= Ax \\ y &= Cx \quad \text{with } x \in \mathbf{R}^n. \end{aligned} \tag{2.6}$$

For this system, the observability rank condition is equivalent to local weak observability (which is itself equivalent to observability) and is characterized by the so-called *Kalman rank condition*:

Theorem 2.2. *For a system of the form (2.6):*

- *The observability rank condition is equivalent to rank*$\mathcal{O}_m = n$ *with* $\mathcal{O}_m = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{pmatrix}$ *the observability matrix;*
- *The observability rank condition is equivalent to the observability of the system.*

The first point results from straightforward computations (*e.g.* as in [29]), while the second one results from the definition of observability (see *e.g.* [34]).

Notice that if System (2.6) satisfies the above observability rank condition, the paire (A, C) is usually called *observable*.
Notice also that the above result also holds for controlled systems with $\dot{x} = Ax + Bu$.
Notice finally that the above observability rank condition is also sufficient for a possible observer design for (2.6) (and necessary and sufficient for a *tunable* observer design - see later).
However, in general, the observability rank condition is not enough for observer design: this is due to the fact that in general, observability depends on the inputs, namely it does not prevent from the existence of inputs for which observability vanishes.
As a simple example, consider the following system:

$$\begin{aligned} \dot{x} &= \begin{pmatrix} 0 & u \\ 0 & 0 \end{pmatrix} x \\ y &= (1\ 0)x \end{aligned} \tag{2.7}$$

it is clearly observable for any constant input $u \neq 0$, but not observable for $u = 0$.
This means that the purpose of observer design requires a look at the inputs.

About “Sufficient” Conditions

Here are discussed sufficient conditions for possible observer designs, related to inputs. Effective designs will be discussed later.
More precisely, usual notions of *universal inputs* and *uniform observability* for systems (2.1) are first introduced (as in [12] for instance), and the stronger notions of *persistency* and *regularity* classically defined for state affine systems [12] are then presented for the more general case of systems (2.1).

Definition 2.8. *Universal inputs [resp. on* $[0,t]$*].*
An input u *is universal (resp. on* $[0,t]$*) for system (2.1) if* $\forall x_0 \neq x_0'$*,* $\exists \tau \geq 0$ *(resp.* $\exists \tau \in [0,t]$*) s.t.* $h(\chi_u(\tau, x_0)) \neq h(\chi_u(\tau, x_0'))$*.*
An input u *is singular if it is not universal.*

As an example, for System (2.7), $u(t) = 0$ is a singular input.

It can be underlined here that for $\mathcal{C}^w$ systems, universal $\mathcal{C}^w$ inputs are dense in the set of $\mathcal{C}^w$ functions for the topology induced by $\mathcal{C}^\infty$ [39].
But one has to also notice that in general characterizing singular inputs is not easy. Things are easier for systems which do not admit such singular inputs:

Definition 2.9. *Uniformly observable systems (resp. locally).*
A system is uniformly observable (UO) if every input is universal (resp. on $[0,t]$*).*

Example 2.1. The System (2.8) below is uniformly observable [18]:

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \\ & & & & 0 \\ \vdots & & & & 1 \\ 0 & \cdots & & & 0 \end{pmatrix} x + \begin{pmatrix} \varphi_1(x_1) \\ \varphi_2(x_1, x_2) \\ \vdots \\ \varphi_n(x_1, \ldots, x_{n-1}) \\ \varphi_{n-1}(x_1, \ldots, x_n) \end{pmatrix} u \tag{2.8}$$
$$y = x_1; \quad x = (x_1 ..., x_n)^T$$

This property means that observability is independent of the inputs and thus can allow an observer design also independent of the inputs, as in the case of LTI systems (see later).
For systems which are not uniformly observable, in general possible observers will depend on the inputs, and not all inputs will be admissible. Restricting the set of inputs to universal ones, as in the case of uniformly observable systems - for which *all* inputs are universal, is actually not enough:
Consider for instance the following system:

$$\dot{x} = \begin{pmatrix} 0 & u \\ -u & 0 \end{pmatrix} x; \quad y = (1 \;\; 0)x$$

For this system, the input defined by $u(t) = 1$ for $t < t_1$ and $u(t) = 0$ for $t \geq t_1$ is clearly universal, but if a disturbance appears after t_1, it is also clear that x cannot be correctly reconstructed.
This shows that universality must be guaranteed over the time, namely must be *persistent*. In order to characterize this persistency, notice first that we have the following property:

Proposition 2.1. *An input u is a universal input on $[0, t]$ for System (2.1) if: and only if $\int_0^t ||h(\chi_u(\tau, x_0) - h(\chi_u(\tau, x_0')||^2 d\tau > 0$ for all $x_0 \neq x_0'$.*

This can be easily checked from Definition 2.8.
Then one can define persistency as follows:

Definition 2.10. *Persistent inputs.*
An input u is a persistent input for a System (2.1) if

$$\exists t_0, T : \forall t \geq t_0, \, \forall x_t \neq x_t', \quad \int_t^{t+T} ||h(\chi_u(\tau, x_t) - h(\chi_u(\tau, x_t')||^2 d\tau > 0$$

This guarantees observability over a given time interval, but if this observability tends to vanish as time goes to infinity, effective observers would in general have to compensate this by a correction gain also going to infinity: Consider for instance the system defined by:

$$\dot{x} = u \; y = \frac{1}{(1+t)^2} x = h(x, t)$$

This system is obviously observable with persistency, but it is also clearly *less and less observable* as $t \to \infty$.
The state x could here be reconstructed by an auxiliary system the form (2.3) for instance given as follows:

$$\dot{\hat{x}} = u - k * (1+t)^2 * (h(\hat{x}, t) - y),$$

which indeed guarantees that $\hat{x} - x \to \infty$, but with a correction gain $k * (1+t)^2$ growing to infinity.
In order to avoid this, one needs a *guarantee* of observability, namely some *regular persistency*:

Definition 2.11. *Regularly persistent inputs.*
An input u is a regularly persistent input for a system (2.1) if:

$$\exists t_0, T, \alpha : \forall x_t, x_t', \; \forall t \geq t_0, \int_t^{t+T} ||h(\chi_u(\tau, x_t) - h(\chi_u(\tau, x_t')||^2 d\tau \geq \alpha \|x_t - x_t'\|^2$$

From the above proposed definitions of persistency and regular persistency, we recover the usual definitions already available for state affine systems (of [12] for instance):

Proposition 2.2. *For state affine systems, regularly persistent inputs are inputs u such that:*

$$\exists t_0, T, \alpha : \int_t^{t+T} \Phi_u^T(\tau, t) C^T C \Phi_u(\tau, t) d\tau \geq \alpha I > 0 \quad \forall t \geq t_0, \tag{2.9}$$

with $\Phi_u(\tau,t)$ the transition matrix classically defined by:

$$\frac{d\Phi_u(\tau,t)}{d\tau} = A(u(\tau))\Phi_u(\tau,t),\ \Phi_u(t,t) = I.$$

This is a straight consequence of the application of definition 2.11 to the case of state affine systems.

Remark 2.1.

- Regularly persistent inputs for state affine systems are those making the system an LTV system *Uniformly Completely Observable* in the sense of Kalman [31] (since uniform complete observability for LTV systems is typically defined by (2.9);
- For general nonlinear systems, the definition is not of easy use, while for state affine or LTV systems, it is independent of initial states.

As an example of input properties, consider the following system:

$$\dot{x} = \begin{pmatrix} 0 & u \\ 0 & 0 \end{pmatrix};\ y = (1\ 0)x$$

For this system, the input for instance defined by:

$$\begin{array}{l} u(t) = 1 \textit{ on } t \in [2kT, (2k+1)T[,\ k \geq 0 \\ u(t) = 0 \textit{ on } t \in [(2k+1)T, (2k+2)T[,\ k \geq 0 \end{array}$$

is regularly persistent, while that defined by:

$$\begin{array}{l} u(t) = 1 \textit{ on } t \in [2kT, (2k+\frac{1}{k+1})T[,\ k \geq 0 \\ u(t) = 0 \textit{ on } t \in [(2k+\frac{1}{k+1})T, (2k+2)T[,\ k \geq 0 \end{array}$$

is not [12].

All this will tell us on some possible observer designs for classes of systems, as discusses in next section.

Remark 2.2.

- If a system, *e.g.* control affine, is not observable in the sense of rank condition, it can be decomposed into observable and non observable subsystems as follows [29]:

  $$\begin{array}{rl} \dot{\zeta}_1 &= f_1(\zeta_1,\zeta_2) + g_1(\zeta_1,\zeta_2)u \\ \dot{\zeta}_2 &= f_2(\zeta_2) + g_2(\zeta_2)u \\ y &= h_2(\zeta_2) \end{array}$$

  where the subsystem in ζ_2 satisfies the observability rank condition. In that case one has to work on ζ_2.

- If the considered system is not observable, but satisfies the following: $\forall u \; such \; that \; x_0 \, and \, x_0'$ are indistinguishable by u :

  $$\chi_u(t, x_0) - \chi_u(t, x_0') \to 0 \, as \, t \to \infty$$

  it satisfies a property of *detectability*, and in that case one may have the opportunity to design an observer in the sense of (i) and (ii).

In summary:

- For uniformly observable systems, one might design uniform observers (but not only)

- For non-uniformly observable systems, one might design non uniform observers (but not only).

The first ones correspond to the so-called *Luenberger observer* for LTI systems [36], while the second ones correspond to the case of *Kalman observers* for LTV systems [31].

2.3 Some "Basic" Designs

Some observers are presented here for particular structures of systems. In the whole section, an observer is to be understood as a *global, exponential, tunable* observer.

2.3.1 Observer designs for Linear Structures

Luenberger Observer (for LTI Systems)

Let us consider here LTI systems of the following form:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \tag{2.10}$$

For those systems we have the following classical (Luenberger) result [36]:

Theorem 2.3. *If System (2.10) satisfies the observability rank condition then there exists an observer of the form:*

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) - K(C\hat{x}(t) - y(t))$$

with K such that $A - KC$ is stable.

Remark 2.3. The rate of convergence can be arbitrarily chosen by appropriate design of K.

This can be established by showing that observability guarantees the existence of a transformation into a so-called observability canonical form, for which the design of an appropriate observer gain is straightforward (see *e.g.* [34]).

Kalman Observer (for LTV Systems)

Let us consider here LTV systems of the following form:

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + Bu(t) \\ y(t) &= C(t)x(t) \end{aligned} \tag{2.11}$$

with $A(t), C(t)$ uniformly bounded.
For those systems we have the following (Kalman-related) result [31, 13, 27, 9, 21]:

Theorem 2.4. *If System (2.11) is uniformly completely observable, then there exists an observer of the form:*

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) - K(t)(C(t)\hat{x}(t) - y(t))$$

with $K(t)$ given by:

$$\begin{aligned} \dot{M}(t) &= A(t)M(t) + M(t)A^T(t) - M(t)C^T(t)W^{-1}C(t)M(t) + V + \delta M(t) \\ M(0) &= M_0 = M_0^T > 0,\ W = W^T > 0 \\ K(t) &= M(t)C^T(t)W^{-1} \end{aligned} \tag{2.12}$$

with either $\delta > 2\|A(t)\|$ for all t, or $V = V^T > 0$.

Remark 2.4.

- The rate of convergence can be tuned by δ or V.
- For $\delta = 0$, we get the classical Kalman observer, the usual related condition for convergence being that (A, V) be *uniformly completely controllable* (dual of uniform complete observability).
- For $\delta = 0$, the observer is optimal in the sense of minimizing w.r.t. z:

  $$\int_0^t [(C(\tau)z(\tau) - y(\tau))^T W^{-1}(C(\tau)z(\tau) - y(\tau)) + v^T(\tau)V^{-1}v(\tau)]d\tau$$
  $$+(z_0 - \hat{x}_0)^T M_0^{-1}(z_0 - \hat{x}_0)$$

  under $\dot{z}(t) = A(t)z(t) + v(t)$ $y(t) = C(t)z(t)$.
  Namely, it provides an explicit solution to the optimization-based approach

mentioned in the introduction.
It is also optimal in the sense of minimizing the mean of the square estimation error for a system affected by state white noises and measurement white noises, uncorrelated to each other, with V and W as respective variance matrices [34].

- The observer gain can also be computed as $K(t) = S^{-1}(t)C^T W^{-1}$ where S is the solution of:

  $$\begin{array}{l} \dot{S}(t) = -A^T(t)S(t) - S(t)A(t) + C^T(t)W^{-1}C(t) - \delta S(t) - S(t)VS(t) \\ S(0) = S^T(0) > 0 \end{array}$$

  which makes it a linear equation in S whenever V is chosen equal to 0. This is also true for all subsequent *Kalman-like* designs, even if they will be expressed in terms of (2.12).

The result of Theorem 2.4 can be established by showing that:

(i) $\exists \alpha_1, \alpha_2, t_0$ such that $\forall t \geq t_0 : \ 0 < \alpha_1 I \leq M^{-1}(t) \leq \alpha_2 I$ basically from the condition of uniform complete observability;

(ii) $V(e,t) = e^T(t)M^{-1}(t)e(t)$ where $e := \hat{x} - x$ is a Lyapunov function for the observer error equation, which is exponentially decaying with a rate of decay tunable via δ or the minimal eigenvalue of V.

This can be shown either when $V = 0$ and $\delta > 2\|A(t)\|$ [9, 27], or when $V = V^T > 0$ and $\delta = 0$ [21].

On the basis of Theorem 2.4, an extension can be intuitively derived for *nonlinear* systems relying on its first order approximation along the estimated trajectories, and known as *Extended Kalman Filter* (see *e.g.* [22]):

Definition 2.12. *Extended Kalman Filter (EKF).*
Given a nonlinear system of the form:

$$\begin{array}{l} \dot{x}(t) = f(x(t), u(t)) \\ y(t) = h(x(t)) \end{array}$$

the corresponding Extended Kalman Filter is given by:

$$\dot{\hat{x}}(t) = f(\hat{x}(t), u(t)) - K(t)(h(\hat{x}(t)) - y(t))$$

where $K(t)$ is given as in the Kalman observer (2.12) with:

$$A(t) := \frac{\partial f}{\partial x}(\hat{x}(t), u(t)), \quad C(t) := \frac{\partial h}{\partial x}(\hat{x}(t))$$

This yields a candidate for a systematic observer design in front of a nonlinear system, but in general the convergence is not guaranteed, except under specific structure conditions (or domain of validity).

2.3.2 Observer Designs for Nonlinear Structures

In this section are presented some observer designs restricted to specific structures of nonlinear systems, and extending observers listed above for linear systems.

Luenberger-Like Design (for UO Systems)

Let us first consider classes of systems for which observability does not depend on the input, namely Uniformly Observable systems.
The idea is basically to rely on a linear time-invariant part in order to design a gain as in Luenberger observers, and either compensate exactly all nonlinear elements when possible, or dominate them via the linear part.

Additive Output Nonlinearity

Consider here a system of the form:

$$\begin{aligned} \dot{x} &= Ax + \varphi(Cx, u) \\ y &= Cx \end{aligned} \tag{2.13}$$

Here the nonlinearity can be constructed from direct measurements and thus compensated in the observer design (as originally proposed in [32, 33] for instance):

Theorem 2.5. *If (A, C) is observable, System (2.13) admits an observer of the form:*

$$\dot{\hat{x}} = A\hat{x} + \varphi(y, u) - K(C\hat{x} - y)$$

with K such that $A - KC$ is stable.

Remark 2.5.
Clearly here, the observer error is exactly linear, and thus the convergence rate can be arbitrarily tuned by appropriate choice of K as in the case of linear systems.

Additive Triangular Nonlinearity

Consider here a system of the form:

$$\begin{aligned} \dot{x} &= A_0 x + \varphi(x, u) \\ y &= C_0 x \\ \text{with} \quad A_0 &= \begin{pmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ & & & 1 \\ 0 & & & 0 \end{pmatrix}, \quad C_0 = (1 \ 0 \cdots \ 0). \end{aligned} \tag{2.14}$$

Here the idea will be to use the uniform observability, and thus a structure as in (2.8), to weight a gain based on the linear part, so as to make the linear dynamics of the observer error to dominate the nonlinear one [19, 12, 21]:

Theorem 2.6. *If φ is globally Lipschitz w.r.t. x, uniformly w.r.t. u and such that:*

$$\frac{\partial \varphi_i}{\partial x_j}(x,u) = 0 \, for \, j \geq i+1, \quad 1 \leq i,j \leq n,$$

System (2.14) admits an observer of the form:

$$\dot{\hat{x}} = A_0\hat{x} + \varphi(\hat{x},u) - \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda^n \end{pmatrix} K_0(C_0\hat{x} - y)$$

with K_0 such that $A_0 - K_0C_0$ is stable, and λ large enough.

Remark 2.6.

- This design is known as *high gain observer* since it relies on the choice of some sufficiently large tuning parameter λ;
- The larger λ is, the faster the convergence is.
- This design can be extended to systems of the following form [15, 20, 21]:

  $$\dot{x}(t) = f(x(t),u(t)), \; y(t) = C_0x(t)$$

  where $\frac{\partial f_i}{\partial x_j} = 0$ for $j > i+1$ and $\frac{\partial f_i}{\partial x_{i+1}} \geq \alpha_i > 0$ for all x,u;
- The design can also be extended to multi-output uniformly observable systems [14].

The result of Theorem 2.6 can be established by showing that $V(e) = e^T P(\lambda)e$ is a Lyapunov function for the observer error equation, exponentially decaying with a rate of decay being tunable via λ, where:

$$e = \hat{x} - x \quad \text{and} \quad P(\lambda) = \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda^n \end{pmatrix}^{-1} P_0 \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda^n \end{pmatrix}^{-1},$$

with P_0 such that:

$$P_0(A_0 - K_0C_0) + (A_0 - K_0C_0)^T = -I.$$

Kalman-Like Design (for Non-UO Systems)

In the case when observability depends on the inputs (systems which are not uniformly observable), the design will be restricted to some appropriate classes of inputs. Then the two possible cases of compensable or non compensable nonlinearities can again be considered.

State Affine Systems

Consider here a system of the form:

$$\begin{aligned} \dot{x}(t) &= A(u(t))x(t) + B(u(t)) \\ y(t) &= Cx(t) \end{aligned} \tag{2.15}$$

with $A(u(t))$ uniformly bounded.
Here the idea is that imposing the input function yields a linear time-varying system. Hence the following Kalman-like result holds [27, 12, 9]:

Theorem 2.7. *If u is regularly persistent for (2.15), then the system admits an observer of the form:*

$$\dot{\hat{x}}(t) = A(u(t))\hat{x}(t) + B(u(t)) - K(t)(C\hat{x}(t) - y(t))$$

with $K(t)$ given by:

$$\begin{aligned} \dot{M}(t) &= M(t)A^T(u(t)) + A(u(t))M(t) - M(t)C^TW^{-1}CM(t) + V + \delta M(t) \\ M(0) &= M^T(0) > 0,\, W = W^T > 0 \\ K(t) &= M(t)C^TW^{-1} \end{aligned}$$

with $\delta > 2\|A(u(t))\|$ or $V = V^T > 0$ as in LTV systems.

Remark 2.7.
The convergence rate can be tuned by appropriate choice of δ or V.

This design can clearly be extended to systems which are affine in the unmeasured states, up to additive output nonlinearity, of the following form [24, 9]:

$$\begin{aligned} \dot{x}(t) &= A(u(t), Cx(t))x(t) + B(u(t), Cx(t)) \\ y(t) &= Cx(t) \end{aligned} \tag{2.16}$$

with $A(u(t), C\chi_u(t, x_0))$ bounded for any x_0.

Theorem 2.8. *If u is regularly persistent for (2.16), in the sense that it makes $v(t) := \begin{pmatrix} u(t) \\ C\chi_u(t, x_0) \end{pmatrix}$ regularly persistent for $\dot{x}(t) = A(v(t))x(t)$ for any x_0, then the system admits an observer of the form:*

$$\dot{\hat{x}}(t) = A(u(t), y(t))\hat{x}(t) + B(u(t), y(t)) - K(t)(C\hat{x}(t) - y(t))$$

with $K(t)$ given by:

$$\begin{aligned} \dot{M}(t) &= M(t)A^T(u(t), y(t)) + A(u(t), y(t))M(t) - M(t)C^TW^{-1}CM(t) \\ &\quad +V + \delta M(t) \\ M(0) &= M^T(0) > 0,\, W = W^T > 0 \\ K(t) &= M(t)C^TW^{-1} \end{aligned}$$

with $\delta > 2\|A(u(t), y(t))\|$ or $V = V^T > 0$.

Systems Affine in the Unmeasured States + Additive Triangular Nonlinearity

Combining structure of System (2.15) or more generally (2.16) with that of System (2.14) leads to consider systems of the following form:

$$\begin{aligned} \dot{x} &= A_0(u, y)x + \varphi(x, u) \\ y &= C_0x \quad \text{with} \\ A_0(u, y) &= \begin{pmatrix} 0 & a_{12}(u, y) & & 0 \\ & & \ddots & \\ & & & a_{n-1n}(u, y) \\ 0 & & & 0 \end{pmatrix} \textit{ bounded}, \quad C_0 = (1\ 0 \cdots 0), \end{aligned} \tag{2.17}$$

and with φ as in Theorem 2.6.
However, this means that the observer will need to rely on high gain, but for a non uniformly observable system. As a consequence, a stronger observability property will be required, roughly corresponding to observability for short times [12, 3]:

Definition 2.13. *Locally regular inputs.*
An input function u is locally regular for (2.17) if $\exists \alpha, \lambda_0 : \forall \lambda \geq \lambda_0, \forall t \geq \frac{1}{\lambda}, \forall x_0$,

$$\int_{t-\frac{1}{\lambda}}^{t} \Phi^T_{v(x_0,.)}(\tau, t)C^TC\Phi_{v(x_0,.)}(\tau, t)d\tau \geq \alpha\lambda \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda^n \end{pmatrix}^{-2}$$

where $v(x_0, .)$ stands for $\begin{pmatrix} u(.) \\ C\chi_u(., x_0) \end{pmatrix}$ and $\Phi_{v(x_0,.)}$ for the transition matrix of $\dot{x}(t) = A(v(x_0, t))x(t)$.

Theorem 2.9. *If φ is globally Lipschitz w.r.t. x, uniformly w.r.t. u and such that:*

$$\frac{\partial \varphi_i}{\partial x_j}(x, u) = 0 \, for \, j \geq i + 1, \quad 1 \leq i, j \leq n,$$

and u is locally regular for (2.14), then the system admits an observer of the form:

$$\dot{\hat{x}} = A_0(u,y)\hat{x} + \varphi(\hat{x},u) - \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda^n \end{pmatrix} K_0(t)(C_0\hat{x} - y)$$

with $K_0(t)$ given by:

$$\begin{aligned} \dot{M}(t) &= \lambda(M(t)A^T(u(t),y(t)) + A(u(t),y(t))M(t) - M(t)C^TW^{-1}CM(t) + \delta M(t)) \\ M(0) &= M^T(0) > 0,\ W = W^T > 0 \\ K(t) &= M(t)C^TW^{-1} \end{aligned}$$

$\delta > 2\|A(u,y)\|$ and λ large enough.

This can be established by showing that [3]:

(i) From local regularity assumption:

$$\exists\lambda > 0,\ \forall\lambda \geq \lambda_0,\ \forall t \geq \frac{1}{\lambda}, \quad 0 < \alpha_1 I \leq M^{-1}(t) \leq \alpha_2 I$$

for α_1, α_2 independent of λ.

(ii) $V(e,t) = e^TP(\lambda,t)e$ is a Lyapunov function for the observer error equation, exponentially decaying, with a rate of decay tunable by λ, where $e = \hat{x} - x$ and

$$P(\lambda,t) = \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda^n \end{pmatrix}^{-1} M^{-1}(t) \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda^n \end{pmatrix}^{-1}$$

2.4 Some "Advanced" Designs

The presentation of possible observer designs in previous section has been restricted to very specific structures of systems. In this section are presented some ways to deal with nonlinear systems which do not satisfy the structures previously presented.

2.4.1 Interconnection-based Design

The first way to extend the class of systems for which an observer can be designed is to interconnect observers in order to design an observer for some interconnected system, when possible. If indeed a system is not under a form

for which an observer is already available, but can be seen as an interconnection between several subsystems each of which would admit an observer if the states of the other subsystems were known, then a candidate observer for the interconnection of these subsystems is given by interconnecting available sub-observers (*e.g.* as in [10]).
As an example, consider the following system:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= u_1 \\ \dot{x}_3 &= x_4 + \varphi(x_2) \\ \dot{x}_4 &= u_2 \\ y &= \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} \end{aligned} \tag{2.18}$$

Clearly here one can consider the system as the interconnection of the following two subsystems:

$$(\Sigma_1) \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u_1 \\ y_1 = x_1 \end{cases} \quad \text{and} \quad (\Sigma_2) \begin{cases} \dot{x}_3 = x_4 + \varphi(v) \\ \dot{x}_4 = u_2 \\ y = x_3 \end{cases}$$

where $v = x_2$ defines the interconnection. It is also clear that (Σ_1) admits an observer (O_1), as well as (Σ_2) admits an observer $(O_2(v))$ if v is considered as a known input for (Σ_2). The idea is to get an observer for the whole system from the interconnection (O_1)+$(O_2(\hat{x}_2))$ where $\hat{x}_2$ is provided by (O_1). It can here be checked that for instance if φ is globally Lipschitz, (O_1)+$(O_2(\hat{x}_2))$ can indeed yield an observer.
Now if (Σ_2) is replaced by:

$$(\Sigma_2') \begin{cases} \dot{x}_3 = \varphi(x_2)x_4 \\ \dot{x}_4 = u_2 \\ y_2 = x_3 \end{cases}$$

it also results from previous section that an observer can be designed for (Σ_2') if x_2 is considered to be a known input, provided that this input is regularly persistent for (Σ_2). If φ is globally Lipschitz, it can again be checked that this is enough for making it possible to get an observer for the whole system by interconnecting sub-observers (*e.g.* as in [9]).
This shows that under appropriate conditions separate possible designs can indeed yield some overall observer. But it does not go that well in any case. Consider for instance the following system:

$$\dot{x}_1 = -\frac{1}{2(t+1)}x_1;\ y_1 = 0 \tag{2.19}$$

$$\dot{x}_2 = -\frac{1}{4(t+1)}x_2 + x_1;\ y_2 = 0 \tag{2.20}$$

It can be seen as an interconnection via x_1 between two subsystems respectively defined by (2.19) and (2.20). Clearly each of them admits an observer (here not tunable) as follows, as long as x_1 is assumed to be known for the second one:

$$\dot{\hat{x}}_1 = -\frac{1}{2(t+1)}\hat{x}_1;\ y_1 = 0 \tag{2.21}$$

$$\dot{\hat{x}}_2 = -\frac{1}{4(t+1)}\hat{x}_2 + x_1;\ y_2 = 0 \tag{2.22}$$

But if we inject $\hat{x}_1$ given by (2.21) into (2.22), one can check that the error equation is not stable.

This just illustrates the fact that in general, the stability of the interconnected observer is not guaranteed by that of each sub-observer, in the same way as separate designs of observer and controller do not in general result in some stable observer-based controller for nonlinear systems (no *separation principle*).
This means that the stability of interconnection of sub-observers requires a specific attention. Conditions can indeed be derived so as to guarantee a possible design by interconnection of separate subdesigns, either in the case of *cascade* interconnection as in the above examples, or even in the case of *full* interconnection [10].

Full Interconnection

Let us first consider the general case of *full* interconnection, via the example of systems made of two subsystems for the sake of illustration, and described by the following representation:

$$(\Sigma)\begin{cases} \dot{x}_1 = f_1(x_1, x_2, u),\ u \subset U \subset I\!R^m;\ f_i\ \mathcal{C}^\infty \text{ function},\ i = 1, 2; \\ \dot{x}_2 = f_2(x_2, x_1, u),\ x_i \in X_i \subset I\!R^{n_i},\ i = 1, 2; \\ y\ \ = (h_1(x_1),\, h_2(x_2))^T = (y_1,\, y_2)^T,\ y_i \in I\!R^{\eta_i},\ i = 1, 2. \end{cases} \tag{2.23}$$

Assume also that $u(.) \in \mathcal{U} \subset \mathcal{L}^\infty(I\!R^+, U)$, and set $\mathcal{X}_i := \mathcal{AC}(I\!R^+, I\!R^{n_i})$ the space of absolutely continuous function from $I\!R^+$ into $I\!R^{n_i}$. Finally, when $i \in \{1, 2\}$, let $\bar{\imath}$ denote its complementary index in $\{1, 2\}$.

The idea here is that System (2.23) can be seen as the interconnection of two subsystems (Σ_i) for $i = 1, 2$ given by:

$$(\Sigma_i) \quad \dot{x}_i = f_i(x_i, v_{\bar{\imath}}, u), \quad y_i = h_i(x_i), \qquad (v_{\bar{\imath}}, u) \in \mathcal{X}_{\bar{\imath}} \times \mathcal{U}. \tag{2.24}$$

Assume that for each system (Σ_i), one can design an observer $(\mathcal{O}_i)$ of the following form:

$$(\mathcal{O}_i) \quad \dot{z}_i = f_i(z_i, v_{\bar{\imath}}, u) + k_i(g_i, z_i)(h_i(z_i) - y_i), \quad \dot{g}_i = G_i(z_i, v_{\bar{\imath}}, u, g_i), \tag{2.25}$$

for smooth k_i, G_i and $(z_i, g_i) \in (I\!R^{n_i} \times \mathcal{G}_i)$, $\mathcal{G}_i$ positively invariant by (2.25). The point is to look for an observer for (2.23) under the form of the following interconnection:

$$(\mathcal{O}) \left\{ \begin{array}{ll} \dot{\hat{x}}_i & = f_i(\hat{x}_i, \hat{x}_{\bar{\imath}}, u) + k_i(\hat{g}_i, \hat{x}_i)(h_i(\hat{x}_i) - y_i);\ i = 1, 2; \\ \dot{\hat{g}}_i & = G_i(\hat{x}_i, \hat{x}_{\bar{\imath}}, u, \hat{g}_i);\ i = 1, 2 \end{array} \right. \tag{2.26}$$

Set $e_i := z_i - x_i$, and for any $u \in \mathcal{U}, v_{\bar{\imath}} \in \mathcal{X}_i$ consider the following system (where $k_i^{v_{\bar{\imath}}}(t)$ denotes gain $k_i(g_i, z_i)$ defined in (2.25)) :

$$\mathcal{E}_i^{(u,v_{\bar{\imath}})} \begin{cases} \dot{e}_i = f_i(z_i, v_{\bar{\imath}}, u) - f_i(z_i - e_i, v_{\bar{\imath}}, u) + k_i^{v_{\bar{\imath}}}(t)(h_i(z_i) - h_i(z_i - e_i)) \\ \dot{z}_i = f_i(z_i, v_{\bar{\imath}}, u) + k_i^{v_{\bar{\imath}}}(t)(h_i(z_i) - h_i(z_i - e_i)) \\ \dot{g}_i = G_i(z_i, v_{\bar{\imath}}, u, g_i). \end{cases}$$

Then sufficient conditions for (2.26) to be an observer for (2.23) have been expressed in [10] as follows:

Theorem 2.10. *[10] If for $i = 1, 2$, any signal $u \in \mathcal{U}, v_{\bar{\imath}} \in \mathcal{AC}(I\!R^+, I\!R^{n_{\bar{\imath}}})$, and any initial value $(z_i^0, g_i^0) \in I\!R^{n_i} \times \mathcal{G}_i$, $\exists V_i(t, e_i), W_i(e_i)$ positive definite functions such that:*

(i) $\forall x_i \in X_i; \forall e_i \in I\!R^{n_i}; \forall t \geq 0,$

$$\frac{\partial V_i}{\partial t}(t, e_i) + \frac{\partial V_i}{\partial e_i}(t, e_i)[f_i(x_i + e_i, v_{\bar{\imath}}(t), u(t)) - f_i(x_i, v_{\bar{\imath}}(t), u(t)) \\ + k_i^{v_{\bar{\imath}}}(t)(h_i(x_i + e_i) - h_i(x_i))] \leq -W_i(e_i)$$

(ii)$\exists \alpha_i > 0; \forall x_i \in X_i; \forall x_{\bar{\imath}} \in I\!R^{n_{\bar{\imath}}}; \forall e_i \in I\!R^{n_i}; \forall e_{\bar{\imath}} \in I\!R^{n_{\bar{\imath}}}; \forall t \geq 0,$

$$\left\| \frac{\partial V_i}{\partial e_i}(t, e_i)[f_i(x_i, x_{\bar{\imath}} + e_{\bar{\imath}}, u(t)) - f_i(x_i, x_{\bar{\imath}}, u(t))] \right\| \leq \alpha_i \sqrt{W_i(e_i)} \sqrt{W_{\bar{\imath}}(e_{\bar{\imath}})},$$

(iii)$\alpha_1 + \alpha_2 < 2,$

then (2.26) is an asymptotic observer for (2.23).

●

This can be established on the basis of Lyapunov arguments.

Cascade Interconnection

In the *weaker* case of *cascade* interconnection, namely when $f_1(x_1, x_2, u) = f_1(x_1, u)$ in (2.23), various results have been proposed for the stability of the interconnected system. Let us report here the weakened assumptions proposed in [10] in this context of observer design:

Theorem 2.11. *Assume that:*

I. System $\dot{x}_1 = f_1(x_1, u)$; $y_1 = h_1(x_1)$ *admits an observer* $(\mathcal{O}_1)$ *as in (2.25) (without* v_2*), s.t.* $\forall u \in \mathcal{U}$ *and* $\forall x_1(t)$ *admissible trajectory of the system associated to* u*:*

$$\lim_{t\to\infty} e_1(t) = 0 \text{ and } \int_0^{+\infty} \|e_1(t)\|dt < +\infty \quad (\text{with } e_1 := z_1 - x_1); \tag{2.27}$$

II. $\exists c > 0;\ \forall u \in U;\ \forall x_2 \in X_2, \|f_2(x_2, x_1, u) - f_2(x_2, x_1', u)\| \leq c\|x_1 - x_1'\|$*;*

III.$\forall u \in \mathcal{U}, \forall v_1 \in \mathcal{AC}(I\!R^+, I\!R^{n_1})$*,* $\forall z_2^0, g_2^0$*,* $\exists v(t, e_2), w(e_2)$ *positive definite functions s.t for every trajectory of* $\mathcal{E}_2^{(u,v_1)}$ *with* $z_2(0) = z_2^0, g_2(0) = g_2^0$*:*

(i) $\forall x_2 \in X_2, e_2 \in I\!R^{n_2}, t \geq 0,$

$$\frac{\partial v}{\partial t}(t, e_2) + \frac{\partial v}{\partial e_2}(t, e_2)[f_2(x_2 + e_2, v_1(t), u(t)) - f_2(x_2, v_1(t), u(t)) \\ + k_2^{v_1}(t)(h_2(x_2 + e_2) - h_2(x_2))] \leq -w(e_2)$$

(ii)$\forall e_2 \in I\!R^{n_2}, t \geq 0;\ v(t, e_2) \geq \bar{w}(e_2)$

(iii)$\forall e_2 \in I\!R^{n_2} \backslash \mathcal{B}(0, r), t \geq 0;\ \left\| \frac{\partial v}{\partial e_2}(t, e_2(t)) \right\| \leq \lambda(1 + v(t, e_2(t)))$ *for some constants* $\lambda, r > 0$ *and* $\mathcal{B}(0, r) := \{e_2 : \|e_2\| \leq r\}$*.*

Then:

$$\begin{aligned} \dot{\hat{x}}_1 &= f_1(\hat{x}_1, u) + k_1(\hat{g}_1, \hat{x}_1)(h_1(\hat{x}_1) - h_1(x_1)) \\ \dot{\hat{x}}_2 &= f_2(\hat{x}_1, \hat{x}_2, u) + k_2(\hat{g}_2, \hat{x}_2)(h_2(\hat{x}_1) - h_2(x_1)) \\ \dot{\hat{g}}_1 &= G_1(\hat{x}_1, u, \hat{g}_1); \quad \dot{\hat{g}}_2 = G_2(\hat{x}_2, \hat{x}_1, u, \hat{g}_2). \end{aligned} \tag{2.28}$$

is an observer for (2.23) where $f_1(x_1, x_2, u) = f_1(x_1, u)$*.*

•

In view of these conditions, and using available observers for systems in some particular forms, one might be able to design observers for further nonlinear systems (see [9] for examples of cascade interconnection, or [10, 5] for examples of full interconnections).

2.4.2 Transformation-based Design

Principle

The observer designs presented till now are still all based on particular structures of the system (either isolated or interconnected). The subsequent idea is that these designs can also give state observers for systems which can be turned into one of these forms by an appropriate transformation. The most common approach in that respect is to consider changes of state coordinates. Such a relationship defines some *system equivalence*:

Definition 2.14. *System equivalence [resp. at x_0].*
A system described by:

$$\left\{\begin{array}{l} \dot{x} = f(x,u) = f_u(x)\, x \in I\!R^n, u \in I\!R^m \\ y = h(x) \in I\!R^p \end{array}\right. \tag{2.29}$$

will be said to be **equivalent** *[resp.* **at x_0***] to the system:*

$$\left\{\begin{array}{l} \dot{z} = F(z,u) = F_u(z) \\ y = H(z) \end{array}\right. \tag{2.30}$$

if there exists a diffeomorphism $z = \Phi(x)$ defined on $\mathbf{R}^n$ [resp. some neighbourhood of x_0] such that:

$$\forall u \in I\!R^m, \quad \frac{\partial \Phi}{\partial x} f_u(x) \mid_{x=\Phi^{-1}(z)} = F_u(z) \qquad \text{and} \qquad h \circ \Phi^{-1} = H.$$

Systems (2.29) and (2.30) are then said to be equivalent by $z = \Phi(x)$.

The interest of such a property for observer design can then be illustrated by the following proposition (*e.g.* as in [5]):

Proposition 2.3. *Given two systems (Σ_1) and (Σ_2) respectively defined by:*
$(\Sigma_1) \left\{\begin{array}{l} \dot{x} = X(x,u) \\ y = h(x) \end{array}\right.$ *and* $(\Sigma_2) \left\{\begin{array}{l} \dot{z} = Z(z,u) \\ y = H(z) \end{array}\right.$
and equivalent by $z = \Phi(x)$,

If:

$$(\mathcal{O}_2) \left\{\begin{array}{l} \dot{\hat{z}} = Z(\hat{z},u) + k(w, H(\hat{z}) - y)) \\ \dot{w} = F(w,u,y) \end{array}\right.$$

is an observer for (Σ_2),

Then:

$$(\mathcal{O}_2) \left\{\begin{array}{l} \dot{\hat{x}} = X(\hat{x},u) + \left(\dfrac{\partial \Phi}{\partial x}\right)^{-1}_{|\hat{x}} k(w, h(\hat{x}) - y) \\ \dot{w} = F(w,u,y) \end{array}\right.$$

is an observer for (Σ_1).

From this indeed, if a system is not of an appropriate structure for an observer design in view of previous sections, but is equivalent to some other system which does have some appropriate structure, then the observer problem can be solved for the original system.

Examples

The idea of Proposition 2.3 has motivated various works on characterizing systems which can be turned into some appropriate structures for observer design, from the linear one up to output injection [32, 11, 33] to several forms of cascade block state affine systems up to nonlinear injections from block to block as in (2.31) below for instance [25, 38, 9, 26, 7, 8, ...].

$$\begin{cases} \dot{z}_1 = A_1(u,y^1)z_1 + \varphi_1(u,y^1) \\ \dot{z}_2 = A_2(u,y^2,z_1)z_2 + \varphi_2(u,y^2,z_1) \\ \quad \vdots \\ \dot{z}_q = A_q(u,y^q,z_1,\ldots z_{q-1})z_q + \varphi_q(u,y^q,\ldots z_{q-1}) \\ y \;= \begin{pmatrix} C_1 z_1 \\ \vdots \\ C_q z_q \end{pmatrix} = \begin{pmatrix} y^1 \\ \vdots \\ y^q \end{pmatrix} \\ \qquad u \in I\!R^m, z_i \in I\!R^{n_i}, y^i \in I\!R^{\nu_i}, \end{cases} \tag{2.31}$$

As a simple illustrative example, let us consider here the problem of turning a nonlinear system:

$$\begin{array}{l} \dot{x} = f(x) \\ y = h(x),\;\; x \in \mathbf{R}^n \end{array}$$

into a linear observable form up to output injection as follows:

$$\begin{array}{l} \dot{x} = Ax + \varphi(Cx) \\ y = Cx \end{array}$$

Necessary and sufficient conditions for this problem to be solvable have been given in terms of differential geometry in [32].
A constructive algorithm to simultaneously check the possibility of the transformation and construct φ can alternatively be given in the spirit of [23] as follows:

1. Get the representation:

   $$y^{(n)} = \Phi(y, \dot{y}, \ldots y^{(n-1)})$$

   and set $z_1 := y$.
2. For $i \geq 1$, define φ_i by: $\frac{\partial \varphi_i}{\partial y} = \frac{\partial z_i^{(n-i+1)}}{\partial y^{(n-i)}}$;
   If φ_i is not only a function of y, the transformation fails and the procedure ends. Else, set: $z_{i+1} := \dot{z}_i - \varphi_i$
3. Continue until $i = n$ or the procedure aborts.

The procedure is clearly sufficient, and it can be checked that it is indeed necessary.

As a second simple example, turning some $n-$dimensional nonlinear control affine system into the appropriate structure for high gain observer design, if possible, is obtained by the following transformation [18, 19]:

$$z = \begin{pmatrix} h(x) \\ L_f h(x) \\ \vdots \\ L_f^{n-1} h(x) \end{pmatrix}$$

Finally, it can be underlined that some enlargement of the class of systems admitting an observer on the basis of the particular structures highlighted in the above presentation can also be obtained by further considering output transformations (*e.g.* as in [33, 23, 4]), or state extension (as in immersions [17, 35, 30, 6]), for instance.

2.5 Conclusion

The purpose in this chapter was to give some overview on techniques of observer design for nonlinear systems. This presentation clearly follows a particular viewpoint on the problem, and does not claim to be exhaustive. In particular notions of observability have been recalled, and some observers have been presented according to two types of designs: uniform and non uniform ones w.r.t. input (or time). Those designs are in particular driven by specific structures of systems, and admit smooth explicit gains. Extensions of such designs to more general structures by interconnexions and transformations have also been discussed. But further comments on *detectability* and related designs have for instance been omitted, as well as various other technical approaches where the design is not necessarily smooth (as in sliding modes [16, 2, ...]) or explicit (as in optimization-based designs [37, 1, ...]).

References

1. M. Alamir (1999), Optimization-based nonlinear observers revisited. *Int. Journal of Control*, 72(13):1204–1217.
2. J.P. Barbot, T. Boukhobza, and M. Djemai (1996), Sliding mode observer for triangular input form. In *Proc. 35th IEEE Conf. on Decision and Control, Kobe, Japan*, pages 1489–90.
3. G. Besançon (1999), Further results on high gain observers for nonlinear systems. In *Proc. 38th IEEE Conf. on Decision and Control, Phoenix, USA*.
4. G. Besançon (1999), On output transformations for state linearization up to output injections of nonlinear systems. *IEEE Trans. on Automatic Control*, 44(11):1975–1981.
5. G. Besançon (1999), A viewpoint on observability and observer design for nonlinear systems. In *New Directions in Nonlinear Observer Design - Lecture Notes in Control and Info. Sciences 244*, pages 2–21. H. Nijmeijer and T. Fossen Eds. Springer-Verlag, London.

6. G. Besançon (2005), Immersion-based observer design for rank-observable nonlinear systems. Technical report, LAG. March.
7. G. Besançon and G. Bornard (1996), State equivalence based observer design for nonlinear control systems. In *Proc. of IFAC World Congress, San Francisco, CA, USA*, pages 287–292.
8. G. Besançon and G. Bornard (1997), On charactering classes of observer forms for nonlinear systems. In *Proc. 4th European Control Conf., Brussels, Belgium.*
9. G. Besançon, G. Bornard, and H. Hammouri (1996), Observer synthesis for a class of nonlinear control systems. *European J. of Control*, 3(1):176–193.
10. G. Besançon and H. Hammouri (1998), On observer design for interconnected systems. *Journal of Math. Systems, Estimation and Control*, 8(3).
11. D. Bestle and M. Zeitz (1983), Canonical form observer design for nonlinear time-variable systems. *Int. Journal of Control*, 38(2):419–431.
12. G. Bornard, F. Celle-Couenne, and G. Gilles (1995), Observability and observers. In *Nonlinear Systems - T.1, Modeling and Estimation*, pages 173–216. Chapman & Hall, London.
13. G. Bornard, N. Couenne, and F. Celle (1988), Regularly persistent observer for bilinear systems. In *New Trends in Nonlinear Control Theory; Lecture Notes in Control and Info. Sci., 122*, pages 130–140. Springer-Verlag, Berlin.
14. G. Bornard and H. Hammouri (1991), A high gain observer for a class of uniformly observable systems. In *Proc. 30th IEEE Conf. on Decision and Control, Brighton, England*, pages 1494–1496.
15. F. Deza, D. Bossanne, E. Busvelle, J.P. Gauthier, and D. Rakotopara (1993), Exponential observers for nonlinear systems. *IEEE Trans. on Automatic Control*, 38(3):482–484.
16. S. Drakunov and V. Utkin (1995), Sliding mode observers. tutorial. In *Proc. 34th IEEE Conf. on Decision and Control, New Orleans, LA, USA*, pages 3376–8.
17. M. Fliess and Kupka (1983), A finiteness criterion for nonlinear input-output differential systems. *Siam Journal on Control and Optimization*, 21(5):721–728.
18. J.P. Gauthier and G. Bornard (1981), Observability for any $u(t)$ of a class of nonlinear systems. *IEEE Trans. on Automatic Control*, 26(4):922–926.
19. J.P. Gauthier, H. Hammouri, and S. Othman (1992), A simple observer for nonlinear systems - applications to bioreactors. *IEEE Trans. on Automatic Control*, 37(6):875–880.
20. J.P. Gauthier and A.K. Kupka (1994), Observability and observers for nonlinear systems. *Siam Journal on Control and Optimization*, 32(4):975–994.
21. J.P. Gauthier and A.K. Kupka (1997), *Deterministic observation theory and applications.* Cambridge Univ Press.
22. A. Gelb (1992), *Applied optimal estimation.* MIT Press, Cambridge.
23. A. Glumineau, C. Moog, and F. Plestan (1996), New algebraic-geometric conditions for the linearization by input-output injection. *IEEE Trans. on Automatic Control*, 41(4):598–603.
24. H. Hammouri and F. Celle (1991), Some results about nonlinear systems equivalence for the observer synthesis. In *New Trends in Systems Theory*, pages 332–339. Birkhäuser.
25. H. Hammouri and J.P. Gauthier (1988), Bilinearization up to output injection. *Systems & Control Letters*, 11:139–149.
26. H. Hammouri and M. Kinnaert (1996), A new formulation for time-varying linearization up to output injection. *Systems & Control Letters*, 28:151–157.

27. H. Hammouri and J. De Leon Morales (1990), Observer synthesis for state-affine systems. In *Proc. 29th IEEE Conf. on Decision and Control, Honolulu, HA, UA*, pages 784–785.
28. R. Hermann and A.J. Krener (1977), Nonlinear controllability and observability. *IEEE Trans. on Automatic Control*, 22(5):728–740.
29. A. Isidori (1995), *Nonlinear Control Systems.* Springer-Verlag, Berlin, 3^{rd} edition.
30. P. Jouan (2003), Immersion of nonlinear systems into linear systems modulo output injection. *Siam Journal on Control and Optimization*, 41:1756–78.
31. R.E. Kalman and R.S. Bucy (1961), New results in linear filtering and prediction theory. 83:95–108.
32. A. J. Krener and A. Isidori (1983), Linearization by output injection and nonlinear observers. *Systems & Control Letters*, 3:47–52.
33. A. J. Krener and W. Respondek (1985), Nonlinear observers with linearizable error dynamics. *Siam Journal on Control and Optimization*, 23(2):197–216.
34. H. Kwakernaak and R. Sivan (1972), *Linear Optimal Control Systems.* Wiley-Interscience, New York.
35. J. Lévine and R. marino (1986), Nonlinear systems immersion, observers and finite dimensional filters. *Systems & Control Letters*, 7:137–142.
36. D.G. Luenberger (1966), Observers for multivariable systems. *IEEE Trans. on Automatic Control*, 11(2):190–197.
37. H. Michalska and D.Q. Mayne (1995), Moving horizon observers and observer-based control. *IEEE Trans. on Automatic Control*, 40(6):995–1006.
38. J. Rudolph and M. Zeitz (1994), A block triangular nonlinear observer normal form. *Systems & Control Letters*, 23:1–8.
39. H.J. Sussmann (1979), Single input observability of continuous time systems. 12:371–393.
40. G. Zimmer (1994), State observation by on-line minimization. *Int. Journal of Control*, 60(4):595–606.

3

Sampled-data Control of Nonlinear Systems

Dina Shona Laila[1], Dragan Nešić[2], and Alessandro Astolfi[1]

[1] Department of Electrical and Electronic Engineering, Imperial College, Exhibition Road, London SW7 2AZ, UK. Email: d.laila@imperial.ac.uk, a.astolfi@imperial.ac.uk

[2] Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3001, Australia. Email: d.nesic@ee.unimelb.edu.au

Summary. This chapter provides some of the main ideas resulting from recent developments in sampled-data control of nonlinear systems. We have tried to bring the basic parts of the new developments within the comfortable grasp of graduate students. Instead of presenting the more general results that are available in the literature, we opted to present their less general versions that are easier to understand and whose proofs are easier to follow. We note that some of the proofs we present have not appeared in the literature in this simplified form. Hence, we believe that this chapter will serve as an important reference for students and researchers that are willing to learn about this area of research.

3.1 Introduction

Technological advances in digital electronics that occurred in the second half of the 20th century have led to a rapid development in computer technology and this has made a great impact on a range of engineering areas, including control engineering. Nowadays, most control systems exploit a digital computer as their crucial part and computer controlled systems are prevalent

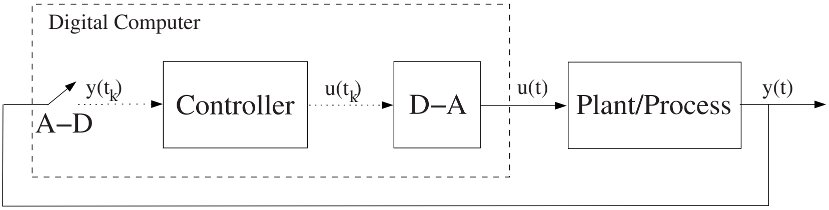

Fig. 3.1. General computer controlled system configuration

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 91–137, 2006.

in engineering practice. Hence, the theory for analysis and design of computer controlled systems is a crucial part of the control engineer's toolbox.

A general configuration of a computer controlled feedback system is illustrated in Figure 3.1. A continuous-time plant (process) is interfaced with the computer via analog-to-digital (A/D) and digital-to-analog (D/A) converters that are often referred to as *sampler* and *hold* devices respectively. The A/D converter produces the samples $y(t_k)$ of the continuous plant output $y(t)$ at sampling times t_k and sends them to a control algorithm within the computer. The control algorithm processes the measured sequence $y(t_k)$ and produces a sequence of control inputs $u(t_k)$. This control sequence is converted in the D/A converter into a piecewise continuous control signal $u(t)$ that is applied to the plant. This is typically done by holding the value of the control signal constant during the sampling intervals (*zero-order-hold*). An internal clock synchronizes the operation of the system. The sampling instants t_k are typically equidistant, *i.e.* $t_k = kT$, $k = 0, 1, 2, \ldots$, where $T > 0$ is the *sampling period*.

The computer controlled system in Figure 3.1 is often referred to as a *sampled-data control system* to emphasize the sampling process as its crucial feature. Note that due to the hybrid nature of sampled-data systems, that involve continuous-time (plant) dynamics and discrete-time (controller) dynamics, their analysis and design are harder than those of continuous-time systems[3]. Indeed, this has led to several distinct approaches to controller design for sampled-data systems.

1. *Emulation:* Design a continuous-time controller for the continuous-time plant model and then discretize the controller for digital implementation. This approach involves an approximation (discretization) of the controller that is valid only for small sampling periods T and, typically, the system loses stability for large sampling periods. Advanced emulation techniques also use controller redesign for digital implementation and they are better behaved for larger sampling periods.
2. *Discrete-time design:* Design a controller in discrete-time using the discrete-time plant model. This method exploits an approximation (discretization) of the plant model that ignores the inter-sample behaviour. While this method does not require fast sampling to maintain stability, performance of the sampled-data system is not automatically guaranteed since the inter-sample behaviour may be unacceptable.
3. *Sampled-data design:* Using an exact sampled-data model of the plant[4], design a controller that achieves both stability and required performance

[3] This is the case, for instance, when the plant is a continuous-time system and the controller is realized via analog electronics using operational amplifiers.

[4] See for instance, [13] where the lifting technique is used to obtain models for sampled-data systems that model the inter-sample behaviour.

for the sampled-data system. This method uses no approximations of the plant model or controller and, hence, it maintains stability and performance for arbitrarily large sampling periods T.

Emulation is regarded as the simplest method, while sampled-data design requires the most advanced techniques. On the other hand, satisfactory system performance can be achieved using the sampled-data design, whereas emulation is typically inferior to the other two methods in terms of stability and/or achievable performance.

Analysis and design of *linear* sampled-data control systems date back to the 1950s, that marked the beginning of the digital revolution. The early works concentrated on input-output approaches involving z-transform and they were parallel to the corresponding continuous-time developments. In the 1960s and 1970s, state space approaches involving state difference equations have become popular and optimal regulation and Kalman filtering for discrete-time systems were developed during that time. This material has become a standard part of many undergraduate curricula. The 1980s and 1990s have seen several new developments for linear systems that have led to H_∞ theory for discrete-time systems, advanced emulation techniques based on optimization, the use of δ-transform and H_∞ sampled-data controller design based on lifting techniques (for more details on all of these developments see [6, 13, 25, 32]).

Linear sampled-data control theory is now a mature area with a range of undergraduate textbooks that cover different analysis and design approaches. On the other hand, *nonlinear* sampled-data control theory is quite underdeveloped compared to its linear counterpart. While it is often possible to use a linear sampled-data control theory for solving nonlinear control problems via the linearization technique, there are many important situations where nonlinearities cannot be neglected. For instance, wide ranges of operating conditions typically prevent control designers from ignoring important nonlinearities, such as saturation, that are commonly present in the system. Moreover, hysteresis, dead-zone and dry friction are but a few examples of common nonlinearities that often can not be ignored in practice (see [52] for details). Indeed, there is a wide area of applications where nonlinear phenomena cannot be avoided. These applications range from vertical take-off and landing (VTOL) aircraft systems, ship or submarine vehicle control, position control for robotic systems in a precision manufacturing process, autonomous vehicle systems, biochemical reactors, power plants and many others. Finally, many control algorithms, such as adaptive and sliding mode controllers, are inherently nonlinear. Therefore, nonlinear sampled-data control systems form an important class of systems that arises in applications. Emulation for nonlinear sampled-data systems has been studied in some details and general results that provide a justification for this approach are available (see [29] and references cited therein). Due to a variety of tools for nonlinear continuous-time controller design (see for instance [23, 24, 53, 52]) and its inherent simplicity, the emulation method is quite attractive to practitioners. Unfortunately, emulated controllers are prone to instability in nonlinear systems. As a result, one typically needs to use smaller sampling periods in emulation design for nonlinear systems. In particular, the required sampling may sometimes exceed the hardware limitations and in such cases one may need to use methods other than emulation.

On the other hand, due to the complexity of the underlying nonlinear sampled-data model, results on sampled-data design for nonlinear systems that would parallel the linear results presented in [13] are scarce (we are not aware of any) and it appears that they will be hard to develop in the future. Hence, it appears that discrete-time design techniques for nonlinear sampled-data systems provide a nice tradeoff between the possible conservatism of emulation design and the difficulty of developing direct sampled-data design.

The literature on discrete-time design methods for nonlinear sampled-data systems can be classified into two large groups:

1. *Exact discrete-time design methods.* The majority of results in this direction, for example [2, 16, 21, 33, 34, 56, 60], assume that the *exact discrete-time plant model is known* and it is available to the designer. Hence, these papers start directly from discrete-time models of the form:

$$x(k+1) = F(x(k), u(k)) \ ,$$

   where $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$ are respectively the state and the control input of the system and $F(\cdot,\cdot)$ is a known vector function. This assumption, however, is rarely justified for nonlinear sampled-data systems (such as the one illustrated by Figure 3.1) as will be discussed in Section 3.2 and, hence, results that belong to this group have very limited applicability.

2. *Approximate discrete-time design methods.* Some earlier research, for instance [15, 17, 19, 31], recognize the fact that the exact discrete-time model for nonlinear systems is typically unavailable to the controller designer and they instead base their controller design on an *approximate discrete-time plant model.* While this approach is closer to reality and it is most natural to use in practice, due to the limited theoretical results, the majority of the published works in this area are *ad hoc* and they do not carefully investigate the interplay between the controller design and the plant model approximation. In particular, we show in Section 3.4 that there may exist controllers that stabilize a seemingly good approximate discrete-time plant model but destabilize the sampled-data system for arbitrarily small sampling periods. Hence, great care is needed when pursuing this approach.

The main purpose of this chapter is to provide a rigorous framework for sampled-data nonlinear controller design via approximate discrete-time plant models. Our framework is fully consistent with what most engineers would do in practice but our analysis provides a framework and guidelines for such design to be successful. Moreover, this framework can be used to justify the emulation method for general nonlinear systems (see Section 3.7.1). Several controller design techniques are presented for classes of nonlinear systems that are fully consistent with our framework. Our approach benefits from selected topics in numerical analysis literature [51, 59]. In particular, we adapted the notion of *consistency*, commonly used in numerical analysis, to develop our controller design framework.

We emphasize that this chapter is not intended to serve as a literature survey and the material presented summarizes just a subset of recent results in nonlinear sampled-data control that reflect the authors research interests. Moreover, we emphasize that our results are often presented in a simpler form than that in the original references in order to achieve clarity and simplicity of exposition. We have tried to achieve this without sacrificing the rigor of our arguments. More complete and detailed results in this area and the closely related works are listed in the references.

3.2 Mathematical Preliminaries

A function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{K}$ if it is continuous, zero at zero and strictly increasing, and of class $\mathcal{K}_\infty$ if it is of class $\mathcal{K}$ and unbounded. Note that linear functions $\varphi(s) = Ks$ for some $K > 0$ are of class $\mathcal{K}_\infty$. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{KL}$ if $\beta(\cdot, \tau)$ is of class $\mathcal{K}$ for each $\tau \geq 0$ and $\beta(s, \cdot)$ is decreasing to zero for each $s > 0$. The function β is of class exp-$\mathcal{KL}$ if there exist $K, \lambda > 0$ such that $\beta(s,t) = Ks \exp(-\lambda t)$. Class $\mathcal{K}$ and $\mathcal{KL}$ functions are useful to characterize stability properties of nonlinear systems [23]. For instance, suppose that there exists $\beta \in \mathcal{KL}$ such that the solutions $\phi(t, x_\circ)$ of the continuous-time system $\dot{x} = f(x)$ satisfy

$$|\phi(t,x_\circ)| \leq \beta(|x_\circ|, t) \qquad \forall t \geq 0,\ x(0) = x_\circ \in \mathbb{R}^n\ .$$

Then, the origin of the system is *globally asymptotically stable (GAS)*. Moreover, if $\beta \in$ exp-$\mathcal{KL}$, then the origin of the system is *globally exponentially stable (GES)*.

A function $f : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is of order $O(T^p), p > 0$, if there exist $\varphi \in \mathcal{K}_\infty$ and $T^* > 0$ such that for all $T \in (0, T^*)$ and all $x \in \mathbb{R}^n$ we have $|f(x,T)| \leq \varphi(|x|)T^p$. We will use the Mean Value Theorem several times in the sequel and we state it below for the sake of completeness.

If x and y are two distinct points in $\mathbb{R}^n$, then the (open) line segment $L(x,y)$ joining two distinct points x and y in $\mathbb{R}^n$ is

$$L(x,y) = \{z | z = \theta x + (1-\theta)y,\ 0 < \theta < 1\}\ .$$

Theorem 3.1 (Mean Value Theorem). *Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable at each point x of an open set $S \subset \mathbb{R}^n$. Let x and y be two points of S such that the line segment $L(x, y) \subset S$. Then there exists a point z of $L(x, y)$ such that*

$$f(y) - f(x) = \left.\frac{\partial f}{\partial x}\right|_{x=z} (y - x) \ .$$

■

3.3 Zero-order-hold Equivalent Models

In this section we present results on discretization of sampled-data systems assuming the use of zero-order-hold devices. These results provide a basis for the controller design framework via approximate discrete-time models presented in the next section. Consider the sampled-data system in Figure 3.1 where we assume that the plant dynamics are linear, *i.e.*

$$\dot{x} = Ax + Bu \ , \tag{3.1}$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ are the state and control vectors respectively. The plant is assumed to be between a sampler (A/D converter) and zero-order-hold (D/A converter). The control signal is assumed to be piecewise constant, *i.e.*

$$u(t) = u(kT) =: u(k), \qquad \forall t \in [kT, (k+1)T), \ k \in \mathbb{N} \tag{3.2}$$

where $T > 0$ is the sampling period. Moreover, we assume that the state measurements $x(k)$, where[5]

$$x(k) := x(kT) \ . \tag{3.3}$$

are available at sampling instants.

A classical approach to controller design for the system (3.1) is to first discretize the model and then design a controller for the discretized model. Using the variations of constant formula for the linear system (3.1) we can compute the solution x at time $t \geq kT$ that starts from the initial state $x(k)$ at time kT, while keeping the control constant $u(t) \equiv u(k)$,

$$x(t) = e^{A(t-kT)} x(k) + \int_{kT}^{t} e^{A(t-s)} Bu(k) ds \ .$$

Evaluating the above equation for $t = (k+1)T$, we have

[5] One can also assume that only outputs $y(k) = Cx(k) + Du(k)$ are measured but in this section we want to keep the presentation as simple as possible and do not consider this case.

$$x(k+1) = \Phi_T x(k) + \Gamma_T u(k) \ , \tag{3.4}$$

where

$$\Phi_T := e^{AT}; \qquad \Gamma_T := \int_0^T e^{As} B ds \ .$$

The discretized model (3.4) describes the sampled-data system (3.1), (3.2), (3.3) *exactly* at sampling instants kT and, in particular, it describes how the state $x(k+1)$ of the system at the time instant $(k+1)T$ depends on the state $x(k)$ at the previous sampling instant kT and control $u(k)$ on the sampling interval $[kT, (k+1)T)$.

Note that the model (3.4) is a linear difference equation that is parameterized by the sampling period T. The sampling period T is assumed to be a design parameter which can be arbitrarily assigned. In practice, there is a range of allowable sampling periods T that depends on the hardware limitations (*e.g.* the DAQ 2000 I/O card can achieve any sampling periods from 0.01 seconds to 30 minutes). Note that the discrete-time model ignores the inter-sample behaviour and any controller that is designed using this model may lead to poor inter-sample behaviour.

Consider now the nonlinear continuous-time control system

$$\dot{x} = f(x, u) \ , \qquad x(0) = x_\circ \ . \tag{3.5}$$

The function f is assumed to be such that, for each initial condition and each constant control, there exists a unique solution defined on some (perhaps bounded) interval of the form $[0, \tau)$. We can compute the solution x at time $t \geq kT$ that starts from the initial state $x(k)$ while keeping the control constant $u(t) \equiv u(k)$ as

$$x(t) = x(k) + \int_{kT}^{t} f(x(s), u(k)) ds \ .$$

Suppose that the solutions are well defined and evaluate the above equations for $t = (k+1)T$,

$$x(k+1) = x(k) + \int_{kT}^{(k+1)T} f(x(s), u(k)) ds =: F_T^e(x(k), u(k)) \ . \tag{3.6}$$

The equation (3.6) represents the exact discrete-time model of the nonlinear sampled-data system (3.5), (3.2), (3.3) and it is the nonlinear counterpart of (3.4). We emphasize that F_T^e is not known in most cases since computing F_T^e explicitly will require an analytic solution of a nonlinear initial value problem. On the other hand, one can easily write down a range of approximate models. For example, the forward Euler approximate model of the sampled-data system (3.5), (3.2), (3.3),

$$x(k+1) = x(k) + T f(x(k), u(k)) =: F_T^{Euler}(x(k), u(k)) \ , \tag{3.7}$$

is often used in the sequel. A range of other approximate models (*e.g.* using Runge-Kutta integration methods) can be found in standard books on numerical analysis [59].

In the sequel, we consider the difference equations corresponding to the exact and approximate discrete-time models of the sampled data system (3.5), (3.2), (3.3) that are denoted respectively as

$$x(k+1) = F_T^e(x(k), u(k)) \tag{3.8}$$

$$x(k+1) = F_T^a(x(k), u(k)) \tag{3.9}$$

and which are parameterized by the sampling period T. We will think of F_T^e and F_T^a as being defined globally for all small T even though the initial value problem (3.5) may exhibit finite escape times. In general, one needs to use small sampling periods T since the approximate plant model is a good approximation of the exact model mainly only for small T.

It turns out that most sampled-data literature [2, 16, 21, 33, 34, 56, 60] uses the following assumption.

Assumption 1 *The exact discrete-time model (3.8) for the sampled-data system (3.5), (3.2), (3.3) is known and it is available to the designer. In other words, the controller design can be carried out using the exact discrete-time model (3.8).*

Indeed, this assumption is the starting point in the exact discrete-time design method discussed in the Introduction. On the other hand, Assumption 1 is not justified in most cases. The exact discrete-time model can not be analytically computed since it requires solving a nonlinear initial value problem explicitly. Hence, our results are useful in cases when the following more realistic assumption holds.

Assumption 2 *The exact discrete-time model (3.8) for the sampled-data system (3.5), (3.2), (3.3) is not known exactly and it is not available to the designer. Therefore, the controller design needs to be carried out using an approximate discrete-time model (3.9).*

We note that Assumption 2 is more natural to use for most nonlinear systems. Moreover, even in the linear case we use approximate models that come from numerically computing the matrices Φ_T and Γ_T in (3.4).

3.4 Motivating Counter-examples

The approximate model (3.9) is parameterized by T and, in general, we need to be able to obtain a family of controllers which is parameterized by T and is defined for all small T. There are two reasons for this: (i) F_T^a is a good approximation for F_T^e only for small T and, hence, the designed controller will have to achieve stability of F_T^a for all small T; (ii) finding a controller that does not depend on T and that stabilizes the approximate family F_T^a for all small T is a harder problem than when the controller is allowed to depend on T. Hence, we will concentrate in the sequel on controllers of the form[6]

$$u(k) = u_T(x(k)) \ . \tag{3.10}$$

The goal of this section is to show that there may exist a family of controllers of the form (3.10) that stabilizes the family of approximate models (3.9) for all small T whereas it *destabilizes* the family of exact models (3.8) for all small T. We identify several indicators of lack of stability robustness that typically lead to these undesirable behaviours. In the next section, we will introduce conditions that rule out each of these non-robustness indicators and this will lead to a framework for sampled-data controller design via approximate discrete-time models.

Examples in this section can be interpreted in the following manner. Assume that we want to pursue an *ad hoc* approach to controller design that many practitioners and researchers have considered. Consider an approximate plant model (3.9), such as (forward) Euler model, that is a good approximation for (3.8) when the two models are regarded as "open-loop". Suppose, moreover, that we want to first reduce T sufficiently to guarantee that F_T^a is a good approximation of F_T^e and then we design a controller (3.10) that stabilizes F_T^a, hoping that it will stabilize F_T^e because T is already small enough. Examples presented in this section show that in general this approach is flawed and no matter how small sampling period T we choose, we can always find a controller (3.10) that stabilizes the approximate model (3.9) but it destabilizes the exact system (3.8). The following examples (taken from [43]) illustrate that a careful investigation is needed if controller design is to be carried out on approximate models.

Example 3.1. **(Control with excessive force)** Consider the sampled-data control of the triple integrator

$$\dot{x}_1 = x_2; \qquad \dot{x}_2 = x_3; \qquad \dot{x}_3 = u \ . \tag{3.11}$$

Although the exact discrete-time model of this system can be computed, we base our control algorithm on the family of the Euler approximate discrete-time models in order to illustrate possible pitfalls in control design based on approximate discrete-time models. The family of Euler approximate discrete-time models for this system is given by (3.7). A minimum time dead beat controller for the Euler discrete-time model is given by

[6] For simplicity we consider only static state feedback controllers, while results on dynamic controllers can be found in the cited references.

$$u = u_T(x) = \left(-\frac{x_1}{T^3} - \frac{3x_2}{T^2} - \frac{3x_3}{T} \right) . \tag{3.12}$$

The closed-loop system (3.7), (3.12) has all eigenvalues equal to zero for all $T > 0$ and hence this discrete-time Euler based closed-loop system is asymptotically stable for all $T > 0$. On the other hand, the closed-loop system consisting of the exact discrete-time model of the triple integrator and controller (3.12) has an eigenvalue at ≈ -2.644 for all $T > 0$. Hence, the closed-loop sampled-data control system is unstable for all $T > 0$.

Note that in Example 3.1 we have the following properties.

1. **Nonuniform bound on overshoot**. The solutions of the family of approximate models with the given controller satisfy for all $T > 0$ a stability estimate of the type

$$|\phi_T(k, x_\circ)| \le b_T e^{-kT} |x_\circ| , \quad k \in \mathbb{N}$$

   and $b_T \to \infty$ as $T \to 0$. Hence, the overshoot in the stability estimate for the family of approximate models is not uniformly bounded in T.
2. **Nonuniform bound on control**. The control is not uniformly bounded on compact sets with respect to the parameter T and in particular we have for all $x \neq 0$ that $|u_T(x)| \to \infty$ as $T \to 0$.

Example 3.2. **(Control with excessive finesse)** Consider the system

$$\dot{x} = x + u . \tag{3.13}$$

Again, the exact discrete-time model of the system can be computed, but we consider a control design based on the "partial Euler" model

$$x(k+1) = (1+T)x(k) + (e^T - 1)u(k) . \tag{3.14}$$

The control

$$u = u_T(x) = -\frac{T(1 + \frac{1}{2}T)x}{e^T - 1} \tag{3.15}$$

stabilizes the family of approximate models (for $T \in (0, 2)$) by placing the eigenvalue of the closed-loop at $1 - \frac{1}{2}T^2$. On the other hand, the eigenvalue of the exact discrete-time closed-loop is located at $e^T - T - \frac{1}{2}T^2 > 1$, $\forall T > 0$.

Note that in Example 3.2 we have the following properties.

- **Nonuniform attractive rate**. For all $T > 0$, the family of approximate discrete-time models satisfies

$$|\phi_T(k, x_\circ)| \le b e^{-kT^2} |x_\circ| , \quad k \in \mathbb{N} ,$$

where $b > 0$ is independent of T. Therefore the overshoot is uniformly bounded in T. However, if we think of $kT = t$ as "continuous-time", then as $T \to 0$, the rate of convergence of solutions is such that for any $t > 0$ we have $e^{-tT} \to 1$. In other words, the rate of convergence in continuous-time is not uniform in the parameter T.

Conditions in our framework for controller design in the next section will rule out all of the above non-robustness indicators.

3.5 Preliminary Results on Stability and Stabilization

This section contains two main results. In Proposition 3.1 we show under natural and general conditions that stability of the exact discrete-time model implies stability of the sampled-data system. Proposition 3.2 provides Lyapunov conditions to analyze the stability of the exact discrete-time model.

These results are important in proving that stability of approximate model will guarantee, under appropriate conditions, stability of the sampled-data system. Indeed, we show in the next section that stability of the approximate model implies, under certain checkable conditions, the stability of the exact model and, consequently, we can conclude stability of the sampled-data system using the results proved in this section.

Note that if Assumption 1 was satisfied, the results of this section could be used to conclude stability of the sampled-data systems directly from stability of its exact discrete-time model. However, since we use Assumption 2, more work will be needed to investigate when stability of the approximate model implies stability of the exact.

Suppose for simplicity that a parameterized family of control laws (3.10) was designed for the system so that the closed-loop sampled-data system becomes

$$\dot{x}(t) = f(x(t), u_T(x(kT))) \qquad t \in [kT, (k+1)T) \ . \tag{3.16}$$

Hence, with the control (3.10) the closed-loop exact model of this sampled-data system is

$$x(k+1) = F^e_T(x(k), u_T(x(k))) = \mathcal{F}^e_T(x(k)) \ . \tag{3.17}$$

Proposition 3.1 given below states that if the sampled-data system (3.16) has bounded inter-sample behaviour (condition 2), then GAS of the exact discrete-time model (condition 1) implies UGAS of the sampled-data system[7]. The proof of this proposition is presented in [46].

[7] Note that the exact discrete-time model (3.17) is time invariant whereas the sampled-data system (3.16) is periodically time varying because of sampling. Hence, we talk about "uniform" GAS for the sampled-data system where uniformity is with respect to the initial time instant $t_\circ$.

Proposition 3.1. *Consider a sampled-data system (3.16) and suppose that the sampling period $T > 0$ is such that the following two conditions hold.*

1. *There exists $\widetilde{\beta} \in \mathcal{KL}$ such that the trajectories of the exact discrete-time closed-loop system (3.17) satisfy*

$$|x(k)| \leq \widetilde{\beta}(|x_\circ|, kT) \qquad \forall k \in \mathbb{N}, \quad x(0) = x_\circ \in \mathbb{R}^n \ . \tag{3.18}$$

2. *There exists $\kappa \in \mathcal{K}_\infty$ such that the solutions of the sampled-data system (3.16) satisfy*

$$|x(t)| \leq \kappa(|x_\circ|) \qquad \forall t \in [t_\circ, t_\circ + T], \ t_\circ \geq 0, \ x(t_\circ) = x_\circ \in \mathbb{R}^n \ . \tag{3.19}$$

Then there exists $\beta \in \mathcal{KL}$ such that the trajectories of the sampled-data system satisfy[8]

$$|x(t)| \leq \beta(|x_\circ|, t - t_\circ) \qquad \forall t \geq t_\circ \geq 0, \quad x(t_\circ) = x_\circ \in \mathbb{R}^n \ . \tag{3.20}$$

Moreover, if $\widetilde{\beta} \in$ exp-$\mathcal{KL}$ and $\kappa \in \mathcal{K}_\infty$ is linear, we can take $\beta \in$ exp-$\mathcal{KL}$. ■

Remark 3.1. If the function f is globally Lipschitz then condition 2 of Proposition 3.1 always holds. It is important to note that condition 2 holds for any locally Lipschitz discrete-time model F_T^e in an appropriate relaxed (semiglobal practical) sense if the sampling period T is sufficiently reduced. We decided not to state these more general conditions to simplify the presentation. The more general semiglobal practical stability results (that are also more natural in this context) can be found in [46].

Condition 2 of Proposition 3.1 holds under natural and general conditions and it only remains to see how one can satisfy condition 1. The following result that will help verifying condition 1 in Proposition 3.1 is presented with a proof.

Proposition 3.2. *Suppose there exists a family of Lyapunov functions $V_T(x)$ parameterized by T and $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{K}_\infty$ such that the following conditions hold for all $x \in \mathbb{R}^n$.*

$$\begin{aligned} \alpha_1(|x|) \leq V_T(x) &\leq \alpha_2(|x|) \ , \\ \frac{V_T(\mathcal{F}_T^e(x)) - V_T(x)}{T} &\leq -\alpha_3(|x|) \ . \end{aligned} \tag{3.21}$$

[8] It was shown in [12] that the state of the sampled-data system (3.16) at any time instant $t_\circ \in [kT, (k+1)T)$ consists of $x(t_0)$ and $u_T(x(k))$. Hence, strictly speaking the stability bound (3.20) is not equivalent to uniform global asymptotic stability of the sampled-data system. However, if $|u_T(x)| \leq \varphi(|x|)$ for some $\varphi \in \mathcal{K}_\infty$, our conditions imply uniform global asymptotic stability of the sampled-data system. To conclude $\beta \in$ exp-$\mathcal{KL}$, we also need that the function φ is linear.

Then there exists $\widetilde{\beta} \in \mathcal{KL}$ such that condition 1 of Proposition 3.1 holds. That is, the solutions of the exact discrete-time model (3.17) satisfy (3.18). Moreover, if there exist $a_i > 0$ and $p > 0$ such that $\alpha_i(s) = a_i s^p$ for $i = 1, 2, 3$, then condition 1 of Proposition 3.1 holds with $\widetilde{\beta} \in$ exp-$\mathcal{KL}$. ■

Proof: Note that (3.21) implies

$$\frac{V_T(\mathcal{F}_T^e(x)) - V_T(x)}{T} \leq -\alpha_3 \circ \alpha_2^{-1}(V_T(x)) =: -\alpha(V_T(x)) \ .$$

Denote $V_T(kT) := V_T(x(kT))$. We introduce a variable $t \in \mathbb{R}$ and define $y(t) := V_T(kT) + (t - kT)\frac{V_T((k+1)T) - V_T(kT)}{T}, t \in [kT, (k+1)T], k \geq 0$. Note that $0 \leq y(kT) = V_T(kT), k \geq 0$ and $y(t)$ is a continuous function of the "time" t. Moreover, it is absolutely continuous in t (in fact, piecewise linear) and we can write for almost all t,

$$\begin{aligned} \frac{d}{dt}y(t) &= \frac{V_T((k+1)T) - V_T(kT)}{T} \\ &\leq -\alpha(V_T(kT)) \ , \qquad \text{for } t \in [kT, (k+1)T), \quad k \geq 0 \ , \\ &\leq -\alpha(y(t)) \ , \qquad \text{for } t \geq 0 \ . \end{aligned} \tag{3.22}$$

Let $v(t) = \beta(v_0, t)$ be the (unique) solution of $\dot{v} = -\alpha(v)$, $v(t_0) = v_0$. It is shown in Lemma 6.1 in [58] that $\beta \in \mathcal{KL}$. By standard comparison theorems (see for instance [30, Theorem 1.10.2]) we have for $y_0 = v_0$ that

$$y(t) \leq v(t) = \beta(y_0, t - t_0), \ \forall t \geq t_0 \ ,$$

which implies using $V_T(kT) = y(kT)$ with $t = kT, t_0 = k_0 = 0, y_0 = V_T(0)$ that

$$|x(k)| \leq \alpha_1^{-1}(V_T(kT)) \leq \alpha_1^{-1}(\beta(V_T(0), kT)) \leq \alpha_1^{-1}(\beta(\alpha_2(|x_0|), kT)), \ k \geq 0 \ ,$$

which proves (3.18) with $\widetilde{\beta}(s,t) := \alpha_1^{-1}(\beta(\alpha_2(s), t))$. Proving that $\widetilde{\beta} \in$ exp-$\mathcal{KL}$ under stronger conditions is easy following the same steps. ■

Remark 3.2. The above results hold for arbitrarily large T. In other words, they are not fast sampling results. However, to satisfy some of these conditions we will need to reduce T in general. For example, to satisfy condition 2 of Proposition 3.1 on a compact subset of $\mathbb{R}^n$ in case f is locally Lipschitz in x, we need to reduce T sufficiently. Similarly, the results of the next section will require fast sampling to show that under certain conditions stability of an approximate model implies stability of the exact model.

3.6 Framework for Controller Design

In this section we show how one can conclude, under certain checkable conditions, that a controller that stabilizes the approximate model F_T^a is guaranteed to also stabilizes the exact model F_T^e. Then, we can conclude that the sampled-data model is also stabilized using the results from the previous section. We will start from the simplest case of exponential stability design which we will prove in detail. While the proof of this result is quite easy to follow, the used conditions are quite strong for general nonlinear systems. In Subsection 3.6.2 we present without proof a more general result on semiglobal practical stability that uses more natural and less restrictive conditions.

3.6.1 Global Exponential Stabilization

Suppose that Assumption 2 holds and we want to achieve global exponential stability (GES) of F_T^e by stabilizing F_T^a. To do this, we assume for convenience that the function $f(\cdot,\cdot)$ in the continuous-time plant model is globally Lipschitz (this can be relaxed).

We need to find conditions that guarantee global exponential stability of the exact discrete-time closed loop system (3.17) via the following discrete-time approximate closed-loop system

$$x(k+1) = F_T^a(x(k), u_T(x(k))) \ , \tag{3.23}$$

where the family of controllers (3.10) that is parameterized by T is designed using the family of approximate discrete-time models (3.9). In the sequel, we refer to the exact (3.17) and approximate (3.23) closed loop systems respectively as (F_T^e, u_T) and (F_T^a, u_T). Using Proposition 3.2, it is reasonable to aim to design the family of controllers (3.10) so that the following holds for some Lyapunov function family (these conditions are also strong and can be relaxed).

$$\begin{aligned} a_1 |x|^c \le V_T(x) &\le a_2 |x|^c \\ \frac{V_T(F_T^a(x, u_T(x))) - V_T(x)}{T} &\le -a_3 |x|^c \ , \end{aligned} \tag{3.24}$$

for some $c > 0$, all x and all $T \in (0, T^*)$ where $T^* > 0$ is fixed. Hence, Proposition 3.2 guarantees that (F_T^a, u_T) is GES for all small $T \in (0, T^*)$. The reason for requiring this condition to hold for all small T is going to become clear soon.

We want to see when the above conditions imply that all conditions of Proposition 3.2 hold for the closed-loop exact discrete-time model if we perhaps further reduce T. In order to see when this can be achieved, add and subtract $\frac{1}{T} V_T(F_T^e(x, u_T(x)))$ to (3.24) yielding

$$\begin{aligned} \frac{V_T(F_T^e(x, u_T(x))) - V_T(x)}{T} &\le -a_3 |x| \\ &\quad + \frac{V_T(F_T^e(x, u_T(x))) - V_T(F_T^a(x, u_T(x)))}{T} \ . \end{aligned} \tag{3.25}$$

Suppose also that for all $x, y \in \mathbb{R}^n$ and all $T \in (0, T^*)$ the following two conditions hold.

$$|V_T(x) - V_T(y)| \leq L\,|x - y| \tag{3.26}$$

and

$$|F_T^e(x, u_T(x)) - F_T^a(x, u_T(x))| \leq T\rho(T)\,|x| \tag{3.27}$$

where $\rho \in \mathcal{K}_\infty$. Then, from (3.25), (3.26) and (3.27) we obtain

$$\begin{aligned}\frac{V_T(F_T^e(x, u_T(x))) - V_T(x)}{T} &\leq -a_3\,|x| + \frac{L\,|F_T^e(x, u_T(x)) - F_T^a(x, u_T(x))|}{T} \\ &\leq -a_3\,|x| + \frac{LT\rho(T)\,|x|}{T} \\ &= -a_3\,|x| + L\rho(T)\,|x| \ .\end{aligned} \tag{3.28}$$

It is now obvious that for all $T \in (0, T_1^*)$ with $T_1^* := \min\{T^*, \rho^{-1}(a_3/2L)\}$ we have that

$$\frac{V_T(F_T^e(x, u_T(x))) - V_T(x)}{T} \leq -\frac{1}{2}a_3\,|x| \ , \tag{3.29}$$

and, hence, we can conclude from Proposition 3.2 that the closed-loop exact model (F_T^e, u_T) is GES. Before discussing this result in detail, we state our findings in the following proposition.

Proposition 3.3. *Suppose there exists $T^* > 0$ such that for all $T \in (0, T^*)$ the following holds.*

1. *The closed-loop approximate model (F_T^a, u_T) satisfies (3.24). Moreover, condition (3.26) holds uniformly in $T \in (0, T^*)$.*
2. *Condition (3.27) holds.*

Then for all $T \in (0, T_1^*)$, with $T_1^* := \min\{T^*, \rho^{-1}(a_3/2L)\}$, we have that the closed-loop exact model (F_T^e, u_T) satisfies (3.21) with $\alpha_i(s) = a_i s$ for $i = 1, 2$ and $\alpha_3(s) = \frac{a_3}{2}s$. ■

The condition (3.27) quantifies the mismatch between the exact and approximate closed-loop models and similar conditions are named *consistency* in the numerical analysis literature [59]. Note that (3.27) is not easy to use since we need to first design u_T to check it. Hence, it would be better if a condition involving the (open-loop) exact (3.8) and approximate (3.9) models is used. This condition is now stated.

Proposition 3.4. *Suppose there exist $\rho_1 \in \mathcal{K}_\infty$, $K > 0$ and $T^* > 0$ such that for all $T \in (0, T^*)$ and all $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ the following conditions hold.*

$$|F_T^e(x, u) - F_T^a(x, u)| \leq T\rho_1(T)[|x| + |u|] \tag{3.30}$$

and

$$|u| := |u_T(x)| \leq K\,|x| \ . \tag{3.31}$$

Then condition (3.27) holds for all $T \in (0, T^*)$*, with* $\rho(s) := \rho_1(s) \cdot [1 + K]$. ■

We emphasize that the conditions (3.30) and (3.31) are easier to use than (3.27).

Combining the statements of Propositions 3.3 and 3.4, these results can be paraphrased as follows. The exact model (F^e_T, u_T) is exponentially stable if the following conditions hold.

1. Lyapunov exponential stability of (F^a_T, u_T) with a globally Lipschitz Lyapunov function (*i.e.* (3.24) and (3.26)).
2. Consistency between the approximate F^a_T and exact F^e_T models of the open-loop systems (*i.e.* (3.30)).
3. Uniform boundedness of control law u_T with respect to small T (*i.e.* (3.31)).

We emphasize that all of the above conditions can be checked without knowing the explicit expression of F^e_T. Indeed, it is obvious that the first and the third conditions only use the knowledge of F^a_T and u_T. The second condition is defined using F^e_T but we note that we do not need to know F^e_T in order to verify that the bound (3.30) holds. Indeed, we can state the following result.

Proposition 3.5. *Suppose that the system (3.5) is globally Lipschitz and* $f(0,0) = 0$*. Suppose, moreover, that* F^a_T *is consistent with* F^{Euler}_T *defined in (3.7). That is, there exists* $T^* > 0$ *and* $\rho_1 \in \mathcal{K}_\infty$ *such that for all* $x \in \mathbb{R}^n$ *and* $u \in \mathbb{R}^m$ *we have*

$$\left|F^a_T(x,u) - F^{Euler}_T(x,u)\right| \leq T\rho_1(T)[|x| + |u|] \ .$$

Then, F^e_T *is consistent with* F^a_T *in the sense of (3.30).* ■

Proof: First, we show that F^{Euler}_T is consistent with F^e_T under the given conditions on f. Indeed, since f is globally Lipschitz and zero at zero, we have $|f(x,u)| \leq L(|x| + |u|)$ and the solution $\phi(t,x,u)$ of the system (3.5) starting from x with the constant control $u(t) \equiv u$ exists for all time, is unique and satisfies

$$|\phi(t,x,u)| \leq \exp(Lt)\,|x| + (\exp(Lt) - 1)\,|u| \qquad \forall t \geq 0, \forall x, u \ .$$

Denote $\phi(t,x,u)$ shortly as $\phi(t)$. Then, using the above bound on $\phi(t)$ and the Lipschitzity of f we can write

$$
\begin{aligned}
\left|F_T^e(x,u) - F_T^{Euler}(x,u)\right| &= \left|x + \int_0^T f(\phi(s),u)ds - x - Tf(x,u)\right| \\
&= \left|\int_0^T [f(\phi(s),u) - f(x,u)]ds\right| \\
&\le \int_0^T |f(\phi(s),u) - f(x,u)|\, ds \\
&\le \int_0^T L\,|\phi(s) - x|\, ds \\
&= \int_0^T L \left|\int_0^s f(\phi(\tau),u)d\tau\right| ds \qquad (3.32) \\
&\le \int_0^T \int_0^s L\,|f(\phi(\tau),u)|\, d\tau ds \\
&\le \int_0^T \int_0^s L[L\,|\phi(\tau)| + L\,|u|]d\tau ds \\
&\le \int_0^T \int_0^T L[L\exp(LT)\,|x| \\
&\quad + L(\exp(LT) - 1)\,|u| + L\,|u|]d\tau ds \\
&= \frac{1}{2}T^2L^2\exp(LT)(|x| + |u|)\ ,
\end{aligned}
$$

which completes the proof of consistency between F_T^e and F_T^{Euler}. Finally, by adding and subtracting F_T^{Euler} and using the triangular inequality, we obtain

$$
\begin{aligned}
|F_T^e(x,u) - F_T^a(x,u)| &= \left|F_T^e(x,u) - F_T^{Euler}(x,u) + F_T^{Euler}(x,u) - F_T^a(x,u)\right| \\
&\le \left|F_T^e(x,u) - F_T^{Euler}(x,u)\right| + \left|F_T^{Euler}(x,u) - F_T^a(x,u)\right|
\end{aligned}
$$

and the conclusion immediately follows since F_T^e is consistent with F_T^{Euler} and by assumption F_T^a is consistent with F_T^{Euler}. ■

Remark 3.3. The conditions in Propositions 3.3 and 3.4 provide a prescriptive framework for controller design via approximate models. Indeed, the first step in this approach is to pick F_T^a that is consistent with F_T^e in the sense of (3.30). Then, one would like to design a family of controllers of the form (3.10) that are bounded in the sense of (3.31) for the family of approximate models that satisfies the Lyapunov conditions (3.24) and (3.26). All of these conditions are checkable without knowing the explicit expression of F_T^e. Note that we do not say how one can design such controllers and that is why we refer to this framework as "prescriptive" rather than "constructive". However, we will show in Section 3.7 that one can obtain a variety of constructive procedures within this framework for certain classes of nonlinear systems, such as separable Hamiltonian systems and systems in strict feedback form.

3.6.2 Semiglobal Practical Stability

The purpose of this subsection is to present several definitions of stability and consistency that are more general than the ones in the previous subsection and use them to provide a more general framework for controller design via approximate models.

Semiglobal practical asymptotic stability property naturally arises when we relax the conditions of global Lipschitzity on f and GES for (F_T^a, u_T) that we used in the previous subsection. For simple illustration, we consider a parameterized family of discrete-time nonlinear systems

$$x(k+1) = F_T(x(k), u_T(x(k))) \ . \tag{3.33}$$

Semiglobal practical asymptotic stability and semiglobal practical asymptotic stability Lyapunov function for the system (3.33) are defined as follows.

Definition 3.1 (Semiglobal practical asymptotic (SPA) stability). *The family of systems (3.33) is SPA stable if there exists $\beta \in \mathcal{KL}$ such that for any strictly positive real numbers (Δ, δ) there exists $T^* > 0$ such that for all $T \in (0, T^*)$, all initial states $x(0) = x_\circ$ with $|x_\circ| \leq \Delta$, the solutions of the system satisfy*

$$|x(k)| \leq \beta(|x_\circ|, kT) + \delta, \quad \forall k \in \mathbb{N} \ . \tag{3.34}$$

■

Definition 3.2 (SPAS Lyapunov function). *A continuously differentiable function $V_T : \mathbb{R}^n \to \mathbb{R}$ is called SPAS Lyapunov function for the system F_T if there exist class $\mathcal{K}_\infty$ functions $\underline{\alpha}(\cdot)$, $\overline{\alpha}(\cdot)$, $\alpha(\cdot)$ such that for any strictly positive real numbers (Δ_x, ν), there exist $L, T^* > 0$ such that for all $T \in (0, T^*)$ and for all $x, y \leq \Delta_x$ and $T \in (0, T^*)$ the following holds.*

$$\underline{\alpha}(|x|) \leq V_T(x) \leq \overline{\alpha}(|x|) \ , \tag{3.35}$$

$$V_T(F_T(x, u_T(x))) - V_T(x) \leq -T\alpha(|x|) + T\nu \tag{3.36}$$

$$|V_T(x) - V_T(y)| \leq L\,|x - y| \tag{3.37}$$

In this case, we say that the pair (V_T, u_T) is Lyapunov SPA stabilizing for the system F_T. ■

We now state a more general notion of consistency.

Definition 3.3 (One-step consistency). *The family F_T^a is said to be one-step consistent with F_T^e if there exist functions $\rho, \varphi_1, \varphi_2 \in \mathcal{K}_\infty$ such that given any strictly positive real numbers (Δ_x, Δ_u) there exists $T^* > 0$ such that, for all $T \in (0, T^*)$, $|x| \leq \Delta_x$, $|u| \leq \Delta_u$ we have*

$$|F_T^e(x, u) - F_T^a(x, u)| \leq T\rho(T)[\varphi_1(|x|) + \varphi_2(|u|)] \ . \tag{3.38}$$

■

Definition 3.4. *The family of controllers u_T is bounded, uniformly in small T, if there exist $\kappa \in \mathcal{K}_\infty$ and for any $\Delta > 0$ there exists $T^* > 0$ such that for all $|x| \leq \Delta$ and $T \in (0, T^*)$ we have*

$$|u_T(x)| \leq \kappa(|x|) .$$

■

Using the above definitions, we can now state the following result.

Theorem 3.2. *Suppose the following conditions hold.*

1. *F^a_T is one-step consistent with F^e_T.*
2. *u_T is bounded, uniformly in small T.*
3. *There exists a SPAS Lyapunov function for the system (F^a_T, u_T).*

Then the system (F^e_T, u_T) is SPA stable and, hence, the sampled-data system (3.16) is SPA stable. ■

The statement of the above theorem is fully consistent with the result presented in the previous subsection but here we use much weaker (and hence more general) conditions that yield weaker conclusions. Hence, Theorem 3.2 is much more widely applicable than the results of the previous subsection.

Remark 3.4. Theorem 3.2 can be strengthened in different ways to either obtain global stability (as opposed to semiglobal) or to achieve local exponential stability. This can be done by combining stronger conditions in Proposition 3.3 with conditions in Theorem 3.2.

Remark 3.5. While conditions of Theorem 3.2 are sufficient (not necessary in general), they are tight in the sense that if we try to relax any of them, then we can find a counterexample where the exact closed-loop is not stabilized for small T. Example 1 and Example 2 in Section 3.4 can be used to illustrate this.

Indeed, Lyapunov SPA stability of the approximate closed-loop implies via Proposition 3.2 that[9] the approximate closed-loop system is SPA stable in the sense of Definition 3.1. This rules out two of the non-robustness indicators shown in Examples 1 and 2: non-uniform overshoot and non-uniform convergence rate. Moreover, the second condition in Theorem 3.2 requires uniform boundedness of the control law in small T. Hence, conditions of Theorem 3.2 rule out all indicators of non-robustness that we observed in Section 3.4. Another example that shows the need for the use of continuous Lyapunov function is presented in [43].

[9] Actually, we need a slightly more general statement than Proposition 3.2 that can be found in [43, 45].

Remark 3.6. Various extensions and variations of Theorem 3.2 have been published in the literature. First, alternative proofs that do not require the knowledge of a Lyapunov function and use SPA stability of closed-loop approximate model can be found in [45] for time invariant systems and in [40] for time-varying systems. These results use a slightly different notion of consistency than the one given in Definition 3.3. A framework for achieving input-to-state stability (ISS) and integral input-to-state stability (iISS) for systems with exogenous inputs via approximate discrete-time models can be found in [38] and [35], respectively. Moreover, similar results for sampled-data differential inclusions are presented in [43].

3.7 Controller Design within the Framework

In this section, we present several simple design tools via approximate discrete-time models that rely on the framework presented in Section 3.6. We emphasize that any techniques for continuous time controller design can be revisited within our framework and new control laws will be obtained as a result (*e.g.* see the backstepping design in Subsection 3.7.4).

In Subsection 3.7.1 we show that emulation of continuous time controllers can be regarded as a special case of controller design that fits within our framework. In this case, we design a continuous-time controller $u^{ct}(x)$ for the continuous-time plant and then implement

$$u(t) = u_T^{dt}(x(k)) \qquad t \in [kT, (k+1)T) \ , \tag{3.39}$$

where[10]

$$u_T^{dt}(x) = u^{ct}(x) \ , \tag{3.40}$$

i.e. the discrete-time controller is identical to the continuous time controller. Note that we can still think of emulation as a design via an approximate model (the continuous-time plant model).

Subsections 3.7.2 and 3.7.3 show that our framework can be used for continuous-time controller *redesign* for sampled-data implementation. In this case, we first design a continuous-time controller $u^{ct}(x)$ for the continuous-time plant model (ignoring sampling) and then in the second step we parameterize the controller in the following manner:

$$u_T^{dt}(x) = u^{ct}(x) + \sum_{i=1}^{M} T^i u_i \ , \tag{3.41}$$

where $M \geq 1$ is a fixed integer and then we use an approximate model, such as the Euler model, to design $u_i = u_i(x)$. This *redesign* of continuous-time controllers can be directed to achieve different objectives and we will present two

[10] We introduce u_T^{dt} to be able to compare emulation with other design techniques.

cases of this controller redesign technique. In Subsection 3.7.2 the Lyapunov function for the continuous-time closed-loop is used as a *control Lyapunov function* for the approximate discrete-time model, assuming the redesigned controller follows the form (3.41). After substituting the term u^{ct} that is known from (3.40), the extra terms u_i's are regarded as new controls. Once the u_i's have been computed, the controller (3.41) is implemented. In Subsection 3.7.3, controller redesign is done starting from a passivity based design for a class of Hamiltonian systems namely the interconnection and damping assignment − passivity based control (IDA-PBC) design method. The modified energy function of the system is used as control Lyapunov function and design is carried out in a similar way as in Subsection 3.7.2.

Backstepping based on the Euler approximate model is presented in Subsection 3.7.4. In this case, we do not design a continuous controller as a first step in design/redesign but rather we use the Euler approximate model directly to design $u_T^{dt}(x)$ using our framework and then implement it using (3.39). It is interesting to observe that although we *do not assume* that the controller has the form (3.41), we show in our example that the obtained controller has the form (3.41) where $u^{ct}(x)$ is a controller that could be obtained using a continuous-time backstepping design (but we do not need to design it first). Moreover, we show in our example that in simulations $u_T^{dt}(x)$ performs better than the emulated $u^{ct}(x)$.

3.7.1 Emulation

Suppose that a static state feedback controller $u = u^{ct}(x)$ has been designed for the continuous-time system (3.5) ignoring sampling, so that there exists a smooth Lyapunov function V satisfying the following conditions:

$$\alpha_1(|x|) \leq V(x) \leq \alpha_2(|x|) \tag{3.42}$$

$$\frac{\partial V}{\partial x} f(x, u^{ct}(x)) \leq -\alpha_3(|x|) \;, \tag{3.43}$$

with $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{K}_\infty$. These conditions guarantee GAS of the continuous-time closed-loop system (f, u^{ct}). Suppose also that $u^{ct}(x)$ is bounded on compact sets of the state space. Then, suppose that the controller is "emulated" using (3.40). Suppose, moreover that the sampled-data system

$$\dot{x}(t) = f(x(t), u^{ct}(x(k))) \qquad t \in [kT, (k+1)T) \;, \tag{3.44}$$

has solutions that are well defined[11] for all initial conditions $x(0) = x_\circ \in \mathbb{R}^n$ and all $t \in [0, T]$. Denote $F_T^{Euler} := x(k) + Tf(x(k), u^{ct}(x(k)))$.

[11] Typically, for locally Lipschitz f the solutions would be defined only in a semiglobal sense, *i.e.* for any bounded set of initial conditions there exists $T^* > 0$ such that for all $T \in (0, T^*)$ and all initial conditions from the set we have that the solutions are well defined for $t \in [0, T]$.

We will show next that the sampled-data system (3.44) is stable in an appropriate sense under appropriate conditions. In particular, we can state the following result.

Theorem 3.3. *Suppose that we have found a (locally bounded) controller $u^{ct}(x)$ and a smooth $V(x)$ that satisfy (3.42) and (3.43). Then, (V, u^{ct}) is a Lyapunov SPA stabilizing pair for F_T^{Euler}. Hence, (V, u^{ct}) is a Lyapunov SPA stabilizing pair for F_T^e, consequently, the sampled-data system (3.44) is SPA stable.* ■

Proof: We first prove that (V, u^{ct}) is a Lyapunov SPA stabilizing pair for the Euler model of the system (3.7). Adding and subtracting $\frac{V(F_T^{Euler})-V(x)}{T}$ to (3.43) and using the Mean Value Theorem twice, we obtain

$$\begin{aligned}
&\frac{V(F_T^{Euler}) - V(x)}{T}\\
&\quad \le -\alpha_3(|x|) + \frac{V(F_T^{Euler}) - V(x)}{T} - \frac{\partial V}{\partial x}(x) f(x, u^{ct}(x))\\
&\quad = -\alpha_3(|x|) + \left[\frac{\partial V}{\partial x}(x + \theta_1 T f(x, u^{ct}(x))) - \frac{\partial V}{\partial x}(x)\right] f(x, u^{ct}(x))\\
&\quad \le -\alpha_3(|x|) + \left|\frac{\partial V}{\partial x}(x + \theta_1 T f(x, u^{ct}(x))) - \frac{\partial V}{\partial x}(x)\right| \cdot \left|f(x, u^{ct}(x))\right|\\
&\quad \le -\alpha_3(|x|) + \theta_1 T \left|\frac{\partial^2 V}{\partial x^2}(x + \theta_2 T f(x, u^{ct}(x)))\right| \cdot \left|f(x, u^{ct}(x))\right|^2\\
&\quad \le -\alpha_3(|x|) + T\kappa(|x|) \ ,
\end{aligned}$$

where $\theta_1, \theta_2 \in (0, 1)$, $\kappa \in \mathcal{K}_\infty$, we assumed that T is bounded and the first and the second derivatives of V are continuous (V is smooth). Hence, (V, u^{ct}) is a Lyapunov SPA stabilizing pair for F_T^{Euler}. Note that F_T^{Euler} is one-step consistent with F_T^e and $u^{ct}(x)$ is assumed to be bounded on compact sets and, hence, bounded uniformly in small T (since $u^{ct}(x)$ is independent of T). Since V has continuous first derivative, it is locally Lipschitz and we can conclude in a similar manner like in the proof of Proposition 3.3 that

$$\frac{V(F_T^e) - V(x)}{T} \le -\alpha_3(|x|) + T\kappa_1(|x|) \ , \tag{3.45}$$

for some $\kappa_1 \in \mathcal{K}_\infty$. Hence, (V, u^{ct}) is a Lyapunov SPA stabilizing pair for the exact model F_T^e. Finally, we conclude that the sampled-data system (3.44) is SPA stable from Theorem 3.2. ■

Remark 3.7. The analysis given above can be carried out with more generality and one can prove that emulation leads to preservation of arbitrary dissipation inequalities in an appropriate sense (see [29] for more details).

The following example will be used to illustrate all our controller design and redesign methods. The reason for considering this simple system in strict feedback form is that we can use backstepping to systematically design a control law and a Lyapunov function that are needed to apply our framework.

Example 3.3. Consider the continuous-time plant

$$\begin{aligned} \dot{\eta} &= \eta^2 + \xi \\ \dot{\xi} &= u \ . \end{aligned} \tag{3.46}$$

We design a continuous-time backstepping controller [24]. Note that the first subsystem can be stabilized with the "control" $\phi(\eta) = -\eta^2 - \eta$ with the Lyapunov function $W(\eta) = \frac{1}{2}\eta^2$. Using this information and applying [24, Lemma 2.8 with c=1], we obtain

$$u^{ct}(\eta, \xi) = -2\eta - \eta^2 - \xi - (2\eta + 1)(\xi + \eta^2) \ , \tag{3.47}$$

which globally asymptotically stabilizes the continuous-time system (3.46) and moreover

$$V(\eta, \xi) = \frac{1}{2}\eta^2 + \frac{1}{2}(\xi + \eta + \eta^2)^2 \tag{3.48}$$

is a Lyapunov function for the continuous-time closed-loop system. Hence, we conclude from Theorem 3.3 that the sampled-data system (3.44) is SPA stable. Simulations for the sampled-data system with the emulated controller are presented in Subsection 3.7.4 and a comparison to other controllers obtained in the sequel is presented.

3.7.2 Continuous-time Controller Redesign

In this subsection we illustrate the Lyapunov based redesign and we refer to [36] for more details. We assume that a continuous-time controller

$$u = u^{ct}(x) \tag{3.49}$$

has been designed and a Lyapunov function V satisfying (3.42), (3.43) was found for the closed-loop continuous-time system. Suppose that we want to implement a controller of the form (3.41) and we want to further design u_i so that the controller is "better" in some sense than u^{ct}. For simplicity, let us assume that

$$u_T^{dt}(x) := u^{ct}(x) + Tu_1(x) \ ,$$

and $u_1(x)$ is a new control input that we want to design (*i.e.* we redesign $u^{ct}(x)$). We do that by using the continuous-time Lyapunov function V as a control Lyapunov function for an approximate discrete-time model F_T^a that is one step consistent with the exact model F_T^e. That is, we consider

$$\frac{V(F_T^a(x, u^{ct}(x) + Tu_1(x))) - V(x)}{T} ,$$

where F_T^a is one step consistent with F_T^e, and u^{ct} and V were obtained from an arbitrary continuous-time design. There are different possible objectives that one may try to achieve by designing u_1 and we discuss here one obvious choice. Let us first note that we can easily compute

$$\frac{V(F_T^a(x, u^{ct}(x))) - V(x)}{T} .$$

One way to design u_1 is to require that

$$\frac{V(F_T^a(x, u^{ct}(x) + Tu_1(x))) - V(x)}{T} < \frac{V(F_T^a(x, u^{ct}(x))) - V(x)}{T} . \tag{3.50}$$

In other words, we can design u_1 to achieve more decrease for the Lyapunov function along solutions of the closed-loop approximate model with the redesigned controller (see [36]). However, not all Lyapunov functions that satisfy (3.42) and (3.43) are appropriate for doing the redesign with the aim of achieving the objective (3.50). Indeed, increasing the rate of convergence in this way may lead to increasing the overshoots for some Lyapunov functions, which is highly undesirable (see [36, Example 4.1]). To avoid creating unacceptable overshoots in this manner, we need to assume that V is "well behaved", that is the overshoot estimates that can be obtained using V for the closed-loop system are acceptable (see [36, Assumption 2.2]). We acknowledge that finding an appropriate V that satisfies this assumption is difficult in general. With this assumption, the above described redesign will yield acceptable overshoots while it will typically improve the rate of convergence of the approximate and sampled data closed-loop systems.

Finally, note that if $u_1 = u_1(x)$ is designed to satisfy (3.50) and it is bounded on compact sets, then we can conclude from our Theorem 3.2 that the sampled-data system with the redesigned controller is SPA stable. We revisit Example 3.3 to illustrate this approach.

Example 3.4. Consider the system in Example 3.3 and assume that we have already designed the controller (3.47) and found the Lyapunov function (3.48). Assume that the plant (3.46) is between a sampler and a zero-order-hold and let us use for redesign, its Euler approximate model

$$\begin{aligned} \eta(k+1) &= \eta(k) + T(\eta^2(k) + \xi(k)) \\ \xi(k+1) &= \xi(k) + Tu(k) . \end{aligned} \tag{3.51}$$

Denote $x := (\eta\ \xi)^T$. Suppose for simplicity that $u^{dt}(x) = u^{ct}(x) + Tu_1(x)$ and it is then not hard to compute

$$\frac{V(F_T^{Euler}(x, u^{ct}(x) + Tu_1)) - V(x)}{T} = -\eta^2 - (\xi + \eta + \eta^2)^2 + Tp_1(u_1, x) + O(T^2)$$

where

$$p_1(u_1, x) = \frac{1}{2}(\eta^2+\xi)^2+(\xi+\eta+\eta^2)(u_1+(\eta^2+\xi)^2)+\frac{1}{2}(2\eta+\eta^2+\xi)^2 \ , \quad (3.52)$$

and $O(T^2)$ contains higher order terms in T. Since T will have to be chosen small, we neglect $O(T^2)$ and we chose u_1 so that the term $p_1(u_1, x)$ is made more negative (note that there are some terms in p_1 that can not be made negative using u_1). One obvious choice is

$$u_1(x) = -(\eta^2 + \xi)^2 - (\xi + \eta + \eta^2) \ , \quad (3.53)$$

which cancels one term and then provides extra damping to yield

$$p_1(u_1(x), x) = \frac{1}{2}(\eta^2 + \xi)^2 - (\xi + \eta + \eta^2)^2 + \frac{1}{2}(2\eta + \eta^2 + \xi)^2 \ .$$

We will simulate this controller in the next subsection and make some comparisons with other designs.

3.7.3 Discrete-time Interconnection and Damping Assignment – Passivity-based Control (IDA-PBC)

In this subsection, the second tool for continuous-time controller redesign is discussed. While in Subsection 3.7.2 we consider general nonlinear system, now we consider a class of nonlinear system namely Hamiltonian systems. The technique used for the controller design is a type of passivity based control design known as IDA-PBC.

IDA-PBC design is a powerful tool for solving the stabilization problem for Hamiltonian systems [47, 48, 50]. Although IDA-PBC design is applicable to a broader class of systems (see [1, 49, 50]), it applies naturally to Hamiltonian systems due to the special structure of this class of systems.

Consider continuous-time Hamiltonian systems whose dynamics can be written as

$$\begin{bmatrix} \dot{q} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} \begin{bmatrix} \nabla_q H \\ \nabla_p H \end{bmatrix} + \begin{bmatrix} 0 \\ G(q) \end{bmatrix} u \ , \quad (3.54)$$

where $p \in \mathbb{R}^n$ and $q \in \mathbb{R}^n$ are the states, and $u \in \mathbb{R}^m$, $m \leq n$, is the control action. The matrix $G(q) \in \mathbb{R}^{n\times m}$ is determined by the way control u enters the system. The function $H(q, p)$ is called the Hamiltonian function of the system, and is defined as the sum of the kinetic energy $K(q, p)$ and the potential energy $P(q)$[12], *i.e.*

[12] Note that in many references the potential energy is commonly denoted with V. However, we use the notation P instead, to avoid confusion with the notations V and V_T that we have used to denote Lyapunov functions.

$$H(q,p) = K(q,p) + P(q) = \frac{1}{2}p^\top M^{-1}(q)p + P(q) , \tag{3.55}$$

where $M(\cdot)$ is the symmetric inertia matrix.

We consider a simple case when system (3.54) is a *separable Hamiltonian system*. For this class of systems, the inertia matrix M is constant, and hence the kinetic energy and the potential energy of the system are decoupled, *i.e.*

$$H(q,p) = K(p) + P(q) = \frac{1}{2}p^\top M^{-1}p + P(q) . \tag{3.56}$$

We also consider only fully actuated systems, *i.e.* when $G(q)$ is full rank ($m = n$). In this setting, $\nabla_q H(q,p) = \nabla_q P(q)$ and $\nabla_p H(q,p) = M^{-1}p$. The idea of IDA-PBC design is to construct a controller for system (3.54) so that the stabilization is achieved assigning a desired energy function

$$H_d(q,p) = K_d(p) + P_d(q) = \frac{1}{2}p^\top M_d^{-1}p + P_d(q) , \tag{3.57}$$

that has an isolated minimum at the desired equilibrium point $(q^e, 0)$ of the closed-loop system. IDA-PBC design consists of two steps. First, design the energy shaping controller u_{es} to shape the total energy of the system to obtain the target dynamics; second, design the damping injection controller u_{di} to achieve asymptotic stability. Hence, an IDA-PBC controller is of the form

$$u = u_{es}(q,p) + u_{di}(q,p) . \tag{3.58}$$

The energy shaping controller u_{es} is obtained by solving the equation

$$\begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} \begin{bmatrix} \nabla_q H \\ \nabla_p H \end{bmatrix} + \begin{bmatrix} 0 \\ G(q) \end{bmatrix} u_{es} = \begin{bmatrix} 0 & M^{-1}M_d \\ -M_d M^{-1} & 0 \end{bmatrix} \begin{bmatrix} \nabla_q H_d \\ \nabla_p H_d \end{bmatrix} . \tag{3.59}$$

The first row of (3.59) is directly satisfied, and the second row can be written as

$$G u_{es} = \nabla_q H - M_d M^{-1} \nabla_q H_d . \tag{3.60}$$

Since we consider G full rank (and hence invertible), u_{es} is obtained as

$$u_{es} = G^{-1}(\nabla_q H - M_d M^{-1} \nabla_q H_d) . \tag{3.61}$$

Moreover, the damping injection controller u_{di} is constructed as

$$u_{di} = -k_v G^\top \nabla_p H_d = -k_v G^\top M_d^{-1} p, \quad k_v > 0 . \tag{3.62}$$

For more details and more general results about IDA-PBC design for continuous-time systems, we refer to [48, 49, 50].

In this subsection, we present a discrete-time IDA-PBC controller redesign to obtain a discrete-time IDA-PBC controller from a controller that is first obtained via continuous-time design. This redesign is based on the Euler approximate model of system (3.54), namely

$$\begin{aligned} q(k+1) &= q(k) + T\nabla_p H(q(k), p(k)) \\ p(k+1) &= p(k) - T\Big(\nabla_q H(q(k), p(k)) - Gu(k)\Big) \ . \end{aligned} \tag{3.63}$$

Suppose all conditions of the continuous-time design hold, and we have assigned the desired energy function (3.57) for the system. As in Subsection 3.7.2, we assume the Lyapunov function to be well behaved [36] and we are now ready to state the following theorem.

Theorem 3.4. *Consider the Euler model (3.63) of the separable Hamiltonian system (3.54) with Hamiltonian (3.56) and matrix G invertible. Suppose the inertia matrix M is diagonal and the desired potential energy V_d is positive definite. Then the discrete-time controller $u^T = u_{es}^T + u_{di}^T$ where*

$$u_{es}^T = G^{-1}\Big(\nabla_q H(q,p) - M_d M^{-1} \overline{\nabla}_q H_d(q,p)\Big) \tag{3.64}$$

$$u_{di}^T = -k_v G^\top \nabla_p H_d(q,p) = -k_v G^\top M_d^{-1} p \ , \tag{3.65}$$

with $k_v > 0$ and $\overline{\nabla}_q H_d(q,p) = \nabla_q H_d(q,p) + T\kappa L_v M^{-1} p$, where $\kappa > 0$ and $L_v = \nabla_{qq} P_d(q) \geq 0$ is a SPA stabilizing controller for the Euler model (3.63). Moreover, there exists a function

$$V(p,q) = H_d(p,q) + \epsilon p^\top q \ , \tag{3.66}$$

with $\epsilon > 0$ sufficiently small which is a SPAS Lyapunov function for the system (3.63), (3.64), (3.65). ■

Remark 3.8. It is known that Euler approximation is not Hamiltonian conserving. To avoid confusion about the motivation of using this method in our construction we emphasize that IDA-PBC design does not involve Hamiltonian conservation as in the numerical analysis context and we need to distinguish these two different issues. Constructing u_{es} is not aimed to conserve the Hamiltonian of the system, but to transform the system to another Hamiltonian system by using feedback and shaping the energy of the system (defining the desired Hamiltonian). Therefore, the use of Euler approximation in this context is justified.

From the construction of the controller (3.64), it is obvious that the discrete-time controller is a modification of the controller obtained by emulation of the continuous-time IDA-PBC controller, with the extra term

$$\begin{aligned} Tu_1 &= -G^{-1} M_d M^{-1}\Big(\overline{\nabla}_q H_d(q,p) - \nabla_q H_d(q,p)\Big) \\ &= -TG^{-1} M_d M^{-1} \kappa L_v M^{-1} p \ . \end{aligned} \tag{3.67}$$

Moreover, assuming that $\epsilon > 0$ is of order T, the contribution of the extra term (3.67) to the Lyapunov difference is

$$\Delta V = -T^2 p^\top \overline{M} p + O(\epsilon T^2) + O(T^3) = -T^2 p^\top \overline{M} p + O(T^3) \ , \qquad (3.68)$$

with $\overline{M} := \kappa M^{-1} L_v M^{-1}$ positive semidefinite. Therefore, it is guaranteed that for $\epsilon > 0$ and $T > 0$ sufficiently small, the Lyapunov difference with the discrete-time redesigned controller is more negative than it is with the emulation of the continuous-time controller.

Remark 3.9. It is obvious that this IDA-PBC redesign construction follows the approximate based design framework presented in Section 3.6. The setting we presented in this subsection is a simple illustration when a strict Lyapunov function for Hamiltonian system can be constructed in a systematic way. In a more general situation, especially for the case of underactuated control [26], finding a strict Lyapunov function is still an open problem.

Example 3.5. Consider the nonlinear pendulum shown in Figure 3.2, which is a separable Hamiltonian system with dynamic model given as

$$\dot{q} = p, \quad \dot{p} = -\sin(q) + u \ . \qquad (3.69)$$

The Hamiltonian of this system is

$$H = K(p) + P(q) = \frac{1}{2}p^2 - \cos(q) \ , \qquad (3.70)$$

and the equilibrium point to be stabilized is the origin. By choosing $M_d =$

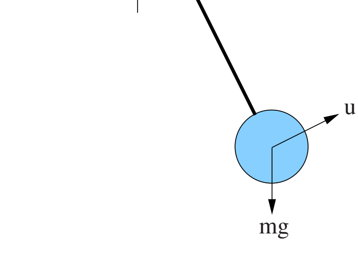

Fig. 3.2. Nonlinear pendulum

$M = I$ and

$$P_d = -\cos(q) + \frac{k_1}{2}q^2 + 1 \ , \quad k_1 \geq 1 \ ,$$

the desired energy function of the system is

$$H_d = K_d(p) + P_d(q) = \frac{1}{2}p^2 - \cos(q) + \frac{k_1}{2}q^2 + 1 \ . \tag{3.71}$$

Applying (3.61) and (3.62), the continuous-time energy shaping and the damping injection controller for system (3.69) are obtained as

$$u_{es}(t) = \nabla_q H - M_d M^{-1} \nabla_q H_d = -k_1 q \ , \tag{3.72}$$

$$u_{di}(t) = -k_v G^\top \nabla_p H_d = -k_v p, \quad k_v > 0 \ . \tag{3.73}$$

Choose the Lyapunov function as

$$V(q,p) = H_d(p,q) + \epsilon q p \tag{3.74}$$

with $\epsilon > 0$ sufficiently small. The Lyapunov derivative is obtained as

$$\begin{aligned} \dot{V}(q,p) &= \dot{H}_d(p,q) + \epsilon(\dot{q}p + q\dot{p}) \\ &= -k_v p^2 + \epsilon(p^2 - q\sin(q) - k_1 q^2 - k_v q p) \\ &\leq -(k_v - \epsilon(1 + \frac{1}{2}k_v))p^2 - \epsilon q \sin(q) - \epsilon(k_1 - \frac{1}{2}k_v)q^2 \ . \end{aligned} \tag{3.75}$$

By choosing k_v and k_1 appropriately, it can be shown that V is a strict AS Lyapunov function for the system (3.69), (3.72), (3.73). Moreover, using Theorem (3.3) we can conclude that the emulation controller $u(k) := u_{es}(k) + u_{di}(k)$ obtained by sample and hold of the continuous-time controller $u(t)$ is a SPA stable controller for the plant (3.69).

Now we redesign the controller (3.72) using Theorem 3.4. Applying (3.64) yields

$$\begin{aligned} u_{es}^T(k) &= \nabla_q H - M_d M^{-1}(\nabla_q H_d + T\kappa L_v M^{-1} p) \\ &= -k_1 q - T\kappa(\cos(q) + k_1)p \ , \end{aligned} \tag{3.76}$$

and (3.65) gives $u_{es}^T(k) = -k_v p$. Applying the discrete-time controller

$$u^T(k) := u_{es}^T(k) + u_{di}^T(k) \tag{3.77}$$

and using the same Lyapunov function (3.74) as in the continuous-time case, we obtain the Lyapunov difference

$$\begin{aligned} \Delta V &:= V(q(k+1), p(k+1)) - V(q(k), p(k)) \\ &\leq -T\left((k_v - \epsilon(1 + \frac{1}{2}k_v))p^2 + \epsilon q \sin(q) + \epsilon(k_1 - \frac{1}{2}k_v)q^2\right) \\ &\quad - T^2\kappa(k_1 + \cos(q))p^2 + O(T^2) \ . \end{aligned} \tag{3.78}$$

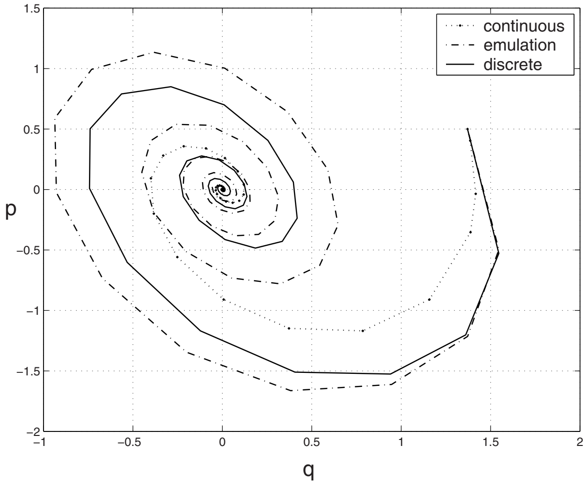

Fig. 3.3. Response of the nonliner pendulum

By choosing k_v, k_1 and κ appropriately, we can show that for sufficiently small $T > 0$ and $\epsilon > 0$, V is a strict SPA Lyapunov function for the Euler model with discrete-time controller.

Taking the trajectory of the continuous-time system as reference, Figure 3.3 shows that applying (3.77) keeps the trajectory of the closed-loop system closer to the reference than using the emulation controller. In the simulation we have used the initial state $(q_\circ, p_\circ) = (\pi/2 - 0.2, 0.5)$, $k_1 = 1$, $k_v = 1$ and $T = 0.35$. Figure 3.4 displays the desired Hamiltonian function when applying only the energy shaping controller to the plant. In continuous-time IDA-PBC, $u_{es}(t)$ conserves the Hamiltonian in closed-loop and hence the closed-loop system is *marginally* stable. Applying the emulation controller $u_{es}(k)$ immediately destroys closed-loop stability. On the other hand, the discrete-time controller $u_{es}^T(k)$ tries to recover Hamiltonian conservation, making the closed-loop system *less unstable* than with $u_{es}(k)$.

Applying each controller to the Euler model of (3.69) and then computing the difference of the Lyapunov differences, we obtain that

$$\Delta V^{u_{es}^T} - \Delta V^{u_{es}} = -T^2\kappa(k_1 + \cos(q))p^2 - \epsilon T^2 \kappa(k_1 + \cos(q))qp + O(T^3) \ . \quad (3.79)$$

Suppose that ϵ is of order T, then we can write

$$\Delta V^{u_{es}^T} - \Delta V^{u_{es}} = -T^2\kappa(k_1 + \cos(q))p^2 + O(T^3) \ , \quad (3.80)$$

which shows that for $\epsilon > 0$ and $T > 0$ sufficiently small, $\Delta V^{u_{es}^T}$ is more negative than $\Delta V^{u_{es}}$ in a practical sense. This explains why the discrete-time controller performs better than the emulation controller.

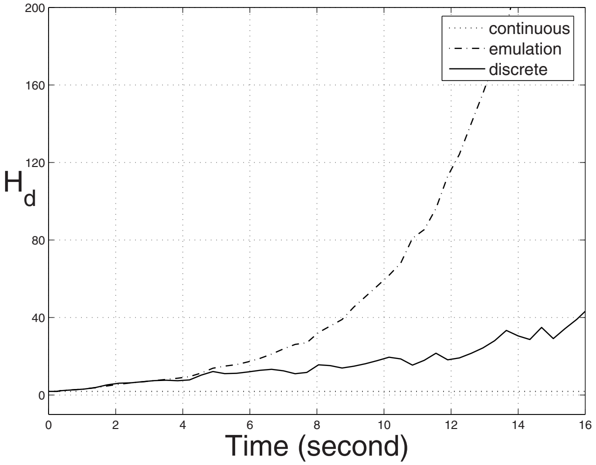

Fig. 3.4. The desired energy function H_d with $k_v = 0$

3.7.4 Backstepping via the Euler Model

Backstepping is a systematic controller design technique for a special class of nonlinear systems in feedback form [24]. The goal is to exploit the special structure of the system to systematically construct a control law u_T for the Euler approximate model of the system and a Lyapunov function V_T that satisfy all conditions of Theorem 3.2 in Section 3.6. Results of this subsection are based on [41].

We consider discrete-time backstepping design based on the Euler model of the system since the Euler model preserves the strict feedback structure of the continuous-time system that is needed in backstepping. Consider the continuous-time system

$$\begin{aligned}\dot{\eta} &= f(\eta) + g(\eta)\xi \\ \dot{\xi} &= u \ .\end{aligned} \tag{3.81}$$

The Euler approximate model of (3.81) has the following form.

$$\eta(k+1) = \eta(k) + T\left[f(\eta(k)) + g(\eta(k))\xi(k)\right] \tag{3.82}$$

$$\xi(k+1) = \xi(k) + Tu(k) \ . \tag{3.83}$$

The main result of this subsection is stated next.

Theorem 3.5. *Consider the Euler approximate model (3.82), (3.83). Suppose that there exists* $\hat{T} \geq 0$ *and a pair* (α_T, W_T) *that is defined for all* $T \in (0, \hat{T})$

and that is a SPA stabilizing pair for the subsystem (3.82), with $\xi \in \mathbb{R}$ regarded as a control. Moreover, suppose that the pair (α_T, W_T) has the following properties.

1. *α_T and W_T are continuously differentiable for any $T \in (0, \hat{T})$.*
2. *there exists $\tilde{\varphi} \in \mathcal{K}_\infty$ such that*
$$|\alpha_T(\eta)| \le \tilde{\varphi}(|\eta|) \ . \tag{3.84}$$
3. *for any $\tilde{\Delta} > 0$ there exist a pair of strictly positive numbers $(\tilde{T}, \tilde{M})$ such that for all $T \in (0, \tilde{T})$ and $|\eta| \le \tilde{\Delta}$ we have*
$$\max\left\{\left|\frac{\partial W_T}{\partial \eta}\right|, \left|\frac{\partial \alpha_T}{\partial \eta}\right|\right\} \le \tilde{M} \ . \tag{3.85}$$

Then, there exists a SPA stabilizing pair (u_T, V_T) for the Euler model (3.82), (3.83). In particular, we can take

$$u_T = -c(\xi - \alpha_T(\eta)) - \frac{\widetilde{\Delta W}_T}{T} + \frac{\Delta \alpha_T}{T} \tag{3.86}$$

where $c > 0$ is arbitrary, and

$$\Delta \alpha_T := \alpha_T(\eta + T(f + g\xi)) - \alpha_T(\eta) \tag{3.87}$$

$$\widetilde{\Delta W}_T := \begin{cases} \frac{\overline{\Delta W}_T}{(\xi - \alpha_T(\eta))}, & \xi \neq \alpha_T(\eta) \\ T\frac{\partial W_T}{\partial \eta}(\eta + T(f + g\xi))g, & \xi = \alpha_T(\eta) \end{cases} \tag{3.88}$$

$$\overline{\Delta W}_T := W_T(\eta + T(f + g\xi)) - W_T(\eta + T(f + g\alpha_T)) \tag{3.89}$$

and the Lyapunov function is

$$V_T(\eta, \xi) = W_T(\eta) + \frac{1}{2}(\xi - \alpha_T(\eta))^2 \ .$$

■

Remark 3.10. The control law (3.86) is in general different from continuous-time backstepping controllers as the next example will illustrate. Interestingly, we show in the next example that our control law can be written in the form

$$u_T^{Euler}(x) = u^{ct}(x) + T u_1^{Euler}(x) \ ,$$

where $u^{ct}(x)$ is a backstepping controller obtained from continuous-time backstepping. We show for the example that u_T^{Euler} yields better performance (better transients and larger domain of attraction) than the emulated backstepping controller $u^{ct}(x)$. While we observed this trend in simulations for any control law designed within our framework, we were unable to prove that this is true in general.

Remark 3.11. Not every backstepping controller that stabilizes the Euler model will stabilize the exact model. Indeed, the dead beat controller in our first motivating example in Section 3.4 can be obtained using backstepping and we saw that it was destabilizing the sampled-data system for all sampling periods T. This further illustrates the importance of using our framework for controller design via approximate discrete-time models.

Example 3.6. [41] We revisit the system in Example 3.3 but now we want to use Theorem 3.5 based on the Euler model (3.51) of the system (3.46).

Again, the control law $\phi(\eta) = -\eta^2 - \eta$ globally asymptotically stabilizes the η-subsystem of (3.51) with the Lyapunov function $W(\eta) = \frac{1}{2}\eta^2$. Using the construction in Theorem 3.5, we obtain the controller

$$u_T^{Euler}(\eta, \xi) = u^{ct}(\eta, \xi) + T u_1^{Euler} \ , \tag{3.90}$$

where $u^{ct}(x)$ is the same as in Examples 3.3 and 3.4 and the following u_1^{Euler} is obtained as

$$u_1^{Euler} = -\frac{1}{2}(\xi - \eta + \eta^2) - (\xi + \eta^2)^2 \ . \tag{3.91}$$

From Theorem 3.5 we see that u_T^{Euler} SPA stabilizes the Euler model (3.51). This can be proven with the Lyapunov function $V(\eta, \xi) = \frac{1}{2}\eta^2 + \frac{1}{2}(\xi + \eta + \eta^2)^2$. Hence, using Theorem 3.2 we conclude that the same controller SPA stabilizes the exact model and consequently the sampled-data system.

Next we compare the controller (3.90) with the controllers that were designed in Examples 3.3 and 3.4. First, note that the terms u_1 in (3.53) and u_1^{Euler} in (3.91) are different. Moreover, all controllers become the same and equal to $u^{ct}(x)$ for $T = 0$. Hence, it makes sense to compare the controllers for small T.

Figure 3.5 shows the time response of the system (3.81) when applying respectively the emulation controller, the redesigned controller and the Euler based discrete-time controller. The response using continuous-time controller is used as reference. In the simulation, we set $x_\circ = (2 \;\; 2)^\top$ and $T = 0.5$. It is shown that the emulation controller destabilizes the system, whereas the redesign and the Euler based controllers maintain the response of the system relatively close to the continuous-time response.

In Figure 3.6 we show the simulation result when we increase the initial state to $x_\circ = (400 \;\; 400)^\top$. We do not plot the response of the system with emulation controller since it is obviously unstable. Interestingly, with the redesign controller and the Euler based controller, the stability of the closed-loop system is preserved although the initial state in this simulation is 200 times larger than the one used in Figure 3.5. In fact, for these two controllers, the stability is still maintained for larger initial states in any direction in the state space.

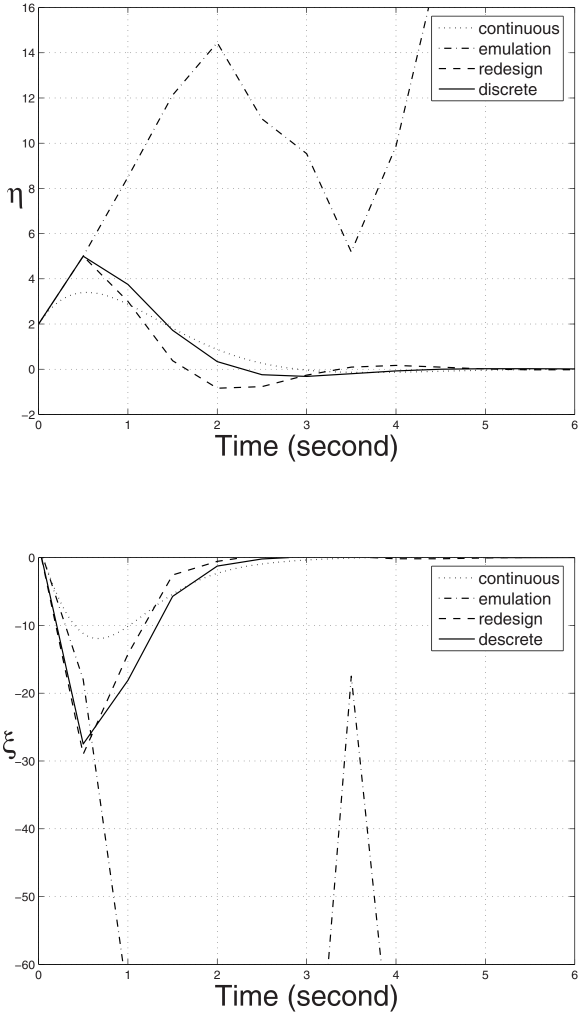

Fig. 3.5. Closed-loop responses of the system (3.81) for small initial states

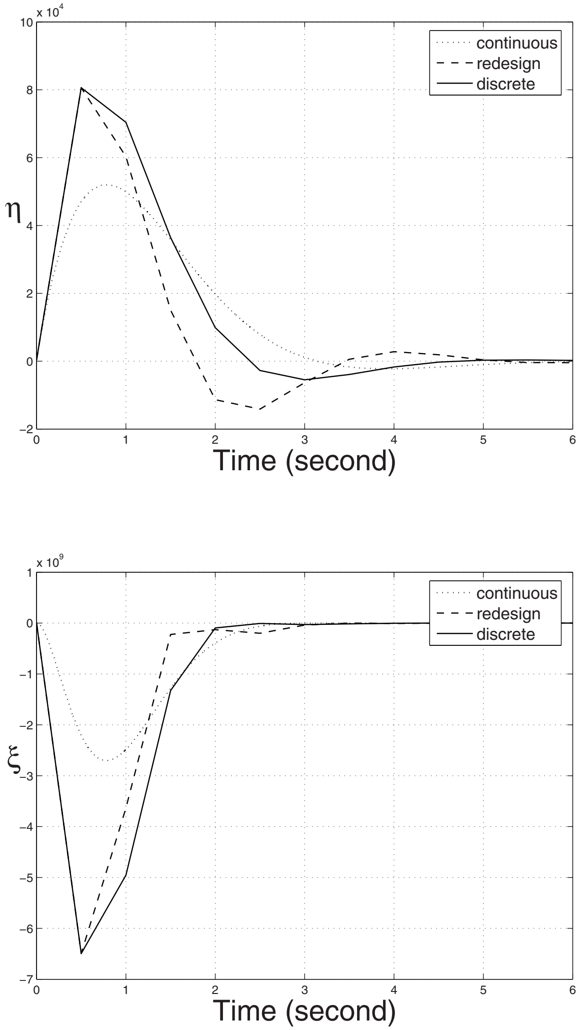

Fig. 3.6. Closed-loop response of the system (3.81) for large initial states

3.8 Design Examples

In this section two examples are presented to illustrate the various design tools we have discussed in Section 3.7. It will also be shown that the design fits with the framework proposed in Section 3.6. In the first example a jet engine system is considered, and the emulation and the Euler based backstepping design are applied to solve the stabilization problem of the jet engine. In the second example, a stabilization design for an inverted pendulum is studied. A backstepping design and an IDA-PBC design are applied to the system. Simulation results are presented to show the performance of each controller designed.

3.8.1 Jet Engine System

A simplified Moore-Greitzer model of a jet engine with the assumption of no stall is given by

$$\begin{aligned} \dot{x}_1 &= -x_2 - \frac{3}{2}x_1^2 - \frac{1}{2}x_1^3 \\ \dot{x}_2 &= -u \ , \end{aligned} \tag{3.92}$$

where x_1 and x_2 are respectively related to the mass flow and the pressure rise through the engine after an appropriate change of coordinates (see [24] for more details). We will apply both the continuous-time and the Euler based backstepping design discussed in Subsection 3.7.4 to this system, and compare the performance of the controller obtained by the Euler based backstepping design with the one obtained by emulation of the continuous-time controller.

Choose $\phi(x_1) = -\frac{3}{2}x_1^2 + x_1$ and $W(x_1) = \frac{1}{2}x_1^2$. Applying [24, Lemma 2.8] and choosing $c = 1$, the continuous-time controller is obtained as

$$u^{ct}(x_1, x_2) = -x_1 + c(x_2 + \frac{3}{2}x_1^2 + \frac{1}{2}x_1^3) + (3x_1 - 1)(-x_2 - \frac{3}{2}x_1^2 - \frac{1}{2}x_1^3) \ . \tag{3.93}$$

Moreover, using the Euler approximate model of (3.92) and applying Theorem 3.5, we obtain the discrete-time Euler-based controller

$$u_T^{Euler}(x_1, x_2) = u^{ct}(x_1, x_2) + Tu_1(x_1, x_2) \ , \tag{3.94}$$

where

$$u_1(x_1, x_2) = \frac{1}{2}(x_2 + x_1 + \frac{3}{2}x_1^2 + x_1^3) \ .$$

We implement the controller (3.94) and the discrete-time emulation of (3.93) to control the continuous-time plant (3.92), comparing the performance. The simulation results with parameters $c = 1$, $x_\circ = (2, \ 3)^\top$ and $T = 0.2$ are illustrated in Figure 3.7.

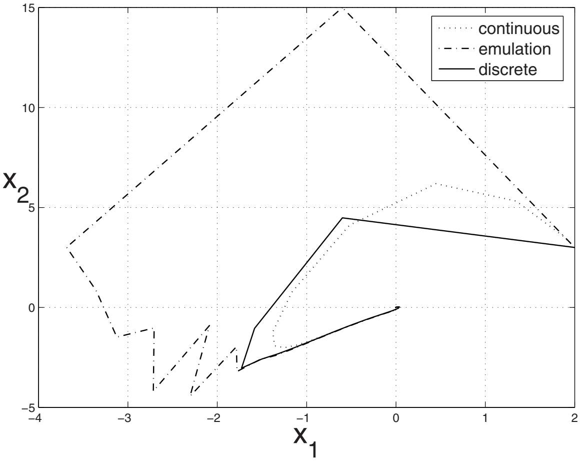

Fig. 3.7. Phase plot of the jet engine system

It is shown that the Euler-based controller outperforms the emulation controller and for the chosen simulation parameters, it keeps the response of the closed-loop system close to the response of the continuous-time closed-loop system.

3.8.2 Inverted Pendulum

Consider a nonlinear dynamic model of an inverted pendulum as illustrated in Figure 3.8, namely

$$\begin{aligned}\dot{q} &= p\\ \dot{p} &= \sin(q) + u\ .\end{aligned} \tag{3.95}$$

This dynamic model is in strict feedback form. This system also belongs to the class of separable Hamiltonian systems with the Hamiltonian function

$$H = \frac{1}{2}p^2 + \cos(q)\ . \tag{3.96}$$

Therefore, we can apply both the backstepping design and the IDA-PBC redesign to construct a stabilizing controller for this system. Note that $q = 0$ is the equilibrium point of this system and is an unstable equilibrium. The control design in this case is aiming at stabilizing this equilibrium point.

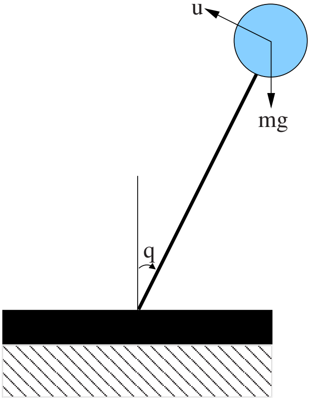

Fig. 3.8. Inverted pendulum

Backstepping Design

Choose $\phi(q) = -q$ and $W(q) = \frac{1}{2}q^2$. The continuous-time controller is obtained as

$$u^{ct}(q,p) = -\sin(q) - (1+c)(p+q) \ . \tag{3.97}$$

Applying Theorem 3.5, we design a discrete-time controller for the Euler approximate model of the system (3.95)

$$\begin{aligned} q(k+1) &= q(k) + Tp(k) \\ p(k+1) &= p(k) + T(\sin(q) + u) \ . \end{aligned} \tag{3.98}$$

We obtain the Euler based controller

$$u_T^{Euler}(q,p) = u^{ct}(q,p) + Tu_1(q,p) \ , \tag{3.99}$$

with

$$u_1(q,p) = -\frac{1}{2}(p-q) \ .$$

Implementing the controller (3.99) and the discrete-time emulation of (3.97) to the continuous-time plant (3.95), we compare the performance of the two controllers, using the continuous-time controller performance as reference. The simulation results with parameters $c = 1$, $(q_\circ,\ p_\circ) = (\frac{\pi}{2} - 0.2,\ \frac{1}{2})$ and $T = 0.5$ are displayed in Figures 3.9 and 3.10.

IDA-PBC Redesign

We apply the results discussed in Subsection 3.7.3 to design a stabilizing controller for the inverted pendulum (3.95). From the Hamiltonian (3.96) we have that $M = I$ and $P = \cos(q)$. To bring the energy to the minimum level at the equilibrium point, we assign a new energy function H_d for the pendulum, by keeping $M_d = M = I$ and choosing the new potential energy $P_d = -\cos(q) + \frac{1}{2}k_1 q^2 + 1$. Hence,

$$H_d = \frac{1}{2}p^2 - \cos(q) + \frac{1}{2}k_1 q^2 + 1 \ . \tag{3.100}$$

Using the continuous-time IDA-PBC design, we obtain the controller

$$u_{ct}(q,p) = u_{es}(q,p) + u_{di}(q,p) \ , \tag{3.101}$$

with

$$u_{es}(q,p) = -2\sin(q) - k_1 q \tag{3.102}$$
$$u_{di}(q,p) = -k_v p \ . \tag{3.103}$$

With this controller, we obtain

$$\dot{H}_d = -k_v p^2 \ . \tag{3.104}$$

Utilizing La Salle Invariance Principle we can show that the closed-loop approximate model is asymptotically stable. Moreover, using Theorem 3.3, we can conclude that the discrete-time controller obtained by emulation of (3.101) is a SPA stabilizing controller for the inverted pendulum (3.95).

Moreover, using Theorem 3.4 we will redesign the emulation controller, to improve the performance of the system. Consider the Euler model (3.98) and applying (3.64) and (3.65), the redesigned controller is obtained as

$$\begin{aligned} u_{dt}^T(q,p) &= u_{ct}(q,p) + T u_1(q,p) \ , \\ u_1(q,p) &= -G^{-1}M_d M^{-1}\kappa L_V(q) M^{-1}p = -\kappa(\cos(q) + k_1)p \ . \end{aligned} \tag{3.105}$$

While in the continuous-time design we can use the desired Hamiltonian as a Lyapunov function and utilize La Salle Invariance Principle to conclude stability of the continuous-time system, the same approach cannot be applied in this controller redesign. In order to apply the framework provided in Theorem 3.2 a strict Lyapunov function for the closed-loop approximate model is required, whereas the desired Hamiltonian does not satisfy this. For that we need to construct a strict Lyapunov function applying (3.66), and we choose such function to be

$$V(q,p) = H_d(q,p) + \epsilon q p \ . \tag{3.106}$$

Applying the controller (3.105) to stabilize the Euler model (3.98), the Lyapunov difference is obtained as

$$\begin{aligned} \Delta V &:= V(q(k+1), p(k+1)) - V(q(k), p(k)) \\ &\leq -T\left((k_v - \epsilon(1 + \frac{1}{2}k_v))p^2 + \epsilon q \sin(q) + \epsilon(k_1 - \frac{1}{2}k_v)q^2\right) \\ &\quad - T^2\kappa(k_1 + \cos(q))p^2 + O(T^2) \ . \end{aligned} \tag{3.107}$$

By choosing k_v, k_1 and κ appropriately, we can show that for sufficiently small $T > 0$ and $\epsilon > 0$, V is a strict SPAS Lyapunov function for the Euler model with the discrete-time controller. Moreover, using Theorem 3.6 we can conclude SPA stability of the exact model and the sampled-data system (3.95), (3.105). The sampled-data simulation results with parameters $k_1 = 1$, $k_v = 2$, $\kappa = 1$, $(q_\circ,\ p_\circ) = (\frac{\pi}{2} - 0.2,\ \frac{1}{2})$ and $T = 0.5$ are illustrated in Figures 3.9 and 3.10.

3.9 Overview of Related Literature

The results we have presented in the earlier sections are only the basic of research that has been done in the topic covered by this chapter. Indeed there is a lot more research done in parallel directions, for both direct discrete-time design and emulation (re)design. In this section, we present an overview of works in the topics related to what we have discussed in this chapter. We emphasize that this section does not serve as a complete review and will not cover all available results in literature. For that, at the first place we apologize to other authors whose works are not cited.

A similar and more general design framework than what has been provided in Section 3.6 is presented in [45]. This framework uses trajectory based analysis and instead of using one step consistency, a multistep consistency property is utilized. More general design frameworks are presented in [43] where nonlinear systems represented as differential inclusion are considered, and in [39] where nonlinear systems with exogenous inputs are studied.

Recently, researchers have started to build design tools within the various frameworks mentioned above. Designs exploring model predictive control or receding horizon techniques are presented in [20, 37].

Although the frameworks consider only time invariant systems, the extension to time-varying systems is straightforward. Results presented in [40] on asymptotic stabilization for time-varying cascaded systems and in [27] on input-to-state stabilization of systems in power form using time varying control are examples of this extension.

A problem that one may face in applying the framework is that it requires the knowledge of a strict Lyapunov function for the system. While for linear systems a strict Lyapunov function is available for free, in the sense that a quadratic Lyapunov function can always be used, it is not the case for nonlinear systems in general. Moreover, when the controller is designed based on an approximate model, powerful tools to analyze stability, either SP-AS or SP-ISS, for continuous-time systems, such as La Salle Invariance Principle and Matrosov Theorem are not directly applicable for sampled-data systems when stability is attained in a semiglobal practical sense (see discussion in Subsection 3.8.2). Hence results from [42] that provide a partial construction

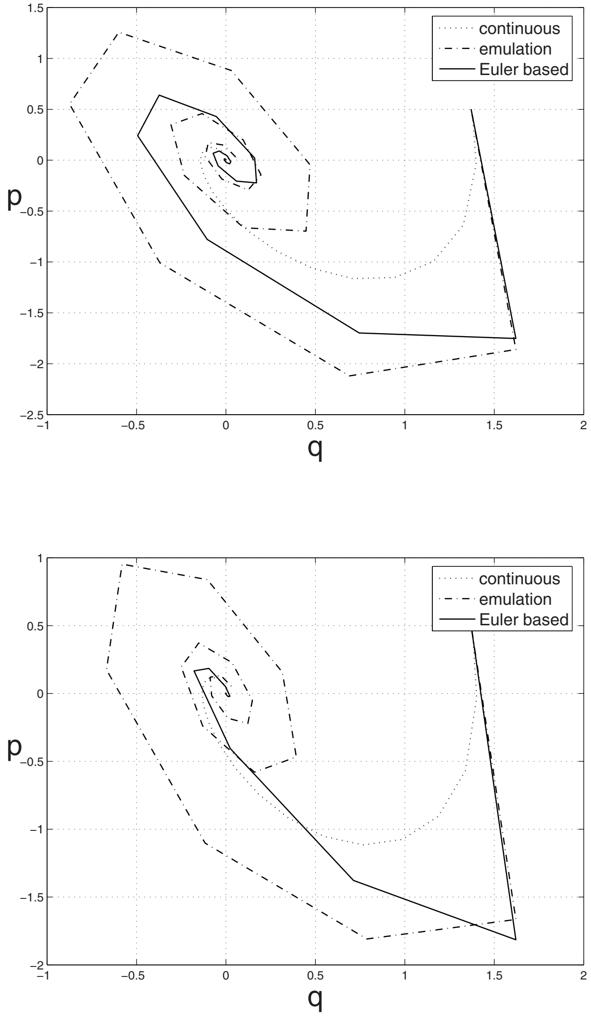

Fig. 3.9. Response of the inverted pendulum with backstepping (above) and IDA-PBC (below) controllers

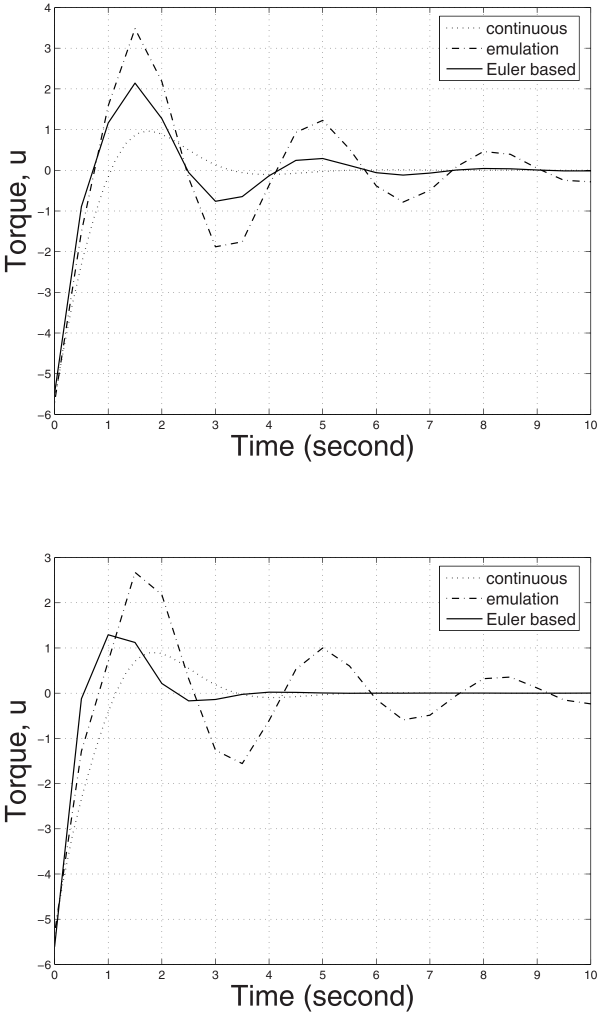

Fig. 3.10. Control signals for the inverted pendulum designed using backstepping (above) and IDA-PBC (below)

of Lyapunov functions, that in some sense generalizes the construction used in Theorem 3.4 are very useful to replace La Salle Invariant Principle. In [28] a Lyapunov function construction for interconnected systems is proposed utilizing a nonlinear small gain theorem. In [44] a result similar to Matrosov theorem is developed.

There are more research and studies related to the topic presented in this chapter that follow a different framework. Approaches using feedback linearization are discussed for instance in [3, 18] and references therein. A geometric framework for feedback linearization is utilized in [8, 11]. Singular perturbation is used as the main tool to solve sampled-data control problems in [7, 9]. Adaptive control approach based on Euler model is used in [31] and robust stabilization using discontinuous control is studied in [22] (see also references therein).

While we only consider static state feedback in this chapter, assuming the availability of all states is sometimes not realistic. The issue of observability, as well as controllability, of discrete-time systems is studied in [55, 57]. Results on discrete-time controller design and stabilization using output feedback are presented for instance in [5, 10, 14, 54]. A framework for designing a discrete-time observer based on the approximate model of the plant is presented in [4]. When implementing the observer to build a controller for the plant, this result can also be considered as a framework for designing a dynamic feedback. This framework can be seen as an extension of the controller design framework presented in Section 3.6.

Due to the increasing interest of research on nonlinear sampled-data control systems, the list of related literature will always grow longer. What we have cited in this section is in any way not a complete list of reference but just a glimpse of available results on various directions that aims to help readers to see the variety and fertility of research in this topic. Interested readers may find in the cited papers, many other references that we have not included in this section.

3.10 Open Problems

There is a wide range of open research problems that one could address.

- Constructive designs for classes of nonlinear systems and their approximate models need to be further developed within our framework. Any continuous-time design technique can be revisited within our framework. If the Euler model is used for design, then the structure of the approximate model is the same as the structure of the continuous-time system and in this case the discrete-time design is easier. However, if higher order approximate models are used for controller design then the structure of the approximate discrete-time model may be very different from the structure of the continuous-time model and design becomes harder. In this case, it seems more natural to use model predictive control that does not exploit the structure of the model to design the controller.

- Case studies and practical implementations of our algorithms are needed to motivate new theoretical issues in this area and to assess the developed theory in practice.
- The quantitative relationship between the choice of approximate model used in design and the performance of the obtained controller is unclear. There is an obvious tradeoff between the complexity of the controller design and the accuracy of the approximations. Typically, the design is easiest for the Euler model but we expect that better performance could be obtained if a better approximation was used for controller design. Understanding this possible improvement in performance appears to be an important issue.
- Obtaining non-conservative estimates of T^* in our theorems would be quite useful for practicing engineers since choosing an appropriate T is an important step in our approach. While we do compute T^* in our proofs, our estimates are very conservative and, hence, not useful in practice. We are not aware of any papers that attempt to address this problem.
- In the presented results, so far we use full state feedback, assuming that all states are available for measurement. In reality, this is not always the case due to the physical meaning of the states or the available sensors and measurement devices may be too expensive. To overcome this situation, observer design and developing results based on output feedback are potential solutions.

References

1. J. A. Acosta, R. Ortega, and A. Astolfi. Interconnection and damping assignment passivity-based control of mechanical systems with underactuation degree one. In *Proc. 6th IFAC NOLCOS*, Stuttgart, 2004.
2. G. L. Amicucci, S. Monaco, and D Normand-Cyrot. Control Lyapunov stabilization of affine discrete-time systems. In *Proc. IEEE Conf. Decis. Contr.*, pages 923–924, San Diago, CA, 1997.
3. A. Arapostathis, B. Jacubczyk, H. G. Lee, S. I. Marcus, and E. D. Sontag. The effect of sampling on linear equivalence and feedback linearization. *Systems and Control Letters*, 13:373–381, 1989.
4. M. Arcak and D. Nešić. A framework for nonlinear sampled-data observer design via approximate discrete-time models and emulation. *Automatica*, 40:1931–1938, 2004.
5. A. El Assoudi, E. H. El Yaagoubi, and H. Hammouri. Nonlinear observer based on the euler discretization. *International Journal of Control*, 75:784–791, 2002.

6. K. J. Aström and B. Wittenmark. *Computer-Controlled System, Theory and Design.* PHI, 1997.
7. J. P. Barbot, M. Djemai, S. Monaco, and D. Normand-Cyrot. Analysis and control of nonlinear singularly perturbed systems under sampling. In C. T. Leondes, editor, *Control and Dynamic Systems: Advances in Theory and Application*, volume 79, pages 203–246. Academic Press, San Diego, 1996.
8. J. P. Barbot, S. Monaco, and D. Normand-Cyrot. A sampled normal form for feedback linearization. *Math. of Control, Signals and Systems*, 9:162–188, 1996.
9. J. P. Barbot, N. Pantalos, S. Monaco, and D. Normand-Cyrot. On the control of regularly perturbed nonlinear systems. *Int. Journal of Control*, 59:1255–1279, 1994.
10. M. Boutayeb and M. Darouach. A reduced-order observer for nonlinear discrete-time systems. *Syst. Cont. Lett.*, 39:141–151, 2000.
11. C. Califano, S. Monaco, and D. Normand-Cyrot. On the problem of feedback linearization. *Syst. Cont. Lett.*, 36:61–67, 1999.
12. T. Chen and B. A. Francis. Input-output stability of sampled-data systems. *IEEE Trans. Automat. Contr.*, 36:50–58, 1991.
13. T. Chen and B. A. Francis. *Optimal Sampled-Data Control Systems.* Springer-Verlag, London, 1995.
14. A. M. Dabroom and H. K. Khalil. Output feedback sampled-data control of nonlinear systems using high-gain observers. *IEEE Trans. Automat. Contr.*, 46:1712–1725, 2001.
15. D. Dochain and G. Bastin. Adaptive identification and control algorithms for nonlinear bacterial growth systems. *Automatica*, 20:621–634, 1984.
16. T. Fliegner. *Contributions to The Control of Nonlinear Discrete-Time Systems.* PhD Thesis, Department of Applied Mathematics, University of Twentee, 1995.
17. G. C. Goodwin, B. McInnis, and R. S. Long. Adaptive control algorithm for waste water treatment and pH neutralization. *Optimal Contr. Applic. Meth.*, 3:443–459, 1982.
18. J. W. Grizzle and P.V. Kokotović. Feedback linearization of sampled-data systems. *IEEE Trans. Auto. Contr.*, 33:857–859, 1988.
19. A. M. Guillaume, G. Bastin, and G. Campion. Sampled-data adaptive control of a class of continuous nonlinear systems. *Int. Journal of Contr.*, 60:569–594, 1994.
20. E. Gyurkovics and A. M. Elaiw. Stabilization of sampled-data nonlinear systems by receding horizon control via discrete-time approximations. *Automatica*, 40:2017–2028, 2004.
21. C. M. Kellet. *Advances in Converse and Control Lyapunov Function, PhD Thesis.* The University of California, Santa Barbara, 2002.
22. C. M. Kellet, H. Shim, and A. R. Teel. Further results on robustness of (possibly discontinuous) sample and hold feedback. *IEEE Trans. on Automatic Control*, 49:1081–1089, 2004.
23. H. K. Khalil. *Nonlinear Control Systems 2nd Ed.* Prentice Hall, 1996.
24. M. Krstić, I. Kanellakopoulos, and P. V. Kokotović. *Nonlinear and Adaptive Control Design.* Wiley, 1995.
25. B. C. Kuo. *Digital Control Systems.* Saunders College, 1992.
26. D. S. Laila and A. Astolfi. Discrete-time IDA-PBC design for separable of Hamiltonian systems. In *Proc. 16th IFAC World Congress*, Prague, 2005.

27. D. S. Laila and A. Astolfi. Input-to-state stability for discrete-time time-varying systems with applications to robust stabilization of systems in power form. *Automatica*, 41:1891–1903, 2005.
28. D. S. Laila and D. Nešić. Lyapunov based small-gain theorem for parameterized discrete-time interconnected ISS systems. *IEEE Trans. on Automatic Control*, 48:1783–1788, 2004.
29. D. S. Laila, D. Nešić, and A. R. Teel. Open and closed loop dissipation inequalities under sampling and controller emulation. *European Journal of Control*, 8:109–125, 2002.
30. V. Lakshmikantham and S. Leela. *Differential and integral inequalities.* Academic Press, New York, 1969.
31. I. M. Y. Mareels, H. B. Penfold, and R. J. Evans. Controlling nonlinear time-varying systems via Euler approximations. *Automatica*, 28:681–696, 1992.
32. R. H. Middleton and G. C. Goodwin. *Digital control and estimation : a unified approach.* Prentice Hall, 1990.
33. S. Monaco and D. Normand-Cyrot. On the conditions of passivity and losslessness in discrete-time. In *Proc. European Control Conference*, Brussels, 1997.
34. D. S. Naidu and A. K. Rao. *Singular Perturbation Analysis of Discrete Control Systems.* Springer Verlag, New York, 1985.
35. D. Nešić and D. Angeli. Integral versions of ISS for sampled-data nonlinear systems via their approximate discrete-time models. *IEEE Trans. Auto. Contr.*, 47:2033–2037, 2002.
36. D. Nešić and L. Grune. Lyapunov based continuous-time nonlinear controller redesign for sampled-data implementation. *Automatica*, 41:1143–1156, 2005.
37. D. Nešić and L. Grune. A receding horizon control approach to sampled-data implementation of continuous-time controller. *Systems and Control Letters*, to appear, 2005.
38. D. Nešić and D. S. Laila. Input-to-state stabilization for nonlinear sample-data systems via approximate discrete-time plant models. In *Proc. 40th IEEE Conf. Decis. Contr.*, pages 887–892, Orlando, FL, 2001.
39. D. Nešić and D. S. Laila. A note on input-to-state stabilization for nonlinear sampled-data systems. *IEEE Trans. Auto. Contr.*, 47:1153–1158, 2002.
40. D. Nešić and A. Loria. On uniform asymptotic stability of time-varying parameterized discrete-time cascades. *IEEE Trans. Auto. Contr.*, 49:875–887, 2004.
41. D. Nešić and A. R. Teel. Backstepping on the Euler approximate model for stabilization of sampled-data nonlinear systems. In *Proc. IEEE CDC.*, pages 1737–1742, Orlando, FL, 2001.
42. D. Nešić and A. R. Teel. Changing supply functions in input to state stable systems: The discrete-time case. *IEEE Trans. Auto. Contr.*, 46:960–962, 2001.
43. D. Nešić and A. R. Teel. A framework for stabilization of nonlinear sampled-data systems based on their approximate discrete-time models. *IEEE Trans. Auto. Contr.*, 49:1103–1122, 2004.
44. D. Nešić and A. R. Teel. Matrosov theorem for parameterized families of discrete-time systems. *Automatica*, 40:1025–1034, 2004.
45. D. Nešić, A. R. Teel, and P. Kokotović. Sufficient conditions for stabilization of sampled-data nonlinear systems via discrete-time approximations. *Syst. Contr. Lett.*, 38:259–270, 1999.
46. D. Nešić, A. R. Teel, and E. Sontag. Formulas relating $\mathcal{KL}$ stability estimates of discrete-time and sampled-data nonlinear systems. *Syst. Contr. Lett.*, 38:49–60, 1999.

47. R. Ortega and E. Garcia-Canseco. Interconnection and damping assignment passivity-based control: A survey. *European J. of Control*, 10, No. 5, 2004.
48. R. Ortega, M. W. Spong, F. Gomez-Estern, and G. Blankenstein. Stabilization of a class of underactuated mechanical systems via interconnection and damping assignment. *IEEE TAC*, 47:1218–1233, 2002.
49. R. Ortega, A. van der Schaft, I. Mareels, and B. Maschke. Putting energy back in control. *IEEE Contr. Syst. Mag.*, 21:18–33, 2001.
50. R. Ortega, A. van der Schaft, B. Maschke, and G. Escobar. Interconnection and damping assignment passivity-based control of port-controlled Hamiltonian systems. *Automatica*, 38:585–596, 2002.
51. J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian Problems*. Chapman & Hall, 1994.
52. S. Sastry. *Nonlinear Systems. Analysis, Stability and Control.* Springer, 1999.
53. R. Sepulchre, M. Janković, and P. V. Kokotović. *Constructive Nonlinear Control.* Springer, London, 1997.
54. H. Shim, J. H. Seo, and A. R. Teel. Nonlinear observer design via passivation of error dynamics. *Automatica*, 39:885–892, 2003.
55. H. Shim and A. R. Teel. Asymptotic controllability and observability imply semiglobal practical asymptotic stabilizability by sampled-data output feedback. *Automatica*, 39:441–454, 2003.
56. C. Simoes. *On Stabilization of Discrete-Time Nonlinear Systems, PhD Thesis.* Department of Applied Mathematics, University of Twente, 1996.
57. E. D. Sontag. A concept of local observability. *Systems and Control Letters*, 5:41–47, 1984.
58. E. D. Sontag. Smooth stabilization implies coprime factorization. *IEEE Trans. Automat. Contr.*, 34:435–443, 1989.
59. A. M. Stuart and A. R. Humphries. *Dynamical Systems and Numerical Analysis.* Cambridge Univ. Press, New York, 1996.
60. M. P. Tzamtzi and S. G. Tzafestas. A small gain theorem for locally input to state stable interconnected systems. *Journal of The Franklin Institute*, 336:893–901, 1999.

4

Stability Analysis of Time-delay Systems: A Lyapunov Approach

Kequin Gu[1] and Silviu-Iulian Niculescu[2]

[1] Department of Mechanical and Industrial Engineering, Southern Illinois University Edwardsville, Edwardsville, Illinois 62026-1805, USA
E-mail: kgu@siue.edu

[2] HEUDIASYC (UMR CNRS 6599), Université de Technologie de Compiègne
Centre de Recherche de Royallieu, BP 20529 60205 Compiègne, France
E-mail: Silviu.Niculescu@hds.utc.fr

Summary. This chapter is devoted to the stability problem of time-delay systems using time-domain approach. Some basic concepts of time-delay systems are introduced. Then, some simple Lyapunov-Krasovskii funtionals, complete Quadratic Lyapunov-Krasovskii functional and discretization scheme are introduced, with connections and extent of conservatism compared. The issue of time-varying delays are also discussed. The concept of Razumikhin Theorem is introduced. An alternative model of coupled difference-differential equations and its stability problem are also introduced.

4.1 Introduction

It is a common pratice to use ordinary differential equations to describe the evolution of physical, engineering or biological system. However, it is also known that such a mathematical description is inadequate for many systems. Indeed, delay-differential equations (or more generally, functional differential equations) are often needed to reflect the fact that the future evolution of system variables not only depends on their current values, but also depends on their past history. Such systems are often known as time-delay systems (also known as hereditary systems, systems with time lag, or systems with aftereffects). This chapter is intended to serve as a tutorial to cover some of the basic ideas of time-delay systems, especially, the stability analysis using Lyapunov approach.

Time-delay systems are distributed parameter systems, or infinite-dimensional systems. To bring out the idea, compare the ordinary differential equation

$$\dot{x}(t) = ax(t), \tag{4.1}$$

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 139–170, 2006.

with a simple time-delay system

$$\dot{x}(t) = ax(t - r). \tag{4.2}$$

In these two systems, a and r are constant scalars, and x, a scalar function of time t, is the state variable. It is well known that for the system represented by (4.1), given any time t_0, then the future value of the state $x(t)$, $t > t_0$ is completely determined by $x(t_0)$, a scalar, which indicates that the system (4.1) is a 1-dimensional system. On the other hand, for the system (4.2), to completely determine $x(t)$, $t > t_0$, it is necessary to know $x(t)$ for all $t_0 - r \leq t \leq t_0$. Therefore, the *state* at time t_0 is an element of the infinite-dimensional functional space $\{x(t) \mid t_0 - r \leq t \leq t_0\}$, and the system (4.2) is an infinite-dimensional system.

Examples of time-delay systems abound in various disciplines of science, engineering and mathematics. Kolmanovskii and Myshkis gave many examples [18]. Other books also contain many practical examples, see, for example, [11] [13] [23]. Here, we will mention only two examples.

Example 4.1. **Network control**. The popularity of internet has brought to the network control problem to prominence. One of the model studied in the literature is the simplified fluid approximation proposed by Kelly [17]

$$\dot{x}(t) = k[w - x(t - \tau)p(x(t - \tau))],$$

where p is a continuously differentiable and strictly increasing function bounded by 1, and k and w are positive constant. The delay τ represents the round-trip time. The function p can be interpreted as the fraction of packets the presence of (potential) congestion. For more details of network model, see [5] [17] [28].

Example 4.2. **Transport delay in chemical reactions.** This example was discussed in [20] and [21]. Consider a first order, exothermic and irreversible chemical reaction from A to B. In practice, the conversion from A to B is not complete. To increase the conversion rate and reduce the costs, a recycle stream is used. The time it takes to transport from the output to the input introduces time delay. The resulting process can be described by the following equations

$$\begin{aligned}
\frac{dA(t)}{dt} &= \frac{q}{V}[\lambda A_0 + (1 - \lambda)A(t - \tau) + A(t)] - K_0 e^{-Q/T} A(t) \\
\frac{dT(t)}{dt} &= \frac{1}{V}[\lambda T_0 + (1 - \lambda)T(t - \tau) - T(t)]\frac{\Delta H}{C\rho} - K_0 e^{-Q/T} A(t) \\
&\quad - \frac{1}{VC\rho}U(T(t) - T_w,)
\end{aligned}$$

where $A(t)$ is the concentration of the component A, $T(t)$ is the temperature, and $\lambda \in [0, 1]$ is the recycle coefficient ($\lambda = 1$ represents no recycle), and τ is the transport delay. The case without time delay τ has been discussed in [2] and [26].

The rest of the chapter is organized as follows. Section 4.2 introduces some basic concepts of time-delay systems. Section 4.3 introduces the concept of stability and Lyapunov-Krasovskii stability Theorem. Section 4.4 introduces some simple Lyapunov-Krasovskii functionals Section 4.5 covers the complete quadratic Lyapunov-Krasovskii functional and its discratization. Section 4.6 compares different Lyapunov functionals, with numerical examples. Section 4.7 discusses time-varying delays. Section 4.8 discusses Razumikhin Theorem. Section 4.9 discusses coupled difference-differential equations and stability. Section 4.10 contains conclusions and discussions.

4.2 Basic Concepts of Time-delay Systems

4.2.1 Systems of Retarded Type

We will concentrate on time-delay systems of *retarded type* in this article. A retarded time-delay system can be represented as

$$\dot{x}(t) = f(t, x_t) \tag{4.3}$$

where $x(t) \in \mathbb{R}^n$, x_t is a function defined in the interval $[-r, 0]$ as

$$x_t(\theta) = x(t+\theta), \qquad -r \leq \theta \leq 0,$$

r is the maximum delay, and f is a *functional*, or a function of functions. In other words, the value of f can be determined if the value of t and the function x_t are given. It is common practice to restrict x_t to be a continuous function. Let $\mathcal{C}$ be the set of all continuous functions defined in the interval $[-r, 0]$, then the initial condition of (4.3) can be expressed as

$$x_{t_0} = \phi, \text{ for some } \phi \in \mathcal{C} \tag{4.4}$$

which means that

$$x(t_0 + \theta) = \phi(\theta), \text{ for } \theta \in [-r, 0].$$

With this notation, the domain of definition of f is $\mathbb{R}\times\mathcal{C}$. The solution of (4.3) with initial condition (4.4) is often denoted as $x(t, t_0, \phi)$, or $x(t, \phi)$ if t_0 is understood.

In some context, it is beneficial to consider the initial condition as consisting of two parts, $x(t_0)$ and $x(t)$ for $t_0 - r \leq t < t_0$. This may be convenient to accommodate the case of a discontinuous ϕ in the initial condition.

Other types of time-delay systems are discussed in, for example, [13] and [18]. For example, if $\dot{x}(t)$ also depends on derivative of x at a time $\tau < t$, then the system is of *neutral type*.

4.2.2 Pointwise Delays

An important special case in pratice can be expressed as

$$\dot{x}(t) = f(t, x(t), x(t-r)). \tag{4.5}$$

In other words, $\dot{x}(t)$ only depends on x at current time and at the time of maximum delay, and is independent of $x(t+\theta)$, $-r < \theta < 0$. Let's consider the case of $t_0 = 0$, so that the initial condition becomes

$$x_0 = \phi, \text{ or } x(t) = \phi(t), \; -r \leq t \leq 0.$$

Such a system admits a simple *method of steps* to generate the future trajectories: Since $x(t-r)$ is already known as the initial condition for $t \in [0, r]$, the equation (4.5) can be considered as an ordinary differential equation in this interval, and $x(t)$, $t \in [0, r]$ can be generated by solving this ordinary differential equation. Once $x(t)$, $t \in [0, r]$ is available, $x(t-r)$, $t \in [r, 2r]$ is also available, and therefore, one can further generate $x(t)$, $t \in [r, 2r]$ by solving ordinary differential equation. Continue this process will allow us to generate $x(t)$ for $t \in [0, \infty)$.

Similarly, we say the system

$$\dot{x}(t) = f(t, x(t), x(t-r_1), x(t-r_2), ..., x(t-r_k)) \tag{4.6}$$

is of *multiple delays*. Furthermore, if there is a common factor τ which divides all delays r_j, $j = 1, 2, ..., k$, then we say the system is of *commensurate delays*. Without loss of generality, we may assume $r_j = j\tau$ in this case. If there does not exist such a factor, in other words, we can find two delays r_i and r_j such that r_i/r_j is irrational, then we say that the delays are *incommensurate*. Obviously, the method of steps can also be used in systems of multiple delays.

Systems with either single delay or multiple delays are known as of *pointwise delays*, *concentrated delays*, or *discrete delay*.

4.2.3 Linear Systems

If the functional f is linear with respect to x_t in (4.3), then we say that the system is *linear*. If it is independent of t, then we say it is *time-invariant*. For a linear time-invariant system, we may define *fundamental solution* $X(t)$ as the solution with initial condition

$$\begin{aligned} &x(0) = I; \\ &x(t) = 0, \; -r \leq t < 0. \end{aligned}$$

where I is the identity matrix of appropriate dimension. If the system is n-dimensional, then $X(t)$ is an $n \times n$-dimensional matrix function of time.

Fundamental solution plays an important role in the study of linear time-delay systems.

Consider, for example, the linear system with single delay

$$\dot{x}(t) = A_0 x(t) + A_1 x(t-r) \tag{4.7}$$

It can be shown, using linearity, that the solution of (4.7) under initial condition $x_0 = \phi$ can be expressed as

$$x(t,\phi) = X(t)\phi(0) + \int_{-r}^{0} X(t-r-\theta)A_1\phi(\theta)d\theta \tag{4.8}$$

4.2.4 Characteristic Quasipolynomials

A linear time-invariant time-delay system is associated with a corresponding *characteristic quasipolynomial* through Laplace Transform. For the system (4.7), the characteristic quasipolynomial is

$$p(s) = \det(sI - A_0 - e^{-rs}A_1).$$

It can be shown that the characteristic quasipolynomial is directly related to the Laplace Transform of the fundamental solution,

$$p(s) = \det(\mathcal{L}[X(t)]).$$

Similar to systems of finite dimension, a time-delay system of retarded type is stable if and only if all the *poles*, or the roots of the characteristic quasipolynomial, are on the left half of the complex plane. However, unlike finite-dimensional systems, a time-delay system has an infinite number of poles, and charaterizing and finding these poles are much more challenging due to the fact that a quasipolynomial involves transcendental functions.

For a linear time-invariant system with multiple delays, the characteristic quasipolynomial can be considered as a polynomial of s, $e^{-r_1 s}$, $e^{-r_2 s}$, ..., $e^{-r_k s}$. For commensurate delays, since $e^{-r_j s} = (e^{-\tau s})^{l_j}$, we can further consider the characteristic quasipolynomial as a polynomial of two variables s and $e^{-\tau s}$. This fact made the stability problem of systems with commensurate delays a much easier problem.

Since the focus in this chapter is on Lyapunov approach, we will not pursue further the stability analysis based on poles.

4.3 Stability

We will start with a formal definition of stability.

Definition 4.1. *For a time-delay system described by (4.3), the trivial solution* $x(t) = 0$ *is said to be* stable *if for any given* $\tau \in \mathbb{R}$ *and* $\varepsilon > 0$*, there exists a* $\delta > 0$ *such that* $||x_\tau||_c < \delta$ *implies* $||x(t)|| < \varepsilon$ *for all* $t \geq \tau$*. It is said to be* asymptotically stable *if it is stable, and for any given* $\tau \in \mathbb{R}$ *and* $\varepsilon > 0$*, there exists, in addition, a* $\delta_a > 0$*, such that* $||x_\tau||_c < \delta_a$ *implies* $\lim_{t\to\infty} x(t) = 0$*. It is said to be* uniformly stable *if it is stable, and* δ *can be made independent of* τ*. It is* uniformly asymptotically stable *if it is uniformly stable and there exists a* $\delta_a > 0$ *such that for any* $\eta > 0$*, there exists a* T *such that* $||x_\tau||_c < \delta_a$ *implies* $||x(t)|| < \eta$ *for* $t > \tau + T$*. It is* globally (uniformly) asymptotically stable *if it is (uniformly) asymptotically stable and* δ_a *can be made arbitrarily large.*

In the above, $|| \cdot ||$ represents the vector 2-norm, and $|| \cdot ||$ is defined as

$$||\phi||_c = \max_{-r\leq\theta\leq 0} ||\phi(\theta)||.$$

The above definition is obviously analogous to finite-dimensional systems. The stability relative to any given solution other than the trivial solution can be transformed to one relative to the trivial solution through a change of variable.

Corresponding to Lyapunov function $V(t, x)$ for finite-dimensional systems, here we need a Lyapunov-Krasovskii functional $V(t, x_t)$ due to the fact that the state is x_t. We have the following Lyapunov-Krasovskii Stability Theorem.

Theorem 4.1. *Suppose* $f : \mathbb{R}\times\mathcal{C} \to\mathbb{R}^n$ *in (4.3) maps* $\mathbb{R}\times$*(bounded sets in* $\mathcal{C}$*) into bounded sets in* $\mathbb{R}^n$*, and that* $u, v, w : \bar{\mathbb{R}}_+ \to \bar{\mathbb{R}}_+$ *are continuous nondecreasing functions. In addition,* $u(s)$ *and* $v(s)$ *are positive for positive* s*, and* $u(0) = v(0) = 0$*. If there exists a continuous differentiable functional* $V : \mathbb{R}\times\mathcal{C} \to\mathbb{R}$ *such that*

$$u(||\phi(0)||) \leq V(t, \phi) \leq v(||\phi||_c), \tag{4.9}$$

and

$$\dot{V}(t, \phi) \leq -w(||\phi(0)||), \tag{4.10}$$

then the trivial solution of (4.3) is uniformly stable. If $w(s) > 0$ *for* $s > 0$*, then it is uniformly asymptotically stable. If, in addition,* $\lim_{s\to\infty} u(s) = \infty$*, then it is globally uniformly asymptotically stable.*

In the above, $\bar{\mathbb{R}}_+$ is the set of nonnegative real scalars. The notation $\dot{V}(t, \phi)$ is defined as

$$\dot{V}(t, \phi) \triangleq \frac{d}{dt} V(t, x_t)|_{x_t=\phi}$$

In other words, we can think of $\dot{V}(\tau, \phi)$ as the derivative of $V(t, x_t)$ with respect to time t, evaluated at the time $t = \tau$, where x_t is the solution of (4.3) with initial condition $x_\tau = \phi$. Indeed, (4.9) and (4.10) are often written as

$$u(||x(t)||) \leq V(t, x_t) \leq v(||x_t||_c),$$
$$\dot{V}(t, x_t) \leq -w(||x(t)||),$$

Notice, although the “state” in this case is x_t, the lower bound of $V(t, x_t)$ and the upper bound of $\dot{V}(t, x_t)$ only need to be functions of $||x(t)||$, and not necessarily be function of $||x_t||_c$. For a proof, the readers are referred to [11], [13] or [18].

4.4 Some Simple Lyapunov-Krasovskii Functionals

This section discusses some simple Lyapunov-Krasovskii functionals for the stability analysis of time-delay systems. The materials of this section may be found from [11] [23] [3].

4.4.1 Delay-independent Stability

Consider the time-delay system (4.7). We may consider the Lyapunov-Krasovskii functional

$$V(x_t) = x^T(t)Px(t) + \int_{-r}^{0} x^T(t+\theta)Sx(t+\theta)d\theta.$$

Where, P and R are symmetric matrices. Obviously,

$$P > 0, \tag{4.11}$$
$$S \geq 0, \tag{4.12}$$

are sufficient to ensure the satisfaction of (4.9). In the above (4.11) means P is positive definite, and (4.12) means S is positive semi-definite. Similarly, we also use “< 0” and “≤ 0” to indicate a matrix is negative definite or semi-definite. Calculating the derivative of V along the system trajectory yields,

$$\dot{V}(x_t) = \begin{pmatrix} x^T(t) \; x^T(t-r) \end{pmatrix} \begin{pmatrix} PA_0 + A_0^T P + S & PA_1 \\ A_1^T P & -S \end{pmatrix} \begin{pmatrix} x(t) \\ x(t-r) \end{pmatrix}.$$

To satisfy (4.10), it is sufficient that

$$\begin{pmatrix} PA_0 + A_0^T P + S & PA_1 \\ A_1^T P & -S \end{pmatrix} < 0. \tag{4.13}$$

Therefore, we can conclude the following.

Proposition 4.1. *The system (4.7) is asymptotically stable if there exists symmetric matrices P and S of appropriate dimension such that (4.11) and (4.13) are satisfied.*

Notice that (4.12) is already implied by (4.13). Inequalities (4.11) and (4.13) are examples of *linear metrix inequalities (LMI)*, where parameters (in this case symmetric matrices P and S). An important development in recent years is that effective numerical methods have been developed to solve LMIs, see [4] for details, and Appendix B of [11] for some facts of LMI most useful for time-delay systems. A number of software packages are available to solve LMIs, see, for example, [8] for LMI Toolbox for MATLAB ®.

It should be observed that the stability conditions (4.11) and (4.13) is independent of the delay r. Such conditions are known as delay-independent stability conditions. Such conditions are obviously have very limited application, because it cannot account for a very common practical situation: a system often tolerate a small delay without losing stability, while a large delay destabilizes the system.

Such simple stability conditions can be extended to more general systems. For example, for systems with multiple delays,

$$\dot{x}(t) = A_0 x(t) + \sum_{j=1}^{k} A_j x(t - r_j), \tag{4.14}$$

we can choose the Lyapunov-Krasovskii functional

$$V(x_t) = x^T(t)Px(t) + \sum_{j=1}^{k} \int_{-r_j}^{0} x^T(t+\theta)S_j x(t+\theta)d\theta. \tag{4.15}$$

Since its derivative along the system trajectory is

$$\dot{V}(x_t) = \psi^T(t)\mathit{\Pi}\psi(t),$$

where

$$\psi^T(t) = \begin{pmatrix} x^T(t) \; x^T(t-r_1) \; \dots \; x^T(t-r_k) \end{pmatrix},$$

$$\mathit{\Pi} = \begin{pmatrix} PA_0 + A_0^T P + \sum\limits_{j=1}^{k} S_j & PA_1 & PA_2 & \dots & PA_k \\ A_1^T P & -S_1 & 0 & \dots & 0 \\ A_2^T P & 0 & -S_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_k^T P & 0 & 0 & \dots & -S_k \end{pmatrix}, \tag{4.16}$$

we arrive at the following stability conditions.

Proposition 4.2. *The system (4.14) is asymptotically stable is there exist symmetric matrices P, S_j, $j = 1, 2, ..., k$, such that*

$$P > 0,$$
$$\mathit{\Pi} < 0$$

are satisfied, where $\mathit{\Pi}$ is defined in (4.16).

A further extension is to systems with distributed delays

$$\dot{x}(t) = A_0 x(t) + \int_{-r}^{0} A(\theta) x(t+\theta) d\theta. \tag{4.17}$$

Analogous to (4.15) for multiple delay case, we can choose

$$V(x_t) = x^T(t) P x(t) + \int_{-r}^{0} [\int_{\theta}^{0} x^T(t+\tau) S(\theta) x(t+\tau) d\tau] d\theta.$$

This gives

$$\begin{aligned}\dot{V}(x_t) &= x^T(t)[PA_0 + A_0^T P + \int_{-r}^{0} S(\theta) d\theta] x(t) \\ &\quad + 2x^T(t) P \int_{-r}^{0} A(\theta) x(t+\theta) d\theta \\ &\quad - \int_{-r}^{0} x^T(t+\theta) S(\theta) x(t+\theta) d\theta.\end{aligned}$$

Add and subtract $x^T(t) \int_{-r}^{0} R(\theta) d\theta x(t)$, where $R(\theta)$ is a symmetric matrix function, we obtain

$$\begin{aligned}\dot{V}(x_t) &= x^T(t)[PA_0 + A_0^T P + \int_{-r}^{0} R(\theta) d\theta] x(t) \\ &\quad + \int_{-r}^{0} \begin{pmatrix} x^T(t) & x^T(t+\theta) \end{pmatrix} \begin{pmatrix} S(\theta) - R(\theta) & PA(\theta) \\ A^T(\theta) P & -S(\theta) \end{pmatrix} \begin{pmatrix} x(t) \\ x(t+\theta) \end{pmatrix} d\theta.\end{aligned}$$

From this, we arrive at the following stability conditions.

Proposition 4.3. *The system (4.17) is asymptotically stable if there exist symmetric matrix P, and symmetric matrix functions S and $R : [-r, 0] \to \mathbb{R}^{n\times n}$, such that*

$$P > 0,$$

$$PA_0 + A_0^T P + \int_{-r}^{0} R(\theta) d\theta < 0,$$

and

$$\begin{pmatrix} S(\theta) - R(\theta) & PA(\theta) \\ A^T(\theta) P & -S(\theta) \end{pmatrix} \leq 0 \text{ for all } \theta \in [-r, 0]$$

are satisfied.

4.4.2 Delay-dependent Stability Using Model Transformation

A simple way of bringing delay r into stability conditions of (4.7) is to transform it to a distributed time-delay system. This is done using the Newton-Raphson formula

$$x(t-r) = x(t) - \int_{-r}^{0} \dot{x}(t+\theta)d\theta$$

for the term $x(t-r)$ in (4.7), and using (4.7) for $\dot{x}(t+\theta)$ in the integral. This result in a new system

$$\dot{x}(t) = (A_0 + A_1)x(t) - A_1 A_0 \int_{-r}^{0} x(t+\theta)d\theta - A_1^2 \int_{-2r}^{-r} x(t+\theta)d\theta. \quad (4.18)$$

The process of obtaining (4.18) from (4.7) is sometimes known as *model transformation*. Before we go on to analyze (4.18), we should point out that the stability of the two systems expressed by (4.7) and (4.18) are not equivalent. Althought the stability of (4.18) implies that of (4.7), the reverse is not necessarily true. It can be seen that the maximum delay of (4.18) is $2r$ rather than r. Indeed, the characteristic equation of (4.7) is

$$\Delta_o(s) = \det(sI - A_0 - e^{-rs}A_1) = 0,$$

and that of (4.18) is

$$\Delta_t(s) = \Delta_a(s)\Delta_o(s) = 0,$$

where

$$\Delta_a(s) = \det\left(I - \frac{1-e^{-rs}}{s}A_1\right).$$

The factor $\Delta_a(s)$ represents *additional dynamics*. It is possible that all the zeros of $\Delta_o(s)$ are on the left half plane while some zeros of $\Delta_a(s)$ are on the right half plane. See [12] for detailed analysis, and [11] and the references therein for additional dynamics in more general setting.

To study the stability of (4.18), we notice that it is in the form of (4.17), and therefore, can use Proposition 4.3, which in this case becomes

$$P > 0,$$

$$P(A_0 + A_1) + (A_0 + A_1)^T P + \int_{-2r}^{0} R(\theta)d\theta < 0,$$

and

$$\begin{pmatrix} S(\theta) - R(\theta) & -PA_1A_0 \\ -(A_1A_0)^T P & -S(\theta) \end{pmatrix} \le 0 \text{ for all } \theta \in [-r, 0],$$

$$\begin{pmatrix} S(\theta) - R(\theta) & -PA_1^2 \\ (A_1^2)^T P & -S(\theta) \end{pmatrix} \le 0 \text{ for all } \theta \in [-2r, -r).$$

We may choose

$$R(\theta) = \begin{cases} R_0, & -r \le \theta \le 0, \\ \\ R_1, & -2r \le \theta < -r, \end{cases}$$

$$S(\theta) = \begin{cases} S_0, & -r \le \theta \le 0, \\ \\ S_1, & -2r \le \theta < -r, \end{cases}$$

to obtain the following stability conditions.

Proposition 4.4. *The system (4.18) is asymptotically stable (which implies that the system (4.17) is asymptotically stable) if there exists symmetric matrices P, S_0, S_1, R_0, and R_1 such that*

$$\begin{aligned} P &> 0, \\ P(A_0 + A_1) + (A_0 + A_1)^T P + r(R_0 + R_1) &< 0, \\ \begin{pmatrix} S_0 - R_0 & -PA_1A_0 \\ -(A_1A_0)^T P & -S_0 \end{pmatrix} &\le 0, \\ \begin{pmatrix} S_1 - R_1 & -PA_1^2 \\ (A_1^2)^T P & -S_1 \end{pmatrix} &\le 0. \end{aligned}$$

We may write the above in a different form by eliminating R_0 and R_1.

Corollary 4.1. *The system (4.18) (and (4.17)) is asymptotically stable if there exist symmetric matrices P, S_0 and S_1 such that*

$$\begin{aligned} P &> 0, \\ \begin{pmatrix} M & -PA_1A_0 & -PA_1^2 \\ & -S_0 & 0 \\ \mathit{Symmetric} & & -S_1 \end{pmatrix} &< 0, \end{aligned}$$

where

$$M = \frac{1}{r}[P(A_0 + A_1) + (A_0 + A_1)^T P] + S_0 + S_1.$$

Proof. We make the conditions in Corollary 4.4 slightly more stringent by replacing "≤" by "<", and eliminating R_0 and R_1 using the technique discussed in [9] or Appendix B of [11] to obtain the resulting LMIs. ■

4.4.3 Implicit Model Transformation

It is also possible to obtain relatively simple delay-dependent stability conditions without explicit model transformation and with less conservatism, although it still uses maximum delay of $2r$. We call such a process as *implicit model transformation*. Here, we will discuss a method very similar to the one proposed by Park [25]. Consider again the system described by (4.7). Choose Lyapunov-Krasovskii functional

$$V(x_t) = x^T(t)Px(t) + \int_{-r}^{0}\int_{\theta}^{0} f^T(x_{t+\xi})Zf(x_{t+\xi})d\xi d\theta + \int_{-r}^{0} x^T(t+\theta)Sx(t+\theta)d\theta$$

where $f(x_t)$ represents the right hand side of (4.7), and by a change of time variable,

$$f(x_{t+\xi}) = A_0x(t+\xi) + A_1x(t+\xi-r).$$

Using the fact that for any defferentiable function ψ and $\theta < 0$,

$$\frac{d}{dt}\int_{\theta}^{0} \psi(f(x_{t+\xi}))d\xi = \psi(f(x_t)) - \psi(f(x_{t+\theta})),$$

we obtain

$$\begin{aligned}\dot{V}(x_t) = {} & \phi_{0r}^T \begin{pmatrix} M & PA_1 + rA_0^TZA_1 \\ [PA_1 + rA_0^TZA_1]^T & rA \end{pmatrix} \phi_{0r} \\ & - \int_{-r}^{0} f^T(x_{t+\theta})Zf(x_{t+\theta})d\theta, \end{aligned} \tag{4.19}$$

where

$$\begin{aligned} M &= PA_0 + A_0^TP + rA_0^TZA_0 + S, \\ \phi_{0r}^T &= \begin{pmatrix} x^T(t) & x^T(t-r) \end{pmatrix}. \end{aligned}$$

For

$$\begin{pmatrix} X & Y \\ Y^T & Z \end{pmatrix} > 0, \tag{4.20}$$

we have

$$\begin{aligned} 0 < {} & \int_{-r}^{0} \begin{pmatrix} x^T(t) & \dot{x}^T(t+\theta) \end{pmatrix} \begin{pmatrix} X & Y \\ Y^T & Z \end{pmatrix} \begin{pmatrix} x(t) \\ \dot{x}(t+\theta) \end{pmatrix} d\theta \\ = {} & rx^T(t)Xx(t) + 2x^T(t)Y(x(t) - x(t-r)) + \int_{-r}^{0} \dot{x}^T(t+\theta)Z\dot{x}(t+\theta)d\theta. \end{aligned} \tag{4.21}$$

Adding (4.21) to (4.19) and using

$$\dot{x}(t+\theta) = f(x_{t+\theta}), \tag{4.22}$$

we obtain

$$\dot{V}(x_t) \leq \phi_{0r}^T \begin{pmatrix} N & PA_1 + rA_0^T ZA_1 - Y \\ \text{Symmetric} & -S + rA_1^T ZA_1 \end{pmatrix} \phi_{0r},$$

where

$$N = PA_0 + A_0^T P + rA_0^T ZA_0 + S + rX + Y + Y^T. \tag{4.23}$$

Therefore, we conclude the following.

Proposition 4.5. *The system (4.7) is asymptotically stable if there exist matrix Y and symmetric matrices P, X and Z such that*

$$P > 0,$$

$$\begin{pmatrix} N & PA_1 + rA_0^T ZA_1 - Y \\ \mathit{Symmetric} & -S + rA_1^T ZA_1 \end{pmatrix} < 0,$$

and (4.20) are satisfied. In the above, N is expressed in (4.23).

Notice, due to the usage of (4.22) for $\theta \in [-r, 0)$, this process involves $x(t+\xi)$ for $-2r \leq \xi \leq 0$, and implicitly involves model transformation in some sense. It can shown that this stability condition is indeed less conservative than both Propositions 4.1 and 4.4. It can also be written in a number of different forms, see [11] for details.

4.5 Complete Quadratic Lyapunov-Krasovskii Functional

It will be shown by numerical examples later on in this section that all the methods discussed in the previous section involves substantial conservatism. Further more, all of them requires the system to be stable if the delay is set to zero. However, there are many practical cases where delay may be used to stabilize the system. See [1] for a simple example. Indeed, a finite difference approximation of derative in control implementation will introduce time delays, which are often used to stabilize the system.

To obtain necessary and sufficient condiiton for stability, it is necessary to use *complete quadratic Lyapunov-Krasovskii functional* as pointed out by Repin [27], Infante and Castelan [15].

4.5.1 Analytical Expression

Recall that the finite dimensional system

$$\dot{x}(t) = Ax(t) \tag{4.24}$$

is asymptotically stable if and only if for any given positive definite W, the Lyapunov equation

$$PA + A^T P = -W$$

has a positive definite solution. Indeed, a quadratic Lyapunov function can be constructed from the solution P,

$$V(x) = x^T P x,$$

which achieves

$$\dot{V}(x) = -x^T W x.$$

Furthermore, the solution P can be explicitly expressed as

$$P = \int_0^\infty X^T(t) W X(t) dt,$$

where $X(t)$ is the fundamental solution of (4.24), which satisfy

$$\begin{aligned} \dot{X}(t) &= AX(t), \\ X(0) &= I. \end{aligned}$$

Let $x(t, \phi)$ be the solution of (4.24) with initial condition $x(0) = \phi$, then $x(t, \phi) = X(t)\phi$, and therefore, we may further write

$$V(\phi) = \int_0^\infty x^T(\tau, \phi) W x(\tau, \phi) d\tau.$$

For a stable time-delay system (4.7), it is also possible to construct a Lyapunov-Krasovskii functional $V(x_t)$ such that

$$\dot{V}(x_t) = -x^T(t) W x(t).$$

Indeed, let $x(t, \phi)$ be the solution of (4.7) with initial condition $x_0(\theta) = \phi(\theta)$, $\theta \in [-r, 0]$, then we can still write

$$V(\phi) = \int_0^\infty x^T(\tau, \phi) W x(\tau, \phi) d\tau.$$

Through some algebra, we can expression $V(\phi)$ explicitly as a quadratic functional of ϕ,

$$\begin{aligned} V(\phi) = {} & \phi^T(0) U(0) \phi(0) \\ & + 2\phi^T(0) \int_{-r}^0 U(-r - \theta) A_1 \phi(\theta) d\theta \\ & + \int_{-r}^0 \int_{-r}^0 \phi^T(\theta_1) A_1^T U(\theta_1 - \theta_2) A_1 \phi(\theta_2) d\theta_1 d\theta_2 \end{aligned} \tag{4.25}$$

where $U : \mathbb{R} \to \mathbb{R}^{n \times n}$ is defined as

$$U(\tau) = \int_0^\infty X^T(t)WX(t+\tau)dt.$$

In order to have $U(\tau)$ well defined, we agree that $X(t) = 0$ for $t < 0$. It can be shown that

$$U^T(\tau) = U(-\tau).$$

The readers are referred to [11] for more details.

4.5.2 Discretization

The analytical expression (4.25) indicates that for any asymptotically stable system, we can always find a complete quadratic Lyapunov-Krasovskii functional. In other words, the existence of such a functional is necessary and sufficient for stability. In order for numerical calculation, we enlarge the class of quadratic Lyapunov-Krasovskii functionals to the form

$$\begin{aligned} V(x_t) &= x^T(t)Px(t) \\ &+ 2x^T(t)\int_{-r}^0 Q(\theta)x(t+\theta)d\theta \\ &+ \int_{-r}^0\int_{-r}^0 x^T(t+\xi)R(\xi,\eta)x(t+\eta)d\xi d\eta \\ &+ \int_{-r}^0 x^T(t+\xi)S(\xi)x(t+\xi)d\xi, \end{aligned} \tag{4.26}$$

where

$$P = P^T,$$

and for all $\xi \in [-r, 0]$, $\eta \in [-r, 0]$,

$$\begin{aligned} Q(\xi) &\in \mathbb{R}^{n\times n}, \\ R(\xi,\eta) &= R^T(\eta,\xi) \in \mathbb{R}^{n\times n}, \\ S(\xi) &= S^T(\xi) \in \mathbb{R}^{n\times n}. \end{aligned}$$

Since $V(x_t)$ is clearly upper-bounded, sufficient conditions for asymptotic stability (we can show they are also necessary) are

$$V(x_t) \geq \varepsilon||x(t)||^2, \tag{4.27}$$

$$\dot{V}(x_t) \leq -\varepsilon||x(t)||^2, \tag{4.28}$$

for some $\varepsilon > 0$.

The search for the existence of functions Q, R and S (in addition to matrix P) is clearly an infinite-dimensional problem. It can be viewed as an infinite-dimensional LMI. To make numerical computation feasible, we will constrain these matrix functions to be *piecewise linear*. Specifically, divide the interval $[-r, 0]$ into N intervals of equal length (nonuniform mesh is also possible, but we will not discuss it here)

$$h = \frac{r}{N},$$

and let the dividing points be denoted as

$$\theta_p = -ph = -\frac{pr}{N},\ p = 0, 1, 2, ..., N.$$

Let

$$\begin{aligned} Q_p &= Q(\theta_p), \\ S_p &= S(\theta_p), \\ R_{pq} &= R(\theta_p, \theta_q). \end{aligned}$$

Then, for $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$, we restrict, for $p = 1, 2, ..., N$,

$$\begin{aligned} Q(\theta_p + \alpha h) &= (1-\alpha)Q_p + \alpha Q_{p-1}, \\ S(\theta_p + \alpha h) &= (1-\alpha)S_p + \alpha S_{p-1}, \end{aligned}$$

and

$$\begin{aligned} &R(\theta_p + \alpha h, \theta_q + \beta h) \\ &= \begin{cases} (1-\alpha)R_{pq} + \beta R_{p-1,q-1} + (\alpha-\beta)R_{p-1,q},\ \alpha \geq \beta, \\ \\ (1-\beta)R_{pq} + \alpha R_{p-1,q-1} + (\beta-\alpha)R_{p,q-1},\ \alpha < \beta. \end{cases} \end{aligned}$$

Through a rather tedious process, we can reduce (4.27) and (4.28) to LMIs. This approach is known as the *discretized Lyapunov functional method*. Here, we will only give the resulting LMI for the case of $N = 1$ in the following. The readers are referred to [11] for the general case.

Proposition 4.6. *The system is asymptotically stable if there exist $n \times n$ real matrices $P = P^T$, Q_p, $S_p = S_p^T$, $R_{pq} = R_{qp}^T$, $p = 0, 1$; $q = 0, 1$, such that*

$$\begin{pmatrix} P & Q_0 & Q_1 \\ & R_{00} + S_0 & R_{01} \\ \mathit{Symmetric} & & R_{11} + S_1 \end{pmatrix} > 0,$$

and

$$\begin{pmatrix} \varDelta_{00} & Q_1 - PA_1 & D_0^s & D_0^a \\ & S_1 & D_1^s & D_1^a \\ & & h(R_{00} - R_{11}) + S_0 - S_1 & 0 \\ \mathit{Symmetric} & & & 3(S_0 - S_1) \end{pmatrix} > 0,$$

where

$$\begin{aligned}
\Delta_{00} &= -PA_0 - A_0^T P - Q_0 - Q_0^T - S_0,\\
D_0^s &= \frac{r}{2}A_0^T(Q_0+Q_1) + \frac{r}{2}(R_{00}+R_{01}) - (Q_0 - Q_1),\\
D_1^s &= \frac{r}{2}A_1^T(Q_0+Q_1) - \frac{r}{2}(R_{10}+R_{11}),\\
D_0^a &= -\frac{r}{2}A_0^T(Q_0-Q_1) - \frac{r}{2}(R_{00}-R_{01}),\\
D_1^a &= -\frac{r}{2}A_1^T(Q_0-Q_1) + \frac{r}{2}(R_{10}-R_{11}).
\end{aligned}$$

4.6 A Comparison of Lyapunov-Krasovskii Functionals

Obviously, the delay-independent stability condition in Proposition 4.1 is very conservative if the delay is known. Although the simple delay-dependent condition in Proposition 4.4 is intended to improve the situation, it is not necessarily less conservative in all the situations. There are indeed systems which satisfy the conditions in Proposition 4.1 but do not satisfy those in Proposition 4.4. See [12] for an example.

As mentioned earlier, it can be shown that the method with implicit model transformation discussed in Proposition 4.5 is indeed less conservative than both Proposition 4.1 and Proposition 4.4.

The discretized Lyapunov functional method can approach analytical results very quickly, and is the least conservative among these methods. The following example is often used in the literature.

Example 4.3. Consider the system

$$\dot{x}(t) = \begin{pmatrix} -2 & 0 \\ 0 & -0.9 \end{pmatrix} x(t) + \begin{pmatrix} -1 & 0 \\ -1 & -1 \end{pmatrix} x(t-r)$$

Various methods are used to estimate the maximum delay $r_{\max}$ without losing stability, and the results are listed in the following table. In the first line, "Analytical" indicates the true maximum delay obtained by the first time a pair of roots of the characteristic quasipolynomial crosses the imaginary axis as the delay increases; "Explicit" means the delay-dependent stability conditions in Propostion 4.4 which uses explicit model transformation; "Implicit" denotes the delay-dependent stability conditions in Proposition 4.5 which uses implicit model transformation, the remaining three columns are the results using discretized Lyapunov functional method with different N, with $N = 1$ covered in Proposition 4.6.

Methods	Analytical	Explicit	Implicit	$N = 1$	$N = 2$	$N = 3$
$r_{\max}$	6.17258	1.00	4.359	6.059	6.165	6.171

The next example shows that there are indeed systems that are unstable without delay, but may becomes stable for some nonzero delays.

Example 4.4. Consider the system

$$\ddot{x}(t) - 0.1\dot{x}(t) + x(t) = -r\frac{x(t) - x(t-r)}{r}$$

The left hand side may be considered as a second order system with negative damping, and the right hand side can be considered as a control to stabilize the system by providing sufficient positive damping and using finite difference to approximate the derivative. If the derivative is used instead of finite difference, then obviously the system would be stable for $r > 0.1$. For such systems, the stability conditions covered in Propositions 4.4 and 4.5 are not applicable since they requires the system to be stable for zero delay. We now write the system in a state space form

$$\frac{d}{dt}\begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -2 & 0.1 \end{pmatrix}\begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x(t-r) \\ \dot{x}(t-r) \end{pmatrix}.$$

The system is stable for $r \in (r_{\min}, r_{\max})$. The following table lists the estimated values using discretized Lyapunov functional method with different N, as well as the analytical values. It can be seen, again, that discretized Lyapunov functional method can approach the analytical results with a rather modest N.

N	1	2	3	Analytical
$r_{\min}$	0.1006	0.1003	0.1003	0.1002
$r_{\max}$	1.4272	1.6921	1.7161	1.7178

4.7 Dealing with Time-varying Delays

Consider a system

$$\dot{x}(t) = A_0 x(t) + A_1 x(t - r(t)), \tag{4.29}$$

where the time-varying delay $r(t)$ satisfies

$$r_m \leq r(t) \leq r_M, \tag{4.30}$$

$$\dot{r}(t) \leq \rho, \tag{4.31}$$

where ρ is a known constant, $0 \leq \rho < 1$.

We choose a complete quadratic Lyapunov-Krasovskii functional

$$V(t, x_t) = V_1(x_t) + V_2(t, x_t),$$

where,

$$\begin{aligned} V_1(x_t) &= x^T(t)Px(t) \\ &+ 2x^T(t)\int_{-r_m}^{0} Q(\theta)x(t+\theta)d\theta \\ &+ \int_{-r_m}^{0}\int_{-r_m}^{0} x^T(t+\xi)R(\xi,\eta)x(t+\eta)d\xi d\eta \\ &+ \int_{-r_m}^{0} x^T(t+\xi)S(\xi)x(t+\xi)d\xi. \end{aligned} \tag{4.32}$$

Let $V_1^*(x_t)$ indicates the derivative of (4.32) along the comparison system

$$\dot{x}(t) = A_0x(t) + A_1x(t-r_m), \tag{4.33}$$

which is in an identical form discussed in the last section. Then,

$$\begin{aligned} \dot{V}_1(x_t) = V_1^*(x_t) &+ 2x^T(t)PA_1[x(t-r(t)) - x(t-r_m)] \\ &+ 2[x(t-r(t)) - x(t-r_m)]^T A_1^T \int_{-r^m}^{0} Q(\theta)x(t+\theta)d\theta. \end{aligned}$$

Let

$$\begin{aligned} V_2(t, x_t) = &\int_{-r_M}^{-r_m} [\int_{\theta}^{0} x^T(t+\zeta)K_1x(t+\zeta)d\zeta]d\theta \\ &+ \int_{-r_M}^{-r_m} [\int_{\theta-r(t+\theta)}^{0} x^T(t+\zeta)K_2x(t+\zeta)d\zeta]d\theta. \end{aligned}$$

Then

$$\begin{aligned} \dot{V}_2(t, x_t) &= (r_M - r_m)x^T(t)(K_1 + K_2)x(t) \\ &- \int_{-r_M}^{-r_m} x^T(t+\theta)K_1x(t+\theta)d\theta \\ &- \int_{-r_M}^{-r_m} (1-\dot{r}(t+\theta))x^T(t+\theta-r(t+\theta))K_2x(t+\theta-r(t+\theta))d\theta. \end{aligned}$$

In view of the fact that

$$\begin{aligned} &x(t-r(t)) - x(t-r_m) \\ &= \int_{-r(t)}^{-r_m} \dot{x}(t+\theta)d\theta \\ &= \int_{-r(t)}^{-r_m} [A_0x(t+\theta) + A_1x(t+\theta+r(t+\theta))]d\theta, \end{aligned}$$

we have

$$\begin{aligned}\dot{V}(t,x_t) = {} & V_1^*(x_t) - 2x^T(t)PA_1 \int_{-r(t)}^{-r_m} [A_0x(t+\theta) + A_1x(t+\theta+r(t+\theta))]d\theta \\ & - 2\int_{-r(t)}^{-r_m} [A_0x(t+\theta) + A_1x(t+\theta+r(t+\theta))]^T d\theta A_1^T \int_{-r_m}^{0} Q(\theta)x(t+\theta)d\theta \\ & + (r_M - r_m)x^T(t)(K_1+K_2)x(t) - \int_{-r_M}^{-r_m} x^T(t+\theta)K_1x(t+\theta)d\theta \\ & - \int_{-r_M}^{-r_m} (1-\dot{r}(t+\theta))x^T(t+\theta-r(t+\theta))K_2x(t+\theta-r(t+\theta))d\theta.\end{aligned}$$

Using (4.30) and (4.31), we can arrive at

$$\begin{aligned}\dot{V}(t,x_t) \leq {} & -\int_{-r(t)}^{-r_m} \begin{pmatrix} x^T(t) & x^T(t+\theta) \end{pmatrix} \begin{pmatrix} \hat{K}_{1a} & PA_1A_0 \\ A_0^TA_1^TP & K_{1a} \end{pmatrix} \begin{pmatrix} x(t) \\ x(t+\theta) \end{pmatrix} d\theta \\ & - \int_{-r(t)}^{-r_m} \begin{pmatrix} \mu^T(t) & x^T(t+\theta) \end{pmatrix} \begin{pmatrix} \hat{K}_{1b} & A_1A_0 \\ A_0^TA_1^T & K_{1b} \end{pmatrix} \begin{pmatrix} \mu(t) \\ x(t+\theta) \end{pmatrix} d\theta \\ & - \int_{-r(t)}^{-r_m} \begin{pmatrix} x^T(t) & \nu^T(t,\theta) \end{pmatrix} \begin{pmatrix} \hat{K}_{2a} & PA_1A_1 \\ A_1^TA_1^TP & (1-\rho)K_{2a} \end{pmatrix} \begin{pmatrix} x(t) \\ \nu(t,\theta) \end{pmatrix} d\theta \\ & - \int_{-r(t)}^{-r_m} \begin{pmatrix} \mu^T(t) & \nu^T(t,\theta) \end{pmatrix} \begin{pmatrix} \hat{K}_{2b} & A_1A_1 \\ A_1^TA_1^T & (1-\rho)K_{2b} \end{pmatrix} \begin{pmatrix} \mu(t) \\ \nu(t,\theta) \end{pmatrix} d\theta \\ & + V_1^*(x_t) + (r_M - r_m)x^T(t)(K_1+K_2)x(t) \\ & + (r(t) - r_m)x^T(t)(\hat{K}_{1a} + \hat{K}_{2a})x(t) \\ & + (r(t) - r_m)\mu^T(t)(\hat{K}_{1b} + \hat{K}_{2b})\mu(t),\end{aligned}$$

where

$$\begin{aligned}\mu(t) &= \int_{-r_m}^{0} Q(\theta)x(t+\theta)d\theta, \\ \nu(t,\theta) &= x(t+\theta-r(t+\theta)),\end{aligned}$$

and

$$\begin{aligned}K_{1a} + K_{1b} &= K_1, \\ K_{2a} + K_{2b} &= K_2.\end{aligned}$$

If we choose K_{1a}, K_{2a} (so that K_{1b}, K_{2b} are also determined), and $\hat{K}_{1a}$, $\hat{K}_{1b}$, $\hat{K}_{2a}$, $\hat{K}_{2b}$ such that

$$\begin{pmatrix} \hat{K}_{1a} & PA_1A_0 \\ A_0^TA_1^TP & K_{1a} \end{pmatrix} \geq 0,$$

$$\begin{pmatrix} \hat{K}_{1b} & A_1A_0 \\ A_0^TA_1^T & K_{1b} \end{pmatrix} \geq 0,$$

$$\begin{pmatrix} \hat{K}_{2a} & PA_1A_1 \\ A_1^TA_1^TP & (1-\rho)K_{2a} \end{pmatrix} \geq 0,$$

$$\begin{pmatrix} \hat{K}_{2b} & A_1A_1 \\ A_1^TA_1^T & (1-\rho)K_{2b} \end{pmatrix} \geq 0,$$

then the four integrals are all less or equal to zero. Therefore, we conclude that the system is asymptotically stable if we can make

$$\begin{aligned} & V_1^*(x_t) + (r_M - r_m)x^T(t)(K_1 + K_2)x(t) \\ & + (r(t) - r_m)x^T(t)(\hat{K}_{1a} + \hat{K}_{2a})x(t) \\ & + (r(t) - r_m)\mu^T(t)(\hat{K}_{1b} + \hat{K}_{2b})\mu(t) \\ & \leq -\varepsilon||x(t)||^2. \end{aligned}$$

A discretized Lyapunov functional approach can be used to achieve this. The above development is similar to [14].

An alternative is to formulate the time-varying delay as a perturbation to a time-invariant delay, and formulate it as an uncertain feedback problem. See, for example, [11] and [23].

It is also possible to lift the restriction of derivative bound (4.31). One simple approach is to use the Razumikhin Theorem based methods. Other approaches include an alternative formulation of Lyapunov-Krasovskii functional method proposed in [7], and the input-output approach along the similar idea as [16].

4.8 Razumikhin Theorem

Razumikhin showed that it is still possible to use function rather than functionals in stability analysis of time-delay system. This is based on the following Razumikhin Theorem.

Theorem 4.2. *Suppose $f : \mathbb{R}\times\mathcal{C} \to \mathbb{R}^n$ in (4.3) takes $\mathbb{R} \times$ (bounded sets of $\mathcal{C}$) into bounded sets of $\mathbb{R}^n$, and u, v, $w : \bar{\mathbb{R}}_+ \to \bar{\mathbb{R}}_+$ are continuous nondecreasing functions, $u(s)$ and $v(s)$ are positive for $s > 0$, and $u(0) = v(0) = 0$, v strictly increasing. If there exists a continuously differentiable function $V : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ such that*

$$u(||x||) \leq V(t, x) \leq v(||x||), \text{ for } t \in \mathbb{R} \text{ and } x \in \mathbb{R}^n, \tag{4.34}$$

and the derivative of V along the solution $x(t)$ of (4.3) satisfies

$$\dot{V}(t,x(t)) \leq -w(||x(t)||) \text{ whenever } V(t+\theta, x(t+\theta)) \leq V(t,x(t)), \quad (4.35)$$

for $\theta \in [-r,0]$, then the system (4.3) is uniformly stable. If, in addition, $w(s) > 0$ for $s > 0$, and there exists a continuous nondecreasing function $p(s) > s$ for $s > 0$ such that condition (4.35) is strengthened to

$$\dot{V}(t,x(t)) \leq -w(||x(t)||) \text{ whenever } V(t+\theta, x(t+\theta)) \leq p(V(t,x(t)), \quad (4.36)$$

for $\theta \in [-r,0]$, then the system (4.3) is uniformly asymptotically stable. If, in addition, $\lim\limits_{s\to\infty} u(s) = \infty$, then the system (4.3) is globally uniformly asymptotically stable.

The basic idea of the above Theorem is to consider the Lyapunov-Krasovskii functional

$$\bar{V}(x_t) = \max_{\theta\in[-r,0]} V(x+\theta),$$

and realize that if $V(x(t)) < \bar{V}(x_t)$, then $\bar{V}(x_t)$ does not grow at the instant t even if $\dot{V}(x(t)) > 0$. Therefore, in order for $\bar{V}(x_t)$ to grow, one only needs to make sure that $\dot{V}(x(t))$ is not positive whenever $V(x(t)) = \bar{V}(x_t)$. For a proof, the readers are referred to [11], [13] or [18].

A direct application of Razumikhin Theorem to time-invariant time-delay systems typically results in a more conservative stability conditions than the counterpart obtained by using the Lyapunov-Krasovskii functional method. However, there are a number of situations where Razumikhin Theorem has advantage. For example, time-varying delay can be easily handled. Consider the following system

$$\dot{x}(t) = A_0x(t) + A_1x(t-r(t)). \quad (4.37)$$

This is the same as (4.7) except the delay is time-varying. Typically, the delay is known within certain range,

$$r_m \leq r(t) \leq r_M.$$

However, there is no restriction on the rate of change of $r(t)$. Let

$$V(x) = x^TPx, \; P > 0.$$

Then, we can calculate

$$\dot{V}(x(t)) = x^T(t)(PA_0 + A_0^TP)x(t) + 2x^TPA_1x(t-r(t)).$$

The system is asymptotically stable if

$$\dot{V}(x(t)) \leq -\varepsilon||x(t)||^2 \text{ for some small } \varepsilon > 0,$$

whenever

$$V(x(t-r(t)) \leq \beta V(x(t)) \text{ for some } \beta > 1.$$

In other words, it is sufficient that

$$\dot{V}(x(t)) - \alpha[\beta V(x(t-r(t)) - V(x(t))] \leq -\varepsilon||x(t)||^2,$$

for some $\varepsilon > 0$, $\alpha \geq 0$ and $\beta > 1$. Using the expression for V and $\dot{V}$, the above becomes

$$\begin{aligned} &\begin{pmatrix} x^T(t) \; x^T(t-r(t)) \end{pmatrix} \begin{pmatrix} PA_0 + A_0^T P + \alpha P & PA_1 \\ A_1^T P & -\alpha\beta P \end{pmatrix} \begin{pmatrix} x(t) \\ x(t-r(t)) \end{pmatrix} \\ &\leq -\varepsilon x^T(t)x(t) \end{aligned}$$

from which we can conclude the following.

Proposition 4.7. *The system (4.37) is asymptotically stable if there exist a real scalar $\alpha > 0$ and symmetric matrix $P > 0$ such that*

$$\begin{pmatrix} PA_0 + A_0^T P + \alpha P & PA_1 \\ A_1^T P & -\alpha P \end{pmatrix} < 0. \tag{4.38}$$

Compared to Proposition 4.1, the above can be obtained from (4.13) by constraining $S = \alpha P$. Therefore, this is obviously more conservative if used for systems with a time-invariant delay. Computationally although (4.38) involves fewer parameters than (4.13), it is actually computationally more difficult because it is no longer an LMI due to the multiplicative term αP. See [11] for handling such computational issue.

Parallel to the Lypunov-Krasovskii functional methods, we can also derive delay-dependent results using explicit and implicit model transformation. See [11] for details.

4.9 Coupled Difference-Differential Equations

4.9.1 Introducation

In this section, we will discuss the system described by coupled difference-differential equations,

$$\dot{x}(t) = Ax(t) + By(t-r), \tag{4.39}$$

$$y(t) = Cx(t) + Dy(t-r), \tag{4.40}$$

where $x(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^n$. This model is also known as the *lossless propagation model* due to the fact that it comes out naturally from simplifying some lossless propagation systems [24]. Most of the materials in this section are based on [10].

The equations (4.39) and (4.40) represent both neutral and retarded time-delay systems with commensurate multiple delays as special cases. For example, for the system described by

$$\sum_{k=0}^{p} F_k \dot{x}(t-kr) = \sum_{k=0}^{p} A_k x(t-kr), \qquad F_0 = I,$$

we may define

$$\begin{aligned} y_k(t) &= x(t-kr+r), \\ z(t) &= \sum_{k=0}^{p} F_k x(t-kr). \end{aligned}$$

This allows us to write the system as

$$\begin{aligned} \dot{z}(t) &= A_0 z(t) + \sum_{k=1}^{p} (A_k - A_0 F_k) y_k(t-r), \\ y_1(t) &= z(t) - \sum_{k=1}^{p} F_k y_k(t-r), \\ y_k(t) &= y_{k-1}(t-r), \qquad k = 2, 3, ..., p, \end{aligned}$$

which is in the standard form of (4.39) and (4.40).

Obviously, the future evolution of the system described by (4.39) and (4.40) is completely decided by $x(t)$ and $y(t+\theta)$, $-r \le \theta < 0$. Naturally, the initial condition to be specified should be described by

$$x(0) = \psi, \tag{4.41}$$

$$y_0 = \phi. \tag{4.42}$$

In (4.42), we have used the notation that y_t represents a time-shift and restriction of y in the interval $[t-r, t)$ defined as

$$y_t(\theta) = y(t+\theta), \ -r \le \theta < 0,$$

and $\phi : [-r, 0) \to \mathbb{R}^n$.

For the pair (ψ, ϕ), we also define the norm as

$$||(\psi, \phi)|| = \max\{||\psi||, \sup_{-r \le \theta < 0} ||\phi(\theta)||\}.$$

We can describe the general Lyapunov-Krasovskii stability condition for the system described by (4.39) and (4.40) as follows, which is also similar to a neutral time-delay system.

Theorem 4.3. *Consider the system described by (4.39) and (4.40) with* $\rho(D) < 1$. *Let* $u, v, w : \bar{\mathbb{R}} \rightarrow \bar{\mathbb{R}}$ *be continuous and nondecreasing functions. In addition,* $u(s)$ *and* $v(s)$ *are positive for positive* s, *and* $u(0) = v(0) = 0$. *If there exists a continuous differentiable functional* $V : (\psi, \phi)$ *such that*

$$\begin{aligned} &u(||\psi||) \leq V(\psi, \phi) \leq v(||(\psi, \phi)||), \\ &\dot{V}(\psi, \phi) \leq -w(\psi), \end{aligned}$$

then the trivial solution of the system is stable. If, in addition, $w(s) > 0$ *for* $s > 0$, *then it is asymptotically stable.*

We can prove the above in a very similar way to the standard neutral time-delay system (for example, Theorem 1.1 in Chapter 8 of [18]) using the fact that $\rho(D) < 1$.

4.9.2 Fundamental Solutions

As in the case of the sytem (4.7), a complete quadratic Lyapunov-Krasovakii functional is essential to give nonconservative stability conditions, and the analytical construction of such a Lyapunov-Krasovskii functional is based on the fundamental solutions.

We will write the solution of the equation

$$\dot{x}(t) = Ax(t) + By(t-r) + \delta(t)I, \tag{4.43}$$

$$y(t) = Cx(t) + Dy(t-r), \tag{4.44}$$

with zero initial conditions

$$x(0) = 0,\ y_0 = 0, \tag{4.45}$$

as

$$\begin{aligned} x(t) &= X_x(t), \\ y(t) &= Y_x(t). \end{aligned}$$

Similarly, the solution of

$$\dot{x}(t) = Ax(t) + By(t-r), \tag{4.46}$$

$$y(t) = Cx(t) + Dy(t-r) + \delta(t)I, \tag{4.47}$$

with zero initial conditions (4.45) are denoted as

$$x(t) = X_y(t),$$
$$y(t) = Y_y(t).$$

We also agree that $X_x(t) = 0, Y_x(t) = 0, X_y(t) = 0, Y_y(t) = 0$ for $t < 0$. The solutions ($X_x(t)$, $Y_x(t)$, $X_y(t)$, $Y_y(t)$) are known as the fundamental solutions of the system described by (4.39) and (4.40). $(X_x(t), Y_x(t))$ can also be regarded as the solution of (4.39) and (4.40) with initial condition

$$x(0) = I,\ y_0 = 0.$$

Similary, $(X_y(t), Y_y(t))$ may be regarded as the solution of (4.39) and (4.40) with initial condition

$$x(0) = 0,$$
$$y(\theta) = \delta(\theta)I,\ \ -r < \theta \leq 0.$$

With this interpretation in mind, we may write $X_y(t)$ and $Y_y(t)$ in terms of $X_x(t)$ and $Y_x(t)$. Indeed, it is easy to see that the solution of (4.46) and (4.47) in the interval $[0, r)$ is $x(t) = 0$, $y(t) = \delta(t)$. Now consider the interval $[r, 2r)$, $y(t - r)$ is zero except the impulse at $t = r$, producing a step of B at time $t = r$. Therefore, solution is $x(t) = X_x(t - r)B$ and $y(t) = Y_x(t - r)B$. Continuing this process yields

$$X_y(t) = \sum_{k=0}^{\infty} D^k X_x(t - kr - r)B \tag{4.48}$$

$$= \sum_{k=0}^{[t/r]-1} D^k X_x(t - kr - r)B, \tag{4.49}$$

$$Y_y(t) = \sum_{k=0}^{\infty} \delta(t - kr)D^k + \sum_{k=0}^{\infty} D^k Y_x(t - kr - r)B \tag{4.50}$$

$$= \sum_{k=0}^{[t/r]} \delta(t - kr)D^k + \sum_{k=0}^{[t/r]-1} D^k Y_x(t - kr - r)B, \tag{4.51}$$

where $[t/r]$ represents the largest integer not to exceed t/r.

With the fundamental solutions, it is easy to write the general solutions of (4.39) and (4.40). Let the solution of (4.39) and (4.40) with initial conditions (4.41) and (4.42) be denoted as

$$x(t) = x(t, \psi, \phi),$$
$$y(t) = y(t, \psi, \phi).$$

Then, using linearity, it is not difficult to see that

$$x(t,\psi,\phi) = X_x(t)\psi + \int_{-r}^{0} X_y(t+\theta)\phi(\theta)d\theta, \tag{4.52}$$

$$y(t,\psi,\phi) = Y_x(t)\psi + \int_{-r}^{0} Y_y(t+\theta)\phi(\theta)d\theta. \tag{4.53}$$

Using Expressions (4.48) and (4.50), they can also be expressed as

$$x(t,\psi,\phi) = X_x(t)\psi + \int_{-r}^{0} \sum_{k=0}^{[(t+\theta)/r]-1} D^k X_x(t+\theta-kr-r)B\phi(\theta)d\theta, \tag{4.54}$$

$$\begin{aligned} y(t,\psi,\phi) &= Y_x(t)\psi + D^{[t/r]+1}\phi(t-[t/r]r-r) \\ &\quad + \int_{-r}^{0} \sum_{k=0}^{[(t+\theta)/r]-1} D^k Y_x(t+\theta-kr-r)B\phi(\theta)d\theta. \end{aligned} \tag{4.55}$$

It can be observed from the above discussions that, for continuous $\phi(\theta)$, $x(t)$ is continuous. However $y(t)$ is in general discontinuous. This is typical of neutral time-delay systems. Also, for the system to be stable, a necessary condition is that the spectrum radius of matrix D is less than 1, $\rho(D) < 1$, another well known fact for neutral time-delay systems.

On the other hand, if $\rho(D) < 1$, then the system would be exponentially stable if and only if $X_x(t)$ and $Y_x(t)$ are exponentially bounded. Indeed, in this case, for any given $\rho(D) < \gamma < 1$, there exists a $K > 0$ such that

$$||D^k|| \le K\gamma^k.$$

Also,

$$X_x(t) \le Me^{-\alpha t},\ M > 0,\ \alpha > 0,$$
$$Y_x(t) \le Ne^{-\beta t},\ N > 0,\ \beta > 0.$$

Then for any bounded initial condition

$$||\psi|| \le L,$$
$$||\phi(\theta)|| \le L,\ -r \le \theta < 0,$$

we have

$$\begin{aligned}
||x(t,\psi,\phi)|| &\leq MLe^{-\alpha t} + \sum_{k=0}^{[t/r]-1} K\gamma^k \int_{-r}^{0} Me^{-\alpha(t-(k-2)r)}||B||Ld\theta \\
&= MLe^{-\alpha t} + \sum_{k=0}^{[t/r]-1} KMLr||B||\gamma^k e^{-\alpha(t-(k-2)r)} \\
&= MLe^{-\alpha t} + KMLr||B||e^{-\alpha(t+2r)} \sum_{k=0}^{[t/r]-1} (\gamma e^{\alpha r})^k \\
&= MLe^{-\alpha t} + KMLr||B||e^{-\alpha(t+2r)} \frac{(\gamma e^{\alpha r})^{[t/r]} - 1}{\gamma e^{\alpha r} - 1} \\
&= MLe^{-\alpha t} + KMLr||B|| \frac{\gamma^{[t/r]} e^{\alpha(r[t/r]-t-2r)} - e^{-\alpha(t+2r)}}{\gamma e^{\alpha r} - 1} \\
&\leq MLe^{-\alpha t} + KMLr||B|| \frac{\gamma^{[t/r]} e^{\alpha(r[t/r]-t-2r)} + e^{-\alpha(t+2r)}}{|\gamma e^{\alpha r} - 1|}.
\end{aligned}$$

In the above, we have assumed $\gamma e^{\alpha r} \neq 1$, which can be satisfied by properly choosing γ. Since $e^{\alpha(r[t/r]-t-2r)} < 1$, and $e^{-\alpha t}$, $\gamma^{[t/r]}$ and $e^{-\alpha(t+2r)}$ all approach zero exponentially, $||x(t,\psi,\phi)|| \to 0$ exponentially. Similarly, we can show that $||y(t,\psi,\phi)|| \to 0$ exponentially.

4.9.3 Lyapunov-Krasovskii functional

We will assume the system described by (4.39) and (4.40) is exponentially stable. We will construct a Lyapunov-Krasovskii functional $V(x(t), y_t)$ such that

$$\dot{V}(x(t), y_t) = -x^T(t)Wx(t), \tag{4.56}$$

for any given positive definite matrix W. For this purpose, one may choose

$$V(\psi,\phi) = \int_0^\infty x^T(t,\psi,\phi)Wx(t,\psi,\phi)dt. \tag{4.57}$$

In other words,

$$V(x(t), y_t) = \int_0^\infty x^T(\xi, x(t), y_t)Wx(\xi, x(t), y_t)d\xi.$$

Then, it is easily shown that

$$\begin{aligned}\dot{V}(x(t), y_t) &= \int_0^\infty \frac{\partial}{\partial t}[x^T(\xi, x(t), y_t)Wx(\xi, x(t), y_t)]d\xi \\ &= \int_0^\infty \frac{\partial}{\partial t}[x^T(\xi + t, \psi, \phi)Wx(\xi + t, \psi, \phi)]d\xi \\ &= \int_0^\infty \frac{\partial}{\partial \xi}[x^T(\xi + t, \psi, \phi)Wx(\xi + t, \psi, \phi)]d\xi \\ &= \int_0^\infty \frac{\partial}{\partial \xi}[x^T(\xi, x(t), y_t)Wx(\xi, x(t), y_t)]d\xi \\ &= x^T(\xi, x(t), y_t)Wx(\xi, x(t), y_t)|_{\xi=0}^{\xi=\infty},\end{aligned}$$

or

$$\dot{V}(x(t), y_t) = -x^T(t)Wx(t). \tag{4.58}$$

Using the general solution (4.52) and (4.53), $V(\psi, \phi)$ can be expressed in an explicit quadratic form of (ψ, ϕ). Indeed, using (4.52) and (4.53) in (4.57), it is easily obtained that

$$\begin{aligned}V(\psi, \phi) = \psi^T U_{xx}\psi + 2\psi^T \int_{-r}^0 U_{xy}(\eta)\phi(\eta)d\eta \\ + \int_{-r}^0 \int_{-r}^0 \phi^T(\xi)U_{yy}(\xi, \eta)\phi(\eta)d\xi d\eta,\end{aligned} \tag{4.59}$$

where

$$U_{xx} = \int_0^\infty X_x^T(\theta)WX_x(\theta)d\theta, \tag{4.60}$$

$$U_{xy}(\eta) = \int_0^\infty X_x^T(\theta)WX_y(\theta - \eta)d\theta, \tag{4.61}$$

$$U_{yy}(\xi, \eta) = \int_0^\infty X_y^T(\theta - \xi)WX_y(\theta - \eta)d\theta. \tag{4.62}$$

These are clearly well defined and finite since both X_x and X_y are exponentially decaying matrix functions. Also, it is easy to see that U_{xx} is positive definite.

4.9.4 Further Comments

The discussions so far established the following fact.

Proposition 4.8. *If the system described by (4.39) and (4.40) is exponentially stable, and $\rho(D) < 1$. Then, there exists a quadratic Lyapunov-Krasovskii functional in the form of*

$$
\begin{aligned}
V(x(t), y_t) = x^T(t)Px(t) + x^T \int_{-r}^{0} Q(\eta)y(t+\eta)d\eta \\
+ \int_{-r}^{0}\int_{-r}^{0} y^T(t+\xi)R(\xi,\eta)y(t+\eta)d\xi d\eta \\
+ \int_{-r}^{0} y^T(t+\eta)S(\eta)y(t+\eta)d\eta,
\end{aligned}
$$

such that

$$\varepsilon||x(t)||^2 \le V(x(t), y_t) \le M||(x(t), y_t)||^2,$$

and

$$\dot{V}(x(t), y_t) \le -\varepsilon||x(t)||^2,$$

for some $\varepsilon > 0$ *and* $M > 0$.

The quadratic form of V and its derivative makes it possible for discretization in a similar scheme as described in [9]. It should be pointed out that even for retarded time-delay systems, the above description has its advantages. First, for systems with multiple commensurate delays, while it is possible to use the scheme described in Chapter 7 of [11] to handle this case, the formulation here is much simpler and the computation would be substantially reduced. Second, in many practical cases, the delay occurs only in a limited part of the system. For a system with single delay (4.7), this means that A_1 has significantly lower rank than the number of states. In this case, we may write $A_1 = FG$, where F has full column rank and G has full row rank. Then, we can write the system as

$$
\begin{aligned}
\dot{x}(t) &= A_0x(t) + Fy(t-r), \\
y(t) &= Gx(t).
\end{aligned}
$$

In this way, since the dimension of y is significantly lower than x, the dimension of LMI resulted from discretization is significantly reduced.

4.10 Conclusions

A number of basic ideas regarding Lyapunov approach of time-delay systems are discussed. The main emphasis is on the presentation of main ideas and motivations. The readers who wish to explore further are referred to references for technical details.

Another interesting topic is dealing with uncertainties. The readers are referred to [11] and [23].

References

1. C. T. Abdallah, P. Dorato, J. Benítez-Read, and R. Byrne. Delayed positive feedback can stabilize oscillatory systems. *Proceedings of 1993 American Control Conference,* San Francisco, CA, 3106–3107, 1993.
2. O. Bilous and N. Admundson. Chemical reactor stability and sensitivity. *AI ChE Journal*, 1:513-521, 1955.
3. E.-K. Boukas and Z. K. Liu. *Deterministic and Stochastic Time-Delayed Systems.* Birkhauser, Boston, 2001.
4. S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* SIAM, Philadelphia, 1994.
5. S. Deb and R. Srikant. Global stability of congestion controllers for the Internet. *IEEE Transactions on Automatic Control,* 48(6):1055–1060, 2003.
6. L. E. El'sgol'ts and S. B. Norkin. *Introduction to the theory and applications of differential equations with deviating arguments,* Mathematics in Science and Eng., **105**, Academic Press, New York, 1973.
7. E. Fridman and U. Shaked. Delay-dependent stability and H_∞ control: constant and time-varying delays. *Int. J. Control*, 76(1):48-60, 2003.
8. P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali. *LMI Control Toolbox for Use with MATLAB*, Mathworks, Natick, MA, 1995.
9. K. Gu. A further refinement of discretized Lyapunov functional method for the stability of time-delay systems. *Int. J. Control*, 74(10):967–976, 2001.
10. K. Gu. Stability analysis of DAE: a Lyapunov approach. *6th SIAM Conference on Control and its Applications*, MS27(4), New Orleans, LA, July 11-14.
11. K. Gu, V. L. Kharitonov, and J. Chen. *Stability of time-delay systems,* Birkhauser, Boston, 2003.
12. K. Gu and S.-I. Niculescu. Additional dynamics in transformed time-delay systems. *IEEE Trans. Auto. Control*, 45(3):572–575, 2000.
13. J. K. Hale and S. M. Verduyn Lunel. *Introduction to Functional Differential Equations*, Springer-Verlag, New York, 1993.
14. Q.-L. Han and K. Gu. Stability of linear systems with time-varying delay: a generalized discretized Lyapunov functional approach. *Asian Journal of Control,* 3(3):170–180, 2001.
15. E. F. Infante and W. V. Castelan. A Lyapunov functional for a matrix difference-differential equation. *J. Diff. Equations,* 29:439–451, 1978.
16. C.-Y. Kao and B. Lincoln. Simple stability criteria for systems with time-varying delays. *Automatica,* 40(8):1429-1434, 2004.
17. F. P. Kelly. Mathematical modelling of the internet, in *Mathematics unlimited - 2001 and beyond* (Eds. B. Engquist, W. Schmid), Springer-Verlag: Berlin, 685-702, 2001.
18. V. Kolmanovskii and A. Myshkis. *Introduction to the Theory and Applications of Functional Differential Equations*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
19. N. N. Krasovskii. *Stability of Motion* [Russian], Moscow, 1959; [English translation] Stanford University Press, Stanford, CA, 1963.
20. B. Lehman. Stability of chemical reactions in a CSTR with delayed recycle stream. *Proc. 1994 Amer. Contr. Conf.*, Baltimore, Maryland, U.S.A., 3521-3522, 1994.

21. B. Lehman and E. I. Verriest. Stability of a continuous stirred reactor with delay in the recycle streams. *Proc. 30th IEEE Conf. Dec. Contr.*, Brighton, England, 1875-1876, 1991.
22. F. Mazenc and S.-I. Niculescu. Remarks on the stability of a class of TCP-like congestion control models. *Proc 42nd IEEE Conf. Dec. Contr.*, Maui, Hawaii, 2003.
23. S.-I. Niculescu. *Delay effects on stability: A robust control approach,* Springer-Verlag: Heidelberg, Germany, LNCIS, vol. **269**, 2001.
24. S.-I. Niculescu and Vl. Rasvan. Delay-independent stability in lossless propagation models with applications, part I and II. *Proc. MTNS 2000*, Perpignan, France, 2000.
25. P. Park. A delay-dependent stability for systems with uncertain time-invariant delays. *IEEE Trans. Auto. Control*, 44(4):876–877, 1999.
26. D. Perlmutter. *Stability of chemical reactors,* Prentice Hall, New Jersey, 1972.
27. Y. M. Repin. Quadratic Lyapunov functionals for systems with delay. [Russian] *Prikl. Mat. Meh.,* 29:564–566, 1965.
28. S. Shakkottai and R. Srikant. How good are deterministic fluid models of internet congestion control. *Proc. IEEE INFOCOM*, New York, NY, USA, 2002.

5

Controllability of Partial Differential Equations

Yacine Chitour[1] and Emmanuel Trélat[2]

[1] Université Paris-Sud,
Laboratoire des Signaux et Systèmes, C.N.R.S., Supélec, 3, Rue Joliot Curie,
91192 Gif-sur-Yvette, France. E-mail: chitour@lss.supelec.fr

[2] Université Paris-Sud, Laboratory AN-EDP, CNRS UMR 8628, Orsay, France.
E-mail: emmanuel.trelat@math.u-psud.fr

5.1 Semigroup Theory, and Cauchy Problems in Banach Spaces

In this section, we recall some basic elements of semigroup theory (see [25]). In particular, all the arguments of the results mentioned below can be found in [25].

5.1.1 Definitions

Let X be a Banach space.

Definition 5.1. *A one-parameter family $(S(t))_{t\geq 0}$ of bounded linear operators from X into X is called a semigroup of bounded linear operators on X if*

- $S(0) = I$,
- $S(t+s) = S(t)S(s)$, *for all* $t, s \geq 0$.

The linear operator $A : D(A) \to X$, defined on the domain

$$D(A) = \left\{ y \in X \mid \lim_{\substack{t\to 0\\ t>0}} \frac{S(t)y - y}{t} \text{ exists} \right\},$$

by

$$Ay = \lim_{\substack{t\to 0\\ t>0}} \frac{S(t)y - y}{t},$$

for $y \in D(A)$, is called the infinitesimal generator of the semigroup $S(t)$.

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 171–198, 2006.

Definition 5.2. *A semigroup $S(t)$ of bounded linear operators is said*

- *uniformly continuous if*
$$\lim_{\substack{t\to 0\\ t>0}} \|S(t)-I\| = 0;$$
- *strongly continuous (or C_0 semigroup) if*
$$\lim_{\substack{t\to 0\\ t>0}} S(t)y = y,$$
for every $y \in X$.

Theorem 5.1. *A linear operator A is the infinitesimal generator of a uniformly continuous semigroup if and only if A is bounded.*

In what follows, $\rho(A)$ denotes the resolvent set of A, that is, the set of complex numbers λ such that $\lambda I - A$ is boundedly invertible. For $\lambda \in \rho(A)$, let
$$R(\lambda, A) = (\lambda I - A)^{-1}$$
denote the resolvent of A.

Theorem 5.2. *Let $S(t)$ be a C_0 semigroup. There exist constants $\omega \geq 0$ and $M \geq 1$ such that*
$$\|S(t)\| \leq M\mathrm{e}^{\omega t},$$
for every $t \geq 0$.
A linear operator A is the infinitesimal generator of $S(t)$ if and only if

(i) A is closed, and $D(A)$ is dense in X;

(ii)$(\omega, +\infty) \subset \rho(A)$, and
$$\|R(\lambda, A)^n\| \leq \frac{M}{(\lambda-\omega)^n},$$
for every λ having a real part $\mathrm{Re}\lambda > \omega$*, and every $n \in I\!N^*$.*

Remark 5.1. Let $S(t)$ be a semigroup satisfying
$$\|S(t)\| \leq M\mathrm{e}^{\omega t},$$
for some $\omega \geq 0$ and $M \geq 1$. Then,
$$\{\lambda \in \mathbb{C} \mid \mathrm{Re}\lambda > \omega\} \subset \rho(A),$$
and
$$R(\lambda, A)y = (\lambda I - A)^{-1}y = \int_0^{+\infty} \mathrm{e}^{-\lambda t} S(t)y \, dt,$$
for every $y \in X$, and every λ such that $\mathrm{Re}\lambda > \omega$.

5.1.2 The Cauchy Problem

Classical Solutions

Let $A : D(A) \to X$ be a linear operator on the Banach space X, such that $D(A)$ is dense in X. Consider the Cauchy problem

$$\begin{aligned} \dot{y}(t) &= Ay(t), \quad \text{for } t \geq 0, \\ y(0) &= y_0 \in D(A). \end{aligned} \tag{5.1}$$

Theorem 5.3. *Suppose that A is the infinitesimal generator of a C_0 semigroup $S(t)$ on X. Then, the Cauchy problem (5.1) has a unique solution*

$$y(\cdot) \in C^0(0,T;D(A)) \cap C^1(0,T;X),$$

given by

$$y(t) = S(t)y_0,$$

for every $t \geq 0$.

Example 5.1. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set having a C^1 boundary. The Cauchy problem

$$\begin{aligned} \dot{y} &= \triangle y \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= 0, \\ y(0) &= y_0 \in H_0^1(\Omega), \end{aligned}$$

has a unique solution

$$y(\cdot) \in C^0(0,+\infty;H_0^1(\Omega)) \cap C^1(0,+\infty;H^{-1}(\Omega)).$$

Moreover, there exist $M, \omega > 0$ such that

$$\|y(t,\cdot)\|_{L^2(\Omega)} \leq M\mathrm{e}^{-\omega t}\|y_0(\cdot)\|_{L^2(\Omega)}.$$

Example 5.2. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set having a C^1 boundary. The Cauchy problem

$$\begin{aligned} \ddot{y} &= \triangle y \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= 0, \\ y(0) &= y_0 \in H_0^1(\Omega),\ \dot{y}(0) = y_1 \in L^2(\Omega), \end{aligned}$$

has a unique solution

$$y(\cdot) \in C^0(0,+\infty;H_0^1(\Omega)) \cap C^1(0,+\infty;L^2(\Omega)) \cap C^1(0,+\infty;H^{-1}(\Omega)).$$

Moreover,

$$\|\dot{y}\|^2_{H^{-1}(\Omega)} + \|y\|^2_{L^2(\Omega)} = \|y_1\|^2_{H^{-1}(\Omega)} + \|y_0\|^2_{L^2(\Omega)}.$$

Note that, if the boundary of Ω is of class C^2, and if

$$y_0 \in H^2(\Omega) \cap H^1_0(\Omega), \text{ and } y_1 \in H^1_0(\Omega),$$

then

$$y(\cdot) \in C^0(0,+\infty; H^2(\Omega) \cap H^1_0(\Omega)) \cap C^1(0,+\infty; H^1_0(\Omega)) \cap C^1(0,+\infty; L^2(\Omega)),$$

and

$$\|\dot{y}\|^2_{L^2(\Omega)} + \|y\|^2_{H^1_0(\Omega)} = \|y_1\|^2_{L^2(\Omega)} + \|y_0\|^2_{H^1_0(\Omega)}.$$

If $y_0 \in X \setminus D(A)$, then, in general, $y(t) = S(t)y_0 \notin D(A)$, and thus, $y(t)$ is not solution of (5.1) in the usual sense. Actually, $y(t)$ is solution in a weaker sense.

Weak Solutions

Let $S(t)$ be a C_0 semigroup on the Banach space X, with generator $A : D(A) \to X$. Let $\beta \in \rho(A)$ (if X is real, consider a real such number β).

Definition 5.3. *Let X_1 denote the Banach space $D(A)$, equipped with the norm*

$$\|y\|_1 = \|(\beta I - A)y\|,$$

and let X_{-1} denote the completion of X with respect to the norm

$$\|y\|_{-1} = \|(\beta I - A)^{-1}y\| = \|R(\beta, A)y\|.$$

It is not difficult to prove that the norm $\| \ \|_1$ on X_1 is equivalent to the graph norm $\|y\|_G = \|y\| + \|Ay\|$. Therefore, from the closed graph theorem,

- $(X_1, \| \ \|_1)$ is complete,
- we get an equivalent norm, for any $\beta' \in \rho(A)$.

On the other part, the space X_{-1} does not depend on the specific value of $\beta \in \rho(A)$.

Example 5.3. Let $\Omega \subset \mathbb{R}^n$ be an open bounded set having a C^2 boundary. Then, $A = -\triangle : H^1_0(\Omega) \cap H^2(\Omega) \to L^2(\Omega)$ is an isomorphism. Set $X = L^2(\Omega)$. Then,

$$X_1 = D(A) = H^1_0(\Omega) \cap H^2(\Omega),$$

and

$$X_{-1} = (H^1_0(\Omega) \cap H^2(\Omega))',$$

where the dual is considered with respect to the pivot space $X = L^2(\Omega)$.

Note that the construction can be generalized so as to obtain a *scale of Banach spaces* $(X_\alpha)_{\alpha\in\mathbb{R}}$.

Definition 5.4. *The adjoint operator* $A^* : D(A^*) \to X'$*, of the operator* A*, is defined by*

$$D(A^*) = \{x \in X' \mid \exists y \in X',\ \forall z \in D(A) \quad \langle x, Az\rangle X', X = \langle y, z\rangle_{X',X}\},$$

and, if $x \in D(A^*)$*, then* $y = A^*x$*.*

Note that, since $D(A)$ is dense in X, there exists at most one such y.

We endow $D(A^*)$ with the graph norm

$$\|y\|_1 = \|(\beta I - A^*)y\|_{X'},$$

where $\beta \in \rho(A^*) = \rho(A)$.

Note that, if X is reflexive, and if $S(t)$ is a C_0 semigroup on X with generator A, then $S(t)^*$ is a C_0 semigroup on X' with generator A^*.

Theorem 5.4. *If* X *is reflexive, then* X_{-1} *is isomorphic to* $D(A^*)'$*.*

Remark 5.2. One has $X_1 \subset X \subset X_{-1}$, with continuous and dense embeddings.

Theorem 5.5. *The operator* $A : D(A) \to X$ *extends to an operator* $A_{-1} : D(A_{-1}) = X \to X_{-1}$*, and the semigroup* $S(t)$ *on* X *extends to a semigroup* $S_{-1}(t)$ *on* X_{-1}*, generated by* A_{-1}*.*

Definition 5.5. *For every* $y_0 \in X$*, the unique solution*

$$y(t) = S(t)y_0$$

of the Cauchy problem

$$\begin{aligned} \dot{y}(t) &= A_{-1}y(t), \quad \textit{for } t \geq 0, \\ y(0) &= y_0, \end{aligned}$$

in the space

$$C^0(0, +\infty; X) \cap C^1(0, +\infty; X_{-1}),$$

is called a weak solution.

Example 5.4. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set having a C^2 boundary. The Cauchy problem

$$\begin{aligned} \dot{y} &= \triangle y \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= 0, \\ y(0) &= y_0 \in L^2(\Omega), \end{aligned}$$

has a unique (weak) solution

$$y \in C^0(0,+\infty;L^2(\Omega)) \cap C^1(0,+\infty;(H_0^1(\Omega) \cap H^2(\Omega))').$$

Moreover, there exist $M,\omega > 0$ such that

$$\|y(t,\cdot)\|_{L^2(\Omega)} \leq M\mathrm{e}^{-\omega t}\|y_0(\cdot)\|_{L^2(\Omega)}.$$

Example 5.5. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set having a C^2 boundary. Consider the Cauchy problem

$$\begin{aligned} \ddot{y} &= \triangle y \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= 0, \\ y(0) &= y_0,\ \dot{y}(0) = y_1. \end{aligned}$$

- If $y_0 \in H^{-1}(\Omega)$ and $y_1 \in (H_0^1(\Omega) \cap H^2(\Omega))'$, then there is a unique solution

$$y(\cdot) \in C^0(0,+\infty;H^{-1}(\Omega)) \cap C^1(0,+\infty;(H_0^1(\Omega) \cap H^2(\Omega))').$$

- If $y_0 \in L^2(\Omega)$ and $y_1 \in H^{-1}(\Omega)$, then there is a unique solution

$$y(\cdot) \in C^0(0,+\infty;L^2(\Omega)) \cap C^1(0,+\infty;H^{-1}(\Omega)).$$

5.1.3 The Nonhomogeneous Initial-value Problem

Consider the Cauchy problem

$$\begin{aligned} \dot{y}(t) &= Ay(t) + f(t), \quad \text{for } t \geq 0, \\ y(0) &= y_0, \end{aligned} \tag{5.2}$$

where $A : D(A) \to X$ generates a C_0 semigroup $S(t)$ on X.

Theorem 5.6. *If $y_0 \in D(A)$, and $f \in L^1(0,T;D(A))$, then (5.2) admits a unique solution*

$$y \in C^0(0,T;D(A)) \cap C^1(0,T;X),$$

given by

$$y(t) = S(t)y_0 + \int_0^t S(t-s)f(s)ds. \tag{5.3}$$

Note that, if $f \in L^1(0,T;X)$, (5.3) still makes sense.

Definition 5.6. • *If $y_0 \in X$ and $f \in L^1(0,T;X)$, then y defined by (5.3) is called mild solution of (5.2).*

- *If $y_0 \in D(A)$ and $f \in C^0(0,T;X)$, and if*

  $$y \in C^0(0,T;D(A)) \cap C^1(0,T;X),$$

  then y defined by (5.3) is called strong solution of (5.2).
- *Assume X reflexive. If $y_0 \in X_{-1} \simeq D(A^*)'$, and $f \in L^1(0,T;X_{-1})$, then y defined by*

  $$y(t) = S_{-1}(t)y_0 + \int_0^t S_{-1}(t-s)f(s)ds,$$

  is called weak solution of (5.2).

Remark 5.3. The condition $f \in C^0(0,T;X)$ does not ensure the existence of strong solutions. However, we have the following result.

Theorem 5.7. *If $y_0 \in D(A)$ and $f \in C^1(0,T;X)$, then (5.2) has a unique strong solution.*

Corollary 5.1. *If $y_0 \in X$ and $f \in C^1(0,T;X_{-1})$ (or $f \in W^{1,1}(0,T;X_{-1})$), then (5.2) has a unique weak solution, such that*

$$y \in C^0(0,T;X) \cap C^1(0,T;X_{-1}).$$

5.2 Controllability and Observability in Banach Spaces

5.2.1 A Short Overview on Controllability of Finite-dimensional Linear Control Systems

We start the section by recalling some well known results in the finite-dimensional context.

Let $T > 0$ be fixed. Consider the linear control system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{5.4}$$

where $x(t) \in \mathbb{R}^n$, A is a $(n \times n)$-matrix, B is a $(n \times m)$-matrix, with real coefficients, and $u(\cdot) \in L^2(0,T;\mathbb{R}^m)$.

Let $x_0 \in \mathbb{R}^n$. The system (5.4) is said to be *controllable from x_0 in time T* if and only if, for every $x_1 \in \mathbb{R}^n$, there exists $u(\cdot) \in L^2(0,T;\mathbb{R}^m)$ so that the solution $x(\cdot)$ of (5.4), with $x(0) = x_0$, associated with the control $u(\cdot)$, satisfies $x(T) = x_1$.

It is well known that the system (5.4) is controllable in time T if and only if the matrix

$$\int_0^T \mathrm{e}^{(T-t)A} BB^* \mathrm{e}^{(T-t)A^*} dt, \tag{5.5}$$

called *Gramian* of the system, is nonsingular (here, M^* denotes the transpose of the matrix M). Since we are in finite dimension, this is equivalent to the existence of $\alpha > 0$ so that

$$\int_0^T \|B^* \mathrm{e}^{(T-t)A^*} \psi\|^2 dt \geq \alpha \|\psi\|^2, \tag{5.6}$$

for every $\psi \in \mathbb{R}^n$ (observability inequality).

It is also well known that, if such a linear system is controllable from x_0 in time $T > 0$, then it is controllable in time T', for every $T' > 0$, and from every initial state $x_0' \in \mathbb{R}^n$. Indeed, another necessary and sufficient condition for controllability is the Kalman condition

$$\mathrm{rank}(B, AB, \ldots, A^{n-1}B) = n,$$

wich is independent on x_0 and T.

5.2.2 Controllability of Linear Partial Differential Equations in Banach Spaces

In this section we review some known facts on controllability of infinite-dimensional linear control systems in Banach spaces (see [34, 35, 31]).

The notation $L(E, F)$ stands for the set of linear continuous mappings from E to F, where E and F are Banach spaces.

We deal with the infinite-dimensional linear control system

$$\begin{aligned} \dot{y}(t) &= Ay(t) + Bu(t), \\ y(0) &= y_0, \end{aligned} \tag{5.7}$$

where the state $y(t)$ belongs to a Banach space X, the control $u(t)$ belongs to a Banach space U, $A : D(A) \rightarrow X$ is the generator of a C_0 semigroup $S(t)$ on X, and $B \in L(U, X_{-1})$.

Admissible Control and Observation Operators

The control operator B is said to be *bounded* if $B \in L(U, X)$, and is called *unbounded* otherwise (note however that B is a bounded operator from U in X_{-1}). Unbounded operators appear naturally when dealing with boundary or pointwise control systems.

a priori, (5.7) makes sense in X_{-1}, and if $u \in L^2(0, T; U)$, then

$$y(t) = S(t)y_0 + L_t u,$$

where

$$L_t u = \int_0^t S(t-s)Bu(s)ds,$$

is a weak solution (from now on, $S_{-1}(t)$ is denoted $S(t)$, for the sake of simplicity). Moreover,

$$L_t u \in X_{-1},$$

and thus

$$y \in H^1(0,T;X_{-1}).$$

The objective is to characterize control operators B such that $y(t) \in X$, for every $t \geq 0$, whenever $y_0 \in X$. Note that, if $y_0 \in X$, then $S(t)y_0 \in X$, for every $t \geq 0$.

Definition 5.7. *$B \in L(U, X_{-1})$ is called admissible control operator for $S(t)$ if the weak solution of (5.7), with $y_0 \in X$, belongs to X, whenever $u \in L^2(0,T;U)$. This is equivalent to requiring*

$$L_T \in L(L^2(0,T;U), X).$$

Note that, if B is admissible, then

$$y \in H^1(0,T;X),$$

and

$$\dot{y} = Ay + Bu \quad \text{in } X_{-1},$$

almost everywhere on $[0,T]$.

Note also that, in the term $L_t u$, the integration is done in X_{-1}, but the result is in X whenever B is admissible.

Definition 5.8. *Let Y denote a Banach space. Let $S(t)$ be a C_0 semigroup on X, with generator A, and let $C \in L(D(A), Y)$. The operator C is called admissible observation operator for $S(t)$ if, for every $T > 0$, there exists $C_T > 0$ such that*

$$\int_0^T \|CS(t)y\|_Y^2 dt \leq C_T \|y\|_X^2, \tag{5.8}$$

for every $y \in D(A)$.

a priori, (5.8) makes sense for $y \in D(A)$. For $y \in X$, one has to replace C with its *Λ-extension*

$$C_\Lambda z = \lim_{\lambda \to +\infty} C\lambda(\lambda I - A)^{-1} z,$$

also called *Lebesgue extension* (introduced in [34]). Then, replacing C with C_Λ, (5.8) makes sense, for every $y \in X$.

Theorem 5.8. *Assume that X and U are reflexive, and that $A : D(A) \to X$ is the generator of a C_0 semigroup $S(t)$ on X. Then, $B \in L(U, X_{-1})$ is an admissible control operator for $S(t)$ if and only if $B^* \in L(D(A^*), U')$ is an admissible observation operator for $S(t)^*$.*

Moreover, the adjoint L_T^* of L_T is given by

$$\begin{aligned} \forall y \in D(A^*) \quad (L_T^* x)(t) &= B^* S(T-t)^* x, \quad \forall t \in [0,T], \\ \forall y \in X' \quad (L_T^* x)(t) &= B_\Lambda^* S(T-t)^* x, \quad \textit{for a.e. } t \in [0,T], \end{aligned}$$

where, as previously,

$$B_\Lambda^* z = \lim_{\lambda \to +\infty} \lambda B^* (\lambda I - A^*)^{-1} z.$$

Note that, for B admissible, $L_T : L^2(0,T;U) \to X$, and $L_T^* : X' \to L^2(0,T;U')$.

Example 5.6. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set having a C^2 boundary. Consider the heat equation with boundary Dirichlet control

$$\begin{aligned} \dot{y} &= \triangle y \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= u(t), \\ y(0) &= y_0 \in L^2(\Omega). \end{aligned}$$

Set $X = L^2(\Omega)$, and $A = \triangle : D(A) \to X$, where

$$D(A) = X_1 = H^2(\Omega) \cap H_0^1(\Omega).$$

The operator A is selfadjoint, and

$$X_{-1} = D(A^*)' = (H^2(\Omega) \cap H_0^1(\Omega))',$$

with respect to the pivot space $L^2(\Omega)$. Then,

$$B^* \phi = -\frac{\partial \phi}{\partial \nu}_{|\partial\Omega},$$

for every $\phi \in D(A^*)$, and B is defined by transposition

$$\langle Bu, \phi \rangle_{(H^2(\Omega)\cap H_0^1(\Omega))', H^2(\Omega)\cap H_0^1(\Omega)} = -\int_{L^2(\partial\Omega)} u \frac{\partial \phi}{\partial \nu}_{|\partial\Omega},$$

for every $u \in L^2(\partial\Omega)$, and every $\phi \in H^2(\Omega) \cap H_0^1(\Omega)$.

Then, B is an admissible control operator, if and only if, B^* is an admissible observation operator, if and only if, for every $T > 0$, there exists $C_T > 0$ such that, for every $\psi_0 \in H^2(\Omega) \cap H_0^1(\Omega)$, the solution of

$$\begin{aligned}\dot{\psi} &= \triangle\psi \quad \text{in } \Omega,\\ \psi_{\partial\Omega} &= 0,\\ \psi(0) &= \psi_0,\end{aligned}$$

satisfies

$$\int_0^T \left\| \frac{\partial\psi}{\partial\nu}_{|\partial\Omega}(t) \right\|^2_{L^2(\partial\Omega)} dt \leq C_T \|\psi_0\|^2_{L^2(\Omega)}.$$

This inequality indeed holds: this is a classical trace regularity result.

Another typical example is provided by second-order equations. The framework is the following (see [31, 32, 33]). Let H be a Hilbert space, and $A_0 : D(A_0) \to H$ be selfadjoint and strictly positive. Recall that $D(A_0^{1/2})$ is the completion of $D(A_0)$ with respect to the norm

$$\|y\|_{D(A_0^{1/2})} = \sqrt{\langle A_0 y, y\rangle_H},$$

and that

$$D(A_0) \subset D(A_0^{1/2}) \subset H,$$

with continuous and dense embeddings. Set

$$X = D(A_0^{1/2}) \times H,$$

and define $A : D(A) \to X$ on

$$D(A) = D(A_0) \times D(A_0^{1/2}),$$

by

$$A = \begin{pmatrix} 0 & I \\ -A_0 & 0 \end{pmatrix}.$$

Note that A is skew-adjoint in X.

Let $B_0 \in L(U, D(A_0^{1/2})'$, where U is a Hilbert, and $D(A_0^{1/2})'$ is the dual of $D(A_0^{1/2})$ with respect to the pivot space U. We investigate the second-order control system

$$\begin{aligned} y_{tt} + A_0 y &= B_0 u,\\ y(0) = y_0,\ y_t(0) &= y_1. \end{aligned} \tag{5.9}$$

It can be written in the form

$$\frac{\partial}{\partial t}\begin{pmatrix} y \\ y_t \end{pmatrix} = A \begin{pmatrix} y \\ y_t \end{pmatrix} + Bu,$$

where

$$B = \begin{pmatrix} 0 \\ B_0 \end{pmatrix}.$$

One has

$$X_{-1} = D(A^*)' = H \times D(A_0^{1/2})',$$

with respect to the pivot space X, where $D(A_0^{1/2})'$ is the dual of $D(A_0^{1/2})$ with respect to the pivot space H. Moreover, $B \in L(U, H \times D(A_0^{1/2})')$, and $B^* \in L(D(A_0) \times D(A_0^{1/2}), U)$ is given by

$$B^* = \begin{pmatrix} 0 \\ B_0^* \end{pmatrix}.$$

Proposition 5.1. *The following statements are equivalent:*

- *B is admissible;*
- *There exists $C_T > 0$ such that every solution of*

$$\begin{aligned} &\psi_{tt} + A_0 \psi = 0, \\ &\psi(0) = \psi_0 \in D(A_0),\ \psi_t(0) = \psi_1 \in D(A_0^{1/2}), \end{aligned}$$

  satisfies

$$\int_0^T \| B_0^* \psi_t(t) \|_U^2 dt \leq C_T \left(\|\psi_0\|^2_{D(A_0^{1/2})} + \|\psi_1\|_H^2 \right).$$

- *There exists $C_T > 0$ such that every solution of*

$$\begin{aligned} &\psi_{tt} + A_0 \psi = 0, \\ &\psi(0) = \psi_0 \in H,\ \psi_t(0) = \psi_1 \in D(A_0^{1/2})', \end{aligned}$$

  satisfies

$$\int_0^T \| B_0^* \psi(t) \|_U^2 dt \leq C_T \left(\|\psi_0\|_H^2 + \|\psi_1\|^2_{D(A_0^{1/2})'} \right).$$

Example 5.7. Consider the boundary controlled wave equation

$$\begin{aligned} y_{tt} &= \triangle y \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= u(t). \end{aligned}$$

Set $H = H^{-1}(\Omega)$, and consider the operator

$$A_0 = -\triangle : D(A_0) = H_0^1(\Omega) \to H.$$

Then, A_0 is an isomorphism from $D(A_0)$ in H, and

$$D(A_0^{1/2}) = L^2(\Omega),$$

and the dual space $(D(A_0^{1/2}))'$ (with respect to the pivot space $H = H^{-1}(\Omega)$) is equal to the dual space $(H^2(\Omega) \cap H_0^1(\Omega))'$ (with respect to the pivot space $L^2(\Omega)$. Indeed, the operator A_0 can be extended as an operator $A_{-1} : L^2(\Omega) \to (H^2(\Omega) \cap H_0^1(\Omega))'$. On the other part, set

$$U = L^2(\partial\Omega),$$

and

$$X = D(A_0^{1/2}) \times H = L^2(\Omega) \times H^{-1}(\Omega).$$

Then, the controlled wave equation writes

$$y_{tt} = -A_0 y + B_0 u \quad \text{in } (D(A_0^{1/2}))',$$

where

$$B_0^* \phi = \frac{\partial}{\partial \nu}(A_0^{-1}\phi)_{|\partial\Omega},$$

for every $\phi \in L^2(\Omega)$.

It is known (see [17, 18, 19]) that there exists $C_T > 0$ such that

$$\int_0^T \left\| \frac{\partial \psi}{\partial \nu}(t) \right\|_{L^2(\partial\Omega)}^2 dt \leq C_T \left(\|\psi_0\|_{H_0^1(\Omega)}^2 + \|\psi_1\|_{L^2(\Omega)}^2 \right),$$

for every $\psi \in C^0(0,T;H^2(\Omega)) \cap C^1(0,T;H^1(\Omega))$ solution of

$$\begin{aligned} &\psi_{tt} = \triangle\psi \quad \text{in } \Omega, \\ &\psi_{\partial\Omega} = 0, \\ &\psi(0) = \psi_0,\ \dot{psi}(0) = \psi_1. \end{aligned}$$

Therefore, the observation operator B^* (and thus, the control operator B) is admissible.

Note that $B_0 \in L(U, D(A_0^{1/2})')$ is given by

$$B_0 u = A_{-1} D u,$$

for every $u \in U$, where D is the Dirichlet mapping, defined by transposition by

$$\int_\Omega Du(x) f(x) dx = \int_{\partial\Omega} u(x) \frac{\partial \phi}{\partial \nu}(x) dx,$$

for all f and ϕ so that

$$\begin{aligned} &\triangle\phi = f \quad \text{in } \Omega, \\ &\phi_{\partial\Omega} = 0. \end{aligned}$$

Necessary and Sufficient Conditions for Controllability in Banach Spaces

We first recall the notations.

Let X be a Banach space. For clarity, denote by $\| \, \|_X$ the norm of X. Let $S(t)$ denote a strongly continuous semigroup on X, of generator $(A, D(A))$. Let X_{-1} denote the completion of X for the norm $\|x\|_{-1} = \|(\beta I - A)^{-1}x\|$, where $\beta \in \rho(A)$ is fixed. The space X_{-1} is isomorphic to $(D(A^*))'$. The semigroup $S(t)$ extends to a semigroup on X_{-1}, still denoted $S(t)$, whose generator is an extension of the operator A, still denoted A. With these notations, A is a linear operator from X to X_{-1}.

Let U be a Banach space. Denote by $\| \, \|_U$ the associated norm. Let $B \in L(U, X_{-1})$ be and admissible control operator. Consider the control system

$$\dot{y}(t) = Ay(t) + Bu(t), \tag{5.10}$$

with $y(0) = y_0 \in X$ and $u(\cdot) \in L^2(0, +\infty; U)$. The solution writes

$$y(t) = S(t)y_0 + \int_0^t S(t-s)Bu(s)ds, \tag{5.11}$$

for every $t \geq 0$. For $T > 0$, the operator $L_T : L^2(0, T; U) \to X_{-1}$ is defined by

$$L_T u = \int_0^T S(t-s)Bu(s)ds. \tag{5.12}$$

Note that, since B is admissible, $L_T \in L(L^2(0, T; U), X)$.

Definition 5.9. *For $y_0 \in X$, and $T > 0$, the system (5.10) is said to be exactly controllable from y_0 in time T if, for every $y_1 \in X$, there exists $u(\cdot) \in L^2(0, T; U)$ so that the solution of (5.10), with $y(0) = y_0$, associated with the control $u(\cdot)$, satisfies $y(T) = y_1$.*

It is clear that the system (5.10) is exactly controllable from y_0 in time T if and only if L_T is onto, that is $\operatorname{Im} L_T = X$. In particular, if the system (5.10) is exactly controllable from y_0 in time T, then it is exactly controllable from any point $y_0' \in X$ in time T. One says that the system (5.10) is exactly controllable in time T.

Definition 5.10. *The system (5.10) is said to be approximately controllable from y_0 in time T if, for every $y_1 \in X$ and every $\varepsilon > 0$, there exists $u(\cdot) \in L^2(0, T; V)$ so that the solution of (5.10), with $y(0) = y_0$, associated with the control $u(\cdot)$, satisfies $\|y(T) - y_1\|_X \leq \varepsilon$.*

As previously, this notion does not depend on the initial point, and the system (5.10) is approximately controllable in time T if and only if $\operatorname{Im} L_T$ is dense in X.

Definition 5.11. *For $T > 0$, the system (5.10) is said to be exactly null controllable in time T if, for every $y_0 \in X$, there exists $u(\cdot) \in L^2(0,T;U)$ so that the solution of (5.10), with $y(0) = y_0$, associated with the control $u(\cdot)$, satisfies $y(T) = 0$.*

Remark 5.4. If the system (5.17) is exactly null controllable in every time T, then it is approximately controllable in every time T.

Theorem 5.9.

- *The system (5.17) is exactly controllable in time T if and only if there exists $\alpha > 0$ so that*

$$\int_0^T \|B^* S^*(t)\psi\|_U^2 dt \geq \alpha \|\psi\|_X^2, \tag{5.13}$$

  for every $\psi \in D(A^*)$ (observability inequality). This is equivalent to saying that L_T^* is bounded below.

- *The system (5.17) is approximately controllable in time T if and only if the following implication holds:*

$$\forall t \in [0,T] \quad B^* S^*(t)\psi = 0 \;\Rightarrow \psi = 0,$$

  for every $\psi \in D(A^*)$. This is equivalent to saying that L_T^* is one-to-one.

- *The system (5.17) is exactly null controllable in time T if and only if there exists $\alpha > 0$ so that*

$$\int_0^T \|B^* S^*(t)\psi\|_U^2 dt \geq \alpha \|S(T)^*\psi\|_X^2, \tag{5.14}$$

  for every $\psi \in D(A^*)$. This is equivalent to saying that $\text{Im } S(T) \subset \text{Im} L_T$.

Remark 5.5. Assume that B is admissible and that the control system (5.10) is exactly null controllable in time T. Let $y_0 \in X$. For every $\psi \in D(A^*)$, set

$$J(\psi) = \frac{1}{2} \int_0^T \|B^* S(t)^*\psi\|_U^2 dt + \langle S(T^*)\psi, y_0 \rangle_X. \tag{5.15}$$

The functional J is strictly convex, and, from the observability inequality (5.14), is coercive. Define the control u by

$$u(t) = B^* S(T-t)^*\psi, \tag{5.16}$$

for every $t \in [0,T]$, and let $y(\cdot)$ be the solution of (5.10), such that $y(0) = y_0$, associated with the control u. Then, one has $y(T) = 0$, and moreover, u is the control of minimal L^2 norm, among all controls whose associated trajectory satisfies $y(T) = 0$.

This remark proves that observability implies controllability, and gives a constructive way to build the control of minimal L^2 norm (see [38]). This is more or less the contents of the Hilbert Uniqueness Method (see [17, 18]). Hence, in what follows, we refer to the control (5.16) as the HUM control.

The same remark holds of course for exact controllability, with the functional

$$J(\psi) = \frac{1}{2}\int_0^T \|B^*S(t)^*\psi\|_U^2 dt - \langle \psi, y_1\rangle_X + \langle S(T)^*\psi, y_0\rangle_X.$$

Example 5.8. For the heat equation of Example 5.6,

$$\begin{aligned} \dot{y} &= \triangle y \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= u(t), \\ y(0) &= y_0 \in L^2(\Omega), \end{aligned}$$

it follows from [8, 15] that, for every $T > 0$, there exists $c_T > 0$ so that, for every $\psi_0 \in H^2(\Omega) \cap H_0^1(\Omega)$, the solution of

$$\begin{aligned} \dot{\psi} &= \triangle \psi \quad \text{in } \Omega, \\ \psi_{|\partial\Omega} &= 0, \\ \psi(0) &= \psi_0, \end{aligned}$$

satisfies

$$\int_0^T \left\| \frac{\partial \psi}{\partial \nu}_{|\partial\Omega}(t) \right\|_{L^2(\partial\Omega)}^2 dt \geq c_T \|\psi_0\|_{L^2(\Omega)}^2.$$

In other words, the heat equation with boundary control is exactly null controllable, in any time $T > 0$.

Example 5.9. Consider the heat equation with distributed control

$$\begin{aligned} \dot{y} &= \triangle y + 1_{\mathcal{O}} u \quad \text{in } \Omega, \\ y_{|\partial\Omega} &= 0, \\ y(0) &= y_0 \in L^2(\Omega), \end{aligned}$$

where $\mathcal{O}$ is an open subset of Ω. It follows from [8, 15] that, for every $T > 0$, there exists $c_T > 0$ so that, for every $\psi_0 \in L^2(\Omega)$, the solution of

$$\begin{aligned} \dot{\psi} &= \triangle \psi \quad \text{in } \Omega, \\ \psi_{|\partial\Omega} &= 0, \\ \psi(0) &= \psi_0, \end{aligned}$$

satisfies

$$\int_0^T \int_{\mathcal{O}} \psi(t,x)^2 dx dt \geq c_T \int_\Omega \psi(T,x)^2 dx.$$

In other words, the heat equation with distributed control is exactly null controllable (in the space $L^2(\Omega)$, in any time $T > 0$.

Note that these observability inequalities are proved in [8, 15] using *Carleman estimates*. In both cases, note also that Holmgren's Uniqueness Theorem implies approximate controllability in $L^2(\Omega)$.

Example 5.10. Consider the wave equation of Example 5.7. It is proved in [3] that it is exactly controllable, under the so-called GCC (Geometric Control Condition), within time T sufficiently large. The time of controllability has to be large enough, because of the finite speed of propagation of the wave equation. The observability inequality has been proved in

- [7], for T large enough, with a condition on $\partial\Omega$, using multipliers methods;
- [11, 18], for T large enough, using multipliers methods;
- [3], using microlocal analysis;

and in many other references.

Note that, in dimension one, the proof of the observability inequality can be achieved easily using Fourier series, for $T \geq 2L$, where L is the length of the interval.

The observability inequality implies a result of exact controllability in the space $L^2(\Omega) \times H^{-1}(\Omega)$.

5.3 Semidiscrete Approximations of Infinite-dimensional Linear Control Systems in Hilbert Spaces

5.3.1 Introduction

Consider the infinite-dimensional linear control system

$$\begin{aligned} \dot{y}(t) &= Ay(t) + Bu(t), \\ y(0) &= y_0, \end{aligned} \tag{5.17}$$

where the state $y(t)$ belongs to a Hilbert space X, the control $u(t)$ belongs to a Hilbert space U, $A : D(A) \rightarrow X$ is an operator, and B is a control operator (in general, unbounded) on U. Discretizing this partial differential equation, using for instance a finite difference, or a finite element scheme, leads to a family of finite dimensional linear control systems

$$\begin{aligned} \dot{y}_h(t) &= Ay_h(t) + Bu_h(t), \\ y(0) &= y_{0h}, \end{aligned} \tag{5.18}$$

where $y_h(t) \in X_h$ and $u_h(t) \in U_h$, for $0 < h < h_0$.

Let $y_1 \in X$; if the control system (5.17) is controllable in time T, then there exists a solution $y(\cdot)$ of (5.17), associated with a control u, such that

$y(T) = y_1$. We address the following question: is it possible to find controls u_h, for $0 < h < h_0$, converging to the control u as the mesh size h of the discretization process tends to zero, and such that the associated trajectories $y_h(\cdot)$, solutions of (5.18), converge to $y(\cdot)$? Moreover, does there exist an efficient algorithmic way to determine the controls u_h?

For controllable linear control systems of the type (5.17), we have available many methods in order to realize the controllability. Among them, the Hilbert Uniqueness Method (HUM), introduced in [17, 18], is adapted to numerical implementations. It consists in minimizing a cost function, namely, the L^2 norm of the control. In Section 5.3.2, we answer to the above question in the case where controllability of (5.17) is achieved using the HUM method. The objective is to establish conditions ensuring *uniform controllability* of the family of discretized control systems (5.18), and to establish a computationally feasible approximation method for realizing controllability.

The question of uniform controllability and/or observability of the family of approximation control systems (5.18) has been investigated by E. Zuazua and collaborators in a series of articles [5, 9, 16, 21, 22, 24, 30, 36, 37, 38, 39], for different discretization processes, on different examples. When the observability constant of the finite dimensional approximation systems does not depend on h, one says that the property of *uniform observability* holds. For classical finite difference schemes, a uniform boundary observability property holds for one dimensional heat equations [21], beam equations [16], Schrödinger equations [39], but does not hold for 1-D wave equations [9]. In this latter case, the observability constant of the one-dimensional semidiscretized wave equation tends to infinity as the mesh size tends to zero. This is due to a pathological behavior of high frequency eigenvalues of the semidiscrete model. Actually, spurious oscillations appear, due to interferences with the mesh, that are responsible for non uniformity. From the point of view of controllability, they cause divergence of controls as the mesh size is tending to zero. These results hold for other numerical schemes, such as the classical $P_1 \times P_1$ finite elements method, and also for two-dimensional linear wave equations (see [9, 22, 36]). In the case of wave equations, several remedies are provided to reestablish uniformity: cutting off high frequencies by Fourier filtering; Tychonoff regularization, which consists in adding a viscosity term to the semidiscrete model; two-grid algorithms, which consist in using different sized grids for filtering solutions; the use of mixed finite elements, namely, $P_1 \times P_0$ finite elements. This latter method, used in [5], is interesting from the practical point of view, because it is simple to implement, and does not require any further filtering procedure or extra corrections. Moreover, from the theoretical point of view it seems natural because it takes into account the natural difference of regularity between u and u_t, for the wave equation. The case of the wave equation is hence quite involved. In contrast, it seems that, for 1-D heat, beam and Schrödinger equations, the dissipative and/or dispersive effects help to recover some uniformity.

The HUM method is not the unique method to discretize the control. We can imagine other ways to realize the controllability for (5.18), with the property $u_h \to u$. Related to this problem is the problem of uniform *stabilizability.* The question is the following: is it true that (5.17) is stabilizable if and only if (5.18) is uniformly stabilizable?

Recall that (5.17) is stabilizable if there exists $K \in L(X, U)$ so that $A+BK$ generates an exponentially stable semigroup $S(t)$, that is,

$$\exists M, \omega > 0 \mid \forall t \geq 0 \quad \|S(t)\| \leq M\mathrm{e}^{-\omega t}.$$

On the other part, (5.18) is said uniformly stabilizable if, for every $h \in (0, h_0)$, there exists a $(m \times n)$ matrix K_h such that the matrix $A_h + B_h K_h$ is uniformly exponentially stable, that is,

$$\exists M, \omega > 0 \mid \forall t \geq 0 \quad \forall h \in (0, h_0) \quad \|\mathrm{e}^{t(A_h + B_h K_h)}\| \leq M\mathrm{e}^{-\omega t}.$$

This question has been widely investigated, in particular, in the context of the *Riccati theory.* In [1, 2, 6, 10, 14, 20, 26], approximation results are provided for the *linear quadratic regulator (LQR) problem* in the *parabolic case*, or in the *hyperbolic damped case*, that show, in particular, the convergence of the controls of the semidiscrete models to the control of the continuous model. However, in the LQR problem, the final point is not fixed. The exact controllability problem is a very different matter. Actually, it appears from the discussion above that the divergence of the controls as the mesh size tends to zero in the exact controllability problem is due to is the requirement to drive the final state exactly to a given point.

Note that, as expected, this problem disappears for the approximate controllability problem, which can be seen as a relaxation of the exact controllability problem (see [38]).

5.3.2 Uniform Controllability of Semidiscrete Approximations of Parabolic Control Systems

We saw previously that controlling an approximation model of a controllable infinite dimensional linear control system does not necessarily yield a good approximation of the control needed for the continuous model. In this section, we report on recent results obtained in [12], in which it is proved that, under the main assumptions that the discretized semigroup is uniformly analytic, and that the control operator is mildly unbounded, the semidiscrete approximation models are uniformly controllable.

The discretization framework used here is in the same spirit as the one of [1, 2, 6, 10, 14, 20, 26].

The question of uniform controllability of the discretized models (5.18) is investigated in the case where the operator A generates an analytic semigroup.

Of course, due to regularization properties, the control system (5.17) is not exactly controllable in general. Hence, we focus on exact null controllability. The main result, Theorem 5.10, states that, for an exactly null controllable parabolic system (5.17), under standard assumptions on the discretization process (that are satisfied for most of classical schemes), if the discretized semigroup is *uniformly analytic* (see [14]), and if the *degree of unboundedness* of the control operator B with respect to A (see [27]) is lower than 1/2, then the approximating control systems are uniformly controllable. A uniform observability and admissibility inequality is proved. Moreover, we provide a minimization procedure to compute the approximation controls. Note that this condition on the degree of unboundedness of B is satisfied for distributed controls (that is, if B is bounded), and, if B is unbounded, it is for instance satisfied for the heat equation with Neumann boundary control, but not with Dirichlet boundary control.

The precise results are as follows.

Let X and U be Hilbert spaces, and let $A : D(A) \to X$ be a linear operator, generating a strongly continuous semigroup $S(t)$ on X. Let $B \in L(U, D(A^*)')$ be a control operator. We make the following assumptions.

($\mathbf{H_1}$) *The semigroup $S(t)$ is analytic.*

Therefore (see [25]), there exist positive real numbers C_1 and ω such that

$$\|S(t)y\|_X \leq C_1 \mathrm{e}^{\omega t}\|y\|_X, \text{ and } \|AS(t)y\|_X \leq C_1 \frac{\mathrm{e}^{\omega t}}{t}\|y\|_X, \tag{5.19}$$

for all $t > 0$ and $y \in D(A)$, and such that, if we set

$$\hat{A} = A - \omega I, \tag{5.20}$$

then the fractional powers $(-\hat{A})^\theta$ of $\hat{A}$ are well defined, for $\theta \in [0, 1]$, and there holds

$$\|(-\hat{A})^\theta S(t)y\|_X \leq C_1 \frac{\mathrm{e}^{\omega t}}{t^\theta}\|y\|_X, \tag{5.21}$$

for all $t > 0$ and $y \in D(A)$.

Of course, the inequalities (5.19) hold as well if one replaces A by A^*, $S(t)$ by $S(t)^*$, for $y \in D(A^*)$.

Moreover, if $\rho(A)$ denotes the *resolvent set* of A, then there exists $\delta \in (0, \pi/2)$ such that

$$\rho(A) \supset \Delta_\delta = \{\omega + \rho \mathrm{e}^{i\theta} \mid \rho > 0,\ |\theta| \leq \frac{\pi}{2} + \delta\}. \tag{5.22}$$

For $\lambda \in \rho(A)$, denote by $R(\lambda, A) = (\lambda I - A)^{-1}$ the resolvent of A. It follows from the previous estimates that there exists $C_2 > 0$ such that

$$\|R(\lambda, A)\|_{L(X)} \leq \frac{C_2}{|\lambda - \omega|}, \text{ and } \|AR(\lambda, A)\|_{L(X)} \leq C_2, \tag{5.23}$$

for every $\lambda \in \varDelta_\delta$, and

$$\|R(\lambda, \hat{A})\|_{L(X)} \leq \frac{C_2}{|\lambda|}, \text{ and } \|\hat{A}R(\lambda, \hat{A})\|_{L(X)} \leq C_2, \tag{5.24}$$

for every $\lambda \in \{\varDelta_\delta + \omega\}$. Similarly, Inequalities (5.23) and (5.24) hold as well, with A^* and $\hat{A}^*$.

(H$_\mathbf{2}$) *The degree of unboundedness of B is lower than $1/2$. In other words, there exists $\gamma \in [0, 1/2)$ such that*

$$B \in L(U, D((-\hat{A}^*)^\gamma)'). \tag{5.25}$$

In these conditions, the domain of B^* is $D(B^*) = D((-\hat{A}^*)^\gamma)$. Moreover, there exists a constant $C_3 > 0$ such that

$$\|B^*\psi\|_U \leq C_3 \|(-\hat{A}^*)^\gamma \psi\|_X, \tag{5.26}$$

for every $\psi \in D((-\hat{A}^*)^\gamma)$.

Note that this assumption implies that the control operator B is *admissible.*

We next introduce adapted approximation assumptions, inspired by [14] (see also [1, 2, 6, 10, 20, 26]). Consider two families $(X_h)_{0<h<h_0}$ and $(U_h)_{0<h<h_0}$ of finite dimensional spaces, where h is the discretization parameter.

(H$_\mathbf{3}$) *For every $h \in (0, h_0)$, there exist linear mappings $P_h : D((-\hat{A}^*)^{1/2})' \to X_h$ and $\widetilde{P}_h : X_h \to D((-\hat{A}^*)^{1/2})$ (resp., there exist linear mappings $Q_h : U \to U_h$ and $\widetilde{Q}_h : U_h \to U$), satisfying the following requirements:*

(H$_\mathbf{3.1}$) *For every $h \in (0, h_0)$, there holds*

$$P_h \widetilde{P}_h = id_{X_h}, \text{ and } Q_h \widetilde{Q}_h = id_{U_h}. \tag{5.27}$$

(H$_\mathbf{3.2}$) *There exist $s > 0$ and $C_4 > 0$ such that there holds, for every $h \in (0, h_0)$,*

$$\|(I - \widetilde{P}_h P_h)\psi\|_X \leq C_4 h^s \|A^*\psi\|_X, \tag{5.28}$$

$$\|(-\hat{A}^*)^\gamma (I - \widetilde{P}_h P_h)\psi\|_X \leq C_4 h^{s(1-\gamma)} \|A^*\psi\|_X, \tag{5.29}$$

for every $\psi \in D(A^*)$, and

$$\|(I - \widetilde{Q}_h Q_h)u\|_U \xrightarrow[h\to 0]{} 0, \tag{5.30}$$

for every $u \in U$, and

$$\|(I - \widetilde{Q}_h Q_h) B^* \psi\|_U \leq C_4 h^{s(1-\gamma)} \|A^* \psi\|_X \tag{5.31}$$

for every $\psi \in D(A^*)$.

Note that (5.29) makes sense since, by assumption, $\gamma < 1/2$, and thus, Im $\widetilde{P}_h \subset D((-\hat{A}^*)^{1/2}) \subset D((-\hat{A}^*)^{\gamma})$.

For every $h \in (0, h_0)$, the vector space X_h (resp. U_h) is endowed with the norm $\| \ \|_{X_h}$ (resp., $\| \ \|_{U_h}$) defined by

$$\|y_h\|_{X_h} = \|\widetilde{P}_h y_h\|_X, \tag{5.32}$$

for $y_h \in X_h$ (resp., $\|u_h\|_{U_h} = \|\widetilde{Q}_h u_h\|_U$, for $u_h \in U_h$). In these conditions, it is clear that

$$\|\widetilde{P}_h\|_{L(X_h,X)} = \|\widetilde{Q}_h\|_{L(U_h,U)} = 1, \tag{5.33}$$

for every $h \in (0, h_0)$. Moreover, it follows from (5.28), (5.29), (5.30), and from the Banach-Steinhaus Theorem, that there exists $C_5 > 0$ such that

$$\|P_h\|_{L(X,X_h)} \leq C_5, \text{ and } \|Q_h\|_{L(U,U_h)} \leq C_5, \tag{5.34}$$

and

$$\|(-\hat{A}^*)^{\gamma}(I - \widetilde{P}_h P_h)\psi\|_X \leq C_5 \|(-\hat{A}^*)^{\gamma} \psi\|_X, \tag{5.35}$$

for every $h \in (0, h_0)$, and every $\psi \in D((-\hat{A}^*)^{\gamma})$.

($\mathbf{H_{3.3}}$) *For every $h \in (0, h_0)$, there holds*

$$P_h = \widetilde{P}_h^*, \text{ and } Q_h = \widetilde{Q}_h^*, \tag{5.36}$$

where the adjoint operators are considered with respect to the pivot spaces X, U, X_h, and U_h.

Note that this assumption indeed holds for most of classical schemes (Galerkin or spectral approximations, centered finite differences, ...).

($\mathbf{H_{3.4}}$) *There exists $C_6 > 0$ such that*

$$\|B^* \widetilde{P}_h \psi_h\|_U \leq C_6 h^{-\gamma s} \|\psi_h\|_{X_h}, \tag{5.37}$$

for all $h \in (0, h_0)$ and $\psi_h \in X_h$.

For every $h \in (0, h_0)$, we define the approximation operators $A_h^* : X_h \to X_h$ of A^*, and $B_h^* : X_h \to U_h$ of B^*, by

$$A_h^* = P_h A^* \widetilde{P}_h, \text{ and } B_h^* = Q_h B^* \widetilde{P}_h. \tag{5.38}$$

Due to Assumption $(H_{3.3})$, it is clear that $B_h = P_h B \widetilde{Q}_h$, for every $h \in (0, h_0)$. On the other part, we set $A_h = (A_h^*)^*$ (with respect to the pivot space X_h). Note that, if A is selfadjoint, then $A_h = P_h A \widetilde{P}_h$.

($\mathbf{H_4}$) *The following properties hold:*

($\mathbf{H_{4.1}}$) *The family of operators* $\mathrm{e}^{tA_h^*}$ *is uniformly analytic, in the sense that there exists* $C_7 > 0$ *such that*

$$\|\mathrm{e}^{tA_h}\|_{L(X_h)} \leq C_7\mathrm{e}^{\omega t}, \text{ and } \|A_h\mathrm{e}^{tA_h}\|_{L(X_h)} \leq C_7\frac{\mathrm{e}^{\omega t}}{t}, \tag{5.39}$$

for every $t > 0$.

Under this assumption, there exists $C_8 > 0$ such that

$$\|R(\lambda, A_h)\|_{L(X_h)} \leq \frac{C_8}{|\lambda - \omega|}, \tag{5.40}$$

for every $\lambda \in \Delta_\delta$. Note that (5.39) and (5.40) hold as well if one replaces A_h with A_h^*.

($\mathbf{H_{4.2}}$) *There exists* $C_9 > 0$ *such that, for every* $f \in X$ *and every* $h \in (0, h_0)$, *the respective solutions of* $\hat{A}^*\psi = f$ *and* $\hat{A}_h^*\psi_h = P_hf$ *satisfy*

$$\|P_h\psi - \psi_h\|_{X_h} \leq C_9h^s\|f\|_X. \tag{5.41}$$

In other words, there holds $\|P_h\hat{A}^{*-1} - \hat{A}_h^{*-1}P_h\|_{L(X,X_h)} \leq C_9h^s$. This is a (strong) rate of convergence assumption.

Remark 5.6. Assumptions (H_3) and ($H_{4.2}$) hold for most of the classical numerical approximation schemes, such as Galerkin methods, spectral methods, centered finite difference schemes, ... As noted in [14], Assumption ($H_{4.1}$) of uniform analyticity is not standard, and has to be checked in each specific case. However, it can be shown to hold, under Assumption (H_1), provided the bilinear form associated with A_h is uniformly coercive (see [4] for the selfadjoint case, and [13, Lemma 4.2] for the general nonselfadjoint case).

Theorem 5.10. *Under the previous assumptions, the control system* $\dot{y} = Ay + Bu$ *is exactly null controllable in time* $T > 0$, *if and only if the family of discretized control systems* $\dot{y}_h = A_hy_h + B_hu_h$ *is uniformly controllable in the following sense. There exist* $\beta > 0$, $h_1 > 0$, *and positive real numbers* c, c', *such that the uniform observability and admissibility inequality*

$$c\|\mathrm{e}^{TA_h^*}\psi_h\|_{X_h}^2 \leq \int_0^T \|B_h^*\mathrm{e}^{tA_h^*}\psi_h\|_{U_h}^2 dt + h^\beta\|\psi_h\|_{X_h}^2 \leq c'\|\psi_h\|_{X_h}^2 \tag{5.42}$$

holds, for every $h \in (0, h_1)$ *and every* $\psi_h \in X_h$.

In these conditions, for every $y_0 \in X$*, and every* $h \in (0, h_1)$*, there exists a unique* $\psi_h \in X_h$ *minimizing the functional*

$$J_h(\psi_h) = \frac{1}{2} \int_0^T \|B_h^* \mathrm{e}^{tA_h^*} \psi_h\|_{U_h}^2 dt + \frac{1}{2} h^\beta \|\psi_h\|_{X_h}^2 + \langle \mathrm{e}^{TA_h^*} \psi_h, P_h y_0 \rangle_{X_h}, \quad (5.43)$$

and the sequence of controls $(\widetilde{Q}_h u_h)_{0<h<h_1}$*, where* u_h *is defined by*

$$u_h(t) = B_h^* \mathrm{e}^{(T-t)A_h^*} \psi_h, \quad (5.44)$$

for every $t \in [0, T]$*, converges weakly (up to a subsequence), in the space* $L^2(0, T; U)$*, to a control* u *such that the solution of*

$$\dot{y} = Ay + Bu, \ y(0) = y_0, \quad (5.45)$$

satisfies $y(T) = 0$*. For every* $h \in (0, h_1)$*, let* $y_h(\cdot)$ *denote the solution of*

$$\dot{y}_h = A_h y_h + B_h u_h, \ y_h(0) = P_h y_0. \quad (5.46)$$

Then,

- $y_h(T) = -h^\beta \psi_h$*;*
- *the sequence* $(\widetilde{P}_h y_h(\cdot))_{0<h<h_1}$ *converges weakly (up to a subsequence), in the space* $L^2(0, T; X)$*, to* $y(\cdot)$ *on* $[0, T]$*;*
- *for every* $t \in (0, T]$*, the sequence* $(\widetilde{P}_h y_h(t))_{0<h<h_1}$ *converges weakly (up to a subsequence), in the space* X*, to* $y(t)$*.*

Furthermore, there holds

$$\int_0^T \|u(t)\|_U^2 dt \leq \frac{1}{c} \|y_0\|_X^2, \quad (5.47)$$

and there exists $M > 0$ *such that, for every* $h \in (0, h_1)$*,*

$$\begin{aligned} &\int_0^T \|u_h(t)\|_{U_h}^2 dt \leq M^2 \|y_0\|_X^2, \quad h^\beta \|\psi_h\|_{X_h}^2 \leq M^2 \|y_0\|_X^2, \\ &\text{and} \ \ \|y_h(T)\|_{X_h} \leq M h^{\beta/2} \|y_0\|_X. \end{aligned} \quad (5.48)$$

Remark 5.7. The left-hand side of (5.42) is a uniform observability inequality for the control systems $\dot{y}_h = A_h y_h + B_h u_h$. The right-hand side of that inequality means that the control operators B_h are uniformly admissible.

Remark 5.8. A similar result holds if the control system $\dot{y} = Ay + Bu$ is exactly controllable in time $T > 0$. However, due to Assumption (H_1), the semigroup $S(t)$ enjoys in general *regularity properties*. Therefore, the solution $y(\cdot)$ of the control system may belong to a subspace of X, whatever the control u is.

For instance, in the case of the heat equation with a Dirichlet or Neumann boundary control, the solution is a smooth function of the state variable x, as soon as $t > 0$, for every control and initial condition $y_0 \in L^2$. Hence, exact controllability does not hold in this case in the space L^2 (for results on exact null controllability, see [8, 15]).

The theorem states that the controls u_h defined by (5.44) tend to a control u realizing the exact null controllability for (5.45). On may wonder under which assumptions the control u is the HUM control such that $y(T) = 0$ (see Remark 5.5). The following result provides an answer.

Proposition 5.2. *With the notations of Theorem 5.10, if the sequence of real numbers $\|\psi_h\|_{X_h}$, $0 < h < h_1$, is moreover bounded, then the control u is the unique HUM control such that $y(T) = 0$.*

A sufficient condition on $y_0 \in X$, ensuring the boundedness of the sequence $(\|\psi_h\|_{X_h})_{0<h<h_1}$, is the following: there exists $\eta > 0$ such that the control system $\dot{y} = Ay + Bu$ is exactly null controllable in time t, for every $t \in [T - \eta, T + \eta]$, and the trajectory $t \mapsto S(t)y_0$ in X, for $t \in [T - \eta, T + \eta]$, is not contained in a hyperplane of X.

An example where this situation indeed occurs is the following. Additionally to the previous assumptions, assume that the operator A admits a Hilbertian basis of eigenvectors e_k, associated with eigenvalues λ_k, for $k \in \mathbb{N}$, satisfying

$$\sum_{k=1}^{+\infty} \frac{-1}{\lambda_k} < +\infty. \tag{5.49}$$

Let $y_0 = \sum_{k\in\mathbb{N}} y_{0k} e_k$ a point of X such that $y_{0k} \neq 0$, for every $k \in \mathbb{N}$. Then, the assumption of Proposition 5.2 is satisfied. Indeed, if the trajectory $t \mapsto S(t)y_0$ in X, for $t \in [T - \eta, T + \eta]$, were contained in a hyperplane of X, there would exist $\Phi = \sum_{k\in\mathbb{N}} \Phi_k e_k \in X \setminus \{0\}$ so that

$$\sum_{k\in\mathbb{N}} \mathrm{e}^{\lambda_k t} y_{0k} \Phi_k = 0,$$

for every $t \in [T - \eta, T + \eta]$. It is well known that the condition (5.49) implies that the functions $\mathrm{e}^{\lambda_k t}$, $k \in \mathbb{N}$, are independent in L^2. Hence, $y_{0k}\Phi_k = 0$, for every $k \in \mathbb{N}$. This yields a contradiction.

Conclusion.

Under standard assumptions on the discretization process, for an exactly null controllable linear control system, if the semigroup of the approximating system is uniformly analytic, and if the degree of unboundedness of the control operator is lower than 1/2, then the semidiscrete approximation models are uniformly controllable.

The problem of providing rates of convergence for the controls of the semidiscrete models is an open problem.

The condition on the degree of unboundedness γ of the control operator B is very stringent, and an interesting open problem is to investigate whether the results of this article still hold whenever $\gamma \geq 1/2$. Note that, if $\gamma < 1/2$, then B is automatically admissible; this does not hold necessarily whenever $\gamma \geq 1/2$, and may cause some technical difficulties. However, there are many important and relevant problems for which $\gamma \geq 1/2$, that are not covered by the previous result, such as, for instance, the heat equation with Dirichlet boundary control. Note that, in this case, although Theorem 5.10 cannot be applied, the finite difference semidiscrete models are uniformly controllable in the one dimensional case (see [9]).

Another open and challenging question, probably much more difficult, is to remove the assumption of uniform analyticity of the discretized semigroup. In the case of the one dimensional wave equation, a result of uniform controllability was proved when using a mixed finite element discretization process (see [5]); the extension to higher dimensions is not clear (see [39]). However, a general result, stating uniform stabilization properties, was derived in [26] for general hyperbolic systems.

Finally, the question of uniform controllability of semidiscrete approximations of controlled partial differential equations is completely open in semilinear (more generally, nonlinear) case. It seems reasonable to investigate, in a first step, whether similar results hold in the case of globally Lipschitzian nonlinearities. Indeed, using fixed point arguments combined with the HUM method (see for instance [37]), it should be possible to reduce the study of the controllability to the linear case.

References

1. H. T. Banks and K. Ito, *Approximation in LQR problems for infinite dimensional systems with unbounded input operators*, J. Math. Systems Estim. Control, 7 (1997), no. 1.
2. H. T. Banks and K. Kunisch, *The linear regulator problem for parabolic systems*, SIAM J. Control Optim., 22 (1984), no. 5, pp. 684–698.
3. C. Bardos, G. Lebeau and J. Rauch, *Sharp sufficient conditions for the observation, control and stabilization of waves from the boundary*, SIAM J. Cont. Optim., 30 (1992), 1024–1065.
4. J. Bramble, A. Shatz, V. Thomee and L. Wahlbin, *Some convergence estimates for semidiscrete Galerkin type approximations for parabolic equations*, SIAM J. Num. Anal., 14 (1977), pp. 218–241.

5. C. Castro and S. Micu, *Boundary controllability of a linear semi-discrete 1-D wave equation derived from a mixed finite element method*, Preprint Univ. Madrid, 2005.
6. J. S. Gibson, *The Riccati integral equations for optimal control problems on Hilbert spaces*, SIAM J. Control Optim., 17 (1979), pp. 537–565.
7. L. F. Ho, *Observabilité frontière de l'équation des ondes*, C. R. Acad. Sci. Paris, 302 (1986), 443–446.
8. O. Yu. Imanuvilov, *Controllability of parabolic equations*, Sb. Math., 186, 6 (1995), pp. 879–900.
9. J. A. Infante and E. Zuazua, *Boundary observability for the space semi-discretizations of the* 1*-D wave equation*, M2AN Math. Model. Numer. Anal., 33 (1999), no. 2, pp. 407–438.
10. F. Kappel and D. Salamon, *An approximation theorem for the algebraic Riccati equation*, SIAM J. Control Optim., 28 (1990), no. 5, pp. 1136–1147.
11. V. Komornik, *Exact controllability and stabilization, the multiplier method*, Wiley, Masson, Paris, 1994.
12. S. Labbé and E. Trélat, *Uniform controllability of semidiscrete approximations of parabolic control systems*, Preprint Univ. Paris-Sud, Orsay (2005).
13. I. Lasiecka, *Convergence estimates for semidiscrete approximations of nonselfadjoint parabolic equations*, SIAM J. Num. Anal., 21, 5 (1977), pp. 894–908.
14. I. Lasiecka and R. Triggiani, *Control theory for partial differential equations: continuous and approximation theories. I. Abstract parabolic systems*, Encyclopedia of Mathematics and its Applications, 74, Cambridge University Press, Cambridge, 2000.
15. G. Lebeau and L. Robbiano, *Contrôle exact de l'équation de la chaleur*, Comm. Partial Differential Equations, 20 (1995), pp. 335–356.
16. L. Leon and E. Zuazua, *Boundary controllability of the finite-difference space semi-discretizations of the beam equation*, A tribute to J.-L. Lions, ESAIM Control Optim. Calc. Var., 8 (2002), pp. 827–862.
17. J.-L. Lions, *Exact controllability, stabilizability and perturbations for distributed systems*, SIAM Rev., 30 (1988), 1–68.
18. J.-L. Lions, *Contrôlabilité exacte, perturbations et stabilisation de systèmes distribués*, Tome 1, Recherches en Mathématiques Appliquées, 8, Masson, 1988.
19. J.-L. Lions and E. Magenes, *Problèmes aux limites non homogènes et applications*, Travaux et Recherches Mathématiques, No. 17, 18, 20, Dunod, 1968.
20. Z. Liu and S. Zheng, *Semigroups associated with dissipative systems*, Chapman & Hall/CRC, Research Notes in Mathematics, 398, 1999.
21. A. Lopez and E. Zuazua, *Some new results related to the null controllability of the 1-d heat equation*, Séminaire sur les Equations aux Dérivées Partielles, Ecole Polytechnique, VIII (1998), pp. 1–22.
22. S. Micu, *Uniform boundary controllability of a semi-discrete 1-D wave equation*, Numer. Math., 91 (2002), no. 4, pp. 723–768.
23. V. J. Mizel and T. I. Seidman, *Observation and prediction for the heat equation*, J. Math. Anal. Appl. 28 (1969), 303–312.
24. M. Negreanu and E. Zuazua, *Uniform boundary controllability of a discrete 1-D wave equation*, Systems Control Lett., 48 (2003), 3-4, pp. 261–279.
25. A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, Applied Mathematical Sciences, 44, Springer-Verlag, New York, 1983.

26. K. Ramdani, T. Takahashi and M. Tucsnak, *Uniformly exponentially stable approximations for a class of second order evolution equations : application to the optimal control of flexible structures*, Preprint Univ. Nancy, 2004.
27. R. Rebarber and G. Weiss, *Necessary conditions for exact controllability with a finite-dimensional input space*, Systems Control Lett., 40 (2000), no. 3, pp. 217–227.
28. D. L. Russell, *Controllability and stabilizability theory for linear partial differential equations: recent progress and open questions*, SIAM Rev., 20, 4 (1978), pp. 639–739.
29. T. I. Seidman, *Observation and prediction for the heat equation, III,* J. Differential Equations, 20 (1976), no. 1, 18–27.
30. L. R. Tcheugoué Tébou and E. Zuazua, *Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity*, Numer. Math., 95 (2003), no. 3, pp. 563–598.
31. M. Tucsnak and G. Weiss, *Simultaneous exact controllability and some applications*, SIAM J. Control Optim., 38 (2000), 5, pp. 1408–1427.
32. M. Tucsnak and G. Weiss, *How to get a conservative well-posed linear system out of thin air. I. Well-posedness and energy balance*, ESAIM Cont. Optim. Calc. Var., 9 (2003), pp. 247–274.
33. M. Tucsnak and G. Weiss, *How to get a conservative well-posed linear system out of thin air. II. Controllability and stability*, SIAM J. Control Optim., 42 (2003), 3, pp. 907–935.
34. G. Weiss, *Admissible observation operators for linear semigroups*, Israel J. Math., 65 (1989), 1, pp. 17–43.
35. G. Weiss, *Admissibility of unbounded control operators*, SIAM J. Control Optim., 27 (1989), 3, pp. 527–545.
36. E. Zuazua, *Boundary observability for the finite-difference space semi-discretizations of the 2-D wave equation in the square*, J. Math. Pures Appl., 9, 78 (1999), no. 5, pp. 523–563.
37. E. Zuazua, *Controllability of partial differential equations and its semi-discrete approximations*, Discrete Contin. Dyn. Syst., 8 (2002), no. 2, pp. 469–513.
38. E. Zuazua, *Optimal and approximate control of finite-difference approximation schemes for the 1-D wave equation*, Rendiconti di Matematica, SIAM Review, 47, 2 (2005), pp. 197-243.
39. E. Zuazua, *Propagation, Observation, Control and Numerical Approximation of Waves approximated by finite difference method*, to appear in SIAM Review.

6

Stability, Told by Its Developers

Antonio Loría and Elena Panteley

C.N.R.S., Laboratoire des Signaux et Systèmes, Supélec, 3, Rue Joliot Curie,
91192 Gif-sur-Yvette, France.
E-mail: loria@lss.supelec.fr, panteley@lss.supelec.fr

> *"The authors of the present manuscript would like to insist on the fact that only the attentive reading of the original documents can contribute to correct certain errors endlessly repeated by different authors."*
>
> J. J. Samueli & J. C. Boudenot[1]

6.1 Introduction

6.1.1 About the Chapter

Lyapunov stability theory is probably the most useful qualitative method to study the behaviour of dynamical systems; it benefits from at least 75 years of sustained development. It started with the memoir of A. M. Lyapunov [32], published in a Western[2] language in [33], and, starting with the 1930s, many refinements to this stability theory have been established.

The purpose of this chapter is to present basic definitions and theorems on stability, mostly on *Lyapunov* stability, through a concise and modest historical survey: it contains a short account of statements made by Lagrange, Lyapunov and other mathematicians: the *developers* of stability theory.

With these notes, we intend to bring some clarifications to important aspects of stability theory which, otherwise, have been somewhat obscured due

[1] Translated from *H. Poincaré (1854-1912), physicien*, Editions Ellipses: Paris, 2005. The citation is taken from the epilogue of the mentioned biography of *the last universalist* –as his biographers call H. Poincaré. The authors give interesting evidence of H. Poincaré's shared discovery – with Lorentz – of restrained relativity – *cf.* Comptes Rendus de l'Académie des Sciences, Paris 9th/June/1905.

[2] The qualifier "Western" is used here to refer to European languages other than Russian as well as to non-Soviet nationals.

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 199–258, 2006.

to not always accurate translations from Russian – on occasions, double – into English and inexact "recursive" citations, *i.e.* citations made by an author A of the work of an author B, based on a text written by a tertiary author C[3]. In contrast, unless explicitly mentioned, our citations are in most cases from *direct* sources and, to avoid further ambiguity, we take special care in citing the *exact* formulations of concepts introduced in early literature and translations are made with a maximum of fidelity, keeping at best the original words, the mathematical notations[4], the numbering of equations, *etc.* We have taken special care in the accuracy of references; in particular, titles are original – Russian titles are phonetically transcripted from Cyrillic characters and translated. When considered necessary, comments are made to explain certain statements in "modern" language. We also emphasise that the chapter is written as a (historical) survey and not as a tutorial, *i.e.* we assume that the reader is familiar with the basic Lyapunov theory (main theorems, basic definitions, invariance principles *etc.*). We hope with this brief historical account to revive otherwise seemingly forgotten fundamental literature on stability theory.

We sacrifice generality for detail of exposition; we focus only on a few definitions of stability for solutions of ordinary differential equations. This excludes, *e.g.*, Input-Output stability [67, 68] and Input-to-State Stability [60]. We shall not deal either with systems described by discontinuous dynamics, systems with delays, systems in discrete time, sampled-data systems *etc.* Throughout we implicitly assume existence and unicity of solutions.

The survey is organised by topics rather than chronologically: In Section 6.1.2 we discuss what stability is in general terms; in Section 6.2 we review Lagrange stability through the work of the mathematician Joseph Louis de la Grange; Section 6.3 is a review of Lyapunov stability; Section 6.4 presents asymptotic stability; in Section 6.5 we discuss global asymptotic stability and, more generally, asymptotic stability for large initial conditions; well-known and less-known invariance-principle-type theorems are presented in Section 6.6; Section 6.7 deals with the fundamental aspect of uniformity; Section 6.8 is a brief account of (a type of) robust stability; some bibliographical notes are provided in Section 6.9 before concluding with some remarks in Section 6.10.

Last but not least, we mention that the reading of this chapter would be incomplete without the independent and original material presented in Appendix A of this book. This material, which has been contributed by A. Teel and L. Zaccarian, has of course interest of its own.

[3] For certain recursive references more than three authors are involved.

[4] Including eventual typographical errors made by the cited author(s).

6.1.2 Stability, Generally Speaking

To understand stability, consider[5]

> *a solution of a differential equation representing a physical phenomenon or the evolution of some system* [. . .] *There always exist two sources of uncertainty in the initial conditions. Indeed, when one attempts to repeat a given experiment, the reproduction of the initial conditions is never entirely faithful: for instance, a satellite can only be placed in orbit from one point and with a velocity that depends on the variable circumstances related to the launching of the rockets* [. . .] *It is thus fundamental to be able to recognise the circumstances under which small variations in the initial conditions will only introduce small variations in what follows of the phenomenon.* •

The example described above is highly illustrative of the concept of stability; however, we shall not be satisfied with a rough exposition of stability theory. To speak of stability in formal terms, whether it concerns a satellite in orbit or any other object (with physical meaning or not), we need to introduce a strict mathematical framework which makes part of the theory of dynamical systems. Stability may be described as a property of the solutions of differential equations[6] by which, given a "*reference*" solution $x^*(t, t_0^*, x_0^*)$ of

$$\dot{x} = f(t, x)\,, \qquad x_0^* = x(t_0^*, t_0^*, x_0^*) \in \mathbb{R}^n\,, \quad t \in \mathbb{R}_{\geq t_0^*}\,, \quad t_0^* \in \mathbb{R}_{\geq 0}\,,$$

any other solution $x(t, t_0, x_0)$ starting close to $x^*(t, t_0^*, x_0^*)$ (*i.e.* such that $t_0^* \approx t_0$ and $x_0^* \approx x_0$), remains close to $x^*(t, t_0^*, x_0^*)$ for later times.

To some readers it may appear at this point that the property of continuity of solutions with respect to initial conditions, and therefore the sufficient conditions for it, may bring an answer to the question of stability, posed above. However, as explained by N. Rouche and J. Mawhin [55], the theorem on continuity of solutions with respect to initial conditions establishes sufficient conditions for a *perturbed* solution to remain "close" to an unperturbed solution over a *finite* interval of time. In the question of stability, which is of our interest, this is insufficient since one requires that "*small variations in the initial conditions [will] only introduce small variations in what follows of the phenomenon*" that is, from the initial time and for *ever* after.

In the previous paragraph, we emphasise the use of the mathematical term: "*perturbed* solution". Even though to some readers perturbation may appeal

[5] This citation is from the formative and enjoyable text [55] which is also published in English, see [56].

[6] As we mentioned earlier, in this document we only speak of differential equations but this does not mean that stability is reserved to solutions of *differential* equations.

rather to an external undesirable phenomenon that makes a system "misbehave" with respect to a desired performance, in classical stability theory of dynamical systems the term perturbation refers to the variation in the initial conditions; hence, we speak of perturbed initial conditions: $t_0 := t_0^* + \Delta_t$ and $x_0 := x_0^* + \Delta_x$.

Solutions of differential equations are commonly referred to as "trajectories"; it is important to understand the precise mathematical meaning of these objects. Following [14, p. 1], we say that

> "a point of the real, n-dimensional space shall be denoted by the coordinates $x_1, \dots, x_n$. [...] In addition to the n-dimensional x-space which is also called *phase space*, we shall refer to the $(n+1)$-dimensional space of the quantities $x_1, \dots, x_n$, t, which will be called *motion space*. [...]
> The notation $\mathsf{x} = \mathsf{x}(t)$ indicates that the components x_i of x are functions of t. If these functions are continuous, then the point $(\mathsf{x}(t), t)$ of the motion space moves along a segment of a curve as t runs from t_1 to t_2, [...]
> The projection of a motion upon the phase space is called the *phase curve*, or *trajectory*, of the motion. In this case the quantity t plays the role of a curve parameter. •

Having introduced the formal terminology we are ready to open our main subject of study: stability. The first to formally have studied the *stability* problem was the Italian-French[7] mathematician J.-L. Lagrange in the context of mechanics:

> (Cited and translated from [27, p. XII] – preface to the 1^{st} ed. [9])
> One shall not find any Figure in this Work. The methods that I hereby expose do not rely neither on constructions, nor on geometric or mechanic reasoning, but only on algebraic operations, subject to a regular and uniform process. Those who like Analysis will see with pleasure how Mechanics becomes a new branch of it and will acknowledge that I have so extended the field. •

Our survey starts with Lagrange's work.

[7] While mostly known as a *French* mathematician, according with Encyclopaedia Britanica, 15th ed., 1989's printing, Joseph Louis de la Grange was born Giuseppe Luigi Lagrangia on the 25th of January 1736 in Turin, Sardinia-Piemonte (now part of Italy). He lived and taught mathematics in Turin until 1766 when he moved to Berlin and, only in 1787, he moved to Paris, where he died. Lagrange had French lineage from the side of his father.

6.2 Lagrange's Stability

> *"Messieurs* DE LA PLACE, COUSIN, LE GENDRE *et moi, ayant rendu compte d'un Ouvrage intitulé :* Méchanique analitique, *par* M. DE LA GRANGE, *l'Académie a jugé cet Ouvrage digne de son approvation, et d'être imprimé sous son Privilège.*
>
> *Je certifie cet Extrait conforme aux registres de l'Académie. A Paris, ce 27 février 1788.*
>
> LE MARQUIS DE CONDORCET*"*

So finishes[8] the preface to the first edition of Lagrange's famous treatise on analytical mechanics which was probably the first to give a formal definition and make precise statements on stability of motion of dynamical systems, *i.e.* of systems of ordinary differential equations; more particularly, of the second order.

Section III of Part I of [9] entitled "General properties of the equilibrium of a bodies system" deals with the concepts of equilibrium and stability:

> (Cited and translated from [27, pp. 69–70]) In a system of bodies in equilibrium, the forces P, Q, R, ..., stemming from gravity, are, as one knows, proportional to the masses of the bodies and, consequently, constant; and the distances p, q, r, ... meet at the centre of Earth. One will thus have, in such case,
>
> $$\Pi = \mathrm{P}p + \mathrm{Q}q + \mathrm{R}r + \ldots;$$
>
> [...] If one now considers the same system in motion, and let u', u'', u''', ... be the velocities, and m', m'', m''', ... be the respective masses of the different bodies that constitute it [the system in motion], the so well-known principle of *conservation of living forces* [...] yield this equation:
>
> $$m'u'^2 + m''u''^2 + m'''u'''^2 + \ldots = \text{const.} - 2\Pi\,.$$
>
> ●

Recognising that Π corresponds to the expression of potential energy and recalling that the "living forces" or *vis viva* corresponds to the kinetic energy, we identify in Lagrange's text, the equation that expresses the principle of energy conservation.

[8] Sirs de la Place –now Laplace, Cousin, le Gendre –now Legendre, having presented a Work entitled "Analytical Mechanics" by Sir de la Grange –now Lagrange, the Academy has judged this Work worth its approval and being printed under its Privilege. I certify this essay according to the annals of the Academy. Paris, 27th February 1788.

In the following paragraph Lagrange makes an interesting citation that he attributes to Courtivron[9] and which, to some extent, already speaks of stability:

> (Cited and translated from [27, p. 70]) Hence, since in the state of equilibrium, the quantity Π is a minimum or a maximum, it follows that the quantity $m'u'^2 + m''u''^2 + m'''u'''^2 + \ldots$, which represents the living force of the whole system, will be at same time a minimum or a maximum; this leads to the following principle of Statics, that, *from all the configurations that the system takes successively, that in which it has the largest or the smallest living force, is that where it would be necessary to place it* [the system] *initially so that it stayed in equilibrium. (See the* Mémoires de l'Académie des Sciences de 1748 et 1749.*)* •

Lagrange continues his essay on the properties of the equilibrium by making his famous statement that the minimum of the potential energy of a mechanical system corresponds to a stable equilibrium point whereas the potential energy function has a maximum at a point corresponding to an unstable equilibrium:

> (Cited and translated from [27, p. 71]) [...] we will show now that if this function [Π] is a minimum, the equilibrium will have stability, that is to say, if the system being supposed initially at the state of equilibrium and then being, no matter how little, displaced from such state, it will tend itself to come back to that position while making infinitely small oscillations: on the contrary, in the case that the same function will be a maximum, the equilibrium will have no stability, and once perturbed, the system will be able to make oscillations that will not be very small, and that may make it to drift farther and farther from its initial state. •

That is how Lagrange talked about stability; in modern terminology stability, as originally defined by Lagrange, may be interpreted as follows.

Definition 6.1 (Lagrange's *original* stability). *Consider a mechanical system with state* $[q, \dot{q}]$*. We say that the point* $q = 0$ *is stable if for any (infinitely small)* $\delta > 0$ *and* $t_0 \geq 0$

$$|q(t_0)| \leq \delta \implies |q(t)| \to 0 \qquad \forall\, t \geq t_0\,.$$

[9] In [27], the cited text is accompanied by a footnote of J. Bertrand, editor of the 3rd edition of Lagrange's treatise, who comments that Lagrange had attributed in [9], the mentioned principle from statics to the the "*little-known geometrician Courtivron*" but that Lagrange had removed Courtivron's name from the second edition to substitute it with the date of publication.

Lagrange's stability states that "[the system] *will tend itself to come back to that* [equilibrium] *position*"; hence, it can be viewed as a notion of *convergence* rather than of stability. Yet, after works on mechanics that succeeded Lagrange's (such as the paper of Dirichlet [29] and references mentioned therein) another mathematical interpretation of Lagrange's statement is the following:

Definition 6.2 (Lagrange's "interpreted" stability). *Consider a mechanical system with state* $[q, \dot{q}]$*. We say that the point* $q = 0$ *is stable if for any (infinitely small)* $\delta > 0$ *and* $t_0 \geq 0$ *there exists* $\varepsilon > 0$ *such that*

$$|q(t_0)| \leq \delta \implies |q(t)| \leq \varepsilon \qquad \forall\, t \geq t_0 \,.$$

Lagrange claims that the minimum of the potential energy corresponds to a stable point. The proof of his statement is based on a series expansion of the function Π and, in the words of G. Lejeune-Dirichlet, "*makes use of the abusive assumption that high-order terms are negligible*". Also according to Dirichlet, Poisson seems to have been the first to point out this inaccuracy and tried to correct it by supposing that the terms of second order dominate largely over terms of higher order than two, in his[10] *Traité de Mécanique, p. 492.* It was thus G. Lejeune-Dirichlet who provided the first rigorous proof[11] of Lagrange's statement in the cited work and actually reformulated the property of stability, making it closer to the property from Definition 6.2.

> (Cited and translated from [29, p. 457])
> The function of coordinates depends only on the nature of forces and can be expressed by a defined number of independent variables λ, μ, ν, ..., in such a way that the equation of living forces will be written as
>
> $$\sum mv^2 = \varphi(\lambda, \mu, \nu, \ldots) + C$$
>
> [...] the condition that expresses that [...] the system is at an equilibrium position, coincides with that which expresses that for these same values [of the coordinates], the total derivative of φ is zero; hence, for each equilibrium position, the function will be a maximum or a minimum. If a maximum really takes place, then the equilibrium is stable, that is, if one displaces infinitely little the points [coordinates] of the system from their initial values, and we give to each a small initial velocity, in the whole course of the motion the displacements of the points of the system, with respect to their equilibrium position, will remain within certain limits [that are] defined and very small. •

[10] G. Lejeune-Dirichlet, contemporary of Poisson, omitted to write a complete reference for Poisson's work; according to [57], the complete reference is [53].

[11] After J. Bertrand, editor of [27], Dirichlet's proof was originally published in the *Journal de Crelle, Vol. 32* and the *Journal de Liouville*, 1st series, Vol. XII, p. 474.

Notice that Dirichlet speaks of *maximum* of the function $\varphi(\lambda, \mu, \nu, \ldots)$ corresponding to a *stable* equilibrium; this makes sense if we consider that in modern notation the potential energy corresponds to $-\varphi$ and the independent coordinates λ, μ, ν, $\ldots$ correspond to the generalised coordinates of a Lagrangian system (see *e.g.* [13]).

Another interesting characteristic of Dirichlet's stability is that he adds to his definition, with respect to that of Lagrange (*cf.* Defintion 6.2), the condition that the initial velocities be small in order to produce small displacements; in modern terms we would put it as follows.

Definition 6.3 (Dirichlet's stability). *Let $x :=$ col$[q, \dot{q}]$. We say that the point $q = 0$ is stable if for each (infinitely small) $\delta > 0$ and $t_0 \geq 0$ there exists an (infinitely small) $\varepsilon > 0$ such that*

$$|x(t_0)| \leq \delta \implies |q(t)| \leq \varepsilon \qquad \forall\, t \geq t_0\,.$$

Coming back to Dirichlet's statement on stability, we remark that the proof of the fact that the minimum of the potential energy corresponds to a stable equilibrium is quite interesting to us since it is close, in spirit, to what we currently know as *Lyapunov theory*:

> (Cited and translated from [29, p. 459])
> Other than the hypothesis already made, that the equilibrium position corresponds to the values $\lambda = 0$, $\mu = 0$, $\ldots$, we will also suppose that $\varphi(0, 0, 0, \ldots) = 0$; $[\ldots]$ hence,
>
> $$\sum mv^2 = \varphi(\lambda, \mu, \nu, \ldots) - \varphi(\lambda_0, \mu_0, \nu_0, \ldots) + \sum mv_0^2\,.$$
>
> Since by hypothesis, $\varphi(\lambda, \mu, \nu, \ldots)$, for $\lambda = 0$, $\mu = 0$, $\ldots$, is zero or maximum, we will be able to determine positive numbers l, m, n, $\ldots$, sufficiently small so that $\varphi(\lambda, \mu, \nu, \ldots)$ be always negative $[\ldots]$ where the absolute values of the variables be respectively constrained not to overpass the limits l, m, n, $\ldots$, $[\ldots]$ Let us suppose that, among all the negative values of the function $[\ldots]$, $-p$, except for the sign, is the smallest: then we can easily show that, if we take λ_0, μ_0, ν_0, $\ldots$ numerically smaller than l, m, n, $\ldots$, and at same time one satisfies the inequality
>
> $$-\varphi(\lambda_0, \mu_0, \nu_0, \ldots) + \sum mv_0^2 < p\,,$$
>
> each of the variables λ, μ, ν, $\ldots$ will remain during the complete duration of the motion below the limits l, m, n, $\ldots$. Indeed, if the contrary took place, since the initial values λ_0, μ_0, ν_0, $\ldots$ satisfy the condition that we have just mentioned, and in view of the continuity of the variables λ, μ, ν, $\ldots$, it would be necessary that at some instant

> one or more numerical values of λ, μ, ν, ... were equal to their respective limits l, m, n, ... , without having any other value overpassing its limit. At this instant, the absolute value of $\varphi(\lambda_0, \mu_0, \nu_0, \ldots)$ would be larger or at least equal to p. Consequently, the second member of the equation of living forces [*i.e.* the kinetic energy term] would be negative, in view of the inequality written above, and which corresponds to the initial state; which is not possible, $\sum mv^2$ being always positive.
>
> •

Dirichlet's proof can be explained in modern terms using the total energy function, in terms of generalised positions $q := \lambda,\ \mu,\ \nu,\ \ldots$ and velocities $\dot{q}$, *i.e.*

$$V(q, \dot{q}) := T(q, \dot{q}) + U(q)$$

where $T(q, \dot{q}) := \sum mv^2$ and $U(q) := -\varphi(\lambda, \mu, \nu, \ldots)$, *i.e.* in general v depends on the generalised velocities and positions and the potential energy is assumed to depend only on the positions. As Dirichlet points out, we can assume without loss of generality that $U(0) = 0$. Dirichlet then posses

$$p := \inf\{U(q) : |\lambda| = l,\ |\mu| = m,\ |\nu| = n \ldots\}\,.$$

Now, consider initial positions $q(t_0)$ and velocities $\dot{q}(t_0)$ such that $V(q(t_0), \dot{q}(t_0)) < p$, the equation of living forces (principle of energy conservation) is

$$V(q(t), \dot{q}(t)) = V(q(t_0), \dot{q}(t_0)) \qquad \forall\, t \geq t_0$$

so we have, necessarily, $V(q(t), \dot{q}(t)) < p$ for all $t \geq t_0$. Equivalently, $T(q(t), \dot{q}(t)) + U(q(t)) < p$ for all $t \geq t_0$. If any of the values λ, μ, ν, ... came to overpass its respective limit, say at $t = t^*$, we would have $U(q(t^*)) \geq p$ and, necessarily, $T(q(t^*), \dot{q}(t^*)) < 0$ which is impossible.

We see clearly that key concepts such as positive definiteness of certain function V as well as negative semi-definiteness of its derivative are implicit in Dirichlet's proof. Indeed, the key property used is the positivity of the kinetic energy T and notice that $V(q(t), \dot{q}(t)) = V(q(t_0), \dot{q}(t_0))$ is equivalent to $\dot{V}(q(t), \dot{q}(t)) = 0$, for the case that V is differentiable; however, $V(q(t), \dot{q}(t)) = V(q(t_0), \dot{q}(t_0))$ being the *integral* of the living forces equation, in Dirichlet's proof it is not required that the energy function be differentiable.

6.2.1 Modern Interpretations of Lagrange-Dirichlet Stability

We have seen above how Lagrange and Dirichlet defined stability in their own words. As we can see from the previous citations, earlier mathematicians stated definitions and theorems in a language that, nowadays, might be qualified as lacking of rigour. This certainly has led to different interpretations.

The terminology "*stability in the sense of Lagrange*" is attributed, by Hahn [14, p. 129], to La Salle [16]; in the latter one reads: "*the boundedness of all solutions for $t \geq 0$ is also a kind of stability, called* Lagrange stability".

In other terms, consider the system

$$\dot{x} = F(t, x) \tag{6.1}$$

where F is continuous, and $F(t, \cdot)$ is locally Lipschitz, uniformly in t and $F(t, 0) \equiv 0$.

Definition 6.4 (Lagrange stability). *The system (6.1) is said to be Lagrange stable if for each $\delta > 0$ and $t_\circ \geq 0$ there exists $\varepsilon > 0$ such that*

$$|x(t_\circ)| \leq \delta \implies |x(t)| \leq \varepsilon \qquad \forall\, t \geq t_\circ \geq 0\,.$$

Theorems on boundedness of solutions can be found for instance in [66, 3, 30], the following result is from [16].

> *Theorem 4—A Lagrange Stability Theorem*
>
> Let Ω be a bounded neighbourhood of the origin and let Ω^c be its complement (Ω^c is the set of all points outside Ω). Assume that $W(x)$ is a scalar function with continuous first partials in Ω^c and satisfying:
>
> 1) $W(x) > 0$ for all x in Ω^c,
>
> 2) $\dot{W}(x) \leq 0$ for all x in Ω^c,
>
> 3) $W(x) \to \infty$ as $\| x \| \to \infty$.
>
> Then each solution of (2) [$\dot{x} = X(x)$] is bounded for all $t \geq 0$. •

Another interpretation of Lagrange-Dirichlet stability is given in [57]: the authors define it similarly to Definition 6.3 but with an *inverted* order in the choice of δ and ε, *i.e.*[12]

> (Cited and translated from [57, p. 23]) We shall say that the origin of the q-space is *stable in the sense of Lagrange-Dirichlet* (*) if to every real positive number ε, we can associate a real positive number $\eta(\varepsilon)$ such that
>
> $$\|r(t_0)\| < \eta(\varepsilon)$$
>
> implies that for all $t \geq t_0$, we have
>
> $$\|q(t)\| < \varepsilon\,.$$
>
> •

[12] In the notation of [57] $r(t) :=$ col$[q(t),\, \dot{q}(t)]$.

The "(*)" in the cite above corresponds to a footnote made by the authors where they mention that "*we shall keep ourselves from confusing this definition with that usually known as stability in the sense of Lagrange*". Also, the authors of [57] remark that the stability in the sense defined above is at the basis of what is known as stability of part of coordinates, a theory largely developed by the Soviet mathematician V. V. Rumiantsev, in the 1950s, and others. We will not develop further on this topic.

6.3 Lyapunov's Stability

> "*J'ai seulement eu en vue d'exposer dans cet Ouvrage ce que je suis parvenu à faire en ce moment et ce qui, peut-être, pourra servir de point de départ pour d'autres recherches de même genre.*"
>
> A. M. Liapounoff, 1907

Such is the final sentence[13] of A. M. Lyapunov's preface to the French translation of his famous memoir on stability of dynamical systems described by ordinary differential equations (*cf.* [33]). It was in this work, or rather in the original Russian version of it –*cf.* [32], that Lyapunov set the basis of the stability theory mostly used (implicitly or explicitly) nowadays in the literature of automatic control. It is following up the work of Lagrange, Dirichlet, Poincaré and other mathematicians who contributed to the foundations of analytical mechanics and *celestial mechanics*, that Lyapunov seems to have come to the theory that we know. In the introduction of his memoir he states:

> (Cited and translated from [34, p. 209]) Let us consider a material [physical] system with k degrees of freedom. Let
>
> $$q_1,\ q_2, \dots q_k$$
>
> be k independent variables by which we agree to define its position. [...]
> Considering such variables as functions of time t, we will denote their first derivatives, with respect to t, by
>
> $$q_1',\ q_2', \dots q_k'\,.$$
>
> In each problem of dynamics, [...] these functions satisfy k second-order differential equations. ●

[13] "*I only had in mind to expose in this Work what I succeeded in doing at this moment and which, maybe, will serve as starting point for other studies of the same type.*"

Then, on the basis of a physical dynamic system, Lyapunov proceeds to introduce his notation for the study of stability of motion for ordinary differential equations:

> Let us assume that a particular solution is found to be
>
> $$q_1 = f_1(t),\ q_2 = f_2(t), \ldots q_k = f_k(t)\,,$$
>
> in which the quantities q_j are expressed by real functions of t, [...] To this particular solution corresponds a determined motion of our system. By comparing it [the motion] [...] to other motions of the system that are plausible under the same forces, we will call it *unperturbed motion*, and all the rest, with respect to which it is compared, will be referred to as *perturbed motions.* ●

As we will see below, Lyapunov is interested in studying the behaviour of any solution, or more generally a given *function* of any solution, with respect to a particular (function of a) "reference" solution. The latter constitutes the unperturbed motion and the former the perturbed motion, that is the word perturbation refers to a (small) change in the initial conditions:

> Denoting by t_0 an arbitrary time instant, let us denote the corresponding values of the quantities q_j, q_j', in an arbitrary motion, by q_{j0}, q_{j0}'.
>
> (Cited and translated from [34, p. 210])
> Let
>
> $$q_{1\,0} = f_1(t_0) + \varepsilon_1, \quad q_{2\,0} = f_2(t_0) + \varepsilon_2, \quad \ldots, \quad q_{k\,0} = f_k(t_0) + \varepsilon_k\,,$$
>
> $$q_{1\,0}' = f_1'(t_0) + \varepsilon_1, \quad q_{2\,0}' = f_2'(t_0) + \varepsilon_2, \quad \ldots, \quad q_{k\,0}' = f_k'(t_0) + \varepsilon_k'\,,$$
>
> where ε_j, ε_j' are real constants. [...] that we will call *perturbations*, will define a perturbed motion.
>
> [...] let Q_1, Q_2, ..., Q_n be given continuous and real functions of the quantities
>
> $$q_1,\ q_2, \ldots q_k, \qquad q_1',\ q_2', \ldots q_k'\,.$$
>
> For the unperturbed motion they will become known functions of t that we will denote respectively F_1, F_2, ... F_n. For a perturbed motion they will become functions of the quantities
>
> $$t,\ \varepsilon_1,\ \varepsilon_2,\ \ldots\ \varepsilon_k, \qquad \varepsilon_1',\ \varepsilon_2',\ \ldots\ \varepsilon_k'\,.$$
>
> When the ε_j, ε_j' are equal to zero, the quantities
>
> $$Q_1 - F_1\,, \quad Q_2 - F_2\,, \quad \ldots, \quad Q_n - F_n$$
>
> will be zero for each value of t. ●

Then, Lyapunov introduces *his* stability –*cf.* [34, p. 210]):

> "if without making the constants ε_j, ε'_j zero, we make them infinitely small, the question that arises is whether it is possible to assign to the quantities $Q_s - F_s$ infinitely small limits, such that these quantities never reach them in absolute value.
>
> The solution to this question, which will be the subject of our study, depends on the nature of the considered unperturbed motion as well as on the choice of the functions $Q_1, Q_2, \ldots, Q_n$ and on the time instant t_0. Hence, this choice being made, the answer to this question will characterise, to some extent, the unperturbed motion, and it is such that it will express the property that we will call *stability* [. . .]"

So speaks Lyapunov about stability; however, in contrast to Lagrange and Dirichlet, Lyapunov gives a formal definition of stability:

(Cited and translated from [34, pp. 210-211])
Let $L_1, L_2, \ldots, L_n$ be positive given numbers. If for all values of these numbers, no matter how small they are, one can choose positive numbers

$$E_1,\ E_2,\ \ldots\ E_k \qquad E'_1,\ E'_2,\ \ldots\ E'_k\,,$$

such that, the inequalities

$$|\varepsilon_j| < E_j, \qquad \left|\varepsilon'_j\right| < E'_j \qquad (j = 1, 2, \ldots\ k)$$

being satisfied, we have

$$|Q_1 - F_1| < L_1\,, \quad |Q_2 - F_2| < L_2\,, \quad \ldots, \quad |Q_n - F_n| < L_n\,,$$

for all values of t greater than t_0, the unperturbed *motion will be called stable* WITH RESPECT TO THE QUANTITIES *$Q_1, Q_2, \ldots, Q_n$; in the opposite case, it will be called unstable with respect to the same quantities.* •

Notice that Lyapunov's stability, as originally stated, is a broad concept; to put it in modern terminology, we introduce the following notation: let $f :=$ col$[f_1, \ldots f_k]$, $f' :=$ col$[f'_1, \ldots f'_k]$, $q :=$ col$[q_1, \ldots q_k]$, $q' :=$ col$[q'_1, \ldots q'_k]$. Such coordinates define motions in the space $\mathbb{R}_{\geq 0} \times \mathbb{R}^{2k}$ as follows. The unperturbed motion, defined by $2k$ independent coordinates and t, is denoted by $(t,\ f(t),\ f'(t))$ with $t \in \mathbb{R}$; the perturbed motion, of the same coordinates and starting at $t_0 \in \mathbb{R}$, is denoted by[14] $(t,\ q(t),\ q'(t))$ with $t \geq t_0$. The initial conditions of the perturbed motion are given by t_0, $q_i(t_0) = f_i(t_0) + \varepsilon_i$, $q'_i(t_0) = f'_i(t_0) + \varepsilon'_i$ with $i \leq k$.

[14] Strictly speaking, q and q' are functions of t and of the initial conditions t_0, $q_0 := q(t_0)$ and $q'_0 := q'(t_0)$. In other words, $q(t)$ and $q'(t)$ is a short-hand notation to denote the trajectories $q(t, t_0, q_0, q'_0)$ and $q'(t, t_0, q_0, q'_0)$.

Stability in the sense of Lyapunov is defined with respect to functions of the perturbed and unperturbed motions. Let $Q : \mathbb{R}^{2k} \to \mathbb{R}^n$ (where n is not necessarily equal to $2k$) be continuous functions of the coordinates q, q' and define the functions $F : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ as

$$F(t) := Q(f(t),\, f'(t)) \qquad \forall\, t \in \mathbb{R}\,.$$

Definition 6.5 (Lyapunov's original statement). *We shall say that the unperturbed motion* $(t,\, f(t),\, f'(t))$ *is Lyapunov stable with respect to Q, if for any (infinitely small)* $\epsilon > 0$ *and* $t_0 \in \mathbb{R}$ *there exists* $\delta > 0$ *such that*

$$|q(t_0) - f(t_0),\, q'(t_0) - f'(t_0)| \leq \delta \implies |Q(q(t),\, q'(t)) - F(t)| \leq \epsilon\,.$$

The following two particular cases, included in the definition above, are worth to be singled out:

- when the function $Q(q, q') = q$ (*i.e.* $n = k$) and the unperturbed motion is the origin of the phase space, *i.e.* $(t,\, f(t),\, f'(t)) = (t, 0, 0)$ then, observing that $Q(0, 0) = 0$, we have $F \equiv 0$ and therefore, we recover the property of stability of part of coordinates, *i.e. for each* $\epsilon > 0$ *and* $t_0 \in \mathbb{R}$ *there exists* $\delta > 0$ *such that*

  $$|q_0,\, q_0'| < \delta \implies |q(t)| < \epsilon \qquad \forall\, t \geq t_0\,,$$

  which is called (for $t_0 \in \mathbb{R}_{\geq 0}$) in [57] "stability in the sense of Lagrange-Dirichlet" –*cf.* Section 6.2.1;

- when Q corresponds to the "identity" operator, *i.e.* $Q(r, s) = (r, s)$ and the unperturbed motion is the origin of the phase space then we have $F \equiv 0$. In this case, Lyapunov's stability reduces to the following: "*for any* $\epsilon > 0$ *and* $t_0 \in \mathbb{R}$ *there exists* $\delta > 0$ *such that the inequalities*

  $$|q_0,\, q_0'| < \delta \implies |q(t, t_0, q_0, q_0'),\, q'(t, t_0, q_0, q_0')| < \epsilon\,.$$

That is, even though Lyapunov's stability is a property of closeness of solutions (like stability in the senses of Lagrange and Dirichlet), it is defined in terms of *functions of the solutions* rather than the solutions themselves. This is a fundamental contribution with respect to previous works on *Analytical Mechanics*. Furthermore, Lyapunov raises the question of stability beyond the realm of physical systems, by considering the stability of *motion* for general differential equations:

> (Cited and translated from [34, pp. 212])
> The solution to our question depends on the study of differential equations of the perturbed motion or, in other words, of the study of the differential equations satisfied by the functions

$$Q_1 - F_1 = x_1, \quad Q_2 - F_2 = x_2, \quad \ldots, \quad Q_n - F_n = x_n\,.$$

> [...] We will assume that the number n and the functions Q_s [are] such that the order of this system is n and that can be put in the form
>
> $$\text{(1)} \qquad \frac{dx_1}{dt} = \mathrm{X}_1\,, \quad \frac{dx_1}{dt} = \mathrm{X}_2\,, \quad \ldots, \quad \frac{dx_1}{dt} = \mathrm{X}_n\,.$$
>
> •

Thus, from the above formulations we recover the definition of Lyapunov stability that we are used to seeing in textbooks on nonlinear systems, such as [19, p. 98], [18, p. 98], [20, p. 112], [65, p. 136] and on ordinary differential equations, *e.g.* [56, p. 6]:

Definition 6.6 (Lyapunov stability). *The origin is a* stable equilibrium *of Equation (6.1) if, for each pair of numbers $\varepsilon > 0$ and $t_\circ \geq 0$, there exists $\delta = \delta(t_\circ, \varepsilon) > 0$ such that*

$$|x(t_\circ)| < \delta \quad \Longrightarrow \quad |x(t)| < \varepsilon \qquad \forall\ t \geq t_\circ \geq 0\,. \tag{6.2}$$

In some texts and articles, starting at least with [15], one also finds the following definition of stability:

Definition 6.7 (Lyapunov stability). *The origin is a* stable equilibrium *of Equation (6.1) if for each $t_\circ \geq 0$ there exists $\varphi \in \mathcal{K}$ such that*

$$|x(t, t_\circ, x_\circ)| \leq \varphi(|x_\circ|) \qquad \forall\ t \geq t_\circ \geq 0\,. \tag{6.3}$$

We recall, from [15], that $\varphi \in \mathcal{K}$ if it is "*defined, continuous, and strictly increasing on $0 \leq r \leq r_1$, resp. $0 \leq r < \infty$, and if it vanishes at $r = 0$: $\varphi(0) = 0$*". It is established in [15, p. 169] that the two definitions, 6.6 and 6.7, are equivalent. See also[15] [19, p. 136], [20, p. 150].

Lyapunov *versus* Lagrange Stability

Comparing Definition 6.6 to Definition 6.4 we can see that Lagrange stability and Lyapunov stability are *different*; specifically, neither one implies the other. That is, a system may have unbounded solutions but a Lyapunov stable equilibrium and Lyapunov instability does not necessarily imply that (perturbed) solutions are unbounded. This is illustrated by the following simple examples.

[15] Strictly speaking, in [19] and [20] the author considers the case when φ, hence δ, are independent of $t_\circ$.

Example 6.1. Consider the van der Pol oscillator:

$$\dot{x}_1 = x_2, \tag{6.4}$$

$$\dot{x}_2 = -x_1 + (1 - x_1^2)x_2\,; \tag{6.5}$$

its phase portrait is depicted in Figure 6.1 for two particular solutions: starting from $t_0 = 0$, $x_1(t_0) = 0.3$, $x_2(t_0) = 0$ and $t_0 = 0$, $x'_1(t_0) = 0$, $x'_2(t_0) = 4$. All solutions starting inside the ε-disc, except at the origin, tend to the *attractor* $\mathcal{A}$, *i.e.* for this ε, no matter how small the initial conditions are, the solutions do not remain in the ε neighbourhood of the origin. Hence, the van der Pol system is not Lyapunov stable at the origin. However, it is Lagrange stable in the sense of Definition 6.4 since the solutions are bounded.

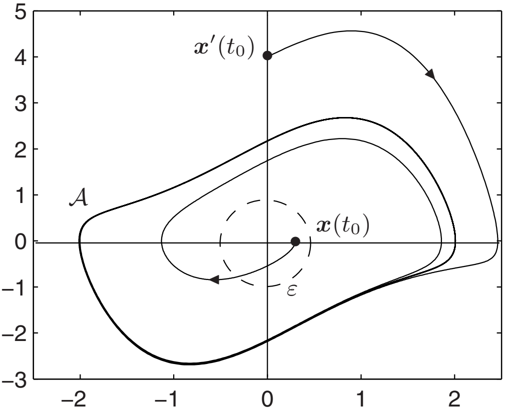

Fig. 6.1. Phase portrait of the van der Pol oscillator

Example 6.2. Consider the pendulum equation,

$$I\ddot{q} + mg\ell\,\sin(q) = 0$$

with mass m, length from its centre of mass to the axis of rotation ℓ and inertia I. The phase portrait of the pendulum (with $I = 1$, $m = 1$ and $l = 1$ is depicted in Figure 6.2. One can verify from this picture that the origin is stable; actually, with the modern tools available, it is easy to show that the origin is Lyapunov stable; as a matter of fact, so are all the equilibria $q = 2n\pi$, $\dot{q} = 0$ with $n \in \mathbb{Z}$. However, for initial velocities sufficiently large in absolute value, the trajectories $q(t)$ grow unboundedly in absolute value; hence, it is not Lagrange stable according to Definition 6.4. Yet it is important to remark that the equilibria $q = 2n\pi$, $\dot{q} = 0$ with $n \in \mathbb{Z}$ are also stable in the sense of Dirichlet –*cf.* Definition 6.3.

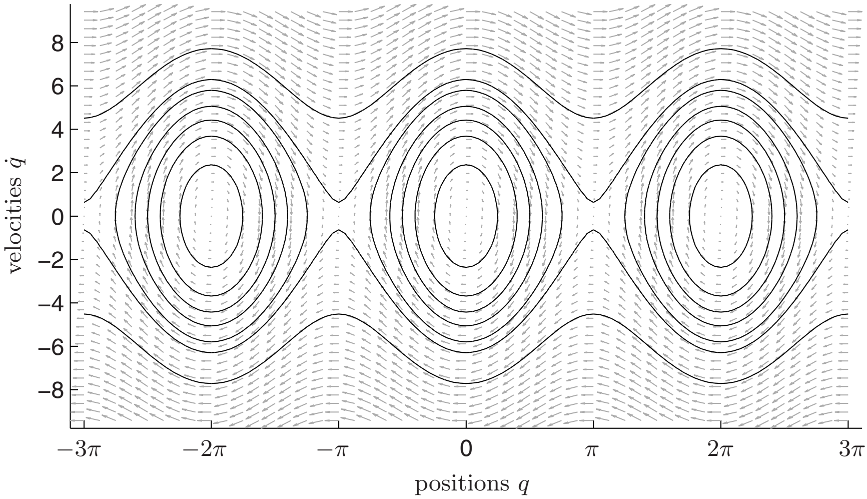

Fig. 6.2. Pendulum: Flowchart on the phase plane and energy-level curves

6.3.1 Lyapunov's Methods to Test for Stability

As is well known, there exist two methods of Lyapunov to investigate the stability of a system or, more precisely, of a solution of a differential equation:

> (Cited and translated from [34, p. 222]) All the processes that we can mention to solve the question that occupies us may be classified in two categories. In one, we shall fit all the processes that reduce to studying directly the perturbed motion and which, consequently, depend on the search for general or particular solutions of the differential equations under consideration.
>
> [. . .]
>
> The group of processes of study fitting in this category shall be called *the first method.* •

The latter corresponds to a method by which the solution of the differential equation is approximated by a series expansion. When we speak of the "first approximation" it means that the series is truncated at the first order. This *particular* version of Lyapunov's first method is what is often called Lyapunov's first method or method of linearization around an equilibrium point; see for instance [19, 20] for recent expositions of the Lyapunov's first method in the first approximation. See also [11] for a brief exposition of Lyapunov's (general) first method.

Considering the second method, which is of greater interest to us, Lyapunov continues by saying that:

> "*In the second one* [category] *we shall fit all sort of processes that are independent of the search for solutions of the differential equations of the perturbed motion.*
>
> *Such is the case, for instance, of the* [well-]*known process of analysis of stability of the equilibrium, in the case that there exists a function of forces.*" •

It seems very probable that Lyapunov refers here to the function of living forces, the principle of energy conservation and, of course, to Lagrange-Dirichlet's "theorem"; after all, Lyapunov's work was largely inspired by Dirichlet's proof –*cf.* citation on page 218. Lyapunov continues:

> (Cited and translated from [34, p. 222])
> These processes will reduce to the search of [...] functions of the variables x_1, x_2, ..., x_n, t, whose total derivatives with respect to t, under the hypothesis that x_1, x_2, ..., x_n are functions of t satisfying the equations (1), must satisfy such and such given conditions.
>
> The group of processes of this category will be called *the second method.* •

The "(1)" in the citation above refers to the second equation on page 213 of these notes.

Lyapunov Functions

Following [34, p. 256], let us consider systems of variables $x_1, x_2, \dots t$ on the following domain. Let T be "*as large as wanted*" and H be "*arbitrarily small but not zero*" [...] "*Thus, to solve the questions of stability, it shall suffice to consider [but] the values of t above some limit T, as large as wanted, replacing the initial values of x_s by their values corresponding to $t = T$*".

In other words, Lyapunov argues that one may consider the solution of a dynamical system starting with any instant T and redefine the initial states making them correspond to the values at that instant, therefore, we can consider, safely, the motion only after $t = T$. Under these conditions,

> (Cited and translated from [34, p. 256])
> we will consider real functions of real variables
>
> $$x_1,\, x_2,\, \dots,\, x_n,\, t, \tag{39}$$
>
> subject to the constraint

$$(40) \qquad t \geq T\,, \quad |x_s| \leq H \qquad (s = 1, 2, \ldots, n)\,.$$

> We will speak of functions that, on such domain, are continuous and uniform and that are zero if
>
> $$x_1 = x_2 = \ldots x_n = 0\,.$$
>
> •

Having set this notation, A. M. Lyapunov introduced what we know now as *Lyapunov functions*:

> (Cited and translated from [34, p. 257]) Let us suppose that the considered function V is such that, under the conditions (40), T being sufficiently large and H sufficiently small, it can only take values of a single sign.
>
> Then, we shall say that it is a *function of fixed sign*; and when it will be needed to indicate its sign, we shall say that it is a *positive function* or a *negative function.*
>
> If, moreover, the function V does not depend on t and if the constant H can be chosen sufficiently small so that, under the conditions (40), the equality $V = 0$ cannot occur unless we have
>
> $$x_1 = x_2 = \ldots x_n = 0\,,$$
>
> we shall call the function V, as if it were a quadratic function, *definite function* or, trying to attract attention on its sign, *positive definite* or *negative definite.*
>
> Concerning functions that depend on t, we shall still use these terms but then, we shall only speak of a *definite* function V under the condition that we can find a function W independent of t, that is positive definite and that in addition one of the expressions
>
> $$V - W \qquad -V - W$$
>
> be a positive function. •

It is in such terms that Lyapunov introduced what we call nowadays, positive definite and negative definite (Lyapunov) functions. Except for the fact that Lyapunov defined the properties of *his* functions, only locally, the definitions above are equivalent to those found in modern texts, as for instance [65, p. 148], [56, p. 7] for functions that depend on t and x; [20, p. 117] for functions that depend on x only. The formulation below, which follows closely that of [65], includes all mentioned cases and is borrowed *verbatim* from [17, pp. 40-41].

Definition 6.8. *A continuous function* $W : \mathbb{R}^n \to \mathbb{R}_+$ *is said to be* locally positive definite *if*

1. $W(0) = 0$*,*

2. $W(x) > 0 \quad \text{for small} \quad \|x\| \neq 0$*.*

A continuous function $W : \mathbb{R}^n \to \mathbb{R}$ *is said to be* globally positive definite *(or simply* positive definite*) if*

1. $W(0) = 0$*,*

2. $W(x) > 0 \quad \forall\, x \neq 0\,.$

For a continuous function $V : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}_+$, *i.e.* which also depends on time, we say that $V(t, x)$ is (resp. locally) positive definite if:

1. $V(t, 0) = 0 \;\; \forall\, t \geq 0\,;$
2. $V(t, x) \geq W(x), \quad \forall\, t \geq 0, \quad \forall\, x \in \mathbb{R}^n$ (resp. for small $\|x\|$)

where $W(x)$ is a (resp. locally) positive definite function.

A continuous function V is negative definite if $-V$ is positive definite.

Lyapunov's Second Method

Lyapunov's memoir contains two now well-known methods: the first, relying on linearization about the equilibrium point and the second, based on what Soviet mathematicians would call later, *Lyapunov* functions. As we have seen the terminology "second method" and "first method" was chosen by Lyapunov himself. He also introduced the method of characteristic exponents which was independently proposed by Poincaré in[16] [52]. In the present survey, we will focus on the second method since, not only it is more useful and general, but it was extended to *global* notions of stability by the mathematicians that succeeded Lyapunov during all of the last century.

> (Cited and translated from [34, pp. 258-259]) Everybody knows Lagrange's theorem on the stability of the equilibrium in the case when there exists a function of the forces, as well as the elegant demonstration given by Lejeune-Dirichlet. The latter relies on considerations that may serve the demonstration of many other analogous theorems.
>
> Guided by these considerations we will establish here the following propositions:

[16] By the way, in that paper H. Poincaré also discovered the theory of chaos.

> THEOREM I. – *If the differential equations of the perturbed motion are such that it is possible to find a definite function* V*, whose derivative* V′ *is a function of fixed sign and opposite to that of* V*, or it is exactly zero, the unperturbed motion is stable.* •

We do not present here Lyapunov's complete proof but we find it interesting to cite the first sentence:

> Let us suppose, to fix the ideas, that the function found V is positive definite and that its derivative V' is negative or identically zero. •

The fact that Lyapunov starts his proof by supposing, without loss of generality, that V is positive definite and its derivative is negative definite has set the convention that is used until now. However, it should be clear that one can also use *negative* Lyapunov functions V as long as the derivative V' is *positive* definite. Therefore, the following is a direct *corollary* of *Lyapunov's* theorem.

> (Cited from [15, p. 102]) Consider the differential equation[17]
>
> $$\dot{\boldsymbol{x}} = f(\boldsymbol{x})\,, \qquad 0 \leq |\boldsymbol{x}| \leq h\,, \quad f \in E \tag{25.1}$$
>
> Theorem 25.1. If there exists a positive definite function $v(\boldsymbol{x})$ whose derivative $\dot{v}(\boldsymbol{x})$ for (25.1) is negative semi-definite or identically zero then the equilibrium of (25.1) is stable. •

Other equivalent statements establishing stability are [65, Theorem 1, Section 5.3.1, p. 158], [54, Theorem 4.2, p. 13], [56, Theorem 4.6, p. 12], [18, Theorem 3.1, p. 101], [19, Theorem 3.1, p. 100], [20, Theorem 4.1, p. 114].

6.4 Asymptotic Stability

Consider again N. Rouche and J. Mawhin's example – *cf.* citation on p. 201, of a satellite put in orbit and for which, as we know, it is practically impossible to repeat *exactly* the same conditions every time. If one wishes to know the circumstances under which small initial errors in the satellite configuration, with respect to a point on its desired orbit, will lead only to small variations, stability in the sense of Lyapunov may be a good criterion to establish such conditions – provided, of course, that we have the right differential equations to model the satellite's motion. In many cases (practical or not), it is required that the small errors vanish as $t \to \infty$, *i.e.* that the perturbed motion approaches the unperturbed one, asymptotically. For instance, in the case of

[17] In Hahn's notation, h is a positive real and E is the set of functions f generating, via $\dot{x} = f(x)$, unique solutions and such that 0 is an isolated equilibrium point.

the satellite, one may want that the latter approaches the ideal orbit as time passes as opposed to only remaining "close" to it.

Lyapunov introduced the property of asymptotic stability in a remark following the proof of his theorem on stability, –*cf.* [34, Theorem I], in the following terms:

> (Cited and translated from [34, p. 261])
> *Remark II.* – If the function V, while satisfying the conditions of the theorem [Theorem I], allows an infinitely small upper bound, and if its derivative represents a definite function, one can show that every perturbed motion, sufficiently close to the unperturbed motion, will approach the latter asymptotically. •

The terminology "admits an infinitely small upper bound" was common in Soviet literature at least until the 1950s; at least since [14] this quality of certain functions is referred to in the literature as "decrescent". Notice also that Lyapunov only says that the derivative of V should be *definite*; however, according to [34, Theorem I] and the way Lyapunov introduced *his* functions, it is understood that he means definite and of opposite sign to that of V.

The definition of asymptotic stability became more precise in the Soviet literature that succeeded Lyapunov. For instance, N. N. Krasovskiĭ in [22, p. 2] says, just before presenting the definitions of stability and asymptotic stability, that "*some of the definitions of refined types of stability follow Četaev's annotations in*" [8, pp. 11-36]. The definition provided in [22] is as follows:

> (Cited from [22, p. 3]) Definition 1.2. The null solution $x = 0$ of the system (1.3)
>
> $$\Big[\quad (1.3) \qquad \frac{dx_i}{dt} = X_i(x_1, \cdots, x_n, t) \qquad (i = 1, \cdots, n) \qquad \Big]$$
>
> is called asymptotically stable and the region G_δ of x-space is said to lie in the region of attraction of the point $x = 0$ (at $t = t_0$), provided that the conditions of definition 1.1 [*cf.*, Definition 6.6] are satisfied, and provided further that
>
> $$\lim_{t\to\infty} x(t, x_0, t_0, t) = 0\,,$$
> $$x(x_0, t_0, t) \in \Gamma, \quad t \ge t_0\,,$$
>
> for all values of the initial point x_0 that lie in G_δ. Here Γ is some subregion of G which is given in advance, and with which the (mathematical model of the) physical problem is intrinsically concerned. •

The property that $|x_\circ| \le \delta$ implies that

$$\lim_{t\to\infty} x(t, t_\circ, x_\circ) = 0$$

was sometimes called *quasi-asymptotic stability* (*cf.* [2, p. 142], *cf.*, [14, p. 7][18]) or *quasi-equi-asymptotic stability –cf.* [66, p. 44] and it may be expressed by the more precise statement: $|x_\circ| < \delta$ *implies that for each* $\eta > 0$ *there exists* $T(\eta) > 0$ *such that*

$$|x(t, t_\circ, x_\circ)| < \eta \quad \forall t > t_\circ + T\,. \tag{6.6}$$

In general the number T depends on $x_\circ$ and on $t_\circ$; not only on η.

This brings us to the following well-adopted definition of asymptotic stability: *the equilibrium is asymptotically stable if it is stable and attractive* (*cf.* [65, Definition 31, p. 141], [56, 55, Definition 2.11, p. 6], [20, Definition 4.1, p. 112]). More precisely, we have:

Definition 6.9 (Asymptotic stability). *We say that the origin of (6.1) is asymptotically stable if it is stable in the sense of Definition 6.6 and there exists* $\delta > 0$ *such that, for each* $\eta > 0$*,* $t_\circ \geq 0$ *there exists* $T(\eta, t_\circ) > 0$*, such that*

$$|x_\circ| < \delta \qquad \Longrightarrow \qquad |x(t, t_\circ, x_\circ)| < \eta \quad \forall t > t_\circ + T\,. \tag{6.7}$$

The notation (6.6), to express the limit condition, was particularly useful to introduce other notions of "stability" such as equi-asymptotic stability, quasi-equi-asymptotic stability and uniform forms of the latter; in the following sections we will discuss uniform asymptotic stability, for the others see [66, 2, 14]. Another form of expressing the condition of quasi-asymptotic stability or *attractivity*, as it was called by W. Hahn in [15], is the following:

Definition 6.10. [19] *The equilibrium of the differential equation*

$$\dot{x} = f(t, x) \qquad t \geq t_\circ$$

is said to be attractive if there exists $\eta > 0$ *and, for each* $x_\circ$ *satisfying* $|x_\circ| < \eta$*, a function* σ *of class* $\mathcal{L}$ *such that*

$$|x(t, t_\circ, x_\circ)| < \sigma(t - t_\circ), \qquad \forall\, t > t_\circ\,.$$

We recall from [15, p. 7] that "*a real function* $\sigma(s)$ belongs to class $\mathcal{L}$ *if it is defined, continuous and strictly decreasing on* $0 \leq s_1 \leq s < \infty$ *and if* $\lim \sigma(s) = 0$ *(*$s \to \infty$*)*". Similarly as in the case of Definition 6.7, the function σ depends in general on $t_\circ$.

[18] Hahn uses $p(t, t_\circ, x_\circ)$ to denote the solutions.

[19] Except for slight changes in the notation, this definition is taken from [15, p. 7, Def. 2.8]

6.5 Globalisation of Asymptotic Stability

Interestingly, *globalisation* of asymptotic stability started in the Union of Soviet *Socialist* Republics (USSR)[20].

Roughly, *global* asymptotic stability means that the asymptotic stability holds for all initial conditions; yet, one may be unsatisfied with such definition and ask what does it *exactly, mathematically,* mean ? Surprisingly as it may appear, different answers have been given to this "innocuous" question since the early 1950s, both by Soviet and Western authors. In early literature (1950s) the terms *asymptotic stability in the whole* and *asymptotic stability in the large* were introduced in Soviet literature to distinguish the case when asymptotic stability holds not only for *infinitely small initial perturbations* (*i.e.* conditions) as originally defined by A. M. Lyapunov. In other terms, as J. P. La Salle puts it in [16]:

> (Cited from [16, pp. 521–522]) [. . .] it is never completely satisfactory to only know that the system is asymptotically stable without some idea of the size of the region of asymptotic stability [. . .] Ideally, we might like to have that the system return to equilibrium regardless of the size of the [initial] perturbation. •

In Western literature, starting probably with Hahn's *Stability of Motion*,

> (cited from [15, p. 109]) [. . .] if the domain of attraction is all of $\mathbb{R}^n$ we speak of *asymptotic stability in the whole*, (*cf.* sec 2) or also of *global asymptotic stability* [. . .] •

and in most of modern literature, as *e.g.* in [65, 59, 64, 58, 20, 19, 18], we use the qualifier "global" in *global asymptotic stability*, to refer to the case when asymptotic stability holds for all initial states in the state-space, *i.e.*, in $\mathbb{R}^n$. More precisely, we have:

Definition 6.11 (Global asymptotic stability). *We say that the origin of (6.1) is globally asymptotically stable if it is stable in the sense of Definition 6.6 and globally attractive, i.e. for each $\delta > 0$, $\eta > 0$ and $t_\circ \geq 0$ there exist $T(\eta, t_\circ) > 0$, such that*

$$|x_\circ| < \delta \qquad \Longrightarrow \qquad |x(t, t_\circ, x_\circ)| < \eta \quad \forall t > t_\circ + T \tag{6.8}$$

holds.

[20] Readers shall understand, however, that the meaning of the term *globalisation* in this chapter differs from the one given in the socio-economical context of the present times.

Unfortunately, through the years different properties have been defined using similar terminology as well as similar properties have been called different names. The main purpose of the following paragraphs is to make clear differences of those properties by recalling the original formulations. See also Appendix A of this textbook.

6.5.1 Global, *i.e.* in the *Large* or in the *Whole* ?

While there is a consensus on the meaning of *global* asymptotic stability there is considerable confusion that has carried on from early literature to modern texts, regarding the qualifiers "in the whole" and "in the large"; in Russian "*v tzelom*" and "*v bolshom*" respectively. On occasions, the notions of *asymptotic stability in the whole* and *asymptotic stability in the large* are *mistakingly* taken as synonyms however,

> (Cited from [22, p. 6])
> the terms *in the large* and *globally* are not synonymous. •

The clarification of the difference between asymptotic stability in the *whole* and in the *large* goes beyond pure semantical interest; mathematically, the two properties are different. As N. N. Krasovskiĭ puts it:[21]

> (Cited and translated from [21, p. 149]) When addressing questions of stability in the large[1] the interest [resides on] the estimate of the domain of stability (in the case when there is no stability in the whole). •

W. Hahn, in [14], explains the difference between asymptotic stability *in the large* and asymptotic stability *in the whole* and warns us against the mistaken translations:

> (Cited from [14, p. 8]) If relation (2.10) [here, (6.6)] is valid for *all* points x_0 from which motions originate, we shall say that the equilibrium is *asymptotically stable in the large* (Aizerman [1], Krasovskiĭ [21]). If relation (2.10) [here, (6.6)] holds for all points of the phase space, the equilibrium is said to be *asymptotically stable in the whole* (Barbashin and Krasovskiĭ [6, 7]). La Salle [16] proposed "complete stability." The distinction between asymptotic stability in the large and asymptotic stability in the whole has often been obliterated by inaccurate translations of the Russian terminology. However, it becomes important in cases where Eq. (2.7) [$\dot{x} = f(t, x)$] is not defined for all points of the phase space. •

[21] The upper index [1] in Krasovskiĭ's citation corresponds to the book [1], which we have not been able to locate.

To remove all ambiguity, we recall that Hahn in [14, p. 2] clearly defines the phase space as the *real $n-$dimensional space*. Hence, according to Hahn, asymptotic stability in the large is important when the differential equation $\dot{x} = f(t,x)$ is defined only in a region of the phase-space, *i.e.* for $x \in \mathcal{R} \subset \mathbb{R}^n$. Such a case may appear from mathematical considerations –*e.g.* consider the system

$$\dot{x} = -\frac{x}{1-x} \qquad \mathcal{R} := (-\infty, 1)\,,$$

or from physical assumptions, *i.e.* when $\dot{x} = f(t,x)$ corresponds to the model of a *physical* system such that the physical quantities x make sense only in a region $\mathcal{R}$. Nevertheless, one should not be misled by Hahn's words and haste to conclude that asymptotic stability in the large makes sense *only* when the system is not defined in the whole phase space but on an open "large" region of $\mathbb{R}^n$. This is discussed in greater detail in the fore-coming sections.

The work of La Salle that succeeded that of Soviet mathematicians did not contribute in clarifying the confusion between stability in the large and stability in the whole; in [16] La Salle introduced *complete stability*:

> (Cited from [16, p. 524]) For many systems it may be important to assure that no matter how large the perturbation, or in a feedback control system, regardless of the size of the error, the system tends to return to its equilibrium state. This is asymptotic stability in the large. In place of this awkward expression we shall say completely stable. The system (2) $[\ \dot{x} = X(x)\ ]$ will be said to be completely stable if the origin is stable and if every solution tends to the origin as t tends to infinity. •

In La Salle's paragraph, perturbation refers to the initial perturbation or, in other terms, to the initial conditions. Clearly, the above-given meaning of asymptotic stability in the large is in contradiction with contemporary Soviet literature[22].

6.5.2 Asymptotic Stability in the Large

Besides rough statements such as those cited above we have not been able to locate, in non-Russian literature, a precise definition of *asymptotic stability in the large*. Furthermore, the term "*v bolshom*" is actually scarcely used in Soviet publications; [12, Section "Stability in the large, in the whole"] is a rare passage dealing with both concepts in certain rigour:

> (Cited and translated from [12, p. 29])
> Definition 6.1. Let Δ_0 be a given positive number. The unperturbed motion Σ is called *asymptotically stable in the large*, if it is

[22] Some of which, well known by the authors of [25]; see *e.g.* Part Six on [26, pp. 117–153] authored by J. P. La Salle and S. Lefschetz.

> stable *a la* Lyapunov and condition (2.5) [$x(x_0, t) \to 0$ as $t \to \infty$] is satisfied for any initial perturbations x_0 from the region
>
> $$|x_0| \leq \mathit{\Delta}_0 \, .$$
>
> •

The previous definition must be distinguished from the definition of (local) asymptotic stability as originally defined by Lyapunov. Note that the latter speaks of the property by which all perturbed motions originating *arbitrarily* close to the unperturbed motion, tend to the latter as time goes to infinity. Hence, for the case of asymptotic stability of the origin, (local) asymptotic stability concerns the case when motions originate in an *infinitely small* open neighbourhood of the trivial solution, rather than in a *given* ball of radius $\mathit{\Delta}_0$. For comparison we cite next, Furasov's definition of (local) asymptotic stability:

> (Cited and translated from [12, p. 13])
> Definition 2.2. The unperturbed motion $\mathit{\Sigma}$ is called *asymptotically stable*, if it is stable *a la* Lyapunov and there exists a positive constant $\mathit{\Delta} \leq \delta(\varepsilon, t_0)$, such that the condition
>
> $$x(x_0, t) \to 0 \quad \text{as} \quad t \to \infty$$
>
> holds for all the solutions of the system, starting in the region
>
> $$\|x_0\| < \mathit{\Delta} \, .$$
>
> •

The constant $\delta(\varepsilon, t_0)$ in the previous definition is the same as in Definition 6.6.

Theorems on Asymptotic Stability in the Large

As Hahn remarked, asymptotic stability in the large was mainly studied by a few Soviet authors including M. A. Aizerman and N. N. Krasovskiĭ. In particular, we find it worth mentioning the reference [21] where the author gives conditions for asymptotic stability in the large, for systems of two equations:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2) \\ \dot{x}_2 &= f_2(x_1, x_2) \, , \end{aligned}$$

in function of the roots of the characteristic equation

$$\begin{vmatrix} \dfrac{\partial f_1}{\partial x_1} - \lambda & \dfrac{\partial f_1}{\partial x_2} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} - \lambda \end{vmatrix} = 0$$

where here, $|\cdot|$ denotes the determinant. Perhaps the best known theorems on asymptotic stability in the large –even though not called such– are the following, formulated by La Salle and repeated in other texts, not always in equivalent forms:

> (Cited from [25, pp. 58-59, Theorems VI and VII] together)
> Let $V(x)$ be a scalar function with continuous first partial derivatives. Let Ω_l designate the region where $V(x) < l$. Assume that Ω_l is bounded and that within Ω_l:
>
> $$\begin{aligned} &V(x) > 0 \qquad \text{for} \qquad x \neq 0\,, && \text{(a)}\\ &\dot{V}(x) < 0 \qquad \text{for all} \qquad x \neq 0 \text{ in } \Omega_l\,, && \text{(b)}^* \end{aligned}$$
>
> then the origin is asymptotically stable, and above all, every solution in Ω_l tends to the origin as $t \to \infty$ (The last conclusion goes beyond Lyapunov's asymptotic stability theorem). •

The comment in brackets is significant: the authors emphasise that [25, Theorem VII] establishes asymptotic stability in a much larger region than "a neighbourhood" of the origin, as originally stated by Lyapunov. Therefore, this theorem comes to determine what La Salle called *the extent of asymptotic stability –cf.* [16, 25].

6.5.3 Asymptotic Stability in the Whole

Concerning asymptotic stability in the whole we recall the original definition of E. A. Barbashin and N. N. Krasovskiĭ, as given in their milestone paper [6]:

> (Cited and translated from [6]) We say, that the trivial solution $x_i = 0$ of systems (1)
>
> $$\Big[\quad \frac{dx}{dt} = \mathrm{X}_i(x_1, x_2, \ldots, x_n), \qquad i = 1, 2, \ldots, n, \qquad (1) \quad \Big]$$
>
> is asymptotically stable for any initial perturbations if it is stable in the sense of Lyapunov (for sufficiently small perturbations) and if each other solution $x_i(t)$ of systems (1) possesses the property $\lim_{t\to\infty} x_i(t) = 0$, $i = 1, 2, \ldots, n$. •

That is, recalling that the "initial perturbations" correspond to the initial states away from zero, we see that E. A. Barbashin and N. N. Krasovskiĭ defined stability with respect to *arbitrary initial conditions* as local stability plus attractivity of the origin for *all* other solutions, *i.e.*, for all initial conditions. Even though the authors do not use, in the definition itself, the terminology *asymptotic stability in the whole* the previous definition can be adopted as such.

For the sake of comparison with the definition given by V. D. Furasov, for asymptotic stability in the large, we recall the following from [12, Section “Stability in the large, in the whole”]:

> (Cited and translated from [12, p. 30])
> Definition 6.2. The unperturbed motion Σ is called *asymptotically stable in the whole* if this motion is stable *a la* Lyapunov and the condition (2.5) [$x(x_0, t) \to 0$ as $t \to \infty$] is satisfied for any initial perturbations x_0 no matter how large they would be. •

The previous definition corresponds, in the same terms, to [4, Definition 12.1].

Thus, asymptotic stability in the whole is the same property known in more recent literature – starting probably with W. Hahn – as *global* asymptotic stability:

> (Cited from [14, p. 8]) [...] the set of all points (x_0, t_0) from which motions originate, satisfying the relation (2.10) [here, (6.6)], forms the *domain of attraction* of the equilibrium. •

> (Cited from [15, p. 109]) If the domain of attraction is all of $\mathbb{R}^n$ we speak of *asymptotic stability in the whole* (*cf.* Sec. 2) or also of *global asymptotic stability.* •

For completeness, let us recall the definition of[23] [64, p. 136]:

> (Cited from [64, pp. 135-136]) The equilibrium point $\mathbf{0}$ at time t_0 is asymptotically stable at time t_0 if (1) it is stable at time t_0, and (2) there exists a number $\delta_1(t_0) > 0$ such that
>
> $$|x(t_0)| < \delta_1(t_0) \quad \Longrightarrow \quad |x(t)| \to 0 \qquad \text{as} \;\; t \to \infty$$
>
> [...] The equilibrium point $\mathbf{0}$ at time t_0 is globally asymptotically stable if $x(t) \to 0$ as $t \to \infty$ (regardless of what $x(t_0)$ is). •

The readers will note that this definition is compatible with Definition 6.11 given above.

Notice also that, even though the original definition of [6] refers to autonomous systems, the same property can be enunciated for time-varying systems.

[23] See also [65] however, it is interesting to note that actually global asymptotic stability does *not* appear in [65] but only global *uniform* asymptotic stability for the more general case of non-autonomous systems.

Theorems on Asymptotic Stability in the Whole

Few texts present theorems for global asymptotic stability of *autonomous* systems; for instance, [14, 15, 54, 65] deal only with the more general case of non-autonomous systems. The following well-known theorem for global asymptotic stability, which is taught in (probably) most of elementary courses on nonlinear control and stability theory, is cited from [20] –see also [18, Theorem 3.2, p. 112] and [19, Theorem 3.2, p. 110]:

> (Cited from [20, p. 124])
> **Theorem 4.2** Let $x - 0$ be an equilibrium point for (4.1) [$\dot{x} = f(x)$]. Let $V : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function such that
>
> $$V(0) = 0 \text{ and } V(x) > 0, \qquad \forall\, x \neq 0 \tag{6.9}$$
> $$|x| \to \infty \quad \Rightarrow \quad V(x) \to \infty \tag{6.10}$$
> $$\dot{V}(x) < 0, \qquad \forall x \neq 0 \tag{6.11}$$
>
> then $x = 0$ is globally asymptotically stable. •

The result given above, as well as its converse, were originally contributed by E. A. Barbashin and N. N. Krasovskiĭ in [6]:

> (Cited and translated from [6, p. 454])
> Theorem 1. If there exists a positively definite, infinitely large function $v(x_1, x_2, \ldots, x_n)$ which has definitely negative derivative then trivial solution of system (1) is asymptotically stable for any initial perturbations. •

Remark 6.1 (on [6, Theorem 1]).

1. In the notation of [6] system "(1)" corresponds to the system

$$\frac{dx}{dt} = X_i(x_1, x_2, \ldots, x_n), \quad i = 1, 2, \ldots, n\,.$$

   where the functions X_i are assumed to be continuously differentiable.
2. After [6], "*a function $v(x_1, x_2, \ldots, x_n)$ is called infinitely large if for any positive number A one can determine a constant N so large that for*

$$\sum_{i=1}^{n} x_i^2 > N$$

   we have $v(x_1, x_2, \ldots, x_n) > A$.

The converse of [6, Theorem 1] is also presented in that article:

> (Cited and translated from [6, p. 454])
> Theorem 2. If the trivial solution $x_i = 0$ is asymptotically stable for any initial perturbations, then there exists a continuously differentiable, infinitely large and positive definite function $v(x_1, x_2, \ldots, x_n)$ which has definitely negative derivative with respect to time.
> [...]
> For the previous theorem it is assumed that solutions of system (1) exist on the interval $-\infty < t \leq 0$. •

The following result, equivalent to that contained in [6, Theorem 1], appeared later in [25]:

> (Cited from [25, p. 67])
> IX. Theorem. Let $V(x)$ be a scalar function with continuous first partial derivatives for all x. Suppose that: (i) $V(x) > 0$ for $x \neq 0$; (ii) $\dot{V}(x) < 0$ for $x \neq 0$; and (iii) $V(x) \to \infty$ as $\|x\| \to \infty$. Then the system [$\dot{x} = X(x),\ X(0) = 0$] is completely stable. •

6.5.4 An Illustrative Example

We close our discussion on the "globalisation" of asymptotic stability with the following example that was contributed by E. A. Barbashin and his pupil, N. N. Krasovskiĭ. The example illustrates, beyond Hahn's words, that asymptotic stability in the large may be significant also in cases when the system is defined in the whole phase-space but such that not all trajectories originating in $\mathbb{R}^n$ tend to the origin. This also makes it clear that asymptotic stability in the large and asymptotic stability in the whole are different properties.

> (Cited and translated from [6])
> Let us consider the system
>
> $$\dot{x} = -\frac{2x}{(1+x^2)^2} + 2y\,, \qquad \dot{y} = -\frac{2y}{(1+x^2)^2} - \frac{2x}{(1+x^2)^2} \tag{2}$$
>
> For this system the following positive-definite function will serve us as a Lyapunov function: $v(x, y) = y^2 + \dfrac{x^2}{1+x^2}$. Next, we have
>
> $$\frac{dv}{dt} = -\frac{4x^2}{(1+x^2)^4} - \frac{4y^2}{(1+x^2)^2}\,.$$
>
> Evidently, dv/dt is a negative-definite function. However, we will show that on the plane (x, y) there is a set of instability for the system (2). Indeed, consider a curve (γ) given by the equation $y = 2 + \dfrac{1}{1+x^2}$. Calculating $\dfrac{dx}{dt}$ and $\dfrac{dy}{dt}$ along this curve,

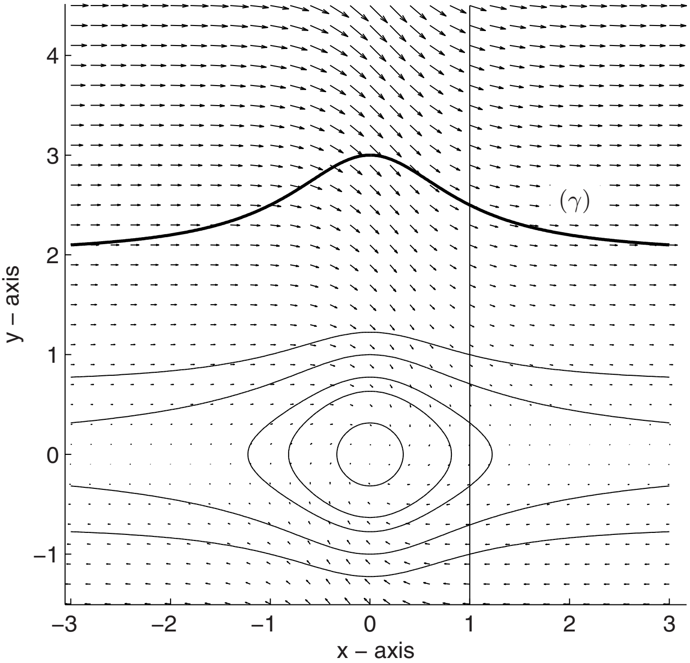

Fig. 6.3. Example from [6]. Flowchart on the phase plane.

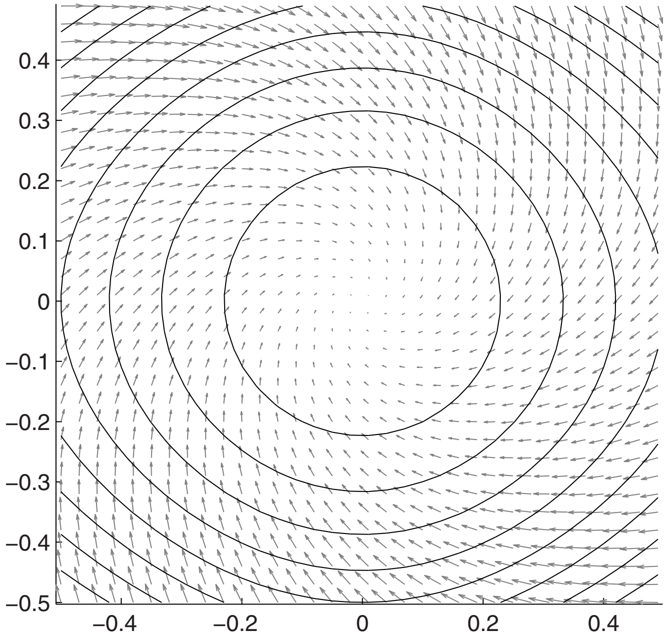

Fig. 6.4. Example from [6]. Estimate of the region of attraction.

$$\frac{dx}{dt} = -\frac{2x}{(1+x^2)^2} + 4 + \frac{2}{1+x^2} \,,$$

$$\frac{dy}{dt} = -\frac{1}{(1+x^2)^2}\left(4 + \frac{2}{1+x^2}\right) - \frac{2x}{(1+x^2)^2} \,,$$

we get

$$\frac{dy}{dx} = -\frac{x}{2(1+x^2)^2}\left(\frac{1 + \frac{1}{x}\left(2 + \frac{1}{1+x^2}\right)}{1 + \frac{1}{2(1+x^2)} - \frac{x}{2(1+x^2)^2}}\right) .$$

Evidently we can choose x_0 so large that for $x \geq x_0$ we will have $dx/dt > 0$ and $dy/dt > -x/(1+x^2)^2$ for the points on the curve (γ). Since the angle coefficient of the tangent to the curve (γ) is equal to $-2x/(1+x^2)^2$, we come to the conclusion that for $x > x_0$, trajectories of (2) cross the curve (γ) from below. Since at the point of intersection of the line $x = x_0$ and (γ) we have dx/dt at the point of the line $x = x_0$ which lay above the curve (γ) and therefore every trajectory of system (2) crosses the upper part of the line $x = x_0$ from left to right as time grows. Now, let us consider the domain G, defined by inequalities $x \geq x_0$, $y \geq 2 + 1/1 + x^2$. From the previous reasoning it is clear that none of the [trajectories starting from the] points of the set G can leave the set G as time passes hence, they cannot approach the origin.

●

In Figure 6.3 is depicted the flow of system [6, Eq. (2)] and the curve γ; we see clearly that the origin is asymptotically stable in the large, *e.g.*, for initial states $|x_0, y_0| < 1$. On the other hand, even though the system is well defined for all x and $y \in \mathbb{R}$ the region on the right hand side of the line $x \equiv 1$ and above the curve γ is clearly contained in the region of *instability*; hence, the origin is not asymptotically stable in the whole. It is also clear that the origin is (locally) asymptotically stable since the origin is stable and, as it is appreciated from Figure 6.4, any solution originating in an (infinitely) small neighbourhood of zero tends to it as time goes to infinity.

Thus, one shall not confuse *local* asymptotic stability in the sense given by Lyapunov, in [34], in which case (6.6) holds only for initial states in an *infinitely small* neighbourhood of the origin; asymptotic stability *in the large*, for which (6.6) holds for all initial states in *a given* region of $\mathbb{R}^n$, and asymptotic stability *in the whole*, which is the same as *global* asymptotic stability and in which case (6.6) holds for *all* initial states in the phase space, *i.e.* in $\mathbb{R}^n$. For *autonomous* systems, the origin is asymptotically stable in the whole if it is asymptotically stable in the large, in the sense of [12, Definition 6.1], for *any* (*i.e.* arbitrarily large) positive number Δ_0. All these properties, as much as their mathematical differences, cover importance when stating sufficient and necessary conditions to establish them, as we see in the following paragraphs.

6.6 On the Trace of "Krasovskiĭ-La Salle's Theorem"

6.6.1 Autonomous Systems

The conditions of the theorem for global asymptotic stability given in the previous section are hard to meet in a number of particular applications. Specifically, finding a Lyapunov function with a negative definite derivative is in general a rather difficult task. In 1960, J. P. La Salle published a number of theorems for asymptotic stability for the case when one does not know a Lyapunov-like function with a negative definite derivative. A similar formulation of the same result was contributed eight years before La Salle, by E. A. Barbashin and N. N. Krasovskiĭ, in [6]. Other formulations of this theorem were presented also for asymptotic stability of time-varying periodic systems, *e.g.* in [22]. Further, a more general result, commonly known as *La Salle's invariance principle* establishes convergence of trajectories to an invariant set –*cf.* [16].

Such theorems, or rather, recent *reformulations* of them, are widely used in control theory and the study of stability of dynamical systems; specially, the so-called La Salle's theorem, some times wrongly called La Salle's invariance principle (the latter is more general, to some extent) is broadly used by control engineers. The importance of these theorems in modern nonlinear control theory motivates us to recall their original formulations. To that end, we start by recalling the contributions of E. A. Barbashin and N. N. Krasovskiĭ.

> (Cited and translated from [6])
> Theorem 4. Let exist an infinitely large definitely positive function $v(x_1, x_2, \dots, x_n)$ and a set M such that
>
> $$\frac{dv}{dt} < 0 \quad \text{not in } M; \qquad \frac{dv}{dt} \leq 0 \ \text{ in } M.$$
>
> Let the set M have the property that on any intersection of the set $v = c$ $(c \neq 0)$ and M there does not exist positive semi-trajectories of system (1). We state that the trivial solution $x_i = 0$ of system (1) is asymptotically stable for any initial perturbations. •

For the sake of illustration[24] let us consider a system of two differential equations described by [6, Equation (1)], *i.e.*

[24] Some readers will recognise in the theorem above important similitude with La Salle's theorem for asymptotic stability; to the point that some doubt might rise on whether the set M is such that $\dot{v} = 0$ for all $x \in M$, rather than $\dot{v} \leq 0$. To

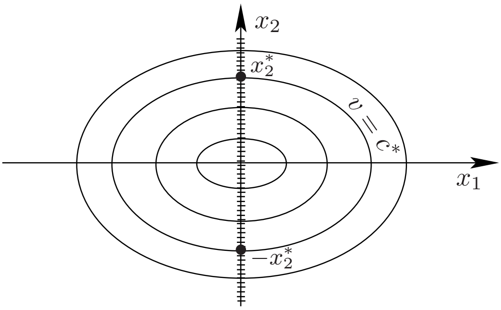

Fig. 6.5. Illustration of Theorem 4 from [6]

$$\frac{dx}{dt} = X_i(x_1, x_2), \quad i = 1, 2\,.$$

Assume that there exists a positive definite, radially unbounded (infinitely large) function $v(x_1, x_2)$ some of whose level curves are showed in Figure 6.5. Assume further that there exists a continuous function $x_1 \mapsto w$ such that $w(0) = 0$, $w(x_1) > 0$ for all $x_1 \neq 0$ and

$$\begin{aligned} \frac{dv}{dt} &= \frac{\partial v}{\partial x_1} X_1(x_1, x_2) + \frac{\partial v}{\partial x_2} X_2(x_1, x_2) \\ &= -w(x_1)\,. \end{aligned}$$

To apply [6, Theorem 4] recalled above, we see that $M = \{x_1 = 0, x_2 \in \mathbb{R}\}$, *i.e.* M corresponds to the vertical axis of the phase space frame. According to the theorem we must verify that the origin is the only element of the intersection of M with $\{v = c\}$ that contains continuous positive semi-trajectories (*i.e.* functions $t \mapsto x$ with $t \geq 0$). In other words, we must verify that the largest invariant[25] set $E \subset \mathbb{R}^n$ contained in $M \cap \{v = c\}$ is the origin. To that end, fix $c = c^* > 0$ arbitrarily; in this case, $M \cap \{v = c^*\} = \{(0, -x_2^*),\ (0, x_2^*)\}$ with $x_2^* \neq 0$, in view of the positivity of v –*cf.* Figure 6.5. Now we pose ourselves the question: is any of the trajectories $x_1(t) \equiv 0$, $x_2(t) \equiv \pm x_2^*$ a solution of the system $\dot{x}_1 = X_1(x_1, x_2)$, $\dot{x}_2 = X_2(x_1, x_2)$? Otherwise, does $0 = X_2(0, x_2)$ has other solutions than $\{x_2 = 0\}$? If the answer is negative, it follows that the origin is globally asymptotically stable.

The previous reasoning, which is at the origin of [17, Corollary 2.1], may certainly be generalised (*cf.* [6]) and is actually at the basis of the rationale behind what is often called *La Salle's theorem* for asymptotic stability. The following more general theorem appears in [16] –see also [25, p. 66], [56, p.

avoid ambiguity, we stress that this is not the case; this theorem is repeated in [4, p. 25] –*cf.* also [5, p. 49].

[25] We remind the reader that a set E is said to be (forward) invariant if it is such that $x(t_0) \in E$ implies that $x(t) \in E$ for all $t \geq t_0$.

51], [55, p. 52]. This theorem is commonly referred to as La Salle's invariance principle.

> (Cited from [16, Theorem 3])
> Let $V(x)$ be a scalar function with continuous first partials for all x. Assume that
>
> 1) $V(x) > 0$ for all $x \neq 0$
>
> 2) $\dot{V}(x) \leq 0$ for all x.
>
> Let E be the set of all points where $\dot{V}(x) = 0$, and let M be the largest invariant set contained in E. Then every solution of (2) [$\dot{x} = X(x)$] bounded for $t \geq 0$ approaches M as $t \to \infty$. •

Note that the invariance principle does *not* establish asymptotic stability but guarantees the attractivity of the set M, assumed invariant. However, in view of condition 2) above, the origin is Lyapunov stable. Further, in the particular case where $M = \{0\}$ and V is radially unbounded, we recover the well known La Salle's theorem for global asymptotic stability and which is equivalent to [6, Theorem 4], published eight years earlier. For completeness, we recall *La Salle's theorem* next:

> (Cited from [25, p. 525])
> *Theorem 4—A Complete Stability Theorem*
>
> *Let $V(x)$ be a scalar function with continuous first partials satisfying*
>
> 1) $V(x) > 0$ *for all* $x \neq 0$
>
> 2) $\dot{V}(x) \leq 0$ *for all* x
>
> 3) $V(x) \to \infty$ *as* $\| x \| \to \infty$.
>
> *If $\dot{V}$ is not identically zero along any solution other than the origin, then the system (2) is completely stable* [globally asymptotically stable]. •

6.6.2 Time-varying Periodic Systems

Extensions of [6, Theorem 4], to the case of non-autonomous periodic systems, have also been published; the first is probably due to N. N. Krasovskiĭ:

> (Cited from [22, Chapter 3, p. 66-67])
> [...] we consider the more general case in which the equations
>
> $$\frac{dx_i}{dt} = X_i(x_1, \ldots, x_n, t) \qquad (i = 1, \ldots, n) \tag{14.1}$$
>
> of [the] perturbed motion are such that the right members $X_i(x, t)$ are periodic functions of the time t with period ϑ, or do not depend

> explicitly on the time t. We further assume that the functions are defined in the region
>
> (14.2) $\qquad \|x\| < H, \quad -\infty < t < \infty \qquad (H = \text{const. or } H = \infty)$
>
> [...]
>
> THEOREM 14.1. *Suppose the equations of perturbed motion (14.1) enjoy the properties that*
>
> *(i) there exists a function $v(x,t)$ which is periodic in the time t with period ϑ or does not depend explicitly on the time;*
>
> *(ii) $v(x,t)$ is positive definite;*
>
> *(iii) $v(x,t)$ admits an infinitely small upper bound in the region (14.2);*
>
> *(iv)*
>
> $$\sup\left(v \textit{ in the region } \|x\| \le H_0, \quad 0 \le t < \vartheta\right)$$
> $$< \inf\left(v \textit{ for } \|x\| = H_1\right) \quad (H_0 < H_1 < H);$$
>
> *(v)*
>
> $$dv/dt \le 0 \textit{ in the region (14.2)};$$
>
> *(vi) the set M of points at which the derivative dv/dt is zero contains no nontrivial half trajectory*
>
> $$x(x_0, t, t) \qquad (0 < t < \infty)$$
>
> *of the system (14.1)*
>
> *Under these conditions, the null solution $x = 0$ is asymptotically stable and the region $\|x\| \le H_0$, lies in the region of attraction of the point $x = 0$.* •

Modern formulations of the latter can be found for instance in [65, p. 179]; specifically, [65, Theorem 5.3.79], which is called by the author “Krasovskiĭ-La Salle’s theorem”, corresponds to [22, Theorem 14.1] given above, to the case when $H = \infty$, *i.e.* the case of global asymptotic stability:

> (Cited from [65, p. 179])
>
> **79 Theorem (Krasovskii-LaSalle)** Suppose that the system (5.1.1) [$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}[t, \boldsymbol{x}(t)], \ t \ge 0$] is periodic. Suppose that there exists a C^1 function $V : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ having the same period as the system such that (i) V is a pdf [positive definite] and is radially unbounded, and (ii)
>
> **80** $\qquad \dot{V}(t, \boldsymbol{x}) \le 0, \quad \forall\, t \ge 0, \quad \forall\, \boldsymbol{x} \in \mathbb{R}^n .$
>
> Define

> **81** $R = \{\boldsymbol{x} \in \mathbb{R}^n : \exists t \geq 0 \quad \text{such that} \quad \dot{V}(t, \boldsymbol{x}) = 0\}\,,$
>
> and suppose R does not contain any trajectories of the system other than the trivial trajectory. Then the equilibrium **0** is globally uniformly asymptotically stable. •

We emphasise that [65, Theorem 5.3.79] cited above, establishes global *uniform* asymptotic stability; as a matter of fact, in this case we can also read global asymptotic stability in the sense of Definition 6.9 since, for periodic systems, this is equivalent to global uniform asymptotic stability:

> (Cited from [15, Theorem 38.5, p. 184]) If the equilibrium of the differential equation $\dot{\boldsymbol{x}} = f(\boldsymbol{x}, t) \quad (\boldsymbol{f} \in E)$ with constant or periodic coefficients is stable then it is uniformly stable. If the equilibrium is asymptotically stable then it is uniformly asymptotically stable. •

[26]Uniformity is understood in the sense that attractivity and stability are independent of the initial time; this is the topic of coming section.

6.7 Uniformity

So far we have discussed the "well-known" concepts of stability, asymptotic stability and non-local versions of the latter. We have made no explicit distinction between autonomous and non-autonomous systems. Making a distinction covers importance when addressing the problem of *uniformity*.

Consider a dynamical system $\dot{x} = f(t, x)$ – to fix the ideas, assume that f is such that the solutions exist on compact intervals of t and are unique – and assume that the origin is asymptotically stable, possibly, globally asymptotically stable. That is, it possesses the properties described in Definitions 6.6, 6.9 and possibly 6.11. What are the consequences of having that δ in (6.2) and T in (6.7) depend on the initial time and state t_0, x_0? Furthermore, under which conditions can one guarantee that stability and attractivity are *uniform* in these parameters? Concerning the first question, we will see in Section 6.8 that *only* uniform asymptotic stability guarantees certain robustness with respect to external perturbations. The second question is addressed below.

[26] In Hahn's notation –*cf.* [14, p. 8], $\boldsymbol{f} \in E$ if $\boldsymbol{f}$ is such that the solutions of $\dot{x} = f(x, t)$ are unique and continuous in the initial conditions. In [15, p. 56] the author redefines "class E", roughly, as the class of functions such that solutions are unique on compacts of x and t and for which the origin is an isolated equilibrium point.

6.7.1 Uniform Stability

Uniform stability is defined as follows – see the equivalent definitions given in [65, p. 137], [54, p. 7], [2, p. 143], [20, Definition 4.4., p. 149]:

Definition 6.12 (Uniform stability). *The origin of the system (6.1) is said to be* **uniformly** *stable if for each* $\varepsilon > 0$ *there exists* $\delta(\varepsilon) > 0$ *such that*

$$|x_\circ| \le \delta \quad \Longrightarrow \quad |x(t; t_\circ, x_\circ)| \le \varepsilon \tag{\#}$$

for all $t \ge t_\circ$ *and all* $t_\circ \ge 0$ *.*

Different authors attribute this property to Persidskiĭ: Antosiewicz, in [2], attributes it to [51] while Rouche *et al*, in [54], attribute it to [49]. Persidskiĭ himself, in [51] refers to [49]:

> (Cited and translated from [49])
> Assume that $x_s = f_s(t, t_1)$, $(s = 1, 2, \ldots, n)$ is a system of continuous functions that satisfies the following system of differential equations of perturbed motion
>
> $$\frac{dx_s}{dt} = y_s(x_1, x_2, \ldots, x_n, t) \quad (s = 1, 2, \ldots, n) \tag{1}$$
>
> and which takes, for $t = t_1$ corresponding values ε_s [*i.e.*, with initial conditions t_1 and $x_s(t_1) = \varepsilon_s$].
>
> If the functions $x_s = f_s(t, t_1)$ are such that for any arbitrarily small number $H > 0$ there exists a number $h > 0$ such that for all values $t \ge t_1$ we will have
>
> $$(x_1^2 + x_2^2 + \cdots + x_n^2) \le H, \tag{2}$$
>
> if
>
> $$(\varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2) \le h \tag{3}$$
>
> then the unperturbed motion will be called stable.
> [. . .]
> In general, the number h is a function of t_1 and H. In the case when for all values $t_1 \ge t_0$ there exists a number h, which is independent of t_1, we will call the unperturbed motion uniformly stable. •

It is important to stress that Persidskiĭ does not assume the usual local Lipschitz, uniform in t, property on the function $y_s(\cdots, t)$ but that there exist continuous positive functions $A_s(t)$ such that

$$|y_s(x_1, x_2, \ldots, x_n, t) - y_s(x_1', x_2', \ldots, x_n', t)| \le A_s(t)(|x_1 - x_1'| + \cdots + |x_n - x_n'|)$$

and y_s are continuous.

Starting with W. Hahn, the notation from Definition 6.7 has been used. More precisely, we have the following:

> (Cited from [14, p. 62])
> THEOREM 17.1: The equilibrium of differential equation (2.7) [$\dot{x} = f(x,t),\ f(0,t) = 0, f \in E$] is uniformly stable if and only if there exists a function $\rho(r)$ with the following properties:
>
> (a) $\rho(r)$ is defined, continuous, and monotonically increasing in an interval $0 \le r \le r_1$;
>
> (b) $\rho(0) = 0$; the function ρ, therefore, belongs to the class K;
>
> (c) the inequality
> $$|p(t, x_0, t_0)| \le \rho(|x_0|)$$
> is valid for $|x_0| < r_1$.
>
> •

We remind the reader that, in Hahn's notation $p(t, x_0, t_0)$ corresponds to the solution of the differential equation $\dot{x} = f(x,t)$ and $f \in E$ if f is such that the solutions of $\dot{x} = f(x,t)$ are unique and continuous in the initial conditions (*cf.* [14, p. 3]).

After Definition 6.7 it is mentioned that the function φ may depend on t_0; in that case, we speak of Lyapunov stability. In the characterisation of [14, THEOREM 17.1] above, the function ρ is independent of the initial conditions, specifically, it is independent of t_0.

Hahn attributes the following result to [50]:

> (Cited from [14]) THEOREM 17.6 If there exists a positive definite decrescent Liapunov function v such that its total derivative $\dot{v}$ for (2.7) is negative semi-definite, then the equilibrium is stable. •

The sufficiency theorem cited above is also attributed by Rouche *et al* [54] to K. P. Persidskiĭ –[50] while Antosiewicz [2] cites [51]. Indeed, Persidskiĭ gives in, [50] and for the first time, necessary and sufficient conditions for uniform stability. That is, Persisdskĭi's original statement is more general than that contained in [14, THEOREM 17.6] however, its precise formulation requires the introduction of other definitions[27] that we do not detail here.

6.7.2 Uniform Global Stability

The concept of uniform global stability has been used implicitly at least, from the 1950s both by Soviet and Western authors; however, an explicit definition remains absent in texts such as [65, 18, 19, 20, 54, 56], among others. Roughly speaking, uniform global stability is the property of *uniform stability* plus

[27] It is worth pointing out that among these definitions, Persidskiĭ uses the terminology "class $\mathcal{L}$" function however, Persidskiĭ's definition is different from "Hahn's" definition of class $\mathcal{L}$ function, which is commonly used nowadays.

uniform global boundedness of the solutions; the first is a local property while the second implies that the overshoot of the norm of the solutions has an upper limit independent of the initial conditions, specifically, of t_0. Thus, uniform global stability is defined as follows:

> (Cited from [24, p. 490])
> **Definition A.4** The equilibrium point $x = 0$ of (A.1) [$\dot{x} = f(x,t)$] is
>
> - uniformly stable, if there exists a class $\mathcal{K}$ function $\gamma(\cdot)$ and a positive constant c independent of t_0, such that
>
> $$|x(t)| \leq \gamma(|x(t_0)|)\,, \quad \forall\, t \geq t_0\,, \forall\, x(t_0) \mid |x(t_0)| < c; \qquad (A.3)$$
>
> - globally uniformly stable, if (A.3) is satisfied with $\gamma \in \mathcal{K}_\infty$ for any initial state $x(t_0)$.
>
> •

We remind the reader that $\gamma \in \mathcal{K}_\infty$ *if* $\gamma \in \mathcal{K}$ *and* $\gamma(s) \to \infty$ *as* $s \to \infty$. In view of the discussion in Section 6.5.1 and [14, Theorem 17.1] cited above, [24, Definition A.4] is equivalent to Hahn's uniform stability *in the whole*:

> (Cited from [14, p. 62])
> Definition 17.2: The equilibrium of (2.7) [$\dot{x} = f(x,t),\ f(0,t) = 0, f \in E$] is said to be uniformly stable in the whole if the assumption of Theorem 17.1 are satisfied for every arbitrarily large r_1. •

As we will see further, uniform global stability is significant to define uniform global asymptotic stability. See also Appendix A of this text.

6.7.3 Uniform Asymptotic Stability

The property of uniform asymptotic stability appears, implicitly, in different articles of I. G. Malkin between 1940 and 1955 in the context of stability with respect to *constantly-acting disturbances* –*cf.* Section 6.8. Hahn attributes the following formulation of uniform asymptotic stability (UAS) to Malkin:

> (Cited from [14])
> Definition 17.4 (Malkin [20]): *The equilibrium of (2.7) is called uniformly asymptotically stable if*
>
> 1. *the equilibrium is uniformly stable*
> 2. *for every $\epsilon > 0$ a number $\tau = \tau(\epsilon)$ depending only on ϵ, but not on the initial instant t_0 can be determined such that the inequality*
>
> $$|p(t, x_0, t_0)| < \epsilon \qquad (t > t_0 + \tau)$$

> holds, provided x_0 belongs to a spherical domain Re_η whose radius η is independent of ϵ. •

Hahn's reference "Malkin [20]" corresponds to the paper [39], on the converse Lyapunov theorem for uniform asymptotic stability and on stability with respect to constantly-acting perturbations. Even though the terminology "uniform asymptotic stability" is actually *not* used in [39], Malkin establishes sufficient conditions for stability with respect to constantly-acting disturbances, which are also sufficient (and under extra regularity conditions, necessary) for uniform asymptotic stability.

The second property in [14, Definition 17.4] is often referred to as *uniform attractivity*[28] –*cf.* [54, p. 8], of which the following interesting characterisation is seemingly due to Hahn:

> (Cited from [14, p. 64]) Necessary and sufficient for the second condition of Definition 17.4 is the existence of a function $\sigma(r)$ with the following properties:
>
> (a) $\sigma(r)$ is defined, continuous, and monotonically decreasing, for all $r \geq 0$,
>
> (b) $\lim\limits_{r\to\infty} \sigma(r) = 0$,
>
> (c) provided the initial points belong to a fixed spherical domain Re_η, the relation
>
> $$|p(t, x_0, t_0)| \leq \sigma(t{-}t_0) \tag{17.6}$$
>
> holds. •

Further, the following characterisation of uniform asymptotic stability is also established by Hahn:

> (Cited from [14, p. 64])
> THEOREM 17.4 (Hahn): *Necessary and sufficient for uniform asymptotic stability of the equilibrium is the existence of two functions $\kappa(r)$ and $\vartheta(r)$ with the following properties:*
>
> *(a) $\kappa(r)$ satisfies assumptions* (*a*) *and* (*b*) *of Theorem 17.1,*
>
> *(b) $\vartheta(r)$ satisfies the corresponding assumptions of Theorem 17.3;*
>
> *(c) in addition, the inequality*
>
> $$|p(t, x_0, t_0)| \leq \kappa(|x_0|)\vartheta(t - t_0) \tag{17.7}$$
>
> *holds, provided that the initial points x_0 belong to a fixed spherical domain* Re_η. •

[28] Hahn uses the terminology *uniformly attracting* as a qualifier for the equilibrium.

6.7.4 Uniform Asymptotic Stability in the Large

Among the consulted references listed in the bibliography the only precise formulation of uniform asymptotic stability in the large that we have located is the following[29].

> (Cited from [22, p. 30])
> DEFINITION 5.3. The null solution $x = 0$ is called uniformly asymptotically stable in the large in the region G if for arbitrary preassigned positive $\eta > 0$ and arbitrary H_0, $\bar{H}_0 \subset G$, there are always a number $T(H_0, \eta)$ and a bounded region H_1, $\bar{H}_1 \subset G$ such that the relations
>
> $$x(x_0, t_0, t) \in H_1 \quad \text{for all} \quad t \geq t_0,$$
> $$\|x(x_0, t_0, t)\|_2 < \eta \quad \text{for all} \quad t \geq t_0 + T(H_0, \eta)\,,$$
>
> hold for every initial moment of time t_0 and for every given value of $x_0 \in H_0$.
>
> A sufficient condition for [uniform] asymptotic stability in the large is the following.
>
> THEOREM 5.2. The null solution $x = 0$ of equations (1.3) [*cf.* p. 220] is asymptotically stable in the large in the region G if there exists a function $v(x, t)$ such that
>
> (i) $v(x, t)$ is positive definite in G
>
> (ii) $v(x, t)$ admits an infinitely small upper bound in G;
>
> (iii) $v(x, t)$ admits an infinitely great lower bound on the boundary of G ($v(x, t)$ is radially unbounded in G);
>
> (iv) The derivative dv/dt along a trajectory of (1.3) is negative-definite in G.
>
> ●

After Krasovskiĭ – *cf.* [22], the previous theorem was established in [7], [6]; he also adds the following comment concerning uniform asymptotic stability in the whole:

> "*for the case in which G is the entire space $\infty < x_i < \infty$; [...] The theorem is incorrect without the assumption (iii); a counter-example appears in*" [6]. ●

To avoid confusion, it is important to stress that [6] deals with asymptotic stability in the whole of *autonomous* systems (*i.e.*, uniformity is obtained for

[29] The symbol $|\cdot|_2$ is used by Krasovskiĭ to denote Euclidean norm.

free) and therefore, the counter-example refers to the case when v depends only on x and is not radially unbounded on $\mathbb{R}^n$; in this case, one is led to conclude asymptotic stability in the large and not asymptotic stability in the whole. The counter-example mentioned by Krasovskiĭ concerns autonomous systems and corresponds to that cited in Section 6.5.4.

6.7.5 Uniform Asymptotic Stability in the Whole

Uniform asymptotic stability in the whole or, uniform global asymptotic stability (UGAS) is one of the strongest stability properties that one may have for the equilibrium of non-autonomous differential equations[30]

$$\dot{x} = f(t, x)\,. \tag{6.12}$$

UGAS was introduced in [7] which is a natural extension, to the case of time-varying systems, of the previous results of the same authors, [6] and that we already discussed:

> (Cited and translated from [7, p. 346])
> We call the solution $x_1 = \ldots = x_n = 0$ of system (1)
>
> $$\left[\qquad \frac{dx_i}{dt} = X_i(x_1, \cdots, x_n, t) \qquad (i = 1, \cdots, n) \qquad (1) \quad \right]$$
>
> [is] uniformly [asymptotically] stable in the whole, if for any numbers $R_1 > 0$ and $R_2 > 0$ one can find a number $T(R_1, R_2)$, depending continuously only on R_1 and R_2, such that, any solution $x_i(x_{10}, \ldots x_{n0}, \tau_0, t)$ $(i = 1, \ldots, n)$ with initial values for $t = \tau_0 \geq t_0$ laying in the region
>
> $$x_{10}^2 + \cdots + x_{n0}^2 \leq R_1^2\,,$$
>
> satisfies inequality
>
> $$x_1^2 + \cdots + x_n^2 < R_2^2 \qquad \text{for} \;\; \tau_0 + T(R_1, R_2)$$
>
> and at same time for any number $R_1 > 0$ there exists a number $R_2 = F(R_1)$, depending continuously only on R_1, such that any trajectory starting from the interior of a sphere of radius R_1 does not escape from a sphere of radius R_2 as time passes. •

The following observations on the previous definition are worth listing:

- note that the term "asymptotically" is omitted by the authors;

[30] Where f is such that the solutions exist on compact intervals and are unique.

- the first part of the definition corresponds to uniform global attractivity while the second part corresponds to uniform boundedness of solutions hence, the constants R_1 and R_2 are not the same in each of the two parts of the definition;
- furhtermore, from the rigorous viewpoint that characterises modern literature, the definition above actually does not explicitly include uniform *stability*;
- in addition, it is not made explicit whether the function F is radially unbounded.

Such imprecisions are not uncommon in early Soviet literature however, one should not hold rigour to these authors. Considering the rest of contents of the paper and citations of [7] by succeeding authors it seems safe to assume that Barbashin and Krasovskiĭ did implicitly assume in their definition, that the origin is required to be uniformly stable and that F is radially unbounded. As a matter of fact, the latter may be inferred from the last sentence of the definition: given *any* $R_1 > 0$ there exists $R_2 = F(R_1)$ such that

$$|x(t_0)| < R_1 \implies |x(t)| \leq F(R_1) \qquad \forall\, t \geq t_0\,;$$

hence, $|x(t_0)| < R_1$ implies that $|x(t_0)| \leq F(R_1)$ which in turn implies that $F(R_1) \geq R_1$. Barbashin and Krasovskiĭ's formulation is closely followed (and formalised) by Hahn –*cf.* [14, Definition 17.5]:

> (Cited from [14, p. 64])
> Definition 17.5 (Barbashin and Krasovskii [2]): The equilibrium of differential equation (2.7) is said to be uniformly asymptotically stable in the whole, if the following two conditions are satisfied:
>
> (a) The equilibrium is uniformly stable in the whole;
>
> (b) for any two numbers $\delta_1 > 0$ and $\delta_2 > 0$ there exists a number $\tau(\delta_1, \delta_2)$ such that
> $$|p(t, x_0, t_0)| < \delta_2$$
> if $t \geq t_0 + \tau(\delta_1, \delta_2)$ and $|x_0| < \delta_1$. •

W. Hahn's citation "[2]" corresponds to the paper [7], which suggests that W. Hahn considered that the authors of the latter assumed implicitly the radial unboundedness of F. In the cas e that radial unboundedness of F is overlooked then item (a) of [14, Definition 17.5] reads *"the equilibrium is uniformly stable"*. That is, in one case we have "*the origin is UGAS if it is uniformly stable and uniformly globally attractive*" while in the second, we have *"the origin is UGAS if it is uniformly* globally *stable and uniformly globally attractive"*. More precisely, the following two definitions of UGAS may be found in the literature:

Definition 6.13. *([54, p. 10]*[31]*,[42, p. 356],[3, Definition 3.6, p. 80],[65, Definition 5.1.38, p. 143]) The origin of (6.12) is said to be uniformly globally asymptotically stable (UGAS) if it is uniformly stable –cf. Definition 6.12 and uniformly globally attractive, i.e. if for any* $r > 0$ *and* $\sigma > 0$ *there exists* $T(\sigma, r) > 0$ *such that, for all* $t_\circ \geq 0$*,*

$$|x_\circ| \leq r \quad \Longrightarrow \quad |x(t; t_\circ, x_\circ)| \leq \sigma \qquad \forall\, t \geq t_\circ + T\,.$$

In the second case the definition of UGAS reads:

Definition 6.14. *([14, Definition 17.5, p. 64],[20, p. 150],[17, Definition 2.7, p. 38]) The origin of the system (6.12) is uniformly globally asymptotically stable if*

1. *it is uniformly* globally *stable, i.e. there exists* $\gamma \in \mathcal{K}_\infty$ *such that*
$$|x(t)| \leq \gamma(|x_\circ|) \tag{6.13}$$
2. *it is uniformly globally attractive.*

Following [14, Theorem 17.1] –*cf.* page 238 of these notes, we see that the first condition in Definition 6.13 is equivalent to the existence of a function $\gamma \in \mathcal{K}$ such that (6.13) holds for sufficiently small initial states (in the domain of definition $\gamma^{-1}(\cdot)$. In contrast to this, in the case of the first condition in Definition 6.14, the same bound holds with $\gamma \in \mathcal{K}_\infty$ and for all initial states in the phase space; this guarantees uniform global boundedness of the solutions. In particular, in the case of the second definition – and only in this case – we are ensured that the norm of the solutions, during the transient behaviour, does not take an overshoot that increases with larger and larger initial times.

The significance of this difference cannot be overestimated; only in the case of the second definition, the consequence [14, Theorem 17.4] –*cf.* p. 240, holds for *all* initial conditions. This fact is proved, as far as we know for the first time, in Appendix A (by A. Teel and L. Zaccarian) of this book via an example of a system whose origin possesses the property of Definition 6.13 but not that of Definition 6.14. By this, one should not haste to conclude that one of the two proposed definitions is invalid or “wrong” in any manner but rather, one shall recognise that the property of Definition 6.14 is stronger.

Besides the cited references, it is also worth mentioning that the definitions of UGAS in [24] and [18] are formulated in terms of $\mathcal{KL}$ bounds directly: [24, Definition A.4, p. 490] states that the origin is UGAS if a bound like [14, Inequality (17.7)] – *cf.* p. 240 – holds with $\kappa \in \mathcal{K}_\infty$ for any initial state while [24, Definition 4.3, p. 168] states that the origin is UGAS if a bound like

[31] The authors attribute this definition to the paper [6] which concerns only autonomous systems.

[14, Inequality (17.7)] holds for all initial states in the phase space. See also Appendix A on p. 285 for other references related to Definitions 6.13 and 6.14.

We close the section with the following classical result –*cf.* [65], [19], [54], [18], [20], [15], [64] and which is originally due to E. A. Barbashin and N. N. Krasovskiĭ:

> (Cited and translated from [7])
> For the solution $x_1 = x_2 \ldots = x_n = 0$ of system (1) to be uniformly [asymptotically] stable in the whole it is necessary and sufficient that in the whole space $-\infty < x_i < \infty$ $(i = 1, \ldots, n)$ there exists positive definite, infinitely large [*i.e.* radially unbounded] function $v(x_1, \ldots, x_n, t)$ which allows an upper limit infinitely small in the point 0 [*i.e.* decrescent] which has definitely negative derivative dv/dt in the whole space $\{x_i\}$. •

It is important to stress that in the cited theorem the authors assume that the functions X_i defined in [7, Equation (1)] are once continuously differentiable.

6.8 Stability with Respect to Perturbations

In this concise survey we have limited our review to the most commonly-used and most elementary forms of stability in the sense of Lyapunov; there are many other forms of stability that we have omitted to discuss and that have been subject of study in the literature of dynamical systems within the 1930s through the 1950s. Among these, we shall briefly discuss a form of robust stability, some times called *total stability* – *cf.* [14, p. 104].

Total stability, or "stability with respect to constantly-acting perturbations", as I. G. Malkin called it, has its roots in Soviet literature from the late 1930s:

> (Cited and translated from [40, p. 20]. See also [41, p. 10][32])
> The influence of small disturbing forces on the stability of motion of a dynamic system was considered first in the following paper: Cetaev N. G., "On Stable Trajectories of Dynamics" Scientific Notes of Kazan University, vol. 4, no. 1 (see also Collection of scientific works of Kazan Aviation Institute No. 5, 1936). To this question are also devoted the following papers: Arteym'ev N. A., "Feasible Motions", Izvestia AN USSR, mathematical series, No. 3, 1939; Malkin I. G., "On Stability under constantly acting perturbations", Applied Mathematics and

[32] We stress that the translation [41] does not include the last two references of Gorshin that Malkin cites in [40, p. 20]; however, [41] corresponds to the translation of the *first* edition of [40], seemingly published in 1952.

> Mechanics, Vol. 8, No. 3, 1944; Gorshin S. I., “On Stability of motion under constantly acting perturbations. Critical cases”, Izvestia AN Kazahskoi SSR No. 56, series “Mathematics and Mechanics”, issue 2; “On Stability of motion under constantly acting perturbations, Izvestia AN Kazahskoi SSR No. 58, 1948. •

Among the references cited by Malkin, at the moment of writing these notes, we were able to locate only [38], where Malkin defines stability with respect to constantly acting disturbances:

> (Cited and translated from [38])
> Consider the system
>
> $$\frac{dx_s}{dt} = X_s(t, x_1, \ldots, x_n) \qquad (s = 1,\, 2,\, \ldots\,,\, n) \tag{1}$$
>
> where X_s are functions defined on the set
>
> $$t \geq 0, \qquad |x_s| \leq H \tag{2}$$
>
> with H a sufficiently small positive constant. [. . .] consider systems
>
> $$\frac{dx_s}{dt} = X_s(t, x_1, \ldots, x_n) + R_s(t, x_1, \ldots, x_n) \tag{3}$$
>
> where R_s are functions defined on the set (2), in which they are bounded and continuous and are not necessarily zero at $x_1 = \ldots = x_n = 0$.
>
> We will also assume that Eq. (3) allows existence of unique integral Cauchy solutions. [. . .]
>
> Definition. *The unperturbed motion (trivial solution $x_1 = \ldots = x_n = 0$) of equation (1) is called stable under constantly-acting disturbances*[1]*, if for any positive number $\varepsilon < H$, no matter how small it would be, there exist two other positive numbers $\eta_1(\varepsilon)$ and $\eta_2(\varepsilon)$ such that, any solution of equation (3) with initial conditions (for $t = t_0$), satisfying inequalities*
>
> $$|x_s^\circ| \leq \eta_1$$
>
> *for arbitrary R_s satisfying in the set $t \geq t_0$, $|x_s| \leq \varepsilon$, inequalities*
>
> $$|R_s(t, x_1, \ldots, x_n)| \leq \eta_2\,,$$
>
> *satisfy for all $t \geq t_0$, inequalities*
>
> $$|x_s| < \varepsilon\,.$$
>
> •

In the previous citation, the super-index [1] in "*disturbances*[1]" corresponds to a citation made by Malkin – here noted [10] – suggesting that the definition of "stability under constantly-acting disturbances" goes back to it. However, at the moment of writing these notes we have not been able to locate [10] and therefore, we are unable to comment on its contents. Besides the references that Malkin cites in his book [40, 41] he gives credit to Duboshin in other papers dealing with stability under constantly-acting disturbances.

In [38] Malkin establishes, for system [38, Eq. (3)], the property defined above under the conditions that there exist a Lyapunov function for the nominal system [38, Eq. (1)] that is positive definite, decrescent and with a negative definite derivative; that is, apart from the boundedness of $\partial V/\partial x$, under the same *sufficient* conditions for *uniform* asymptotic stability, *now* well-known to the control community (*cf. e.g.* [15, 65, 20]):

> (Cited and translated from [38])
> Theorem 1. If for the differential equations of the perturbed motion (1) there exists a positive definite function V, of which the complete derivative with respect to time, composed with these equations, is negative definite and if in the domain (2) the partial derivatives $\partial V/\partial x$ are bounded then, unperturbed motion is stable under constantly-acting perturbations. •

It is worth stressing that Malkin often omitted the qualifier "uniformly" however, it should be understood that in [38, Theorem 1] above, in the phrase "*the partial derivatives* $\partial V/\partial x$ *are bounded*" it is actually meant "*the partial derivatives* $\partial V/\partial x$ *are bounded uniformly in t*".

In the later paper [39], Malkin establishes a converse stability result under the standing assumption that the right-hand side of [38, Eq. (1)] is continuously differentiable with respect to all arguments and has bounded partial derivatives: *uniform* asymptotic stability implies the existence of a positive-definite decrescent Lyapunov function with negative definite derivative and for which $\partial V/\partial x$ is uniformly bounded. Malkin concludes in [39] that for systems with continuously differentiable right-hand sides and having bounded partial derivatives[33], uniform asymptotic stability implies stability with respect to constantly acting perturbations, *i.e.* total stability.

Examples of systems that are asymptotically stable but not uniformly, *i.e.* that satisfy Definition 6.9, and that are not stable with respect to perturbations have been reported for instance in [46] where the authors give an example of a linear time-varying system that is asymptotically stable but not uniformly asymptotically stable hence, even though this is not discussed in [46], one cannot expect this system to be stable with respect to certain bounded perturbations. In [4, p. 178][34] it is shown that the solutions of

[33] However, local Lipschitz in x, uniformly in t suffices.

[34] See also [5].

$$\dot{x} + (a - \sin(\ln(t+1)) - \cos(\ln(t+1)))x = u(t)$$

with a such that $0 < 1 < a < 1 + \frac{e^{-\pi}}{2}$ and $u(t) = e^{-a(t+1)}$, tend to infinity as $t \to \infty$; however, the solution of the system without input (*i.e.* with $u \equiv 0$) satisfies – *cf.* [4]:

$$|x(t)| \leq e^{2(t_0+1)} e^{-(a-1)(t-t_0)} |x(t_0)|$$

that is, the origin of the system without input is GAS since it is globally attractive and stable but it is neither uniformly attractive nor the solutions are globally *uniformly* bounded. Barbashin attributes this example to [45, 48]. In the recent paper [61] the authors present an interesting example of an *autonomous* system that is globally asymptotically stable but that can be destabilised by an integrable input. See also [63] for an example of a system that satisfies a global sector-growth condition, is exponentially stable and may be destabilised by a decaying exponential.

We wrap up the section with an example of a system that is uniformly globally stable and whose trajectories converge exponentially fast but with a convergence rate that depends on the initial times; hence, it is not uniformly asymptotically stable and one can construct non-vanishing perturbations that destabilise the system.

Proposition 6.1. *(Taken from [47][35]) Consider the system $\dot{x} = f(t,x)$ where*

$$f(t,x) = \begin{cases} -a(t)sgn(x) & \text{if} \quad |x| \geq a(t) \\ -x & \text{if} \quad |x| \leq a(t) \end{cases} \tag{6.14}$$

and $a(t) = \frac{1}{t+1}$. This system has the following properties:

1. *The function $f(t,x)$ is globally Lipschitz in x, uniformly in t.*
2. *For each $r > 0$ and $t_\circ \geq 0$ there exist strictly positive constants κ and λ such that for all $t \geq t_\circ$,*

$$|x(t_\circ)| \leq r \qquad \Rightarrow \qquad |x(t)| \leq \kappa |x(t_\circ)| e^{-\lambda(t-t_\circ)} \tag{6.15}$$

   where $\kappa(t_\circ) \to \infty$ as $t_\circ \to \infty$.

3. *The origin is* not *totally stable. Furthermore, for any $\delta > 0$ there exist $(t_\circ, x_\circ)$ and $g(t,x)$ satisfying $\|g(t,x)\| \leq \delta$, such that the trajectories of $\dot{x} = f(t,x) + g(t,x)$ grow* unboundedly.

Proof of Proposition 6.1

Proof of 1. We will prove that

[35] The proposition without proof appears in [31].

$$|f(t,y) - f(t,z)| \le |y - z|, \qquad \forall\, y,\, z \in \mathbb{R}\,. \tag{6.16}$$

For this, without loss of generality, let $z \ge y$ and consider the following three possible cases.

Case 1: Let $y \le -a(t) < 0$, then $f(t,y) = a(t)$ and consider the following:

1. If $z \le -a(t)$ then, $|f(t,y) - f(t,z)| = 0$.
2. If $|z| \le a(t)$ then, $|f(t,y) - f(t,z)| = |a(t) + z| \le |-y+z|$.
3. If $z \ge a(t)$ then, $|f(t,y) - f(t,z)| = |a(t) + a(t)| \le |-y+z|$.

Case 2: Let $|y| \le a(t)$, then, $f(t,y) = -y$. Consider the following possibilities:

1. If $|z| \le a(t)$ then, $|f(t,y) - f(t,z)| = |-y+z|$.
2. If $|z| \ge a(t)$ then, $|f(t,y) - f(t,z)| = |-y+a(t)| \le |-y+z|$.

Case 3: Let $y \ge a(t)$ then, $f(t,y) = -a(t)$ and we consider the following cases:

1. If $z \ge a(t)$ then $|f(t,y) - f(t,z)| = 0$.
2. If $|z| \le a(t)$ then $|f(t,y) - f(t,z)| = a(t) - z = |y - z|$.
3. If $z \le -a(t)$ then $|f(t,y) - f(t,z)| = a(t) + a(t) \le |y - z|$. ▲

Proof of 2. In order to prove (6.15) we consider two cases of initial conditions separately, starting with the case where $|x(t_\circ)| \le a(t_\circ)$.

It is easy to see that for all $t \ge 0$ the function $a(t)$ satisfies the inequality

$$\dot{a}(t) \ge -a(t)$$

and therefore for all $t \ge t_\circ$ we have

$$\begin{aligned}\frac{d(|x(t)| - a(t))}{dt} &= \operatorname{sgn}(x(t)) f(t, x(t)) - \dot{a}(t)\\ &\le -|f(t,x(t))| + a(t) \;\le\; 0\,.\end{aligned}$$

Invoking the comparison theorem we obtain $x(t) \le a(t)$ for all $t \ge t_\circ$. Therefore, in this case we see that the system (6.14) has the form $\dot{x} = -x$ for all $t \ge t_\circ$ and hence

$$|x(t)| \le |x(t_\circ)| e^{-(t-t_\circ)} \qquad \text{for all} \quad t \ge t_\circ. \tag{6.17}$$

Now, let us consider the case $|x(t_\circ)| > a(t_\circ)$. Let $[t_\circ, t_1)$ be a time-interval such that $|x(t)| > a(t)$ for all $t \in [t_\circ, t_1)$. On this time interval we have,[36]

$$\dot{x} = -a(t)\operatorname{sgn}(x) \quad \text{and} \quad \frac{d\,|x(t)|}{dt} = -a(t)$$

[36] Notice that on the interval $(t_\circ, t_1)$, $x(t)$ does not change its sign.

and, therefore

$$|x(t)| = |x(t_\circ)| - \ln\left(\frac{t+1}{t_\circ+1}\right) \tag{6.18}$$

Now, if $t_1 = +\infty$, *i.e.* if (6.18) is valid for all $t \geq t_\circ$ it follows that there exists a time moment t_* such that $x(t_*) = 0$. We conclude that the time instant t_1 such that $|x(t_1)| = a(t_1)$ must be finite and, moreover

$$\frac{1}{t_1+1} = |x(t_\circ)| - \ln\left(\frac{t_1+1}{t_\circ+1}\right)$$

i.e. $t_1 = t_1(t_\circ, x(t_\circ))$. Notice that from the analysis of the first case it follows that $|x(t)| \leq a(t)$ for all $t \geq t_1$, hence $\dot{x} = -x$ for all $t \geq t_1$ and consequently

$$|x(t)| \leq |x(t_1)|e^{-(t-t_1)} \leq |x(t_1)|e^{t_1-t_\circ}e^{-(t-t_\circ)} \tag{6.19}$$

Now let us estimate $t_1 - t_\circ$. From (6.18) we have

$$|x(t_\circ)| - |x(t_1)| = \ln\left(\frac{t_1+1}{t_\circ+1}\right) = \ln\left(1 + \frac{t_1-t_\circ}{t_\circ+1}\right)$$

hence

$$1 + \frac{t_1-t_\circ}{t_\circ+1} = e^{|x(t_\circ)|-|x(t_1)|} \leq e^{|x(t_\circ)|}$$

and

$$t_1 - t_\circ \leq \left(e^{|x(t_\circ)|} - 1\right)(t_\circ+1)$$

therefore for all $|x_\circ|$ and all $t \geq t_1$

$$\begin{aligned}|x(t)| &\leq |x(t_\circ)|\exp\left(\left[e^{|x(t_\circ)|} - 1\right](t_\circ+1)\right)e^{-(t-t_\circ)}\\ &\leq |x(t_\circ)|e^{(e^r-1)(t_\circ+1)}e^{-(t-t_\circ)}\end{aligned}$$

Now, let us estimate $|x(t)|$ for $t \in [t_\circ, t_1)$. From (6.18) we have

$$\begin{aligned}|x(t)| &= |x(t_\circ)| - \ln\left(\frac{t_1+1}{t_\circ+1}\right)\\ &\leq |x(t_\circ)|e^{(t_1-t_\circ)}e^{-(t-t_\circ)}\\ &\leq |x(t_\circ)|\exp\left(\left[e^{|x(t_\circ)|} - 1\right](t_\circ+1)\right)e^{-(t-t_\circ)}\\ &\leq |x(t_\circ)|e^{(e^r-1)(t_\circ+1)}e^{-(t-t_\circ)}\end{aligned} \tag{6.20}$$

Now, combining bounds (6.17), (6.19) and (6.20) we obtain that for all $t_\circ \geq 0$ and each $r > 0$ the following bound is satisfied for all $|x(t_\circ)| \leq r$ and all $t \geq t_\circ$:

$$|x(t)| \leq \kappa|x(t_\circ)|e^{-(t-t_\circ)}$$

where $\kappa = e^{(e^r-1)(t_\circ+1)}$. ▲

Proof of 3. We show that for any (arbitrarily small) number $\delta > 0$, there exists a function $g(t, x)$ satisfying $|g(t, x)| \leq \delta$ for all $t \geq t_\circ$, initial conditions $(t_\circ, x_\circ) \in \mathbb{R}_{\geq 0} \times B_\delta$, such that, the solution $x(t, t_\circ, x_\circ)$ of the differential equation

$$\dot{x} = f(t, x) + g(t, x)$$

where $f(t, x)$ is defined in (6.14), satisfies

$$\lim_{t \to \infty} |x(t, t_\circ, x_\circ)| = +\infty \,. \tag{6.21}$$

Let $g(t, x) = \delta \mathrm{sgn}(x)$ and pick $(t_\circ, x_\circ)$ such that $a(t_\circ) < x_\circ < \delta$. From (6.14), it follows that there exist a time interval $[t_\circ,\, t_\circ + \Delta)$ such that $|x(t)| > a(t)$ for all $t \in [t_\circ,\, t_\circ + \Delta)$. We show next that $\Delta = +\infty$. Over this time interval we have

$$\dot{x} = [-a(t) + \delta]\mathrm{sgn}(x)$$

or equivalently

$$\frac{d|x(t)|}{dt} = -a(t) + \delta \,.$$

Now, since $a(t_\circ) < \delta$, we also have $a(t) < \delta$ for all $t \geq t_\circ$ and therefore, $\frac{d|x(t)|}{dt} \geq \nu$ where $\nu := \delta - a(t_\circ)$. Invoking the comparison principle we obtain that

$$|x(t)| \geq |x(t_\circ)| + \nu(t - t_\circ), \qquad \forall\, t \in [t_\circ,\, t_\circ + \Delta) \,. \tag{6.22}$$

From the inequality above and the continuity of $x(t)$ it follows that $|x(t_\circ + \Delta)| > |x(t_\circ)| > a(t_\circ) > a(t + \Delta)$. Repeating the same reasoning from the initial conditions $t'_\circ := t_\circ + \Delta$ and $x'_\circ := x(t_\circ + \Delta)$ we conclude that $\Delta = +\infty$ and therefore, (6.22) holds for all $t \geq t_\circ$. Consequently,

$$\lim_{t \to \infty} |x(t, t_\circ, x_\circ)| = +\infty \,.$$

■

6.9 Further Bibliographical Remarks

1. As exposed in Section 6.2.1, the definition of stability in the sense of Lagrange and of Dirichlet has also evolved through years up to the, rough but mostly accepted, meaning of "boundedness of solutions". See the books [3, 30, 66] for a modern treatment of Lagrange stability.
2. Lyapunov's original work was published in 1892 in Kharkov Ukraine – *cf.* [32]. His work was reprinted and translated several times and, since we have been unable to find an original copy of his memoir, we rely on the citations that we have found. Many Soviet authors (*e.g.* Oziraner, Rumiantsev, Persidskiĭ, Vinograd, Barbashin, Krasovskiĭ) refer to [35]

as well as other "Gostehiszdat" editions. The French translation [33] is due to E. Davaux and was revised and corrected by Lyapunov himself hence, to some extent, it has the value of the original. This translation was facsimiled and reprinted by Princeton University – *cf.* [34]. Only in 1992, an English translation – based on [34], was made available: [36], which was reprinted and commented by A. T. Fuller in [37]. According to the electronic catalogue of the Math. Library of the Université Paris Sud XI, Orsay, France, the Ukrainian journal *Coobschenya Kharkovskogo Matematicheskogo Obasestvo* (Communications of the Mathematical Society of Kharkov) where [32] was published, became *Zapiski Kharkevskogo Matematicheskogo Tovaristvo to Naukovo*, which is no longer printed since 1983. See also [11] for other historical notes. Finally, some readers might be interested in knowing that, at the moment of writing these notes, reprints of [33, 34] were available at `http://www.gabay.com/` and `http://www.pupress.princeton.edu/` respectively.

3. Many other notions of Lyapunov stability, absent in modern texts such as [20, 65], were introduced in the 1950s and 1960s and stemmed from pure mathematical considerations: in this respect J. L. Massera mentions in [44] that "*in a personal communication, J. Kurzweil has indicated to me the usefulness of introducing still other definitions of asymptotic stability which make it possible to prove interesting necessary and sufficient conditions*"; also, in [43] the author introduced *equiasymptotic* stability, for the case when asymptotic stability (for time-varying systems) is uniform in the initial states but not in the initial times. Thus, there exist a number of definitions of stability that are poorly used nowadays, at least in modern control theory. Further interested readers are invited to see for instance [54, 66, 44, 2] and [15, p. 176].
4. In contemporary and succeeding papers and books (that we do not cite here) improvements to the converse results of Malkin and others were carried out. Notably for the global case and with relaxation of certain regularity conditions on f (see for instance [44] for a compilation of many results by J.-L. Massera, as well as the recent reference [62] for generalisations of converse theorems and literature review on the latter.
5. The late 1940s and 1950s mark an important period in the theory of stability since it was then that this theory seems to have been "westernised", for instance through the book-keeping of S. Lefschetz, J. P. La Salle and W. Hahn, as well as contributions of the latter. Soviet literature became known in Western Europe and in the USA: workshops and collections of papers were entirely devoted to Lyapunov's theory (see *e.g.* as well as [28], Part Six on [26, pp. 117–153]); for reference, we remark that it was in 1947 that the French translation of Lyapunov's work was facsimiled by Princeton University Press. Unfortunately, as we have stressed, other translations of Soviet literature were not always sufficiently accurate: we have in [6], a perfect example of mathematical misinterpretations of fun-

damental concepts and results of stability theory that due to not always accurate translations and (inexact) recursive citations: the title of this paper, which reads in Russian "*On the stability of motion in the whole"*, is cited as *On the stability of motion in the large*" – see for instance [14, 15, 54] and even [22]!

The mistake is not necessarily to be attributed to the multiple authors that have cited the English translation of [6] but it may come from the "official" translation of the Soviet journals by professional translators to whom escape the subtleties of the Soviet mathematical terminology. Interestingly, the ambiguous use of the latter referring to the property of asymptotic stability for *all* initial conditions and for *large* initial conditions also appears in the *Russian* literature: A. A. Krasovskiĭ[37] says in [23, §3.2. Criteria for stability of motion "in the whole"]:

> (Cited and translated from [23, p. 95])
> For nonlinear systems the only known formal analytical method to investigate stability "v bolshom" (for finite and even infinite region of attraction) is that of Lyapunov functions and the first method of Lyapunov. •

The handbook [23] is a multi-author document edited by A. A. Krasovskiĭ and in particular, §3.2. is authored by the latter. We emphasise that the author uses quotation for "*v bolshom*"; suggesting that the term was somewhat *shyly* used (from a linguistic viewpoint).

6. We finish these bibliographical remarks with a *personal* opinion on two classical texts: [56], [14]. The first is simply one of the best-written *textbooks* addressed to young graduate students, on stability of dynamic systems, that we have come across. On the other hand, Hahn's book has also many merits: it is contemporary to the fundamental work of the Soviet protagonists of stability theory and it is written in a formal and precise, yet accessible, language. It exposes the basis of Lyapunov stability theory in, sometimes, a more rigorous and less ambiguous manner than the original articles in Russian. In each of these textbooks, readers will find a significant bibliography that can be qualified as "historical".

   For all these reasons we strongly suggest doctoral students, interested in the fundamentals of stability theory, to study the texts [56, 14] for a "good start". Readers of French language will be delighted by the original book [55].

[37] Not to be confused with N. N. Krasovskiĭ.

6.10 Conclusions

In this chapter we have reviewed basic, "well-known" concepts related to stability theory of dynamical systems; mostly, of Lyapunov stability theory. Even though such theory has been initially developed by Soviet mathematicians, the enrichment that stems from Western literature cannot be overestimated. In view of the geopolitical and economical context of the second half of the last century a very wide number of publications originally written in Russian language, were translated into Western languages, mainly English. Unfortunately, due to not always accurate translations certain terminology was lost or, worse, changed meaning. We hope that this *bona fide* review will contribute to clarify certain concepts that are used widely in the study of dynamical systems, thereby removing *part* of the ambiguity present in the literature.

We stress that in choosing the material here presented we have sacrificed generality for detail of exposition; also, we have limited our own interpretations and emphasised the use of *verbatim* citations taken, in most cases, from direct sources. In this respect, we shall conclude by borrowing Voltaire's words:

> *"Attentive readers, who spread their thoughts among themselves, always go beyond the author"*
>
> —Voltaire, 1763[†].

Acknowledgements

The initial motivation to write this survey came from discussions with R. Kelly who pointed out to the first author, certain inconsistencies regarding stability definitions available in the literature. Our interest was further triggered by personal discussions with A. Teel on the subject. The authors greatly acknowledge this enriching exchange as well as the brainstorming with A. Chaillet, on the contents of these notes. At the moment of proof-editing this textbook, we received, from A. Teel, a friendly review of this chapter which led to several improvements in its exposition.

[†] Original citation in French: *"Des lecteurs attentifs, qui se communiquent leurs pensées, vont toujours plus loin que l'auteur"*, in *Traité sur la tolérence à l'occasion de la mort de Jean Calas*, Voltaire, 1763.

References

Caveat: The references preceded by '*' correspond to those that are cited following other authors and that we did *not* have at the moment of writing these notes hence, we cannot speak of their contents by ourselves. Russian titles are phonetically transcripted from Cyrillic characters and translated: the term "English title" refers to titles in English language, as used in bibliographies of *other* authors; the term "English translation" corresponds to our *own* translation of the corresponding title.

* 1. M. A. Aizerman (1952), *Teoriya avtomaticheskogo regulirovniya dvigatelei.* Gostehiszdat, Moscow. English title: *Theory of automatic control of motors.*
2. H. A. Antosiewicz (1958), *Contributions to the theory of nonlinear oscillations*, volume IV, chapter VIII. A survey on Lyapunov's second method, pages 141–166. Princeton University Press, Princeton. Edited by S. Lefschetz.
3. A. Bacciotti and L. Rosier (2001), *Liapunov Functions and stability theory*, volume 267 of *Lecture notes in control and information sciences.* Springer Verlag.
4. E. A. Barbashin (1967), *Vedenie v teoriyu ustoichivosti.* Izdatelstvo Nauka, Moscow.
5. E. A. Barbashin (1970), *Introduction to the theory of stability.* Wolters-Noordhoff, Groningen, The Netherlands. Translated by Transcripta Service, London, edited by T. Lukes.
6. E. A. Barbashin and N. N. Krasovskiĭ (1952), Ob ustoichivosti dvizheniya v tzelom. *Dokl. Akad. Nauk. USSR*, 86(3):453–456. English title: "On the stability of motion in the large". English translation: "On the stability of motion in the *whole*".
7. E. A. Barbashin and N. N. Krasovskiĭ (1954), O suschestvovanii functsiĭ Lyapunova v sluchae asimptoticheskoĭ ustoĭchivostĭ v tzelom. *Prikl. Mat. i Mekh.*, 18:345–350. English title: "On the existance of Lyapunov fuctions in the case of asymptotic stability in the large". English translation: On the existance of Lyapunov fuctions in the case of asymptotic stability in the whole.
8. N. G. Četaev (1956), *Ustoichivost' dvizheniya.* Gostehiszdat. English title: *Stability of motion.*
* 9. J.-L. de la Grange (1788), *Méchanique Analitique.* Chez la veuve DESAINT, Paris, 1ère edition. Avec appprobation et privilège du Roi, M. DCC. LVIII. (In French).
* 10. G. N. Duboshin (1940), K voprosu ob ustoĭchivoisti dvizheniya otnositelno postoyanno deĭstvuyuschih vozmuscheniĭ. *Trudi GAISh*, Tom XIV(1). In Russian. English translation: On the question of stability of motion with respect to constantly-acting disturbances.
11. A. T. Fuller (1992), Guest editorial to the translation of *The general problem of stability of motion* by A. M. Lyapunov. *Int. J. of Contr.*, 55. No. 3, pp. 521-527.
12. V. D. Furasov (1977), *Ustoichivost' dvijeniya, otzenki i stabilizatsia.* Nauka, Moscow. English translation: Stability of motion, estimates and stabilization.
13. H. Goldstein (1974), *Classical Mecanics.* Adisson Wesley.
14. W. Hahn (1963), *Theory and Application of Liapunov's Direct Method.* Prentice Hall, New York.
15. W. Hahn (1967), *Stability of motion.* Springer-Verlag, New York.

16. J. P. La Salle (1960), Some extensions of Liapunov's second method. *IRE Trans. Circs. Th.*, CT-7(4):520–527.
17. R. Kelly, V. Santibáñez, and A. Loría (2005), *Control of robot manipulators in joint space.* Series *Advanced textbooks in control engineering.* Springer Verlag, London.
18. H. Khalil (1992), *Nonlinear systems.* Macmillan Publishing Co., 1st ed., New York.
19. H. Khalil (1996), *Nonlinear systems.* Macmillan Publishing Co., 2nd ed., New York.
20. H. Khalil (2002), *Nonlinear systems.* Prentice Hall, 3rd ed., New York.
21. N. N. Krasovskiĭ (1954), O povedenii v tzelom integralih krivih systemi dvuh differentzialnih uravneniĭ. *Prikl. Mat. i Mekh.*, 18:149–154. English title: On the behaviour in the large of integral curves of a system of two differential eqations. English translation: On the behaviour in the *whole* of integral curves of a system of two differential eqations.
22. N. N. Krasovskii (1963), *Problems of the theory of stability of motion.* Stanford Univ. Press. Translation of Russian edition, Moscow 1959.
23. A. A. Krasovskiĭ, ed. (1987), *Spravochnik po teorii avtomaticheskogo upravleniya.* Nauka, Moscow. Translated title: Handbook on the theory of automatic control.
24. M. Krstić, I. Kanellakopoulos, and P. Kokotović (1995), *Nonlinear and Adaptive control design.* John Wiley & Sons, Inc., New York.
25. J. P. La Salle and S. Lefschetz (1961), *Stability by Lyapunov's direct method.* Academic Press, New York.
26. J. P. La Salle and S. Lefschetz, editors (1962), *Recent soviet contributions to mathematics.* The Macmillan Company, New York.
27. J.-L. Lagrange (1888), *Mécanique Analytique.* Gauthier-Villars et fils, Paris, 4ème edition. Publiée par Gaston Darboux. Includes the papers by Dirichlet and Poisson.
28. S. Lefschetz, editor (1958), *Contributions to the theory of nonlinear oscillations*, volume 1-V. Princeton University Press, Princeton.
29. G. Lejeune-Dirichlet (1888), *Sur la stabilité de l'équlibre*, chapter in Méchanique Analytique de J.-L. Lagrange –Note II., p. 457 du Tome Premier – *cf.* [27]. Gauthier-Villars et fils.
30. G. A. Leonov, D. V. Ponomarenko, and V. B. Smirnova (1996), *Frequency domain methods for nonlinear analysis: theory and applications.* World Scientific.
31. A. Loría R. Kelly, and A. Teel (2005), Discussions on "Uniform parametric convergence in the adaptive control of mechanical systems". *European J. of Contr.*, 11(2).

* 32. A. M. Lyapunov (1892), Obschaya zadacha ob ustoichivosti dvizhenya. *Coobschenya Kharkovskogo Matematicheskogo Obasestva.* Journal title sometimes cited as: Comm. Math. Soc. Kharkov.

* 33. A. M. Liapounoff (1907), Problème de la stabilité de mouvement. *Annales de la faculté de sciences de Toulouse*, 9:203–474. French translation of [32] made by E. Davaux and revised by the author.

34. A. M. Liapounoff (1947), Facsimile reprint of [33]. Published in Ann. Math. Studies, vol. AM-17, Princeton University Press.

* 35. A. M. Lyapunov (1950), *Obschaya zadacha ob ustoichivosti dvizhenya.* Gostehiszdat, Moscow-Leningrad. Reprint of [32].

36. A. M. Lyapunov (1992), The general problem of stability of motion. *Int. J. of Contr.*, 55. No. 3. English translation of [34].

* 37. A. M. Lyapunov (1996), *The general problem of stability of motion.* Academic Press, New York. Reprint of [36], edited and commented by T. Fuller.

38. I. Malkin (1944), Ob ustoĭchivosti pri postoyanno deĭstvuyuschih vozmyscheniyah. *Prikl. Mat. i Mekh.*, Tom. VIII:241–245. English translation: *On stability under constantly acting disturbances.*

39. I. G. Malkin (1954), K voprosu ob obratimosti teoremi Liapunova ob asimptoticheskoi ustoĭchivosti. *Prikl. Mat. i Mekh.*, 18:129–138. English title: *On the reciprocal of Lyapunov's theorem on asymptotic stability.*

40. I. G. Malkin (1966), *Teoriya ustoichivosti dvizheniya.* Izdatel'vo Nauka, Moskva – Leningrad, 2nd edition. English title: see below.

41. I. G. Malkin (Unspecified year), *Theory of Stability of Motion*, volume AEC-tr-3352 of *Translation Series, Physics and Mathematics.* United States Atomic Energy Commission, Technical Information Service Extension, Oak Ridge, Tennessee. Translated from a publication of the state publishing house of technical theoretical literature, Moscow – Leningrad, 1952. Translated by the Language Service Bureau, Washinton, D. C., under contract AT(40-1)-2274.

42. R. Marino and P. Tomei (1995), *Nonlinear control design.* Prentice Hall, UK.

43. J. L. Massera (1949), On Lyapounoff's conditions of stability. *Annals of Mathematics*, 50:705–721.

44. J. L. Massera (1956), Contributions to stability theory. *Annals of Mathematics*, 64(1):182–206.

* 45. J. L. Massera and J. J. Shäffer (1958), Linear differential equations and functional analysis. *I. Ann. of Math.*, 57(3).

46. A. P. Morgan and K. S. Narendra (1977), On the uniform asymptotic stability of certain linear nonautonomous differential equations. *SIAM J. on Contr. and Opt.*, 15(1):5–24.

47. E. Panteley and A. Teel. Notes from private collaboration between the authors, 1998. Santa Barbara, CA.

* 48. O. Perron (1930), Die stabilitätsfrage bei Differentialgleischungen. *Math. Zeit.*, 32.

49. K. P. Persidskiĭ (1933), Ob ustoichivosti dvizhenya po pervomu priblizheniyu. *Mat. Sbornik*, 40(3):284–293. English title: On the stability of motion in the first aproximation.

50. K. P. Persidskiĭ (1937), Ob odnoi teoreme Lyapunova. *Dokl. Akad. Nauk. USSR*, XIV(9):541–543. English title: On a theorem by Lyapunov. Title also used: Über einen satz von Liapounoff.

51. K. P. Persidskiĭ (1946), K teorii ustoichivosti reshenii differentzialnih uravnenii. *Uspehi Matematicheskih Nauk*, 1:250–255. Part 5: Doktorskie dissertatsii. English translation: On the theory of stability of solutions of differential equations –summary of doctoral (Sc.) dissertation.

52. H. Poincaré (1890), Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:5–266.

* 53. S. D. Poisson (1838), *Traité de Mécanique.* Société belge de Librairie, Bruxelles.

54. N. Rouche, P. Habets, and M. Laloy (1977), *Stability theory by Liapunov's direct method*, volume 22 of *Appl. Math. Sc.* Springer-Verlag, New York.

55. N. Rouche and J. Mawhin (1973), *Equations différentielles ordinaires, Tome 2: Stabilité et solutions périodiques* . Masson et C$^{\text{ie}}$, Paris.

56. N. Rouche and J. Mawhin (1980), *Ordinary differential equations II: Stability and periodical solutions.* Pitman publishing Ltd., London.
57. N. Rouche and K. Peiffer (1967), Le théorème de Lagrange-Dirichlet et la deuxème méthode de Lyapunoff. *Annales de la Société Scientifique de Bruxelles*, 81(I):19–33.
58. S. Sastry (1999), *Nonlinear systems – Analysis stability and control*, volume 10 of *Interdisciplinary applied mathematics.* Springer, New York.
59. J.J. Slotine and W. Li (1991), *Nonlinear Control Analysis.* Prentice Hall.
60. E. Sontag (1989), Smooth stabilization implies coprime factorization. *IEEE Trans. on Automat. Contr.*, 34(4):435–443.
61. E. D. Sontag and M. Krichman (2003), An example of a GAS system which can be destabilized by an integrable perturbation. *IEEE Trans. on Automat. Contr.*, 48(6):1046–1049.
62. A. Teel and L. Praly (2000), A smooth Lyapunov function from a class-KL estimate involving two positive semi-definite functions. *ESAIM: COCV*, 5:313–368.
63. A. R. Teel and J. Hespanha (2004), Examples of GES systems that can be driven to infinity by arbitrarily small additive decaying exponentials. *IEEE Trans. on Automat. Contr.*, 40(3):1407–1410.
64. M. Vidyasagar (1978), *Nonlinear systems analysis.* Prentice Hall, New Jersey.
65. M. Vidyasagar (1993), *Nonlinear systems analysis.* Prentice Hall, New Jersey.
66. T. Yoshizawa (1966), *Stability theory by Lyapunov's second method.* The Mathematical Society of Japan, Tokyo.
67. G. Zames (1966), On the Input–Output stability of time varying nonlinear feedback systems, Part I. *IEEE Trans. on Automat. Contr.*, 11:228–238.
68. G. Zames (1966), On the Input–Output stability of time varying nonlinear feedback systems, Part II. *IEEE Trans. on Automat. Contr.*, 11:465–476.

7

Structural Properties of Linear Systems – Part II: Structure at Infinity

Henri Bourlès

SATIE, Ecole Normale Supérieure de Cachan and Conservatoire National des Arts et Métiers, 61 avenue du Président Wilson, 94230 Cachan, France. E-mail: Henri.Bourles@satie.ens-cachan.fr

7.1 Introduction

This chapter is the sequel of [7]. Its topic is the structure at infinity of discrete and continuous linear time-varying systems in a unified approach.

In the time-invariant case, the linear systems in [7] are implicitly assumed to be perpetually existing and the smoothness of their behavior is not studied. In practice, however, that behavior must be sufficiently smooth (to avoid undesirable saturations of the variables, or even the destruction of the system), and the system has a limited useful life. These constraints can be taken into account by studying the *structure at infinity* of the system under consideration. As this system is existing during a limited period, it is called a *temporal system* [8].

A list of *errata* and *addenda* for [7] is given at the end of the chapter.

7.2 Differential Polynomials and Non-commutative Formal Series

7.2.1 Differential Polynomials: A Short Review

The notation is the same as in [7], except for the derivation which is now denoted by γ (instead of δ, to avoid confusing with the Dirac distribution). In what follows, the "coefficient field" $\mathbf{K}$ is a differential field, equipped with an α-derivation γ such that α is an automorphism of $\mathbf{K}$, $a^{\gamma} = 0$ implies $a^{\alpha} = a$ (for any $a \in \mathbf{K}$) and $\alpha\,\delta = \delta\,\alpha$. The subfield of constants of $\mathbf{K}$ is denoted by $\mathbf{k}$. The ring of differential polynomials $\mathbf{K}\,[\partial; \alpha, \gamma]$ is denoted by $\mathbf{R}$. This ring is equipped with the commutation rule

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 259–284, 2006.

$$\partial a = a^{\alpha} \partial + a^{\gamma}. \tag{7.1}$$

As shown in ([7], Section 6.2.3, Proposition 6.4), $\mathbf{R}$ is an euclidean domain, thus it is a two-sided Ore domain, and it has a field of left fractions and a field of right fractions which coincide ([7], Section 6.2.3); this field is denoted by $\mathbf{Q} = \mathbf{K}(\partial; \alpha, \gamma)$.

7.2.2 Local Complete Rings

Let $\mathbf{A} \neq 0$ be a ring. The set of all left ideals of $\mathbf{A}$ is inductively ordered by inclusion, thus (by Zorn's lemma), any left ideal of $\mathbf{A}$ is included in a maximal left ideal ("Krull's theorem"). The intersection of all maximal left ideals of $\mathbf{A}$ is called the *Jacobson radical* of $\mathbf{A}$; it turns out that this notion is left/right symmetric. A left ideal $\mathbf{a}$ of $\mathbf{A}$ is maximal if, and only if $\mathbf{A}/\mathbf{a}$ is a field ([1], Section I.9.1)[1]. The following result is proved in ([22], Sections 4 and 19):

Proposition 7.1. *(i) If* $\mathbf{A}$ *has a* unique *maximal left ideal* $\mathbf{m}$*, then* $\mathbf{m}$ *is a two-sided ideal, it is also the unique maximal right ideal of* $\mathbf{A}$ *and it consists of all noninvertible elements of* $\mathbf{A}$*. (ii) Conversely, if the set of all noninvertible elements of* $\mathbf{A}$ *is an additive group* $\mathbf{m}$*, then* $\mathbf{m}$ *is the unique maximal left ideal of* $\mathbf{A}$ *(and is the Jacobson radical of that ring).*

Definition 7.1. *A ring* $\mathbf{A}$ *which has a unique maximal left ideal* $\mathbf{m}$ *is called a* local ring *(and is also denoted by* $(\mathbf{A}, \mathbf{m})$*, to emphasize the role of* $\mathbf{m}$*);* $\mathbf{A}/\mathbf{m}$ *is called the* residue field *of* $\mathbf{A}$*.*

Let $(\mathbf{A}, \mathbf{m})$ be a local ring. The set $\{\mathbf{m}^i, i \geq 0\}$ is a filter base, and $\mathbf{A}$ is a topological ring with $\{\mathbf{m}^i, i \geq 0\}$ as a neighborhood base of 0 ([3], Section III.6.3). This topology of $\mathbf{A}$ is called the $\mathbf{m}$-adic topology. Assuming that

$$\bigcap_{i \geq 0} \mathbf{m}^i = 0, \tag{7.2}$$

the $\mathbf{m}$-adic topology is Hausdorff, and it is metrizable since the basis $\{\mathbf{m}^i, i \geq 0\}$ is countable ([4], Section IX.2.4).

Definition 7.2. *The local ring* $(\mathbf{A}, \mathbf{m})$ *is said to be* complete *if it is complete in the* $\mathbf{m}$-adic topology*.*

7.2.3 Formal Power Series

Set $\sigma = 1/\partial$ (σ can be viewed as the "integration operator": see Section 7.4.2) and $\beta = \alpha^{-1}$; $\mathbf{S} := \mathbf{K}[[\sigma; \beta, \gamma]]$ denotes the ring of *formal power series in* σ, equipped with the commutation rule ([12], Section 8.7)

[1] In this chapter, as in [7], *field* means *skew field*.

$$\sigma a = a^{\beta}\sigma - \sigma a^{\beta\gamma}\sigma, \tag{7.3}$$

deduced from (7.1). An element a of $\mathbf{S}$ is of the form

$$a = \sum_{i\geq 0} a_i\,\sigma^i, \quad a_i \in \mathbf{K}.$$

Definition 7.3. *Let a be a nonzero element of $\mathbf{S}$ and set $\omega(a) = \min\{i : a_i \neq 0\}$; the natural integer $\omega(a)$ is called the* order *of a.*

Proposition 7.2. *(i) The ring $\mathbf{S}$ is a principal ideal domain and is local with maximal left ideal $\mathbf{S}\,\sigma = \sigma\,\mathbf{S} = (\sigma)$. (ii) The units of $\mathbf{S}$ are the power series of order zero; any nonzero element $a \in \mathbf{S}$ can be written in the form $a = \upsilon\,\sigma^{\omega(a)} = \sigma^{\omega(a)}\,\upsilon'$, where υ and υ' are units of $\mathbf{S}$. (iii) Let a and b be nonzero elements of $\mathbf{S}$; then $b \parallel a$ (i.e. b is a* total divisor *of a) if, and only if $\omega(b) \leq \omega(a)$; every nonzero element of $\mathbf{S}$ is invariant (see [7], Section 6.2.3). (iv) The local ring $(\mathbf{S}, (\sigma))$ is complete.*

Proof. 1) It is easy to check that the only nonzero elements of $\mathbf{S}$ are the powers $\sigma^i, i \geq 0$, and their associates, thus (i), (ii) and (iii) are obvious; note that the residue field of $\mathbf{S}$ is $\mathbf{S}/(\sigma) \cong \mathbf{K}$. 2) Condition (7.2) is satisfied with $\mathbf{m} = (\sigma)$ (see Exercise 7.1). Let (a_n) be a Cauchy sequence of $\mathbf{S}$, $a_n = \sum_{i\geq 0} a_{n,i}\,\sigma^i$. For any integer $k \geq 1$, there exists a natural integer N such that for all $n, m \geq N$, $a_n - a_m \in \left(\sigma^k\right)$, *i.e.* for any integer i such that $0 \leq i \leq k-1$ we have $a_{n,i} = a_{m,i}$. Let b_i be the latter quantity and $b = \sum_{i\geq 0} b_i\,\sigma^i \in \mathbf{S}$. The sequence (a_n) converges to b in the (σ)-adic topology. ■

Exercise 7.1. Prove that for any $i \geq 0$, $\left(\sigma^i\right) = (\sigma)^i$ and that $\bigcap_{i\geq 0}\left(\sigma^i\right) = 0$.

7.2.4 A Canonical Cogenerator

For any $\mu \in \mathbb{N}$ (where $\mathbb{N}$ is the set of natural integers), set $\tilde{C}_\mu = \frac{\mathbf{S}}{(\sigma^\mu)}$ and let $\tilde{\delta}^{\mu-1}$ be the canonical image of $1 \in \mathbf{S}$ in $\tilde{C}_\mu$; $\tilde{C}_\mu$ is isomorphic to a submodule of $\tilde{C}_{\mu+1}$ under right multiplication by σ and $\tilde{\delta}^{(\mu)}\sigma = \sigma + \left(\sigma^{\mu+1}\right) = \sigma\tilde{\delta}^{(\mu)}$; identifying $\tilde{\delta}^{(\mu-1)}$ with $\sigma\tilde{\delta}^{(\mu)}$, $\tilde{C}_\mu$ is embedded in $\tilde{C}_{\mu+1}$, and

$$\tilde{C}_\mu = \oplus_{i=1}^{\mu}\mathbf{K}\tilde{\delta}^{(i-1)}. \tag{7.4}$$

Set

$$\tilde{\Delta} := \varinjlim_{\mu}\tilde{C}_\mu = \oplus_{\mu\geq 0}\mathbf{K}\tilde{\delta}^{(\mu)}. \tag{7.5}$$

The left $\mathbf{S}$-module $\tilde{\Delta}$ becomes a left $\mathbf{L}$-vector space, setting $\sigma^{-1}\tilde{\delta}^{(\mu)} = \tilde{\delta}^{(\mu+1)}$, thus $\tilde{\Delta}$ becomes also a left $\mathbf{R}$-module by restriction of the ring of scalars. Considering σ and ∂ as operators on $\tilde{\Delta}$, σ is a left inverse of ∂, but σ has no left inverse since $\sigma\tilde{\delta} = 0$.

Exercise 7.2. (i) Prove that $\tilde{\Delta} = E\left(\tilde{C}_{\mu}\right)$ for any $\mu \geq 1$ (where $E\left(.\right)$ is the injective hull of the module in parentheses). (ii) Prove that $\tilde{C}_1$ is the only simple $\mathbf{S}$-module. (iii) Prove that $\tilde{\Delta}$ is the *canonical cogenerator* of ${}_{\mathbf{S}}\mathbf{Mod}$. (Hint: for (i), first show that $\tilde{\Delta}$ is divisible and then proceed as in the proof of ([7], Section 6.5.1, Proposition 6.26, parts (3) and (4)). For (ii), use ([7], Section 6.3.1, Excercise 6.8(ii)). For (iii), use ([7], Section 6.5.1, Theorem 6.9(i)).)

Remark 7.1. Let $\mathbf{E}$ be the endomorphism ring of $\tilde{\Delta}$ ([7], Section 6.5.1). If $\mathbf{S}$ is commutative (*i.e.* $\mathbf{K} = \mathbf{k}$), then $\mathbf{E} \cong \mathbf{S}$, since $\mathbf{S}$ is complete ([23], Section 3.H), thus these two rings can be identified (this is the main result of the "Matlis theory").

7.2.5 Matrices over S

Unimodular Matrices

Let us study the general linear group $\mathbf{GL}_n\left(\mathbf{S}\right)$, *i.e.* the set of unimodular matrices belonging to $\mathbf{S}^{n\times n}$ [11]:

Proposition 7.3. *(i) Let* $U = \sum_{i\geq 0} \Upsilon_i\,\sigma^i \in \mathbf{S}^{n\times n}$*, where* $\Upsilon_i \in \mathbf{K}^{n\times n}, i \geq 0$. *The matrix* U *belongs to* $\mathbf{GL}_n\left(\mathbf{S}\right)$ *if, and only if* Υ_0 *is invertible, i.e.* $|\Upsilon_0| \neq 0$. *(ii) Let* $U \in \mathbf{GL}_n\left(\mathbf{S}\right)$ *and* $k \in \mathbb{N}$. *There exist two matrices* $U_k, U'_k \in \mathbf{GL}_n\left(\mathbf{S}\right)$ *such that* $\sigma^k\,U = U_k\,\sigma^k$ *and* $U\,\sigma^k = \sigma^k\,U'_k$.

Proof. (i) If Υ_0 is invertible, U can be written in the form $\Upsilon_0\left(I_n - X\right), X \in \sigma\,\mathbf{S}^{n\times n}$. The matrix $I_n - X$ is invertible with inverse $\sum_{i\geq 0} X^i$. Conversely, if U is invertible, there exists $L = \sum_{i\geq 0} \Lambda_i\,\sigma^i \in \mathbf{S}^{n\times n}$ such that $U\,L = I_n$. This implies $\Upsilon_0\,\Lambda_0 = I_n$, thus Υ_0 is invertible. (ii) Let $U = \sum_{i\geq 0} \Upsilon_i\,\sigma^i \in \mathbf{GL}_n\left(\mathbf{S}\right)$. By (7.3), $\sigma\,U = \left(\sum_{i\geq 0} \Theta_i\,\sigma^i\right)\sigma$ with $\Theta_0 = \Upsilon_0^{\beta}$. The matrix Υ_0^{β} (whose entries are the images of the entries of Υ_0 by the automorphism β) is invertible, therefore $\sum_{i\geq 0} \Theta_i\,\sigma^i$ is unimodular by (i). Finally, (ii) is obtained by induction. ■

Smith Canonical Form over S.

Let $B^{+} \in \mathbf{S}^{q\times k}$. By Proposition 7.2 and ([7], Section 6.3.3, Theorem 6.4), there exist matrices $U \in \mathbf{GL}_q(\mathbf{S})$ and $V \in \mathbf{GL}_k(\mathbf{S})$ such that $U\,B^{+}\,V^{-1} = \Sigma$ where

$$\Sigma = \operatorname{diag}\left(\sigma^{\mu_1}, ..., \sigma^{\mu_r}, 0, ..., 0\right), \quad 0 \leq \mu_1 \leq ... \leq \mu_r. \tag{7.6}$$

Let $\mu_i, 1 \leq i \leq s$ be the zero elements in the list $\{\mu_i, 1 \leq i \leq r\}$ (if any). The following proposition is obvious:

Proposition 7.4. *Σ is the Smith canonical form of B^{+} over* $\mathbf{S}$. *The noninvertible invariant factors σ^{μ_i} $(s+1 \leq i \leq r)$ of B^{+} coincide with its elementary divisors.*

Exercise 7.3. Prove that the Smith canonical form of a matrix $B^{+} \in \mathbf{S}^{q\times k}$ can be obtained using only elementary operations (*i.e.* secondary operations are not necessary).

Exercise 7.4. *Dieudonné determinant over* $\mathbf{S}$. Let $\mathbf{F}$ be a skew field. The "Dieudonné determinant" $|.|$ of a square matrix over $\mathbf{F}$ has the following properties [13]: (a) $|A| = 0$ if, and only if A is singular; (b) if $|A| \neq 0$, $|A| \in \mathbf{F}^{\times}/\mathbf{C}(\mathbf{F}^{\times})$, where $\mathbf{F}^{\times}$ is the multiplicative group consisting of all nonzero elements of $\mathbf{F}$ and $\mathbf{C}(\mathbf{F}^{\times})$ is the *derived group* of $\mathbf{F}^{\times}$, *i.e.* the subgroup generated by all elements $x^{-1}y^{-1}xy$, $x, y \in \mathbf{F}^{\times}$ ([1], Section I.6.2); (c) for any $\lambda \in \mathbf{F}^{\times}$, $|\lambda|$ is the canonical image of λ in $\mathbf{F}^{\times}/\mathbf{C}(\mathbf{F}^{\times})$; (d) $|.|$ is multiplicative, *i.e.* $|A\,B| = |A|\,|B|$; (e) if X is square and nonsingular, $\left|\begin{bmatrix} X & Y \\ Z & T \end{bmatrix}\right| = |X|\,\left|T - Z\,X^{-1}\,Y\right|$. (i) Let $\mathbf{U}$ be the multiplicative group consisting of all units of $\mathbf{S}$; prove that $\mathbf{C}(\mathbf{L}^{\times}) \subset \mathbf{U}$. (ii) Set $\mathbf{1} = \mathbf{U}/\mathbf{C}(\mathbf{L}^{\times})$ and let U be an elementary matrix ([7], Section 6.3.3); show that $|U| \in \mathbf{1}$. (iii) Prove that for any matrix $A \in \mathbf{GL}_n(\mathbf{S})$, $|A| \in \mathbf{1}$. (iv) For any nonsingular matrix $A \in \mathbf{S}^{n\times n}$, prove that there exists $|v| \in \mathbf{1}$ such that $|A| = |\sigma^{\mu}|\,|v|$, where $\mu = \sum_{s+1 \leq i \leq n} \mu_i$ and the σ^{μ_i} $(s+1 \leq i \leq n)$ are the elementary divisors of A. (Hint: for (iii) and (iv), reduce A to its Smith canonical form, and for (iii) show that A is a product of elementary matrices using the result to be proved in Exercise 7.3.)

Canonical Decomposition of an S-Module

Let M^{+} be a finitely generated (f.g.) $\mathbf{S}$-module. The following theorem is an immediate consequence of (7.6) (see [7], Section 6.3.3, Theorem 6.5).

Theorem 7.1. *(i) The following relations hold:*

$$(a) \quad M^{+} = \mathcal{T}\left(M^{+}\right) \oplus \Phi^{+}, \quad (b) \quad \mathcal{T}\left(M^{+}\right) \cong \bigoplus_{s+1 \leq i \leq r} \frac{\mathbf{S}}{\left(\sigma^{\mu_i}\right)}$$

$$(c) \quad \Phi^{+} \cong M^{+}/\mathcal{T}\left(M^{+}\right)$$

where $\mathcal{T}\left(M^{+}\right)$ is the torsion submodule of M^{+}, the module Φ^{+} is free, and $1 \leq \mu_{s+1} \leq ... \leq \mu_r$; the elements σ^{μ_i} $(s+1 \leq i \leq r)$ are uniquely determined from M^{+}. (ii) The module M^{+} can be presented by a short exact sequence

$$0 \longrightarrow \mathbf{S}^r \xrightarrow{\bullet B^{+}} \mathbf{S}^k \longrightarrow M^{+} \longrightarrow 0. \tag{7.7}$$

The terminology in the first part of the definition below is taken from ([2], Section VII.4.8).

Definition 7.4. *(i) The elements σ^{μ_i} $(s+1 \leq i \leq r)$ –or the ideals generated by them– are the* nonzero invariant factors *of M^{+}, and they coincide with its* nonzero elementary divisors*; the number of times a same element σ^{μ_i} is encountered in the list $\{\sigma^{\mu_i}, s+1 \leq i \leq r\}$ is the* multiplicity of that elementary divisor*; $rk\Phi^{+} = k - r$ is the* multiplicity of the elementary divisor 0*. (ii) The integer $\#\left(M^{+}\right) = \sum_{s+1 \leq i \leq r} \mu_i$ is called the* degree *of M^{+}.*

7.2.6 Formal Laurent Series

The Quotient Field L

The quotient field of $\mathbf{S}$ is $\mathbf{L} = \mathbf{K}\left((\sigma;\beta,\gamma)\right)$, the field of *formal Laurent series in* σ, equipped with the commutation rule (7.3). An element a of $\mathbf{L}$ is of the form

$$a = \sum_{i \geq \nu} a_i\, \sigma^i, \quad a_i \in \mathbf{K}, \quad a_\nu \neq 0,$$

where ν belongs to the ring $\mathbb{Z}$ of integers.

The rings $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{S}$ can be embedded in $\mathbf{L} = \mathbf{K}\left((\sigma;\beta,\gamma)\right)$; all these rings are integral domains, and are noncommutative except if $\mathbf{K} = \mathbf{k}$.

Smith-MacMillan Canonical Form over L

Let $G \in \mathbf{L}^{p\times m}$ be a matrix of rank r.

Theorem 7.2. *There exist matrices $W \in \mathbf{GL}_p(\mathbf{S})$ and $V \in \mathbf{GL}_m(\mathbf{S})$ such that*

$$W\, G\, V^{-1} = diag\left(\sigma^{\nu_1}, ..., \sigma^{\nu_r}, 0, ..., 0\right), \quad \nu_1 \leq ... \leq \nu_r, \tag{7.8}$$

and the integers $\nu_i \in \mathbb{Z}$ $(1 \leq i \leq r)$ are uniquely determined from G.

Proof. Let σ^k be a least common denominator of all entries of G and $A^+ = \sigma^k\, G \in \mathbf{S}^{p \times m}$. According to Proposition 7.4, there exist matrices $U \in \mathbf{GL}_p(\mathbf{S})$ and $V \in \mathbf{GL}_m(\mathbf{S})$ such that $U\, A^+\, V^{-1} = \Sigma$, where Σ is given by (7.6). Therefore, $U\, \sigma^k\, G\, V^{-1} = \Sigma$, and by Proposition 7.3, $U\, \sigma^k = \sigma^k\, W$ where $W = U_k' \in \mathbf{GL}_p(\mathbf{S})$. Thus, the equality in (7.8) holds with $\nu_i = \mu_i - k$ $(1 \le i \le r)$. ■

Definition 7.5. *The matrix in the right-hand side of the equality in* (7.8) *is called the* Smith-MacMillan canonical form *of* $G = G(\partial)$ *over* $\mathbf{L}$.

7.3 Transmission Poles and Zeros at Infinity

7.3.1 Transfer Matrix of an Input-Output System

Let M be an input-output system with input u and output y ([7], Section 6.4.1); M is a f.g. left $\mathbf{R}$-module, the input $u = (u_i)_{1 \le i \le m}$ (such that the module $M/[u]_{\mathbf{R}}$ is torsion) is assumed to be *independent*, and $y = (y_i)_{1 \le i \le p}$.

Since $\mathbf{R}$ is a two-sided Ore domain, the functor $\mathbf{Q} \bigotimes_{\mathbf{R}} -$ is well-defined, it is covariant from the category of left $\mathbf{R}$-modules to the category of left $\mathbf{Q}$-vector spaces, and it is exact (*i.e.* $\mathbf{Q}$ is a flat $\mathbf{R}$-module: see [12], Sections 0.9 and Appendix 2). Let $\hat{M} = \mathbf{Q} \bigotimes_{\mathbf{R}} M$ and $\hat{\varphi} : M \to \hat{M}$ be the canonical map defined by $\hat{\varphi}(w) = \hat{w} = 1_{\mathbf{Q}} \otimes_{\mathbf{R}} w$, where $1_{\mathbf{Q}}$ is the unit-element of $\mathbf{Q}$; then $\ker \hat{\varphi} = \mathcal{T}(M)$ ([7], Section 6.3.1, Excercise 6.6).

Definition 7.6. *[15]* $\mathbf{Q} \bigotimes_{\mathbf{R}} -$ *is called the* Laplace functor*.*

The following theorem is due to Fliess ([14], [15]).

Theorem 7.3. *(i) $\hat{u}$ is a basis of $\hat{M}$. (ii) There exists a unique matrix $G(\partial) \in \mathbf{Q}^{p \times m}$ such that $\hat{y} = G(\partial)\, \hat{u}$.*

Proof. There exists a short exact sequence

$$0 \to [u]_{\mathbf{R}} \to M \to M/[u]_{\mathbf{R}} \to 0;$$

by exactness of the functor $\mathbf{Q} \bigotimes_{\mathbf{R}} -$, this yields the short exact sequence

$$0 \to [\hat{u}]_{\mathbf{Q}} \to \hat{M} \to 0 \to 0$$

since the module $M/[u]_{\mathbf{R}}$ is torsion; therefore, $\hat{M} = [\hat{u}]_{\mathbf{Q}}$. In addition, $\dim [\hat{u}]_{\mathbf{Q}} = \operatorname{rk}\, [u]_{\mathbf{R}} = m$ ([7], Section 6.4.1), thus $\hat{u}$ is a basis of $\hat{M}$. (ii) is an obvious consequence of (i). ■

Definition 7.7. *[15] The matrix $G(\partial) \in \mathbf{Q}^{p \times m}$ is called the* transfer matrix *of the input-output system.*

7.3.2 Structure at Infinity of a Transfer Matrix

Embedding $\mathbf{Q}$ into $\mathbf{L}$, a transfer matrix $G(\partial) \in \mathbf{Q}^{p\times m}$ can be considered as an element of $\mathbf{L}^{p\times m}$, thus its Smith-MacMillan canonical form over $\mathbf{L}$ can be determined. The following definition, taken from ([11], [8]), generalizes notions which are classical in the context of time-invariant linear systems ([19], [28], [27]).

Definition 7.8. *(i) The Smith-MacMillan canonical form* (7.8) *of* $G(\partial)$ *over* $\mathbf{L}$ *is called its* Smith-MacMillan canonical form at infinity. *(ii) Define the finite sequences* $(\bar{\varsigma}_i)_{1\leq i\leq r}$ *and* $(\bar{\pi}_i)_{1\leq i\leq r}$ *as:* $\bar{\varsigma}_i = \max(0,\nu_i)$ *and* $\bar{\pi}_i = \max(0,-\nu_i)$. *Among the natural integers* $\bar{\varsigma}_i$ *(resp.* $\bar{\pi}_i$*), those which are nonzero (if any) are called the* structural indexes *of the* zeros at infinity *(resp. of the* poles at infinity*) of the matrix* $G(\partial)$*; they are put in increasing (resp. decreasing) order and denoted by* $\varsigma_i, 1\leq i\leq \rho$ *(resp.* $\pi_i, 1\leq i\leq s$*). (iii) If* $\rho\geq 1$ *(resp.* $s\geq 1$*),* $G(\partial)$ *is said to have* ρ *zeros (resp.* s *poles) at infinity, the* i*-th one of* order ς_i *(resp.* π_i*). (iii) If* $\nu_1 > 0$, $G(\partial)$ *is said to have a* blocking zero at infinity of order ν_1. *The natural integer* $\#(TP_\infty) = \sum_{1\leq i\leq s}\pi_i$ *(resp.* $\#(TZ_\infty) = \sum_{1\leq i\leq \rho}\varsigma_i$*) is called the* degree *of the poles (resp. the zeros) at infinity of* $G(\partial)$.

Definition 7.9. *The poles (resp. the zeros) at infinity of* $G(\partial)$ *are called the* transmission poles *(resp. the* transmission zeros*)* at infinity *of the input-output system with transfer matrix* $G(\partial)$. *See also Exercise 7.17, Section 7.5.6.*

The matrix $G(\partial)$ can be expanded as:

$$G(\partial) = \sum_{i\geq \nu_1}\Theta_i\,\sigma^i, \quad \Theta_{\nu_1}\neq 0.$$

Definition 7.10. *The transfer matrix* $G(\partial)$ *is said to be* proper *(resp.* strictly proper*) if* $\nu_1\geq 0$ *(resp.* $\nu_1\geq 1$*). It is said to be* biproper *if it is invertible, proper and with a proper inverse [16].*

The following notion, introduced in [30] in the time-invariant case, was generalized in [25] to time-varying systems.

Definition 7.11. *The integer* $c_\infty(G) = \#(TP_\infty) - \#(TZ_\infty)$ *is called the* content at infinity *of* $G(\partial)$.

Note that $c_\infty(G) = -\sum_{1\leq i\leq r}\nu_i$, where the integers $\nu_i\in\mathbb{Z}$ are defined according to (7.8).

Exercise 7.5. Let $G \in \mathbf{Q}^{p \times m}$ and denote by g_{ij} the entries of G $(1 \leq i \leq p, 1 \leq j \leq m)$. Write $g_{ij} = a_{ij}^{-1} b_{ij}$, $0 \neq a_{ij} \in \mathbf{R}$, $b_{ij} \in \mathbf{R}$. Let $a \in \mathbf{R}$ be an l.c.l.m. of all elements a_{ij} (*i.e.* a least common left denominator of all elements g_{ij}); this means that

$$\mathbf{R}\, a = \bigcap_{\substack{1 \leq i \leq p \\ 1 \leq j \leq m}} \mathbf{R}\, a_{ij}$$

([7], Section 6.2.3). Let $aG = C$, $L \in \mathbf{R}^{p \times p}$ be a g.c.l.d. of aI_p and C ([7], Section 6.3.3, Excercise 6.13), set $D_l = L^{-1} a I_p$ and $N_l = L^{-1} C$. Prove that $(D_l(\partial), N_l(\partial))$ is a left-coprime factorization of $G(\partial)$ over $\mathbf{R}$, *i.e.* $G = D_l^{-1} N_l$ and $\{D_l, N_l\}$ are left-coprime over $\mathbf{R}$.

Let $(D_l(\partial), N_l(\partial))$ be a left-coprime factorization of $G(\partial)$ over $\mathbf{R}$. The input-output system $M = [y, u]_{\mathbf{R}}$ defined by the equation $D_l(\partial)\, y = N_l(\partial)\, u$ is observable and controllable ([7], Section 6.4.1, Excercise 6.20). Let $\dim_{\mathbf{K}} (M / [u]_{\mathbf{R}})$ be the dimension of the $\mathbf{K}$-vector space $M / [u]_{\mathbf{R}}$ (*i.e.* the "order" of the above input-output system).

Definition 7.12. *The natural integer* $\dim_{\mathbf{K}} (M / [u]_{\mathbf{R}}) + \# (TP_\infty)$ *is called the* MacMillan degree *of* $G(\partial)$*, and is denoted by* $\delta_M(G)$.

Remark 7.2. If $G(\partial)$ is a *polynomial matrix,* $\delta_M(G) = \# (TP_\infty) = c_\infty(G) + \# (TZ_\infty)$.

Exercise 7.6. [25] Let $G_1(\partial) \in \mathbf{Q}^{p \times r}$ and $G_2(\partial) \in \mathbf{Q}^{r \times m}$ be two matrices of rank r. Prove that $c_\infty(G_1 G_2) = c_\infty(G_1) + c_\infty(G_2)$. (Hint: using the Dieudonné determinant and its properties established in Exercise 7.4, first show that $c_\infty(\bar{G}_1 U \bar{G}_2) = c_\infty(\bar{G}_1) + c_\infty(\bar{G}_2)$ when $\bar{G}_1(\partial) = \operatorname{diag}\{\sigma^{\mu_i}\}_{1 \leq i \leq r}$, $\bar{G}_2(\partial) = \operatorname{diag}\{\sigma^{\nu_i}\}_{1 \leq i \leq r}$ and $U \in \mathbf{GL}_r(\mathbf{S})$.)

Exercise 7.7. Let $A(\partial) \in \mathbf{Q}^{p \times m}$ and $B(\partial) \in \mathbf{Q}^{q \times m}$ be two matrices of rank r and set $F(\partial) = \begin{bmatrix} A(\partial) \\ B(\partial) \end{bmatrix}$. (i) Assuming that the Smith-MacMillan form at infinity of $A(\partial)$ and of $B(\partial)$ are $\operatorname{diag}(\sigma^{\nu_1}, ..., \sigma^{\nu_r}, 0, ..., 0)$ and $\operatorname{diag}(\sigma^{\lambda_1}, ..., \sigma^{\lambda_r}, 0, ..., 0)$, respectively, show that the Smith-MacMillan form at infinity of $F(\partial)$ is $\operatorname{diag}(\sigma^{\varepsilon_1}, ..., \sigma^{\varepsilon_r}, 0, ..., 0)$ with $\varepsilon_i = \min\{\nu_i, \lambda_i\}$, $1 \leq i \leq r$. (ii) Deduce from (i) that if $F(\partial)$ has no pole at infinity, then $A(\partial)$ and $B(\partial)$ have the same property and that $c_\infty(F) \geq \max\{c_\infty(A), c_\infty(B)\}$. (iii) Show that $\begin{bmatrix} A(\partial) \\ I_m \end{bmatrix}$ and $A(\partial)$ have the same poles at infinity (with the same orders). (Hint: for (i), $B(\partial)$ and $B(\partial)\, V$ have the same Smith-MacMillan form over $\mathbf{S}$ if $V \in \mathbf{GL}_m(\mathbf{S})$; use row elementary operations once $A(\partial)$ and $B(\partial)\,V$ have been reduced to their Smith-MacMillan form over $\mathbf{S}$, where V is a suitable element of $\mathbf{GL}_m(\mathbf{S})$.)

Exercise 7.8. *Proper model matching [25].* Let $A(\partial) \in \mathbf{Q}^{p\times m}$ and $B(\partial) \in \mathbf{Q}^{q\times m}$. The *model matching problem* is to determine a matrix $H(\partial) \in \mathbf{Q}^{q\times p}$ such that $H(\partial)\,A(\partial) = B(\partial)$. The *proper model matching problem* is to determine a *proper* solution $H(\partial) \in \mathbf{Q}^{q\times p}$ to the model matching problem. (i) Let $F(\partial)$ be defined as in Exercise 7.7. Show that the model matching problem has a solution if, and only if

$$\operatorname{rk} A(\partial) = \operatorname{rk} F(\partial). \tag{7.9}$$

(ii) Let $H(\partial) \in \mathbf{Q}^{q\times p}$; show that $H(\partial)$ is proper if, and only if, $c_\infty\left(\begin{bmatrix} H(\partial) \\ I_p \end{bmatrix}\right)$ $= 0$. (iii) Considering the model matching problem, and assuming that $m = r = \operatorname{rk} A(\partial)$, show that there exists an invertible matrix $Q(\partial) \in \mathbf{Q}^{r\times r}$ such that $F(\partial) = \bar{F}(\partial)\,Q(\partial)$ where $\bar{F}(\partial)$ has no pole and no zero at infinity and $Q(\partial)$ has the same structural indexes at infinity as $F(\partial)$. (iv) Let $\bar{F}(\partial) = \begin{bmatrix} \bar{A}(\partial) \\ \bar{B}(\partial) \end{bmatrix}$; the model matching problem can be written in the form $\begin{bmatrix} H(\partial) \\ I_p \end{bmatrix} \bar{A}(\partial) = \bar{F}(\partial)$. Prove that $H(\partial)$ is proper if and only if

$$c_\infty(A) = c_\infty(F). \tag{7.10}$$

(v) To summarize: (7.9) and (7.10) are a *necessary and sufficient condition* for the *proper model matching problem to have a solution* when $A(\partial)$ is full column rank. (Hint: use the results to be proved in exercises 7.6 and 7.7.)

7.4 Impulsive Systems and Behaviors

7.4.1 Temporal Systems

Definition of a Module by Generators and Relations

Let $M = \operatorname{coker} \bullet B(\partial)$ be a system, where $B(\partial) \in \mathbf{R}^{q\times k}$. The system equations can be written

$$\begin{cases} B(\partial)\,w = e, \\ \quad e = 0. \end{cases} \tag{7.11}$$

The above module M is said to be defined by generators and relations ([7], Section 6.3.1, Definition 6.1). Equations (7.11) correspond to the exact sequence

$$\mathbf{R}^q \overset{\bullet B}{\to} \mathbf{R}^k \overset{\varphi}{\to} M \to 0. \tag{7.12}$$

The module of generators is $\mathbf{R}^k = [\mathring{w}]_{\mathbf{R}}$ where $\mathring{w} = (\mathring{w}_i)_{1\leq i\leq k}$ is the canonical basis of $\mathbf{R}^k$; the module of relations is Im $\bullet\, B = [\mathring{e}]_{\mathbf{R}}$ where $\mathring{e} = B(\partial)\,\mathring{w} = (\mathring{e}_j)_{1\leq j\leq q}$. Let w_i and e_j be the canonical image of $\mathring{w}_i$ and $\mathring{e}_j$, respectively, in the quotient $M = \mathbf{R}^k / [\mathring{e}]_{\mathbf{R}}$ $(1 \leq i \leq k, 1 \leq j \leq q)$. Equations (7.11) are satisfied, and the second one (*i.e.* $e = 0$) expresses the fact that the relations existing between the system variables are active.

Continuous-time Temporal System

Assuming that $\mathbf{K} = \mathrm{Re}$, set $\mathbb{T} = \mathrm{Re}$ and $\mathbb{T}_0 = [0, +\infty[$. In place of (7.11), consider the equations

$$\begin{cases} B(\partial)\, w(t) = e(t)\,,\ t \in \mathbb{T}, \\ \qquad e(t) = 0,\ t \in \mathbb{T}_0. \end{cases} \tag{7.13}$$

The relations between the system variables are now active only during the time period $\mathbb{T}_0$, *i.e.* the system is *formed at initial time zero* (due, *e.g.*, to a failure or a switch). On the complement $\mathbb{T} \setminus \mathbb{T}_0$ of $\mathbb{T}_0$ in $\mathbb{T}$, e can be any C^∞ function. Let us give a provisional definition of a temporal system (the final one is given in Section 7.4.5):

Definition 7.13. *[8] The system of differential equations* (7.13) *is the* temporal system with matrix of definition $B(\partial)$.

Once the input and output variables have been chosen, one can assume without loss of generality that the first line of (7.13) corresponds to a polynomial matrix description (PMD) ([7], Section 6.4.1), *i.e.*

$$B(\partial) = \begin{bmatrix} D(\partial) & -N(\partial) & 0 \\ Q(\partial) & W(\partial) & -I_p \end{bmatrix}, \tag{7.14}$$

with module of generators $\mathbf{R}^k = [\mathring{w}]_{\mathbf{R}}$ where $\mathring{w} = \begin{bmatrix} \mathring{\xi}^T & \mathring{u}^T & \mathring{y}^T \end{bmatrix}^T$. In that case, the temporal system under consideration is called an *input-output temporal system* (this definition is consistent with the one in ([7], Section 6.4.1) of an *input-output system.*

Discrete-time Temporal System

Assuming that $\mathbf{K} = \mathrm{Re}$, set $\mathbb{T} = \mathbb{Z}$ and $\mathbb{T}_0 = \{..., -2, -1, 0\}$. Definition 7.13 still holds (since "discrete-time differential equations" are well-defined: see [7], Section 6.1), and it means that the relations between the system variables are active only *up to final time zero.* On $\mathbb{T} \setminus \mathbb{T}_0$, the sequence $(e\,(t))_{t \in \mathbb{T}}$ can have any values.

Such temporal systems are encountered in various fields, notably in economy; see [8] and the references therein for more details.

7.4.2 A Key Isomorphism

Continuous-time Case

From the analytic point of view, the temporal system $\mathbf{\Sigma}$ defined by (7.13) is formed as follows: take for e in the first line of (7.13) any $\mathcal{C}^\infty$ function; then multiply e by $1 - \Upsilon$, where Υ is the Heaviside function (*i.e.* $\Upsilon\,(t) = 1$ for $t > 0$ and 0 otherwise). Let $W = \mathcal{C}^\infty\,(\mathrm{Re};\mathrm{Re})$ and set

$$\varDelta = \oplus_{\mu \geq 0} \mathrm{Re}\ \delta^{(\mu)} \tag{7.15}$$

where δ is the Dirac distribution. The $\mathbf{R}$-module generated by $S_0 := (1 - \Upsilon)\,W$ is (as Re-vector space): $S = S_0 \oplus \varDelta$. The operator ∂ is an automorphism of the Re-vector space S, and $\sigma = \partial^{-1}$ is the operator defined on S by: $(\sigma w)\,(t) = \int_{+\infty}^{t} w\,(\tau)\,d\tau$. The space S is an $\mathbf{L}$-vector space (and thus an $\mathbf{S}$-module which is an $\mathbf{R}$-module, by restriction of the ring of scalars), and S_0 is an $\mathbf{S}$-submodule of S. The $\mathbf{R}$-module $\varDelta$ is not an $\mathbf{S}$-module, but $\varDelta \cong_{\mathrm{Re}} S/S_0$; the set S/S_0 (denoted by $\bar{\varDelta}$ in the sequel) is clearly an $\mathbf{L}$-vector space (and thus an $\mathbf{R}$-module which is an $\mathbf{S}$-module). The nature of the above isomorphism, denoted by τ, can be further detailed:

Lemma 7.1. *The isomorphism τ, defined as: $\varDelta \ni \lambda\,\delta \overset{\backsim}{\longrightarrow} \lambda\,\bar{\delta} \in \bar{\varDelta}$, is* $\mathbf{R}$-*linear.*

Proof. First, notice that any element of $\varDelta$ (resp. $\bar{\varDelta}$) is uniquely expressible in the form $\lambda\,\delta$ (resp. $\lambda\,\bar{\delta}$) for some $\lambda \in \mathbf{R}$, thus τ is a well-defined $\mathbb{Z}$-isomorphism. In addition, for any $x \in \varDelta$, such that $x = \lambda\,\delta, \lambda \in \mathbf{R}$, and any $\mu \in \mathbf{R}$, $\tau\,(\mu\,x) = \tau\,(\mu\lambda\,\delta) = \mu\lambda\,\bar{\delta} = \mu\,\tau\,(x)\,.$ ■

Therefore,

$$\varDelta \cong_{\mathbf{R}} \bar{\varDelta} \tag{7.16}$$

We have $\sigma\delta = \Upsilon - 1$; setting $\bar{\delta} = \tau\,(\delta)$, we obtain $\sigma\bar{\delta} = 0$, thus $\tilde{\delta}$ and $\bar{\delta}$ can be identified, as well as the $\mathbf{S}$-modules $\tilde{\varDelta}$ and $\bar{\varDelta}$. As a result, by (7.15), (7.5), we can write

$$\tilde{\varDelta} = \bar{\varDelta} = \oplus_{\mu \geq 0} \mathrm{Re} \tilde{\delta}^{(\mu)}. \tag{7.17}$$

The canonical epimorphism $S \to \tilde{\varDelta}$ is denoted by $\tilde{\phi}$. Let θ be the Re-linear projection $S_0 \oplus \varDelta \to \varDelta$; the following diagram is commutative:

$$\begin{tikzcd} S_0 \oplus \varDelta \arrow[r, "\tilde{\phi}"] \arrow[d, "\theta"] & \tilde{\varDelta} \\ \varDelta \arrow[ur, "\tau"] & \end{tikzcd} \tag{7.18}$$

Discrete-time Case

Let $\varUpsilon$ be the sequence defined by $\varUpsilon(t) = 1$ for $t > 0$ and 0 otherwise. From the analytic point of view, the temporal system $\mathbf{\Sigma}$ defined by (7.13) is formed as follows: take for e in the first line of (7.13) any sequence; then multiply e by $\varUpsilon$. Set $W = \mathrm{Re}^{\mathbb{Z}}$ and $S_0 = \varUpsilon W$. Let $\varDelta$ be defined as in (7.15), but where $\delta := \partial\varUpsilon$ is the "Kronecker sequence", such that $\delta(t) = 1$ for $t = 0$ and 0 otherwise (thus, $\varDelta$ is the $\mathbf{R}$-module consisting of all sequences with left and finite support). The $\mathbf{R}$-module generated by S_0 is (as Re-vector space) $S = S_0 \oplus \varDelta$. The operator ∂ is an automorphism of the Re-vector space S, and $\sigma = \partial^{-1}$ is the operator defined on S by: $(\sigma w)(t) = \sum_{j=-\infty}^{t-1} w(j)$; S is an $\mathbf{L}$-vector space. The $\mathbf{R}$-isomorphism (7.16) still holds; the same identifications as in the continuous-time case can be made and the same notation can be used. Obviously, the continuous- and discrete-time cases are now completely analogous.

7.4.3 Impulsive Behavior

Assuming that $\mathbf{K} = \mathrm{Re}$, consider a temporal system $\mathbf{\Sigma}$ with matrix of definition $B(\partial) \in \mathbf{R}^{q\times k}$.

Proposition 7.5. *The following properties are equivalent: (i) For any $e \in S_0^q$, there exists $w \in S^k$ such that* (7.13) *is satisfied. (ii) The matrix $B(\partial)$ is full row rank.*

Proof. (i) ⇒ (ii): If the matrix $B(\partial)$ is not full row rank, $\bullet B(\partial)$ is not injective, *i.e.* there exists a nonzero element $\eta(\partial) \in \mathbf{R}^q$ (*i.e.* a $1 \times q$ matrix with entries in $\mathbf{R}$) such that $\eta(\partial)\, B(\partial) = 0$. Therefore, for $w \in S^k$ and $e \in S_0^q$ to satisfy (7.13), e must satisfy the "compatibility condition" $\eta(\partial)\, e = 0$ (see [21], [20], where this compatibility condition is further detailed). (ii) ⇒ (i): By (7.8) with $B = G$, assuming that $q = r$, (7.13) is equivalent to

$$\begin{bmatrix} \mathrm{diag}\,\{\sigma^{\nu_i}\}_{1\le i\le r} & 0 \end{bmatrix} v = h \tag{7.19}$$

where $v = V(\sigma)\, w$ and $h = W(\sigma)\, e$; (7.19) is equivalent to $\sigma^{\nu_i} v_i = h_i$, $1 \le i \le q$. For any $\nu_i \in \mathbb{Z}$ and any $h_i \in S_0$, $v_i = \partial^{\nu_i} h_i$ belongs to S.

Therefore, (i) holds because h spans S_0^q as e spans the same space (since S_0 is an **S**-module). ■

In the sequel, the matrix $B(\partial)$ is assumed to be full row rank (*i.e.* $q = r$).

Notation 1 *For any scalar operator* ω *and any integer* $l \geq 1$, $\omega_{(l)}$ *denotes the operator* $diag(\omega, ..., \omega)$, *where* ω *is repeated* l *times.*

Definition 7.14. *[8] Let* $\mathcal{W} \subset S^k$ *be the space spanned by the elements* w *satisfying* (7.13) *as* e *spans* S_0^q. *The* impulsive behavior *of* $\mathbf{\Sigma}$ *is:* $\mathcal{B}_\infty = \theta_{(k)}\mathcal{W}$.

Definition 7.15. *[8] The* pseudo-impulsive behavior *of* $\mathbf{\Sigma}$ *is:* $\mathcal{A}_\infty = \tau_{(k)}\mathcal{B}_\infty$.

7.4.4 Impulsive System

Assuming that the right-hand member of the equality in (7.8) is the Smith-MacMillan form at infinity of $B(\partial)$ with $q = r$, $m = k$ and $W = U$, set

$$\varPi(\sigma) = \text{diag}\left\{\sigma^{\bar{\pi}_i}\right\}_{1 \leq i \leq r}, \quad \varSigma(\sigma) = \text{diag}\left\{\sigma^{\bar{\varsigma}_i}\right\}_{1 \leq i \leq r} \tag{7.20}$$

so that $\text{diag}\{\sigma^{\nu_i}\}_{1 \leq i \leq r} = \varPi^{-1}(\sigma)\,\varSigma(\sigma) = \varSigma(\sigma)\,\varPi^{-1}(\sigma)$. By (7.8),

$$B(\partial) = A^{-1}(\sigma)\,B^{+}(\sigma) = B'^{+}(\sigma)\,A'^{-1}(\sigma) \tag{7.21}$$

where

$$A = \varPi\,U, \quad B^{+} = \begin{bmatrix} \varSigma & 0 \end{bmatrix} V, \tag{7.22}$$

$$A' = V^{-1}(\varPi \oplus I_{k-r}), \quad B'^{+} = U^{-1}\begin{bmatrix} \varSigma & 0 \end{bmatrix}. \tag{7.23}$$

The above expressions, the results to be proved in exercises 7.9 and 7.10, and Definitions 7.16 and 7.17 below, are valid when **K** is any differential field.

Exercise 7.9. (i) The above pair $(A(\sigma), B^{+}(\sigma))$ (resp. $(B'^{+}(\sigma), A'(\sigma))$) is a left-coprime (resp. right-coprime) factorization of $B(\partial)$ over **S**. (ii) Let $\left(A_1(\sigma), B_1^{+}(\sigma)\right)$ and $\left(A_2(\sigma), B_2^{+}(\sigma)\right)$ be two left-coprime factorizations of $B(\partial)$ over **S**; then, there exists a unimodular matrix $W(\sigma)$ over **S** such that $B_2^{+}(\sigma) = W(\sigma)\,B_1^{+}(\sigma)$ and $A_2(\sigma) = W(\sigma)\,A_1(\sigma)$. (iii) A similar result holds for right-coprime factorizations of $B(\partial)$ over **S**; make it explicit. (Hint: see, *e.g.*, [31], Section 4.1, (43).)

Let $(A(\sigma), B^{+}(\sigma))$ be any left-coprime factorization of $B(\partial)$ over **S**. According to the result to be proved in Exercise 7.9(ii), the module $M^{+} = \text{coker} \bullet B^{+}(\sigma)$ is uniquely defined from $B(\partial)$.

Definition 7.16. *[8] (i) The* $\mathbf{S}$*-module* $M^{+} = coker \bullet B^{+}(\sigma)$ *is called the* impulsive system*. (ii) The torsion submodule of* M^{+}*, written* $\mathcal{T}(M^{+})$*, is called the* module of uncontrollable poles at infinity*.*

Exercise 7.10. Let $B(\partial) = [C(\partial)\ D(\partial)]$, $D(\partial) \in \mathbf{R}^{r \times m_2}$ and $C(\partial) \in \mathbf{R}^{r \times m_1}$. Assuming that $\operatorname{rk} B(\partial) = r$, let $(A(\sigma), B^{+}(\sigma))$ be a left-coprime factorization of $B(\partial)$ over $\mathbf{S}$, and set $B^{+}(\sigma) = \begin{bmatrix} C^{+}(\sigma) & D^{+}(\sigma) \end{bmatrix}$, $C(\sigma) \in \mathbf{S}^{r \times m_1}$, $D(\sigma) \in \mathbf{S}^{r \times m_2}$. (i) Prove that the Smith-MacMillan form of $B^{+}(\sigma)$ over $\mathbf{S}$ is $\begin{bmatrix} I_r & 0 \end{bmatrix}$ if, and only if $B(\partial)$ has no zero at infinity. (ii) Deduce that $\{C^{+}(\sigma), D^{+}(\sigma)\}$ are left-coprime if, and only if $\begin{bmatrix} C(\partial) & D(\partial) \end{bmatrix}$ has no zero at infinity.

Definition 7.17. *(i) Let* $C(\partial) \in \mathbf{R}^{r \times m_1}$ *and* $D(\partial) \in \mathbf{R}^{r \times m_2}$ *be such that rk* $\begin{bmatrix} C(\partial) & D(\partial) \end{bmatrix} = r$*. The matrices* $C(\partial)$ *and* $D(\partial)$ *are said to be* left-coprime at infinity *if* $\begin{bmatrix} C(\partial) & D(\partial) \end{bmatrix}$ *has no zero at infinity. (ii)* Right-coprimeness at infinity *is defined analogously.*

The connection between the pseudo-impulsive behavior $\mathcal{A}_{\infty}$ and the impulsive system M^{+} is given below, with the notation $(.)^{*} := \operatorname{Hom}_{\mathbf{S}}\left(., \tilde{\Delta}\right)$ ([7], Section 6.5.1). Let us assume that $\mathbf{K} = \mathrm{Re}$.

Theorem 7.4. *[8]* $\mathcal{A}_{\infty} = (M^{+})^{*}$.

Proof. By Definition 7.15 and the commutativity of the diagram (7.18), $\mathcal{A}_{\infty}$ is the $\mathbf{E}$-module consisting of all elements $\tilde{w} = \tilde{\phi}_{(k)} w$ for which there exists $h \in S_0^q$ such that (7.19) is satisfied (recall that $q = r$). With the notation in the proof of Proposition 7.5, this equation is equivalent to $\sigma^{\nu_i} v_i = h_i, 1 \leq i \leq q$. For any index i such that $\nu_i \leq 0$, $v_i = \sigma^{-\nu_i} h_i$ belongs to S_0, thus $\tilde{v}_i = 0$ (where $\tilde{v}_i := \tilde{\phi}\, v_i$). Therefore, $\mathcal{A}_{\infty}$ is the $\mathbf{S}$-module consisting of all elements $\tilde{w} = V^{-1}(\sigma)\, \tilde{v}$ such that $\tilde{v} \in \tilde{\Delta}^k$ satisfies $\begin{bmatrix} \Sigma(\sigma) & 0 \end{bmatrix} \tilde{v} = 0$; as a result, $\mathcal{A}_{\infty} = \ker B^{+}(\sigma) \bullet$. ■

Remark 7.3. Let $\mathbf{K}$ be *any differential field.* The equality in the statement of Theorem 7.4 still makes sense, thus this equality becomes the *definition* of $\mathcal{A}_{\infty}$ (for the construction of $\mathcal{B}_{\infty}$ when $\mathbf{K}$ is a *differential ring*, see [8]).

Proposition 7.6. *[8] Let* $\mathbf{K}$ *be any differential field; for any natural integer* $\mu, \tilde{C}_{\mu}^{*} = \tilde{C}_{\mu}$.

Proof. For $\mu = 0$, $\tilde{C}_{\mu} = \tilde{C}_{\mu}^{*} = 0$. For $\mu \geq 1$, $\tilde{C}_{\mu}^{*}$ is the set of all elements $x \in \tilde{\Delta}$ such that $\sigma^{\mu} x = 0$. Obviously, $\tilde{\delta}^{(i-1)}$ belongs to $\tilde{C}_{\mu}^{*}$ if, and only if $1 \leq i \leq \mu$. By (7.4), $\tilde{C}^*_\mu \subset \tilde{C}_\mu$. Let us prove by induction the reverse inclusion. By (7.3), for any $a \in \mathbf{K}$, $\sigma a \tilde{\delta} = \left(a^\beta \sigma - \sigma a^{\beta\gamma} \sigma\right) \tilde{\delta} = 0$, which implies that $\tilde{C}_1 = \mathbf{K}\tilde{\delta} \subset \tilde{C}^*_1$. Assuming that $\tilde{C}_\mu \subset \tilde{C}^*_\mu, \mu \geq 1$, let $a \in \mathbf{K}$; then, $\sigma^{\mu+1} a \tilde{\delta}^{(\mu)} = \sigma^\mu \left(a^\beta - \sigma a^{\beta\gamma}\right) \tilde{\delta}^{(\mu-1)}$; by hypothesis, $a^\beta \tilde{\delta}^{(\mu-1)}$ and $\sigma a^{\beta\gamma} \tilde{\delta}^{(\mu-1)}$ belong to $\tilde{C}^*_\mu$, thus $\sigma^{\mu+1} a \tilde{\delta}^{(\mu)} = 0$, which implies that $\tilde{C}_{\mu+1} \subset \tilde{C}^*_{\mu+1}$. ■

As $\tilde{\Delta}$ is a cogenerator of $_{\mathbf{S}}\mathbf{Mod}$ (see Section 7.2.4, Exercise 7.2), the following result is a consequence of ([7], Section 6.5.2, Proposition 6.28) and of Proposition 7.6, assuming that the impulsive system M^+ has the structure in Definition 7.4:

Proposition 7.7. *(i) There exist sub-behaviors $\mathcal{A}_{\infty,c} \simeq \left(\Phi^+\right)^*$ and $\mathcal{A}_{\infty,u} \simeq \left(\mathcal{T}\left(M^+\right)\right)^*$ of $\mathcal{A}_\infty$ such that $\mathcal{A}_\infty = \mathcal{A}_{\infty,c} \oplus \mathcal{A}_{\infty,u}$. (ii) The sub-behavior $\mathcal{A}_{\infty,c}$ satisfying this property is unique and such that $\mathcal{A}_{\infty,c} \cong_{\mathbf{E}} \tilde{\Delta}^{k-r}$ ($\mathcal{A}_{\infty,c}$ is called the "controllable pseudo-impulsive behavior"). (iii) $\mathcal{A}_{\infty,u} \cong_{\mathbf{E}} \prod_{i=s+1}^{r} \tilde{C}_{\mu_i}$ (this sub-behavior, unique up to* **E***-isomorphism, is said to be "uncontrollable").*

Remark 7.4. (i) According to Theorem 7.4, $\mathcal{A}_\infty$ is a "behavior" in the sense specified in ([7], Section 6.5.2), *i.e.* a *kernel*, whereas $\mathcal{B}_\infty$ (deduced from $\mathcal{A}_\infty$ using Definition 7.15) cannot be expressed in a so simple way (in this meaning, the expression "impulsive behavior" is an abuse of language). The notion of "sub-behavior" of a pseudo-impulsive behavior $\mathcal{A}_\infty$ is defined in accordance with the general definition in ([7], Section 6.5.2). (ii) A construction of $\mathcal{B}_\infty$ using the "causal Laplace transform" (in the continuous-time case) and the "anti-causal Z-transform" (in the discrete-time case) is developed in [5] and [6]. This construction has connections with the pioneer work of Verghese [29], and with the recent contributions [21] and [20] (where the approach is quite different and limited to the case of systems with constant coefficients).

Assuming that $\mathbf{K} = \mathrm{Re}$, set for any integer $\mu \geq 1$

$$C_\mu = \tau^{-1}\left(\tilde{C}_\mu\right) = \oplus_{i=1}^{\mu} \mathrm{Re}\delta^{(i-1)}. \tag{7.24}$$

The following theorem is an obvious consequence of Proposition 7.7:

Theorem 7.5. *The impulsive behavior $\mathcal{B}_\infty$ can be expressed as: $\mathcal{B}_\infty = \mathcal{B}_{\infty,c} \oplus \mathcal{B}_{\infty,u}$, where $\mathcal{B}_{\infty,c} := \tau_{(k)}^{-1} \mathcal{A}_{\infty,c} \cong_{\mathrm{Re}} \Delta^\kappa$ and $\mathcal{B}_{\infty,u} = \tau_{(k)}^{-1} \mathcal{A}_{\infty,u} \cong_{\mathrm{Re}} \prod_{i=1}^{\rho} C_{\mu_i}$ (the space $\mathcal{B}_{\infty,c}$, which is uniquely determined, is called the "controllable impulsive behavior", and the impulsive behavior $\mathcal{B}_{\infty,u}$, unique up to* Re-*isomorphism, is said to be "uncontrollable").*

7.4.5 Generalization of the Notion of Temporal System

Up to now, the notion of temporal system has been defined when the coefficient field is $\mathbf{K} = \mathrm{Re}$. Our aim in this section is to generalize this notion to a coefficient field $\mathbf{K}$ which is any differential field.

Strict Equivalence

Let $\mathbf{D}$ be a principal ideal domain[2] and M_i be a f.g. $\mathbf{D}$-module with matrix of definition $B_i \in \mathbf{D}^{q_i \times k_i}$ $(i = 1, 2)$, assumed to be full row rank. The following result is classical ([12], Section 0.6, Theorem 6.2 and Proposition 6.5):

Proposition 7.8. *The two following properties are equivalent: (i) $M_1 \cong M_2$; (ii) B_1 and B_2 satisfy a "comaximal relation", i.e. there exist two matrices $A_1 \in \mathbf{D}^{q_1 \times q_2}$ and $A_2 \in \mathbf{D}^{k_1 \times k_2}$ such that*

$$\begin{bmatrix} B_1 & A_1 \end{bmatrix} \begin{bmatrix} -A_2 \\ B_2 \end{bmatrix} = 0, \tag{7.25}$$

$\begin{bmatrix} B_1 & A_1 \end{bmatrix}$ is right-invertible (i.e. $\{B_1, A_1\}$ are left-coprime) and $\begin{bmatrix} -A_2 \\ B_2 \end{bmatrix}$ is left-invertible (i.e. $\{B_2, A_2\}$ are right-coprime).

Definition 7.18. *[9] Consider two PMDs $\{D_i, N_i, Q_i, W_i\}$ with matrices over $\mathbf{R} = \mathbf{K}\,[\partial; \alpha, \gamma]$, partial state ξ_i, input u and output y $(i = 1, 2)$. These PMD are said to be* strictly equivalent *if the diagram below is commutative:*

$$\begin{tikzcd}
{[\xi_1, u]_{\mathbf{R}}} \arrow[rr, "\chi"] & & {[\xi_2, u]_{\mathbf{R}}} \\
& {[u, y]_{\mathbf{R}}} \arrow[ul, "i_1"] \arrow[ur, "i_2"'] &
\end{tikzcd} \tag{7.26}$$

where $[u, y]_{\mathbf{R}} = [u]_{\mathbf{R}} + [y]_{\mathbf{R}}$, i_1 and i_2 are the canonical injections and χ is an $\mathbf{R}$-isomorphism.

Proposition 7.9. *The two above PMDs are strictly equivalent if, and only if there exist matrices M_1, X_1, M_2, X_2 over $\mathbf{R}$, of appropriate dimension, such that*

$$\begin{bmatrix} M_1 & 0 \\ -X_1 & I_p \end{bmatrix} \begin{bmatrix} D_2 & -N_2 \\ Q_2 & W_2 \end{bmatrix} = \begin{bmatrix} D_1 & -N_1 \\ Q_1 & W_1 \end{bmatrix} \begin{bmatrix} M_2 & X_2 \\ 0 & I_m \end{bmatrix}, \tag{7.27}$$

$$\left\{ \begin{bmatrix} D_1 & -N_1 \end{bmatrix}, M_1 \right\} \text{ are left coprime over } \mathbf{R}, \tag{7.28}$$

$$\{D_2, M_2\} \text{ are right coprime over } \mathbf{R}. \tag{7.29}$$

[2] More general rings can be considered: see [12].

Proof . 1) Set $\tilde{B}_i = \begin{bmatrix} D_i & -N_i \end{bmatrix}$ $(i = 1, 2)$. By Proposition 7.8, the isomorphism χ in (7.26) exists if, and only if the two matrices $\tilde{B}_1$ and $\tilde{B}_2$ satisfy a comaximal relation, *i.e.* there exist two matrices $A_1 = M_1$ and $A_2 = \begin{bmatrix} M_2 & X_2 \\ Y_2' & Z_2 \end{bmatrix}$ such that $\tilde{B}_1 \, A_2 = M_1 \, \tilde{B}_2$ with the suitable coprimeness properties. 2) The existence of the canonical injections i_1 and i_2 in (7.26) means that (a) we have $\begin{bmatrix} \xi_1 \\ u \end{bmatrix} = \begin{bmatrix} M_2 & X_2 \\ Y_2' & Z_2 \end{bmatrix} \begin{bmatrix} \xi_2 \\ u \end{bmatrix}$, thus $Y_2' = 0$ and $Z_2 = I_m$; (b) we have $y = Q_2 \, \xi_2 + W_2 \, u = Q_1 \, \xi_1 + W_1 \, u$. The latter quantity can be expressed as: $Q_1 \, \xi_1 + W_1 \, u = Q_1 \, (M_2 \, \xi_2 + X_2 \, u) + W_1 \, u + X_1 \, (D_2 \, \xi_2 - N_2 \, u)$, thus $Q_2 = Q_1 \, M_2 + X_1 \, D_2$, $W_2 = Q_1 \, X_2 + W_1 - X_1 \, N_2$, *i.e.* the equality (7.27) holds. 3) (7.28) and (7.29) mean that $\left\{ \tilde{B}_1, A_1 \right\}$ and $\left\{ \tilde{B}_2, A_2 \right\}$ are, respectively, left- and right-coprime. ■

When $\mathbf{K} = \mathrm{Re}$, Proposition 7.9 is essentially due to Fuhrmann [17]; other equivalent formulations have been developed (see, *e.g.*, [19], Section 8.2).

Full Equivalence

Let us consider the matrix of definition $B(\partial)$ in (7.14), associated with a PMD, let $(A(\sigma), B^+(\sigma))$ be a left-coprime factorization of $B(\partial)$ over $\mathbf{S}$, and write

$$B^+(\sigma) = \begin{bmatrix} D^+(\sigma) & -N^+(\sigma) & Z^+(\sigma) \\ Q^+(\sigma) & W^+(\sigma) & Y^+(\sigma) \end{bmatrix} \tag{7.30}$$

according to the sizes in (7.14). The impulsive system associated with $B(\partial)$ is $M^+ = \operatorname{coker} \bullet B^+(\sigma)$ (see Section 7.4.4, Definition 7.16), and $M^+ = [\xi^+, u^+, y^+]_{\mathbf{S}}$ where

$$B^+(\sigma) \begin{bmatrix} \xi^+ \\ u^+ \\ y^+ \end{bmatrix} = 0.$$

Definition 7.19. *Consider two PMDs* $\{D_i, N_i, Q_i, W_i\}$ *with matrices over* $\mathbf{R}$ *and denote by* M_i^+ *the associated impulsive system* $(i = 1, 2)$. *These PMDs are* fully *equivalent if (i) they are strictly equivalent and (ii) there exists an* $\mathbf{S}$*-isomorphism* $M_1^+ \cong M_2^+$.

Exercise 7.11. (i) Write the comaximal relation (7.27) (over $\mathbf{R}$) in the form (7.25) (where each B_i $(i = 1, 2)$ is of the form (7.14) with matrices $\{D_i, N_i, Q_i, W_i\}$) and determine the form of A_i $(i = 1, 2)$. (ii) Let $\left(J_1, \begin{bmatrix} B_1^+ & A_1^+ \end{bmatrix}\right)$ be a left-coprime factorization over $\mathbf{S}$ of $\begin{bmatrix} B_1 & A_1 \end{bmatrix}$ and $\left(\begin{bmatrix} -A_2^+ \\ B_2^+ \end{bmatrix}, J_2\right)$ be a right-coprime factorization over $\mathbf{S}$ of $\begin{bmatrix} -A_2 \\ B_2 \end{bmatrix}$. Prove that

$$\begin{bmatrix} B_1^+ & A_1^+ \end{bmatrix} \begin{bmatrix} -A_2^+ \\ B_2^+ \end{bmatrix} = 0. \tag{7.31}$$

(iii) Using the result to be proved in Exercise 7.10 (Section 7.4.4), show that (7.31) is a comaximal relation if, and only if $\{B_1, A_1\}$ and $\{A_2, B_2\}$ are, respectively, left- and right-coprime at infinity. (iv) Show that the following properties are equivalent: (a) $\left(J_1, B_1^+\right)$ is a left-coprime factorization over $\mathbf{S}$ of B_1; (b) $\delta_M\left(B_1\right) = \delta_M\left(\begin{bmatrix} B_1 & A_1 \end{bmatrix}\right)$. (v) Similarly, show that the following properties are equivalent: (a') $\left(B_2^+, J_2\right)$ is a right-coprime factorization over $\mathbf{S}$ of B_2; (b') $\delta_M\left(B_2\right) = \delta_M\left(\begin{bmatrix} -A_2 \\ B_2 \end{bmatrix}\right)$. (vi) Finally, using (7.21) and (7.23), conclude that two PMDs $\{D_i, N_i, Q_i, W_i\}$ $(i = 1, 2)$ are *fully equivalent* (written $\{D_1, N_1, Q_1, W_1\} \overset{f}{\sim} \{D_2, N_2, Q_2, W_2\}$) if, and only if (1) they are strictly equivalent, (2) $\{B_1, A_1\}$ and $\{A_2, B_2\}$, as defined in (i), are, respectively, left- and right-coprime at infinity, (3) the MacMillan degree conditions in (iv) and (v) are satisfied. (Hint: for (vi), $\left(B_2^+, J_2\right)$ is a right-coprime factorization of B_2 over $\mathbf{S}$ if, and only if $c_\infty\left(\begin{bmatrix} B_2^+ \\ J_2 \end{bmatrix}\right) = 0$, *i.e.* $c_\infty\left(\begin{bmatrix} B_2 \\ I \end{bmatrix}\right) = -c_\infty\left(J_2\right)$: see Exercise 7.6 (Section 7.3.2). In addition, $c_\infty\left(\begin{bmatrix} B_2 \\ I \end{bmatrix}\right) = \delta_M\left(\begin{bmatrix} B_2 \\ I \end{bmatrix}\right) = \delta_M\left(B_2\right)$ from Remark 7.2 and Exercise 7.7(iii) (Section 7.3.2). Conclude.)

Full equivalence is defined in [26] (see also [18]) in the case $\mathbf{K} = \mathrm{Re}$ in accordance with the necessary and sufficient condition to be proved in Exercise 7.11(vi) –in a slightly different but equivalent form.

An Algebraic Definition of a Temporal System

Definition 7.20. *A* temporal input-output system $\mathbf{\Sigma}$ *is an* equivalence class of fully equivalent PMDs. *The* impulsive system M^+ *of* $\mathbf{\Sigma}$ *is defined (up to* $\mathbf{S}$*-isomorphism) according to Definition 7.16 in Section 7.4.4, its* pseudo-impulsive behavior $\mathcal{A}_\infty$ *is defined according to Remark 7.3 in Section 7.4.4, and its* system M *is defined (up to* $\mathbf{R}$*-isomorphism) as* $[\xi, u]_{\mathbf{R}}$*, where* ξ *is the partial state of any PMD belonging to* $\mathbf{\Sigma}$*.*

7.5 Poles and Zeros at Infinity

In this section, $\mathbf{K}$ is any differential field and we consider an input-output temporal system $\mathbf{\Sigma}$ (in the meaning of Definition 7.20). The transfer matrix of $\mathbf{\Sigma}$ is $G(\partial) = Q(\partial) D^{-1}(\partial) N(\partial) + W(\partial)$, where $\{D, N, Q, W\}$ is any PMD belonging to $\mathbf{\Sigma}$.

The *transmission poles and zeros at infinity* of $\mathbf{\Sigma}$ are those of its input-output system M (see Definition 7.9, Section 7.3.2).

The definition of the other poles and zeros at infinity of $\mathbf{\Sigma}$ is made from its *impulsive system* M^{+} [11], and it is similar to the definition of the various *finite poles and zeros* from its *system* M ([7], Section 6.4.2). Poles and zeros at infinity, as defined below, are torsion $\mathbf{S}$-modules (transmission poles and zeros at infinity can also be defined in this manner: see Exercise 7.17 below). Let T^{+} be any of these modules; the associated pseudo-impulsive behavior is $\left(T^{+}\right)^{*} = \operatorname{Hom}_{\mathbf{S}}\left(T^{+}, \tilde{\Delta}\right)$, from which the associated impulsive behavior can be calculated in the case $\mathbf{K} = \mathrm{Re}$, using the commutative diagram (7.18). For explicit calculations, the matrix of definition $B^{+}(\sigma)$ of M^{+} is assumed to be given by (7.30).

7.5.1 Uncontrollable Poles at Infinity

The module of *uncontrollable poles at infinity* is $\mathcal{T}\left(M^{+}\right)$, according to Definition 7.16 (Section 7.4.4). Its elementary divisors are those of the matrix $B^{+}(\sigma)$.

Exercise 7.12. Assume that $\mathbf{K} = \mathrm{Re}$. Let $\mathbf{\Sigma}_1$ be the system $(\partial + 1)\, y = u$, $\mathbf{\Sigma}_2$ be the system $\bar{y} = \partial^2 \bar{u}$, and consider the interconnection $\bar{y} = u$ (*i.e.*, $\mathbf{\Sigma}_2 \rightarrow \mathbf{\Sigma}_1$), active only for $t \in \mathbb{T}_0$. (i) What kind of "pole-zero cancellation at infinity" does arise when the temporal system is formed? (ii) Check that $\mathcal{T}\left(M^{+}\right) \cong \tilde{C}_1 = \mathrm{Re}\, \tilde{\delta}$. (iii) "Where" in the temporal system does the uncontrollable impulsive behavior $\mathrm{Re}\, \delta$ arise?

7.5.2 System Poles at Infinity

Let $\breve{M}^{+} = \mathbf{L} \bigotimes_{\mathbf{S}} M^{+}$ and $\breve{\varphi} : M^{+} \rightarrow \breve{M}$ be the canonical map defined by $\breve{\varphi}\left(w^{+}\right) = \breve{w}^{+} = 1_{\mathbf{L}} \otimes_{\mathbf{S}} w^{+}$, where $1_{\mathbf{L}}$ is the unit-element of $\mathbf{L}$.

Lemma 7.2. *(i) The* $\mathbf{L}$*-vector space* $\breve{M}^{+}$ *is of dimension* m *and* $\breve{M}^{+} = \left[\breve{u}^{+}\right]_{\mathbf{L}}$. *(ii) The module* $\left[u^{+}\right]_{\mathbf{S}}$ *is free of rank* m *and* $M^{+} / \left[u^{+}\right]_{\mathbf{S}}$ *is torsion. (iii) Considering equality (a) in Theorem 7.1(i),* $\left[u^{+}\right]_{\mathbf{S}} \subset \Phi^{+}$.

Proof. The module M^{+} is defined by $B^{+}(\sigma)\, w^{+} = 0$, therefore $A^{-1}(\sigma)$ $B^{+}(\sigma)\, \breve{w}^{+} = 0$, *i.e.* $B(\partial)\, \breve{w}^{+} = 0$. Thus, from Section 7.3.1, $\breve{M}^{+} = \mathbf{L} \bigotimes_{\mathbf{Q}} \hat{M} = \mathbf{L} \bigotimes_{\mathbf{Q}} [\hat{u}]_{\mathbf{Q}} = \left[\breve{u}^{+}\right]_{\mathbf{L}}$, and (i) is proved. By (i), $\left[u^{+}\right]_{\mathbf{S}}$ is of rank m, and $\left[u^{+}\right]_{\mathbf{S}}$ is free since $\left[u^{+}\right]_{\mathbf{S}}$ is generated by m elements[3]; in addition, $\mathbf{L} \bigotimes_{\mathbf{S}} \left(M^{+} / \left[u^{+}\right]_{\mathbf{S}}\right) = 0$, thus $M^{+} / \left[u^{+}\right]_{\mathbf{S}}$ is torsion. (iii) is a consequence of the freeness of $\left[u^{+}\right]_{\mathbf{S}}$ (see addendum 6 in Section 7.7). ■

[3] Over a left Ore domain, the rank of a module is the cardinal of a maximal linearly independent subset of that module ([12], Section 0.9).

Definition 7.21. *The module of* system poles at infinity, *denoted by* SP_∞, *is* $M^+ / [u^+]_{\mathbf{S}}$.

The elementary divisors of $M^+ / [u^+]_{\mathbf{S}}$ are those of the submatrix

$$\begin{bmatrix} D^+(\sigma) & Z^+(\sigma) \\ Q^+(\sigma) & Y^+(\sigma) \end{bmatrix}$$

of $B^+(\sigma)$.

Exercise 7.13. Show that the temporal system in Exercise 7.12 has one transmission pole at infinity of order 1 and that $SP_\infty \cong \tilde{C}_2$.

7.5.3 Hidden Modes at Infinity

The module of uncontrollable poles at infinity is also called the module of *input-decoupling zeros at infinity* and is denoted by IDZ_∞.

Definition 7.22. *The module of* output-decoupling zeros at infinity *(denoted by* ODZ_∞*) is* $M^+ / [y^+, u^+]_{\mathbf{S}}$.

The elementary divisors of $M^+ / [y^+, u^+]_{\mathbf{S}}$ are those of the submatrix $\begin{bmatrix} D^+(\sigma) \\ Q^+(\sigma) \end{bmatrix}$.

Exercise 7.14. Assume that $\mathbf{K} = \mathrm{Re}$. Consider the systems $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ in Exercise 7.12 and the interconnection $y = \bar{u}$ (*i.e.*, $\mathbf{\Sigma}_1 \rightarrow \mathbf{\Sigma}_2$), active only for $t \in \mathbb{T}_0$. (i) What kind of "pole-zero cancellation at infinity" does arise when the temporal system is formed? (ii) Check that $M^+ / [y^+, u^+]_{\mathbf{S}} \cong \tilde{C}_1 = \mathrm{Re}\,\tilde{\delta}$. (iii) "Where" in the temporal system does the "unobservable impulsive behavior" $\mathrm{Re}\,\delta$ arise?

Definition 7.23. *The module of input-output decoupling zeros at infinity (denoted by* $IODZ_\infty$*) is* $\mathcal{T}(M^+) / (\mathcal{T}(M^+) \cap [y^+, u^+]_{\mathbf{S}})$.

Definition 7.24. *Considering equality (a) in Theorem 7.1(i) (Section 7.2.5), the module of hidden modes at infinity (denoted by* HM_∞*) is*

$$M^+ / \left(\Phi^+ \cap \left[y^+, u^+\right]_{\mathbf{S}}\right) .$$

Remark 7.5. The above module HM_∞ is uniquely determined up to isomorphism, as shown by Theorem 7.6 below.

Let us denote by $\varepsilon(T^+)$ the set of all elementary divisors of a f.g. torsion $\mathbf{S}$-module T^+.

Exercise 7.15. (i) Let T_1^+ and T_2^+ be submodules of a f.g. **S**-module, such that T_1^+ and T_2^+ are torsion and $T_1^+ \cap T_2^+ = 0$. Prove that $\varepsilon\left(T_1^+ \oplus T_2^+\right) = \varepsilon\left(T_1^+\right) \overset{\bullet}{\cup} \varepsilon\left(T_2^+\right)$, where $\overset{\bullet}{\cup}$ is the disjoint union[4]. (ii) Let M_1^+, M_2^+, M_3^+ be f.g. torsion **S**-modules such that $M_1^+ \subset M_2^+ \subset M_3^+$ and M_{i+1}^+/M_i^+ is torsion $(i = 1, 2)$. Prove that $\#\left(M_3^+/M_1^+\right) = \#\left(M_3^+/M_2^+\right) + \#\left(M_2^+/M_1^+\right)$ (Hint: for (i): denoting by B_i^+ a matrix of definition of $T_i^+, i = 1, 2$, the diagonal sum[5] $B_1^+ \oplus B_2^+$ is a matrix of definition of $T_1^+ \oplus T_2^+$; for (ii): use ([7], Proposition 6.7(iii)) and see the proof of ([9], Lemma(a)).)

The following result is classical ([1], Section II.1.5) and will be useful in the sequel.

Lemma 7.3. *Let* **A** *be a ring and* M_i *and* N_i *be submodules of an* **A***-module, such that* $N_i \subset M_i$ $(i = 1, 2)$ *and* $M_1 \cap M_2 = 0$. *Then*

$$\frac{M_1 \oplus M_2}{N_1 \oplus N_2} \cong \frac{M_1}{N_1} \oplus \frac{M_2}{N_2}.$$

The theorem below is more precise than ([11], Theorem 2(1)) :

Theorem 7.6. *The following equality holds*[6]*:*

$$\varepsilon\left(HM_\infty\right) = \varepsilon\left(IDZ_\infty\right) \overset{\bullet}{\cup} \varepsilon\left(ODZ_\infty\right) \setminus \varepsilon\left(IODZ_\infty\right).$$

Proof. We have

$$\frac{M^+}{\Phi^+ \cap [y^+, u^+]_{\mathbf{S}}} = \frac{\mathcal{T}\left(M^+\right) \oplus \Phi^+}{\Phi^+ \cap [y^+, u^+]_{\mathbf{S}}} \cong \mathcal{T}\left(M^+\right) \oplus \frac{\Phi^+}{\Phi^+ \cap [y^+, u^+]_{\mathbf{S}}}, \tag{7.32}$$

this isomorphism holding because $(\Phi^+ \cap [y^+, u^+]_{\mathbf{S}}) \cap \mathcal{T}\left(M^+\right) = 0$. In addition,

$$\begin{aligned} \frac{M^+}{[y^+, u^+]_{\mathbf{S}}} &= \frac{\mathcal{T}\left(M^+\right) \oplus \Phi^+}{\left(\mathcal{T}\left(M^+\right) \cap [y^+, u^+]_{\mathbf{S}}\right) \oplus \left(\Phi^+ \cap [y^+, u^+]_{\mathbf{S}}\right)} \\ &\cong \frac{\mathcal{T}\left(M^+\right)}{\mathcal{T}\left(M^+\right) \cap [y^+, u^+]_{\mathbf{S}}} \oplus \frac{\Phi^+}{\Phi^+ \cap [y^+, u^+]_{\mathbf{S}}} \end{aligned} \tag{7.33}$$

by Lemma 7.3. The theorem is a consequence of (7.32), (7.33) and of the result to be proved in Exercise 7.15(i). ■

[4] This notion was already used in [7]. For example, $\{x, y\} \overset{\bullet}{\cup} \{x, z\} = \{x, x, y, z\}$.

[5] See [7], footnote 10.

[6] The reader may notice that the same relation holds between the sets of all elementary divisors of the modules of *finite* hidden modes (defined in addendum 9, Section 7.7), i.d.z., o.d.z. and i.o.d.z. This relation is more precise than equality (6.37) in ([7], Section 6.4.2).

7.5.4 Invariant Zeros at Infinity

Definition 7.25. *The module of* invariant zeros at infinity *(denoted by IZ_∞) is $\mathcal{T}\left(M^+/\left[y^+\right]_{\mathbf{S}}\right)$.*

The elementary divisors of $\mathcal{T}\left(M^+/\left[y^+\right]_{\mathbf{S}}\right)$ are those of the submatrix $\begin{bmatrix} D^+(\sigma) & -N^+(\sigma) \\ Q^+(\sigma) & W^+(\sigma) \end{bmatrix}$.

Exercise 7.16. Assume that $\mathbf{K} = \mathrm{Re}$. Let $\mathbf{\Sigma}_1$ be the system $y = \partial^2 u$, $\mathbf{\Sigma}_2$ be the system $\partial^3 \bar{y} = \partial \bar{u}$, and consider the interconnection $y = \bar{u}$ (*i.e.*, $\varSigma_1 \rightarrow \varSigma_2$), active only for $t \in \mathbb{T}_0$. (i) What kind of "pole-zero cancellation at infinity" does arise when the temporal system is formed? (ii) Calculate IDZ_∞, ODZ_∞, $IODZ_\infty$, HM_∞, SP_∞ and IZ_∞. (Answers: $IDZ_\infty = IODZ_\infty = 0$, $ODZ_\infty \cong HM_\infty \cong SP_\infty \cong IZ_\infty \cong \tilde{C}_2$.)

7.5.5 System Zeros at Infinity

To the author's knowledge, a module system zeros at infinity was not defined However, the *degree of the system zeros at infinity* (denoted by $\#\left(SZ_\infty\right)$ is defined as follows:

Definition 7.26. *$\#\left(SZ_\infty\right) = \#\left(TZ_\infty\right) + \#\left(HM_\infty\right)$.*

7.5.6 Relations between the Various Poles and Zeros at Infinity

Exercise 7.17. Consider the two torsion modules TP_∞ and TZ_∞ below:

$$TP_\infty = \frac{\varPhi^+ \cap \left[y^+, u^+\right]_{\mathbf{S}}}{\left[u^+\right]_{\mathbf{S}}}, \quad TZ_\infty = \mathcal{T}\left(\frac{\varPhi^+ \cap \left[y^+, u^+\right]_{\mathbf{S}}}{\varPhi^+ \cap \left[y^+\right]_{\mathbf{S}}}\right).$$

(i) Prove that these modules are uniquely determined up to isomorphism. (ii) Call TP_∞ and TZ_∞ the *module of transmission poles at infinity* and the *module of transmission zeros at infinity*, respectively. Prove that this definition is consistent with Definition 7.9 (Section 7.3.2). (iii) Notice that $HM_\infty \subset SP_\infty$ and that $HM_\infty \cong SP_\infty / TP_\infty$ (see ([7], Proposition 6.7(iii)). Is it possible to deduce from this isomorphism a connection between the *elementary divisors* of TP_∞, those of HM_∞ and those of SP_∞? (Hint: for (i), write

$$\frac{\left[y^+, u^+\right]_{\mathbf{S}}}{\left[y^+\right]_{\mathbf{S}}} \cong \frac{\mathcal{T}\left(M^+\right) \cap \left[y^+, u^+\right]_{\mathbf{S}}}{\mathcal{T}\left(M^+\right) \cap \left[y^+\right]_{\mathbf{S}}} \oplus \frac{\varPhi^+ \cap \left[y^+, u^+\right]_{\mathbf{S}}}{\varPhi^+ \cap \left[y^+\right]_{\mathbf{S}}},$$

and deduce from this isomorphism that TZ_∞ is uniquely determined up to isomorphism; do a similar reasoning for TP_∞. For (ii), see [9], Sections 3.3 and 3.6, where the modules of *finite* transmission poles and zeros are considered. For (iii), see Exercise 7.13: the answer is negative.)

Exercise 7.18. [11] Prove the following relations: (i) $\#(SP_\infty) = \#(TP_\infty) + \#(HM_\infty)$; (ii) $\#(TZ_\infty) + \#(IODZ_\infty) \leq \#(IZ_\infty) \leq \#(SZ_\infty)$; (iii) if $G(\partial)$ is full row rank, $\#(TZ_\infty) + \#(IDZ_\infty) \leq \#(IZ_\infty)$; (iv) if $G(\partial)$ is full column rank, $\#(TZ_\infty) + \#(ODZ_\infty) \leq \#(IZ_\infty)$; (v) if $G(\partial)$ is square and invertible, $\#(IZ_\infty) = \#(SZ_\infty)$. (Hint: use Lemma 7.3 and the results to be proved in exercises 7.15 and 7.17; see also the proof of ([9], Theorem 1 and 2).)

Exercise 7.19. [10] Assume that $\mathbf{K} = \mathrm{Re}(t)$. Consider the temporal system whose matrix of definition $B(\partial)$ is given by (7.14) with $D(\partial) = \begin{bmatrix} 1 & 0 \\ t\partial^3 & \partial^2 \end{bmatrix}$, $N(\partial) = \begin{bmatrix} 0 \\ (t-1)\partial \end{bmatrix}$, $Q(\partial) = \begin{bmatrix} t\partial & t^2\partial \end{bmatrix}$, $W(\partial) = t^2\partial$. Calculate SP_∞, IDZ_∞, ODZ_∞, $IODZ_\infty$, HM_∞, IZ_∞, TP_∞, ZP_∞ (see exercise 7.17) and $\#(SZ_\infty)$. Interpretation? (Answers: $SP_\infty \cong \tilde{C}_1 \oplus \tilde{C}_1$, $IDZ_\infty \cong ODZ_\infty \cong IODZ_\infty \cong HM_\infty \cong TP_\infty \cong \tilde{C}_1$, $ZP_\infty = 0$ and $\#(SZ_\infty) = 1$. For the detailed interpretations, see [10].)

7.6 Concluding Remarks

Using the algebraic tools explained in this chapter, several results, classical for linear time-invariant systems, have been extended to linear continuous- or discrete-time-varying systems, *e.g.*: (i) the necessary and sufficient condition for the proper model matching problem to have a solution (Exercise 7.8), (ii) the necessary and sufficient condition for two PMDs to be fully equivalent (Exercise 7.11), (iii) the relations between the various poles and zeros at infinity (Exercise 7.18). Hidden modes at infinity are related to "pole/zero cancellations at infinity" (exercises 7.12 and 7.16). The algebraic definition of a temporal system (Definition 7.20) is new.

Impulsive behaviors of linear continuous- or discrete-time-varying temporal systems are further studied in [8], assuming that $\mathbf{K}$ is a differential ring such as $\mathrm{Re}[t]$ and that a suitable regularity condition is satisfied.

7.7 *errata* and *addenda* for [7]

1) p. 240, 17th line from top, change "of a_i by f" to: "$f(\mathbf{A}\, a_i)$"

2) p. 242, change the two sentences beginning at 6th line from top to: "Let $\psi : \mathbf{A} \to M$ be the epimorphism $\lambda \to \lambda\, w$. As $\ker \varphi = \mathbf{a}$, there exists an isomorphism $M \cong \mathbf{A}/\mathbf{a}$ by Proposition 6.7(i). Conversely, a quotient $\mathbf{A}/\mathbf{a}$ is cyclic, generated by $\varphi(1)$, where $\varphi : \mathbf{A} \to \mathbf{A}/\mathbf{a}$ is the canonical epimorphism."

3) p. 248, 7th line from top, add after "r": "(this Jordan block is also denoted by J_π in the sequel)"

4) p. 249, 8th line from bottom, change ", since $|U|$ is a *rational function* of the entries of U" to: "in general"

5) p. 251, 4th line from top, change the sentence in parentheses to: "not necessarily square, but such that all its entries outside the main diagonal are zero"

6) p. 257, after 13th line from top, add: "Consider equality (a) in Theorem 6.5(i). We have $[u]_{\mathbf{R}} \subset \mathcal{T}(M) \oplus \Phi$ and $[u]_{\mathbf{R}} \cap \mathcal{T}(M) = 0$, thus $[u]_{\mathbf{R}} \subset \Phi$."

7) p. 259, 9th line from top, change "controllable quotient" to: "controllable subsystem"

8) p. 260, 11th line from top, change "is equivalent to" to: "implies"

9) p. 265, after 6th line from top, add: "Consider equality (a) in Theorem 6.5(i); the *module of hidden modes* is $M/\left([y,u]_{\mathbf{R}} \cap \Phi\right)$."

10) p. 265, 10th line from top, change "is defined as:" to: "satisfies the equality"

11) p. 269, 2nd line from bottom and p. 270, 2nd line from top, change "$\operatorname{Hom}_{\mathbf{A}}\left(\mathbf{D}^k, W\right)$" to: "$\operatorname{Hom}_{\mathbf{A}}\left(\mathbf{A}^k, W\right)$"

12) p. 272, 9th line from bottom, change "$\mathbf{D}\,(p+k)$" to: "$\mathbf{D}\left(p^{n+k}\right)$"

13) p. 272, 8th line from bottom, change "$p+k$" to: "$n+k$"

14) p. 276, 1st line from top, add before "A *sub-behavior*": "A *subsystem* of M is a quotient M/N of M. "

15) p. 276, 7th line from top, change "quotient" to "subsystem"

References

1. Bourbaki N. (1970) *Algèbre, Chapitres 1 à 3*, Hermann.
2. Bourbaki N. (1981) *Algèbre, Chapitres 4 à 7*, Masson.
3. Bourbaki N. (1971) *Topologie générale, Chapitres 1 à 4*, Hermann.
4. Bourbaki N. (1974) *Topologie générale, Chapitres 5 à 10*, Hermann.
5. Bourlès, H. (December 2002) "A New Look on Poles and Zeros at Infinity in the Light of Systems Interconnection", *Proc. 41st Conf. on Decision and Control*, Las Vegas, Nevada, pp. 2125-2130.
6. Bourlès H. (September 2003), "Impulsive Behaviors of Discrete and Continuous Time-Varying Systems: a Unified Approach", *Proc. European Control Conf.*, Cambridge, U.K.
7. Bourlès H. (2005) "Structural Properties of Discrete and Continuous Linear Time-Varying Systems: A Unified Approach", in: *Advanced Topics in Control Systems Theory*, Lamnabhi-Lagarrique F., Loría A., Panteley E. (eds.), Chap. 6, pp. 225-280, Springer.

8. Bourlès H. (2005) "Impulsive systems and behaviors in the theory of linear dynamical systems", *Forum Math.*, vol. 17, n°5, pp. 781–808.
9. Bourlès H., Fliess M. (1997) "Finite poles and zeros of linear systems: an intrinsic approach", *Internat. J. Control*, vol. 68, pp. 897-922.
10. Bourlès H., Marinescu B. (July 1997), "Infinite Poles and Zeros of Linear Time-Varying Systems: Computation Rules", *Proc. 4th European Control Conf.*, Brussels, Belgium.
11. Bourlès H., Marinescu B. (1999) "Poles and Zeros at Infinity of Linear Time-Varying Systems", *IEEE Trans. on Automat. Control*, vol. 44, pp. 1981-1985.
12. Cohn P. M. (1985) *Free Rings and their Relations*, Academic Press.
13. Dieudonné J. (1943) "Les déterminants sur un corps non commutatif", *Bulletin de la S.M.F.*, tome 71, n°2, pp. 27-45.
14. Fliess M. (1990) "Some Structural Properties of Generalized Linear Systems", *Systems and Control Letters*, vol. 15, pp. 391-396.
15. Fliess M. (1994) "Une Interprétation Algébrique de la Transformation de Laplace et des Matrices de Transfert", *Linear Algebra Appl.*, vol. 203-204, pp. 429-442.
16. Fliess M., Lévine J., Rouchon P. (1993) "Index of an implicit time-varying linear differential equation: a noncommutative linear algebraic approach", *Linear Algebra Appl.*, vol. 186, pp. 59-71.
17. Fuhrmann P.A. (1977) "On Strict System Equivalence and Similarity", *Internat. J. Control*, vol. 25, pp. 5-10.
18. Hayton G.E., Pugh A.C., Fretwell P. (1988) "Infinite elementary divisors of a matrix polynomial and implications", *Internat. J. Control*, vol. 47, pp. 53-64.
19. Kailath T. (1980) *Linear Systems*, Prentice-Hall.
20. Karampetakis N.P. (2004) "On the solution space of discrete time AR-representations over a finite horizon", *Linear Algebra Appl.*, vol. 383, pp. 83-116.
21. Karampetakis N.P., Vardulakis A.I.G. (1993) "On the Solution Space of Continuous Time AR Representations", *Proc. European Control Conf.*, Groningen, The Netherlands, pp. 1784-1789.
22. Lam T.Y. (2001) *A First Course in Noncommutative Rings* (2nd ed.), Springer.
23. Lam T.Y. (1999) *Lectures on Modules and Rings*, Springer.
24. Lang S. (2002) *Algebra*, Springer.
25. Marinescu B., Bourlès H. (2003) "The Exact Model Matching Problem for Linear Time-Varying Systems: an Algebraic Approach", *IEEE Trans. on Automat. Control*, vol. 48, pp. 166-169.
26. Pugh A.C., Karampetakis N.P., Vardulakis A.I.G., Hayton G.E. (1994) "A Fundamental Notion of Equivalence for Linear Multivariable Systems", *IEEE Trans. on Automat. Control*, vol. 39, pp. 1141-1145.
27. Vardulakis A.I.G. (1991) *Linear Multivariable Control*, Wiley.
28. Vardulakis A.I.G., Limebeer D.J.N., Karkanias N. (1982) "Structure and Smith-MacMillan form of a rational matrix at infinity", *Internat. J. Control*, vol. 35, pp. 701-725.
29. Verghese, G. C. (1979) *Infinite-Frequency Behaviour in Generalized Dynamical Systems*, Ph. D. dissertation, Electrical Engineering Department, Stanford University.
30. Verghese G.C., Kailath T. (1981) "Rational Matrix Structure", *IEEE Trans. on Automat. Control*, vol. 18, pp. 220-225.
31. Vidyasagar M (1985) *Control System Synthesis –A Factorization Approach*, MIT Press.

A

On the Literature's Two Different Definitions of Uniform Global Asymptotic Stability for Nonlinear Systems

Andrew R. Teel[1] and Luca Zaccarian[2]

Dipartimento di Informatica, Sistemi e Produzione, University of Rome, Tor Vergata, 00133 Rome, Italy. E-mail: zack@disp.uniroma2.it

In this appendix we discuss two different definitions of uniform global asymptotic stability (UGAS), both used in the literature. In the first one, UGAS is defined to be uniform local stability (ULS) plus uniform global attractivity (UGA). In the second one, it is defined to be ULS+UGA plus uniform global boundedness (UGB). We reemphasize, by means of an explicit example, that UGB is not necessarily implied by ULS and UGA, even for smooth time-varying nonlinear systems where the right-hand side's derivative with respect to the state is bounded uniformly in time. Thus, the two definitions are truly different for nonlinear time-varying systems.

A.1 Different UGAS Definitions

Two definitions of uniform global asymptotic stability for the origin of a time-varying nonlinear system $\dot{x} = f(t, x)$ have appeared in the literature. Both definitions include the concept of *uniform local stability (ULS)*: for each $\varepsilon > 0$ there exists $\delta > 0$ such that, for any $t_\circ \geq 0$, $|x(t_\circ)| \leq \delta$ implies $|x(t)| \leq \varepsilon$ for all $t \geq t_\circ$. Both definitions also include the concept of *uniform global attractivity (UGA)*: for each pair of strictly positive real numbers (r, ε) there exists $T > 0$ such that, for any $t_\circ \geq 0$, $|x(t_\circ)| \leq r$ implies $|x(t)| \leq \varepsilon$ for all $t \geq t_\circ + T$. These two concepts comprise the definition of uniform global asymptotic stability given in [3, Definition 36.9],[1] [7, page 10], [9, page 143],

Work supported in part by AFOSR grant number F49620-03-1-0203, NSF grant number ECS-0324679, by ENEA-Euratom, ASI and MIUR under PRIN and FIRB projects.

[1] In [3, equation (36.10)], the author also states that UGAS "could be formulated" as ULS+UGA+UGB, but he doesn't provide a comparison between the two alternative definitions.

A. Loría et al. (Eds.): Advanced Topics in Control Systems Theory, LNCIS 328, pp. 285–289, 2006.

[1, Definition 3.6] and [4, Definition 3.2]. In [2], [6], [10, Definition 2.10], [11, Definition 9.5] and [5, Definition 4.4], the authors add the concept of *uniform global boundedness (UGB)*: for each $r > 0$ there exists $M > 0$ such that, for any $t_\circ \geq 0$, $|x(t_\circ)| \leq r$ implies $|x(t)| \leq M$ for all $t \geq t_\circ$.

The purpose of this note is to give an example of a time-varying system with right-hand side that is locally Lipschitz (or smooth) in (x, t) and locally Lipschitz in x uniformly in t that shows

$$\boxed{\text{ULS \& UGA \& UGB} \neq \text{ULS \& UGA.}} \tag{A.1}$$

In particular, the definitions mentioned above are truly different for nonlinear systems. It is worth mentioning that the two definitions agree for linear time-varying systems $\dot{x} = A(t)x$ with $A(\cdot)$ uniformly bounded, which implies that the right-hand side is globally Lipschitz in x uniformly in t. It also should be noted that Willems suggested the relationship (A.1) in [10] without giving a precise example to illustrate this fact, but hinting at a linear example with a right-hand side unbounded in time.

A.2 A Locally Lipschitz (Uniform-in-time) System that is ULS and UGA but not UGB

In the following example, we will guarantee uniform local stability by making the system uniformly asymptotically stable when the state value is less than one in magnitude. In particular, we use $\frac{d}{dt}|x(t)| \leq -|x(t)|^3$. Uniform global attractivity will be guaranteed by periodically enforcing intervals of time on which the system is able to converge to magnitude less than one from every initial condition. Again, $\frac{d}{dt}|x(t)| \leq -|x(t)|^3$ is sufficient. Uniform global boundedness will be precluded by periodically enforcing intervals of time on which the system is capable of escaping to infinity. For this, $\frac{d}{dt}|x(t)| \geq |x(t)|^3$ is sufficient. Actual finite escape times are prevented, without inducing UGB, by returning to $\frac{d}{dt}|x(t)| \leq -|x(t)|^3$ whenever the state exceeds a certain threshold that grows unbounded with time. More specifically, we consider a scalar time-varying system of the form

$$\dot{x} = f(t, x) := \varphi(t, x)x^3, \tag{A.2}$$

where $\varphi(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is defined based on the sets indicated in Figure A.1. It should be a bounded continuous function which is globally Lipschitz (or smooth with bounded gradient). Moreover, it should be greater than or equal to one in the black region, and less than or equal to minus one in the white region. Below we propose a globally Lipschitz section of $\varphi(\cdot, \cdot)$, in terms of the distance to the set of points indicated by the black region in Figure A.1. A more explicit, smooth construction is given in [8].

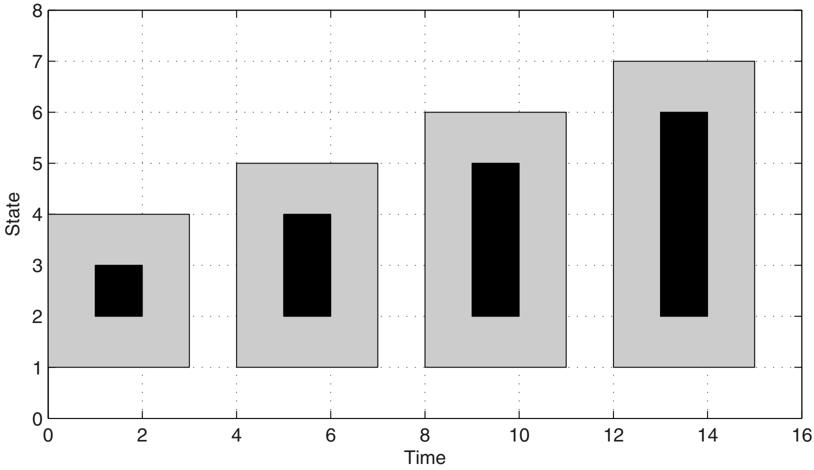

Fig. A.1. Domains of definition of the function $\varphi(\cdot,\cdot)$ in the (t, x) plane.

Based on the properties of $\varphi(\cdot,\cdot)$ it is straightforward to conclude the following:

- The origin of system (A.2) is ULS. Indeed in the unit ball around the set $\{(x,t) : x = 0\}$, the dynamics satisfies $\frac{d}{dt}|x(t)| \leq -|x(t)|^3$ so that $|x(t_\circ)| \leq 1$ implies $|x(t)| \leq |x(t_\circ)|$.
- the origin of system (A.2) is UGA since
  1. the time to converge to $[-1, 1]$ for any system that satisfies $\frac{d}{dt}|x(t)| \leq -|x(t)|^3$ is smaller than a half unit of time, no matter the size of the initial condition;
  2. the width of each vertical white strip is one time unit;
  3. the maximum time to reach the left of a vertical white strip is four units of time.
- the system is not UGB since
  1. the time to escape to infinity from $x_\circ = 2$ for any system that satisfies $\frac{d}{dt}|x(t)| \geq |x(t)|^3$ is less than a quarter unit of time;
  2. the width of each black rectangle is one time unit.

  In particular, for each $i \in \mathbb{Z}_{\geq 0}$ the trajectory starting at $(x(t_\circ), t_\circ) = (2, 4i + 1)$ satisfies $x(t_\circ + 0.5) > 3 + i$.

There are various ways to find a function $\varphi(\cdot,\cdot)$ with the properties mentioned below Equation (A.2). One way is in terms of the distance $d((t, x), \mathcal{B})$ to the black region $\mathcal{B}$. For example

$$\varphi(t,x) := 1 - 2\min\{1, d((t,x),\mathcal{B})\} \ , \tag{A.3}$$

has all of the desired properties. The simulated behavior of system (A.2), (A.3), using the ∞ norm to evaluate the distance $d((t,x),\mathcal{B})$, is shown in Figure A.2.

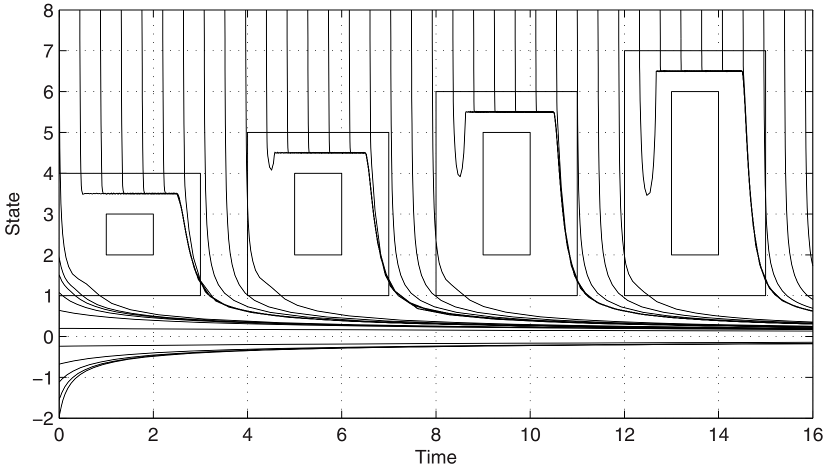

Fig. A.2. Trajectories of system (A.2), (A.3) starting from different initial conditions and times.

References

1. A. Bacciotti and L. Rosier. *Liapunov functions and stability in control theory.* Springer, The Netherlands, 2nd edition, 2005.
2. E.A. Barbashin and N.N. Krasovskii. On the existence of a function of Lyapunov in the case of asymptotic stability in the large. *Prikl. Mat. Meh.*, 18:345–350, 1954.
3. W. Hahn. *Stability of Motion.* Springer-Verlag, 1967.
4. H.K. Khalil. *Nonlinear Systems.* Prentice Hall, USA, 2nd edition, 1996.
5. H.K. Khalil. *Nonlinear Systems.* Prentice Hall, USA, 3rd edition, 2002.
6. J.L. Massera. Contributions to stability theory. *Ann. of Math.*, 64:182–206, 1956.
7. N. Rouche, P. Habets, and M. Laloy. *Stability Theory by Liapunov's Direct Method.* Springer-Verlag, New York, USA, 1977.
8. A.R. Teel and L. Zaccarian. On "uniformity" in definitions of global asymptotic stability for time-varying nonlinear systems. *Automatica*, 2006, submitted.
9. M. Vidyasagar. *Nonlinear Systems Analysis.* Prentice-Hall, Englewood Cliffs, New Jersey, 2nd edition, 1993.

10. J.L. Willems. *Stability theory of dynamical systems.* Thomas Nelson and sons, Great Britain, 1970.
11. T. Yoshizawa. *Stability theory and the existence of periodic solutions and almost periodic solutions.* Springer-Verlag, New York(USA), 1975.

Lecture Notes in Control and Information Sciences

Edited by M. Thoma and M. Morari

Further volumes of this series can be found on our homepage:
springer.com

Vol. 327: Fournier, J.-D.; Grimm, J.; Leblond, J.;
Partington, J.R. (Eds.)
Harmonic Analysis and Rational Approximation
301 p. 2006 [3-540-30922-5]

Vol. 326: Wang, H.-S.; Yung, C.-F.; Chang, F.-R.
H_∞ Control for Nonlinear Descriptor Systems
164 p. 2006 [1-84628-289-6]

Vol. 325: Amato, F.
Robust Control of Linear Systems Subject to Uncertain
Time-Varying Parameters
180 p. 2006 [3-540-23950-2]

Vol. 324: Christofides, P.; El-Farra, N.
Control of Nonlinear and Hybrid Process Systems
446 p. 2005 [3-540-28456-7]

Vol. 323: Bandyopadhyay, B.; Janardhanan, S.
Discrete-time Sliding Mode Control
147 p. 2005 [3-540-28140-1]

Vol. 322: Meurer, T.; Graichen, K.; Gilles, E.D. (Eds.)
Control and Observer Design for Nonlinear Finite
and Infinite Dimensional Systems
422 p. 2005 [3-540-27938-5]

Vol. 321: Dayawansa, W.P.; Lindquist, A.;
Zhou, Y. (Eds.)
New Directions and Applications in Control Theory
400 p. 2005 [3-540-23953-7]

Vol. 320: Steffen, T.
Control Reconfiguration of Dynamical Systems
290 p. 2005 [3-540-25730-6]

Vol. 319: Hofbaur, M.W.
Hybrid Estimation of Complex Systems
148 p. 2005 [3-540-25727-6]

Vol. 318: Gershon, E.; Shaked, U.; Yaesh, I.
H_∞ Control and Estimation of State-muliplicative
Linear Systems
256 p. 2005 [1-85233-997-7]

Vol. 317: Ma, C.; Wonham, M.
Nonblocking Supervisory Control of State Tree
Structures
208 p. 2005 [3-540-25069-7]

Vol. 316: Patel, R.V.; Shadpey, F.
Control of Redundant Robot Manipulators
224 p. 2005 [3-540-25071-9]

Vol. 315: Herbordt, W.
Sound Capture for Human/Machine Interfaces:
Practical Aspects of Microphone Array Signal Processing
286 p. 2005 [3-540-23954-5]

Vol. 314: Gil', M.I.
Explicit Stability Conditions for Continuous Systems
193 p. 2005 [3-540-23984-7]

Vol. 313: Li, Z.; Soh, Y.; Wen, C.
Switched and Impulsive Systems
277 p. 2005 [3-540-23952-9]

Vol. 312: Henrion, D.; Garulli, A. (Eds.)
Positive Polynomials in Control
313 p. 2005 [3-540-23948-0]

Vol. 311: Lamnabhi-Lagarrigue, F.; Loría, A.;
Panteley, E. (Eds.)
Advanced Topics in Control Systems Theory
294 p. 2005 [1-85233-923-3]

Vol. 310: Janczak, A.
Identification of Nonlinear Systems Using Neural
Networks and Polynomial Models
197 p. 2005 [3-540-23185-4]

Vol. 309: Kumar, V.; Leonard, N.; Morse, A.S. (Eds.)
Cooperative Control
301 p. 2005 [3-540-22861-6]

Vol. 308: Tarbouriech, S.; Abdallah, C.T.; Chiasson, J. (Eds.)
Advances in Communication Control Networks
358 p. 2005 [3-540-22819-5]

Vol. 307: Kwon, S.J.; Chung, W.K.
Perturbation Compensator based Robust Tracking
Control and State Estimation of Mechanical Systems
158 p. 2004 [3-540-22077-1]

Vol. 306: Bien, Z.Z.; Stefanov, D. (Eds.)
Advances in Rehabilitation
472 p. 2004 [3-540-21986-2]

Vol. 305: Nebylov, A.
Ensuring Control Accuracy
256 p. 2004 [3-540-21876-9]

Vol. 304: Margaris, N.I.
Theory of the Non-linear Analog Phase Locked Loop
303 p. 2004 [3-540-21339-2]

Vol. 303: Mahmoud, M.S.
Resilient Control of Uncertain Dynamical Systems
278 p. 2004 [3-540-21351-1]

Vol. 302: Filatov, N.M.; Unbehauen, H.
Adaptive Dual Control: Theory and Applications
237 p. 2004 [3-540-21373-2]

Vol. 301: de Queiroz, M.; Malisoff, M.; Wolenski, P. (Eds.)
Optimal Control, Stabilization and Nonsmooth Analysis
373 p. 2004 [3-540-21330-9]

Vol. 300: Nakamura, M.; Goto, S.; Kyura, N.; Zhang, T.
Mechatronic Servo System Control
Problems in Industries and their Theoretical Solutions
212 p. 2004 [3-540-21096-2]

Vol. 299: Tarn, T.-J.; Chen, S.-B.; Zhou, C. (Eds.)
Robotic Welding, Intelligence and Automation
214 p. 2004 [3-540-20804-6]

Vol. 298: Choi, Y.; Chung, W.K.
PID Trajectory Tracking Control for Mechanical Systems
127 p. 2004 [3-540-20567-5]

Vol. 297: Damm, T.
Rational Matrix Equations in Stochastic Control
219 p. 2004 [3-540-20516-0]

Vol. 296: Matsuo, T.; Hasegawa, Y.
Realization Theory of Discrete-Time Dynamical Systems
235 p. 2003 [3-540-40675-1]

Vol. 295: Kang, W.; Xiao, M.; Borges, C. (Eds)
New Trends in Nonlinear Dynamics and Control,
and their Applications
365 p. 2003 [3-540-10474-0]

Vol. 294: Benvenuti, L.; De Santis, A.; Farina, L. (Eds)
Positive Systems: Theory and Applications (POSTA 2003)
414 p. 2003 [3-540-40342-6]

Vol. 293: Chen, G. and Hill, D.J.
Bifurcation Control
320 p. 2003 [3-540-40341-8]

Vol. 292: Chen, G. and Yu, X.
Chaos Control
380 p. 2003 [3-540-40405-8]

Vol. 291: Xu, J.-X. and Tan, Y.
Linear and Nonlinear Iterative Learning Control
189 p. 2003 [3-540-40173-3]

Vol. 290: Borrelli, F.
Constrained Optimal Control
of Linear and Hybrid Systems
237 p. 2003 [3-540-00257-X]

Vol. 289: Giarré, L. and Bamieh, B.
Multidisciplinary Research in Control
237 p. 2003 [3-540-00917-5]

Vol. 288: Taware, A. and Tao, G.
Control of Sandwich Nonlinear Systems
393 p. 2003 [3-540-44115-8]

Vol. 287: Mahmoud, M.M.; Jiang, J.; Zhang, Y.
Active Fault Tolerant Control Systems
239 p. 2003 [3-540-00318-5]

Vol. 286: Rantzer, A. and Byrnes C.I. (Eds)
Directions in Mathematical Systems
Theory and Optimization
399 p. 2003 [3-540-00065-8]

Vol. 285: Wang, Q.-G.
Decoupling Control
373 p. 2003 [3-540-44128-X]

Vol. 284: Johansson, M.
Piecewise Linear Control Systems
216 p. 2003 [3-540-44124-7]

Vol. 283: Fielding, Ch. et al. (Eds)
Advanced Techniques for Clearance of
Flight Control Laws
480 p. 2003 [3-540-44054-2]

Vol. 282: Schröder, J.
Modelling, State Observation and
Diagnosis of Quantised Systems
368 p. 2003 [3-540-44075-5]

Vol. 281: Zinober A.; Owens D. (Eds)
Nonlinear and Adaptive Control
416 p. 2002 [3-540-43240-X]

Vol. 280: Pasik-Duncan, B. (Ed)
Stochastic Theory and Control
564 p. 2002 [3-540-43777-0]

Vol. 279: Engell, S.; Frehse, G.; Schnieder, E. (Eds)
Modelling, Analysis, and Design of Hybrid Systems
516 p. 2002 [3-540-43812-2]

Vol. 278: Chunling D. and Lihua X. (Eds)
H_∞ Control and Filtering of
Two-dimensional Systems
161 p. 2002 [3-540-43329-5]

Vol. 277: Sasane, A.
Hankel Norm Approximation
for Infinite-Dimensional Systems
150 p. 2002 [3-540-43327-9]

Vol. 276: Bubnicki, Z.
Uncertain Logics, Variables and Systems
142 p. 2002 [3-540-43235-3]

Vol. 275: Ishii, H.; Francis, B.A.
Limited Data Rate in Control Systems with Networks
171 p. 2002 [3-540-43237-X]

Vol. 274: Yu, X.; Xu, J.-X. (Eds)
Variable Structure Systems:
Towards the 21$^{\mathrm{st}}$ Century
420 p. 2002 [3-540-42965-4]

Vol. 273: Colonius, F.; Grüne, L. (Eds)
Dynamics, Bifurcations, and Control
312 p. 2002 [3-540-42560-9]

Vol. 272: Yang, T.
Impulsive Control Theory
363 p. 2001 [3-540-42296-X]

Vol. 271: Rus, D.; Singh, S.
Experimental Robotics VII
585 p. 2001 [3-540-42104-1]

Vol. 270: Nicosia, S. et al.
RAMSETE
294 p. 2001 [3-540-42090-8]

Vol. 269: Niculescu, S.-I.
Delay Effects on Stability
400 p. 2001 [1-85233-291-316]