Systems & Control: Foundations & Applications

Franco Blanchini
Stefano Miani

Set-Theoretic Methods in Control

Second Edition

Birkhäuser

Systems & Control: Foundations & Applications

More information about this series at http://www.springer.com/series/4895

Franco Blanchini • Stefano Miani

Set-Theoretic Methods in Control

Second Edition

Birkhäuser

Franco Blanchini
Mathematics and Computer Science
University of Udine
Udine, Italy

Stefano Miani
Electrical, Management and Mechanical Engineering
University of Udine
Udine, Italy

ISSN 2324-9749 ISSN 2324-9757 (electronic)
Systems & Control: Foundations & Applications
ISBN 978-3-319-17932-2 ISBN 978-3-319-17933-9 (eBook)
DOI 10.1007/978-3-319-17933-9

Library of Congress Control Number: 2015937619

Mathematics Subject Classification (2010): 26E25, 34D20, 37B35, 49L20, 93B03, 93B51, 93B52, 93C05, 93C30, 93D05, 93D09, 93D15, 93D30

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

To Ulla Tahir

–Franco Blanchini

To Christina, Giovanna, Pietro and Lorenza

–Stefano Miani

Preface

Many control problems can be naturally formulated, analyzed, and solved in a set-theoretic context. Sets appear naturally when three aspects, which are crucial in control systems design, are considered: constraints, uncertainties, and design specifications. Furthermore, sets are the most appropriate language to specify several system performances, for instance when we are interested in determining the domain of attraction, in measuring the effect of a persistent noise in a feedback loop or in bounding the error of an estimation algorithm.

From a conceptual point of view, the peculiarity of the material presented in this book lies in the fact that sets are not only terms of the formulation, but they play an active role in the solution of the problems as well. Generally speaking, in the control theory context, all the techniques which are theoretically based on some properties of subsets of the state-space could be referred to as set-theoretic methods. The most popular and clear link is that with the Lyapunov theory and positive invariance. Lyapunov functions are positive-definite energy-type functions of the state variables which have the property of being decreasing in time and are a fundamental tool to guarantee stability. Besides, their sublevel sets are positively invariant and thus their shape is quite meaningful to characterize the system dynamics, a key point which will be enlightened in the present book. The invariance property will be shown to be fundamental in dealing with problems such as saturating control, noise suppression, model-predictive control, and many others.

The main purpose of this book is to describe the set-theoretic approach for the control and analysis of dynamic systems from both a theoretical and practical standpoint. The material presented in the book is only partially due to the authors' work. Most of it is derived from the existing literature starting from some seminal works of the early 70s concerning a special kind of dynamic games. By its nature, the book has many intersections with other areas in control theory including constrained control, robust control, disturbance rejection, and robust estimation. None of these is fully covered, but for each of them we will present a particular view only. However, when necessary, the reader will be referred to specialized literature for a complementary reading.

The present work could be seen as a new book on Lyapunov methods, but this would not be an accurate classification. Although Lyapunov's name, as well as the string "set," will appear hundreds of times, our aim is that of providing a different view with respect to the existing excellent work, which typically introduces the invariance concept starting from that of Lyapunov function. Here, we basically do the opposite: We show how to synthesize Lyapunov functions starting from sets which are specifically constructed to face relevant problems in control.

Although the considered approach is based on established mathematical and dynamic programming concepts, it is apparent that the approach is far from being considered obsolete. The reason is that these methods, proposed several decades ago, were subsequently abandoned because they were clearly unsuitable for the limited computer technology of the time.

In the authors' mind, it was important to revise those techniques in a renewed light, especially in view of the modern computing possibilities. Besides, many connections with others theories which have been developed in recent years (often based on the same old ideas) have been pointed out.

Concerning the audience, the book is mostly oriented towards faculty and advanced graduate students. A good background on control-and-system theory is necessary to the reader to access the book. Although, for the sake of completeness, some of its parts are mathematically involved, the "hard-to-digest" initial mathematical digressions can be left to an intuitive level without compromising the reading and understanding of the sequel. To this aim, an introduction has been written to simplify as much as possible the comprehension of the book. In such chapter, the reasons for dealing with non-differentiable Lyapunov functions are discussed and preliminary examples are proposed to make the (scaring) notations of the following sections more reader-friendly. In the same spirit, many exercises have been put at the end of each chapter.

The present second edition is identical in spirit, but deeply revised in many parts. In particular, it includes new examples and ideas[1]. Many changes are due to the precious and constructive comments of many colleagues. The new edition presents a new chapter about switching systems, which was only a section of the chapter related topics.

The outline of the new book, depicted in the figure at the end of the present section, is as follows.

Basic mathematical notations and acronyms, an intuitive description of the main book content and the link with Lyapunov theory and Nagumo's theorem, are provided in Chapter 1.

In Chapter 2, Lyapunov's methods, including non-smooth functions and converse stability results, are detailed together with their connections with invariant set theory. Some links with differential games and differential inclusion theories are also indicated.

[1] And hopefully less mistakes.

Background material on convex sets and convex analysis, used in the rest of the book, is presented in Chapter 3.

Set invariance theory fundamentals are developed in Chapter 4 along with methods for the determination of appropriate invariant sets, essentially ellipsoids and polytopes, for dynamic systems analysis and design.

Dynamic programming ideas and techniques are presented in Chapter 5 and some algorithms for backward computation of Lyapunov functions are derived.

The ideas presented in Chapters 4 and 5 are at the basis of the following three chapters.

Their application to dynamic system analysis is reported in Chapter 6, where it is shown how to compute reachable sets and how these tools result extremely helpful in the stability and performance analysis of polytopic systems.

The control of parameter-varying systems by means of robust or gain-scheduled controllers is looked at in Chapter 7, where it is shown how to derive such controllers starting from quadratic or polytopic functions.

Time constraints are dealt with in Chapter 8. Special emphasis is put on controllability and reachability issues and on the computation of a domain of attraction under bounded or rate constrained inputs. An extension of such techniques to tracking problems is presented.

We dedicated a whole chapter to the problem of switching and switched systems. Definitely the relevant theory in this topic is much wider than the material proposed here. Still, we believe that the set-theoretic point of view of the subject can be inspiring for the reader.

Chapter 10 presents a set-theoretic solution to different optimal and sub-optimal control problems such as the minimum-time, the bounded disturbance rejection, the constrained receding horizon, and the recent relatively optimal control.

Basic ideas in the set-theoretic estimation area are reported in Chapter 11, where it is basically shown how it is possible to bound the error estimate via sets, though paying a high price in terms of computational complexity, especially when polytopes are to be considered.

Finally, some topics, which can be solved by set-theoretic methods, are presented in Chapter 12: adaptive control, estimation of the domain of attraction, switched and planar systems.

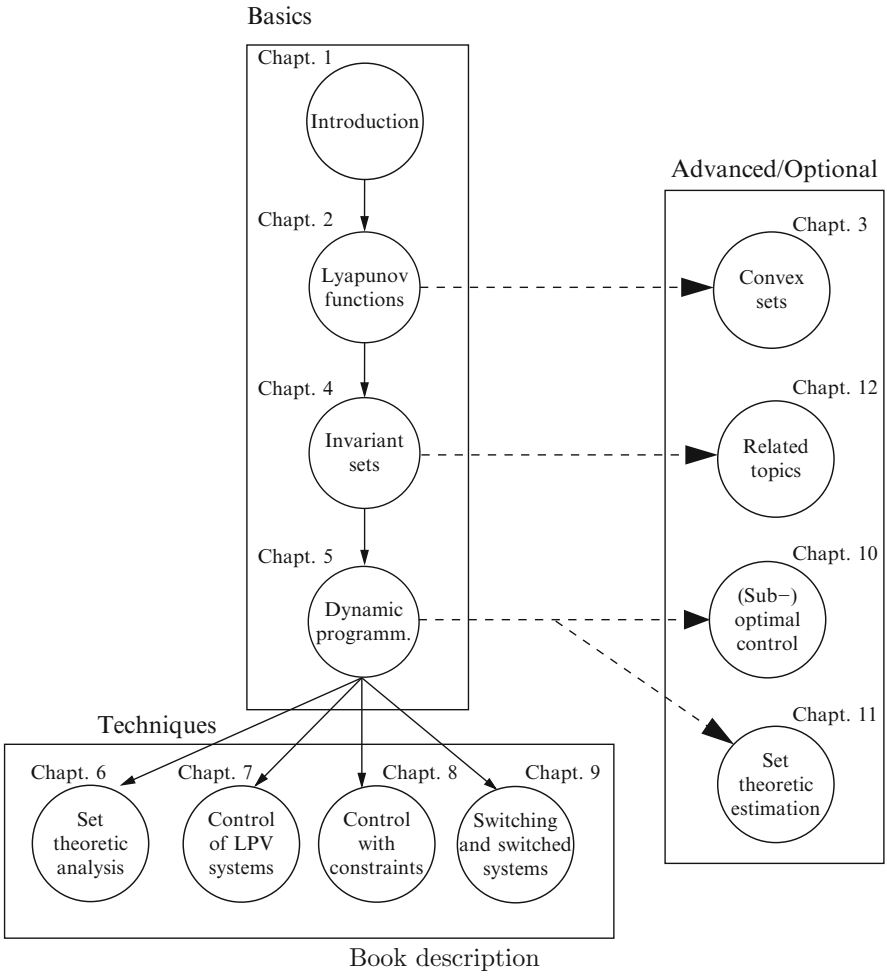
A concluding "Appendix" illustrates some interesting properties of the Euler auxiliary system, the discrete-time dynamic system which is used throughout the book in many proofs and the basic functioning of the numerical algorithm used for the backward computation of polytopes for linear parameter-varying systems.

There are many people the authors should thank (including the members of their own families) and a full citation would be impossible. Special thanks are due to Dr. Sasa Raković and to Prof. Fabio Zanolin, for their help. We also thank Prof. Maria Elena Valcher, Dr. Felice Andrea Pellegrino, Dr. Angelo Alessandri, Prof. Fouad Mesquine, Dr. Mirko Fiacchini, Prof. Patrizio Colaneri, Prof. Sorin Olaru, and Dr Sergio Grammatico for their constructive comments. We thank Dr. Carlo Savorgnan, from the University of Udine, who wrote the appendix on the MAXIS-G code. Finally the authors gratefully acknowledge the precious contribution of

Dr. Giulia Giordano in proofreading and improving the quality of the book during the writing of the second edition.

Udine, Italy
August 2014

Franco Blanchini
Stefano Miani



Contents

Chapter 1
Introduction

1.1 Notations

The book will cover several topics requiring many different mathematical tools. Therefore adopting a completely coherent notation is impossible. Several letters will have different meaning in different sections of the book. Coherence is preserved inside single sections as long as it is possible. Typically, but not exclusively, Greek letters α, β, ... will denote scalars, Roman letter a, b, ... vectors, Roman capital letters A, B matrices, script letters $\mathcal{A}$, $\mathcal{B}$, ... sets. A_i will denote both the ith row or the ith column of matrix A. Besides the standard mathematical conventions, the following notations will be used.

- $\mathbb{R}$ is the set of real numbers.
- $\mathbb{R}_+$ is the set of non-negative real numbers.
- A^T denotes the transposed of matrix A.
- $eig(A)$ denotes the set of the eigenvalues of matrix A.
- Given function $\Psi : \mathbb{R}^n \to \mathbb{R}$ and $\alpha \leq \beta$, we denote the sets

$$\mathcal{N}[\Psi, \alpha, \beta] \doteq \{x : \ \alpha \leq \Psi(x) \leq \beta\}$$

and

$$\mathcal{N}[\Psi, \beta] \doteq \mathcal{N}[\Psi(x), -\infty, \beta] = \{x : \ \Psi(x) \leq \beta\}$$

- Given a smooth function $\Psi : \mathbb{R}^n \to \mathbb{R}$ its gradient $\nabla\Psi(x)$ is the column vector

$$\nabla\Psi(x) = \left[\frac{\partial\Psi}{\partial x_1}(x) \ \frac{\partial\Psi}{\partial x_2}(x) \ \ \ldots \ \ \frac{\partial\Psi}{\partial x_n}(x)\right]^T$$

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_1

- If $x, z \in \mathbb{R}^n$ we denote the directional upper derivative of $\Psi : \mathbb{R}^n \to \mathbb{R}$

$$D^+\Psi(x,z) = \lim_{h\to 0+} \sup \frac{\Psi(x+hz) - \Psi(x)}{h}$$

(in the case of a smooth function $\Psi(x)$ it simply reduces to $\nabla\Psi(x)^T z$). We will also (ab)use (of) this notation when a function $z = f(x,w,u)$ has to be considered and, to keep the notation simple, we will write

$$D^+\Psi(x,w,u) = D^+\Psi(x,f(x,w,u))$$

to mean the upper directional derivative with respect to $f(x,w,u)$.
- If A and B are matrices (or vectors) of the same dimensions, then

$$A < (\leq, >, \geq) B$$

has to be intended componentwise $A_{ij} < (\leq, >, \geq) B_{ij}$ for all i and j.
- In the space of symmetric matrices

$$Q \prec (\preceq, \succ, \succeq) P$$

denotes that $P - Q$ is positive definite (positive semi-definite, negative definite, negative semi-definite).
- We will denote by $\|\cdot\|$ a generic norm. We will use this notation in all cases in which specifying the norm is of no importance.
- More specifically, $\|x\|_p$, with integer $1 \leq p < \infty$, denotes the p-norm

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$$

and

$$\|x\|_\infty = \max_i |x_i|$$

- If $P \succ 0$ is a symmetric square matrix, then

$$\|x\|_P = \sqrt[2]{x^T P x}$$

- Given any vector norm $\|\cdot\|_*$, the corresponding induced matrix norm is

$$\|A\|_* \doteq \sup_{x\neq 0} \frac{\|Ax\|_*}{\|x\|_*}$$

- For $x \in \mathbb{R}^n$, the sign and saturation vector functions $\mathrm{sgn}(x)$ and $\mathrm{sat}(x)$ are defined, respectively, by the component-wise assignations

$$[\mathrm{sgn}(x)]_i \doteq \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{if } x_i = 0 \\ -1 & \text{if } x_i < 0 \end{cases}$$

$$[\mathrm{sat}(x)]_i \doteq \begin{cases} x_i & \text{if } |x_i| \leq 1 \\ \mathrm{sgn}(x_i) & \text{if } |x_i| > 1 \end{cases}$$

  The saturation function can be generalized to the weighted case $\mathrm{sat}_a[x]$ or the unsymmetrical case where $\mathrm{sat}_{a,b}[x]$ where a and b are vectors as follows:

$$[\mathrm{sat}_{a,b}(x)]_i \doteq \begin{cases} x_i & \text{if } a_i \leq x_i \leq b_i \\ a & \text{if } x_i < a_i \\ b & \text{if } x_i > b_i \end{cases}$$

  and $\mathrm{sat}_a[x] \doteq \mathrm{sat}_{-a,a}[x]$.
- With a slight abuse of notation, we will often refer to a function $y(\cdot) : \mathbb{R}^q \mapsto \mathcal{Y} \subset \mathbb{R}^p$ by writing "the function $y(t) \in \mathcal{Y}$" or even $y \in \mathcal{Y}$, if the meaning is clear from the context.
- A locally Lipschitz function $\Psi : \mathbb{R}^n \to \mathbb{R}$ is positive definite if $\Psi(0) = 0$ and $\Psi(x) > 0$ for all $x \neq 0$. It is positive semi-definite if the strict inequality is replaced by the weak one. A function $\Psi(x)$ is negative (semi-)definite if $-\Psi(x)$ is positive (semi-)definite.

  The above definitions admit local versions in a neighborhood $\mathcal{S}$ of the origin. In this case, the statement "for all $x \neq 0$" is replaced by "for all $x \in \mathcal{S}, x \neq 0$".

1.1.1 Acronyms

In the paper, very few acronyms will be used, with few exceptions. We report some of the acronyms next.

EAS	Euler Auxiliary System;
LMI(s)	Linear Matrix Inequality (Inequalities);
LPV	Linear Parameter-Varying;
RAS	Region of Asymptotic Stability;
DOA	Domain of Attraction;
GUAS	Globally Uniformly Asymptotically Stable;
UUB	Uniformly Ultimately Bounded.

1.2 Basic ideas and motivations

The goal of this book is providing a broad overview of important problems in system analysis and control that can be successfully faced via set-theoretic methods.

1.2.1 The spirit of the book

We immediately warn the reader who is mainly interested in plug-and-play solutions to problems or in "user friendly" recipes for engineering problems that she/he might be partially disappointed by this book. The material presented in most parts of the book is essentially conceptual. By no means the book lacks numerical examples and numerical procedures presented in detail. But it turns out that in some cases the provided examples evidence the limits of the theory, especially from a computational standpoint, if approached with a "toolbox" spirit. However, we hope that if the reader will be patient enough to read the following subsections, she/he will be convinced that the book can provide an useful support.

The set-theoretic approach applies naturally in many contexts in which its language is essential even to state the problem. Therefore the set-theoretic framework is not only a collection of methods, but it is mainly a natural way to formulate study and solve problems.

As a simple example consider the "problem" of actuator limitations whose practical meaning is out of questions. The main issue in this regard is indeed: how to formulate the "problem" in a meaningful way. It is known that, as long as a controlled system state is close to the desired equilibrium point, actuator limitation is not an issue at all. Clearly, troubles arise when the state is "far" from the target. However, to properly formulate the problem in an engineering spirit, one must decide what "far" means and provide the problem specification. A possible way to proceed is the typical analysis problem in which a control is fixed and its performance is evaluated by determining the domain of attraction under the effect of saturation. If one is interested in a synthesis problem, then a possible approach is trying to find a controller which includes a certain initial condition or a set of initial conditions in its domain of attraction. A more ambitious problem is determining a controller which maximizes the domain of attraction. From the above simple problem formulation it is apparent that the set of states which can be brought to the origin, the domain of attraction, is essential in the problem specification. The same considerations can be done if output or state constraints are considered, since a quite natural requirement is to meet the constraints for specified initial conditions. As it will be seen, this is equivalent to requiring that these initial states belong to a proper set in the state space which is a domain of attraction for the closed-loop system.

The problem of constrained control can be actually solved in the disturbance rejection framework by seeking a stabilizing compensator which guarantees constraint satisfaction when a certain disturbance signal (or a class of disturbance

signals) is applied with zero initial conditions. Although, in principle, no sets are involved at all in this problem, there are strong relations with the set-theoretic approach. For instance, if one considers the tracking problem of reaching a certain constant reference without constraint violation, the problem can be cast in the set-theoretic language after state translation, by assuming the target state as the new origin and by checking if the initial state (formerly the origin) is in the domain of attraction.

In other contexts, such as the rejection of unknown-but-bounded disturbances under constraints, the set-theoretic approach plays a central role. Indeed, classical results on dynamic programming show how the problem of keeping the state inside prescribed constraint-admissible sets under the effect of persistent unknown-but-bounded disturbances can be formulated and solved exactly (up to computational complexity limits) in the set-theoretic framework.

The set-theoretic language, beside being the natural one to state several important problems, also provides the natural tool for solving them as, for instance, in the case of uncertain systems with unknown-but-bounded time-varying parameters, for which Lyapunov theory plays a fundamental role. One key point of Lyapunov's work is that the designer has to choose a class in which a candidate Lyapunov function needs to be found. Several classes of functions are available and, without doubts, the most popular are those based on quadratic forms. Very powerful tools are available to handle these functions. However, it is known (as it will clearly be evidenced) that quadratic functions have strong theoretical limitations. Other classes of functions, for instance the polyhedral ones, do not have such limitations and several methods to compute them are based on the set-theoretic approach, as we will see later.

In this book, several problems will be considered, without privileging any of them. Indeed we are describing tools that can be exploited in several different situations (although, for space reasons, some of these will be only sketched).

1.2.2 Solving a problem

It is quite useful to briefly dwell on the statement "solving a problem," since this is often used with different meanings. As long as we are talking about a problem which is mathematically formulated, a distinction has to be made between its "general formulation" and "the instance of the problem," being the latter referred to a special case, namely to a problem with specific data. When we say that a problem "is solved" (or can be "solved") we are referring to the general formulation. For instance, the analytic integration problem which consists in finding a primitive of a function is a generically unsolved problem although many special instances ($\int x dt = x^2/2 + C$) are solvable.

We could discuss for years on the meaning of solving a problem. Physicists, doctors, mathematicians, and engineers have different feelings about this. Therefore we decided to insert a "pseudo-definition" of problem solving in order to clarify our approach.

Our pseudo-definition of "solving a problem" sounds as follows.

We say that a given problem, mathematically formulated, is solved if there exists an algorithm that can be implemented on a computer such that, given any instance of the problem, in a finite number of steps (no matter how many) it leads to one of the following conclusions:

- the instance can be solved (and, hopefully, a solution is provided);
- there is no solution with the given data.

The discussion here would be almost endless since nothing about the computability has been said, and indeed computability will not be the main issue of the book. Certainly, we will often consider the computational complexity of the proposed algorithms, but we will not assume that an "algorithm" must necessarily possess good "computational complexity," as, for instance, that of being solvable in a time which is a polynomial function of the data dimensions.

We remark that, although we absolutely do not underestimate the importance of the complexity issue, complexity will not be considered of primary importance in this book. Basically we support this decision by two considerations:

- if we claimed that a problem can be solved if there exists a polynomial algorithm, then we would implicitly admit that the major part of the problems is unsolvable;
- complexity analysis is quite useful in all disciplines in which large instances are the normal case (operation research and networks). We rather believe that this is not the case of control area.

Unfortunately, as it will be shown later, finding tools which solve a problem in a complete way requires algorithms that can be very demanding from a computational viewpoint and therefore complexity aspects cannot be completely disregarded. In particular, the issue of the trade-off between conservativeness and complexity, that will be discussed next, will be a recurring theme of the work.

1.2.3 Conservative or intractable?

Constructive control theory is based on mathematical propositions. Typical conditions have the form "condition $\mathbf{C}$ implies property $\mathbf{P}$" or, in lucky cases, "condition $\mathbf{C}$ is equivalent to property $\mathbf{P}$," where $\mathbf{P}$ is any property pertaining to a system and $\mathbf{C}$ is any checkable (at least by means of a computer) mathematical condition. Clearly, when the formulation is of the equivalence type (often referred to as characterization), the control theoretician is more satisfied. For instance, for linear discrete-time and time-invariant systems, stability is equivalent to the state matrix having eigenvalues with modulus strictly less than 1. This condition is often called a characterization, since the family of asymptotically stable systems is the same family of systems whose matrix A has only eigenvalues included in the open unit disk.

There is a much simpler condition, which can be stated in terms of norms and which states that a discrete-time linear system is asymptotically stable if $\|A\| < 1$, where $\|\cdot\|$ is any matrix induced norm, for instance $\|A\|_\infty \doteq \max_i \sum_j |A_{ij}|$. This gain-type condition is generically preferable since, from the computational point of view, it is easier to compute the norm of A rather than its eigenvalues. However, the gain condition is a sufficient condition only, since if $\|A\| \geq 1$ nothing can be inferred about the system stability, as in the next case

$$A(\mu, \nu) = \begin{bmatrix} 0 & \mu \\ \nu & 0 \end{bmatrix}$$

In the book we will say that a criterion based on a condition $\mathbf{C}$ is *conservative* to establish property $\mathbf{P}$ if $\mathbf{C}$ implies $\mathbf{P}$, but it is not equivalent to. In lucky cases it is possible to establish a measure of conservativeness. We say that a criterion based on a condition $\mathbf{C}$ is *arbitrarily conservative* to establish property $\mathbf{P}$ if, besides being conservative, there are examples in which condition $\mathbf{C}$ is "arbitrarily violated," but still property $\mathbf{P}$ holds. This is the case of the previous example, since $\|A(\mu, \nu)\|_\infty = \max\{\mu, \nu\}$ so $\|A(\mu, \nu)\| < 1$ can be arbitrarily violated (for instance for $\nu = 0$ and arbitrarily large μ) and still the matrix could be asymptotically stable. If we can "measure the violation," then we can also measure conservativeness.

The counterpart of conservativeness is intractability. Certainly the example provided is not so significant since computing the eigenvalues of a matrix is not a problem as long as the computers work. But we can easily be trapped in the complexity issue if we consider a more sophisticated problem, for instance establishing the stability of a system of the form $x(k+1) = A(w(k))x(k)$ where $A(w(k))$ takes its values in the discrete set $\{A_1, A_2\}$ (this is a switching system, a family that will be considered in the book). Since $A(w(k))$ is time-varying, the eigenvalues play a marginal role[1]. Conversely, the condition $\|A(w)\| < 1$ remains valid as a conservative sufficient condition for stability. If we are interested in a non-conservative (sufficient and necessary) condition, we can exploit the following result: $x(k+1) = A(w(k))x(k)$ is stable if and only if there exists a full column rank matrix F such that the norm $\|A\|_F \doteq \|FA\|_\infty \leq 1$ [Bar88a, Bar88b, Bar88c]. The matrix F can be numerically computed and it will be shown how to manage the computation via set-theoretic algorithms. However, it will also be apparent that the number of rows forming F, which depends on the problem, can be very large. Actually it turns out that the problem of establishing stability of $x(k+1) = A(w(k))x(k)$ or, equivalently, computing the spectral radius of the pair $\{A_1, A_2\}$, is computationally intractable [TB97].

It is known that computer technology has improved so much[2] that hard problems can be faced in at least reasonable instances. However, there is a further issue. Assume that we are considering a design problem and we are interested in finding an optimal compensator. Assume that we can spend two days and two nights in

[1] $|\lambda| < 1$ for $\lambda \in \sigma(A_1) \bigcup \sigma(A_2)$ is a necessary condition only.

[2] Otherwise this book would not have reason to exist.

computing a compensator of order 200 which is "optimal." It is expected that no one (or few people) will actually implement this compensator since in many cases she/he will be satisfied by a simple compensator, for instance a PID. We can use the hard solutions to evaluate the approximate solutions. It is almost a paradox, but our approach is supported by considering that recurrent situations, such as the one described below, in which it happens that reasonably simple solutions are quite close to the optimal ones.

Quite frequently, situations of this kind arise: an "optimal" compensator of order 200 is computed and it is then established that by means of a PID one can achieve a performance which is 5% worse than the optimal one so that, seemingly, the "optimal control evaluation" has been almost useless. But we can find a solid argument (and a very good motivation to proceed): we should be happy to use a simple PID based controller because, thanks to the fact that the optimal solution was found, we are now aware *of the limits of performance*, and that the PID is just 5% sub-optimal.

However, there are cases in which the simple solution is not so close to the "optimal" one and therefore it is reasonable and recommendable to seek for a compromise. This typically happens in the case of linear compensators which normally suffer from the fact that they "react proportionally" to the distance of the state from the target point, which is known to be a source of performance degradation. If a high gain feedback is adopted, then the compensator works smoothly when close to the origin but the performance can deteriorate in transients of large magnitude. Conversely, reducing the gain to limit the saturation leads to a weak action. A simple way to overcome the problem is to use small gains when far from the origin and large gain as the origin is approached. This can be achieved in several ways, for instance by switching among compensators, and in this case the switching law must be suitably coordinated, as we will see later on, to ensure the stability of the scheme.

1.2.4 How to avoid reading this book

The book is not structured as a manual or a collection of recipes to be accessed in case of specific problems, but rather the opposite: it is a collection of ideas and concepts. Organizing and ordering them has certainly been the major effort as it will be pointed out soon.

To avoid a waste of time to the reader who is not interested in the details, we have introduced Section 1.3, in which the essentials of the book are presented in form of a summary with examples. Therefore, Section 1.3 could be very useful to decide whether to continue or to drop further reading[3]. In such a section we have sketched, in a very intuitive way, which is the context of the book, which are the main results and concepts and which is the spirit of the presentation.

[3] With the hope that the final decision will be the former.

We think that accessing Section 1.3 could be sufficient at least to understand the basics of the message the authors are trying to send, even in the case of postponed (or abandoned) reading.

1.2.5 How to benefit from reading this book

If, eventually, the decision is to read, we would like to give the reader some hints:

- Do not be too scared by the mathematics you will find at the beginning. It has been introduced for the sake of completeness. For instance, if you do not like the Dini superior derivative just think in terms of regular derivative of a differentiable function.
- Do not be too concerned with proofs. We have, clearly, inserted them (or referred to available references), but we have not spent too much effort in elegance. We have rather concentrated on enlightening the main ideas.
- If you find the book interesting, please give a look at the exercises at the end of each chapter while reading. We have tried our best to stimulate ideas.
- Please note that a strong effort has been put in emphasizing the main concepts. We could not avoid the details, but do not sacrifice time to follow them if this compromises the essential.
- Always remind that we are humans and therefore error-prone. We are 100% sure that the book will include errors, questionable sentences, or opinions.

1.2.6 Past work referencing

This has been a crucial aspect, especially in view of the fact that the book includes material which has been known for more than 40 years. As the reader can see, the reference list is full of items, but we assume as an unavoidable fact that some relevant references will be missing. This is certainly a problem that can have two types of consequences:

- misleading readers, who will ignore some work;
- disappointing authors, who will not see their work recognized.

The provided references are our good-faith best knowledge of the literature (up to errors or specific decisions of not including some work for which we will accept the responsibility).

The first edition of the book reported the following sentence: “Clearly, any comment/remark/complain concerning forgotten of improperly cited references will be very much appreciated.” We did not receive many complaints, but certainly we discovered many references which should have been included but they were not. We added all of them and we further added many references of work subsequent to the first edition. We were also happy to notice that many new papers found a theoretical support from our work. In any case the reported sentence remains valid.

1.3 Outline of the book

We generically refer to all the techniques which exploit properties of suitably chosen or constructed sets in the state space as set-theoretic methods. The set-theoretic approach appears naturally or can be successfully employed in many problems of different nature. As a consequence, it was absolutely no obvious how to present the material and how to sequence the chapters (actually this was the major concern in structuring this work). Among the several aspects which are related to the set-theoretic approach, the dominant one is certainly Lyapunov theory which is considered next. Other fundamental issues are constrained control problems and robust analysis and design.

1.3.1 The link with Lyapunov theory

Lyapunov theory is inspired by the concept of energy and energy-dissipation (or preservation). The main idea of the theory is based on the fact that if an equilibrium point of a dynamical system is the local minimum of an energy function and the system is dissipative, then the equilibrium is (locally) stable. There is a subsequent property that comes into play, and more precisely the fact that the sublevel sets of a Lyapunov function $\Psi(x)$ (i.e., the sets $\mathcal{N}[\Psi, \kappa] = \{x : \Psi(x) \leq \kappa\}$) are positively invariant for κ small enough. This means that if the initial state is inside one of this sets at time t, then it will be in the set for all $t' \geq t$.[4] This fact turns out to be very useful in many applications which will be examined later on.

The concept of positive invariance is, in principle, not associated with a Lyapunov function. There are examples of invariant sets that do not derive from any Lyapunov function. Therefore the idea of set-invariance can originate a theory which is much more general than Lyapunov theory. For instance, the standard definition of a Lyapunov function requires positive definiteness. As a consequence the sublevel sets $\{x : \Psi(x) \leq \kappa\}$, for $\kappa > 0$ are bounded sets which include the origin as an interior point. But this is not necessary in many problems in which suitable invariant sets do not need to have (or even should not have) this property. A Lyapunov function is typically used to assure stability or boundedness of the solution of a system. An unstable system admits positively invariant sets. For instance, Chetaev type of criteria to establish instability are based on the existence of suitable positively invariant sets. As we will see in Section 4, invariant sets are sometimes associated or represented by means of the so-called Lyapunov-like functions.

But there are cases in which no functions at all are involved. Consider, for instance, the case of a positive system, precisely a system such that, if the initial

[4] If $t = 0$, then it will belong to the set for positive values of t', hence the name "positive invariance."

state has non-negative components, then the same property is preserved by the future states. This property can be alternatively stated by claiming that the state-space positive orthant is positively invariant. It is clear that the claim has no stability implications, since a positive system can be stable or unstable. Still the positive invariance conditions are quite close (at least from a technical standpoint) to the known derivative conditions in Lyapunov theory.

We borrow a simple preliminary example from nonlinear mechanics.

Example 1.1. Consider the following nonlinear system

$$\ddot{\theta}(t) = \alpha \sin(\mu\theta(t)) - \beta \sin(\nu\theta(t))$$

A standard procedure to investigate the behavior of the system is multiplying both members by $\dot{\theta}$

$$\dot{\theta}\ddot{\theta} - \alpha \sin(\mu\theta(t))\dot{\theta} + \beta \sin(\nu\theta(t))\dot{\theta} = 0$$

and integrating the above so as to achieve

$$\Psi(\theta, \dot{\theta}) \doteq \frac{1}{2}\dot{\theta}^2 + \frac{\alpha}{\mu}\cos(\mu\theta(t)) - \frac{\beta}{\nu}\cos(\nu\theta(t)) = C$$

This means that $\Psi(\theta, \dot{\theta})$ is constant along any system trajectory and thus, a qualitative investigation of such trajectories can be obtained by simply plotting the level curves of the function Ψ in the θ–$\dot{\theta}$ space. For $\alpha = 2$, $\beta = 1$, $\mu = 2$, and $\nu = 1$, the level curves are depicted in Figure 1.1. From the picture it can be inferred that the equilibrium point $\theta = 0$ is unstable (a conclusion that can be derived via elementary analysis), and that there are other two equilibrium points which are stable (not asymptotically), $(\pm\bar{\theta}, 0)$, with $\bar{\theta} \approx 1.3$. However, if the system is initialized close to the origin, there are two types of trajectories. For instance, if

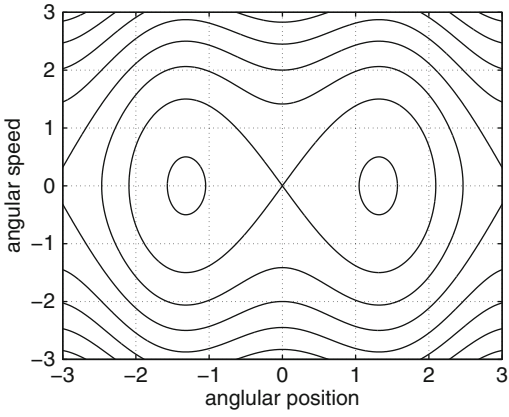


Fig. 1.1 The level surfaces of the function Ψ

$\theta(0) = \epsilon$ $(-\epsilon)$, $\epsilon > 0$ and $\dot{\theta}(0) = 0$, then the system trajectories are periodic and encircle the left (right) equilibrium point. Conversely, for any initial condition $\theta(0) = 0$ and $\dot{\theta}(0) = \epsilon$ $(-\epsilon)$, the trajectory encircles both equilibria.

The type of investigation in the example can be clearly extended to cases which are not so lucky and the property that the trajectories evolve along the set $\Psi = C$ is not true anymore. The invariance property of some suitably chosen set can provide useful information about the qualitative behavior. For instance, if a damping is introduced in the nonlinear system

$$\ddot{\theta}(t) = \alpha \sin(\mu\theta(t)) - \beta \sin(\nu\theta(t)) - \gamma\dot{\theta}(t)$$

one gets

$$\frac{d}{dt}\Psi(\theta, \dot{\theta}) = -\gamma\dot{\theta}(t)^2$$

so that, due to the energy dissipation, the system will eventually "fall" in one of the stable equilibrium points (with the exception of a zero-measure set of initial conditions from which the state converges to the origin in an unrealistic behavior).

The next natural question concerning the link between set invariance and Lyapunov theory is the following: since the existence of a Lyapunov function implies the existence of positively invariant sets, is the opposite true? More precisely, given an invariant set, is it possible to derive a Lyapunov function from it? The answer is negative, in general. For instance, for positive systems the positive orthant is invariant, but it does not originate any Lyapunov function. However, for certain classes of systems (e.g., those with linear or affine dynamics) it is actually possible to derive a Lyapunov function from a compact invariant set which contains the origin in its interior, as in the following example.

Example 1.2. Consider the linear system $\dot{x} = Ax$ with

$$A = \begin{bmatrix} -1 & \alpha \\ -\beta & -1 \end{bmatrix}$$

where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are uncertain, constant, parameters. To check whether this system is stable for any choice of the values of α and β in the given range, it is sufficient to consider the unit circle (actually any circle) and check whether it is positively invariant. An elementary way to achieve this is to use the Lyapunov function $\Psi(x) = x^T x/2$ and notice that the Lyapunov derivative

$$\dot{\Psi}(x) = x^T\dot{x} = x^T Ax = -x_1^2 - x_2^2 + (\alpha - \beta)x_1 x_2 < 0$$

for $(x_1, x_2) \neq 0$. An interpretation of the inequality can be deduced from Figure 1.2 (left). The time derivative of $\Psi(x(t))$ for x on the circle is equal to the scalar product between the gradient, namely the vector which is orthogonal to the circle surface and

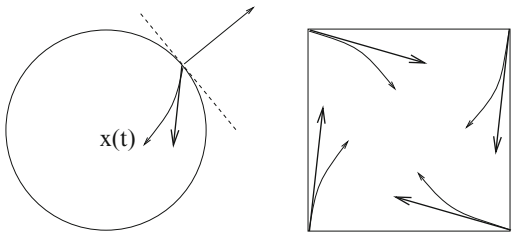


Fig. 1.2 The subtangentiality conditions for the circle and the square

points outside, and the velocity vector $\dot{x}$ (represented by a thick arrow in the figure). Intuitively, the fact that such a scalar product is negative, namely that the derivative points inside, implies that any system trajectory originating on the circle surface goes inside the circle (the arrowed curve). This condition will be referred to as subtangentiality condition.

Up to now, standard quadratic functions have been considered together with standard derivatives. As an alternative, one might think about, or for some mysterious reasons be interested in, other shapes. If, for example, a unit square is investigated, it is possible to reason in a similar way but with a fundamental difference: *the unit square has corners!* An important theorem, due to Nagumo in 1942 [Nag42], comes into play. Consider the right top vertex of the square which is $[1\ 1]^T$ (the other three vertices can be handled in a similar way). The corresponding derivative is

$$\dot{x} = \begin{bmatrix} (-1+\alpha) \\ (-1-\beta) \end{bmatrix},$$

which "points towards the interior of the square" as long as $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$. Intuitively, this means that any trajectory passing through the vertex "goes inside." It is also very easy to see that for any point of any edge the trajectory points inside the square. Consider, for instance, any point $[x_1\ x_2]^T$ on the right edge, $x_1 = 1$ and $|x_2| \leq 1$. The time derivative of x_1 results in

$$\dot{x}_1 = -x_1 + \alpha x_2 \leq -1 + \alpha|x_2| \leq -1 + \alpha \leq 0.$$

This means that $x_1(t)$ is non-increasing when the state x is on the right edge, so that no trajectory can cross it from the left to right. By combining the edge and the vertex conditions, one can expect that no trajectory originating in the square will leave it (it will be shown later that for linear uncertain systems one needs only to check "vertex conditions"). It is rather obvious that, by homogeneity, the same consideration can be applied to any scaled square. This fact allows to consider the norm

$$\Psi(x) = \|x\|_\infty = \max_i |x_i|$$

which is such that $\Psi(x(t))$ is non-increasing along the trajectories of the system and then results in a Lyapunov function for the system. Given that different shapes (say,

different level sets) can be considered, the obvious next question is the following: how can we deal with this kind of functions since $\Psi(x)$ is non-differentiable and the standard Lyapunov derivative cannot be applied? We will reply to this question in two ways. From a theoretical standpoint we will introduce a powerful tool, the Dini derivative, which is suitable for locally Lipschitz Lyapunov functions (therefore including all kinds of norms). From a practical standpoint, it will be shown that for the class of piecewise linear positive definite functions there exist *linear programming* conditions which are equivalent to the fact that $\Psi(x(t))$ is non-increasing. These conditions are basically derived by the same type of analysis sketched before and performed on the unit ball of Ψ (this type of Lyapunov functions are called set-induced).

1.3.2 Uncertain systems

Enlightening the importance of Lyapunov theory for the analysis and control of uncertain systems is definitely not an original contribution of this book. However, the issue of uncertainty is of primary importance and it will be deeply investigated in the book. Uncertainty will be analyzed not only in the standard way (i.e., by Lyapunov second method) but also by means of a set-theoretic approach which will provide a broader view and will allow to face several problems which are not directly solvable by means of the standard Lyapunov theory. In particular, reachability and controllability problems under uncertainties and their applications will be considered. These problems will be faced by means of a dynamic programming approach. As a very simple example consider the next inventory problem.

Example 1.3. The following equation

$$x(k+1) = x(k) - d(k) + u(k)$$

represents a typical (and, probably, the simplest) inventory model. The variable u is the control representing the production rate, while d is the demand rate. The state variable x is the amount of stored goods. Consider the problem of finding a control u over a horizon $0, 1, \ldots, T-1$ such that, given $x(0) = x_0$, the following constraints will be satisfied: $0 \leq x(k)$, $x(T) = \bar{x}$, and $0 \leq u(k) \leq \bar{u}$. If $d(k)$ is assumed to be a known function, then the problem is a standard reachability problem in which the following constraints have to be taken into account

$$x(k) = x_0 - \sum_{i=0}^{k-1} d(i) + \sum_{i=0}^{k-1} u(i) \geq 0$$

along with the control constraints $0 \leq u(k) \leq \bar{u}$ and the final condition $x(T) = \bar{x}$. If we assume $d(k)$ uncertain and we adopt an *unknown-but-bounded* uncertainty specification, for instance $d^-(k) \leq d(k) \leq d^+(k)$, then the scenario changes

completely. Three kinds of policies can be basically considered. The first is the open-loop strategy, in which the whole sequence is chosen as a function of the initial state $u(\cdot) = \Phi(x_0)$. The second is the state feedback strategy, precisely $u(k) = \Phi(x(k))$ while the third is the full information strategy, $u(k) = \Phi(x(k), d(k))$, in which the controller is granted the knowledge of $d(k)$, at the current time. These three strategies are strictly equivalent if d is known in advance, in the sense that if the problem is solvable by one of them then it is solvable by the other two, but under uncertainty the situation changes. It is immediate that only the third type of strategy can lead to the terminal goal $x(T) = \bar{x}$. To hope to produce something useful by means of the other two strategies we have to relax our request to a more reasonable target like $|x(T) - \bar{x}| \leq \beta$, where β is a tolerance factor.

The open-loop problem can then be solved if and only if one can find an open-loop sequence such that $0 \leq u(k) \leq \bar{u}$ and

$$x(k) = x_0 - \sum_{i=0}^{k-1} d^+(i) + \sum_{i=0}^{k-1} u(i) \geq 0$$

$$x(T) = x_0 - \sum_{i=0}^{T-1} d^+(i) + \sum_{i=0}^{T-1} u(i) \geq \bar{x} - \beta$$

$$x(T) = x_0 - \sum_{i=0}^{T-1} d^-(i) + \sum_{i=0}^{T-1} u(i) \leq \bar{x} + \beta$$

In this case the solution is simple since, in view of the fact that the problem is scalar, one can consider the "worst case" action of the disturbance (which is $d^+(i)$ for the upper bound and $d^-(i)$ for the lower bound). For multi-dimensional problems, the situation is more involved because there is no clear way to detect the "worst case." Then a possibility, in the linear system case, is to compute the "effect of the disturbance," namely the reachability set at time k, with $u = 0$ and $x_0 = 0$. In this simple case we have that such a set is

$$\mathcal{D}_i = \left\{ \sum_{i=0}^{k-1} d(i), \quad \text{for all possible sequences } \ d(i) \right\}$$

namely the interval $\left[\sum_{i=0}^{T-1} d^-(i), \sum_{i=0}^{T-1} d^+(i)\right]$ (as we will see the situation is more involved in the general case). Then, for instance, non-negative constraint satisfaction reduces to the condition

$$x_0 - \delta + \sum_{i=0}^{k-1} u(i) \geq 0, \quad \text{for all } \ \delta \in \mathcal{D}_i$$

We will see that this kind of trick is very useful in model predictive control (a technique which embeds an open-loop control computation in a feedback scheme) in the presence of uncertainties.

The feedback problem is more involved and the solution procedure works as follows. Consider the set of all non-negative states at time $T-1$ which can be driven in one step to the interval $[x_T^-, x_T^+] \doteq [\bar{x}-\delta, \bar{x}+\delta]$. This is the set

$$\begin{aligned}\mathcal{X}_{T-1} = \{x \geq 0 : \exists u,\ 0 \leq u \leq \bar{u},\ \text{such that } x - d + u \in [x_T^-, x_T^+],\\ \forall d^-(T-1) \leq d \leq d^+(T-1)\}\end{aligned}$$

It will be seen that such a set is convex. In the scalar case it is an interval $\mathcal{X}_{T-1} = [x_{T-1}^-, x_{T-1}^+]$ (the impatient reader can have fun in determining the extrema). Once $\mathcal{X}_{T-1}$ has been computed, the procedure repeats exactly backward by determining the set of all non-negative states at time $T-2$ that can be driven in one step to the interval $\mathcal{X}_{T-1} = [x_{T-1}^-, x_{T-1}^+]$ and so on.

It is apparent that the control strategy requires two stages:

- **off-line stage**: the sequence of sets $\mathcal{X}_k$ is sequentially determined backward in time and stored;
- **on-line stage**: the sets of the sequence, $\mathcal{X}_k$, are used at time k to determine the control value $u(k) = \Phi(x(k))$ inside the following set

$$\Omega_k(x) = \{u :\ \ 0 \leq u \leq \bar{u},\ \ x(k) - d(k) + u(k) \in \mathcal{X}_{k+1},\ \forall d^-(k) \leq d \leq d^+(k)\}$$

  which is referred to as control map. It is not difficult to realize that feasibility is assured by construction if and only if $x(0) \in \mathcal{X}_0$.

Though the operation of storing the sets $\mathcal{X}_k$ can be computationally demanding, this solution (being of the feedback nature) presents several advantages over the open-loop one. It is very simple to find examples in which the open-loop solution does not exist while the feedback solution does. For instance, for $1 \leq d(k) \leq 3$, $\bar{u} = 4$ and the target interval $0 \leq x \leq 4$, the target can be met for arbitrary $T > 0$ and all $0 \leq x_0 \leq 4$ by using the feedback strategy, but no open-loop strategy exists for $T > 4$.

The previous closed-loop solution is a typical dynamic programming algorithm [Ber00]. The basic idea of dynamic programming is determining the solution backward in time by starting from the target.

There is an interesting connection between dynamic programming and the construction of Lyapunov functions for uncertain systems. Let us consider the case of a simple linear time-varying uncertain system

$$x(k+1) = A(w(k))x(k)$$

where $A(w)$, for $w \in \mathcal{W}$, and $\mathcal{W}$ is a compact set. Being w time-varying, the natural way to check robust stability of the system is to seek for a Lyapunov function.

If we resort to quadratic functions, then the problem is basically that of checking if for some positive definite P it is possible to assure

$$x^T A(w)^T P A(w) x < x^T P x, \quad \forall w \in \mathcal{W}$$

However, this is a sufficient condition only. Indeed there are examples of linear uncertain systems for which no quadratic Lyapunov function can be found, but are indeed stable. Then the question is shifted to the following one: is it possible to find a suitable non-quadratic Lyapunov function? The set-theoretic approach provides a constructive way to do this. Again, one possibility is to apply dynamic programming ideas. Given an arbitrary convex set $\mathcal{X}$ containing the origin in its interior, consider the next sequence, computed backward in time (i.e. $k = 0, -1, -2, \ldots$)

$$\begin{aligned} \mathcal{X}_0 &= \mathcal{X} \\ \mathcal{X}_{k-1} &\doteq \{x \in \mathcal{X} : \; A(w)x \in \mathcal{X}_k, \;\; \forall w \in \mathcal{W}\} \end{aligned}$$

Under appropriate technical assumptions, it will be seen that each element of this sequence is compact and convex and such that $\mathcal{X}_k \subseteq \mathcal{X}$. Furthermore the sequence is nested, say $\mathcal{X}_{k-1} \subseteq \mathcal{X}_k$. If the sequence converges to a set (which includes the origin) which we will denote by $\mathcal{X}_{-\infty}$, then such limit set exhibits some fundamental properties:

- it is the set of *all states* inside $\mathcal{X}$ which remain inside $\mathcal{X}$ if propagated by the system $x(k+1) = A(w(k))x(k)$;
- it is the *largest* positively invariant set in $\mathcal{X}$;
- if $\mathcal{X}_{-\infty}$ includes 0 *as an interior point*, then it is possible to associate with this set a positively homogeneous function, whose level surfaces are achieved by scaling the boundary of $\mathcal{X}_{-\infty}$ (a simple example will be presented soon), which is a Lyapunov function for the system (then the system is stable);
- if $\mathcal{X}_{-\infty}$ is zero-symmetric (i.e., $x \in \mathcal{X}_{-\infty}$ implies $-x \in \mathcal{X}_{-\infty}$), the associated function is a norm.

We will show later how the possibility of deriving Lyapunov functions by means of invariant sets leads to a stability criterion which is not only sufficient for stability, but necessary as well. We will also show that the technique can be used for continuous-time systems and for stabilization problems, namely when a proper control action has to be found.

The set-theoretic approach can be successfully exploited to achieve other kinds of results. Consider, for instance, the following asymptotically stable single-input single-output system

$$x(k+1) = Ax(k) + Ed(k), \;\; y(k) = Cx(k)$$

and assume that $|d(k)| \leq 1$ is a bounded unknown signal.

If its performance in terms of disturbance rejection is investigated, namely

$$\mu_{max} = \sup_{x(0)=0,\ k\geq 0} |y(k)|$$

then it is well known that

$$\mu_{max} = \sum_{h=0}^{\infty} |CA^h E| \tag{1.1}$$

However, it is possible to solve this problem in the following alternative set-theoretic way by considering the strip

$$Y(\mu) = \{x : |Cx| \leq \mu\}$$

It will be shown that $\mu \geq \mu_{max}$ if and only the set of states reachable from 0 with bounded input d is included in the strip $Y(\mu)$ (see Fig. 1.3). A different, but equivalent, condition is to verify whether there exists a (robustly) positively invariant set $\mathcal{S}$ for the system $x(k+1) = Ax(k) + Ed(k)$, including 0 and which is included in the strip or, in other words, that the maximal positively invariant set included in $Y(\mu)$ includes the origin. As we shall see, both the computation of the 0 reachable sets and of the largest positively invariant set are in general much harder than the computation of the series (1.1). The main point is that when (even linear) uncertain systems are dealt with, the series formula (1.1) is not valid anymore, while the computation of invariant sets is a viable solution. It will also be shown how this kind of ideas are useful to synthesize controllers which minimize the worst case peak ratio, i.e. minimize μ_{max}.

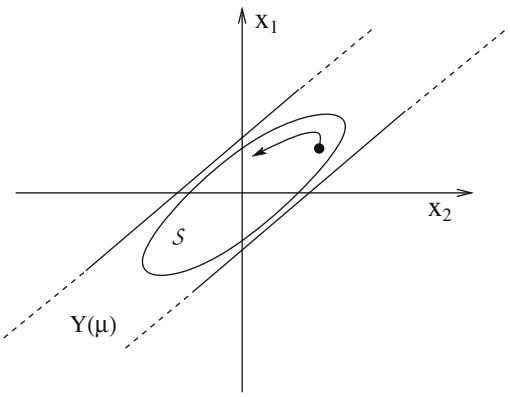


Fig. 1.3 The set-theoretic interpretation μ.

1.3.3 Constrained control

Dealing with constraints in control design is an old issue, but it is still a major challenge in modern control theory and application. The classical approach based on feasibility regions in the state space, presented in the 70s, is still receiving attention. Let us briefly show, by means of an example, which are the basic issues in this context.

Example 1.4. Consider the problem of damping an oscillating system (Fig. 1.4) for which certain admissible ranges are prescribed for the position y (with respect the equilibrium), the speed $\dot{y}$ and the force u. Let $M = K = 1$ and assume that such ranges are described by $|y| \leq \bar{y}$, $|\dot{y}| \leq \bar{v}$, $|u| \leq \bar{u}$ and that these constraints are hard, namely no violation is allowed, so that the control must be designed accordingly. The first (obvious) point to consider is that, once the control law is applied, constraints violation depend on the initial condition. It is apparent that there are many different ways to face this problem:

- compute a stabilizing control and find the set of initial states for which no violation occurs;
- compute a stabilizing control for which no violation occurs for a prescribed (set of) initial state(s);
- compute a stabilizing control which "maximizes" (in some sense) the set of initial states which do not produce violations.

It is clear that there are other possible ways to formulate the problem. If only the control input is constrained, one can consider the saturation operator

$$u_{sat} = \mathrm{sat}_{\bar{u}}(u)$$

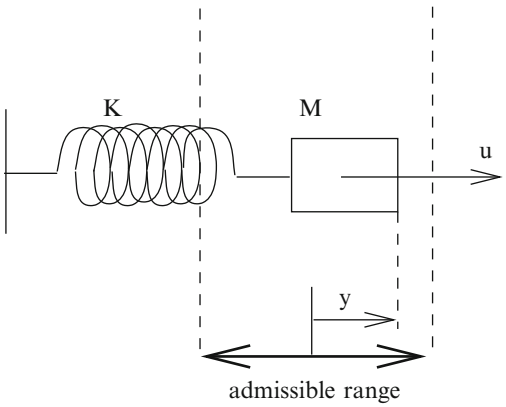


Fig. 1.4 The oscillating system

that limits the control value (we remind that $sat_{\bar{u}}(u) = u$ if $|u| \leq \bar{u}$ otherwise $sat_{\bar{u}}(u) = \bar{u}\ \mathrm{sgn}(u)$). One can then adopt any control, for instance a linear gain $u = Kx$, and saturate it: $u_{sat} = sat_{\bar{u}}[Kx]$. Though this feedback prevents constraints violation, nothing can be said about output constraints (and even about stability). For instance, let the state constraints be given by a square, precisely $|y| \leq \bar{y} = 1$, $|\dot{y}| \leq \bar{v} = 1$, $\bar{u} = 1$. The resulting equations are

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 + u \end{aligned}$$

The first idea is to seek for a control which introduces a damping

$$u = -\kappa x_2$$

with $\kappa > 0$. If no violation is allowed, then any initial state must be restricted in the strip

$$|x_2| \leq \frac{1}{\kappa}$$

Unfortunately, meeting the constraints at time t does not assure that the constraints will be met in the future. For instance, for $\kappa = 1$ all the states in the square satisfy all the constraints but, if we consider the right upper corner, $[1\ 1]^T$, it is immediate to see that the derivative is $\dot{x} = [1\ \ -2]^T$ and then, being $\dot{x}_1 > 0$, the square is abandoned.

One of the most efficient approaches is based on the invariant sets theory. What we need is a set $\mathcal{S}$ which is compatible with the constraints and which is positively invariant. This is going to be a safe set, precisely a set of states for which not only the constraints are satisfied, but constraint fulfilling is guaranteed in the future as well.

A possible choice of candidate invariant sets is the family of ellipsoids, which are quite simple to use. For instance, in our case, it is easy to see that the unit circle is positively invariant for the system with the control $u = -x_2$ since the function $\Psi(x_1, x_2) = (x_1^2 + x_2^2)/2$ has non-positive Lyapunov derivative $\dot{\Psi}(x_1, x_2) = -x_2^2$.

Note that the unit circle is the "largest" ellipse that can be included in the square, and therefore it is, in some sense, the optimal choice inside the family of ellipsoidal sets. Note also that there are other possible controls under which the unit circle is invariant and compatible with constraints. To investigate this problem, consider the Lyapunov derivative "with no feedback applied"

$$\dot{\Psi}(x_1, x_2, u) \doteq \nabla\Psi(x)(Ax + Bu) = x_2 u$$

Then any control action which makes $\dot{\Psi} \leq 0$ assures that $\Psi(x_1(t), x_2(t))$ is non-increasing, so any circle is an invariant set. In this lucky case saturated control, $u = -\mathrm{sat}[\kappa x_2]$, is appropriate since it yields $\dot{\Psi}(x_1, x_2) = -\mathrm{sat}[\kappa x_2]x_2 \leq 0$. Any set having the property of becoming positively invariant provided that a certain control

action is applied is named *controlled-invariant*. This circle is a special case because it is by itself positively invariant for the autonomous system (i.e., with no control action, $u = 0$).

In principle the function $\Psi(x_1(t), x_2(t))$ is not a Lyapunov function in a strict sense since its derivative is negative semidefinite (it is necessarily 0 for $x_2 = 0$). Actually it is possible to assure convergence by means of a slightly different Lyapunov function, $\Psi_\alpha(x_1(t), x_2(t)) = (x_1^2 + 2\alpha x_1 x_2 + x_2^2)/2$. For a small $\alpha > 0$ the unit ball is an ellipse which is arbitrarily close to the unit circle, hence "almost optimal." The Lyapunov derivative is

$$\dot{\Psi}_\alpha(x_1, x_2) = [x_1 + \alpha x_2 \;\; x_2 + \alpha x_1](Ax + Bu) = -\alpha x_1^2 + \alpha x_2^2 + (x_2 + \alpha x_1)u$$

Elementary computations show that one can make the expression above strictly negative by means of the control

$$u = -\kappa[x_2 + \alpha x_1]$$

for $\kappa > 0$ large enough. Indeed

$$\dot{\Psi}_\alpha(x_1, x_2) = -\alpha(1 + \kappa\alpha)x_1^2 - 2\alpha\kappa x_2 x_1 - (\kappa - \alpha)x_2^2$$

which is negative definite if and only if $\kappa > \alpha$ and $(1 + \kappa\alpha)\alpha\kappa - 2\alpha\kappa > 0$. Note that the smallest is α the smallest is the value of κ required to assure that $\Psi_\alpha(x_1, x_2)$ is negative definite. The function $\Psi_\alpha(x_1, x_2)$ becomes a Lyapunov function for the system once a proper controller is assigned, namely it is a *control Lyapunov function*. Its sublevel sets are, again, controlled invariant sets (with the difference that now the control action is necessary since they are not invariant for $u = 0$)

Therefore by means of a small modification (associated with the new parameter α) of the original function one gets

- a safe set arbitrarily close to the optimal ellipsoid (i.e., the circle);
- constraint satisfaction, because κ can be arbitrary small, hence the control-admissible strip

$$|x_2 + \alpha x_1| \leq \frac{1}{\kappa}$$

  becomes arbitrarily large;
- assured convergence because the derivative is negative $\dot{\Psi}_\alpha(x_1, x_2) < 0$ for $(x_1, x_2) \neq 0$.

Is it possible to do something better? Indeed it is. We can abandon the family of ellipsoidal sets and choose a "more general" class. It can be shown that the largest controlled invariant set is a convex set, but not an ellipsoidal one. If $|u| \leq \bar{u} = 1$, then the largest controlled invariant set $\mathcal{S}$ can be shown to be the one which is depicted in Fig. 1.5 and compared with the circle. This region is delimited by part

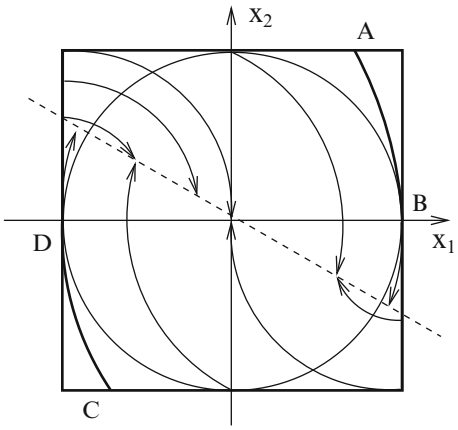


Fig. 1.5 The largest controlled invariant set

of the square edges and by the arcs in the first and third quadrants (thick lines). An intuitive explanation can be given as follows. Consider the full-force trajectories namely those associated with $u = \pm 1$. It can be shown[5] that these are circles centered in the points $(-1, 0)$ (for $u = -1$) and $(1, 0)$ (for $u = +1$). These trajectories are tracked in the clockwise sense. The two curved lines, *A–B* and *C–D*, are achieved by full-force trajectories ($u = -1$ and $u = 1$, respectively). It is quite intuitive that from any point on the boundary of the set there is a proper choice of $u = \pm 1$ such that the corresponding full-force trajectory drives the state inside this set. It is also easy to realize that any point in the square which is not in this region must escape, no matter which control (constrained as $|u| \leq 1$) is applied.

Once it has been established that the depicted convex set is the largest region, the next questions have to be faced:

Q_1: How can a control in the feedback form be found?
Q_2: Can this simple idea be extended to more than two-dimensions? If so, how can it be extended?
Q_3: Is it possible to find the boundary (even in two-dimensions) if the system is uncertain, so that no trajectory can be a priory computed?
Q_4: Given that the maximal region is not much bigger than the circle, why shouldn't we be satisfied with the circle?

The first question can be easily faced if one allows for discontinuous control. The feedback

$$u = -\text{sgn}[x_2 + \alpha x_1],$$

[5]For instance for $u = 1$, write the system as $x_2 dx_2 + (x_1 - 1)dx_1 = 0$ by means of variable separation to derive $x_2^2 + (x_1 - 1)^2 = \gamma$.

where sgn[·] is the sign function, does the job. It applies the full force input $u = -1$ above and $u = 1$ below the line $x_2 + \alpha x_1 = 0$ (the dashed line in the figure). As a result, any trajectory starting in the square reaches such line. Once this line is reached it cannot be abandoned, but still convergence to the origin has to be proved and thus let us see what happens on the line. Consider the function $\Psi_2(x_1, x_2) = x_1^2$. This cannot be given the dignity of candidate Lyapunov function since it is not positive definite (only positive semidefinite) so we will call it a Lyapunov-like function. The Lyapunov derivative is

$$\dot{\Psi}_2(x_1, x_2) = \alpha x_1 x_2 < 0$$

which is negative as long as the state $x \neq 0$ is on the line $x_2 + \alpha x_1 = 0$ (precisely $\dot{\Psi}_2(x_1, x_2) = -\alpha^2 x_1^2$). There is a continuous version of this control which is

$$u = -sat[\kappa(x_2 + \alpha x_1)]$$

which works properly for $\kappa > 0$ large enough.

By chance, in this case we have that the same control which was previously deduced by the control Lyapunov function $\Psi_1(x_1, x_2)$ works for the maximal invariant set although, in general, this is not the case. A natural question is whether we can associate a Lyapunov function Ψ_3 with the largest invariant set. Let us consider the function which is intuitively constructed as a positively homogeneous function (i.e., $\Psi_3(\lambda x_1, \lambda x_2) = \lambda \Psi_3(x_1, x_2)$) and which is equal to 1 on the boundary of the largest controlled invariant set $\mathcal{S}$. Such a function turns out to be

$$\Psi_3(x_1, x_2) = \max \left\{ |x_1|, |x_2|, \text{sgn}[x_1 x_2] \frac{|x_1| + \sqrt{4x_1^2 + 3x_2^2}}{3} \right\}$$

It is not difficult to see that such a function is equal to one on the boundary of $\mathcal{S}$, and it is clearly positively homogeneous of order one. Such function is called a Minkowski function of $\mathcal{S}$ (see Fig. 1.6) Since $\mathcal{S}$ is symmetric, it is a norm (in general, if symmetry is dropped, it is called gauge function). It is apparent that this function is not differentiable, but only piecewise differentiable. This implies that the gradient of this function is defined only on the interior of the regions denoted by $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_3$ and their opposite. What about the corners? Actually one can replace the gradient by the subgradient and instead of the regular Lyapunov derivative one can use the directional derivative

$$D^+\Psi(x(t)) \doteq \lim_{\tau \to 0^+} \frac{\Psi(x(t+\tau)) - \Psi(x(t))}{\tau}$$

which is always defined if Ψ is convex and $x(t)$ is a regular solution of the differential equation. Unfortunately one cannot always assume that the solution is "regular." For instance, the closed-loop system with the saturated control is Lipschitz, but it becomes discontinuous if we consider sign-type functions. To be

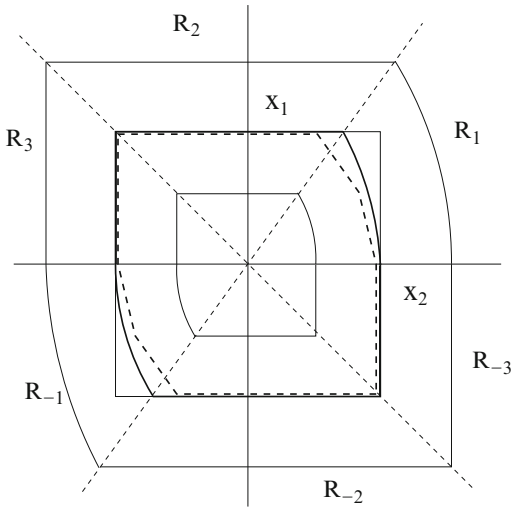


Fig. 1.6 The set $\mathcal{S}$, its polyhedral approximation and the level curves of Ψ_3

able to include "discontinuities" we will use, for the sake of generality, the Dini derivative (basically, replacing $\lim$ by $\limsup$). Although this generality will cause some difficulties, still the essential ideas can be presented in an intuitive way (and they will be).

Question Q_2 points out some limits of the presented theory. The simple ideas presented in the example, indeed, cannot be so easily generalized (in a constructive way) to more general cases. The idea of using the "extremal trajectory" to delimit the largest invariant set cannot easily be extended in dimension 3 and above. However, if we resort to a special category of sets, namely the polyhedral ones, then there are available algorithms which compute, in a systematic way, invariant sets in this family. These algorithms are based on linear programming. The main idea is depicted in Figure 1.6 and basically amounts to replace the original set $\mathcal{S}$ by a polyhedral set (the dashed one). It will be shown that such an approximation can be arbitrarily faithful but, in turn, arbitrarily complex.

Question Q_3, concerning uncertainties, is also a problem that has to be faced. For uncertain systems, extra care should be put to extend the idea of extremal trajectories leading to the boundary of invariant sets, even in two-dimensions. To face the general problem, as we will see, the approach based on dynamic programming, previously described, is effective, since it enables to face, in some sense, the "worst case" with no conceptual limitation (there are only computational limitation due to the system dimensions).

Question Q_4 is extremely important and, notwithstanding its apparent "malice," we cannot (and it is not our intention to) hide the fact that quite often the described algorithms provide modest improvements to a much higher price from the computational complexity standpoint. But, as already stressed, these operational consuming techniques lead to necessary and sufficient conditions that can be used

to evaluate the performances of approximate solutions. This critical view is one of the most important features of the book. We reply to the fourth malicious question as follows.

- Perhaps it depends on the application. One may be ready to use a sophisticated software to have "the best solution" if the application is extremely important and with stringent specifications.
- Computing the largest invariant set provides a reference to your approximated solution based on circle. If you denote by v_{cir} the circle volume and by v_{max} the volume of the maximal set, you can say that the approximate solution is of a factor v_{cir}/v_{max} worse than the "optimal" one. Note that the volumes can be computed even in high dimension (in the example it is an elementary exercise) via randomized algorithms [TCD04].
- By analyzing the optimal solution you know something very important about the "approximate solution," since the feedback $u = -\text{sat}[\kappa(x_2 - \alpha x_1)]$ deduced from the "modified circle" indeed assures, in this special case, the maximal domain of attraction.
- Concerning the oscillating system just presented, *given the outcome in terms of the difference between the two sets*, we would suggest to use the approximated one, unless such system had to be used in a stringent performance application.

1.3.4 Required background

We believe that this book is accessible to any student with a normal control and system theory background, including fundamentals of linear system theory and basics of Lyapunov theory. We also believe that it is accessible to all control theoreticians, especially mathematicians who are not be too frustrated by "intuitive arguments," and practical engineers who are not scared by the formalism of the first part. The notions which are necessary for the reading are

- very basic notions of topology (neighborhood, compactness, continuity, Lipschitz continuity, etc.) in finite dimension;
- basic calculus;
- linear algebra;
- notions of convexity and convex functions;
- basics of linear and convex programming.

1.4 Related topics and reading

In writing this book we planned to fill a hole (in space–time). The topic of this book has been addressed since thirty years, for instance in Schweppe's book [Sch73] that, basically, was written with the same spirit. We do claim that the concepts expressed

there remain still valid and this work covers only a part of that book. As far as the dynamic programming is concerned, we would like to point out the excellent work by Bertsekas [Ber00], now available in a renewed version.

The concepts presented here, in particular invariance and controlled invariance, can be faced in a formal way and the interested reader is referred to [Aub91]. Fundamental concepts on Lyapunov theory can be found in [RHL77].

A suggested complementary reading is the book [BEGFB04], which is a renewed version of a very successful edition. That book is mainly concerned with ellipsoidal sets and quadratic functions. The existence of that work has been fundamental in the tuning of the present book since it has allowed us not to enter in deep details of invariance of ellipsoidal sets since much material can be found there. A further reference of a book specialized to the case of ellipsoidal sets is [KV97]

A subject which is related to the material in this book is the geometric theory of linear systems in particular the properties of controlled invariant subspaces [BM92].

One of the fundamental issues which will be dealt with is the control of constrained systems. A nice complementary reading is the book [HL01]. Several parts of the book will be devoted to the control of uncertain systems based on Lyapunov methods. Classical books on this matter are [FK96b, Qu98]. Other references (written with different spirit) about the control of uncertain systems are [SPS98, ZDG96, DDB95, CGTV09].

The book will dwell briefly on special problems such as that of state estimation. The literature on this topic is huge. Among the books concerned with the problem, we point out [Che94] which is the most similar in spirit. A new emerging topic at the moment of writing the first edition of this book is the control of switching systems. We have now dedicated a new chapter to the topic which is far from being exhaustive. Excellent references on this topic are [Lib03, SG05, SG11].

Chapter 2
Lyapunov and Lyapunov-like functions

As shown in the introduction, Lyapunov functions are crucial in the present book aims, given the strict relation between Lyapunov functions and invariant sets. In this chapter, basic notions of Lyapunov and Lyapunov-like functions will be presented. Before introducing the main concept, a brief presentation of the class of dynamic models which will be considered and some preliminary mathematical notions are given.

It is assumed that the reader is familiar with basic concepts of differential and difference equations.

2.1 State space models

The dynamic systems which will be considered in the book are those governed by ordinary differential equations of the form

$$\dot{x}(t) = f(x(t), u(t), w(t)) \tag{2.1}$$

$$y(t) = h(x(t), w(t)) \tag{2.2}$$

or by difference equations of the form

$$x(t+1) = f(x(t), u(t), w(t)) \tag{2.3}$$

$$y(t) = h(x(t), w(t)) \tag{2.4}$$

where $x(t) \in \mathbb{R}^n$ is the system state, $u(t) \in \mathbb{R}^m$ is the control input, $y(t) \in \mathbb{R}^p$ is the system output, and $w(t) \in \mathbb{R}^q$ is an external input (uncontrolled and whose nature will be specified later). In particular, systems of the form

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_2

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t) \tag{2.5}$$

$$y(t) = Cx(t) \tag{2.6}$$

will be often considered. The distinction between the two external inputs is due to practical consideration. Indeed w typically represents model variations while d represents a disturbance signal. Clearly this is a special case because the two signals can be combined in a single input $\hat{w} = (w, d)$.

The book is mainly devoted to control problems. The class of regulators which will be considered is the following:

$$\dot{x}_c(t) = f_c(x_c(t), y(t), w(t)) \tag{2.7}$$

$$u(t) = h_c(x_c(t), y(t), w(t)) \tag{2.8}$$

In the case of discrete-time systems the expression becomes

$$x_c(t+1) = f_c(x_c(t), y(t), w(t)) \tag{2.9}$$

$$u(t) = h_c(x_c(t), y(t), w(t)) \tag{2.10}$$

The reason why no control action has been considered in the output equation (2.2) and (2.4) is the well posedness of the feedback connection. In some parts of the book this assumption will be dropped and a direct dependence of the output from the input will be considered, resulting in

$$y(t) = h(x(t), u(t), w(t)).$$

It is well known that the connection of the systems (2.1)–(2.2) and (2.7)–(2.8) results in a dynamic system of augmented dimension whose state is the compound vector

$$z(t) = \begin{bmatrix} x(t) \\ x_c(t) \end{bmatrix} \tag{2.11}$$

This state augmentation is crucial for the Lyapunov theory which has the state space as natural environment. Indeed, a dynamic feedback (as long as the dimension of $x_c(t)$ is known) can always be regarded as the static feedback

$$v(t) = f_c(x_c(t), y(t), w(t))$$

$$u(t) = h_c(x_c(t), y(t), w(t))$$

for the augmented system

$$\dot{z}(t) = \begin{bmatrix} \dot{x}(t) \\ \dot{x}_c(t) \end{bmatrix} = \begin{bmatrix} f(x(t), u(t), w(t)) \\ v(t) \end{bmatrix} \tag{2.12}$$

with output

$$y(t) = h(x(t), w(t))$$

Therefore (with few exceptions) we will usually refer to static feedback control actions. Obviously the same considerations can be done for the discrete-time version of the problem.

One of the goals of the present work is to deal with constraints which are imposed on both control and output, typically of the form

$$u(t) \in \mathcal{U} \tag{2.13}$$

and

$$y(t) \in \mathcal{Y}, \tag{2.14}$$

where $\mathcal{U} \subset \mathbb{R}^m$ and $\mathcal{Y} \subset \mathbb{R}^p$ are assigned "admissible" sets.

As far as the input $w(t)$ is concerned, it will play different roles, depending on the problem. More precisely, it will act as a noise entering the system and/or as an uncertain time-varying parameter affecting the system or it will be regarded as a reference signal. A typical specification for such function is given in the form

$$w(t) \in \mathcal{W}. \tag{2.15}$$

The set $\mathcal{W}$ will therefore be either the set of possible variation of an unknown-but-bounded noise/parameter or the set of variation of the admissible reference signals.

It is definitely worth mentioning that the joint presence of the control $u(t)$ and the signal $w(t)$ can be interpreted in terms of dynamic game theory in which u (the "good guy") plays against w (the "bad guy") [BO99]. In this context, a successful design problem is a "good end movie" in which u prevails over w, which in turn applies the worst actions to prevent u from assuring its goal. Two possibilities can occur:

(a) the control $u = u(x)$ plays first, say it is unaware of what the "bad guy" w is doing, and u has to minimize all the possible "damage" caused by w;
(b) the disturbance w plays first, say the control $u = u(x, w)$ is aware of the "bad guy" move and can exploit such knowledge to counteract its effects.

Obviously, the possibility for the compensator (i.e., the strategy adopted by u) of exploiting the information of w is an advantage which might produce a big difference in the result of the game. We will dwell again later on this game-theoretic concept.

2.1.1 Differential inclusions

An important concept that will be considered is the concept of differential inclusion. Basically, a differential inclusion is an expression of the form

$$\dot{x}(t) \in F(x(t)) \tag{2.16}$$

where $F(x)$ is not a single vector valued function, but a set-valued function, i.e. $F(x)$ is a set for all x. In the case of an uncertain system

$$\dot{x}(t) = f(x(t), w(t)), \quad w(t) \in \mathcal{W}$$

the set F is then given by $F(x) = \{f(x,w), \quad w \in \mathcal{W}\}$ so that (2.16) is a generalization. The formalism (2.16) is very useful, and the differential inclusions theory is quite effective in dealing with several mathematical problems.

A well-known example is the case of a system of differential equations with discontinuous terms. The problem is mathematically relevant and stems from many different applications such as that of relay systems, switching systems or sliding mode control. Assume that

$$\dot{x}(t) = f(x(t))$$

is not continuous in a point x. It is possible to embed this system in a "minimal differential inclusion" of the form (2.16) and claim that, by definition, any absolutely continuous function (defined later) $x(t)$ which satisfies (2.16) is a solution of the system.

Example 2.1 (Oven control). Consider the following relay system

$$\dot{x}(t) = -\lambda x(t) + u(t)$$

with control

$$u(t) = \begin{cases} 0 \text{ if } x \geq \bar{x} \\ \bar{u} \text{ if } x < \bar{x} \end{cases}$$

The system is a well-known example of control of a heating plant (such as an oven) where $\bar{x}$ is the desired temperature. Clearly, the value $x = \bar{x}$ represents a discontinuity of the function. The corresponding differential inclusion is

$$\dot{x} \in F(x) := \begin{cases} -\lambda x & \text{if } x > \bar{x} \\ [-\lambda\bar{x}, -\lambda\bar{x} + \bar{u}] & \text{if } x = \bar{x} \\ -\lambda x + \bar{u} & \text{if } x < \bar{x} \end{cases}$$

Let us assume that the desired temperature is $\bar{x} \geq 0$ (reasonable for both Celsius and Fahrenheit scale) and that $-\lambda\bar{x} + \bar{u} > 0$ (a desirable assumption). Then, it is not

difficult to see that the (unique in this case) absolutely continuous function which satisfies the inclusion for $x(0) \leq \bar{x}$ is

$$x(t) = \min\{(x(0) - \bar{u}/\lambda)e^{-\lambda t} + \bar{u}/\lambda, \bar{x}\}$$

which becomes, by definition, the solution of the relay system. The determination of the analogous expression of $x(t)$ for $x(0) \geq \bar{x}$ is left to the reader.

The theory of differential inclusions, by exploiting the set valued maps theory, and especially the concept of semicontinuity (or continuity) of set-valued maps, is a framework to prove the existence of solutions in the general case (see Chapter 2 of [AC84] for details).

2.1.2 *Model absorbing*

An idea that sometimes is extremely useful is the approximation achieved by absorbing a nonlinear uncertain system in a linear (controlled) differential inclusion. Consider the system

$$\dot{x}(t) = f(x(t), u(t), w(t)) \tag{2.17}$$

The idea is determining a family of matrices

$$\{A(p), B(p), \ \ p \in \mathcal{P}\}$$

such that for all $w \in \mathcal{W}$

$$f(x, u, w) = A(p)x + B(p)u, \quad \text{for some } \ p \in \mathcal{P} \tag{2.18}$$

for all x and u. If such $A(\cdot)$ and $B(\cdot)$ exist, we say that system (2.17) is absorbed in (2.18) . Then we can claim that, no matter how u is taken, any trajectory of the original system (2.17) is also a trajectory of (2.18) (the opposite is clearly not true in general). As a consequence, if we are able to determine the qualitative behavior of the absorbing system, we can determine (in a conservative way) the behavior of the original system. The important property of the mentioned trick is that if one is able to stabilize (2.18), or to prove its stability in the uncontrolled case, then stability is assured also for (2.17) [Liu68] (see also [LR95, SA90, SA91]).

Example 2.2 (Magnetic levitator). Consider the following simplified equation of a magnetic levitator

$$\ddot{y}(t) = -k\frac{i(t)^2}{y(t)^2} + g = f(y(t), i(t))$$

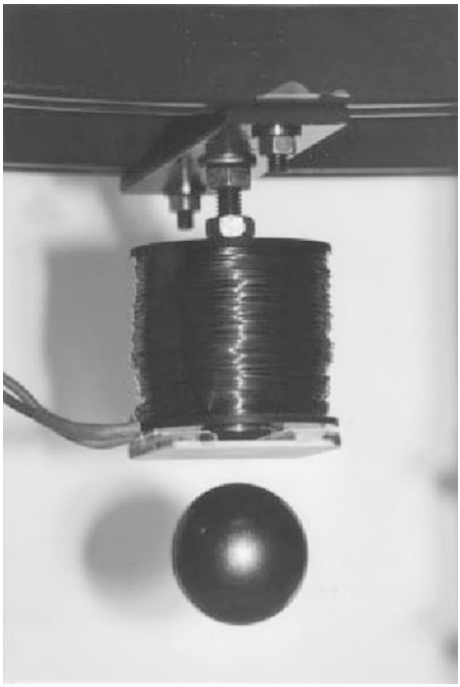


Fig. 2.1 The magnetic levitator

shown in Fig. 2.1. The variable y is the distance of a steel sphere from a controlled magnet and i is the current impressed by an amplifier. Note that y is oriented downwards (i.e., increasing y means lowering the sphere). Let $(\bar{y}, \bar{i})$ be the pair of equilibrium positive values, say such that $f(\bar{y}, \bar{i}) = 0$. Then, $f(y, i)$ can be written as follows:

$$f(y,i) = \int_{(\bar{y},\bar{i})}^{(y,i)} \left[\frac{2ki^2}{y^3} dy - \frac{2ki}{y^2} di \right]$$

The integral in the previous expression is to be thought as a curve-integral evaluated on any continuous curve connecting $(\bar{y}, \bar{i})$ to (y, i). Now, it is reasonable to assume that the variables are bounded as

$$0 < y^- \leq y \leq y^+, \quad 0 < i^- \leq i \leq i^+, \quad 0 < k^- \leq k \leq k^+,$$

and then get the absorbing model

$$\ddot{y}(t) = p_1(t)(y(t) - \bar{y}) - p_2(t)(i(t) - \bar{i}) \tag{2.19}$$

By the mean value theorem, the minimal and maximal values of $p_1(t)$ and $p_2(t)$ can be taken as

$$\frac{2k^- {i^-}^2}{{y^+}^3} \leq p_1(t) \leq \frac{2k^+ {i^+}^2}{{y^-}^3}, \quad \frac{2k^- i^-}{{y^+}^2} \leq p_2(t) \leq \frac{2k^+ i^+}{{y^-}^2}$$

The original nonlinear model is then handled as a model of the type (2.18), known as Linear Parameter-Varying (LPV) systems. LPV models are very important and they will be often considered in the present book. The previous approach is also known as quasi-LPV system modeling in which a "nonlinearity" is removed by introducing a suitable parameter (see, for instance, [SR00]). The next example comes from chemistry.

Example 2.3 (A simple chemical reaction network). Consider the following very simple chemical reactions

$$\emptyset \to A, \quad \emptyset \to B, \quad A + B \to C, \quad C \to \emptyset$$

In which two reactants A and B produce C. C is subject to a linear degradation. We assume that the supply of A is controlled while the inflow of B is exogenously determined.

Assuming mass action kinetics, these reactions correspond to the following set of equations, in which we have denoted by the lowercase letters a, b, and c the concentrations of A, B, and C respectively:

$$\begin{aligned}
\dot{a}(t) &= -ka(t)b(t) + a_0(t) \\
\dot{b}(t) &= -ka(t)b(t) + b_0(t) \\
\dot{c}(t) &= ka(t)b(t) - hc(t)
\end{aligned}$$

k and h are positive rate constants, $a_0(t)$ is the controlled supply of A and $b_0(t)$ is the spontaneous flow of B. Consider a certain equilibrium value corresponding to a constant value of $b_0(t) = \bar{b}_0$. The steady-state conditions are

$$\begin{aligned}
0 &= -k\bar{a}\bar{b} + \bar{a}_0 \\
0 &= -k\bar{a}\bar{b} + \bar{b}_0 \\
0 &= k\bar{a}\bar{b} - h\bar{c}
\end{aligned}$$

thus $\bar{a}_0 = \bar{b}_0$, $\bar{c} = \bar{b}_0/h$ while $\bar{a}$ and $\bar{b}$ can be chosen according to $\bar{a}\bar{b} = b_0/k$. Hence the reaction steady state can be chosen with one degree of freedom. Once the equilibrium point is fixed, by introducing the variables $x_1 = a - \bar{a}$, $x_2 = b - \bar{b}$, $x_3 = c - \bar{c}$, $u = a_0 - \bar{a}_0$ and by noticing that

$$ka(t)b(t) - k\bar{a}\bar{b} = kb(t)(a(t) - \bar{a}) + k\bar{a}(b(t) - \bar{b})$$

it is readily seen that the system can be rewritten as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} -kw(t) & -k\bar{a} & 0 \\ -kw(t) & -k\bar{a} & 0 \\ kw(t) & k\bar{a} & -h \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t) \tag{2.20}$$

with the "parameter" $w(t) = b(t)$. We may obviously assume that $\bar{a} > 0$. To provide bounds for $w(t) = b(t)$ we can assume that the system will be confined in a working region

$$0 < b^- \leq b(t) \leq b^+$$

Note that this system is not asymptotically stable. This can be seen by linearizing it and noticing that the Jacobian has the same form of (2.20) with $w = \bar{v}$, or simply by noticing that, if a_0 and b_0 are fixed, we get

$$\dot{a} - \dot{b} = a_0 - b_0$$

hence even small values of $a_0 - b_0$ produce a linear divergence of the variable $a(t) - b(t) = (a_0 - b_0)t + a(0) - b(0)$. This system needs a stabilizing control.

The two previous examples open the way to a circular argument. The LPV representation is valid as long as the system remains in a proper region. But, since the two systems are unstable, confining the state in such a region can be achieved only by a proper feedback control. Note that this applies to stable systems as well, unless we assume to be extremely close to the equilibrium value.

Therefore, an appropriate quasi-LPV system control should be approached as follows:

1. select a working region in the state-space;
2. find bounds for the "parameters" valid in such a region;
3. make sure that the controller keeps the state inside this region by analyzing the LPV closed-loop system.

As it will be seen later on the Lyapunov approach is the ideal one to solve the problem of confining the state inside an assigned region.

2.1.3 The pitfall of equilibrium drift

The equilibrium point definition is quite popular in elementary system theory classes. It is well known that, given a (certain) dynamic system $\dot{x} = f(x, u)$ and a pair $\bar{x} \in \mathbb{R}^n$ and $\bar{u} \in \mathbb{R}^m$ such that $0 = f(\bar{x}, \bar{u})$, then a translation is possible by introducing the new variables $z(t) \doteq x(t) - \bar{x}$ and $v(t) \doteq u(t) - \bar{u}$, thus achieving a new system

$$\dot{z}(t) = F_{\bar{x},\bar{u}}(z(t), v(t)) \doteq f(z(t) + \bar{x}, v(t) + \bar{u})$$

In the case of an uncertain system, the additional input w is present and therefore, when talking about equilibria, one needs to refer to a nominal value $\bar{w}$ of w to derive, under the condition $0 = f(\bar{x}, \bar{u}, \bar{w})$, a model of the form

$$\dot{z}(t) = F_{\bar{x},\bar{u},\bar{w}}(z(t), v(t), r(t))$$

where $r(t) \doteq w(t) - \bar{w}$. It is quite clear that the equilibrium is changed if the value of the input w is not the nominal.

The pitfall we are talking about is exactly the just mentioned one: a linear model changes its nature to a linear + disturbance model in the uncertain case.

Consider the LPV system $\dot{x} = A(w)x + B(w)u$, for which $\bar{x} = 0$ and $\bar{u} = 0$ is clearly an equilibrium pair, and consider a different equilibrium condition, with $\bar{x} \neq 0$, which is subject to the condition

$$0 = A(\bar{w})\bar{x} + B(\bar{w})\bar{u}$$

Denoting by $\delta A(w) = A(w) - A(\bar{w})$ and by $\delta B(w) = B(w) - B(\bar{w})$, the original LPV system can be written as

$$\dot{z}(t) = A(w(t))z(t) + B(w(t))v(t) + \underbrace{\delta A(w(t))\bar{x} + \delta B(w(t))\bar{u}}_{\Delta(\bar{x},\bar{u},w(t))},$$

which is not a pure LPV system anymore, being affected by an additive term Δ. In particular, if the nominal value of the control $u = \bar{u}$ is applied, the steady state variations of w change the equilibrium state. The theory (often without even mentioning the problem) typically deals with the new term Δ as an additional uncertain noise.

Consider again the levitator model in Example 2.2. Equation (2.19) is correct, but it does present the pitfall. Indeed, for $i = \bar{i}$ and $y = \bar{y}$, the system is in equilibrium conditions. However, if the parameter k changes, the corresponding values of $\bar{i}$ and $\bar{y}$ change as well, according to the physical law $k\bar{i}^2 = g\bar{y}^2$. If a non-nominal value $k \neq \bar{k}$ is considered, we have also to bear in mind that the equilibrium values are not anymore $\bar{y}$ and $\bar{i}$ such that $\bar{y} = \bar{i}\sqrt{\bar{k}/g}$.

Example 2.4 (A circuit with uncertain parameters). Consider the linear electric circuit represented in Fig. 2.2 and whose equations are

$$\dot{x}_1 = -\frac{1}{C_1}\left(\frac{1}{R_1} + \frac{1}{R_3} + \frac{1}{R_5}\right)x_1 + \frac{1}{C_1 R_3}u_1 + \frac{1}{C_1 R_5}u_2$$

$$\dot{x}_2 = -\frac{1}{C_2}\left(\frac{1}{R_2} + \frac{1}{R_4} + \frac{1}{R_6}\right)x_2 + \frac{1}{C_2 R_6}u_1 + \frac{1}{C_2 R_4}u_2$$

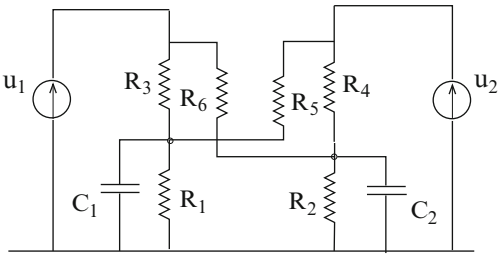


Fig. 2.2 The electric circuit

where x_1 and x_2 are the capacitor voltages. The circuit represents two voltage generators u_1 and u_2 which supply power to the load R_1 and R_2 (whose values are assumed much greater than the transportation resistances R_3, R_4, R_5, and R_6). If symmetry of the circuit is assumed, say $R_1 = R_2$, $R_3 = R_4$, $R_5 = R_6$ and $C_1 = C_2$, then any constant input couple $\bar{u}_1$ and $\bar{u}_2$ results in the steady state values for the state variables

$$\bar{x}_1 = \bar{x}_2 = \bar{x} = \mu(\bar{u}_1 + \bar{u}_2)$$

Therefore, the power is equally supplied to R_1 and R_2, no matter which of the sources u_1 and u_2 pushes harder. However, variations of the provided parameters can deeply unbalance the energy distribution between the two loads, and completely change the equilibrium point (and create power circulation between the generators). The problem becomes even harder if looked at in another way. Assume that $\bar{x}_1$ and $\bar{x}_2$ are fixed. If the input matrix B is non-singular, then the couple $\bar{x}_1$ and $\bar{x}_2$ can be obtained by imposing suitable values to $\bar{u}_1$ and $\bar{u}_2$, precisely $\bar{u} = B^{-1}\Lambda\bar{x}$, where $-\Lambda$ is the diagonal state matrix. It is obvious that if B becomes near-singular due to parameter changes, then the components of $\bar{u}$ may become arbitrarily large. However, if the parameters are unknown, the situation is completely different. We will discuss this matter later when the effects of the uncertainties on B will be considered.

To summarize, two simple facts are worth evidencing:

- The equilibrium shift can be non-negligible, depending on the specific problem faced.
- Models of the form

  $$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

  are more realistic than pure LPV (since the previously introduced term Δ can be included in Ed) when uncertainties have to be taken into account. Such models will be often considered in the book.

2.2 Lyapunov derivative

This chapter is dense of non-standard concepts. However, the reader who is not interested in mathematical formalism can leave the contents at an intuitive level without essentially compromising the reading of the rest of the book.

2.2.1 Solution of a system of differential equations

Although this book is mathematically written, it is mainly devoted to engineering problems. As a consequence, theoretical conditions of existence of solutions are not the main concern. However, to render the exposition rigorous, we need to specify what we mean by solution of a differential equation.

Consider a system (possibly resulting from a feedback connection) of the form

$$\dot{x}(t) = f(x(t), w(t)) \tag{2.21}$$

Clearly if we assume that the signal w is continuous and f is sufficiently regular, then we may always think at (2.21) as a regular relation between $f(x(t), w(t))$ and $\dot{x}(t)$ which exists in the usual sense.

Unfortunately assuming a continuous w unacceptably restricts the class of problems since we could not even consider step inputs. In the sequel, it will always be assumed that $w(t)$ is a piecewise continuous function of time, which means that it has a finite (possibly zero) number of discontinuity points in any finite interval and in each of them it admits finite left and right limits. Then the proper way to handle (2.21) is to consider the equivalent integral equation

$$x(t) = x(0) + \int_0^t f(x(\sigma), w(\sigma))d\sigma \tag{2.22}$$

which is strictly equivalent to (2.21) as long as the derivative $\dot{x}$ exists.

Then, under proper assumptions on f, the solution of (2.22) exists. Unfortunately, it is not always possible to rely on the continuity of the function f, since we will sometimes refer to systems with discontinuous controllers, an event which causes some mathematical difficulties.

For the sake of completeness, we introduce a general notion of solution of a differential equation. Let us start with the next definition [Hal69]

Definition 2.5. A function $f : [a, b] \to \mathbb{R}$, with $[a, b] \subset \mathbb{R}$ a finite interval, is absolutely continuous if for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any countable number of disjoint sub-intervals $[a_i, b_i]$ such that $\sum_i (b_i - a_i) \leq \delta$ we have

$$\sum_i |f(b_i) - f(a_i)| \leq \epsilon$$

It is known that an absolutely continuous function is differentiable everywhere and

$$f(t) = \int_a^t f'(\sigma)d\sigma$$

Actually we have the characterization that a function is absolutely continuous if and only if it is the integral of some Lebesgue-measurable function m

$$f(t) = \int_a^t m(\sigma)d\sigma$$

The next is the most general (reasonable) definition of solution of a differential equation.

Definition 2.6 (Solution of a differential equation). Given a function $x : \mathbb{R}^+ \to \mathbb{R}^n$, which is component-wise absolutely continuous[1] in any compact interval, x is said to be a solution of (2.21) if it satisfies it for almost all $t \geq 0$.

The same definition holds for the solution of a differential inclusion which is an absolutely continuous function satisfying (2.16) almost everywhere. The above definition is quite general, but it is necessary to deal with problems in which the solution $x(t)$ is not differentiable in the regular sense.

In most of the book (but with several exception) we will refer to differential equations admitting regular (i.e., differentiable everywhere) solutions. As far as the existence of global solutions (i.e., defined on all the positive axis $\mathbb{R}^+$) is concerned, we will not enter in this question, since it will always be assumed that the system (2.21) is globally solvable. In particular, equations with finite escape time will not be considered (see Exercise 1 for an example).

Another issue we have to explore is the domain of definition of f. Let us consider the following unit magnetic levitator model[2]

$$\ddot{y}(t) = F(y(t)) = 1 - \frac{1}{y(t)^2}$$

Obviously, function F is not defined for $y = 0$. So it is legitimate to consider F only for positive values $y > 0$. If we start from an initial condition $\dot{y}(0) = 0$ and we take $0 < y(0) < 1$, the system will accelerate upwards, thus decreasing the distance of the sphere from the magnet: $\dot{y}(t) < 0$, and this will produce a decreasing of $y(t)$ and this will further decrease the acceleration .... In simple words this equation will meet its doom: the state will reach the condition $y(t) = 0$ at some (finite) t and after this instant the differential equation is not defined, so this is the end of the story.

[1] See [RHL77] for more details.

[2] For instance built by using a sphere of unit mass with a unitary constant κ, using unitary current on a planet in which the gravity is one.

This is understandable, since the levitator model is realistic (even accurate) only in a neighborhood of the equilibrium point.

Thus in general, a differential equation is defined on a domain, $\mathcal{D}$ but its solution may well escape such a domain. As we have pointed out the "non escaping theme" is a fundamental one in the book. But for the moment being let us point out that mathematicians usually refer to "solution in the set" in the sense that its general properties-existence are a concern as long as $x(t)$ is in $\mathcal{D}$. The fact that the solution might reach the boundary of $\mathcal{D}$ and stop to exist may be a subject of investigation, but not a worry.

Note that mathematical singularities may be removed by slightly changing the model as

$$F(y) = 1 - \frac{1}{(y+\epsilon)^2}$$

with a small ϵ, so that in the contact condition $y = 0$ the force is high but finite. In this case the condition $y = 0$ does not imply mathematical singularities, but just the fact that the model is not valid anymore, because $y > 0$ is a physical constraint. Introducing $\epsilon > 0$ may improve the quality of the model when the sphere is close to the magnet, but not necessarily when it is close to the equilibrium point.

A similar case is the unit-tank model with a hole in the bottom

$$\dot{h}(t) = -\sqrt{h(t)}$$

where h is the fluid height. The domain of the definition is $h \geq 0$. The solution for $h(0) > 0$ reaches the boundary $h = 0$ (in finite time), but the solution does not cease to exist.

2.2.2 The beauty of Lyapunov theory

No one would doubt that the system in Figure 2.3, which is formed by three masses connected by a rope after a sufficiently long time will be found in its minimum energy configuration. More precisely, denoting by (x_i, y_i), $i = 1, 2, 3$, the coordinates of the three masses, the final configuration will be the solution of

$$\begin{aligned} &\min\ E = \min\ gm_1y_1 + gm_2y_2 + gm_3y_3, \\ &s.t. \\ &(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \leq r_i^2,\ \ i = 1, \ldots, 4, \end{aligned}$$

where (x_i, y_i) are the coordinates of the ith point, namely it will minimize the potential energy if the rope is assumed not elastic. Slightly perturbing this systems would not prevent it from returning in this configuration.

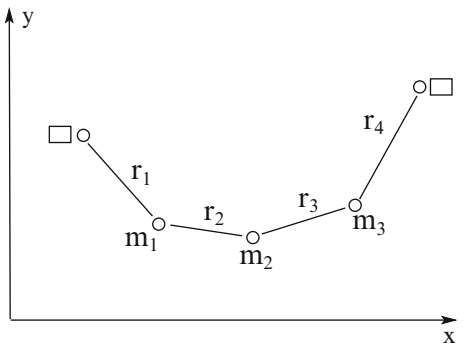


Fig. 2.3 The masses connected by a rope

The reason why we are sure about this fact is that during the evolution any additional energy put in the system will be dissipated and the system will eventually return in the minimal energy configuration. A theoretical mathematician who only looks at the mathematical formulation would say that if the free points (x_1, y_1), (x_2, y_2) (x_3, y_3) are persistently "moved" within the constraint set, by decreasing E without violating the constraints, namely by applying a gradient method, then at some point no further move could be possible after reaching the (unique) constrained minimum. No equations are necessary.

Similarly, any floating object, without any auxiliary force but gravity and Archimedes' force, is known to reach a minimal-energy configuration just because of the dissipation of energy. This minimum may be non-unique (and, in this case, engineers are often concerned with the problem of assuring the proper minimum is attained).

In general, for mechanical systems, writing an energy function is possible and often easy. The main idea of Lyapunov theory is that any system for which there exists a function $\Psi(x)$ of its state variables which has a (possibly local) minimum in a point $\bar{x}$ and is "dissipative" (namely $\Psi(x(t))$ decreases in a neighborhood of $\bar{x}$) exhibits a "stable behavior."

Example 2.7 (A simple compartmental system). Consider the system

$$\dot{x} = -g(x - y) + g_0$$
$$\dot{y} = g(x - y) - f(y)$$

where $g(\cdot)$ and $f(\cdot)$ are continuous and strictly increasing. Such a model is a typical compartmental system which is often encountered in fluid networks and systems biology (Fig. 2.4). The term g_0 represents a flow into compartment 1, g is a flow from compartment 1 to compartment 2 or vice versa, which is increasing with the difference $x - y$, f is an outflow from compartment 2. Assuming g_0 constant, the equilibrium conditions are

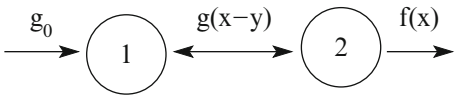


Fig. 2.4 The compartmental model

$$0 = -g(\bar{x} - \bar{y}) + g_0$$
$$0 = g(\bar{x} - \bar{y}) - f(\bar{y})$$

from which it is possible to derive $\bar{y}$ as the unique solution of the equation $f(\bar{y}) = g_0$ and then $\bar{x}$ as the unique solution of $g(\bar{x} - \bar{y}) = g_0$.

What about the behavior of the system? What kind of "energy function" should we consider? Let us pretend that an oracle tells us that a possible choice is

$$\Psi(x, y) = (x - \bar{x})^2 + (y - \bar{y})^2$$

and let us try it. Denoting by $z = x - \bar{x}$ and $w = y - \bar{y}$, the time derivative $\frac{d}{dt}\Psi(z(t), w(t))$ results in

$$\begin{aligned}
&\frac{d}{dt}\Psi(z(t), w(t)) = 2z\dot{z} + 2w\dot{w} \\
&= 2z[-g(z + \bar{x} - w - \bar{y}) + g_0] + 2w[g(z + \bar{x} - w - \bar{y}) - f(w + \bar{y})] \\
&= 2(z - w)[-g(z + \bar{x} - w - \bar{y}) + g_0] + 2w[-f(w + \bar{y}) + g_0]
\end{aligned}$$

Notwithstanding the fact that $z(t)$ and $w(t)$ are unknown, it is not difficult to see that, since $g(\cdot)$ and $f(\cdot)$ are increasing and taking into account the equilibrium conditions, the above expression is always negative unless $w = z = 0$.

It is then a consequence that $\Psi(x(t), y(t))$ is decreasing for $x \neq \bar{x}$ and $y \neq \bar{y}$, no matter how the increasing functions f and g are chosen. Moreover, it can be proven that $\Psi(x(t), y(t))$ converges to 0, hence $x(t) \to \bar{x}$ and $y(t) \to \bar{y}$ (a formal proof will be given later in the general case in Section 2.3).

The above arguments are quite straightforward and nice. However, there are two open questions:

- how has $\Psi(x, y)$ been chosen?
- are there other possible choices?

The first question is one of the main issues of Lyapunov theory. In the case of this simple system, the inspiration for choosing $\Psi = z^2 + w^2$ comes from the observation that the Jacobian is symmetric. As it will be seen later, a linear system with a symmetric matrix is asymptotically stable if and only if $x^T x$ is a Lyapunov function[3].

[3]Note that the function Ψ assures that the convergence to the equilibrium is global, while the linearization would only prove local stability.

The "how" question has no general answer, unless special classes of systems and candidate Lyapunov functions are considered. The second question has a simple and affirmative answer and this aspect will be also deeply investigated.

The Lyapunov theory provides solid tools, but also interesting problems to face. This book will attempt to provide ideas in this direction.

2.2.3 The upper right Dini derivative

Consider a function $\Psi : \mathbb{R}^n \to \mathbb{R}$, defined and locally Lipschitz on the state space. As long as one is interested in the behavior of this function in terms of its monotonicity, one needs to exploit the concept of Lyapunov derivative along the system trajectory. For any solution $x(t)$, consider the composed function

$$\psi(t) \doteq \Psi(x(t))$$

This new function $\psi(t)$ is not usually differentiable. Under the additional assumption of regularity on Ψ and x we could write $\dot{\psi}(t) = \nabla\Psi(x(t))^T\dot{x}(t)$. The basic point of Lyapunov standard theory is that, if we have $x(t) = x$ and $w(t) = w$, then

$$\frac{d}{dt}\Psi(x(t))\bigg|_{x(t)=x,\ w(t)=w} = \nabla\Psi(x)^T f(x,w)$$

This implies that the derivative of $\Psi(x(t))$ can be computed *without the knowledge of* $x(t)$, but as a function of the current state and input values.

In our more general framework, we can just say that the composition of a locally Lipschitz function Ψ and an absolutely continuous function is also absolutely continuous, and therefore it is differentiable almost everywhere. To avoid problems due to the lack of differentiability, the following definition is introduced:

Definition 2.8 (Dini derivative). The upper right Dini derivative $D^+\psi(t)$ of ψ at t is

$$D^+\psi(t) \doteq \limsup_{h\to 0^+} \frac{\psi(t+h) - \psi(t)}{h} \tag{2.23}$$

When the function $\psi(t)$ is differentiable in the regular sense, the standard derivative

$$D^+\psi(t) = \dot{\psi}(t)$$

is obtained. Other Dini derivatives can also be defined. They are four in total and they are denoted as D^+, D_+, D^-, and D_- where '$+$' and '$-$' indicate a limit from the right or from the left, whereas the upper or lower position of the symbol means upper or lower limit. The following inequalities (obviously) hold

$$D_+ \leq D^+$$
$$D_- \leq D^-$$

The only Dini derivative which will be used in the sequel will be the upper right one, say we will limit our attention to D^+. The reason for the lack of interest in the other derivatives is that a) the main focus here is on causal systems evolving in the positive direction, and thus the future trend alone is of interest b) most of the analysis will be carried out in a "worst-case" setting and, since in most of the sequel we will be interested in decreasing monotonicity properties, the "+" choice is preferable. In fact the theory would work as well with any of the three remaining derivatives.

If an absolutely continuous function $\psi(t)$ defined on $[t_1, t_2]$ has upper right Dini derivative $D^+\psi(t)$ which is non-positive almost everywhere, then it is non-increasing on such interval as in the case of differentiable functions. The assumption of absolute continuity is fundamental, because there exist famous examples of continuous functions whose derivative is 0 almost everywhere, but which are indeed increasing[4].

2.2.4 *Derivative along the solution of a differential equation*

Let us consider again a locally Lipschitz function $\Psi : \mathbb{R}^n \to \mathbb{R}$ and the solution $x(t)$ of the differential equation (2.21). A key point of the Lyapunov theory is that, as long as one wishes to analyze certain monotonicity properties of a function $\psi(t)$ resulting from the composition of $\Psi(\cdot)$ and a solution $x(\cdot)$ of (2.21), $\psi(t) = \Psi(x(t))$, it is not necessary to know $x(\cdot)$ as a function of time, but just the current values x and w are needed. Let us introduce the upper directional derivative of Ψ with respect to (2.21), say the limsup of the variation rate $\Psi(x, w)$ in the direction given by the vector $f(x, w)$.

Definition 2.9 (Directional derivative). The upper directional derivative of Ψ with respect to (2.21) is

$$D^+\Psi(x, f(x, w)) \doteq \limsup_{h \to 0^+} \frac{\Psi(x + hf(x, w)) - \Psi(x)}{h} \tag{2.24}$$

In the sequel we will sometimes use, with an abuse of notation, the expression

$$D^+\Psi(x, w) \doteq D^+\Psi(x, f(x, w))$$

[4]The interested reader can find details on the web using "devil's staircase" as a keyword.

when no confusion can arise. The next fundamental property holds (see [RHL77], Appendix. 1, Th. 4.3).

Theorem 2.10. *Let $x(t)$ be a solution of the differential equation (2.21), $\Psi : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function and let $\psi(t)$ denote the composed function $\psi(t) = \Psi(x(t))$. Then*

$$D^+\psi(t) = D^+\Psi(x(t), w(t)) \tag{2.25}$$

for almost all t.

Theorem 2.11. *Let $x(t)$ be a solution of the differential equation (2.21), $\Psi : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function and let $\psi(t)$ denote the composed function $\psi(t) = \Psi(x(t))$. Then*

$$\psi(t_2) - \psi(t_1) = \int_{t_1}^{t_2} D^+\psi(\sigma)\, d\sigma = \int_{t_1}^{t_2} D^+\Psi(x(\sigma), w(\sigma))\, d\sigma \tag{2.26}$$

for all $0 \leq t_1 \leq t_2$.

According to what has been previously mentioned, both theorems are valid for any of the Dini derivatives D^+ D^+ D_- D^-, since $x(t)$ is absolutely continuous (see [RHL77] Corollary 3.4 and Remark 3.5)

2.2.5 Special cases of directional derivatives

There are special but important cases in which the Lyapunov derivative admits an explicit expression, since the directional derivative can be written in a simple way. The most famous and popular case is that in which the function Ψ is continuously differentiable. We formally state what we have already written.

Proposition 2.12. *Assume now that Ψ is continuously differentiable on $\mathbb{R}^n$. Then*

$$D^+\Psi(x, w) = \nabla\Psi(x)^T f(x, w). \tag{2.27}$$

(we remind that $\nabla\Psi(x) = [\partial\Psi(x)/\partial x_1\ \partial\Psi(x)/\partial x_2\ \ldots\ \partial\Psi(x)/\partial x_n]^T$).

Another important case arises when the function $\Psi(x)$ is a proper (i.e., locally bounded) possibly non-differentiable convex function defined in an open domain.

Definition 2.13 (Subgradient). The vector $z \in \mathbb{R}^n$ is a subgradient of Ψ, at x_1 if (see Fig. 2.5):

$$\Psi(x) - \Psi(x_1) \geq z^T(x - x_1), \quad \text{for all } x \in \mathbb{R}^n \tag{2.28}$$

The set of all the subgradients at x_1 is the subdifferential $\partial\Psi(x_1)$.

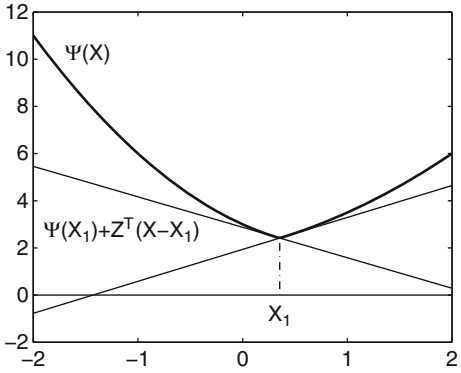


Fig. 2.5 Subgradient of Ψ at point x_1

In general the subdifferential $\partial\Psi(x)$ is a set-valued map. Note that for a differentiable (convex) function $\partial\Psi(x)$ is a singleton including the gradient at x : $\partial\Psi(x) = \{\nabla\Psi(x)\}$. Clearly the provided definition of subgradient holds for any function, but if the function is non-convex the sub-differential may be empty in some points[5].

For a convex (possibly non-differentiable) Lyapunov function $\Psi(x)$ the above definition allows to compute the Lyapunov derivative as

$$D^+\Psi(x,w) = \sup_{z\in\partial\Psi(x)} z^T f(x,w). \tag{2.29}$$

Another interesting case is that of maximum-type convex functions. Assume that $g_1(x),\ g_2(x),\ldots,g_m(x)$ is a given family of continuously differentiable convex functions. The "maximum function" is the convex function defined as

$$g(x) = \max_i g_i(x)$$

Define as $\mathcal{I}(x)$ the set of indices where the maximum is achieved:

$$\mathcal{I}(x) = \{i : g_i(x) = g(x)\}$$

Then

$$D^+\Psi(x,w) = \max_{i\in\mathcal{I}(x)} \nabla g_i(x)^T f(x,w). \tag{2.30}$$

These expressions will be useful in the sequel, in particular when piecewise-linear functions will be dealt with.

[5]For example, the function $-x^2$ has empty subdifferential for all x.

2.3 Lyapunov functions and stability

In this section, some basic notions concerning Lyapunov stability of differential systems of equations are recalled. The concept of Lyapunov function is widely known in system theory. The main purpose of this section is not that of providing a further presentation of the theory. Conversely, but that of focusing on one aspect, precisely the relationship between the concept of Lyapunov (and Lyapunov-like) functions and the notion of invariant set.

Generically speaking, a Lyapunov function for a system is a positive definite function having the property that it is decreasing along the system trajectories. This property can be checked *without knowing the system trajectories* by means of the Lyapunov derivative. If a function Ψ of the state variables is non-increasing along the system trajectory, as an obvious consequence the set

$$\mathcal{N}[\Psi, \nu] = \{x : \ \Psi(x) \leq \nu\}$$

is positively invariant and precisely, if $x(t_1) \in \mathcal{N}[\Psi, \nu]$, then $x(t) \in \mathcal{N}[\Psi, \nu]$ for all $t \geq t_1$. Furthermore, if the function is strictly decreasing and the derivative is bounded away from zero, namely $\dot{\Psi}(x(t)) < -\gamma$, with $\gamma > 0$ in a set of the form

$$\mathcal{N}[\Psi, \alpha, \beta] = \{x : \ \alpha \leq \Psi(x) \leq \beta\},$$

then the condition $x(t_1) \in \mathcal{N}[\Psi, \alpha, \beta]$ implies (besides $x(t) \in \mathcal{N}[\Psi, \beta]$, $t \geq t_1$) that $x(t)$ reaches the smaller set $\mathcal{N}[\Psi, \alpha]$ in finite time[6]. Properties such as the mentioned one form the core of the book.

We remind once again that in the sequel it will always be assumed that

- any system of differential equations under consideration admits a solution for every initial condition and each piecewise continuous input;
- any solution is globally defined (i.e., defined on $\mathbb{R}^+$).

2.3.1 Global stability

Let us introduce the next definitions.

Definition 2.14 (Radially unbounded function). A locally Lipschitz function $\Psi : \mathbb{R}^n \to \mathbb{R}$ is radially unbounded if

$$\lim_{\|x\| \to \infty} |\Psi(x)| = \infty.$$

[6] $\Psi' \leq -\gamma$ means that this will happen at some time $t \geq t_1$, with $t \leq (\beta - \alpha)/\gamma + t_1$.

Definition 2.15 (κ-function). A continuous function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ is said to be a κ-function if it is continuous, strictly increasing and $\phi(0) = 0$.

Consider a model of the form

$$\dot{x}(t) = f(x(t), w(t)), \quad w(t) \in \mathcal{W}, \tag{2.31}$$

and assume that the following condition is satisfied

$$f(0, w) = 0, \quad \text{for all } w \in \mathcal{W} \tag{2.32}$$

which is well known to be equivalent to the fact that $x(t) \equiv 0$ is a trajectory of the system. Denote by $x(t)$ any solution of (2.31) corresponding to $x(0) \in \mathbb{R}^n$ and $w(t) \in \mathcal{W}$.

Definition 2.16 (Global uniform asymptotic stability). The system (2.31) is said to be Globally Uniformly Asymptotically Stable if it is

Locally Stable: for all $\nu > 0$ there exists $\delta > 0$ such that if $\|x(0)\| \leq \delta$ then

$$\|x(t)\| \leq \nu, \quad \text{for all } t \geq 0; \tag{2.33}$$

for all functions $w(t) \in \mathcal{W}$.

Globally Attractive: for all $\mu > 0$ and $\epsilon > 0$ there exists $T(\mu, \epsilon) > 0$ such that if $\|x(0)\| \leq \mu$ then

$$\|x(t)\| \leq \epsilon, \quad \text{for all } t \geq T(\mu, \epsilon) \tag{2.34}$$

for all functions $w(t) \in \mathcal{W}$.

Since local stability and global attractivity properties are requested to hold for all functions w, the above property is often referred to as Robust Global Uniform Asymptotic Stability. The meaning of the above definition is the following: For any neighborhood of the origin the evolution of the system is bounded inside it provided the system is initialized sufficiently close to 0 and, moreover, the state converges to zero uniformly in the sense that for all the initial states inside a μ-ball, the ultimate capture of the state inside any ϵ-ball occurs in a time that admits an upper bound independent from $w(t)$.

Definition 2.17 (Positive definite function). A function $\Psi : \mathbb{R}^n \to \mathbb{R}$ is positive definite if $\Psi(0) = 0$ and there exists a κ-function ϕ_0 such that

$$\Psi(x) \geq \phi_0(\|x\|)$$

Definition 2.18 (Global Lyapunov Function). A locally Lipschitz function $\Psi : \mathbb{R}^n \to \mathbb{R}$ is a said to be a Global Lyapunov Function (GLF) for the system if it is positive definite, radially unbounded and there exists a κ-function ϕ such that

$$D^+\Psi(x, w) \leq -\phi(\|x(t)\|) \tag{2.35}$$

For differential equations $\dot{x} = f(x)$, with a continuous f, the previous condition can be written as $D^+\Psi(x) < 0$ for $x \neq 0$. In the presence of uncertainty this is not sufficient (see Exercise 4).

The following theorem is a well-established result in system theory. Its first formulation is due to Lyapunov [Lya66] and several other versions have been introduced in the literature.

Theorem 2.19. *Assume that system (2.31) admits a global Lyapunov function Ψ. Then it is globally uniformly asymptotically stable.*

Proof. From Theorem 2.11 we have that

$$\Psi(x(t)) - \Psi(x(0)) = \int_0^t D^+(x(\sigma), w(\sigma))\, d\sigma \leq -\int_0^t \phi(\|x(\sigma)\|)\, d\sigma \tag{2.36}$$

To show uniform stability, let $\nu > 0$ be arbitrary and let ξ be any positive value such that $\mathcal{N}[\Psi, \xi] \subseteq \mathcal{N}[\|\cdot\|, \nu]$ (the ν-ball of any norm). Since $\Psi(x)$ is positive definite, such a $\xi > 0$ exists and moreover there exists $\delta > 0$ such that $\mathcal{N}[\|\cdot\|, \delta] \subseteq \mathcal{N}[\Psi, \xi]$. Then, for all $\|x(0)\| \leq \delta$, $\Psi(x(0)) \leq \xi$. Since $\Psi(x(t))$ is monotonically non-increasing, then $x(t) \in \mathcal{N}[\Psi, \xi] \subseteq \mathcal{N}[\|\cdot\|, \nu]$, say $\|x(t)\| \leq \nu$.

To prove uniform convergence, it has to be shown that for any given $\mu > 0$ and any arbitrary small $\epsilon > 0$, there exists $T(\mu, \epsilon)$ (which does not depend on $w(t)$ and $x(0)$) such that all the solutions originating from the μ-ball of the norm, $x(0) \in \mathcal{N}[\|\cdot\|, \mu]$, are ultimately confined in the ϵ-ball $x(t) \in \mathcal{N}[\|\cdot\|, \epsilon]$, for $t \geq T(\mu, \epsilon)$.

Take $\rho^* < \infty$ such that

$$\mathcal{N}[\|\cdot\|, \mu] \subseteq \mathcal{N}[\Psi, \rho^*]$$

(for instance, one can take $\rho^* = \max_{\|x\| \leq \mu} \Psi(x)$, the smallest value such that the inclusion holds). For any arbitrary small ϵ, take $\rho_* > 0$ such that

$$\mathcal{N}[\Psi, \rho_*] \subset \mathcal{N}[\|\cdot\|, \epsilon]$$

(again, one can always take the largest of such values which is necessarily greater than zero since Ψ is positive definite). Consider the set

$$\mathcal{N}[\Psi, \rho_*, \rho^*] = \{x : \rho_* \leq \Psi(x) \leq \rho^*\}$$

which is compact, being Ψ radially unbounded, and let ζ be

$$\zeta \doteq \min_{x \in \mathcal{N}[\Psi, \rho_*, \rho^*]} \phi(\|x\|) > 0$$

For $x(0) \in \mathcal{N}[\|\cdot\|, \mu] \subset \mathcal{N}[\Psi, \rho^*]$, in view of the previous considerations we have $x(t) \in \mathcal{N}[\Psi, \rho^*]$. Then necessarily, in a finite time $\bar{t}$, we must have $x(\bar{t}) \in \mathcal{N}[\Psi, \rho_*]$. Indeed if this condition is not satisfied, we must have $x(t) \in \mathcal{N}[\Psi, \rho_*, \rho^*]$. From the

integral equality (2.36), we achieve the bound (independent of $w(t)$)

$$\Psi(x(t)) = \Psi(x(0)) + \int_0^t D^+(x(\sigma), w(\sigma))\, d\sigma \leq \Psi(x(0)) - \zeta t \leq$$
$$\leq \rho^* - \zeta t$$

it is immediate to see that the rightmost term becomes smaller than ρ_* when

$$t \geq T(\mu, \epsilon) = [\rho^* - \rho_*]/\zeta$$

thus necessarily $x(t)$ reaches $\mathcal{N}[\Psi, \rho_*]$ no later than $T(\mu, \epsilon)$.

To complete the proof it is then sufficient to recall that, since $\Psi(x(t))$ is non-increasing, $x(t) \in \mathcal{N}[\Psi, \rho_*]$, therefore $\|x(t)\| \leq \epsilon$, for $t \geq T(\mu, \epsilon)$.

A stronger notion of stability, which will be often used in the sequel, is the following:

Definition 2.20 (Global exponential stability). System (2.31) is said to be Globally Exponentially Stable if there exist $\mu, \gamma > 0$ such that for all $\|x(0)\|$ the condition

$$\|x(t)\| \leq \mu \|x(0)\| e^{-\gamma t}, \tag{2.37}$$

holds for every $t \geq 0$ and every function $w(t) \in \mathcal{W}$.

The factor γ in the definition above will be named the convergence speed, while the factor μ will be named the transient estimate. Exponential stability can be assured by the existence of a Lyapunov function whose decreasing rate along the system trajectories is expressed in terms of the magnitude of the function. Let us assume that the positive definite function $\Psi(x)$ is upper and lower polynomially bounded, namely for some positive reals α and β and some positive integer p

$$\alpha \|x\|^p \leq \Psi(x) \leq \beta \|x\|^p, \quad \text{for all } \; x \in \mathbb{R}^n. \tag{2.38}$$

The following theorem holds true.

Theorem 2.21. *Assume that system (2.31) admits a positive definite locally Lipschitz function Ψ, which has polynomial growth as in (2.38) and*

$$D^+\Psi(x, w) \leq -\gamma \Psi(x) \tag{2.39}$$

for some positive γ. Then it is globally exponentially stable.

Proof. Consider the integral equality (2.26) and bound it as

$$\Psi(x(t+T)) = \Psi(x(t)) + \int_t^{t+T} D^+\Psi(x(\sigma), w(\sigma)) d\sigma$$

$$\leq \Psi(x(t)) - \gamma \int_t^{t+T} \Psi(x(\sigma))d\sigma \leq \Psi(x(t)) - \gamma T \Psi(x(t+T))$$

where the last inequality follows by the fact that $\Psi(x(t))$ is non-increasing. This implies

$$\Psi(x(t+T)) \leq \frac{1}{1+T\gamma}\Psi(x(t))$$

Therefore, for all integer k

$$\Psi(x(kT)) \leq \left[\frac{1}{1+T\gamma}\right]^k \Psi(x(0))$$

Let now $t > 0$ be arbitrary and let $T = t/k$, with k integer so that the above can be rewritten as

$$\Psi(x(t)) \leq \left\{\left[\frac{1}{1+\gamma t/k}\right]^{-\frac{k}{\gamma t}}\right\}^{-\gamma t} \Psi(x(0)) = \left\{[1+\gamma t/k]^{\frac{k}{\gamma t}}\right\}^{-\gamma t} \Psi(x(0))$$

The number inside the curly brackets converges to e as $k \to \infty$ and then, since the inequality holds for any k,

$$\Psi(x(t)) \leq e^{-\gamma t}\Psi(x(0))$$

Finally, exploiting condition (2.38), after simple mathematics the following inequality holds

$$\|x(t)\| \leq \sqrt[p]{\frac{\beta}{\alpha}}\, e^{-\frac{\gamma}{p}t}\, \|x(0)\|$$

which implies exponential convergence with convergence speed γ/p.

The previous theorem admits a trivial proof if one assumes that $\Psi(x(t))$ is differentiable in the regular sense, since the inequality (2.39) becomes the differential inequality $\dot{\Psi}(x(t)) \leq -\gamma\Psi(x)$, implying $\Psi(x(t)) \leq e^{-\gamma t}\Psi(x(0))$.

It is rather obvious that global exponential stability implies global uniform asymptotic stability.

2.3.2 *Local stability and ultimate boundedness*

Global stability can be somewhat a too ambitious requirement in practical control theory, basically for the next two reasons.

- requiring convergence with arbitrary initial conditions can be too restrictive;
- persistent disturbances can prevent the system from asymptotically approaching the origin, thus the best we can get is convergence to a set.

For the above reasons, it is very useful to introduce the notion of local stability and uniform ultimate boundedness.

Definition 2.22 (Uniform local asymptotic stability). Let $\mathcal{S}$ be a neighborhood of the origin. The system (2.31) is said to be Uniformly Locally Asymptotically Stable with basin (or domain) of attraction $\mathcal{S}$ if, for every function $w(t) \in \mathcal{W}$, the next two conditions hold:

Local Stability : for all $\mu > 0$ there exists $\delta > 0$ such that $\|x(0)\| \leq \delta$ implies $\|x(t)\| \leq \mu$ for all $t \geq 0$.

Local Uniform Convergence : for all $\epsilon > 0$ there exists $T(\epsilon) > 0$ such that if $x(0) \in \mathcal{S}$, then $\|x(t)\| \leq \epsilon$, for all $t \geq T(\epsilon)$;

Definition 2.23 (Uniform ultimate boundedness). Let $\mathcal{S}$ be a neighborhood of the origin. The system (2.31) is said to be Uniformly Ultimately Bounded in $\mathcal{S}$ if for all $\mu > 0$ there exists $T(\mu) > 0$ such that, for every $\|x(0)\| \leq \mu$,

$$x(t) \in \mathcal{S}$$

for all $t \geq T(\mu)$ and all functions $w(t) \in \mathcal{W}$.

To assure the conditions of the above definitions, the following concepts of Lyapunov functions *inside* and *outside* $\mathcal{S}$ are introduced.

Definition 2.24 (Lyapunov function inside a set). Let $\mathcal{S}$ be a neighborhood of the origin. The locally Lipschitz positive definite function Ψ is said to be a Lyapunov function inside $\mathcal{S}$ for system (2.31) if there exists $\nu > 0$ such that

$$\mathcal{S} \subseteq \mathcal{N}[\Psi, \nu]$$

and for all $x \in \mathcal{N}[\Psi, \nu]$ the inequality

$$D^+\Psi(x, w) \leq -\phi(\|x(t)\|)$$

holds for some κ-function ϕ and all $w \in \mathcal{W}$.

Definition 2.25 (Lyapunov function outside a set). Let $\mathcal{S}$ be a neighborhood of the origin. The locally Lipschitz positive definite function Ψ is said to be a Lyapunov function outside $\mathcal{S}$ for system (2.31) if there exists $\nu > 0$ such that

$$\mathcal{N}[\Psi, \nu] \subseteq \mathcal{S}$$

and for all $x \notin \mathcal{N}[\Psi, \nu]$ the inequality

$$D^+\Psi(x, w) \leq -\phi(\|x(t)\|)$$

holds for some κ-function ϕ.

The next two theorems hold.

Theorem 2.26. *Assume that system (2.31) satisfying condition (2.32) admits a positive definite locally Lipschitz function Ψ inside $\mathcal{S}$. Then, it is locally stable with basin (domain) of attraction $\mathcal{S}$.*

Theorem 2.27. *Assume that system (2.31) admits a positive definite locally Lipschitz function Ψ outside $\mathcal{S}$. Then it is uniformly ultimately bounded in $\mathcal{S}$.*

It is rather intuitive that there are as many possible stability definitions as the number of possible permutations of the requirements (Global-Local-Uniform-Exponential and so on ...). For instance, one can define exponential local stability by requiring condition (2.37) to be satisfied only for $x(0) \in \mathcal{S}$. Similarly, exponential ultimate boundedness in the set $\mathcal{S}$ can be defined by imposing that $\mathcal{N}[\|\cdot\|, \nu] \subset \mathcal{S}$ and $\|x(t)\| \leq \max\{\mu e^{-\gamma t}\|x(0)\|, \nu\}$. The problem is well known and in the literature some classifications of stability concepts have been proposed (see [RHL77] section VI), further investigation in this sense is beyond the scope of this book.

Remark 2.28. It is apparent that a Lyapunov function is not only a stability certificate. Indeed the sets $\mathcal{N}[\Psi, \nu]$ (for appropriate values of ν) are positively invariant, a property whose consequences are among the main concerns of the book.

2.4 Control Lyapunov function

In the previous section, the main results of the Lyapunov theory for a dynamical system with an external input have been presented. These concepts will now be extended to systems of the form (2.1)–(2.2) with a controlled input. Essentially, a Control Lyapunov Function can be defined as a positive definite (locally Lipschitz) function which becomes a Lyapunov function whenever a proper control action is applied.

As previously observed, any finite-dimensional dynamic feedback controller can be viewed as a static output feedback for a properly augmented system. Therefore, in this section, systems of the following form

$$\begin{cases} \dot{x}(t) = f(x(t), u(t), w(t)) \\ y(t) = h(x(t), w(t)) \end{cases} \tag{2.40}$$

will be considered and associated with a static feedback. To introduce the main definitions one has to refer to a class $\mathcal{C}$ of controllers. The main classes considered here are

Output feedback : $u(t) = \Phi(y(t))$;
State feedback : $u(t) = \Phi(x(t))$;
Output feedback with feed forward : $u(t) = \Phi(y(t), w(t))$;
State feedback with feed forward : $u(t) = \Phi(x(t), w(t))$ (full information);

Definition 2.29 (Control Lyapunov function). Given a class of controllers $\mathcal{C}$ and a set $\mathcal{P}$, a locally Lipschitz positive definite function Ψ is said to be a global Control Lyapunov Function (a CLF outside $\mathcal{P}$ or a CLF function inside $\mathcal{P}$) if there exists a controller in $\mathcal{C}$ such that:

- for each initial condition $x(0)$ there exists a solution $x(t)$ for any admissible $w(t)$ and each of such solutions is defined for all $t \geq 0$;
- the function Ψ is a global Lyapunov function (a Lyapunov function outside $\mathcal{P}$ or a Lyapunov function inside $\mathcal{P}$) for the closed-loop system.

An interesting generalization of the previous definition is achieved by considering the control constraints (2.13)

$$u(t) \in \mathcal{U}.$$

In this case Ψ is said to be a global Control Lyapunov Function (a CLF outside $\mathcal{P}$ or a CLF inside $\mathcal{P}$) if there exists a controller (in a specified class $\mathcal{C}$) such that, beside the conditions in Definition 2.29, the control constraints are also satisfied. Note also that the problem with state constraints can be addressed as well. If one assumes that

$$x(t) \in \mathcal{X}$$

is a hard constraint to be satisfied, it can be immediately argued that, as long as $\mathcal{N}[\Psi, \mu] \subseteq \mathcal{X}$, for some μ, and Ψ is a control Lyapunov function (either global, inside or outside $\mathcal{P}$), then the constraints can be satisfied by means of a proper control action as long as $x(0) \in \mathcal{N}[\Psi, \mu]$.

2.4.1 *Associating a control law with a Control Lyapunov Function: state feedback*

In this section, the state feedback and the full information feedback cases will be mainly analyzed. According to the previous considerations the following domain will be considered

$$\mathcal{N}[\Psi, \alpha, \beta] = \{x : \ \alpha \leq \Psi(x) \leq \beta\} \tag{2.41}$$

By letting $\alpha = 0$ or $\beta = +\infty$, the above class includes all the "meaningful" cases of control Lyapunov functions (inside a set, outside a set or global). Assume that a locally Lipschitz function Ψ is given and consider the next inequality

$$D^+\Psi(x,u,w) \doteq \limsup_{h\to 0^+} \frac{\Psi(x+hf(x,u,w)) - \Psi(x)}{h} \leq -\phi(\|x\|). \tag{2.42}$$

Again, the above is a simplified version of the "appropriate notation"

$$D^+\Psi(x,u,w) = D^+\Psi(x,f(x,u,w))$$

Consider the next two conditions in the set (2.41):

- for all x there exists u such that (2.42) is satisfied for all $w \in \mathcal{W}$;
- for all x and $w \in \mathcal{W}$ there exists u such that (2.42) is satisfied.

These conditions are clearly necessary for Ψ to be a control Lyapunov function with state or full information feedback, because, by definition, they are satisfied by assuming $u = \Phi(x)$ or $u = \Phi(x,w)$, respectively. A fundamental question is then the following: if these conditions hold, how can a feedback function Φ be found? The problem can be thought of in the following terms. Let us first analyze the state feedback case. Consider the set

$$\Omega(x) = \{u : \ \text{(2.42) is satisfied for all } w \in \mathcal{W}\}$$

which is known as the regulation map or control map. Then the question becomes the following: does there exist a state feedback control function $u = \Phi(x)$ such that

$$\Phi(x) \in \Omega(x)?$$

Put in this term, this question appears to be a philosophic one, because, as long as the set $\Omega(x)$ is not empty, it is always possible to associate to x a point $u \in \Omega(x)$, and "define" such a function Φ. The matter is different if one requires the function to satisfy certain regularity properties such as that of being continuous. From a mathematical point of view, continuity is important because it implies solvability. From a practical standpoint, continuity is important because discontinuous controllers cannot be always applied. Furthermore it is fundamental to derive some formulas suitable for the implementation.

A positive answer to our question can be given for control-affine systems, namely systems of the form

$$\dot{x}(t) = a(x(t),w(t)) + b(x(t),w(t))u(t) \tag{2.43}$$

where a and b are continuous and $a(0,w) = 0$ for all $w \in \mathcal{W}$. In this case, if a continuously differentiable positive definite function Ψ is given and for such

function (2.42) is satisfied for all x, from the differentiability of Ψ one has that (2.42) can be written as follows:

$$\nabla\Psi(x)^T[a(x,w)+b(x,w)u] \leq -\phi(\|x\|),$$

Then the set $\Omega(x)$ turns out to be

$$\Omega(x) = \left\{u:\ \nabla\Psi(x)^T b(x,w)u \leq -\nabla\Psi(x)^T a(x,w) - \phi(\|x\|),\ \text{for all } w \in \mathcal{W}\right\} \tag{2.44}$$

Such a non-empty set is convex for each x and the continuity of a and b and $\nabla\Psi(x)$ allow to state the next theorem.

Theorem 2.30. *Assume that the set $\Omega(x)$ as in (2.44) is non-empty. Then there always exists a function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ continuous everywhere, possibly with the exception of the origin, such that*

$$\Phi(x) \in \Omega(x) \tag{2.45}$$

Proof. See [FK96a]

The previous theorem considers the fundamental concept of *selection* of a set-valued map. A set-valued map f from $\mathcal{X}$ to $\mathcal{Y}$ is a multivalued function which associates with any element x of $\mathcal{X}$ a subset Y of $\mathcal{Y}$. A selection is a single-valued function which maps x into one of the elements in $Y = f(X)$. In our case, $\Omega(x)$ is the set-valued map of all the feasible control values which assure a certain decreasing rate to the Lyapunov function.

In the case of full-information control, the appropriate set-valued map must be defined in the state-disturbance product space:

$$\Omega(x,w) = \left\{u:\ \ \nabla\Psi(x)^T b(x,w)u \leq -\nabla\Psi(x)^T a(x,w) - \phi(\|x\|)\right\} \tag{2.46}$$

If this set is not empty for all x and w, then it is possible to seek for a function

$$\Phi(x,w) \in \Omega(x,w) \tag{2.47}$$

which is a stabilizing full-information control. In view of the convexity of the set $\Omega(x,w)$ and the continuity of a and b and $\nabla\Psi(x)$, it can be shown that a continuous selection always exists, namely, that Theorem 2.30 can be stated by replacing (2.45) with (2.47).

Although mathematically appealing, the engineering question which obviously comes next is the following: Is it possible to determine this function in an analytic form?

To this aim, let us think first to the full information case. Let us also consider, for the moment being, that the region of interest is of the form $\mathcal{N}[\Psi,\epsilon,\kappa]$ with k finite and ϵ small. Assume that there exists a continuous function $\hat{\Phi}(x,w)$ which satisfies (2.47) and consider the following minimum effort control [PB87]

$$\Phi_{ME}(x,w) = \arg \min_{u \in \Omega(x,w)} \|u\|_2 \tag{2.48}$$

($\|\cdot\|_2$ is the Euclidean norm). Such a control function always exists and it has the obvious property that

$$\|\Phi_{ME}(x,w)\|_2 \leq \|\hat{\Phi}(x,w)\|_2$$

for any admissible controller $\hat{\Phi}(x,w)$, therefore it is bounded in $\mathcal{N}[\Psi,\epsilon,\kappa]$. The minimum effort control admits an analytic expression which can be easily derived as follows. For fixed x and w, equation (2.46) represents a linear inequality for u:

$$\nabla\Psi(x)^T b(x,w)u \leq -c(x,w) \tag{2.49}$$

where

$$c(x,w) \doteq \nabla\Psi(x)^T a(x,w) + \phi(\|x\|)$$

The vector u of minimal norm which satisfies (2.49) can be determined analytically as follows

$$\Phi_{ME}(x,w) = \begin{cases} -\frac{b(x,w)^T\nabla\Psi(x)}{\|\nabla\Psi(x)^T b(x,w)\|^2}\, c(x,w) & \text{if } c(x,w) > 0 \\ 0 & \text{if } c(x,w) \leq 0 \end{cases} \tag{2.50}$$

The singularity due to the condition $\nabla\Psi(x)^T b(x,w) = 0$, for some x and w, is not a problem since this automatically implies that $c(x,w) \leq \phi(\|x\|) < 0$ (hence $\Phi_{ME} = 0$: this is basically the reason of working inside the set $\mathcal{N}[\Phi,\epsilon,\kappa]$ with small but positive ϵ, say to exclude $x = 0$). This expression admits an immediate extension to the state feedback case if one assumes that the term b does not depend on w and precisely

$$\dot{x}(t) = a(x(t),w(t)) + b(x(t))u(t)$$

In this case the condition becomes

$$\nabla\Psi(x)^T b(x)u \leq -c(x,w) \tag{2.51}$$

where

$$c(x,w) \doteq \nabla\Psi(x)^T a(x,w) + \phi(\|x\|)$$

which has to be satisfied for all w, by an appropriate choice of u.

Take $\epsilon > 0$ very small, (to avoid singularities) and define

$$\hat{c}_\epsilon(x) = \max_{w \in \mathcal{W}} \nabla\Psi(x)^T a(x,w) + (1-\epsilon)\phi(\|x\|)$$

which is a continuous function of x [PB87, FK96a].[7] The condition to be considered is then

$$\nabla\Psi(x)^T b(x)u \leq -\hat{c}_\epsilon(x) \tag{2.52}$$

which yields the following expression for the control:

$$\Phi_{ME}(x) = \begin{cases} -\frac{b(x)^T\nabla\Psi(x)}{\|\nabla\Psi(x)^T b(x)\|^2}\,\hat{c}_\epsilon(x) & \text{if } \hat{c}_\epsilon(x) > 0 \\ 0 & \text{if } \hat{c}_\epsilon(x) \leq 0 \end{cases} \tag{2.53}$$

The minimum effort control (2.53) belongs to the class of gradient-based controllers of the form

$$u(t) = -\gamma(x)b(x)^T\nabla\Psi(x) \tag{2.54}$$

where $\gamma(x) \in \mathbb{R}^+$. This type of controllers is well known and includes other types of control functions. For instance, if the control effort is not a concern, one can just consider (2.54) with $\gamma(x) > 0$ "sufficiently large function" [BCL83]. It can be shown that, as long as (2.53) is a suitable controller, any function (if it exists) having the property

$$\gamma(x) \geq \bar{\gamma}_\epsilon(x) \doteq \max\left\{\frac{\hat{c}_\epsilon(x)}{\|\nabla\Psi(x)^T b(x)\|^2}, 0\right\} \tag{2.55}$$

is also a suitable controller. The following proposition holds.

Proposition 2.31. *Function $\bar{\gamma}_\epsilon(x)$ in (2.55) is continuous.*

Proof. The expression is obviously continuous in each point x for which $\nabla\Psi(x)^T b(x) \neq 0$. Let us consider a point $\hat{x}$ in which $\nabla\Psi(\hat{x})^T b(\hat{x}) = 0$. In such a point

$$\nabla\Psi(\hat{x})^T a(\hat{x}, w) + \phi(\|\hat{x}\|) \leq 0$$

for all w, by definition, and therefore

$$\hat{c}_\epsilon(\hat{x}) \leq -\epsilon\phi(\|\hat{x}\|)$$

This means that $\hat{c}_\epsilon(x)$ is strictly negative in a neighborhood of $\hat{x}$ hence $\bar{\gamma}_\epsilon(x)$ is identically 0 in such neighborhood.

The problem becomes more involved if one admits that also the term b depends on the uncertain parameter. In this case finding a state feedback controller is related to the following min–max problem

[7] $\hat{c}(x)$ is referred to as Marginal Function.

$$\min_{u\in\mathcal{U}} \max_{w\in\mathcal{W}} \left\{\nabla\Psi(x)^T[a(x,w)+b(x,w)u]\right\} \leq -\phi(\|x\|), \tag{2.56}$$

where, to keep things as general as possible, it is also assumed that $u \in \mathcal{U}$. If this condition is pointwise-satisfied, then there exists a robustly stabilizing feedback control. However, determining the minimizer function $u = \Phi(x)$ can be very hard.

Example 2.32 (Electric circuit revisited). Consider again the system considered in Example 2.4 which can be written as

$$\dot{x} = -\Lambda(w)x + B(w)u$$

where $w \in \mathcal{W}$ is the vector including the resistor and capacitor values ($\mathcal{W}$ is typically an hyper-rectangle). In this case

$$A(w) = -\Lambda(w) = -diag\{\Lambda_1(w), \Lambda_2(w)\}$$

where $\Lambda(w)$ is a strictly positive and diagonal matrix. To avoid the singularity problems described in Example 2.4, it is assumed that $\det(B(w)) \geq \nu > 0$ for all $w \in \mathcal{W}$.

Consider a nominal equilibrium point $\Lambda(\bar{w})\bar{x} = B(\bar{w})\bar{u}$ corresponding to a nominal value $\bar{w} \in \mathcal{W}$, and define the variables $z = x - \bar{x}$, $v = u - \bar{u}$, so that

$$\dot{z} = -\Lambda(w)z + B(w)v + \Delta(w)$$

For any fixed $\bar{x}$ and $\bar{u}$, it is clear that there exists a positive value μ such that

$$\|\Delta(w)\| = \|(\Lambda(\bar{w}) - \Lambda(w))\bar{x} + (B(w) - B(\bar{w}))\bar{u}\| \leq \mu$$

Let us consider the problem of keeping the state x as close as possible to the nominal value $\bar{x}$ (thus $z \simeq 0$) and, to this aim, consider the Lyapunov function $\Psi(z) = (z_1^2 + z_2^2)/2$, whose derivative is

$$\dot{\Psi}(z) = -\Lambda_1(w)z_1^2 - \Lambda_2(w)z_2^2 + z^T B(w)v + z^T\Delta(w)$$

If w is available, since B is invertible, the problem becomes quite easy since it is readily seen that the control

$$v(t) = -\gamma B(w)^{-1}z(t)$$

is such that

$$\dot{\Psi}(z,w) = -\Lambda_1 z_1^2 - \Lambda_2 z_2^2 - \gamma z^T z + z^T\Delta(w)$$

and, by taking γ large enough, it is possible to ensure that $\dot{\Psi}(z)$ is strictly negative outside any ϵ-ball (the computations are straightforward), so that the state is uniformly ultimately bounded inside $\mathcal{N}[\Psi, \epsilon]$.

If w is not available for control, the situation is quite different. A solution is that of trusting the nominal value $\bar{w}$ and consider

$$u(t) = -\gamma B(\bar{w})^{-1} z(t)$$

To see how the proposed "trust-based" control law works, let $B(w) = B(\bar{w}) + \delta B(w)$ and assume $B(w)B(\bar{w})^{-1}$ remains close to the identity, say

$$\|B(w)B(\bar{w})^{-1} - I\| = \|\delta B(w)B(\bar{w})^{-1}\| \leq \nu < 1$$

(this is a typical assumption when B is uncertain). If the previously introduced Lyapunov function is analyzed, then its Lyapunov derivative results in

$$\begin{aligned}
\dot{\Psi}(z, w) &= -\Lambda_1 z_1^2 - \Lambda_2 z_2^2 - \gamma z^T [B(\bar{w}) + \delta B(w)] B(\bar{w})^{-1} z + z^T \Delta(w) \\
&= -\Lambda_1 z_1^2 - \Lambda_2 z_2^2 - \gamma z^T z - \gamma z^T \delta B(w) B(\bar{w})^{-1} z + z^T \Delta(w) \\
&\leq -\Lambda_1 z_1^2 - \Lambda_2 z_2^2 - \gamma(1 - \nu) z^T z + z^T \Delta(w)
\end{aligned}$$

Again, the same argument used for the nominal case applies and then the state can be confined in any arbitrarily small neighborhood of the desired value via state feedback provided that $\|B(w)B(\bar{w})^{-1} - I\|$ is "small enough."

Going back to the min–max problem in Eq. (2.56), where the control (the good guy) had to play ignoring the action of the opponent, it is worth mentioning an important fact, precisely its relation with the corresponding full information problem

$$\max_{w \in \mathcal{W}} \min_{u \in \mathcal{U}} \left\{ \nabla \Psi(x)^T [a(x, w) + b(x, w)u] \right\} \leq -\phi(\|x\|), \tag{2.57}$$

in which the "min" and the "max" are reversed (say the disturbance plays first). If condition (2.57) is satisfied, then there exists a full information control. In fact condition (2.56) always implies (2.57). There are important classes of systems for which the two conditions are equivalent. For instance, in the case in which b does depend on x only, the two problems are equivalent. This means that the existence of a full-information stabilizing controller implies the existence of a pure state feedback controller [Meı79]. A further class of control affine uncertain systems for which the same property holds is that of the so-called convex processes [Bla00].

There are other forms of controllers which can be associated with a control Lyapunov function. An interesting class, strongly related to the minimum effort control, is the limited-effort control. Assume that the control is constrained as

$$\|u(t)\| \leq 1. \tag{2.58}$$

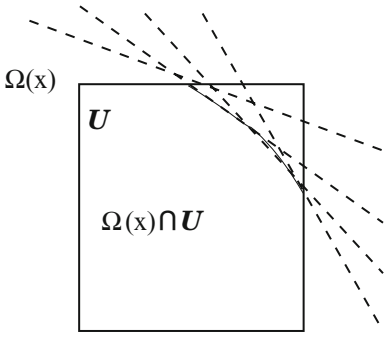


Fig. 2.6 The set $\Omega(x) \bigcap \mathcal{U}$

Note that this expression is nothing but a state (and disturbance) independent constraint which adds to the constraint coming from (2.44) (the dashed ones depicted in Figure 2.6). The figure represents the set $\Omega(x) \bigcap \mathcal{U}$ derived by considering, for instance, Example 2.32 by allowing some parameter variations. The square represents the set $\mathcal{U} = \{u : \|u(t)\|_\infty \leq 1\}$ while the dashed curved line represents the envelope of lines of the form

$$\dot{\Psi}(z, u, w) = -z^T \Lambda(w) + z^T B(w)u \leq -\phi(\|x\|),$$

each associated with a certain value of the parameter.

Note that, in general, if Ψ is smooth, the resulting set is convex, since it is the intersection of half-planes and a convex $\mathcal{U}$, no matter how $a(x, w)$ and $b(x, w)$ are chosen.

Assuming the magnitude equal to 1 is not a restriction since the actual magnitude or weighting factors can be discharged on $b(x, w)$. Two norms are essentially worth of consideration, and precisely the 2-norm $\|u\|_2 = \sqrt[2]{u^T u}$ and the ∞-norm $\|u\|_\infty = \max_j |u_j|$. Consider the case of local stability and assume that a control Lyapunov function inside $\mathcal{N}[\Psi, \kappa]$ exists and that the associated stabilizing control law does not violate the constraint (2.58).

This in turn means that, when minimizing the Lyapunov derivative; hence there exists a minimizer which satisfies the imposed bound and results in a negative derivative, say the solution to the minimization problem

$$\min_{\|u\| \leq 1} \nabla\Psi(x)^T [a(x, w) + b(x)u]$$

is strictly negative (if the origin, obviously, is excluded). For instance, if the 2-norm is considered, this control is

$$u = \begin{cases} -\frac{b(x)^T \nabla\Psi(x)}{\|b(x)^T \nabla\Psi(x)\|} & \text{if } \nabla\Psi(x)^T b(x) \neq 0, \\ 0 & \text{if } \nabla\Psi(x)^T b(x) = 0 \end{cases} \tag{2.59}$$

whereas, in the case of infinity norm, one obtains

$$u = -\text{sgn}[b(x)^T \nabla \Psi(x)] \tag{2.60}$$

(the sign function sgn is defined in the notation section 1.1). Both the controls (2.59) and (2.60) are discontinuous, but can be approximated by the continuous controllers:

$$u = -\frac{b(x)^T \nabla \Psi(x)}{\|b(x)^T \nabla \Psi(x)\| + \delta}$$

for δ sufficiently small and

$$u = -\text{sat}[\gamma b(x)^T \nabla \Psi(x)]$$

for γ sufficiently large, respectively (sat is the vector saturation function). It is left to the reader as an exercise to show that for all ϵ there exists a sufficiently small δ (sufficiently large γ) such that the system is ultimately bounded inside $\mathcal{N}[\Psi, \epsilon]$.

The literature about control Lyapunov functions is huge, since this concept is of fundamental importance in the control of nonlinear systems. Here we do not enter into further details, but we rather refer to specialized work such as [FK96a, FK96b, Qu98].

2.4.2 *Associating a control law with a Control Lyapunov Function: output feedback*

As far as the output feedback case is concerned, the problem of determining a control to associate with a Control Lyapunov Function is much harder and the problem will be marginally faced with the aim of explaining the reasons of these difficulties. Consider the system

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

and a candidate Control Lyapunov function Ψ defined on $\mathbb{R}^n$. Consider a static output feedback of the form

$$u(t) = \Phi(y(t))$$

Since only output information is available, the control value u that assures a negative Lyapunov derivative must assure this property for a given output value y for all

possible values of x and w that produce that output. Let $\mathcal{Y}$ be the image of h: $\mathcal{Y} = \{y = h(x,w),\ \ x \in \mathbb{R}^n, w \in \mathcal{W}\}$. Given $y \in \mathcal{Y}$, define the preimage set as

$$h^{-1}(y) = \{(x,w),\ \ x \in \mathbb{R}^n,\ w \in \mathcal{W} :\quad y = h(x,w)\ \}$$

The condition for Ψ to be a control Lyapunov function under output feedback is then the following. Consider the set

$$\Omega(y) = \{u :\ \ D^+\Psi(x,w,u) \leq -\phi(\|x\|),\ \ \text{for all}\ \ (x,w) \in h^{-1}(y)\}$$

Then a necessary and sufficient condition for $\Phi(y)$ to be a proper control function is that

$$\Phi(y) \in \Omega(y), \quad \text{for all}\ \ y \in \mathcal{Y} \tag{2.61}$$

This theoretical condition is simple to state, but useless in most cases, since the set $\Omega(y)$ can be hard (not to say impossible) to determine. It is not difficult to find similar theoretical conditions for a control of the form $\Phi(y,w)$, which are again hard to apply.

2.4.3 Finding a control Lyapunov function

So far, we considered the problem of associating a control with a Control Lyapunov Function. Such task is evidently subsequent to "the Problem" of finding such a function. Although there are lucky cases in which the function can be determined, in general such problem is very difficult to solve even with state feedback (the output feedback case is much worse). Special classes of systems are, for instance, those that admit the so-called strict feedback form (and the "extended strict feedback form") and the reader is referred to specialized literature for further details [FK96a, FK96b, Qu98].

As an exception, in the case of linear uncertain systems, there are effective algorithms to determine a control Lyapunov function that will be discussed later. This fact is meaningful since we have seen that a nonlinear plant can be often merged (locally) in a linear uncertain dynamics, as in the levitator case. However, a general constructive Lyapunov theory for the control of uncertain systems is missing. Therefore it seems reasonable to approach the problem in a case-by-case spirit. Indeed there are significant classes of systems originating from practical problems for which the construction of a control Lyapunov function is possible (often derived by intuition rather than by exploiting a general theory).

2.4.4 Classical methods to find Control Lyapunov Functions

A well-known way to find a control Lyapunov function for a system is to derive it from its linearization. We briefly remind some basic concepts from the general control theory on this topic. Any sufficiently smooth nonlinear system having an equilibrium point, which is assumed to be the origin, can be written as

$$\dot{x}(t) = f(x(t), u(x)) = Ax(t) + Bu(t) + R(x(t), u(x))$$

where $R(x,u)$ is an infinitesimal of order greater than one. A locally stabilizing control can be found by considering a linear feedback $u = Kx$ such that $A_{CL} = A + BK$ has eigenvalues with negative real part. The so achieved closed-loop system satisfies the Lyapunov equation

$$A_{CL}^T P + PA_{CL} = -Q$$

where Q is an arbitrary positive definite symmetric matrix. The Lyapunov function for the linearized system is

$$\Psi(x) = x^T Px, \quad \text{with } P > 0$$

Such a function is the candidate Lyapunov function for the original nonlinear system. The essential idea is that $\Psi(x)$ is not only a Lyapunov function for the closed-loop linear system, but it is also a Lyapunov function for the closed-loop nonlinear system valid in a proper neighborhood. Obviously, $\Psi(x)$ is a control-Lyapunov function for the original system, that is the following inequality

$$2x^T PAx + 2x^T PBu < 0 \tag{2.62}$$

is true for some $u(x)$, and therefore

$$2x^T P(Ax + R(x,u)) + 2x^T PBu < 0$$

is satisfied, for some $u(x)$, for all x in a neighborhood of the origin.

We do not know anything about the size of such neighborhood, which clearly depends on $R(x,u)$, but the following proposition holds for all candidate quadratic Lyapunov functions.

Proposition 2.33. *A quadratic positive definite function $\Psi(x) = x^T Px$ is a control Lyapunov function for the nonlinear system if (2.62) is satisfied for all x for some $u(x)$.*

The proof of the above proposition, which basically states that (2.62) can be used to find candidate quadratic (local) control Lyapunov function for the nonlinear system, is quite straightforward. The proposition essentially states that if a quadratic (local) control Lyapunov has to be found, one should start with the linear part.

Note that if P is determined, the control gain K is not the only choice, since it is possible to adopt a gradient-based control

$$u = -\gamma B^T P x$$

which is a linear feedback gain (see exercise 12). Note that any candidate smooth control Lyapunov function for the nonlinear system must be a control Lyapunov function for the linearized plant.

If the system is affected by uncertainties, the problem is even harder since the "linearization" does not provide a simple linear model, not even a certain equilibrium and one has to resort to results specifically tailored on the class of models under consideration.

An important class of uncertain models is given by systems with matching conditions such as

$$\dot{x}(t) = f(x(t)) + Bu(t) + Bw(t)Cx(t), \quad \|w(t)\| \leq \omega.$$

Assume that a control Lyapunov function is found for the system without uncertainties:

$$\nabla\Psi(x)[f(x) + Bu_N(x)] \leq -\phi(\|x\|)$$

for some κ function ϕ and some nominal feedback $u_N(x)$. Without restriction we incorporate Bu_u in f: $[f(x) + Bu_N(x)] = F(x)$ so

$$\nabla\Psi(x)[F(x) + Bu] \leq -\phi(\|x\|) \tag{2.63}$$

is satisfied in region containing 0, where u is a new control action which will be used to robustly stabilize the system. As expected, we take a gradient-based control $u = -\beta^2 B^T \nabla\Psi(x)^T$. Considering the perturbed system we get

$$\begin{aligned}
\dot{\Psi}(x) &= \nabla\Psi(x)F(x) + \nabla\Psi(x)BwCx - \beta^2\nabla\Psi(x)BB^T\nabla\Psi(x)^T \\
&\leq -\phi(\|x\|) + \nabla\Psi(x)BwCx - \beta^2\nabla\Psi(x)BB^T\nabla\Psi(x)^T \pm \frac{x^T C^T w^T wCx}{4\beta^2} \\
&= -\phi(\|x\|) - \left\|\beta B^T\nabla\Psi(x)^T - \frac{wCx}{2\beta}\right\|^2 + \frac{\|wCx\|^2}{4\beta^2} \\
&\leq -\phi(\|x\|) + \frac{\|Cx\|^2}{4\beta^2}\|w\|^2 \leq -\phi(\|x\|) + \frac{\|Cx\|^2}{4\beta^2}\omega^2
\end{aligned}$$

The last term is negative in a neighborhood of 0 and null in $x = 0$ for a large enough $\beta > 0$.

In the linear case, assuming A Hurwitz (e.g., pre-stabilized)

$$\dot{x}(t) = Ax(t) + Bu(t) + Bw(t)Cx(t)$$

it is possible to take a quadratic function such that $2x^T PAx \leq -x^T Qx$, $Q > 0$, so that

$$2x^T PAx + 2x^T PBwCx - 2\beta^2 x^T PBB^T Px \leq -x^T Qx + \frac{\|Cx\|^2}{4\beta^2}\omega^2$$

and then, by taking β such that, for all x,

$$\beta^2 \geq \frac{\|Cx\|^2}{4x^T Qx}\omega^2$$

it is immediate to see that the origin is globally stable with the linear state feedback control law

$$u(t) = -\beta^2 B^T Px$$

Example 2.34 (Cart-pendulum). Let us consider the following linearized model

$$\dot{x}(t) = Ax(t) + Bu(t) + Bw(t)Cx$$

of the cart-pendulum system in the inverted position (Fig. 2.7) where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \alpha^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -\epsilon & 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ -\rho\nu \\ 0 \\ \nu \end{bmatrix}, \quad C = \begin{bmatrix} 0\,0\,0\,1 \end{bmatrix}$$

and

$$\alpha^2 = 15.09, \;\; \nu = 5.78, \;\; \rho = 1.53, \;\; \epsilon \approx 0$$

The state variables are, in order, the pendulum angle, the angular speed, the cart position, and the cart speed. The parameter w represents the high uncertainty in the friction of the cart on the tracks and its nominal value is $w_0 \approx 4.9$ (but this value is extremely unreliable). The equilibrium point is clearly unstable. It is worth noticing that the uncertainty w affecting the system can be seen as an output feedback gain of the same linearized system with output C and input matrix B. This system is non-minimum phase. This means that high values of the friction tend to destabilize the system. Physically this fact can be interpreted by considering that high friction values tend to stick the cart preventing its motion hence the possibility of stabilizing the pendulum. Therefore we need an "authoritative control action." On the other hand, a strong control action, means high gain. If we consider the control

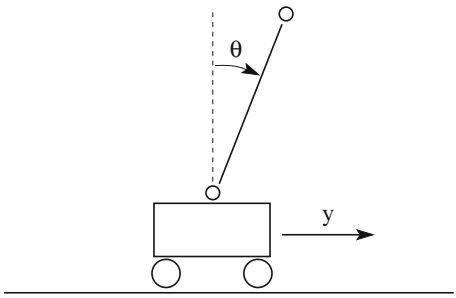


Fig. 2.7 The cart-pendulum system

$$u = -\beta^2 B^T P x$$

with $x^T P x$ control Lyapunov function, then β^2 large does not destabilize the system, in other words, the loop has infinite gain margin.

The Lyapunov matrix was derived by solving a standard LQ problem. However we do not claim any practical optimality. Other gradient-based controllers worked properly as well. Experimental results on the system[8] showed that indeed the gradient-based control worked quite well on the system, while other techniques such as pole assignment gave quite unsatisfactory results.

2.4.5 *Polytopic systems*

In this book much emphasis will be put on uncertain linear systems of the form

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

where $A(w)$, $B(w)$ are continuous functions of the parameter $w \in \mathcal{W}$. A special case of considerable interest is that of polytopic systems, in which matrices $A(w)$ and $B(w)$ belong to a polytope of matrices

$$\begin{aligned} A(w) &= \textstyle\sum_{i=1}^{s} A_i w_i, \\ B(w) &= \textstyle\sum_{i=1}^{s} B_i w_i \end{aligned} \qquad \text{with} \quad w_i \geq 0, \quad \textstyle\sum_{i=1}^{s} w_i = 1 \tag{2.64}$$

and

$$d(t) \in \mathcal{D}$$

[8] Experiments can be seen on Blanchini's web page.

where $\mathcal{D}$ is a polytope with vertex set $\{d_1,\ d_2, \dots d_r\}$.

$$\mathcal{D} = \{d = \sum_{i=1}^{r} \beta_i d_i,\ \ \beta_i \geq 0,\ \ \sum_{i=1}^{r} \beta_i = 1\}$$

To keep things more general, the following class of systems, including the previous one as a special case, will also be considered

$$\dot{x}(t) = \sum_{i=1}^{s} w_i(t) f_i(x(t), u(t)) \tag{2.65}$$

Note that distinct external signals w and d can be absorbed in a single vector. For example, the system

$$\dot{x} = wx + u + d$$

with $|w| \leq 1$ and $|d| \leq 1$ can be rewritten as (2.65) with

$$f_1 = x + u + 1,\ \ f_2 = x + u - 1,\ \ f_3 = -x + u + 1,\ \ f_4 = -x + u - 1$$

To check the negativity of the Lyapunov derivative for the above class of systems it is possible to exploit a "vertex result," which will be stated next in its general form. It is assumed that a candidate Lyapunov Ψ function is assigned and that its Lyapunov derivative can be expressed as in (2.29)

$$D^+\Psi(x, f(x, u, w)) = \sup_{z \in \partial\Psi} z^T f(x, u, w) \tag{2.66}$$

Note that this category comprises most of the commonly used Lyapunov functions such as the smooth and the convex ones.

Proposition 2.35. *Let Ψ be a candidate Lyapunov function which is positive definite, locally Lipschitz. Assume its subgradient (see Definition 2.13) is a convex non-empty set for all $x \in \mathbb{R}^n$ and its Lyapunov derivative can be expressed as in (2.66). Then Ψ is a control Lyapunov function (global, inside $\mathcal{X}$ or outside $\mathcal{X}$) for system (2.65) (possibly under constraints $u \in \mathcal{U}$) assuring the condition*

$$D^+\Psi\left(x, f(x, \Phi(x), w)\right) \doteq D^+\Psi\left(x, \Phi(x), w\right) \leq -\phi(\|x\|)$$

with a proper control function $\Phi(x)$ and a κ-function ϕ if and only if

$$D^+\Psi(x, f_i(x(t), \Phi(x(t)))) \leq -\phi_i(\|x\|),\ \ i = 1, 2, \dots, r.$$

for some κ-functions ϕ_i.

Proof. The condition is obviously necessary, because $w_i(t) \equiv 1$ and $w_j(t) \equiv 0$, $j \neq i$, is a possible realization of the uncertainty.

To prove sufficiency, consider the expression of the derivative (2.66)

$$\begin{aligned} & D^+\Psi\ (x, \Phi(x), w) = \\ & = \sup_{z \in \partial\Psi(x)} z^T \left[\sum_{i=1}^{s} w_i f_i(x, \Phi(x)) \right] = \sup_{z \in \partial\Psi(x)} \sum_{i=1}^{s} w_i \ \left[z^T f_i(x, \Phi(x)) \right] \leq \\ & \leq \sum_{i=1}^{s} w_i \sup_{z \in \partial\Psi(x)} \left[z^T f_i(x, \Phi(x)) \right] \leq - \min_i \ \phi_i(\|x\|) \end{aligned}$$

Since $\phi(\cdot) \doteq \min_i \ \phi_i(\cdot)$ is a κ-function, the proof is completed.

It is fundamental to stress that the control function Φ must be common for all i. The existence of control functions, each "working" for some i, is not sufficient. For instance, the system

$$\dot{x}(t) = x(t) + \delta(t)u(t), \quad |\delta| \leq 1$$

cannot be stabilized since $\delta(t) = 0$ is a possible realization. However, both the "extremal" systems $\dot{x} = x + u$ and $\dot{x} = x - u$ admit the control Lyapunov function $|x|$, as it is easy to check, although associated with different control actions (for instance, $u = -2x$ and $u = 2x$, respectively).

The previous property holds, under additional assumptions, when a full information control of the form $u = \Phi(x, w)$ is considered. Suitable systems are, for instance, the elements of the class

$$\dot{x}(t) = \sum_{i=1}^{s} w_i(t)\, f_i(x(t)) + Bu(t) \tag{2.67}$$

which are linear in the control with a constant input matrix B. It can be seen (using the previous type of machinery) that the existence of control laws $u = \Phi_i(x)$ that make $\Psi(x)$ a Lyapunov function for all the extreme systems is a sufficient and necessary condition for Ψ to be a control Lyapunov function. Such a function can be associated with the full information control

$$u = \Phi(x, w) = \sum_{i=1}^{s} w_i\, \Phi_i(x)$$

2.4.6 The convexity issue

We briefly report on the problem of "convexity seeking," an issue which is often regarded as a major deal in control synthesis. Convex problems have nice properties and they are easier to solve than non-convex problems. Let us see how this issue affects the Lyapunov theory.

Consider, for instance, the output feedback problem for the system

$$\begin{aligned}\dot{x}(t) &= f(x(t), w(t)) + Bu(t)\\ y(t) &= Cx(t)\end{aligned}$$

and a given candidate Lyapunov function $\Psi(x)$. Consider, for brevity, the standard inequality for the Lyapunov derivative for exponential stability

$$\nabla\Psi(x)^T\left[f(x(t), w(t)) + B\Phi(y)\right] \leq -\beta\Psi(x)$$

Then it can be seen that the set of all the functions $\Phi(y)$ that satisfy the inequality form a convex set (in the sense that if $\Phi_1(y)$ and $\Phi_2(y)$ are suitable controllers, then $\alpha\Phi_1(y) + (1-\alpha)\Phi_2(y)$ is suitable as well, for every $0 \leq \alpha \leq 1$). This convexity property is quite strong and desirable in the computation of a control, for a given $\Psi(x)$, as we will see later.

Let us consider the problem of determining a Lyapunov function $\Psi(x)$. At least in the space of smooth functions, for a given system (we assume that the control, if any, is fixed) the set of Lyapunov function is convex i.e., if $\Psi_1(x)$ and $\Psi_2(x)$ are suitable positive definite Lyapunov functions that satisfy

$$\nabla\Psi_i(x)^T f(x(t), w(t)) \leq -\beta\Psi_i(x), \quad i = 1, 2$$

then $\alpha\Psi_1(x) + (1-\alpha)\Psi_2(x)$ satisfies the same inequality for $0 \leq \alpha \leq 1$ (see, for instance, [Joh00]).

Unfortunately, when the control term is present and both Ψ and a proper controller Φ have to be determined, convexity is usually lost. This convexity issue will appear again several times in the book.

2.4.7 Fake Control Lyapunov functions

In this section we sketch a simple concept which is associated with some pitfall in the choice of a Lyapunov based control. Roughly we call a "fake control Lyapunov function" a function which is used as such but does not satisfy the proper requirements.

The first pathological case considered here is related to notions in other fields such as the greedy or myopic strategy in Dynamic optimization. Let us introduce

a very heuristic approach to control a system. Given a plant (we do not introduce uncertainties for brevity)

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

and a positive definite function $\Psi(x)$, adopt the following heuristically justified strategy: pick the controller that maximizes the decreasing rate of $\Psi(x(t))$, regardless of the fact that the basic condition (2.42) is satisfied. Assuming a constraint of the form $u(t) \in \mathcal{U}$ and assuming, for the sake of simplicity, Ψ to be a differentiable function, this reduces to

$$u = \Phi(x) = \arg\min_{u \in \mathcal{U}} \nabla\Psi(x)^T f(x, u)$$

If one assumes that an integral cost

$$\int_0^\infty \Psi(x(t))dt$$

has been assigned, this type of strategy is known as greedy or myopic strategy. Indeed, at each time it minimizes the derivative in order to achieve the "best" instantaneous results. It is well known that this strategy is far from achieving the optimum of the integral cost (with the exception of special cases, see, for instance, [MSP01]). Not only, it may also produce instability.

To prove this fact, it is sufficient to consider the simple case of a linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bu(t)$$

with a scalar input u and the function

$$\Psi(x) = x^T Px$$

If the linear system $(A, B, B^T P)$[9] admits zeros with positive real parts, the just introduced strategy may lead to instability. Now let us first consider a gradient-based controller which tries to render the derivative $\dot{\Psi}(x) = 2x^T P(Ax + Bu)$ negative and large. According to the previous considerations, the gradient-based control in this case is

$$u(t) = -\gamma B^T Px(t)$$

with $\gamma > 0$ chosen sufficiently large to decrease $\dot{\Psi}$. However, due to the non-minimum phase nature of the system, this control can lead to instability. Things

[9] That is, with state matrix A, input matrix B, and output matrix $C = B^T P$.

do not get any better if one considers a limitation for the input, such as $|u| \leq 1$. In this case the pointwise minimizer is

$$u = \arg\min_{|u|\leq 1} 2x^T P(Ax + Bu) = -\text{sgn}[x^T PB]$$

If system $(A, B, B^T P)$ has right-half-plane zeros, the system becomes locally unstable at the origin. The problem with the chosen "candidate control Lyapunov function" is that it cannot, by its nature, have negative derivative everywhere. An attempt of making the derivative as negative as possible, in all points in which this is possible, produces a destabilizing effect.

Example 2.36 (Destabilizing a stable system via wrong feedback). Consider the (open-loop stable!) linear system with

$$A = \begin{bmatrix} 1 & -2 \\ 2 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad P = I$$

It is immediate that the gradient based controller associated with the fake Lyapunov function $x^T Px = x^T x$ is

$$u = -\gamma B^T Px = -\gamma x_2$$

destabilizes the system for $\gamma \geq 1$. The discontinuous control

$$u = -\text{sgn}[x_2]$$

produces similar destabilizing effects (see Exercise 14) as well as its "approximation"

$$u = -\text{sat}[\gamma x_2].$$

where $\gamma > 0$ is a "large number."

Another classical case of bad "candidate control Lyapunov function" choice is presented next. In this case, the function does not satisfy the positive definiteness property and its use to derive a feedback control (in general) results in an unstable closed-loop behavior.

Example 2.37 (A non-positive definite function). Consider a linear continuous-time single-input single-output system

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

with $CB \neq 0$, for which a state feedback control law is sought as to satisfy the output specification

$$|y(t)| \leq 1,$$

for all initial conditions such that $|y(0)| = |Cx(0)| \leq 1$ (and possibly, stabilizing). In the attempt to find such control law, the candidate Lyapunov function $\Psi(x) = |Cx|$, whose gradient is $\nabla\Psi^T = C\text{sgn}(Cx)$, is considered. If a gradient-based controller of the form (2.59) is chosen,

$$u(x) = -\gamma(x)\frac{B^T C^T}{(CB)^2}\text{sgn}(Cx)$$

it is possible to make the Lyapunov derivative negative if

$$\text{sgn}(Cx)CAx - \gamma(x) < 0$$

One possible choice for $\gamma(x)$ is

$$\gamma(x) = \text{sgn}(Cx)CAx + \rho|Cx|$$

for some $\rho > 0$, leading to the linear feedback law

$$u(x) = -\frac{B^T C^T}{(CB)^2} C\left(A + \rho I\right)x.$$

Unfortunately, the only result which we are assured to get is that the function $\Psi(x) = |Cx|$ will decrease along the system trajectories, precisely $D^+\Psi(x) \leq -\rho\Psi(x)$, but there is no guarantee that the closed-loop will be stable (see Exercise 11).

2.5 Lyapunov-like functions

In the previous subsections, candidate positive definite Lyapunov functions were considered, motivated by the fact that the origin is typically the target point for the system state in the stability analysis or stabilization problem. However, many analysis or regulation problems are not concerned with stability but rather the main goal is to establish the qualitative behavior of a given system. Lyapunov-like functions may provide a useful tool for this purpose. The definition of a Lyapunov-like function can be simply stated as follows.

Definition 2.38 (Lyapunov-like function). The locally Lipschitz function $\Psi(x)$ is a Lyapunov-like function for system (2.21) inside the domain $\mathcal{D}$ if for $x \in \mathcal{D}$ we have that

$$D^+\Psi(x,w) \leq 0, \tag{2.68}$$

for all $w(t) \in \mathcal{W}$.

The interest in such functions lies in the fact that the level surfaces of the function Ψ provide a qualitative information of the behavior of the system inside the set $\mathcal{D}$. More precisely, if $x(t) \in \mathcal{N}[\Psi,\nu] \bigcap \mathcal{D}$, the state cannot leave the set $\mathcal{N}[\Psi,\nu]$ without previously abandoning $\mathcal{D}$. If $\mathcal{D}$ is an invariant set, then the state cannot leave the set $\mathcal{N}[\Psi,\nu]$ at all.

Example 2.39 (Magnetic levitator revisited). Consider the following system of differential equations

$$\begin{aligned}\dot{x}_1(t) &= x_2(t)\\ \dot{x}_2(t) &= g - \frac{d^2 g}{(x_1(t)+d)^2}\end{aligned}$$

representing, with a suitable parameterization, the unstable vertical motion of an uncontrolled magnetic levitator (see Fig. 2.1 and Example 2.2). The constant d is the equilibrium distance of the steel sphere from the magnet and g is the gravity. The variable x_1 is the vertical distance of the object from the equilibrium, while x_2 represents the velocity. The unstable vertical motion in the positive sector will now be studied by means of a pair of Lyapunov-like functions.

Consider the function $\Psi_1(x) = -x_1 x_2$ and let $\mathcal{D}$ be the positive quadrant $x_1, x_2 \geq 0$. The Lyapunov derivative of $\Psi_1(x)$ is

$$D^+\Psi_1(x) = -\left(x_2^2 + \frac{x_1^3 + 2dx_1^2}{(x_1+d)^2} g\right) < 0, \quad \textit{for} \quad x \in \mathcal{D}.$$

The level surfaces of $\Psi_1(x)$ are hyperbola whose asymptotes are the principal axes (see Fig. 2.8). This implies that as long as $x(t_1)$ is the positive quadrant, and $\Psi_1(x(t_1)) = k < 0$ the set $\mathcal{N}[\Psi,k]$ cannot be escaped. But this implies that the interior of the positive quadrant is positively invariant because for each interior point $x(t_1)$ there exist $k < 0$ such that $\Psi_1(x(t_1)) = k$. Moreover, being $\Psi_1(x) = 0$ for each x on the axes, these cannot be reached for $t \geq t_1$.

Consider also the Lyapunov-like function

$$\Psi_2(x) = -gx_1 + \frac{1}{2}x_2^2$$

whose level sets $\mathcal{N}[\Psi_2,\mu]$, for $\mu < 0$, are delimited in the positive quadrant by arcs of parabolas each originating on the x_1 axis at the point $x_1 = -\mu/g$ (see again Fig. 2.8). It is therefore clear that the trajectories of the system originating in the positive quadrant have the behavior described in Fig. 2.8 (dashed-line), because they must intersect both parabolas and hyperbolas from left to right.

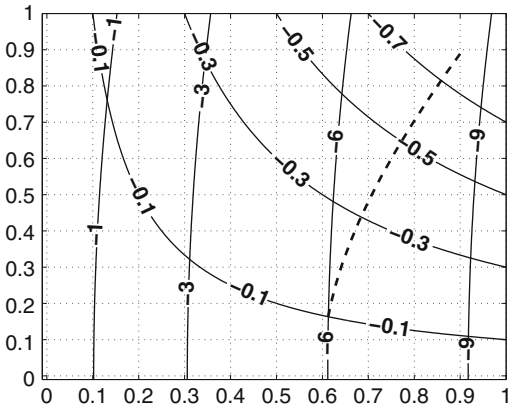


Fig. 2.8 The level surfaces of Ψ_1 and of Ψ_2

Lyapunov like functions are very useful to prove unstable behavior of a system and the example above is one of such cases. There are more lucky cases in which the level surfaces of a Lyapunov-like function describe exactly the system trajectory.

Example 2.40 (Oscillator). The system of equation

$$\begin{aligned}\dot{x}_1(t) &= x_2(t)\\ \dot{x}_2(t) &= x_1(t) - x_1(t)^3\end{aligned}$$

admits the Lyapunov-like function

$$\Psi(x) = \frac{1}{2}x_2^2 + \left(\frac{x_1^4}{4} - \frac{x_1^2}{2}\right)$$

which is such that

$$D^+\Psi(x) = 0, \quad \text{for all } x$$

Therefore the system trajectories lie on the level surfaces of this function. These level surfaces show that the system is an oscillator admitting periodic trajectories which encircle either or both the equilibrium points $(1,0)$ and $(-1,0)$ (see Figure 2.9)

Example 2.41 (Decentralized job balancing). Consider the problem of m agents who have to assign a fixed amount of work among n operators. Each agent controls her/his group of n_i operators represented by the n nodes. The groups are represented by the nodes included in the polygons in Fig. 2.10 and they might have non-empty intersections (so $\sum n_i \geq n$). In the case of the figure, agent 1 controls nodes 1, 2, and 5, agent 2 controls 6, 7, and 8, agent 3 controls 2, 3, 6 and 7, and agent 4 controls 3, 4, 7, and 8.

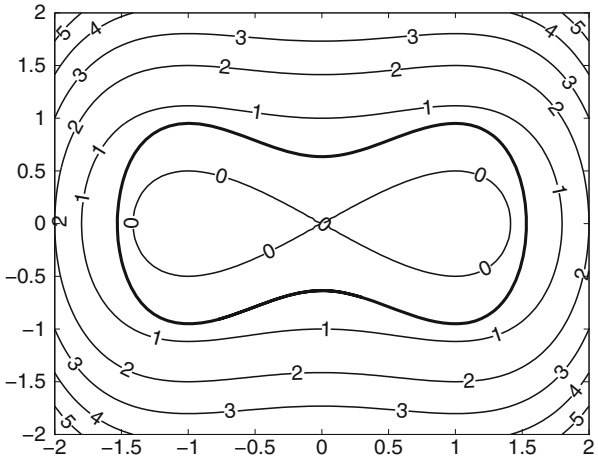
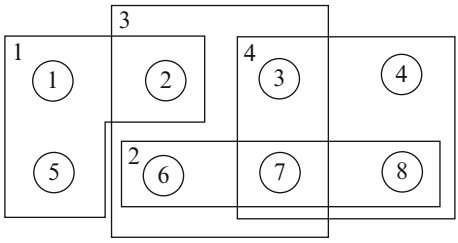


Fig. 2.9 The level surfaces of $\Psi(x)$ and the trajectory for $x(0) = [0.3 \quad -0.7]^T$ (bold) for system in Example 2.40

Fig. 2.10 The agent-operator problem



Denote by $B_k = (n \times n_k)$ the incidence matrix of the partial graph having all the n nodes and all the arcs connecting the n_k nodes in the kth group. For instance, group 1 includes nodes 1, 2, and 5, so B_1 has three columns corresponding to the arcs $1 \to 2$, $2 \to 5$, and $5 \to 1$

$$B_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}^T$$

The agent k can impose a distribution flow confined in the subspace spanned by matrix B_k. Such a flow is limited to the subset of its nodes. The incidence matrix $B = [B_1 \; B_2 \; B_3 \; B_4]$ is the input matrix of the model of our re-distribution problem whose equations are

$$\dot{x}(t) = \sum_{k=1}^{m} B_k u_k(t) = Bu \tag{2.69}$$

where u_k is the action of the kth agent, and $x_i(t)$ represent the workload of the ith operator. The matrices B_k have the property that

$$\bar{1}^T B_k = 0, \quad \text{and} \quad \text{rank}(B_k) = n_k - 1$$

so that each single agent can distribute load among its operators without changing the total workload of the group. The workload can be arbitrarily distributed in the group. Again, u_1 can distribute the load of x_1, x_2, and x_5 but cannot change $x_1 + x_2 + x_5$. Note that the overall matrix B has not full row rank, since $\bar{1}^T B = 0$.

We assume that the agents do not talk to each other and have information only about the loads of their operators. Accordingly, we consider the network-decentralized strategy [ID90, Ift99, BMU00] $u_k = -\gamma B_k^T x$ with $\gamma > 0$, namely

$$u = -\gamma B^T x$$

This means that the agents can feed back only the states of their operators. For instance, agent 1 control is $u = -\gamma B_1^T x$ and she/he needs information only about the loads of operators 1, 2, and 5 because only the corresponding columns in B_1^T have non-zero elements. We wish to show that, asymptotically, the system converges to an "optimal load distribution."

Let us consider a general problem of the form (2.69). For any initial condition x_0, the system evolves on the affine manifold

$$\mathcal{A} = \{x = x_0 + Bv, \ \ \forall v\}$$

For the system of Fig. 2.10, this just means that $\bar{1}^T x(t) = \bar{1}^T x_0$.

The control $u = -\gamma B^T x(t)$ yields

$$\dot{x}(t) = -\gamma B B^T x(t)$$

We show that the state converges to the minimum norm value inside $\mathcal{A}$, i.e. it asymptotically minimizes $\|x\|^2$. The minimum is clearly attained for a value $\bar{v}$ such that $\bar{x} = x_0 + B\bar{v}$ is orthogonal to the image space of B, namely

$$B^T \bar{x} = 0$$

Let us consider the Lyapunov-like function

$$\Psi(x) = x^T B B^T x$$

Its Lyapunov derivative is

$$\dot{\Psi}(x) = 2x^T B B^T \dot{x} = -2\gamma x^T B B^T B B^T x = -2\gamma \|B B^T x\|^2$$

which is negative, unless $BB^Tx = 0$, say $x^TBB^Tx = 0$, which is equivalent to $B^Tx = 0$. This means that

$$\Psi(x(t)) \to 0$$

hence $x(t) \to \bar{x}$.

In the case of the load distribution problem, the overall network is connected, hence the kernel of B has dimension 1 and it has the form $\mathcal{A} = \{x = \lambda\bar{1}\}$, so, asymptotically, we will have $x_1 = x_2 = \cdots = x_n$, namely a consensus (see [RBA07] for a survey) is reached.

The following example is similar, but with the difference that the state is asymptotically unbounded.

Example 2.42 (Clock synchronization). Consider the system of n clocks each having its own speed ω_i. Their time indications are grouped in the vector $x(t)$. These clocks may communicate pairwise according. The overall communication is described by a graph having an arc connecting two nodes if and only if they communicate. Let us denote its incidence matrix as B. Each clock compares its time with that of its neighbors and adjusts its speed proportionally:

$$\dot{x}_i(t) = \omega_i + u_i = \omega_i + \underbrace{\sum_{\mathcal{N}_i} -\gamma(x_i(t) - x_j(t))}_{u_i}$$

where $\mathcal{N}_i$ is the set of nodes communicating with node i.

The derivative $\dot{x}_i(t)$ represents the change of the ith clock speed with respect to the "absolute" time. This basically means that each clock can modify its speed by means of the "control action" u_i. Clearly, in real implementations, each clock has no access to the "true" time (otherwise the problem would be trivial), and in order to synchronize its time with the rest of the group, the best it can do is to correct its time based on its own "time perception." If we assume that the relative speeds are not too different, the model can be valid.

If we accept this simplified model, we arrive to the equation

$$\dot{x}(t) = \omega - \gamma BB^Tx(t) = \omega - \gamma Lx(t)$$

where $\omega = [\omega_1\ \omega_2 \ldots \omega_n]^T$ and $L = BB^T$ is the so-called Laplacian matrix [RBA07]. The Laplacian is a symmetric positive semi-definite matrix and, if the graph is connected, it has rank $n - 1$, as B. Its kernel is generated by the vector $\bar{1}$

$$L\bar{1} = BB^T\bar{1} = 0$$

Matrix L^2 is symmetric positive semi-definite and it has the same kernel. Moreover, there exists α such that

$$x^T L^2 x = \|Lx\|^2 \leq \alpha^2 x^T L x$$

Note that in the case of no adjustment, the clock times would spread apart since

$$x(t) = x(0) + \omega t,$$

unless ω has strictly equal components. But even in the last case the initial mismatch among the components of $x(0)$ would be maintained.

Let us see what happens with $\gamma > 0$. The ideal situation is the case in which $x_1(t) = x_2(t) = \cdots = x_n(t)$, namely $x(t) = \xi(t)\bar{1}$: the clocks are synchronized. As in the previous example, we use the Lyapunov-like function

$$\Psi(x) = x^T L x = x^T B B^T x = \|B^T x\|^2$$

as a measure of the mismatch. Note indeed that, for a connected graph, $\Psi(x) = 0$ is equivalent to $x = \xi\bar{1}$, as desired.

Now we separate the "common" and "mismatched" part of vector ω

$$\omega = \frac{\bar{1}^T \omega}{n} \bar{1} + \Delta = \bar{\omega}\bar{1} + \Delta$$

where $\bar{\omega} = \bar{1}^T \omega / n$ is the average and Δ is a (hopefully small) mismatch vector. Assume that a bound is known for Δ: $\|\Delta\| \leq \delta$. Observe that $BB^T \omega = L\omega = BB^T \Delta$. Then we have

$$\dot{\Psi}(x) = -2\gamma x^T L^2 x + 2x^T L\omega \leq -2\gamma \|Lx\|^2 + 2|x^T L\Delta| \tag{2.70}$$

$$\leq -2\gamma \|Lx\|^2 + 2\|Lx\|\|\Delta\| \leq -2[\gamma\|Lx\| - \delta]\ \|Lx\| \tag{2.71}$$

Then the derivative is negative if $\|Lx\|^2 < (\delta/\gamma)^2$. But we have seen that $\|Lx\|^2 \leq \alpha^2 x^T L x = \alpha^2 \Psi(x)$, so the derivative is negative if

$$\Psi(x) \leq \frac{\delta^2}{\gamma^2 \alpha^2}$$

This means that

$$\limsup_{t\to\infty} \Psi(x(t)) \leq \frac{\delta^2}{\gamma^2 \alpha^2}$$

On the other hand, this is equivalent to saying that asymptotically

$$\|B^T x\| \leq \frac{\delta}{\gamma\alpha}$$

which means the mismatch becomes arbitrarily small according to our measure $\|B^T x\|$ if we take $\gamma > 0$ large.

To complete our investigation, we consider another Lyapunov-like function, the average

$$a(x) = \frac{\bar{1}^T x}{n}$$

Then, since $\bar{1}^T L = 0$,

$$\dot{a}(x) = \frac{\bar{1}^T \dot{x}}{n} = \frac{\bar{1}^T \omega}{n} = \bar{1}\bar{\omega}$$

Hence

$$a(x(t)) = a(x(0)) + \bar{\omega} t$$

so the average time evolves with the average of the initial speeds.

The synchronization problem will be reconsidered later in the book.

2.6 Discrete-time systems

The main concepts of this chapter have been presented in the context of continuous-time systems. The same concepts hold in the case of discrete-time systems, although there are several technical differences. In particular some difficulties, typical of the differential equations, such as existence, uniqueness, and finite escape time of the solution, do not trouble anymore.

Let us now consider the case of a system of the form

$$x(t+1) = f(x(t), w(t)), \tag{2.72}$$

where now functions $x(t)$ and $w(t) \in \mathcal{W}$ are indeed sequences, although they will often be often referred to as "functions." Consider a function $\Psi : \mathbb{R}^n \to \mathbb{R}$ defined and continuous on the state space. It is known that, as a counterpart of the Lyapunov derivative, the Lyapunov difference has to be considered. Again, for any solution $x(t)$, it is possible to consider the composed function

$$\psi(t) \doteq \Psi(x(t))$$

and its increment

$$\Delta\psi(t) \doteq \psi(t+1) - \psi(t) = \Psi(x(t+1)) - \Psi(x(t))$$

so that the Lyapunov difference is simply defined as

$$\Delta\psi(t) = \Psi(f(x(t), w(t))) - \Psi(x(t)) \doteq \Delta\Psi(x(t), w(t)) \tag{2.73}$$

Therefore, the monotonicity of function Ψ along the system trajectories can be inferred by considering the function $\Delta\Psi(x, w)$, thought of as a function of x and w.

Consider a model of the form (2.72) and again assume that the following condition is satisfied

$$f(0, w) = 0, \quad \text{for all } w \in \mathcal{W} \tag{2.74}$$

namely, $x(t) \equiv 0$ is a trajectory of the system.

For this discrete-time model, the same definition of global uniform asymptotic stability (Definition 2.16) holds unchanged. Definition 2.18 of global Lyapunov function remains unchanged up to the fact that the Lyapunov derivative is replaced by the Lyapunov difference.

Definition 2.43 (Global Lyapunov Function, discrete-time). A continuous function $\Psi : \mathbb{R}^n \to \mathbb{R}$ is said to be a Global Lyapunov Function (GLF) for system (2.72) if it is positive definite, radially unbounded and there exists a κ-function ϕ such that

$$\Delta\Psi(x, w) \leq -\phi(\|x(t)\|) \quad \text{for all } w \in \mathcal{W} \tag{2.75}$$

The next theorem, reported without proof, is the discrete-time counterpart of Theorem 2.19.

Theorem 2.44. *Assume the system (2.72) admits a global Lyapunov function Ψ. Then it is globally uniformly asymptotically stable.*

Even in the discrete-time case, it is possible to introduce a stronger notion of stability, namely exponential stability.

Definition 2.45 (Global Exponential Stability, discrete-time). System (2.72) is said to be Globally Exponentially Stable if there exist a positive $\lambda < 1$ and a positive μ such that, for all $\|x(0)\|$,

$$\|x(t)\| \leq \mu\|x(0)\|\lambda^t \tag{2.76}$$

for all $t \geq 0$ and all sequences $w(t) \in \mathcal{W}$.

The coefficient λ is the discrete-time convergence speed and μ is the discrete-time transient estimate. As in the continuous-time case, exponential stability can be assured by the existence of a Lyapunov function whose decreasing rate, or better, whose Lyapunov difference along the system trajectories is bounded by a term proportional to the function value. Let us assume that the positive definite function $\Psi(x)$ is upper and lower polynomially bounded, as in (2.38). The following theorem, reported without proof, holds.

Theorem 2.46. *Assume that system (2.72) admits a positive definite continuous function Ψ, which has polynomial growth as in (2.38) and*

$$\Delta\Psi(x,w) \leq -\beta\Psi(x) \tag{2.77}$$

for some positive $\beta < 1$. Then it is globally exponentially stable with speed of convergence $\lambda = (1-\beta)(< 1)$.

Note that the condition of the theorem may be equivalently stated as

$$\Psi(f(x,w)) \leq \lambda\Psi(x)$$

with $0 \leq \lambda < 1$ for all x and $w \in \mathcal{W}$.

The concept of Uniform Local Stability and Uniform Ultimate Boundedness for discrete-time system are expressed by Definitions 2.22 and 2.23 which hold without modifications. Given a neighborhood of the origin $\mathcal{S}$ the concepts of Lyapunov functions *inside* and *outside* $\mathcal{S}$ sound now as follows.

Definition 2.47 (Lyapunov function inside a set, discrete-time). The continuous positive definite function Ψ is said to be a Lyapunov function inside $\mathcal{S}$ for system (2.72) if there exists $\nu > 0$ such that $\mathcal{S} \subseteq \mathcal{N}[\Psi,\nu]$ and for all $x \in \mathcal{N}[\Psi,\nu]$ the inequality

$$\Delta\Psi(x,w) \leq -\phi(\|x\|)$$

holds for some κ-function ϕ and for all $w \in \mathcal{W}$.

Definition 2.48 (Lyapunov function outside a set, discrete-time). The continuous positive definite function Ψ is a Lyapunov function outside $\mathcal{S}$ for system (2.72) if there exists $\nu > 0$ such that $\mathcal{N}[\Psi,\nu] \subseteq \mathcal{S}$ and for all $x \notin \mathcal{N}[\Psi,\nu]$ the inequality

$$\Delta\Psi(x,w) \leq -\phi(\|x\|)$$

holds for some κ-function ϕ and all $w \in \mathcal{W}$; moreover

$$\Psi(f(x,w)) \leq \nu$$

for all $x \in \mathcal{N}[\Psi,\nu]$ and all $w \in \mathcal{W}$.

Note that the last condition in the previous definition has no analogous statement in the continuous-time Definition 2.25 and its meaning is that once the set $\mathcal{N}[\Psi,\nu]$ is reached by the state, it cannot be escaped. This condition is automatically satisfied in the continuous-time case by a function as in Definition 2.25. The next two theorems hold.

Theorem 2.49. *Assume system (2.72) admits a positive definite continuous function Ψ inside $\mathcal{S}$. Then it is Locally Stable with basin (domain) of attraction $\mathcal{S}$.*

Theorem 2.50. *Assume system (2.72) admits a positive definite continuous function Ψ outside $\mathcal{S}$. Then it is uniformly ultimately bounded in $\mathcal{S}$.*

Example 2.51 (Newton–Raphson method for nonlinear systems of equations). Consider the problem of solving the nonlinear system $f(x) = 0$. Assume that $\bar{x}$ is an isolated root $f(\bar{x}) = 0$ and that f is twice continuously differentiable and its Jacobian $J(x)$ is continuously differentiable and $\|J(x)^{-1}\| \leq \nu$ in a neighborhood of $\bar{x}$. We further assume that

$$\|(J(x) - J(\bar{x}))(x - \bar{x})\| \leq \mu_1 \|x - \bar{x}\|^2$$

is infinitesimal of the second order. Note also that

$$f(x) = J(\bar{x})(x - \bar{x}) + \Theta(x - \bar{x})$$

with Θ infinitesimal

$$\|\Theta(x - \bar{x})\| \leq \mu_2 \|x - \bar{x}\|^2$$

Consider the following iterative method (Newton–Raphson method)

$$x_{k+1} = x_k - J(x_k)^{-1} f(x_k) + w$$

where $\|w\| \leq \omega$ is the roundoff error. We initially consider $w = 0$ and we show that x_k converges to the root, $x_k \to \bar{x}$, if x_0 is close enough to $\bar{x}$. Consider $\|x - \bar{x}\|$ as candidate Lyapunov function.

$$\begin{aligned}
&\|x_{k+1} - \bar{x}\| = \left\|x_k - J(x_k)^{-1} f(x_k) - \bar{x}\right\| = \\
&= \left\|x_k - J(x_k)^{-1} \left[J(\bar{x})(x_k - \bar{x}) + \Theta(x - \bar{x})\right] - \bar{x}\right\| = \\
&= \left\|x_k - J(x_k)^{-1} \left[J(x_k)(x_k - \bar{x}) + (J(\bar{x}) - J(x_k))(x_k - \bar{x}) + \Theta(x - \bar{x})\right] - \bar{x}\right\| = \\
&= \left\|J(x_k)^{-1} \left[(J(\bar{x}) - J(x_k))(x_k - \bar{x}) + \Theta(x - \bar{x})\right]\right\| \leq \nu(\mu_1 + \mu_2)\|x_k - \bar{x}\|^2
\end{aligned}$$

Take any positive $\lambda_0 < 1$ ($\lambda_0 \simeq 1$). It can be verified that for $\|x_0 - \bar{x}\| < \frac{\lambda_0}{\nu(\mu_1+\mu_2)}$ also $\|x_1 - \bar{x}\|$ satisfies the same condition $\|x_1 - \bar{x}\| < \frac{\lambda_0}{\nu(\mu_1+\mu_2)}$. Hence, recursively,

$$\|x_k - \bar{x}\| < \frac{\lambda_0}{\nu(\mu_1 + \mu_2)} \doteq \rho_{conv}$$

for every $k > 0$, say the sequence x_k is bounded if we start in the ball centered in $\bar{x}$ of radius ρ_{conv}. We have also that

$$\|x_{k+1} - \bar{x}\| \leq \lambda_0 \|x_k - \bar{x}\|$$

which means that we have exponential convergence with speed $\lambda_0 < 1$. Actually, convergence is faster. Indeed convergence along with the condition[10]

$$\|x_{k+1} - \bar{x}\| \leq \nu(\mu_1 + \mu_2)\|x_k - \bar{x}\|^p$$

with $p = 2$ implies that $\|x_k - \bar{x}\|$ converges faster than λ^k, with $0 < \lambda < 1$.

Finally, if we consider the roundoff error we derive

$$\|x_{k+1} - \bar{x}\| \leq \nu(\mu_1 + \mu_2)\|x_k - \bar{x}\|^2 + \omega$$

For ω not too large it can be seen that x_k converges to a neighborhood

$$\|x_k - \bar{x}\| \leq \rho$$

where ρ is the smallest positive root (if any) of the equation

$$\rho = \nu(\mu_1 + \mu_2)\rho^2 + \omega$$

Note that, for ω large, no root can exist as expected.

Consider now the case of a controlled discrete-time system which has to be equipped with a controller of the form (2.9)–(2.10). As previously observed, any dynamic finite-dimensional feedback controller can be viewed as a static output feedback for a properly augmented system. Therefore, it is possible to consider a system of the form

$$\begin{cases} x(t+1) = f(x(t), u(t), w(t)) \\ y(t) = h(x(t), w(t)) \end{cases} \tag{2.78}$$

with a static feedback control law. As in the continuous-time case, the class $\mathcal{C}$ of controllers needs to be specified. In the following, controllers in one of these classes will be considered: output feedback, state feedback, output feedback with feedforward, state feedback with feedforward.

The definition of control Lyapunov function is identical to that reported in the continuous-time case in Definition 2.29. In practice, a Control Lyapunov function (Global, Inside, Outside) is a Lyapunov function once a proper control is applied. If control constraints of the form (2.13) need to be taken into account, then it is sufficient to include them in the control Lyapunov function definition. If constraints on the state, $x(t) \in \mathcal{X}$, are present, then the key condition is that $x(0) \in \mathcal{N}[\Psi, \mu] \subseteq \mathcal{X}$, for some μ. Again, dealing with difference equations, there is no need to specify that the closed-loop system has to admit a solution which, in the discrete-time case,

[10] In numerical analysis this condition is known as convergence of order p, so the Newton–Raphson method is quadratic.

exists as long as the control function is well defined. Other technical differences appear when a control law has to be associated with a Control Lyapunov Function.

Let us consider a domain of the form

$$\mathcal{N}[\Psi, \alpha, \beta] = \{x : \ \alpha \leq \Psi(x) \leq \beta\} \tag{2.79}$$

where, by possibly assuming $\alpha = 0$ or $\beta = +\infty$, all the "meaningful" cases of Lyapunov function inside a set (outside a set or global) are included. Assume that a positive definite continuous function Ψ is given and consider the next inequality

$$\Delta\Psi(x, u, w) \doteq \Psi(f(x, u, w)) - \Psi(x) \leq -\phi(\|x\|). \tag{2.80}$$

The problem can be thought of in either of the following ways. If, for all x, there must exist u such that (2.80) is satisfied for all $w \in \mathcal{W}$, then this condition implies that Ψ is a control Lyapunov function with state feedback. Conversely, if u is allowed to be a function also of w, the condition becomes: for all x and $w \in \mathcal{W}$ there exists u such that (2.80) is satisfied. In this second case one is dealing with a control Lyapunov function for the full information feedback.

To characterize the control function for the state feedback, consider the discrete-time control map or regulation map

$$\Omega(x) = \{u : \ \text{(2.80) is satisfied for all } w \in \mathcal{W}\},$$

conceptually identical to that already introduced in the continuous-time case Any proper state feedback control function $u = \Phi(x)$ has to be such that

$$\Phi(x) \in \Omega(x). \tag{2.81}$$

In the full-information control case, the following set

$$\Omega(x, w) = \{u : \ \text{(2.80) is satisfied}\}$$

comes into play. The control function in this case must be such that

$$\Phi(x, w) \in \Omega(x, w) \tag{2.82}$$

The question of the regularity of function $\Phi(x)$ is not essential from the mathematical point of view. Nevertheless, it may be important, since discontinuous controllers may have practical troubles, such as actuator over-exploitation.

Unfortunately, though continuity is not an issue anymore, in the discrete-time case the problem of determining a feedback control in an analytic form does not admit in general a solution, as it does instead in the continuous-time case. The main reason is that even in the case of a smooth control Lyapunov function the gradient does not play any role. Once the gradient is known, in the continuous-time case,

basically the control is chosen in order to push the system in the opposite direction as much as possible. In the discrete-time case this property does not hold. Let us consider a very simple example.

Example 2.52. Let us seek for a state feedback for the scalar system

$$\dot{x}(t)[x(t+1)] = f(x(t), w(t)) + u, \quad |w| \leq 1. \tag{2.83}$$

Assume that f is bounded as $|f(x(t), w(t))| \leq \xi|x|$ and consider the control Lyapunov function $\Psi(x) = x^2/2$. The continuous-time problem is straightforward. Since the "gradient" is x, it is sufficient to take a control pushing towards the origin, for instance $u = -\kappa x$, with κ large enough (at least $\kappa > \xi$). The Lyapunov derivative results then in

$$\dot{\Psi}(x) = x\,(f(x, w) + u(t)) \leq -(\kappa - \xi)x^2,$$

so that the closed-loop system is globally asymptotically stable.

The discrete-time version of the problem is rather different. For the Lyapunov function $\Psi(x) = x^2/2$ to decrease, the basic condition to be respected is:

$$\frac{(f(x, w) + u(t))^2}{2} - \frac{x^2}{2} \leq -\phi(|x|), \quad \text{for all } |w| \leq 1$$

The only information which can be derived by the above condition is that

$$u(x) \in \Omega(x) = \{x : \ (f(x, w) + u(t))^2 \leq x^2 - 2\phi(|x|), \ \text{ for all } |w| \leq 1\},$$

a condition that heavily involves function f. Furthermore, it is not difficult to show that the bound $|f(x(t), w(t))| \leq \xi|x|$ does not assure that $\Omega(x)$ is non-empty for all x (say, the system is stabilizable). For instance, the system

$$x(t+1) = [a + bw(t)]x + u, \quad |w| \leq 1, \tag{2.84}$$

is not stabilizable by any feedback for all values of the constant a and b. The necessary and sufficient stabilizability condition via state feedback is $|b| < 1$.

The previous example shows that there are no analogous controllers to those proposed for uncertain systems by [BCL83]. Another point worth being evidenced is the difference, in the discrete-time case, between state feedback and full information control. Indeed, for the above example, when $|b| > 1$ there is no stabilizing state feedback $u = \Phi(x)$, but the system is always stabilizable by means of the full information feedback $u = -f(x, w) = -[a+bw]x$. This implies that the equivalence between state and full information feedback, which holds under some technical assumptions [Mei79, Bla00] in the continuous-time case, is not true for discrete-time systems.

To conclude the section, the analogous of Proposition 2.35 is stated. Consider a discrete-time polytopic system of the form

$$x(t+1) = f(x(t), u(t), w(t)) = \sum_{i=1}^{s} w_i(t)\, f_i(x(t), u(t))$$

with $\sum_{i=1}^{s} w_i = 1$, $w_i \geq 0$. The following proposition holds.

Proposition 2.53. *The* convex *positive definite function Ψ is a control Lyapunov function (global outside or inside a set $\mathcal{X}$) if and only if there exists a single control function Φ such that*

$$\Delta\Psi_i(x, \Phi(x)) \doteq \Psi(f_i(x, \Phi(x))) - \Psi(x) \leq -\phi_i(\|x\|)$$

for proper κ-functions ϕ_i.

Proof. The proof is immediate, since

$$\begin{aligned}\Delta\Psi(f(x, \Phi(x), w)) = \Psi\left(\sum_{i=1}^{s} w_i\, f_i(x, \Phi(x))\right) - \Psi(x) \leq \\ \leq \sum_{i=1}^{s} w_i\ [\Psi\left(f_i(x, \Phi(x))\right) - \Psi(x)] \leq \\ \leq -\sum_{i=1}^{s} w_i\ \phi_i(\|x\|) \leq -\min_i \phi_i(\|x\|)\end{aligned}$$

and $\min_i \phi_i(\|x\|)$ is a κ-function.

It has to be stressed that convexity of $\Psi(x)$ is fundamental here, while it was not necessary in the analogous Proposition 2.35.

2.6.1 Converse Lyapunov theorems

Lyapunov theory not only provides fundamental tools for solving engineering problems, but it is conceptually fundamental as well. Nevertheless, the main weak point of this theory, precisely the not-always-so-clear way to find a suitable candidate Lyapunov function, is unfortunately[11] an unsolved problem.

[11] Actually for mathematicians and theoretical engineers for which this circumstance is a source of fun.

In the previous sections it has been shown that finding a suitable (control) Lyapunov function for a given dynamic system is crucial for the stability (stabilizability) of the considered system. Unfortunately, if such function cannot be found, nothing can be said. Indeed, in basic system theory, it is a rather established fact that a candidate positive definite $\Psi(x)$, even smooth, whose derivative $\dot{\Psi}(x)$ is not sign definite, leads to the only conclusion that $\Psi(x)$ is the wrong candidate.

Therefore, a fundamental question arises: Is the Lyapunov approach the right one for stability? This question has a theoretical affirmative fundamental reply. Indeed Lyapunov-type theorems admit several converse theorems which basically state that if a system is asymptotically stable (under appropriate assumptions), then it admits a Lyapunov function. Some famous results in this sense are due to Persisdki and Kurzweil and to Massera. The reader is again referred to the book [RHL77]. We report here a "robust converse Lyapunov theorem," proved in [LSW96] and which includes several previous results as special cases.

Consider the system

$$\dot{x} = f(x(t), w(t))$$

where $w(t) \in \mathcal{W}$ and $\mathcal{W}$ is a compact set, f is continuous with respect to w and locally Lipschitz in x uniformly with respect to w[12].

Theorem 2.54. *For the above system the following two conditions are equivalent.*

Global uniform asymptotic stability

- *There exists a κ-function $\phi(\cdot)$ such that, for every $\epsilon > 0$ and any $\|x(0)\| \leq \phi(\epsilon)$, the solution $\|x(t)\| \leq \epsilon$ for all $t \geq 0$ and all $w \in \mathcal{W}$.*
- *For any $\rho > 0$ and $\epsilon > 0$ there exists T such that, for all $\|x(0)\| \leq \rho$, $\|x(t)\| \leq \epsilon$, $t \geq T$, for all $w(t) \in \mathcal{W}$.*

Existence of a smooth Lyapunov function

- *There exist a smooth function $\Psi(x)$ and two κ-functions $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ such that*

$$\alpha_1(\|x\|) \leq \Psi(x) \leq \alpha_2(\|x\|)$$

  for all x.
- *There exists a κ-function $\alpha_3(\cdot)$ such that*

$$D^+\Psi(x) = \nabla\Psi(x)^T f(x, w) \leq -\alpha_3(\|x\|)$$

This theorem was previously proved by Meilakhs [Mei79] under the stronger assumption of exponential stability. It is also to be mentioned that the formulation in [LSW96] is more general than that provided here since the case of convergence to a closed set $\mathcal{A}$ is considered. Basically, the statement in [LSW96] sounds as follows:

[12] For each closed ball $\mathcal{S} \subset \mathbb{R}^n$ there exists ν such that $\|f(x_1, w) - f(x_2, w)\| \leq \nu\|x_1 - x_2\|$ for all $x_1, x_2 \in \mathcal{S}$ and all $w \in \mathcal{W}$.

If the system converges to the closed set $\mathcal{A}$, then there exists a Lyapunov like function which is zero inside this set and positive elsewhere for which

$$D^+\Psi(x) = \nabla\Psi(x)^T f(x,w) \leq -\alpha_3(\text{dist}(x,\mathcal{A}))$$

where $\text{dist}(x,\mathcal{A})$ is the distance of x from $\mathcal{A}$ (which will be formally introduced later in Section 4.2). In particular, this implies that if a system is uniformly ultimately bounded inside a compact neighborhood of the origin, $\mathcal{A}$, then there exists a Lyapunov function outside this set. The discrete version of the theorem is reported in [JW02].

2.6.2 *Literature Review*

In this section, some basic notions concerning Lyapunov theory have been reported. As it has been underlined several times, the main focus has been the geometrical interpretation of the Lyapunov theory and the examination of some formal concepts.

Needless to say, the literature on Lyapunov theory is so huge that it is not possible to provide but a limited review on the subject. Nevertheless, it is mandatory to remind some seminal works as well as some fundamental textbooks as specific references. Beside the already mentioned work of Lyapunov himself [Lya66], it is fundamental to quote the works of La Salle and Lefschetz [LL61], Krasowski [Kra63], Hahn [Hah67], and Hale [Hal69] as pioneering works concerning the stability of motion.

An important work on Lyapunov theory is the book [RHL77]. The reader is referred to this book for further details on the theory of stability and for a complete list of references.

The Lyapunov direct method provides sufficient conditions to establish the stability of a dynamic system. A major problem in the theory is that it is non-constructive in most cases. How to construct Lyapunov and control Lyapunov functions, will be one of the main deals of this book. Lyapunov theory has also played an important role in robustness analysis and robust synthesis of control systems. In connection with the robust stabilization problem, pioneering papers for the construction of quadratic functions are [HB76, Mei79, BCL83, BPF83, Gut79, Lei81, RK89].

Converse Lyapunov Theorems for certain systems are provided in [Hal69, RHL77] and the extensions to uncertain system are in [Mei79, LSW96, JW02]. An interesting connection among robust stabilizability, optimality, and existence of a control Lyapunov function is presented in [FK96a]. The problem of associating a feedback control function with a Control Lyapunov Function has been considered by Artstein [Art83] and a universal formula can be found in [Son98]. This problem has been considered in the context of systems with uncertainties in [FK96b].

There are results concerning the robust stability (stabilization) for linear uncertain systems which show that a linear uncertain system is stable [MP86a, MP86b, MP86c] or robustly stabilizable [Bla95] if and only if there exists a Lyapunov (or a control Lyapunov) function which is a norm (polyhedral or "polynomial"). This problem will be later reconsidered in a constructive way.

It is worth mentioning that the concept of Lyapunov-like function is in some sense related with the concept of partial stability. Basically, a system is partially stable with respect to part of its state variables if these remain bounded and converge, regardless what the remaining do. For further details on this matter, the reader is referred to [Vor98].

2.7 Exercises

1. An equation with finite escape time admits solutions which cannot be defined for all $t \geq 0$. Show that the equation $\dot{x} = x^2$ has finite escape time.
2. Prove the claim of Example 2.1 about the proposed expression of the solution.
3. Show why condition (2.38) is essential for exponential stability. Can you figure out an example in which (2.38) fails and (2.39) holds and you do not have exponential stability? (hint: try the scalar system $\dot{x} = -x^3$ and $\Psi(x) = \exp(-1/x^2)$).
4. For an uncertain system $\dot{x} = f(x, w)$ the existence of a positive definite function $\Psi(x)$ such that $D^+\Psi(x, w) < 0$ for $x \neq 0$ is not sufficient to prove asymptotic stability (and indeed we wrote $D^+\Psi(x, w) < -\phi(\|x\|)$ with ϕ a κ-function in expression (2.35)). Consider $\dot{x}(t) = -w(t)x(t)$, with $0 < w(t) < 1$ and let $\Psi(x) = x^2$. Take, for instance, $x(0) > 0$ and show that for some $w(t)$, $x(t)$ decreases but it does not converge to 0 (hint: let $w(t) = e^{-t}$).
5. In the previous exercise, the interval for w is not compact. Can the problem be fixed under compactness assumptions?
6. Show that if $\Psi(x)$ is a Lyapunov function assuring exponential stability, then $\Psi(x)^m$ (as long as it still satisfies (2.38)) is also a Lyapunov function. How do the convergence measure factors assured by the two function relate to each other?
7. Sketch the proof of Theorems 2.26 and 2.27.
8. Find an example of a system and a differentiable positive definite function which is a control Lyapunov function if the class of control $\mathcal{C}$ is of state-feedback type but it is not a control Lyapunov function if static output feedback controllers are considered.
9. Although we generally desire continuous controllers, discontinuity is often necessary (hence the evidenced difficulties). Show that the (certain) system $\dot{x} = x + |x|u$ is stabilizable (for instance by means of the control Lyapunov function $x^2/2$) but no controller $u = \Phi(x)$ continuous at 0 exists. Can we achieve UUB in a small interval $[-\epsilon, \epsilon]$ with a continuous Φ?

10. Consider the system

$$\dot{x}(t) = a(x(t), w(t))x(t) + b(x(t))u(t)$$

and assume that $\Psi(x(t))$ is a global control Lyapunov function. Show that, for each arbitrarily small $\epsilon > 0$ and each arbitrarily large $\kappa > 0$, there exists a constant $\gamma > 0$ such that, if the control

$$u = -\gamma b(x)^T \nabla \Psi(x)$$

is applied, then there exists T such that for all $x(0) \in \mathcal{N}[\Psi, \kappa]$, $x(t) \in \mathcal{N}[\Psi, \epsilon]$ for $t \geq T$ [BCL83, Meı79].

11. Check that, given the SISO system

$$A = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \end{bmatrix},$$

the state feedback control law of Example 2.37 leads to an unstable closed-loop system for any $\rho > 0$. Show that, on the other hand, for any initial condition $x(0)$, the proposed greedy linear control law is such that $\Psi(x(t)) = |Cx(t)| = e^{-\rho t}|Cx(0)|$. Find a two-dimensional dynamic system, which is reachable and observable, for which the proposed control law results in a stable closed-loop (hint: take (A, B, C) with stable zeros and relative degree 1 $CB \neq 0$)

12. Consider the linear uncertain system

$$\dot{x}(t) = A(w(t))x(t) + Bu(t)$$

with state matrix globally bounded: $\|A(w)\| \leq \mu$, $w \in \mathcal{W}$, and assume that it admits the global quadratic control Lyapunov function $\Psi(x) = x^T P x$ associated with a continuous control function $\Phi(x)$

$$2x^T PA(w)x + 2x^T B\Phi(x) \leq -\alpha^2 \|x\|^2.$$

Show that the system is globally quadratically stabilizable via a linear controller of the form

$$u = -\gamma B^T P x(t),$$

with γ sufficiently large [Meı74, BPF83].

13. Show that $|b| < 1$ is the stabilizability limit via state feedback for system (2.84).

14. Consider the example of Section 2.4.7 with the control $u = -\text{sgn}(x_2)$. Show that the origin with this control is unstable. Hint: consider the rectangle $\mathcal{D}_1(\epsilon) = \{x : |x_1| \leq 1/4, \ |x_2| \leq \epsilon\}$ and show that $\Psi_1 = x_2^2/2$ is a Lyapunov-like function inside $\mathcal{D}_1(\epsilon)$, if ϵ is small enough, so that the trajectory can escape the rectangle only through the vertical edges. Then consider the Lyapunov-like function $\Psi_2 = -x_1$ in the rectangle $\mathcal{D}_2(\epsilon) = \{x \in \mathcal{D}_1 : x_1 \geq \epsilon/2\}$ (or the opposite one) so that no trajectory originating in $\mathcal{D}_2(\epsilon)$ can reach the origin without violating the constraint $x_1 \leq 1/4 \ldots$ (or the constraint $x_1 \geq 1/4$).
15. Sketch the proof of Theorem 2.46.
16. Show that convexity in Proposition 2.53 is essential (Hint: take $x(t+1) = u(t) \in \mathbb{R}^2$, $u = K(\theta_i)x$, $i = 1, 2$, rotation matrices, and a non-convex Lyapunov function whose sublevel sets are kind of "stars" ...).
17. Show that the solution of $\dot{x}(t) = -1/x(t)$ for $x(0) > 0$ reaches the boundary of the existence domain in finite time.

Chapter 3
Convex sets and their representation

3.1 Convex functions and sets

The purpose of this section is to remind the reader essential notions about convex sets and functions which will be useful in the sequel. For a more detailed exposition on convexity, the reader is referred to specialized literature [Roc70, RW98, BV04].

Definition 3.1 (Convex set). A set $\mathcal{S} \in \mathbb{R}^n$ is said to be convex if for all $x_1 \in \mathcal{S}$ and $x_2 \in \mathcal{S}$ we have that

$$\alpha x_1 + (1-\alpha)x_2 \in \mathcal{S} \quad \text{for all } \; 0 \leq \alpha \leq 1. \tag{3.1}$$

It is henceforth assumed that the empty set $\emptyset$ is convex[1]. The point

$$x = \alpha x_1 + (1-\alpha)x_2$$

with $0 \leq \alpha \leq 1$ is called a **convex combination** of the pair x_1 and x_2. The set of all such points is the segment connecting x_1 and x_2. Basically a set is convex if it includes all the segments connecting all the pairs of its points (if any). Some definitions will be recurrently used in the sequel and it is henceforth useful to group them here.

Definition 3.2 (Convex hull). Given a set $\mathcal{S}$, its convex hull is the intersection of all the convex sets containing $\mathcal{S}$.

With a slight abuse of notation, given two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, both in $\mathbb{R}^n$, we will often call convex hull of $\mathcal{S}_1$ and $\mathcal{S}_2$ the convex hull of $\mathcal{S}_1 \bigcup \mathcal{S}_2$ (see Figure 3.1).

[1] To keep the exposition simple, it is assumed that any considered family of convex sets (i.e., ellipsoids, polytopes) includes the empty set.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_3

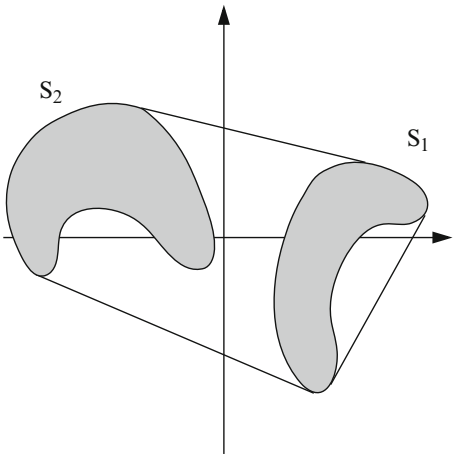


Fig. 3.1 The convex hull of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$

The convex hull of a set $\mathcal{S}$ can be alternatively seen as the smallest convex set containing $\mathcal{S}$.

Definition 3.3 (Centered cone). A set $\mathcal{C} \subset \mathbb{R}^n$ is called a cone centered in $x_0 \in \mathbb{R}^n$ if

$$x_0 + \lambda\,(x - x_0) \in \mathcal{C}, \quad \text{for all } \; x \in \mathcal{C} \quad \text{and } \; \lambda > 0. \tag{3.2}$$

A cone centered in $x_0 = 0$ is simply called a *cone*. A set $\mathcal{C}$ is a convex centered cone if it is a centered cone and convex or, equivalently,

$$x_1, x_2 \in \mathcal{C} \;\text{ implies }\; \alpha(x_1 - x_0) + \beta(x_2 - x_0) \in \mathcal{C}, \;\text{ for all }\; \alpha, \beta \geq 0.$$

The definition of convexity for functions is the following.

Definition 3.4 (Convex function). A real function ψ defined on a convex subset $\mathcal{S}$ of $\mathbb{R}^n$, $\psi : \mathcal{S} \to \mathbb{R}$, is convex if the condition

$$\psi(\alpha x_1 + (1-\alpha)x_2) \leq \alpha\psi(x_1) + (1-\alpha)\psi(x_2) \tag{3.3}$$

holds for all $x_1, x_2 \in \mathcal{S}$ and all $0 \leq \alpha \leq 1$.

It is immediately seen that any sublevel set $\mathcal{N}[\psi, \kappa]$ of a convex function is a convex set. However the opposite is not true, in the sense that functions whose sublevel sets are convex are not necessarily convex. For instance, the function $\phi(x) = \|x\|^2/(1 + \|x\|^2)$ is not convex but its sublevel sets are spheres thus they are convex. In Fig. 3.2 the function $\phi(x)$ and its sublevel sets (bold line) when $x \in \mathbb{R}^2$ are depicted.

The following definition allows a better characterization of a function having convex sublevel sets.

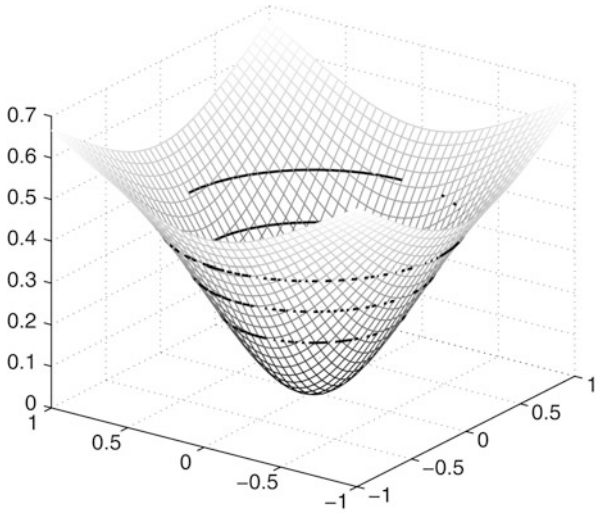


Fig. 3.2 Quasi-convex function $\phi(x) = \|x\|^2/(1+\|x\|^2)$ and level sets

Definition 3.5 (Quasi-convex function). A function $\psi : \mathbb{R}^n \to \mathbb{R}$ is quasi-convex if the condition

$$\psi(\alpha x_1 + (1-\alpha)x_2) \leq \max\{\psi(x_1), \psi(x_2)\} \tag{3.4}$$

holds for all $x_1, x_2 \in \mathbb{R}^n$ and $0 \leq \alpha \leq 1$.

Convexity implies quasi-convexity, while the vice versa is not true. It is easy to show that a function is quasi-convex if and only if its sublevel sets are convex.

A function ψ is said concave if $-\psi$ is convex.

In the cases in which a convex function ψ (or a quasi-convex function) can be naturally defined only on a bounded domain $\mathcal{S}$ (for instance, the function $1/\sqrt{1-\|x\|^2}$ with $\mathcal{S} = \{x : \|x\| < 1\}$) one can always extend the definition over $\mathbb{R}^n$ by just assuming that the function is equal to $+\infty$ outside $\mathcal{S}$. Obviously, in this case, functions with values on the extended real axis $\{-\infty\} \bigcup \mathbb{R} \bigcup +\{\infty\} = [-\infty, \infty]$ have to be considered. In this case, the set of all values in which the function ψ is not infinite is named the effective domain and denoted by *dom* $\psi = \{x : \psi(x) < \infty\}$. A convex function is always continuous on every open set included in its effective domain [Roc70] pag. 82.

There are many important properties associated with convexity. Perhaps, one of the most important is the following [Roc70].

Proposition 3.6. *Any local minimum of the optimization problem*

$$\min_{x \in \mathcal{S}} \psi(x)$$

where $\mathcal{S} \subseteq \mathbb{R}^n$ is a closed convex set and ψ is a quasi-convex function, is a global minimum.

A concept which turns out to be important is the support function of a convex set.

Definition 3.7 (Support function). Given a convex set $\mathcal{S} \subseteq \mathbb{R}^n$, the function $\phi_{\mathcal{S}} : \mathbb{R}^n \to [-\infty, +\infty]$ defined as

$$\phi_{\mathcal{S}}(z) = \sup_{x \in \mathcal{S}} z^T x \tag{3.5}$$

is said to be the support function.[2]

The support function (often referred to as support functional) of a convex set has to be thought in the extended way, in the sense it may assume values in $[-\infty, +\infty]$. A convex and closed set can be represented in terms of its support function. If $\mathcal{S}$ is a convex and closed set, then

$$\mathcal{S} = \left\{ x : z^T x \leq \phi_{\mathcal{S}}(z) \ \ \forall z \in \mathbb{R}^n \right\} \tag{3.6}$$

3.1.1 Operations between sets

There are some basic operations between sets which are necessary to describe the algorithms that will be presented later. Let $\mathcal{A}$ and $\mathcal{B}$ denote generic subsets of $\mathbb{R}^n$, let λ denote a real number, and let $M(x)$ be a map $M : \mathbb{R}^n \to \mathbb{R}^m$

Definition 3.8 (Operations on convex sets).

Sum of sets $\mathcal{A}$ and $\mathcal{B}$ (a.k.a. Minkowski sum, see Fig. 3.3): the set

$$\mathcal{C} = \{x = a + b, \ \ a \in \mathcal{A}, \ \text{ and } \ b \in \mathcal{B}\}$$

Scaled set of $\mathcal{A}$: the set

$$\lambda\mathcal{A} = \{x = \lambda a, \ \ a \in \mathcal{A}\}, \quad \lambda \geq 0$$

Erosion of a set $\mathcal{A}$ with respect to $\mathcal{B}$ [3] (see Fig. 3.4): the set

$$\tilde{\mathcal{A}}_{\mathcal{B}} = \{x : x + b \in \mathcal{A}, \ \text{ for all } b \in \mathcal{B}\}$$

Image of a set under a map M (see Fig. 3.5): the set

$$M(\mathcal{A}) = \{y = M(x), \ \ x \in \mathcal{A}\}$$

Projection of $\mathcal{A}$ on a subspace $\mathcal{X}$ (see Fig. 3.6): the set

$$\mathcal{B} = \{b \in \mathcal{X} : \ \exists\, a \in \mathcal{A} : a = b + c, \ \text{ with } \ c \in \mathcal{X}^{\perp}\}$$

[2] In general the support functional is defined in the dual space; since we are working in $\mathbb{R}^n$, we do not need to introduce any distinction.

[3] Often referred to as Pontryagin difference.

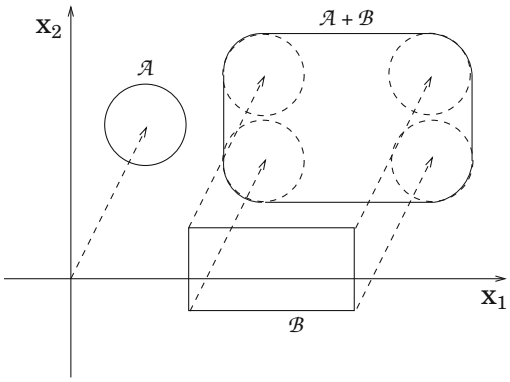


Fig. 3.3 The sum of two sets

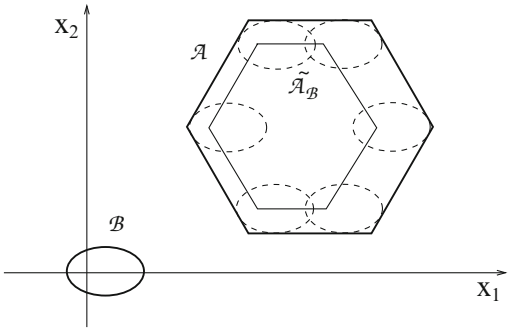


Fig. 3.4 The erosion of $\mathcal{A}$ with respect to $\mathcal{B}$

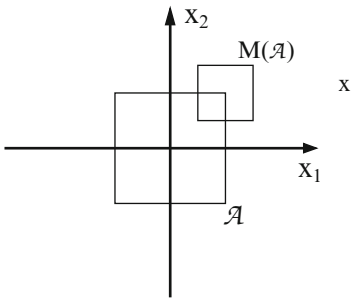


Fig. 3.5 The image of $\mathcal{A} = \{x : \|x\|_\infty \leq 1\}$ under the map $M(x) = 0.5x + [1\ \ 1]^T$

where $\mathcal{X}^\perp$ denotes the subspace orthogonal to $\mathcal{X}$.

It is immediately seen that, if $\mathcal{A}$ and $\mathcal{B}$ are convex, then

- their sum is convex;
- the scaled set is convex;
- the erosion is convex;
- if $M(x)$ is an affine map (the sum of a constant and a linear map), then $M(\mathcal{A})$ is convex.

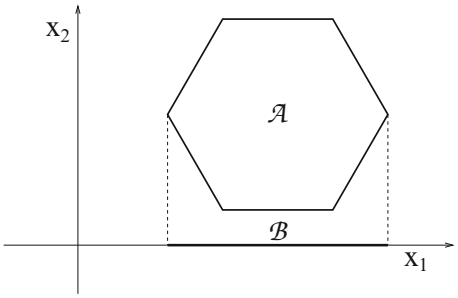


Fig. 3.6 The projection of $\mathcal{A}$ on $\mathcal{X} = \{x : \ x_2 = 0\}$

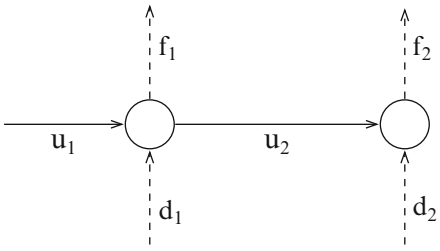


Fig. 3.7 The network considered in Example 3.9

Convexity preserving in the case of a map requires strong properties. For instance that of being affine is sufficient, but not actually necessary. We refer the reader to [BV04] for an extended description of convexity preserving maps.

Example 3.9. Consider the simple network in Figure 3.7 in which two agents u and d operate by a choosing their components within given constraints (representing the capacity of the arcs)

$$u \in \mathcal{U} = \{u : \ u_i^- \leq u_i \leq u_i^+, \ i = 1, 2\}, \quad d \in \mathcal{D} = \{d : \ d_i^- \leq d_i \leq d_i^+, \ i = 1, 2\}.$$

The resulting outcoming flow is represented by vector f given by

$$f = Bu + d$$

where

$$B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

As for the interaction between u and d, two extreme cases can be considered:

- u and d cooperate;
- u and d are in competition.

In the first case, the achievable flows are

$$f \in B\mathcal{U} + \mathcal{D}$$

In the other case the situation is different. If, for instance, u plays the role of the good player and d that of the opposer, it is interesting to consider the "worst case" set of all the flows f such that no matter which is the move of $d \in \mathcal{D}$ there exists $u \in \mathcal{U}$ (depending on d) such that $Bu + d = f$. This set is precisely

$$f \in [\tilde{B\mathcal{U}}]_{\mathcal{D}}$$

It turns out that in both cases the resulting set is a polyhedron (this example will be reconsidered later, after the introduction of some elementary notions about polyhedral sets).

It is worth reminding here some basic properties which concern the operations just introduced. Let $\mathcal{A}$ and $\mathcal{B}$ be compact and convex sets. Then:

- $\mathcal{A} \subseteq \mathcal{A} + \mathcal{B}$ if $0 \in \mathcal{B}$;
- $\tilde{\mathcal{A}}_{\mathcal{B}} \subseteq \mathcal{A}$ if and only if $0 \in \mathcal{B}$;
- $\tilde{\mathcal{A}}_{\mathcal{B}} + \mathcal{B} \subseteq \mathcal{A}$
- Let $\bar{\mathcal{B}}$ be the closure of the convex hull of $\mathcal{B}$. Then $\tilde{\mathcal{A}}_{\mathcal{B}} = \tilde{\mathcal{A}}_{\bar{\mathcal{B}}}$.

3.1.2 Minkowski function

An important definition which will be often used in the sequel is that of a C-set.

Definition 3.10 (C-set). A C-set is a convex and compact subset of $\mathbb{R}^n$ including the origin as an interior point.

Given a C-set $\mathcal{S} \subset \mathbb{R}^n$, it is always possible to define a function, named after Minkowski, which is essentially the function whose sublevel sets are achieved by linearly scaling the set $\mathcal{S}$ (see Fig. 3.8).

Definition 3.11 (Minkowski function). Given a C-set $\mathcal{S}$, its Minkowski function[4] is

$$\psi_{\mathcal{S}}(x) = \inf\{\lambda \geq 0 : \ x \in \lambda \mathcal{S}\}$$

The Minkowski function $\psi_{\mathcal{S}}$ satisfies the following properties [Lue69], p. 131.

Proposition 3.12 (Minkowski function properties).

- *It is positive definite:* $0 \leq \psi_{\mathcal{S}}(x) < \infty$ *and* $\psi_{\mathcal{S}}(x) > 0$ *for* $x \neq 0$.
- *It is positively homogeneous of order* 1*:* $\psi_{\mathcal{S}}(\lambda x) = \lambda \psi_{\mathcal{S}}(x)$ *for* $\lambda \geq 0$.
- *It is sub-additive:* $\psi_{\mathcal{S}}(x_1 + x_2) \leq \psi_{\mathcal{S}}(x_1) + \psi_{\mathcal{S}}(x_2)$.
- *It is continuous.*
- *Its unit ball is* $\mathcal{S} = \mathcal{N}[\psi_{\mathcal{S}}, 1]$.
- *It is convex.*

[4] Also known as Minkowski functional.

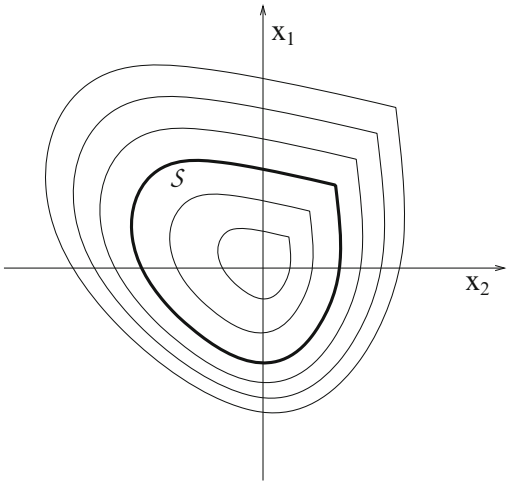


Fig. 3.8 The level surfaces of the Minkowski function

A function having the above properties is named gauge function, because it introduces a measure of the distance from the origin. Since the closed unit ball $\mathcal{N}[\psi, 1]$ of a gauge function is a C-set, gauge functions and C-sets are in a one-to-one correspondence.

If a C-set $\mathcal{S}$ is 0-symmetric, that is

$$x \in \mathcal{S} \Rightarrow -x \in \mathcal{S},$$

for all $x \in \mathbb{R}^n$, then

$$\psi_{\mathcal{S}}(x) = \psi_{\mathcal{S}}(-x),$$

and in this case $\psi_{\mathcal{S}}$ is a norm. Therefore, the concept of a gauge function generalizes the concept of a norm to functions with similar properties but with possibly non-zero-symmetric unit balls.

The concept of C-set introduced here (proper C-set) could be relaxed to that of a non-proper C-set, namely a convex and closed set $\mathcal{S}$ (not necessarily compact) including 0 in the interior. In this case Definition 3.11 of Minkowski function still holds. Function $\psi_{\mathcal{S}}$ is convex and positive semi-definite. In the case of a symmetric set $\mathcal{S}$, $\psi_{\mathcal{S}}$ is a semi-norm [5].

The following definition turns out to be useful to introduce the concept of duality between support functions and Minkowski functions.

[5] A semi-norm has the properties of a norm except positive definiteness: it is a non-negative function $\psi(x) \geq 0$ such that $\psi(\lambda x) = |\lambda|\psi(x)$, $\psi(x + y) \leq \psi(x) + \psi(y)$, but $x \neq 0 \not\Rightarrow \psi(x) > 0$ (e.g., in $\mathbb{R}^2$, $|x_1 + x_2|$ is a semi-norm).

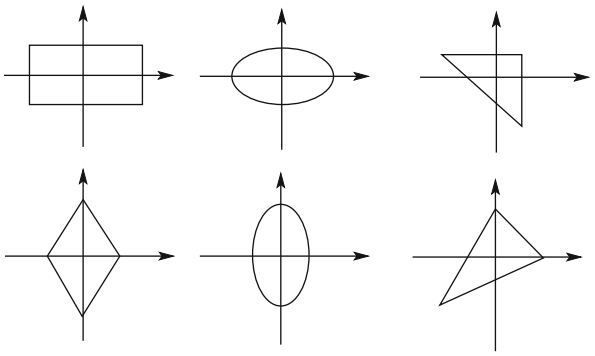


Fig. 3.9 Examples of C-sets (top) and their polar sets (down)

Definition 3.13 (Polar of a C-set). Consider a C-set $\mathcal{P}$ (Fig. 3.9). Then its polar set is defined as

$$\mathcal{P}^* = \{x : \ z^T x \leq 1, \ \text{ for all } \ z \in \mathcal{P}\} \tag{3.7}$$

The following properties hold for any C-set $\mathcal{P}$ (see [Roc70] Theorem 14.5).

Proposition 3.14.

- $[\mathcal{P}^*]^* = \mathcal{P}$.
- *If $\psi_{\mathcal{P}}(x)$ is the Minkowski function of $\mathcal{P}$ and $\phi_{\mathcal{P}^*}(x)$ is the support functional of the polar set $\mathcal{P}^*$, then*

$$\psi_{\mathcal{P}}(x) = \phi_{\mathcal{P}^*}(x).$$

3.1.3 The normal and the tangent cones

Throughout this book, convex functions will often be considered as candidate Lyapunov functions for dynamical systems. As mentioned in the previous chapter, this requires the exploitation of the concept of directional derivative. Assume a convex function $\psi(x)$, defined on a convex open set, is given and let the difference quotient be defined as:

$$R(\psi, x, f, \tau) = \frac{\psi(x + \tau f) - \psi(x)}{\tau}$$

A fundamental property is that for each pair of vectors x and f, the difference quotient is a non-decreasing function of τ, as long as ψ is convex. Therefore, the

directional derivative defined in (2.24) in the previous section can be equivalently replaced by the formula (see [Roc70] for details)

$$D^+\psi(x,f) = \lim_{\tau\to 0+} R(\psi,x,f,\tau) = \inf_{\tau\geq 0} R(\psi,x,f,\tau) \tag{3.8}$$

(note that $f = f(x,w)$ if we are considering the Lyapunov derivative of $\dot{x} = f(x,w)$). The existence of the limit is assured by the monotonicity of the difference quotient R. In the case of convex functions which are not differentiable, it is convenient to use the notion of subgradient already introduced: given the subdifferential

$$\partial\psi(x) = \{z : \ z^T(y-x) \leq \psi(y) - \psi(x), \ \ \forall y \in \mathbb{R}^n\}$$

we have that

$$D^+\psi(x,f) = \sup_{z\in\partial\psi(x)} z^T f \tag{3.9}$$

The subdifferential of a convex function is a convex set for all x. An important notion related to that of subdifferential is that of normal cone.

Definition 3.15 (Normal cone). Given a closed and convex set $\mathcal{S}$, the normal cone at $\mathcal{S}$ in x is defined as follows (see Fig. 3.10):

$$\mathcal{H}_{\mathcal{S}}(x) = \{z : \ z^T(y-x) \leq 0, \ \text{ for all } \ y \in \mathcal{S}\}$$

The normal cone is trivially $\{0\}$ if $x \in \mathit{int}\ \mathcal{S}$, the interior of $\mathcal{S}$. Assume now that a closed and convex set with a non-empty interior is given as the sublevel set of a convex function

$$\mathcal{S} = \mathcal{N}[\psi,\kappa]$$

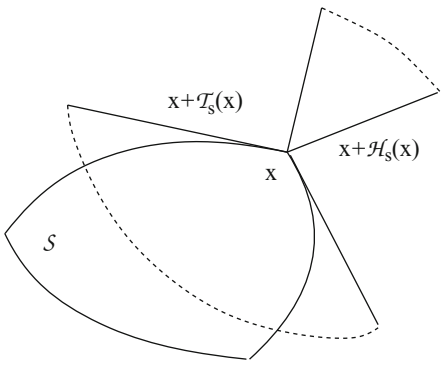


Fig. 3.10 The normal and tangent cones

Assume that κ is not a minimum of ψ. Then $\psi(x) < \kappa$, for all $x \in int\ \mathcal{S} \neq \emptyset$, and consider a point x on its boundary, i.e. such that $\psi(x) = \kappa$. Then the normal cone is the closure of the cone generated by the vectors of the subgradient of ψ in x (see [Roc70]).

$$\mathcal{H}_{\mathcal{S}}(x) = \{z = \lambda w,\ \ \lambda \geq 0,\ \ \text{for some}\ \ w \in \partial\psi(x)\}.$$

For a convex differentiable function ψ, under the previous assumption that κ is not the minimum value, if x is on the boundary of $\mathcal{N}[\psi, k]$, then the set $\mathcal{H}_{\mathcal{S}}(x)$ is generated by the gradient vector $\nabla\psi(x)$. The plane of all vectors y such that $\nabla\psi(x)(y - x) = 0$ is the tangent plane to $\mathcal{S}$ in x. The following notion of tangent cone generalizes that of tangent plane.

Definition 3.16 (Tangent cone). Given a closed and convex set $\mathcal{S}$ the tangent cone at $\mathcal{S}$ in x is defined as follows (see Fig. 3.10):

$$\mathcal{T}_{\mathcal{S}}(x) = cl\ \left\{ \bigcup_{h>0} \frac{1}{h}(\mathcal{S} - x) \right\} \tag{3.10}$$

where $cl\{\cdot\}$ denotes the closure of the set[6]

Given a cone $\mathcal{C}$ we say that $\mathcal{C}^*$ is the polar cone if

$$\mathcal{C}^* = \{z :\ \ z^T x \leq 0,\ \ \text{for all}\ \ x \in \mathcal{C}\} \tag{3.11}$$

A fundamental relation between the normal and the tangent cone is that they are polar to each other as per the following proposition (see [AC84], Section 5.1, Proposition 2, pag. 220).

Proposition 3.17. *Let $\mathcal{S}$ be a closed convex set with a non-empty interior and $x \in \mathcal{S}$. Then $\mathcal{T}_{\mathcal{S}}(x)$ and $\mathcal{H}_{\mathcal{S}}(x)$ are both closed and convex cones and*

$$\mathcal{T}_{\mathcal{S}}(x) = \mathcal{H}_{\mathcal{S}}(x)^*$$

and

$$\mathcal{H}_{\mathcal{S}}(x) = \mathcal{T}_{\mathcal{S}}(x)^*.$$

Note that for $x \in int\ \mathcal{S}$ the proposition is trivial but coherent because $\mathcal{H}_{\mathcal{S}}(x) = \{0\}$ and $\mathcal{T}_{\mathcal{S}}(x) = \mathbb{R}^n$. Note also that, for the sake of simplicity in the representation, the cones depicted in Fig. 3.10 are not centered in 0, but in x, so they are actually the

[6] This is the closure of the set of all vectors of the form $(s - x)/h$, with $s \in \mathcal{S}$ and $h > 0$.

translated versions of the normal and the tangent ones. As a final remark, it has to be pointed out that the definition of normal cone $\mathcal{T}_S(x)$ given here is valid for a convex set only. A more general definition, valid for general closed sets, will be given later.

3.2 Ellipsoidal sets

A famous class of convex sets, definitely the most popular in the dynamic systems control area, is that of the ellipsoidal sets or ellipsoids. Given a vector x_0, named the center, and a positive definite matrix P an ellipsoid is a set of the form

$$\mathcal{E}(x_0, P, \mu) \doteq \{x : \sqrt{(x - x_0)^T P (x - x_0)} \leq \mu\} \tag{3.12}$$

If the ellipsoid is centered in the origin, then it is possible to write:

$$\mathcal{E}(P, \mu) \doteq \mathcal{E}(0, P, \mu) = \mathcal{N}[\sqrt{x^T P x}, \mu] = \mathcal{N}[\|x\|_P, \mu]$$

When $\mu = 1$, for brevity, the following notation will be adopted: $\mathcal{E}(P) \doteq \mathcal{E}(0, P, 1) = \mathcal{N}[\|x\|_P, 1]$. By defining the root of a positive definite matrix P as the unique positive symmetric matrix $R = P^{1/2}$ such that $R^2 = P$, it is possible to derive an alternative dual representation for an ellipsoidal set:

$$\mathcal{D}(x_0, Q, \mu) = \{x = x_0 + \mu\, Q^{\frac{1}{2}} z, \quad \text{where} \quad \|z\|_2 \leq 1\} \tag{3.13}$$

Again, dropping the first argument implies that the ellipsoid is centered in 0, $\mathcal{D}(Q, \mu) = \mathcal{D}(0, Q, \mu)$, and dropping the third argument means that the radius is 1 $\mathcal{D}(Q) = \mathcal{D}(0, Q, 1)$.

By defining $P^{-1/2} = R^{-1} = [P^{1/2}]^{-1}$, the following proposition holds true:

Proposition 3.18. *If $Q = P^{-1}$, expressions (3.12) and (3.13) represent the same set. In particular $\mathcal{E}(P) = \mathcal{D}(Q)$*

Proof. Consider the transformation $z = P^{1/2}(x - x_0)/\mu$. Then x is in (3.12) if and only if $z^T z \leq 1$. In the new reference frame, this condition represents the unit ball, the set of all z such that $\|z\|_2 \leq 1$. By applying the inverse transformation it can be seen that this set is in a one-to-one correspondence with the set of all x such that

$$x = x_0 + \mu P^{-1/2} z, \quad \text{for some} \quad \|z\|_2 \leq 1$$

which is equivalent to (3.13)

A further representation for an ellipsoid is achievable by means of its support functional (3.6). It is easy to see that the support function of an ellipsoid $\mathcal{E}(P, \mu)$ is

$$\phi_{\mathcal{E}}(z) = \mu \sqrt{z^T P^{-1} z} = \mu \sqrt{z^T Q z} = \mu \|z\|_Q$$

and therefore

$$\mathcal{E}(P,\mu) = \mathcal{D}(Q,\mu) = \left\{x : z^T x \leq \mu\sqrt{z^T Q z}, \text{ for all } z\right\} \tag{3.14}$$

(actually, it is possible to replace "for all z" with "for all $\|z\|_2 = 1$" in the previous expression). An ellipsoidal set is uniquely determined by the entries of matrix P and by the components of x_0. Since P is symmetric, the complexity of the representation (the number of required free parameters) is

$$n(n+1)/2 + n = n(n+3)/2.$$

An ellipsoidal set has some special properties which are useful in the application.

Proposition 3.19.

- *The intersection of an ellipsoid and a subspace $\mathcal{X} \subset \mathbb{R}^n$ is an ellipsoid.*
- *If M is a full row rank matrix and $\mathcal{E}$ is an ellipsoid, then $M\mathcal{E}$ is an ellipsoid.*
- *The projection of an ellipsoid on a subspace $\mathcal{X}$ is an ellipsoid (inside the projection subspace).*

Proof. To show the first property consider representation (3.12) and apply the transformation $z = P^{1/2}x$ and $z_0 = P^{1/2}x_0$, so that the ellipsoid becomes the sphere determined by $(z - z_0)^T(z - z_0) \leq \mu^2$. The subspace $\mathcal{X}$ is mapped in the subspace $\mathcal{Z} = P^{1/2}\mathcal{X}$. Let B be any basis matrix for $\mathcal{Z}$.

Vector z_0 can be decomposed as $z_0 = Bw_0 + h_0$ where h_0 is in the orthogonal of the subspace spanned by B, so that $h_0^T B = 0$. Let us consider the identical decomposition for $z = Bw + h$. Define $h_* = h - h_0$. Then

$$(z - z_0)^T(z - z_0) = [(B(w - w_0) + h_*]^T[(B(w - w_0) + h_*] \leq \mu^2$$

which is equivalent to

$$(w - w_0)^T(B^T B)(w - w_0) \leq \mu^2 - h_*^T h_*$$

Since B is a basis matrix, $(B^T B)$ is positive definite. Therefore the above expression represents an ellipsoid (possibly empty) in the w-space.

To prove the second property consider an $m \times n$ full row rank matrix (so that $m \leq n$) and the set $\mathcal{Y} = M\mathcal{E}$, where $\mathcal{E}$ is represented by means of (3.12). Apply the transformation $z = P^{1/2}x/\mu$ and $z_0 = P^{1/2}x_0/\mu$, so that the ellipsoid becomes the sphere determined by $(z - z_0)^T(z - z_0) \leq 1$. This is the unit ball centered in z_0 namely the set of all $z = z_0 + w$, with $\|w\|_2 \leq 1$. Consider the transformed matrix $y = Mx = \hat{M}z$ where $\hat{M} = \mu M P^{-1/2}$, and its singular value decomposition $\hat{M} = U[\Sigma\ 0]W$, with U and W orthonormal matrices and Σ diagonal square. Then the image is the set of all vectors of the form

$$y = U[\Sigma\ 0]W(z_0 + w) = y_0 + U[\Sigma\ 0]\hat{w}, \quad \|\hat{w}\|_2 \leq 1$$

where $y_0 = \hat{M} z_0$ and $\hat{w} = Ww$ is any arbitrary unit vector (in view of the fact that W is orthonormal). Denoting by $\hat{w}_1$ the first m components of $\hat{w}$, it turns out that the set of all y is

$$y = y_0 + U\Sigma\, \hat{w}_1, \quad \text{for all } \|\hat{w}_1\|_2 \leq 1.$$

This can be written equivalently as

$$\|[U\Sigma]^{-1}(y - y_0)\|_2 = \|\hat{w}_1\|_2 \leq 1$$

Since $[U\Sigma]^{-1} = \Sigma^{-1}U^T$, this means that the set of all admissible y is characterized by

$$\|[U\Sigma]^{-1}(y - y_0)\|_2^2 = (y - y_0)^T U\Sigma^{-2}U^T(y - y_0) \leq 1$$

therefore it is an ellipsoid ($\hat{P} = U\Sigma^{-2}U^T$ is positive definite).

The third property follows immediately from the second, because the projection on a subspace is a linear operator that can be represented by the linear map $y = B^T x$ where B is an orthonormal basis.

Other operations between ellipsoids do not preserve the ellipsoidal structure. In particular

- the sum of two ellipsoids is not in general an ellipsoid;
- the intersection of two ellipsoids is not in general an ellipsoid;
- the erosion of an ellipsoid (even with respect to an ellipsoid) is not in general an ellipsoid.
- the convex hull of the union of two ellipsoids is not an ellipsoid;

Therefore, the above operations increase the complexity of the resulting set. It can be easily seen that the negative assertions above hold even if ellipsoids centered in the origin are considered.

As a final point, notice that, given an ellipsoidal set $\mathcal{E}(0, P, 1) = \mathcal{E}(P, 1)$ centered in zero, its Minkowski functional is the quadratic norm

$$\|x\|_P = \sqrt{x^T P x},$$

and its support function is

$$\|x\|_Q = \sqrt{x^T Q x}.$$

where $Q = P^{-1}$. In agreement with Proposition 3.14 we have that

$$\mathcal{E}(P) = \mathcal{D}(Q)$$

and

$$\mathcal{E}(Q) = \mathcal{D}(P)$$

are dual to each other while $\|x\|_P$ and $\|x\|_Q$ are dual norms.

3.3 Polyhedral sets

An important family of convex sets of common interest is that of polyhedral sets. The main advantage of polyhedral sets is the fact that they form a closed family with respect to the mentioned operations. Their main disadvantage is that the complexity of representation is not fixed by the space dimension.

Definition 3.20 (Polyhedral set). A convex polyhedral set is a set of the form

$$\mathcal{P}(F,g) = \{x : \ Fx \leq g\} = \{x : \ F_i x \leq g_i, \ \ i = 1,2,\ldots,s\} \tag{3.15}$$

where F_i denotes the i-th row of the $s \times n$ matrix F and g_i the i-th component of the $s \times 1$ vector g.

A polyhedral set includes the origin if and only if $g \geq 0$ and includes the origin as an interior point if and only if $g > 0$ ($g_i > 0, \ \forall i$). We will use the notation $\bar{1}$ to mean the vector with all components equal to 1

$$\bar{1}^T = \ [\,1\ 1\ \ldots\ 1\,] \tag{3.16}$$

Thus a polyhedral set including the origin can be always represented as

$$\mathcal{P}(F,\bar{1}) = \mathcal{P}(F) = \{x : \ Fx \leq \bar{1}\} = \{x : \ F_i x \leq 1, \ \ i = 1,2,\ldots,s\} \tag{3.17}$$

which can be achieved from (3.15) by dividing both sides of each inequality by $g_i > 0$. A 0-symmetric convex polyhedral set can be always represented in the form

$$\bar{\mathcal{P}}(F,g) = \{x : \ \ -g \leq Fx \leq g\} = \{x : \ \ -g_i \leq F_i x \leq g_i, \ \ i = 1,2,\ldots,s\} \tag{3.18}$$

Again, if $\bar{\mathcal{P}}(F,g)$ includes 0 as an interior point, up to a normalization, it can be represented as

$$\bar{\mathcal{P}}(F,\bar{1}) = \bar{\mathcal{P}}(F) = \{x : \ -\bar{1} \leq Fx \leq \bar{1}\} = \{x : \ -1 \leq F_i x \leq 1, \ i = 1,2,\ldots,s\} \tag{3.19}$$

A convex polyhedron admits a vertex representation of the form

$$\mathcal{V}(X_w, X_y) = \{x = X_w w + X_y y, \quad \sum_{i=1}^{p} w_i = 1, \ \ w \geq 0, \ y \geq 0\} \tag{3.20}$$

The columns of matrix X_w represent the set of finite vertices, while those of matrix X_y represent the set of infinite directions or "infinite vertices." In the symmetric case, the following representation is possible:

$$\bar{\mathcal{V}}(X_w, X_y) = \{x = X_w w + X_y y, \quad \sum_{i=1}^{p} |w_i| \leq 1, \ \ y \ \text{ arbitrary}\} \tag{3.21}$$

In this case, the finite (respectively infinite) vertices are given by the columns of the matrices X_w (respectively X_y) and their opposite.

Definition 3.21 (Polytope). A bounded polyhedral set is called a polytope.

A necessary and sufficient condition for (3.20) or (3.21) to represent a polytope is that $X_y = 0$. It is not difficult to see that expression (3.18) or expression (3.19) represents bounded sets if and only if F has full column rank. A condition for $\mathcal{P}(F)$ to be a polytope is given in Exercise 11.

In the book, the following notation

$$\mathcal{P}(F) = \{x : \ \ Fx \leq \bar{1}\} \tag{3.22}$$

and its dual

$$\mathcal{V}(X) = \{x = Xz, \ \ \bar{1}^T z = 1, \ z \geq 0\} \tag{3.23}$$

will often be used. The corresponding notations for symmetric sets are

$$\bar{\mathcal{P}}(F) = \{x : \ \ \|Fx\|_\infty \leq 1\} \tag{3.24}$$

and its dual

$$\bar{\mathcal{V}}(X) = \{x = Xz, \ \ \|z\|_1 \leq 1\} \tag{3.25}$$

The duality between the plane (3.22) and vertex (3.23) representations (or the analogous (3.24) and (3.25)) can be explained as follows.

Consider a polytope $\mathcal{P} = \mathcal{P}(F) = \mathcal{V}(X)$ including the origin as an interior point[7] and its polar set $\mathcal{P}^*$(as defined in expression (3.7)). Then the following properties hold

$$\mathcal{P}(F)^* = \mathcal{V}(F^T) \tag{3.26}$$

and

$$\bar{\mathcal{V}}(X)^* = \bar{\mathcal{P}}(X^T) \tag{3.27}$$

Let us explain why (3.26) is true (the case of (3.27) is left to the reader). The Minkowski (gauge) functions associated with $\mathcal{P}(F)$ and $\mathcal{V}(X)$ are

$$\psi_{\mathcal{P}(F)}(x) = \max\{Fx\} \doteq \max_i\{F_i\, x\} \tag{3.28}$$

[7] Boundedness and the inclusion of 0 as an interior point are assumed for simplicity.

and

$$\psi_{\mathcal{V}(X)}(x) = \min\{\bar{1}^T w : \ x = Xw, \ w \geq 0\}, \tag{3.29}$$

respectively. The polarity between $\mathcal{P}(F)$ and $\mathcal{V}(X)$ can be seen from the linear programming duality. Consider the support function of $\mathcal{P}(F)$

$$\begin{aligned}\phi_{\mathcal{P}(F)}(z) &= \max \ \{z^T x : \ Fx \leq \bar{1}\} \\ &= \min \ \{\bar{1}^T w : \ s.t. \ F^T w = z, \ w \geq 0\} = \psi_{\mathcal{V}(F^T)}(z)\end{aligned}$$

which is the Minkowski function of $\mathcal{V}(F^T)$. In view of Proposition 3.14, $\mathcal{V}(F^T)$ and $\mathcal{P}(F) = \mathcal{V}(X)$ polar. As a consequence of this fact the support function is achieved by the expressions (3.28) and (3.29) if F and X are replaced by X^T and F^T, respectively.

Important cases of polyhedral sets are introduced next.

Definition 3.22 (Simplex). Given a full row rank matrix $X \in \mathbb{R}^{n\times(n+1)}$ a simplex is the convex hull of the sets $\mathcal{S}_i = x_i$, where each of the x_i is the i-th column vector of matrix X.

Definition 3.23 (Diamond set). Given a full rank matrix $X \in \mathbb{R}^{n\times n}$, a diamond is the convex hull of the sets $\mathcal{S}_i = \pm x_i$, where each of the x_i is the i-th column vector of matrix X.

Definition 3.24 (Simplicial cone). Given a full rank matrix $X \in \mathbb{R}^{n\times n}$, a *simplicial cone* $\mathcal{C} \subset \mathbb{R}^n$ is a set of the form

$$\mathcal{C} = \{x = Xp : \ p \geq 0\}. \tag{3.30}$$

A simplicial cone is always generated by the simplex having the origin amongst its vertices, say the simplex generated by the $n \times (n+1)$ matrix $X_1 = [X \ \ 0]$, where 0 is the $n \times 1$ vector with zero entries.

A further important concept is the minimality of the representation.

Definition 3.25 (Minimal representation). A plane or vertex representation is minimal if and only if there is no other representation of the same set involving a smaller (with respect to dimensions) F or X.

A minimal representation of a set can be achieved by removing from the plane (vertex) representation all the redundant planes (vertices), whose definition is reported next.

Definition 3.26 (Redundant plane). Given the polyhedral set $\mathcal{P}(F, g)$, let $\mathcal{P}(\tilde{F}_{-i}, \tilde{g}_{-i})$ be the set obtained by removing the i-th plane F_i from matrix F (and the corresponding component g_i of vector g). The plane F_i is said to be redundant if

$$\max_{x\in\mathcal{P}(\tilde{F}_{-i},\tilde{g}_{-i})} F_i x \leq g_i$$

Definition 3.27 (Redundant vertex). Given the polyhedral set $\mathcal{V}(X)$, let $\tilde{X}_{-i}$ be the vertex matrix obtained by removing vertex x_i from the vertex matrix X. The vertex x_i is said to be redundant if

$$\min\{1^T\zeta : \ x_i = \tilde{X}_{-i}\zeta\} \leq 1$$

Note that checking any of the presented conditions requires the solution of a linear programming problem.

Once again it has to be stressed that the representation complexity of polyhedral sets, differently from ellipsoidal sets, is not a function of the space dimension only, but it may be arbitrarily high. The usual complexity index of a polyhedral set is the number of rows of matrix F, for the plane representation $\mathcal{P}(F)$, and the number of vertices, for the vertex representation $\mathcal{V}(X)$. As far as the complexity issue is concerned, none of these representations can be regarded as more convenient. Indeed, as we have seen, for any polyhedral C-set with n_p planes and n_v vertices, the dual has exactly $n_p^* = n_v$ planes and $n_v^* = n_p$ vertices. For high dimensional problems, passing from a representation to the other, namely determining the vertices from the planes and vice versa, is a hard task. It is worth saying that in general the algorithms which involve polyhedra computations are very demanding in terms of computational complexity and thus it is often mandatory to work with minimal representations to keep the complexity as low as possible. Unfortunately, computing a minimal representation can also turn out to be a computational demanding task.

The counterpart of the computational troubles of polyhedral set is their flexibility. Indeed any convex and compact set can be arbitrarily closely approximated by a polyhedron (see [Lay82]). In particular, if $\mathcal{S}$ is a C-set, then for all $0 < \epsilon < 1$ there exists a polytope $\mathcal{P}$ such that

$$(1-\epsilon)\mathcal{S} \subseteq \mathcal{P} \subseteq \mathcal{S} \tag{3.31}$$

(internal approximation) or

$$\mathcal{S} \subseteq \mathcal{P} \subseteq (1+\epsilon)\mathcal{S} \tag{3.32}$$

(external approximation). The same concept can be expressed in terms of Hausdorff distance. Denote by $\mathcal{B}$ the unit ball of any norm. The Hausdorff distance between two sets $\mathcal{S}$ and $\mathcal{R}$ can be expressed as

$$\delta_H(\mathcal{S},\mathcal{R}) = \min \ \{\alpha \geq 0 : \ \mathcal{R} \subseteq \mathcal{S} + \alpha\mathcal{B}, \ \ \mathcal{S} \subseteq \mathcal{R} + \alpha\mathcal{B}\}$$

Then we can always find a polyhedral approximation arbitrarily close to any convex C-set in the sense of Hausdorff. In particular given $\epsilon > 0$ and a C-set $\mathcal{S}$, we can find an internal polytope $\mathcal{P}$ such that

$$\mathcal{S} \subseteq \mathcal{P} + \epsilon\mathcal{B}$$

or an external polytope such that

$$\mathcal{P} \subseteq \mathcal{S} + \epsilon\mathcal{B}$$

The class of polyhedral sets is closed with respect to the basic operations already considered according to the next proposition.

Proposition 3.28. *If $\mathcal{A}$ and $\mathcal{B}$ are polyhedra, $\lambda \geq 0$ and M is an affine map, then*

- *the image $M(\mathcal{A})$ is a polyhedron;*
- *the preimage $M^{-1}(\mathcal{A})$ is a polyhedron;*
- *the projection on a subspace of $\mathcal{A}$ is a polyhedron;*
- *$\mathcal{A} \bigcap \mathcal{B}$ is a polyhedron;*
- *the intersection of $\mathcal{A}$ with a subspace $\mathcal{X} \subset \mathbb{R}^n$ is a polyhedron;*
- *the scaled set $\lambda\mathcal{A}$ is a polyhedron;*
- *the sum $\mathcal{A} + \mathcal{B}$ is a polyhedron;*
- *the erosion of $\mathcal{A}$ with respect to $\mathcal{B}$ $\tilde{\mathcal{A}}_{\mathcal{B}}$ is a polyhedron*[8]*;*
- *the convex hull $conv\{\mathcal{A} \bigcup \mathcal{B}\}$ is a polyhedron.*

Proof. The fact that $M(\mathcal{A})$ is a polyhedron follows from the representation (3.23) $\mathcal{A} = \mathcal{V}(X)$, since it is immediate to see that

$$M\mathcal{V}(X) = \mathcal{V}(MX)$$

To show that $M^{-1}(\mathcal{A}) = \{x : Mx \in \mathcal{A}\}$ is also polyhedral, consider the representation (3.15). It follows that the preimage is identified by the inequalities

$$FMx \leq g,$$

This can be synthetically written as follows:

$$M^{-1}(\mathcal{A}) = \mathcal{P}(FM, g).$$

To show that the projection of $\mathcal{A}$ is a polyhedron, it is sufficient to remind that the projection operator is a linear map.

To prove that the intersection of two polyhedra is itself a polyhedron, consider the representation (3.15). Then, the intersection of two polyhedra is the set of vectors which satisfies all the inequalities associated with the two elements. If $\mathcal{A} = \mathcal{P}(F^A, g^A)$ and $\mathcal{B} = \mathcal{P}(F^B, g^B)$, then

$$\mathcal{P}(F^B, g^B) \bigcap \mathcal{P}(F^A, g^A) = \mathcal{P}\left(\begin{bmatrix} F^A \\ F^B \end{bmatrix}, \begin{bmatrix} g^A \\ g^B \end{bmatrix}\right)$$

[8] This property holds even if $\mathcal{B}$ is a generic closed set.

To show that the intersection of $\mathcal{A}$ with a subspace $\mathcal{X}$ is a polyhedron, it is sufficient to notice that any subspace can be represented by the kernel of a proper matrix B^T, where B is a basis of the orthogonal subspace, namely $\mathcal{X} = \{x : \ B^T x = 0\}$. This is equivalent to the inequalities

$$0 \leq B^T x \leq 0,$$

say the intersection of $\mathcal{A}$ with a subspace is nothing but the intersection between $\mathcal{A}$ and another polyhedron and thus, by the previously shown statement, it is a polyhedron.

The scaled set $\lambda\mathcal{A}$ is a polyhedron because scaling is equivalent to applying the linear operator λI. Then from (3.23) we see that

$$\lambda\mathcal{V}(X) = \mathcal{V}(\lambda X)$$

and for $\lambda > 0$

$$\lambda\mathcal{P}(F) = \mathcal{P}(F/\lambda)$$

(note that if $\lambda = 0$, $\lambda\mathcal{P}(F) = \{0\}$, and then the representation $\mathcal{P}(F)$ cannot be used).

To show that $\mathcal{A} + \mathcal{B}$ is a polyhedron, consider for brevity the case in which both $\mathcal{A}$ and $\mathcal{B}$ are bounded and represented as in (3.23). If $\mathcal{A} = \mathcal{V}(X^A)$ and $\mathcal{B} = \mathcal{V}(X^B)$, then

$$\mathcal{A} + \mathcal{B} = \mathcal{V}(X^{AB})$$

where X^{AB} is the matrix achieved by summing in all possible ways a column of X^A and a column of X^B. The sum $\mathcal{A} + \mathcal{B}$ is formed by all vectors of the form

$$x = X^A w^A + X^B w^B = \sum_i X_i^A w_i^A + \sum_j X_j^B w_j^B,$$

with

$$\sum_i w_i^A = 1, \quad \sum_i w_j^B = 1, \quad \text{and} \quad w_i^A \geq 0, \quad w_j^B \geq 0.$$

Then

$$x = X^A w^A + X^B w^B = \sum_{ij} X_i^A w_i^A w_j^B + \sum_{ij} X_j^B w_i^A w_j^B =$$

$$= \sum_{ij} w_i^A w_j^B (X_i^A + X_j^B) = X^{AB} w^{AB}$$

where the component of the vector w^{AB} are of the form $w_i^A w_j^B \geq 0$ and $\sum_k w_k^{AB} = 1$. Then the sum $\mathcal{A} + \mathcal{B} \subseteq \mathcal{V}(X^{AB})$, the convex hull of all the points $X_i^A + X_j^B$. On the other hand, all these points belong to the sum because both X_i^A and X_j^B do. Therefore $conv\{\mathcal{A} + \mathcal{B}\} = \mathcal{V}(X^{AB})$. The proof of the statement is completed by keeping in mind that $\mathcal{A} + \mathcal{B}$ is a convex set so that $\mathcal{A} + \mathcal{B} = conv\{\mathcal{A} + \mathcal{B}\}$

The erosion of $\mathcal{A}$ with respect to $\mathcal{B}$ is defined as

$$\tilde{\mathcal{A}}_{\mathcal{B}} = \{x : \ x + b \in \mathcal{A}, \ \text{ for all } b \in \mathcal{B}\}.$$

Consider the plane representation (3.15). Then $x \in \mathcal{A}_{\mathcal{B}}$ if and only if, for all i,

$$F_i(x + b) \leq g_i, \ \text{ for all } \ b \in \mathcal{B},$$

which is equivalent to

$$F_i x \leq g_i - \max_{b \in \mathcal{B}} \ F_i b$$

then

$$\tilde{\mathcal{P}}(F, g)_{\mathcal{B}} = \mathcal{P}(F, \tilde{g})$$

so that, denoting by $\phi_{\mathcal{B}}$ the support functional of $\mathcal{B}$, $\tilde{g}$ is the vector whose components are

$$\tilde{g}_i \doteq g_i - \max_{b \in \mathcal{B}} \ F_i b = g_i - \phi_{\mathcal{B}}(F_i)$$

The convex hull $conv\{\mathcal{A} \bigcup \mathcal{B}\}$ is nothing else than the convex hull of all the vertices of $\mathcal{A}$ and $\mathcal{B}$ (if, for brevity, the case of bounded polyhedra alone is considered). In terms of the representation (3.23) this results in

$$conv\{\mathcal{A} \bigcup \mathcal{B}\} = \mathcal{V}([X_A \ X_B]).$$

Remark 3.29. If $\mathcal{A}$ and $\mathcal{B}$ are 0 symmetric polyhedra, then all the statements of the previous proposition are valid by replacing the word "polyhedra" by "symmetric polyhedra."

Example 3.30. Let us consider the simple network introduced in Example 3.9 Simple computations show that the plane representation of $B\mathcal{U}$ is $\mathcal{P}(M, g)$, where

$$M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & -1 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad g = \begin{bmatrix} u_1^+ \\ u_2^+ \\ -u_1^- \\ -u_2^- \end{bmatrix}$$

Its vertex representation is $\mathcal{V}(R)$ where

$$R = \begin{bmatrix} u_1^- - u_2^- & u_1^- - u_2^+ & u_1^+ - u_2^- & u_1^+ - u_2^+ \\ u_2^- & u_2^+ & u_2^- & u_2^+ \end{bmatrix}$$

The vertices of the set $\mathcal{D}$ are

$$D = \begin{bmatrix} d_1^- & d_1^- & d_1^+ & d_1^+ \\ d_2^- & d_2^+ & d_2^- & d_2^+ \end{bmatrix}$$

The "cooperation set" is given by $B\mathcal{U} + \mathcal{D}$ and it turns out to be $\mathcal{V}(Y)$, where the matrix Y is achieved by summing pair of vertices chosen from $B\mathcal{U}$ and $\mathcal{D}$, respectively, in all possible ways. Therefore the set has, in principle, 16 vertices (although four of them are redundant). For brevity the matrix Y is not reported.

The "competition set" $[B\mathcal{U}]_{\mathcal{D}}$ is achievable by the plane description. Since $\mathcal{D}$ is a square, the computation of the plane representation $\mathcal{P}[M, \tilde{g}] = [\tilde{B\mathcal{U}}]_{\mathcal{D}}$ is quite simple and, more precisely, M is the same matrix reported above and

$$\tilde{g} = \begin{bmatrix} u_1^+ - (d_1^+ - d_2^-) \\ u_2^+ - (d_2^+) \\ -u_1^- + (d_1^- - d_2^-) \\ -u_2^- + (d_2^-) \end{bmatrix}$$

Note that this polyhedron may be empty.

As a final point, the following result, that provides an interesting condition to check the inclusion between polyhedra (see for instance [DH99]), is reported.

Proposition 3.31. *The inclusion*

$$\mathcal{P}[F^{(1)}, g^{(1)}] \subset \mathcal{P}[F^{(2)}, g^{(2)}]$$

holds if and only if there exists a non-negative matrix H such that

$$HF^{(1)} = F^{(2)}$$
$$Hg^{(1)} \leq g^{(2)}$$

Proof. If the inclusion holds then, denoting by $F_i^{(k)}$ the ith row of $F^{(k)}$, we get for all i

$$\mu_i \doteq \max F_i^{(2)} x$$
$$\text{s.t. } F^{(1)} x \leq g^{(1)}$$

is such that

$$\mu_i \leq g_i^{(2)}$$

This is a linear programming problem. Its dual is:

$$\begin{aligned} \mu_i = \min & \, hg^{(1)} \\ \text{s.t. } & hF^{(1)} = F_i^{(2)} \\ & h \geq 0 \end{aligned}$$

Denote by $h^{(i)}$ the non-negative row vector which is a solution of the dual and let H be the square matrix whose ith row is $h^{(i)}$. Then H satisfies the required conditions.

Conversely, assume that the mentioned matrix H exists. Then for all $x \in \mathcal{P}[F^{(1)}, g^{(1)}]$, namely $F^{(1)}x \leq g^{(1)}$, we have

$$F^{(2)}x = HF^{(1)}x \leq Hg^{(1)} \leq g^{(2)}$$

(the first inequality holds because H is non-negative) so $x \in \mathcal{P}[F^{(2)}, g^{(2)}]$.

In the case of polyhedral C-sets, we can assume $g^{(1)} = \bar{1}$ and $g^{(2)} = \bar{1}$, so the inequality becomes

$$H\bar{1} \leq \bar{1}.$$

The previous proposition admits a "dual version".

Proposition 3.32. *If $\mathcal{V}(X^{(1)})$ and $\mathcal{V}(X^{(2)})$ are polyhedral C-sets, then*

$$\mathcal{V}(X^{(1)}) \subseteq \mathcal{V}(X^{(2)})$$

if and only if there exists a matrix $P \geq 0$ such that

$$\begin{aligned} PX^{(2)} &= X^{(1)} \\ \bar{1}^T P &\leq \bar{1}^T \end{aligned}$$

The proof proceeds along the same lines of that of Proposition 3.31.

3.4 Other families of convex sets

Obviously the family of convex sets in $\mathbb{R}^n$ is much wider than that of quadratic and polyhedral ones. The problem in general is that, to be effective in applications, a class of sets must be supported by efficient tools. As we have seen, convex ellipsoids

and polyhedra are associated with quadratic and piecewise linear functions, respectively. A first question is whether there exists a class of functions which includes the family of convex quadratic and piecewise linear functions. The answer is obviously that the class of piecewise quadratic functions [RJ98, BMDP02] includes both polyhedral and quadratic functions (as well as piecewise affine function [Mil02b]). A piecewise quadratic function is a function of the form

$$\Psi(x) = \max_i \ \{x^T P_i x + q_i^T x + r_i\} \tag{3.33}$$

where P_i are positive definite matrices, q_i are vectors and r_i scalars. The single component $x^T P_i x + q_i^T x + r_i$ is convex and therefore Ψ is convex. Clearly any set of the form $\mathcal{N}[\Psi, \kappa]$ is convex and it can represent any ellipsoid or any polyhedron, by an appropriate choice of P_i, q_i and r_i.

A further family of convex sets can be achieved by "smoothing" a polyhedral set. The application of such smoothing property will be presented later. Consider a polyhedral 0-symmetric C-set $\mathcal{S}$ represented by the notation (3.24)

$$\mathcal{S} = \{x : \ \|Fx\|_\infty \leq 1\}.$$

and its associated Minkowski function, denoted by $\Psi(x)$. For p positive integer, define the function

$$\Psi_{2p}(x) \doteq \sqrt[2p]{\sum_{i=1}^{r} (F_i x)^{2p}}$$

where r is the number of rows of F (the even exponent allows to avoid using absolute values inside the root). It turns out that

$$\lim_{p\to\infty} \Psi_{2p}(x) \to \Psi(x),$$

uniformly on every compact set. This is immediate to see since

$$\max_i\{F_i x\} \leq \sqrt[2p]{\sum_{i=1}^{r} (F_i x)^{2p}} \leq \sqrt[2p]{r \ \max_i\{F_i x\}^{2p}} \leq \sqrt[2p]{r} \ \max_i\{F_i x\}$$

and $\sqrt[2p]{r} \to 1$, as $p \to \infty$. Therefore, the unit ball $\mathcal{S}_{2p}$ of $\Psi_{2p}(x)$ converges to $\mathcal{S}$ from inside. The function $\Psi_{2p}(x)$ is smooth everywhere for $x \neq 0$.

This smoothing procedure can be extended to non-symmetric sets. Consider for p integer the function

$$\sigma_p(\xi) = \begin{cases} 0 & \text{if} \quad \xi \leq 0, \\ \xi^p & \text{if} \quad \xi > 0. \end{cases}$$

Then the approximating smoothing function for the Minkowski function of a polyhedral C-set of the form (3.17)

$$\Psi(x) = \max_i \ F_i x$$

is given by (it is not necessary to have even numbers now)

$$\Psi_p(x) \doteq \sqrt[p]{\sum_{i=1}^{r} \ \sigma_p(F_i x)}.$$

Other types of convex sets can be efficiently adopted. In particular those achieved as sublevel surfaces of special positive definite convex functions. An interesting generalization of polyhedral sets and functions is achieved by allowing for complex numbers [BT80]. It is interesting to see that these sets, projected on the real space, can provide real convex sets with (at least partially) smooth surfaces and then are suitable to reduce complexity.

We remind also the composite quadratic functions [HL03], and in particular the max of quadratics,

$$\Psi(x) = \max_i \ x^T P_i x$$

(or their root $\sqrt{\Psi(x)}$) where P_i, $i = 1, 2, \ldots, m$ is a family of positive definite symmetric matrices. These functions are convex and piecewise-smooth and, as we will see, they are an interesting counterpart to the polyhedral functions.

Convex sets and functions are naturally derived as cost-to-go functions of constrained optimal control problems, as we will see later on.

3.5 Star-shaped sets and homogeneous functions

Convex sets are an important family, however convexity can be a too high requirement in many applications. One important class of sets which includes C-sets as a special case is that of the so-called star-shaped sets, defined as follows.

Definition 3.33. A compact and closed set $\mathcal{S}$ including the origin in its interior is star-shaped if any ray originating at 0 encounters its boundary in a single point, namely if for any $z \neq 0$, there exists a value λ_z such that $x = \lambda z \in \mathcal{S}$ if and only if $0 \leq \lambda \leq \lambda_z$.

A star-shaped set can be associated with a positive definite function which is positively homogeneous of order 1 (Fig. 3.11), i.e. such that $\psi(\lambda x) = \lambda \psi(x)$, for any $\lambda \geq 0$.

$$\psi(x) = inf\{\lambda : \ x \in \lambda \mathcal{S}\}$$

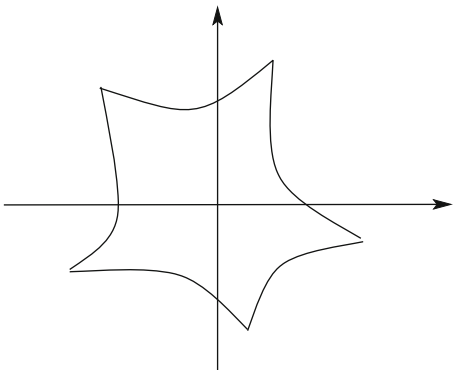


Fig. 3.11 A star-shaped set

Conversely, any positive definite, locally bounded and positively homogeneous function defines a star-shaped set

$$\mathcal{S} = \{x : \ \psi(x) \leq 1\}$$

Note that the degree of homogeneity is not necessarily one. Precisely the function $\psi(x)^p$, $p > 0$ integer, defines the same set

$$\mathcal{S} = \{x : \ \psi^p(x) \leq 1\}$$

This kind of sets and functions turns out to be useful when we are dealing with homogeneous systems, as we will see later.

One important class of star-shaped sets is that associated with homogeneous polynomial [Zel94, CGTV03]. In general these functions are not necessarily convex. For further reading on this topic the reader is referred to [Che11a].

An interesting case of star-shaped function which is not convex is the min type function

$$\Psi(x) = \min_i \ \ x^T P_i x$$

where P_i, $i = 1, 2, \ldots, m$ is a family of positive definite symmetric matrices. While any single $x^T P_i x$ is convex, if we take the minimum, then convexity is lost.

3.6 Exercises

1. Show that a function Ψ is quasi-convex if and only if its sublevel sets $\mathcal{N}[\Psi, \kappa]$ are convex.
2. Show that the sum of two convex functions ψ_1 and ψ_2 is convex. Show by means of a counterexample that the sum of two quasi-convex functions is not quasi-convex. What about $\max\{\psi_1(x), \psi_2(x)\}$ with ψ_1 and ψ_2 quasi-convex?
3. Show that a convex positive definite function is radially unbounded. Is this true in the case of a quasi-convex function?

4. Is the support function of the sum of two C-sets equal to the sum of their support functions?
5. Is the Minkowski function of the sum of two C-sets equal to the sum of their Minkowski functions?
6. Show that the definition of the polar of a convex set (Def. 3.13) is consistent with that of a polar convex cone (centered in 0) provided in Section 3.1.3.
7. Show that the normal cone of a set $\mathcal{N}[\Psi, \kappa]$, having non-empty interior and with Ψ convex, is not necessarily given by the subgradient formula if κ is a (global) minimum of Ψ. (Hint: take a convex differentiable function which is constant in a portion of the domain ...)
8. Show that the sum of two ellipsoids is in general not an ellipsoid. What about the sum of spheres?
9. Show that the erosion of an ellipsoid $\mathcal{E}$ with respect to a convex set $\mathcal{B}$ is not an ellipsoid in general (hint: Take $\mathcal{E}$ a disk and $\mathcal{B}$ a segment centered at the origin).
10. Given an ellipsoid containing the origin in its interior (not necessarily the center) find the expression of its Minkowski function.
11. Show that $\mathcal{P}(F)$ is bounded if and only if the polar $\mathcal{P}(F)^* = \mathcal{V}(F^T)$ includes 0 as an interior point (is the statement achievable by commuting $\mathcal{P}(F)$ and $\mathcal{V}(F^T)$ meaningful?[9])
12. Which is the expression (3.20) for the sum of two polyhedra (not necessarily bounded)?
13. Which is the expression (3.20) for the convex hull of the union of two polyhedra (not necessarily bounded)?
14. Under which condition the set $\mathcal{P}[M, \tilde{g}] = [B\mathcal{U}]_{\mathcal{D}}$ in Example 3.30 is non-empty?
15. The complexity of the representation $\mathcal{P}(F, g)$ of a polyhedron is not increased by the erosion operation. Can you figure out an example in which the complexity is actually reduced?
16. The complexity of the representation of $\mathcal{V}(X^{(1)}) + \mathcal{V}(X^{(2)})$ is the product of their number of vertices of $n_v^{(1)} n_v^{(2)}$. Can you find an example in which the minimal representation of the sum does not preserve this complexity? And another in which it does?
17. Show that if $\Psi(x)$ is a convex function defined on a convex set $\mathcal{P}$, then it reaches it maximum on the boundary, and that, if $\mathcal{P}$ is a polytope, then it achieves the maximum on one of its vertices. What about the minimum?
18. Show pictorially an example of a two-dimensional set $\mathcal{V}(X)$ with redundant vertices. Show that a vertex x_i is redundant if and only if the plane x_i^T is redundant for $\mathcal{P}(X^T, \bar{1})$.
19. Give a set-theoretic interpretation of the set $[B\mathcal{U}]_{\mathcal{D}}$ in example 3.9 if u is a player who wants to meet a certain flow $f = Bu + d$ by choosing $u \in \mathcal{U}$ no matter how its opponent chooses $d \in \mathcal{D}$.

[9] Of course not, why?

Chapter 4
Invariant sets

This chapter contains the basic definitions and results concerning invariant sets in control and it is the core of the book. Indeed, the invariance concept is at the basis of many control schemes that will be considered later. Such a concept naturally arises when dealing with Lyapunov functions, as we have seen, since any Lyapunov function has positively invariant sublevel sets. However, the invariance concept does not require the introduction of the notion of Lyapunov functions and indeed there exist invariant sets that are not obviously related to any Lyapunov function.

4.1 Basic definitions

The idea of positive invariance can be easily understood by referring to a simple autonomous system in state space form:

$$\dot{x}(t) = f(x(t)) \tag{4.1}$$

It is assumed that the above system of equations is defined in a proper open set

$$\mathcal{O} \subseteq \mathbb{R}^n$$

and that there exists a globally defined solution (i.e., for all $t \geq 0$) for every initial condition $x(0) \in \mathcal{O}$. Although the concept has been already considered, positive invariance is formally defined as follows.

Definition 4.1 (Positive invariance). The set $\mathcal{S} \subseteq \mathcal{O}$ is said to be positively invariant w.r.t. (4.1) if every solution of (4.1) with initial condition $x(0) \in \mathcal{S}$ is globally defined and such that $x(t) \in \mathcal{S}$ for $t > 0$.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_4

The above definition is all that is needed when the problem is well-posed, say there is a unique solution corresponding to each given initial condition $x(0) \in \mathcal{S}$. For the sake of generality, it is worth saying that if pathological situations have to be taken into account (say if one wants to consider the case in which the differential equation may have multiple solutions for the same initial condition) then the following weak version of the concept comes into play:

Definition 4.2 (Weak positive invariance). The set $\mathcal{S} \subseteq \mathcal{O}$ is said to be weakly positively invariant w.r.t. (4.1) if, among all the solutions of (4.1) originating in $x(0) \in \mathcal{S}$, there exists at least one globally defined solution which remains inside $\mathcal{S}$, namely $x(t) \in \mathcal{S}$ for $t > 0$.

The reason for the introduction of the (rarely used throughout the book) weak invariance concept is basically that of establishing a link with the abundant mathematical work in this area (see, among the recent literature [AC84, Aub91]). Indeed, from an engineering point of view, the existence of at least **a** solution in $\mathcal{S}$ is not that stunning, since nothing is said about all other possible solutions to the given set of equations. To avoid having the reader dropping the book we guarantee that in the 99.999% of the dynamic systems which will be considered well-posedness will be assumed, namely the existence of a unique solution for any $x(0) \in \mathcal{S}$, a case in which the weak definition collapses to the standard one. The latter can be simply restated as $x(t_1) \in \mathcal{S} \Rightarrow x(t) \in \mathcal{S}$ for $t \geq t_1$. It has to be pointed out that the role of the word "positive" is referred to the fact that the property regards the future. If $x(t_1) \in \mathcal{S}$ implies $x(t) \in \mathcal{S}$, for all t, this property is known as invariance and $\mathcal{S}$ is said to be an invariant set. Invariance is a too special concept to be considered. Therefore in the book we will always refer to positive invariance (although we will sometimes write "invariance" for brevity).

Set invariance plays a fundamental role not just for autonomous systems but also for differential equations of the form

$$\dot{x}(t) = f(x(t), w(t)) \tag{4.2}$$

or controlled differential equations of the form

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), w(t)) \\ y(t) &= g(x(t), w(t)) \end{aligned} \tag{4.3}$$

where $w(t) \in \mathcal{W}$ is an exogenous input and $u(t) \in \mathcal{U}$ is a control input as considered in Chapter 2. To keep (as promised a few lines above) the exposition simple, it will always be assumed that, unless differently specified, (4.2) admits a unique globally defined solution for any $w(\cdot) : \mathbb{R}^+ \mapsto \mathcal{W}$ and all initial conditions $x(0) \in \mathcal{O} \subset \mathbb{R}^n$ and that the control associated with (4.3) is *admissible*, in the sense that it assures the closed-loop system has this property.

The presence of the external input w is required to account for performance specifications as well as for uncertainty entering the system. Thus, to embed the previously introduced invariance concept in the uncertain setting, the following definition is stated:

Definition 4.3 (Robust positive invariance). The set $\mathcal{S} \subseteq \mathcal{X}$ is said to be robustly positively invariant if, for all $x(0) \in \mathcal{S}$ and any $w(t) \in \mathcal{W}$, [1] the condition $x(t) \in \mathcal{S}$ holds for all $t \geq 0$.

To deal with synthesis problems a further definition, that of robust controlled invariance, has to be introduced. It is worth recalling that, for the control Lyapunov functions discussed in Chapter 2, synthesis requires the specification of the class of adopted controllers $\mathcal{C}$ (output-feedback, state feedback, ...). In a similar fashion, being the considered sets $\mathcal{S}$ defined in the plant state space, only static controllers (possibly of a suitably augmented plant) will be considered, since no additive dynamics can be admitted.

Definition 4.4 (Robust controlled positive invariance). The set $\mathcal{S} \subseteq \mathcal{X}$ is said to be robust controlled positively invariant if there exists a control in the class $\mathcal{C}$ (assuring the existence and uniqueness of the solution for the closed-loop system) such that, for all $x(0) \in \mathcal{S}$ and $w(t) \in \mathcal{W}$, the condition $x(t) \in \mathcal{S}$ holds for all $t \geq 0$.

Note that this definition requires the existence of a control such that the problem is well-posed. For instance, assume $\dot{x} = u$ and $\mathcal{S}$ the positive real axis. Then any continuous control function $u = \Phi(x)$ such that $\Phi(0) > 0$ would be in principle suitable since, for $x(0) \geq 0$, $x(t)$ remains positive for $t > 0$. However $u = 1 + x^2$ is not acceptable, because the resulting equation has finite escape time.

The positive invariance notions just introduced are quite useful and several applications will be shown later. In the next section, a fundamental result which characterizes the invariance of a closed set will be presented.

4.2 Nagumo's Theorem

Nagumo's theorem is of a fundamental importance in the characterization of positively invariant sets for continuous-time systems. The best way to state the theorem is to consider the notion of tangent cone. In Section 3.1, a definition of a tangent cone to a set was proposed (see Eq. (3.10)), which is valid for convex sets only. To state the theorem in its general form, a new definition of tangent cone to a set due to Bouligand [Bou32], equivalent to the previous one in the case of convex sets, is introduced. To this aim, the notion of distance has to be defined first, as per the next definition:

[1] Formally $w : \mathbb{R}^+ \to \mathcal{W}$.

Definition 4.5 (Distance from a set). Given a set $\mathcal{S} \subset \mathbb{R}^n$ and a point $y \in \mathbb{R}^n$, the distance is defined as:

$$dist(y, \mathcal{S}) = \inf_{w \in \mathcal{S}} \|y - w\|_*$$

where $\|\cdot\|_*$ is any relevant norm.

With this in mind, the "Bouligand" definition of the tangent cone to a closed set is as follows [Bou32].

Definition 4.6 (Bouligand's tangent cone). Given a closed set $\mathcal{S}$, the tangent cone to $\mathcal{S}$ at x is defined as follows:

$$\mathcal{T}_{\mathcal{S}}(x) = \left\{ z : \liminf_{\tau \to 0} \frac{dist(x + \tau z, \mathcal{S})}{\tau} = 0 \right\} \tag{4.4}$$

It is worth stressing that, although the distance function depends on the chosen norm, the set $\mathcal{T}_{\mathcal{S}}(x)$ does not. It has to be pointed out that there exist other definitions of a tangent cone, due to Bony [Bon69] and Clarke [Cla83], which lead to similar results. The three definitions are equivalent in the case of convex sets.

Also, it is easy to see that if $\mathcal{S}$ is convex, so is $\mathcal{T}_{\mathcal{S}}(x)$, and "$\liminf$" can be replaced by "$\lim$" in (4.4). Furthermore if $x \in int\{\mathcal{S}\}$, then $\mathcal{T}_{\mathcal{S}}(x) = \mathbb{R}^n$, whereas if $x \notin \mathcal{S}$, then $\mathcal{T}_{\mathcal{S}}(x) = \emptyset$ (remember that $\mathcal{S}$ is closed). Thus the tangent cone $\mathcal{T}_{\mathcal{S}}(x)$ is non-trivial only on the boundary of $\mathcal{S}$.

We are now able to state one basic result concerning positive invariance. This theorem was introduced for the first time in [Nag42] and it was reconsidered later in different formulations (see, for instance, [Bre70, Gar80, Yor67]). Here, the standard version in terms of tangent cone (see also [Aub91, AC84, Cla83, FH76a] for details) is presented.

Theorem 4.7 (Nagumo 1942 [Nag42]). *Consider the system $\dot{x}(t) = f(x(t))$ and assume that for each initial condition $x(0)$ in an open set $\mathcal{O}$ it admits a (not necessarily unique) solution defined for all $t \geq 0$. Let $\mathcal{S} \subset \mathcal{O}$ be a closed set. Then, $\mathcal{S}$ is weakly positively invariant for the system if and only if the velocity vector satisfies Nagumo's condition:*

$$f(x) \in \mathcal{T}_{\mathcal{S}}(x), \quad \textit{for all } x \in \mathcal{S}. \tag{4.5}$$

As expected, weak (positive) invariance turns into positive invariance if uniqueness of the solution is assumed, thus getting the following corollary.

Corollary 4.8. *If all the assumptions and conditions of Theorem 4.7 are verified, under the "more strict" assumption of uniqueness of the solution for every $x(0) \in \mathcal{O}$, then the positive invariance of $\mathcal{S}$ is equivalent to condition (4.5).*

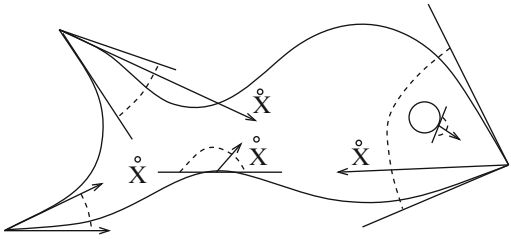


Fig. 4.1 Nagumo's conditions applied to a fish shaped set

Nagumo's condition (4.5), also known as *sub-tangentiality condition*, is meaningful only for $x \in \partial\mathcal{S}$, since for $x \in int\{\mathcal{S}\}$, $\mathcal{T}_{\mathcal{S}}(x) = \mathbb{R}^n$. Thus the condition (4.5) can be replaced by

$$f(x) \in \mathcal{T}_{\mathcal{S}}(x), \quad \text{for all } \ x \in \partial\mathcal{S}.$$

The theorem has a simple geometric interpretation (see Fig. 4.1). Indeed, in plain words, it says that if the velocity vector $\dot{x} = f(x),\ x \in \partial\mathcal{S}$ "points inside or it is tangent to $\mathcal{S}$," then the trajectory $x(t)$ remains in $\mathcal{S}$. It is worth pointing out that Nagumo's condition without the assumption of the uniqueness of the solution is not sufficient to assure positive invariance. Consider, for instance, the set $\mathcal{S} = \{0\}$ (i.e., the set including the origin only). Its tangent cone for $x = 0$ is $\{0\}$. The equation $\dot{x}(t) = \sqrt{x(t)}$ does fulfil the requirements of the theorem. However, for $x(0) = 0$, only the zero solution $x(t) = 0,\ t \geq 0$ remains inside $\mathcal{S}$. In fact, there are infinitely many non-zero solutions escaping from $\mathcal{S}$, each of them being of the form

$$x(t) = \begin{cases} 0 \ \text{ for } \ t \leq t_0, \\ (t - t_0)^2/4, \ \text{ for } \ t > t_0, \end{cases}$$

where t_0 is any non-negative real number. There are several proofs of Nagumo's theorem. Perhaps the easiest one is that provided by Hartman [Har72] (who was not aware of the previous result by Nagumo). Here, a simple proof in the very special case of what is called a "practical set" is provided. This proof is inspired by the one proposed in the book [Kra68].

Definition 4.9 (Practical set). Let $\mathcal{O}$ be an open set. The set $\mathcal{S} \subset \mathcal{O}$ is said to be a practical set if

1. it is defined by a finite set of inequalities of the form

$$\mathcal{S} = \{x : \ g_k(x) \leq 0, \ \ k = 1, 2, \ldots, r\}$$

or in the equivalent form

$$\mathcal{S} = \{x : \ \hat{g}(x) = \max_{k=1,2,\ldots,r} g_i(x) \leq 0\}$$

where $g_i(x)$ are continuously differentiable functions defined on $\mathcal{O}$;

2. for all $x \in \mathcal{S}$ there exists z such that

$$g_i(x) + \nabla g_i(x)^T z < 0, \quad \text{for all } i;$$

3. there exists a Lipschitz continuous vector field $\phi(x)$ such that for all $x \in \partial\mathcal{S}(x)$

$$\nabla g_i(x)^T \phi(x) < 0$$

Practical sets form a large class which is of significance in engineering. The first two assumptions are purely technical, and basically concern the regularity. In particular, the second one essentially corresponds to the constraint qualification conditions as reported in [Lue69], Sect. 9.4. Such an assumption implies that the interior of the set is given by

$$int\{\mathcal{S}\} = \{x : \ g_k(x) < 0, \ \ k = 1, 2, \ldots, r\}$$

Under these assumptions, denote by $Act(x)$ the set of active constraints

$$Act(x) = \{i : \ g_i(x) = 0\}$$

Note that $Act(x)$ is non-empty only on the boundary in view of the second assumption. It turns out that for all $x \in \partial\mathcal{S}$ the tangent cone is given by

$$\mathcal{T}_{\mathcal{S}}(x) = \{z : \ \nabla g_i(x)^T z \leq 0, \ \ \text{for all } i \in Act(x)\} \tag{4.6}$$

(the expression is valid in the interior where $Act(x)$ is empty, if one admits that any $z \in \mathbb{R}^n$ is a solution of the resulting empty set of inequalities). Similar assumptions have been considered by [FZ87].

The third assumption is strong but "reasonable" and it basically requires the existence of a regular vector field that from each point of the boundary "points inside." In particular, for convex sets with non-empty interior, one of such vector fields ϕ is

$$\phi(x) = \bar{x} - x$$

where $\bar{x} \in int\ \{\mathcal{S}\}$. The same vector field can also be associated with star-shaped sets, namely sets having an interior point $\bar{x}$ such that any ray originating in $\bar{x}$ encounters the boundary of $\mathcal{S}$ in at most one point, provided that $\bar{x} - x$ is in the interior of the tangent cone for all $x \in \partial\mathcal{S}$ (mind that this amounts to say that the interior of the tangent cone is non-empty for any $x \in \partial\mathcal{S}$). For instance, the set reported in Figure 4.1 (assuming that a finite set of inequalities describing the "fish" boundaries exists) is practical, while pictorial examples of "non-practical" sets are reported in Figure 4.2 (still Nagumo's theorem applies to these sets).

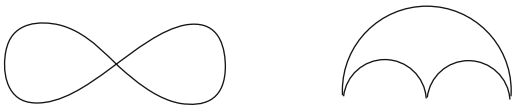


Fig. 4.2 Examples of "non-practical" sets

4.2.1 Proof of Nagumo's Theorem for practical sets and regular f

To provide a reasonably simple proof we assume that $\mathcal{S} \subset \mathcal{O}$ is a practical set and that the function f is continuous and that the system $\dot{x} = f(x)$ admits a unique solution for any initial state in $\mathcal{O}$.

Proof (Sufficiency of the condition). To prove the theorem, we consider the auxiliary system

$$\dot{x}(t) = f_\epsilon(x(t)) = f(x(t)) + \epsilon\phi(x(t))$$

with $\epsilon > 0$. Let us denote by $x_\epsilon(t)$ its solution. As a first step, we show that $x_\epsilon(t) \in \mathcal{S}$ for all $t \geq 0$. Since $\mathcal{S}$ is closed, we need to prove that any solution $x_\epsilon(t)$ originating from a point x^0, $x_\epsilon(0) = x^0 \in \partial\mathcal{S}$, i.e. on the boundary, remains inside $\mathcal{S}$ in a proper interval $[0, \tau]$, with $\tau > 0$. This is immediate because if $x_0 \in \partial\mathcal{S}$ then $\hat{g}(x) = 0$ and, by the assumption, we have that the derivative of $g_i(x_\epsilon(t))$ at $t = 0$ is

$$\dot{g}_i(x_\epsilon(0)) = \nabla g_i(x^0)^T f_\epsilon(x^0) = \underbrace{\nabla g_i(x^0)^T f(x^0)}_{\leq 0} + \epsilon \underbrace{\nabla g_i(x^0)^T \phi(x^0)}_{<0} < 0,$$

for all $i \in Act(x^0)$. Then in a right neighborhood of 0 we have $g_i(x_\epsilon(t)) < 0$ for $i \in Act(x^0)$ and, by continuity, $g_i(x_\epsilon(t)) < 0$ for $i \notin Act(x^0)$, then $g_i(x_\epsilon(t)) < 0$ for all i. Then, for $x_\epsilon(0) \in \partial\mathcal{S}$ we have that $x_\epsilon(t) \in \mathcal{S}$ in an interval $[0, \tau]$, $\tau > 0$. Since the system is time-invariant, the same situation holds for any initial time t_0, therefore $x_\epsilon(t)$ cannot cross the boundary at any time, so that for all $\epsilon > 0$ $x_\epsilon(t) \in \mathcal{S}$ if $x_\epsilon(0) \in \mathcal{S}$. So the first step of the proof is completed having shown that $\mathcal{S}$ is positively invariant for the modified system.

Let us now consider the original system and its solution $x(t)$ with $x(0) = x^0 \in \mathcal{S}$. For $\epsilon = 1/k$ consider the corresponding sequence of solutions $x_{1/k}(t)$ of the auxiliary system. As $k \to \infty$, $x_{1/k}(t) \to x(t)$ converges uniformly on each interval $[0, \tau]$. This implies that for each $0 \leq t \leq \tau$, $x(t) \in \mathcal{S}$, because if we assume $x(t) = x^1 \notin \mathcal{S}$ we arrive to a contradiction. Indeed, since $\mathcal{S}$ is closed there would be a neighborhood $\mathcal{B}_\delta$ of radius δ and center x^1 such that $\mathcal{B}_\delta \bigcap \mathcal{S} = \emptyset$. But in view of the mentioned convergence, for k large enough, we would have both $x_{1/k}(t) \in \mathcal{B}_\delta$ and, as shown before, $x_{1/k}(t) \in \mathcal{S}$: a contradiction. The fact that for $x(0) \in \mathcal{S}$ we have $x(t) \in \mathcal{S}$ in the interval $[0, \tau]$ implies also that the inclusion holds for all $t > 0$.

Necessity of the condition Conversely let us assume that for some $x^0 \in \mathcal{S}$ we have $f(x^0) \notin \mathcal{T}_{\mathcal{S}}(x^0)$. Since for $x^0 \in int\{\mathcal{S}\}$ we would have from (4.6) that $\mathcal{T}_{\mathcal{S}}(x^0) = \mathbb{R}^n$, this implies that $x^0 \in \partial\mathcal{S}$. Let $i \in Act(x^0)$ such that $g_i(x^0) = 0$ and $\nabla g_i(x^0)^T f(x^0) > 0$. Then we have that the derivative of the function $g_i(x(t))$ if $x(t) = x^0$ is given by

$$\frac{d}{dt} g_i(x(t))\Big|_{x(t)=x^0} = \nabla g_i(x^0)^T f(x^0) > 0$$

and therefore, in a right neighborhood of τ, $\tau \in [t, t+h]$, we have $g_i(x(\tau)) > 0$, so that $x(\tau) \notin \mathcal{S}$.

Note that the results presented in Section 2.5 concerning Lyapunov functions can be regarded as special cases of Nagumo's Theorem. Indeed if the set $\mathcal{S}$ is defined by a differentiable Lyapunov function g as $\mathcal{S} = \mathcal{N}[g, 0]$, then the derivative condition $\dot{g}(x)\nabla g(x)^T f(x) \leq 0$ is nothing but a special case of Nagumo's condition. The contribution of Nagumo's theorem is twofold. First, it considers a much more general class of sets than those described by a single inequality, say of the form $\mathcal{S} = \mathcal{N}[g, 0]$. Second, it provides a condition which has to be satisfied *on set boundary only*.

4.2.2 Generalizations of Nagumo's theorem

There are natural generalizations of Nagumo's theorem to non-autonomous systems, reported next.

Let us first consider the case of a simple time-varying system

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

We wish to establish the invariance of a closed set.

First the concept of positive invariance has to be re-defined since the system is not time-invariant

Definition 4.10. The closed set $\mathcal{S}$ is positively invariant if for all t_0 the condition $x(t_0) \in \mathcal{S}$ implies $x(t) \in \mathcal{S}$ for $t \geq t_0$.

Definition 4.10 is not equivalent to Definition 4.1.

Investigating the positive invariance of a set $\mathcal{S}$ is possible by adding a fictitious equation $\dot{\theta}(t) = 1$, so the system becomes

$$\begin{aligned} \dot{x}(t) &= f(x(t), \theta(t)) \\ \dot{\theta}(t) &= 1, \end{aligned}$$

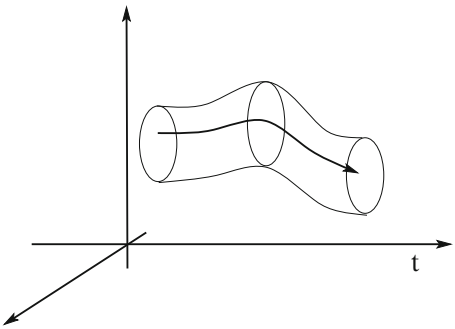


Fig. 4.3 The tube

which is autonomous with the new variable θ. Then Nagumo's theorem can be applied. Note that the set $\mathcal{S}$ has changed its nature, since in the new state-space, it is a cylinder. In particular it is not compact, only closed. We can further generalize the theory by taking time-varying sets. Take a set of the form

$$\mathcal{S}(t) = \{x : \ \phi(x,t) \leq 0\}$$

where $\phi(x,t)$ is smooth. Assume that at any t

$$\frac{\partial \phi(x,t)}{\partial x} \neq 0, \quad \text{for } \ \phi(x,t) = 0.$$

Under this "boundary-regularity" assumption, we have that $\phi(x,t) = 0$ is the equation of the boundary of the set at time t (Fig. 4.3). The problem of remaining in such a set if we are inside is often called "state in a tube."

The question whether the condition $x(\tau) \in \mathcal{S}(\tau)$ implies $x(t) \in \mathcal{S}(t)$ for $t \geq \tau$ naturally leads to the condition

$$\nabla \phi(x,t)\dot{x} + \frac{\partial \phi(x,t)}{\partial t}\dot{\theta} = \nabla \phi(x,t) f(x(t),t) + \frac{\partial \phi(x,t)}{\partial t} 1 \leq 0$$

at the boundary. This type of expression is well known in the context of time-varying Lyapunov functions.

We are especially interested in the robust version of Nagumo's theorem which is reported next. The proofs can be inferred from that previously given in the case of a practical set and the previous augmentation by means of the equation $\dot{\theta}(t) = 1$. The reader is referred to [AC84, Aub91].

Theorem 4.11. *Consider the dynamic system $\dot{x}(t) = f(x(t), w(t))$, $w(t) \in \mathcal{W}$, with $\mathcal{W}$ a compact set and f continuous. Assume that, for each initial condition $x(0)$ in an open set $\mathcal{X}$ and each piecewise continuous function $w(t)$, such system admits a unique and globally defined (i.e., for all $t \geq 0$) solution. Let $\mathcal{S} \subseteq \mathcal{X}$ be a closed set. Then $\mathcal{S}$ is robustly positively invariant for such a system if and only if for all $x \in \mathcal{S}$*

$$f(x,w) \in \mathcal{T}_{\mathcal{S}}(x), \quad \textit{for all} \ \ w \in \mathcal{W}. \tag{4.7}$$

The continuity of f in Theorem 4.11 is fundamental. If it is not verified, then the meaning of solution of a system of differential equations has to be defined.

Example 4.12. Consider the system

$$\dot{x} = -\text{sgn}[x] + w, \quad |w| \leq 1/2$$

The solution of this system may be defined by absorbing the system in a differential inclusions (an exhaustive book on differential inclusion is [AC84], see Chapter 2 for details). Intuitively it can be argued that, as long as $x(t) \neq 0$, the solution of this system is

$$x(t) = x(0) - \text{sgn}(x(0))\, t + \int_0^t w(\sigma)\, d\sigma,$$

which has the property $|x(t)| \leq \max\{0, |x(0)| - 1/2t\}$, and therefore converges in finite time to 0 for all the specified $w(\cdot)$. Once 0 is reached the solution remains null, thus the set $\{0\}$ is necessarily robustly positively invariant. This claim can be easily proved in view of the fact that any interval of the form $[-\epsilon, \epsilon]$ is positively invariant (note that no discontinuity of f appears on the extrema). However, Nagumo conditions are not satisfied for the set $\{0\}$. Therefore Nagumo's result holds as long as it is possible to define a solution in the regular sense.

In the case of a controlled system $\dot{x}(t) = f(x(t), u(t), w(t))$, $w(t) \in \mathcal{W}$, with f continuous, $\mathcal{S}$ is said to be robustly controlled invariant if there exists a continuous control function $\Phi(x)$ (or a control in a proper class $\mathcal{C}$) assuring the existence of a unique and globally defined solution for all $x(0) \in \mathcal{X}$ and such that, for all $x \in \mathcal{S}$,

$$f(x, \Phi(x), w) \in \mathcal{T}_{\mathcal{S}}(x), \quad \text{for all } \; w \in \mathcal{W}. \tag{4.8}$$

Note that it is *assumed* that the control Φ is such that the corresponding closed-loop system admits a solution. In general, assuring such a condition can be a real headache for an (even academic) engineer. In principle, condition (4.8) could be formulated in a pointwise sense as follows: "for each x there must exist a control value u such that $f(x, u, w) \in \mathcal{T}_{\mathcal{S}}(x)$." If formulated as above, the problem then becomes that of verifying the existence of a sufficiently regular feedback function $u = \Phi(x)$ (or $u = \Phi(x, w)$) which guarantees the global existence of a unique solution. To explain the problem, let us define the set of admissible pointwise control values

$$\Omega(x) = \{u : f(x, u, w) \in \mathcal{T}_{\mathcal{S}}(x), \quad \text{for all } \; w \in \mathcal{W}\}. \tag{4.9}$$

Then one should find a sufficiently regular function Φ such that

$$\Phi(x) \in \Omega(x)$$

As already mentioned in Chapter 2, this is a selection problem and theorems analogous to Theorem 2.30 apply (under appropriate conditions). The reader is referred to specialized literature [AC84, Aub91].

One of the basic properties of the class of positively invariant sets is that they are closed w.r.t. the union and intersection operations, as per the following fundamental proposition.

Proposition 4.13. *Let $\mathcal{S}$ and $\mathcal{P}$ be closed positively invariant subsets of an open set $\mathcal{O}$ for $\dot{x}(t) = f(x(t), w(t))$, $w(t) \in \mathcal{W}$. Then*

- *$\mathcal{S} \bigcap \mathcal{P}$ is positively invariant.*
- *$\mathcal{S} \bigcup \mathcal{P}$ is positively invariant.*

If $\mathcal{S}$ and $\mathcal{P}$ are controlled-invariant for $\dot{x}(t) = f(x(t), u(t), w(t))$, $w(t) \in \mathcal{W}$, then

- *$\mathcal{S} \bigcup \mathcal{P}$ is controlled invariant.*

The proofs of the above assertions are trivial and thus omitted. It is easy to see that the intersection of controlled-invariant set is not controlled-invariant, in general.

4.2.3 Examples of application of Nagumo's Theorem

As a simple example of application of Nagumo's theorem, a qualitative analysis of a dynamical competition model is now presented.

Example 4.14 (A competition model). Consider a system of the form

$$\begin{aligned}
\dot{x}_1(t) &= x_1(t)\,[1 - x_1(t) - \alpha x_2(t) - \beta x_3(t)] \\
\dot{x}_2(t) &= x_2(t)\,[1 - \beta x_1(t) - x_2(t) - \alpha x_3(t)] \\
\dot{x}_3(t) &= x_3(t)\,[1 - \alpha x_1(t) - \beta x_2(t) - x_3(t)]
\end{aligned}$$

where α and β are positive constants. This model describes the competition among three populations, sharing the same environment. The variable x_i represents the total number of individuals of population i, while $\alpha > 0$ and $\beta > 0$ are positive parameters [Bel97].

The evolution of this system clearly depends on the specific parameters α and β and can be computed (approximately) via numerical integration. However, as a preliminary qualitative analysis of these equations, we can try to answer the following questions.

- The variables of this system have physical meaning as long as they are non-negative. Are these equations consistent in this sense, namely, is the condition $x(t) \geq 0$ for all $t > 0$ assured for $x(0) \geq 0$?
- Is the solution of this system bounded for any given $x(0) \geq 0$?

Only a positive answer to these questions may lead to the conclusion that the model is realistic. In particular, the second property is important if it is assumed that the environment has limited resource so that none of the populations can diverge.

We can provide such a positive answer by proving that the set

$$\mathcal{S} = \{x: \ \|x\|_2 \leq \rho, \ \ x \geq 0\}$$

is positively invariant for ρ large enough.

To check the positive invariance one has to consider the expression of the cone at the boundary. Assume that x is on the positive orthant boundary. The tangent cone is the set

$$\mathcal{T}_{\mathcal{S}}(x) = \{z: \ z_i \geq 0, \ \text{ for all } i \ \text{ such that } \ x_i = 0\}$$

It is immediate to see that $f(x) \in \mathcal{T}_{\mathcal{S}}(x)$ because $\dot{x}_i = 0$ whenever $x_i = 0$.

Let us now consider the part of the boundary which is on the sphere of radius ρ. If x is in the interior of the positive orthant, the tangent cone is the plane $\mathcal{T}_{\mathcal{S}}(x) = \{z: \ x^T z \leq 0\}$ tangent to the sphere. Consider the smallest positive constant γ such that $\|x\|_3 \geq \gamma \|x\|_2$ for all x, which turns out to be $\gamma = 1/\sqrt[6]{3}$. Bearing in mind that in the interior of the orthant $x_i > 0$, we derive

$$\begin{aligned} x^T \dot{x} &= x_1^2 + x_2^2 + x_3^2 - [x_1^3 + x_2^3 + x_3^3] + \\ &\quad - \alpha[x_1^2 x_2 + x_2^2 x_3 + x_3^2 x_1] - \beta[x_1^2 x_3 + x_2^2 x_1 + x_3^2 x_2] \\ &\leq x_1^2 + x_2^2 + x_3^2 - [x_1^3 + x_2^3 + x_3^3] = \\ &= \|x\|_2^2 - \|x\|_3^3 \leq (\|x\|_2^2 - \gamma^3 \|x\|_2^3) = (\rho^2 - \gamma^3 \rho^3) \leq 0 \end{aligned}$$

provided that

$$\rho \geq 1/\gamma^3 = \sqrt{3}$$

As for the remaining points on the boundary (i.e., those which are both on the orthant boundary and the sphere boundary), the tangent cone is given by

$$\mathcal{T}_{\mathcal{S}}(x) = \{z \in \mathbb{R}^n: \ \ z_i \geq 0, \ \text{ for all } i \ \text{ s.t. } \ x_i = 0, \ \text{ and } x^T z \leq 0\}$$

By means of the same considerations above it can be seen that even for such points the condition $\dot{x} \in \mathcal{T}_{\mathcal{S}}(x)$ is satisfied.

The analysis we carried out so far is strictly preliminary and there are several additional properties that can be shown. For instance, there exists a single nontrivial[2] equilibrium point given by

[2]Namely such that all the populations are non-zero.

$$\bar{x} = \frac{1}{1+\alpha+\beta}\begin{bmatrix}1\\1\\1\end{bmatrix},$$

which is included in $\mathcal{S}$ provided that the radius $\rho > 0$ is large enough. This can be shown by proving that the central ray of the orthant

$$\mathcal{R}_c = \{x = [\,1\;1\;1\,]^T\lambda, \quad \lambda \geq 0\}$$

is positively invariant and that for any non-zero initial condition taken on this ray the solution converges to the equilibrium point (throughout the ray). Note that this is coherent with our results because the norm of the equilibrium point is $\|\bar{x}\|_2 \leq \sqrt{3} \leq \rho$, since α and β are both positive.

If the dynamical model is "reduced" by one dimension, for instance by setting $x_3 = 0$ (i.e., after the extinction of the third population), the theory of invariant sets can provide a deeper insight on the situation since a graphical representation is possible (see [Bel97] and Exercise 5). The case in which $x_3 = 0$ represents a special evolution of our model because the subset of $\mathcal{S}$ of all points for which $x_i = 0$ is positively invariant, which means that no population can recover spontaneously after extinction (as nature sadly imposes). In the one-dimensional case, i.e. $x_2 = x_3 = 0$, the well-known equation $\dot{x}_1 = x_1[1 - x_1]$, present in most basic books on dynamical systems, is recovered. For further details on the possible funny behaviors of this kind of systems, the reader is referred to [Bel97].

4.2.4 Contractive Sets

Consider now a C-set $\mathcal{S} \subset \mathbb{R}^n$ and its Minkowski function $\Psi_{\mathcal{S}}(x)$. An equivalent statement of positive invariance sounds as follows: If $\Psi_{\mathcal{S}}(x(0)) \leq 1$, then $\Psi_{\mathcal{S}}(x(t)) \leq 1$ for all $t > 0$. In particular, we must have that $\Psi_{\mathcal{S}}(x(t))$ is *non-increasing on the boundary*. It is possible to strengthen this notion by imposing a certain "speed of entrance" which will be useful later to assure speed of convergence to the system.

Definition 4.15 (Contractive set, continuous-time). The C-set $\mathcal{S}$ is contractive for the system

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

$w \in \mathcal{W}$, $u \in \mathcal{U}$ if and only if there exists a Lipschitz control function $\bar{u}(x) \in \mathcal{U}$, defined for $x \in \partial\mathcal{S}$, such that for each point on the boundary $x \in \partial\mathcal{S}$ the following condition holds

$$D^+\Psi_{\mathcal{S}}(x, f(x, \bar{u}(x), w)) \leq -\beta, \quad \text{for all } w \in \mathcal{W}, \tag{4.10}$$

for some $\beta > 0$. In this case $\mathcal{S}$ is also said to be β-contractive.

Remark 4.16. If we allowed for $\beta = 0$ in the definition we would have controlled-invariance. We also stress that the definition of contractivity can be generalized to the much more general class of star-shaped closed sets see Definition 3.33.

The fact that the function $\bar{u}(x)$ defined on the boundary is Lipschitz, namely that there exists $L > 0$ such that $\|\bar{u}(x_1) - \bar{u}(x_2)\| \leq L\|x_1 - x_2\|$, is assumed for the easier exposition, but it is not strictly necessary. Clearly, contractivity is a stronger property than positive invariance. It implies that, when x is on the boundary, not only the trajectory remains in the set but also that it "can be pushed inside" with a guaranteed boundary-crossing speed. The consequence for linear (actually homogeneous) systems is fundamental because, by scaling the boundary, we can see that this crossing speed implies a "global convergence speed to the set," and even a convergence speed to the origin when no additive disturbances are present.

4.2.5 *Discrete-time systems*

Consider the discrete-time system

$$x(t+1) = f(x(t))$$

and a set $\mathcal{S}$. The definition of positive invariance of $\mathcal{S}$ holds without changes, even in the more general version when systems of the form $x(t+1) = f(x(t), w(t))$, $w \in \mathcal{W}$ are considered.

As it can be easily understood, there is no evident extension of Nagumo's "boundary-type" conditions for discrete-time systems. Intuitively the natural counterpart of the Nagumo's condition $\dot{x} \in \mathcal{T}_{\mathcal{S}}(x)$, namely the derivative on the boundary "pointing inside," would be

$$f(x) \in \mathcal{S} \quad \text{for all } x \in \partial\mathcal{S}$$

which means, roughly, the state on the boundary "jumps inside." However, this condition is not sufficient to assure $f(x) \in \mathcal{S}$ for all $x \in \mathcal{S}$. Indeed, it is easy to provide discrete-time examples in which the above boundary condition can be satisfied, yet the set is not positively invariant [Bla99].

Therefore the only reasonable "discrete-time extension" of Nagumo's theorem is the next tautology: $\mathcal{S}$ is positively invariant if and only if

$$f(\mathcal{S}) \subseteq \mathcal{S}$$

Luckily enough, the situation is completely different if we restrict our attention to the class of homogeneous systems (including the linear ones)[3] and the sets under considerations are convex C-sets $\mathcal{S}$.

Definition 4.17 (Positively homogeneous system). The system $x(t+1) = f(x(t), u(t), w(t))$ is said to be positively homogeneous of order $p > 0$ if

$$f(\lambda x, \lambda u, w) = \lambda^p f(x, u, w) \quad \text{for all} \quad w \in \mathcal{W}$$

The following proposition holds true.

Theorem 4.18. *The C-set $\mathcal{S}$ is controlled invariant for a positively homogeneous system $x(t+1) = f(x(t), u(t), w(t))$ if and only if for all $x \in \partial\mathcal{S}$ there exists a control $u(x)$ (or $u(x, w)$) such that*

$$f(x, u, w) \in \mathcal{S} \text{ for all } w \in \mathcal{W}$$

Proof. Only the case in which the control law does not depend on w is considered. The proof of the case in which $u = u(x, w)$ is an easy extension. Consider any point $x \in \mathcal{S}$ and the ray originating from the origin and passing through x, and let $\bar{x}$ be the unique intersection of this ray with the boundary of the set $\partial\mathcal{S}$. In terms of Minkowski functional, it is possible to write $x = \Psi_{\mathcal{S}}(x)\bar{x} = \lambda\bar{x}$ with $\lambda = \Psi(x) \leq 1$. Then, consider the control $u(x) = \Psi_{\mathcal{S}}(x)\bar{u}(\bar{x})$, where $u(\bar{x})$ is the control that drives $\bar{x}$ inside $\mathcal{S}$. In view of the positively homogeneity assumptions one gets:

$$f(x, u, w) = f(\lambda\bar{x}, \lambda\bar{u}, w) = \lambda^p f(\bar{x}, \bar{u}, w) \in \mathcal{S} \text{ for all } w \in \mathcal{W}.$$

Note that the theorem is valid even if the input values are constrained as $u \in \mathcal{U}$, with $\mathcal{U}$ a C-set. Furthermore, the theorem holds true for other classes of homogeneous systems, for instance those which are affected by an exogenous input $d \in \mathcal{D}$ with $\mathcal{D}$ a C-set. One of such cases is

$$x(t+1) = A(w)x(t) + B(w)u(t) + Ed(t)$$

(although the homogeneous control used in the proof might be not appropriate). The theorem provides, as a special case, the positive-invariance condition for uncontrolled systems. It is also worth mentioning that the distinction between positive invariance and weak-positive invariance has no reason to exist (not even as a mathematical fantasy) since there are no questions concerning the uniqueness of the solution in the case of difference equations.

[3] If the reader is scared by this idle generality, we let her/him know immediately that the subclass of relevance, that of linear uncertain systems $x(t+1) = A(w)x(t) + B(w)u(t)$, is the only one which will be actually considered.

As already done in the continuous-time case, it is possible to strengthen the notion of invariance by introducing a "speed of entrance."

Definition 4.19 (Contractive set, discrete-time). The C-set $\mathcal{S}$ is contractive for the system

$$x(t+1) = f(x(t), u(t), w(t)),$$

where $w \in \mathcal{W}$, $u \in \mathcal{U}$, if and only if there exists a control function $u(x) \in \mathcal{U}$ such that, for every $x \in \mathcal{S}$, the following condition holds

$$\Psi_{\mathcal{S}}(f(x,u,w)) \leq \lambda, \quad \text{for all } \ w \in \mathcal{W}$$

where $\Psi_{\mathcal{S}}(x)$ is the Minkowski function of $\mathcal{S}$, for some $0 \leq \lambda < 1$. In this case the set $\mathcal{S}$ is said to be λ-contractive.

Note that the control action, differently from the continuous-time case, has to be defined on the whole $\mathcal{S}$ (say, not just on the boundary), for the same already analyzed reasons concerning the non-existence of boundary-type conditions for discrete-time systems. Finally we stress that, the case in which u is a function of w can be easily dealt with by considering $u = u(x,w)$, and in this case the set $\mathcal{S}$ is referred to as gain-scheduling contractive, with obvious meaning of the term.

4.2.6 Positive invariance and fixed point theorem

There is an interesting connection between positive invariance and the fixed point theorem we wish to analyze. For the simple exposition, we limit our attention to convex and compact sets but the results can be further generalized. Let us first consider a discrete-time system of the form

$$x(t+1) = f(x(t)),$$

where f is continuous and defined on a compact and convex set $\mathcal{S}$. Then the next theorem is a re-statement of the well-known fixed-point-theorem.

Theorem 4.20 (Brouwer fixed-point theorem). *If the convex and compact set $\mathcal{S}$ is positively invariant for the continuous map $f : \mathcal{S} \to \mathcal{S}$, then it necessarily includes a stationary point for the system. More precisely, there exists $\bar{x} \in \mathcal{S}$ such that*

$$f(\bar{x}) = \bar{x}$$

The previous theorem is a milestone in mathematics and no proof is provided here. Rather, we point out that convexity[4], boundedness and closedness are crucial assumptions. It is easy to see that, if any of these assumption is dropped, then the remaining two are not sufficient for the claim (see Exercise 3).

The previous property can be, noteworthy, extended to continuous-time systems, as per the next theorem.

Theorem 4.21. *Consider a continuous-time system of the form*

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

with f locally Lipschitz and defined on a compact and convex set $\mathcal{S}$ which is positively invariant. Then $\mathcal{S}$ includes at least one stationary point. More precisely, there exists $\bar{x} \in \mathcal{S}$ such that

$$f(\bar{x}) = 0$$

The proof of the theorem is based on the following Lemma[5].

Lemma 4.22. *If $v(x)$ is a continuous vector field defined on a compact convex set $\mathcal{S}$, then one of the following conditions holds:*

- *$v(x)$ vanishes in some point $x \in \mathcal{S}$;*
- *there is a point $\tilde{x}$ on the boundary in which Nagumo's conditions are violated, precisely $v(x) \notin \mathcal{T}_{\mathcal{S}}(x)$.*

Proof. Consider the minimum-distance function $\mu(x)$ which associates with x the point $y \in \mathcal{S}$ of minimum distance (which is unique if we choose the Euclidean norm), formally

$$\mu(x) \doteq \arg\min_{z \in \mathcal{S}} \|z - x\|.$$

Function μ is continuous. Define the function $\mu(x + v(x))$ which is also continuous and maps $\mathcal{S}$ in $\mathcal{S}$. Therefore, according to the fixed-point theorem, it admits a fixed point

$$\hat{x} = \mu(\hat{x} + v(\hat{x}))$$

Two cases are possible:

1. $\hat{x} + v(\hat{x}) \in \mathcal{S}$: in this case we have

[4] Actually, convexity can be relaxed to the requirement that there is a continuous invertible map G such that $G(\mathcal{S})$ is convex.

[5] The provided proof is sketched in [Mil95], to which the reader is referred for an interesting dissertation on game theory.

$$\hat{x} + v(\hat{x}) = \mu(\hat{x} + v(\hat{x})) = \hat{x}$$

and consequently $v(\hat{x}) = 0$.

2. $\hat{x} + v(\hat{x}) \notin \mathcal{S}$: this means that $\hat{x}$ is on the boundary (being the closest point to $\hat{x} + v(\hat{x})$) and necessarily $v(\hat{x})$ belongs to the normal cone $v(\hat{x}) \in \mathcal{H}_{\mathcal{S}}(\hat{x})$ in view of the Kuhn–Tucker conditions [Lue69], precisely

$$v(\hat{x})x \leq v(\hat{x})\hat{x}, \text{ for all } x \in \mathcal{S}$$

As a consequence, unless $v(\hat{x}) = 0$, $v(\hat{x})$ cannot belong also to the tangent cone $\mathcal{T}_{\mathcal{S}}(\hat{x})$ to $\mathcal{S}$, being the two cones polar to each other (see Proposition 3.17).

The proof of Theorem 4.21 is now immediate. Indeed if $\mathcal{S}$ is a convex and compact positively invariant set, then Nagumo's conditions (which are necessary and sufficient) must be satisfied and therefore $f(x) \in \mathcal{T}_{\mathcal{S}}(x)$ for all $x \in \mathcal{S}$ which means, in view of Lemma 4.22, that $f(x)$ vanishes somewhere in $\mathcal{S}$.

Another proof of the mentioned result can be found, in the more general setting of "weakly controlled invariant" sets, in [FH76b] (see Th. 3.3).

A third way to prove the theorem is based on the following idea, which is going to be only sketched for brevity. Each time-invariant system can be considered as periodic of arbitrary period $T > 0$. Consider, for $T > 0$ the map $\phi_T(x)$ defined as the value of the solution $x(T)$ with initial condition x. This map is continuous and maps $\mathcal{S}$ in $\mathcal{S}$. Therefore for each T there exists a fixed point $\hat{x}^{(T)}$, i.e. $\hat{x}^{(T)} = \phi_T(\hat{x}^{(T)})$. Now consider a "sequence of periods" $T_k > 0$ converging to zero and the corresponding sequence of fixed points $\hat{x}^{(T_k)} \in \mathcal{S}$ (we admit the notation is terrible). Since $\mathcal{S}$ is compact, $\hat{x}^{(T_k)}$ has an accumulation point $\bar{x}$. Without restriction, assume $\hat{x}^{(T_k)} \to \bar{x}$ (by possibly extracting a sub-sequence with this property). For each $\hat{x}^{(T_k)}$ consider the periodic trajectory originating in $\hat{x}^{(T_k)}$ and returning in $\hat{x}^{(T_k)}$ after time T_k. Since $T_k \to 0$, the maximum distance of each of these periodic trajectories from its originating point $\hat{x}^{(T_k)}$ goes to 0.

Now by simple-but-tedious considerations, one can argue that all these periodic trajectories "collapse to $\bar{x}$" which results to be a stationary point. The details are left to the interested reader as an exercise.

An important extension of the theorem is the fact that the set $\mathcal{S}$ is not necessarily convex but it has to be closed compact and homeomorphic to a convex set, namely that there exists a continuous invertible functions Φ which maps $\mathcal{S}$ in a compact and convex set $\hat{\mathcal{S}}$.

A known application of positive invariance is the question of the existence of a stationary point in a flow.

Example 4.23 (Stagnation point in a flow). Positive invariance is often referred to as flow invariance and we wish to give an intuitive idea of this concept with an example.

Consider Fig. 4.4, left, in which a planar flow (fluid electrical or magnetic) is defined with the following properties. The vector field

$$\dot{x} = V_x(x, y)$$
$$\dot{y} = V_y(x, y)$$

is stationary and smooth. The flow "enters through the lateral" vertical faces $V_x < 0$ in R (right) and $V_x > 0$ in L (left) and exits through the other two horizontal faces $V_y < 0$ in T (top) and $V_y > 0$ in B (bottom). We assume also that the curved surfaces are smooth and at the boundary the flow is tangent and directed as in the figure, so that at the two top boundaries L-T and R-T we have $V_y < 0$ and in the bottom boundaries L-B and R-B we have $V_y > 0$. Note that we cannot invoke any symmetry, since we do not assume neither a special flow distributions at the boundaries nor homogeneity of the space. we need only regularity of the flow.

The considered domain is not convex but it is a star-shaped set if we place the origin in its center and its boundary can be described by a continuous function of the angle in polar coordinates $\rho(\theta)$ or, equivalently, as a function of the unit vector $\rho(\cos(\theta), \sin(\theta))$. Such function expresses the distance of the point of the boundary from the origin in any direction θ. This means that the considered set is isomorphic to a convex set. To prove this, we just need to take the continuous invertible function which maps $\bar{x}$ on the boundary on a point of the unit ball and all x aligned with $\bar{x}$ and scaled proportionally as

$$(\hat{x}, \hat{y}) = \rho(\cos(\theta), \sin(\theta))^{-1}(x, y)$$

To show that there is at least a stagnation point, namely a point in which both "speed" components are 0, we need just to consider the modified flow

$$\dot{x} = V_x(x, y)$$
$$\dot{y} = -V_y(x, y)$$

which behaves as in Fig. 4.4, right. The new "flow" has no physical meaning but it is mathematically well defined. It is quite an easy exercise to show that the considered

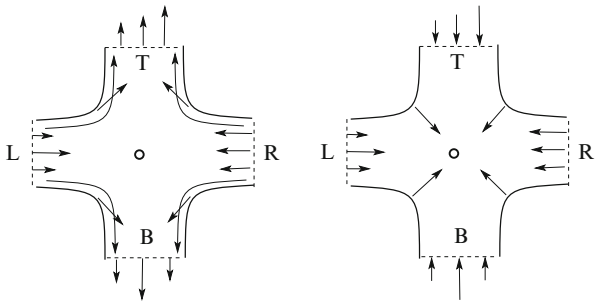


Fig. 4.4 The flow problem (left) and its complementary problem (right)

region is positively invariant for the new flow. Therefore there exists at least one point in which the modified flow has a stationary point. But this corresponds to a stationary point of the original flow.

We finally would like to stress that the existence of equilibria can be inferred from other assumptions. For instance, it turns out that if all the solutions $x(t)$, with $x(0) \in \mathbb{R}^n$ of a dynamical system $\dot{x} = f(x)$, with sufficiently regular f are ultimately bounded inside a compact set, then there is necessarily an equilibrium point $\bar{x}$, (i.e., $f(\bar{x}) = 0$) in such a set (see, for instance, [Srz85, Hal88, RW02]).

4.3 Convex invariant sets and linear systems

The results presented in the previous sections can be particularized to linear systems, leading to interesting results, especially as far as the construction of Lyapunov functions is concerned. Later in the book, it will indeed be shown how the derived controlled contractive sets, or level sets of Lyapunov functions, can be equipped by controllers $u(x)$ naturally associated with the points on the boundary (in fact with the vertices, in the case of a polytopic set $\mathcal{S}$).

As the reader can imagine, positive invariance concepts are extremely general, but the effective (constructive) results are limited to special classes of functions and systems. In particular some remarkable tools are available for the class of C-sets as candidate invariant sets for the class of linear (possibly uncertain) systems of the form

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t) + Ed(t), \quad w(t) \in \mathcal{W}, \quad d(t) \in \mathcal{D} \tag{4.11}$$

or, in the discrete-time case, of the form

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t), \quad w(t) \in \mathcal{W}, \quad d(t) \in \mathcal{D} \tag{4.12}$$

where the additive disturbance bounding set $\mathcal{D}$ is a C-set. We assume that $u \in \mathcal{U}$, a convex and closed set including the origin in its interior.

The existence of a contractive set is very important, because it allows the construction of Lyapunov functions according to the following theorem.

Theorem 4.24. *Assume that the C-set $\mathcal{S}$ is contractive for the system (4.11) (resp. (4.12)) and let $\Psi_{\mathcal{S}}(x)$ be its Minkowski function. The following statements hold*

- *If $E = 0$ and $\mathcal{U} = \mathbb{R}^n$, then $\Psi_{\mathcal{S}}(x)$ is a global control Lyapunov function.*
- *If $\mathcal{U} = \mathbb{R}^n$, then $\Psi_{\mathcal{S}}(x)$ is a control Lyapunov function outside $\mathcal{S}$.*
- *If $E = 0$, then $\Psi_{\mathcal{S}}(x)$ is a control Lyapunov function inside $\mathcal{S}$.*

Proof. Let us first consider the discrete-time version of the theorem. Consider any point $x \in \mathbb{R}^n$ with $x \neq 0$. There exists a unique point $\bar{x}$ on the boundary of $\mathcal{S}$ such that

$$x = \Psi_{\mathcal{S}}(x)\bar{x}$$

(i.e., $\bar{x}$ the intersection of the ray originating in 0 and passing through x). Consider the control

$$\Phi(x) = \Psi_{\mathcal{S}}(x)\bar{u}(\bar{x})$$

where $\bar{u}(\bar{x})$ is such that $A(w)\bar{x} + B(w)\bar{u}(\bar{x}) + Ed \in \lambda\mathcal{S}$ or, equivalently, $\Psi_{\mathcal{S}}(A(w)\bar{x} + B(w)\bar{u}(\bar{x}) + Ed) \leq \lambda$ for all $w \in \mathcal{W}$ and $d \in \mathcal{D}$ (such a value $\bar{u}(\bar{x})$ does exist by assumption). If $\mathcal{U} = \mathbb{R}^n$, the control $u(x) = \Phi(x)$ is admissible for all x and then, for all $w \in \mathcal{W}$ and $d \in \mathcal{D}$,

$$\begin{aligned}
&\Psi_{\mathcal{S}}(A(w)x + B(w)\Phi(x) + Ed) = \\
&= \Psi_{\mathcal{S}}\left(A(w)\Psi_{\mathcal{S}}(x)\bar{x} + B(w)\Psi_{\mathcal{S}}(x)\bar{u}(\bar{x}) + \Psi_{\mathcal{S}}(x)E(d/\Psi_{\mathcal{S}}(x))\right) \\
&= \Psi_{\mathcal{S}}(x)\Psi_{\mathcal{S}}\left(A(w)\bar{x} + B(w)\bar{u}(\bar{x}) + Ed'\right)
\end{aligned}$$

where $d' = d/\Psi_{\mathcal{S}}(x)$. If we show that

$$\Psi_{\mathcal{S}}\left(A(w)\bar{x} + B(w)\bar{u}(\bar{x}) + Ed'\right) \leq \lambda \tag{4.13}$$

for suitable x, we get the inequality

$$\Psi_{\mathcal{S}}(A(w)x + B(w)\Phi(x) + Ed) \leq \lambda\Psi_{\mathcal{S}}(x) \tag{4.14}$$

which proves the claim.

For $E = 0$ and u unconstrained (4.13) holds by definition, no matter how x is chosen so the first statement is then proved (actually the case $x = 0$ is not considered but it is quite trivial).

The second statement is also immediate. Indeed, for $x \notin int\mathcal{S}$, say $\Psi_{\mathcal{S}}(x) \geq 1$, $d' = d/\Psi_{\mathcal{S}}(x) \in \mathcal{D}$, which is a C-set and again (4.13) holds and thus (4.14). We notice that, for $x \notin \mathcal{S}$ it is then possible to apply the control $\Phi(x)$ (thus the global control $\hat{\Phi}(x) = \max\{\Psi_{\mathcal{S}}(x), 1\}\bar{u}(\bar{x})$ can be used).

The third statement is essentially a consequence of Theorem 4.18. The inequality (4.13) holds by construction, thus the only issue is the constraint on u. However, the scaled control $\Phi(x) = \Psi_{\mathcal{S}}(x)\bar{u}(\bar{x})$ is admissible if $\Psi_{\mathcal{S}}(x) \leq 1$, so that (4.14) holds for $x \in \mathcal{S}$.

To prove the theorem in the continuous-time case, note that if $\mathcal{S}$ is contractive for (4.11), then for each point $\bar{x} \in \partial\mathcal{S}$ there exists $\bar{u}$ such that

$$\limsup_{h \to 0} \frac{\Psi_{\mathcal{S}}(\bar{x} + h(A(w)\bar{x} + B(w)\bar{u}(\bar{x}) + Ed)) - \Psi_{\mathcal{S}}(\bar{x})}{h} \leq -\beta$$

As done for the discrete-time case, the control can be "extended" as $u(x) = \bar{u}(\bar{x})\Psi_{\mathcal{S}}(x)$ depending on which of the three following cases is being considered:

i) $E = 0$ and u unbounded: extended to all $x \in \mathbb{R}^n$;
ii) $E \neq 0$ and u unbounded: extended to all $x \notin \mathcal{S}$;
iii) $E = 0$ and u bounded: extended to all $x \in \mathcal{S}$.

Consider, for brevity, case i). Then, by scaling, one gets

$$\limsup_{h\to 0} \quad \frac{\Psi_{\mathcal{S}}(x + h[A(w)x + B(w)u(x)]) - \Psi_{\mathcal{S}}(x)}{h}$$
$$= \Psi_{\mathcal{S}}(x) \frac{\Psi_{\mathcal{S}}(\bar{x} + h[A(w)\bar{x} + B(w)\bar{u}(\bar{x})] - \Psi_{\mathcal{S}}(\bar{x})}{h} \leq -\beta \Psi_{\mathcal{S}}(x)$$

and therefore the control $u(x)$ assures the decreasing conditions in a pointwise sense. Since $\bar{u}(\bar{x})$ is locally Lipschitz on $\partial\mathcal{S}$, then $\Phi(x)$ is continuous (actually locally Lipschitz) and thus the statement is proved.

The above theorem can be extended quite naturally to the full information control case, $u = \Phi(x, w)$, as well as to the output feedback control case, assuming a linear output function $y = Cx$,

$$u = \Phi(Cx, w) \tag{4.15}$$

The next lemma relates continuous and discrete-time systems and will be very useful in the sequel. Its importance lies in the fact that it establishes a relation between the stabilization of the continuous-time system and of the associated Euler Auxiliary System (EAS).

Definition 4.25 (Euler Auxiliary System). Given the continuous-time system (4.11) and $\tau > 0$, the discrete-time system

$$x(t+1) = [I + \tau A(w(t))]x(t) + \tau B(w(t))u(t) + \tau E d(t) \tag{4.16}$$

is the Euler Auxiliary System.

Before getting confused with the parameter τ just introduced, the reader is referred to Remark 4.27 where the relation between this parameter and the sampling period of a possible digital implementation is better explained.

Lemma 4.26. *The following conditions are equivalent.*

i) There exist a non-negative $\lambda < 1$ and $\tau > 0$ and a locally Lipschitz function $u(x)$ such that, for all $x \notin int\{\mathcal{S}\}$ (respectively: if $E = 0$, for all x; if $E = 0$ and $u \in \mathcal{U}$, for all $x \in \mathcal{S}$),

$$\Psi_{\mathcal{S}}(x + \tau(A(w)x + B(w)u(x) + Ed)) \leq \lambda \Psi_{\mathcal{S}}(x)$$

ii) There exist $\beta > 0$ and a locally Lipschitz function $u(x)$ such that, for all $x \notin int\{\mathcal{S}\}$ (respectively: if $E = 0$, for all x; if $E = 0$ and $u \in \mathcal{U}$, for all $x \in \mathcal{S}$),

$$D^+\Psi_{\mathcal{S}}(x, A(w)x + B(w)u + Ed) \leq -\beta\Psi_{\mathcal{S}}(x), \quad \textit{for all } w \in \mathcal{W}, \ d \in \mathcal{D}$$

Furthermore, if i) holds, then β as in ii) can be derived as

$$\beta = \frac{1-\lambda}{\tau}$$

Proof. The implication i) $\Rightarrow$ ii) along with the last assertion will be shown here, whereas for the converse statement ii) $\Rightarrow$ i) the reader is referred to [Bla95].

Since $\Psi_{\mathcal{S}}$ is convex, the quotient ratio is a non-decreasing function of τ, namely

$$\frac{\Psi(x+\tau_1 z) - \Psi(x)}{\tau_1} \leq \frac{\Psi(x+\tau_2 z) - \Psi(x)}{\tau_2}$$

for $\tau_1 \leq \tau_2$ [Roc70]. In view of the above decreasing condition,

$$\begin{aligned} & D^+\Psi_{\mathcal{S}}(x, A(w)x + B(w)u(x) + Ed) \\ \leq & \frac{\Psi_{\mathcal{S}}\left(x + \tau(A(w)x + B(w)u(x) + Ed)\right) - \Psi_{\mathcal{S}}(x)}{\tau} \leq \\ \leq & \frac{\lambda\Psi_{\mathcal{S}}(x) - \Psi_{\mathcal{S}}(x)}{\tau} \leq -\frac{1-\lambda}{\tau}\Psi_{\mathcal{S}}(x) \end{aligned}$$

which is what was to be shown.

The importance of the Lemma is that it allows the computation of Lyapunov functions and invariant sets for continuous-time systems by exploiting discrete-time algorithms. To avoid confusion, we report immediately the following.

Remark 4.27. Parameter τ has nothing to do with a sampling time. All the properties and the constructions that will be based on the EAS will be proper of the continuous-time system. So if the control $u(x)$ has to be digitally implemented, the sampling time T is not required nor recommended to be equal to τ. In practice the condition $T << \tau$ should be satisfied.

Further relevant properties of the EAS will be reported in the appendix.

The next lemma concerns the power of a Lyapunov function. It shows that if Ψ is a Lyapunov function, so is Ψ^p

Lemma 4.28. *Assume $\Psi(x)$ is a control Lyapunov function for the dynamic system $\dot{x}(t) = f(x(t), u(t), w(t))$ (global, inside or outside a C-set $\mathcal{S}$). Assume that Ψ^p is locally Lipschitz, where $p > 0$ is a real number*[6]*. Then $\Psi^p(x)$ is also a control Lyapunov function for the system.*

[6]The request for Ψ^p to be locally Lipschitz is for coherence with our definition of a Lyapunov function and could be removed with suitable care.

Proof. By definition $D^+\Psi(x,f(x,\Phi(x),w)) \leq -\phi(\|x\|)$ and, since Ψ is positive definite, $\Psi(x) \geq \phi_0(\|x\|)$ for some κ-functions ϕ and ϕ_0. Consider the Lyapunov derivative

$$\begin{aligned} D^+\Psi^p(x) &= \limsup_{h\to 0^+} \frac{\Psi^p(x+hf(x,\Phi(x),w)) - \Psi^p(x)}{h} \\ &= \limsup_{h\to 0^+} \frac{\Psi^p(x+hf) - \Psi^p(x)}{\Psi(x+hf) - \Psi(x)} \ \frac{\Psi(x+hf) - \Psi(x)}{h} \\ &= \lim_{h\to 0^+} \frac{\Psi^p(x+hf) - \Psi^p(x)}{\Psi(x+hf) - \Psi(x)} \ \limsup_{h\to 0^+} \frac{\Psi(x+hf) - \Psi(x)}{h} \\ &\leq -p\Psi^{p-1}(x)\phi(\|x\|) \leq -p\phi_0(\|x\|)^{p-1}\phi(\|x\|) \end{aligned}$$

since $\phi_1(\cdot) \doteq p\phi_0(\cdot)^{p-1}\phi(\cdot)$ is a κ-function, the assertion is proved.

Basically the lemma states that what characterizes a control Lyapunov function is the shape of its level surfaces, rather than the corresponding values. For instance, if a quadratic Lyapunov function of the form $\Psi(x) = x^TPx$ is considered, also its root, namely the quadratic norm $\|x\|_P \doteq \sqrt{x^TPx}$ is a Lyapunov function. This property has been reported to emphasize that classes of functions such as the piecewise linear, the quadratic and the homogeneous polynomial functions of order p can be all equivalently replaced by positively homogeneous functions of order one and then analyzed and compared in the common class of norms.

As a final remark, let us note that if a polynomially bounded function assures exponential stability

$$D^+\Psi(x,f) \leq -\beta\Psi(x)$$

then the Lyapunov function $\Psi^p(x)$ (assuming it polynomially bounded) assures the condition

$$D^+\Psi^p(x,f) \leq -\beta p\Psi^p(x)$$

and therefore, up to a change of coefficient, exponential convergence is assured by both $\Psi(x)$ and $\Psi^p(x)$.

In the sequel, several techniques to compute contractive sets and the associated Lyapunov function will be presented. Many of the proposed results are known in the literature. The main feature of the book, however, is that of presenting a different perspective. In some sense, the main aspect which will be emphasized is that of positive invariance and contractivity. We saw that these concepts are related to that of a Lyapunov function. In particular, it will be shown how to generate Lyapunov functions from contractive sets, rather than the opposite.

The next lemma is a preliminary result, which holds for convex and positively homogeneous functions of order 1 (see [BM00] for a proof), we need to prove the subsequent key Lemma 4.30.

Lemma 4.29. *Let $\Psi_{\mathcal{S}}(x)$ be the Minkowski function of the convex C-set $\mathcal{S}$. Then, denoting by $\partial\Psi_{\mathcal{S}}(x)$ the sub-differential of $\Psi_{\mathcal{S}}(x)$, the following holds:*

$$\Psi_{\mathcal{S}}(x) = w^T x, \quad \text{for all } w \in \partial\Psi_{\mathcal{S}}(x).$$

With the above in mind, the following result can be presented, which will be used in the sequel to prove the equivalence between stability and exponential stability for a class of systems.

Lemma 4.30. *Consider the dynamic system*

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

with $w(t) \in \mathcal{W}$. The C-set $\mathcal{S}$ is β-contractive for the above system associated with the control $u = \Phi(x) \in \mathcal{U}$ if and only if it is controlled-invariant for the system

$$\dot{x}(t) = \beta x(t) + f(x(t), u(t), w(t))$$

associated with the same control action $u = \Phi(x)$.

Proof. Consider the expression of the directional derivative of a convex function (3.9) and recall that, by the just introduced Lemma, $w^T x = \Psi_{\mathcal{S}}(x)$ for all $w \in \partial\Psi_{\mathcal{S}}(x)$. Then we have

$$D^+\Psi_{\mathcal{S}}(x, f(x,u,w)) = \sup_{w \in \partial\Psi_{\mathcal{S}}(x)} w^T [f(x,u,w)] =$$

$$\sup_{w \in \partial\Psi_{\mathcal{S}}(x)} \left[w^T \left(\beta x + f(x,u,w)\right)) - \beta \underbrace{w^T x}_{=\Psi_{\mathcal{S}}(x)} \right] =$$

$$\sup_{w \in \partial\Psi_{\mathcal{S}}(x)} w^T \left[\beta x + f(x,u,w)\right] - \beta\Psi_{\mathcal{S}}(x)$$

Therefore

$$D^+\Psi_{\mathcal{S}}(x, f(x,u,w)) \leq -\beta\Psi_{\mathcal{S}}(x)$$

is equivalent to

$$D^+\Psi_{\mathcal{S}}(x, \beta x + f(x,u,w)) \leq 0,$$

say the contractivity of $\mathcal{S}$ implies its controlled invariance for the modified system.

The Lemma admits a discrete-time version.

Lemma 4.31. *The C-set $\mathcal{S}$ is λ contractive for the dynamic system*

$$x(t+1) = f(x(t), u(t), w(t))$$

with $w(t) \in \mathcal{W}$ and $u(t) \in \mathcal{U}$ if and only if it is controlled-invariant for the modified system

$$x(t+1) = \frac{f(x(t), u(t), w(t))}{\lambda}$$

4.4 Ellipsoidal invariant sets

Ellipsoids are the most commonly exploited sets as candidate invariant regions since they are associated with powerful tools such as the Lyapunov equation or Linear Matrix Inequalities (LMIs). In this section, an overview of the main results will be provided. The reader is referred to the excellent book [BEGFB04] for a specialized exposition.

4.4.1 Ellipsoidal invariant sets for continuous-time systems

Consider an ellipsoidal set of the form[7]

$$\mathcal{E}(P,1) = \{x : \sqrt{x^T P x} \leq 1\} = \mathcal{N}[\sqrt{x^T P x}, 1],$$

namely the unit ball of the quadratic norm

$$\|x\|_P = \sqrt{x^T P x},$$

and apply Nagumo's condition for the system

$$\dot{x}(t) = f(x(t), w(t))$$

in a point x on the boundary. The tangent cone, in this case the tangent plane, is given by

$$\mathcal{T}_{\mathcal{E}}(x) = \{z : \ x^T P z \leq 0\}$$

[7] Assuming $\mu = 1$ in the expression $\mathcal{E}(P,\mu)$ is not a restriction because it is always possible to scale P to achieve $\mathcal{E}(P,\mu) = \mathcal{E}(P/\mu, 1)$.

thus the positive invariance condition becomes

$$x^T P f(x,w) \leq 0, \quad \text{for all } \; x \in \partial\mathcal{E}(P,1).$$

In the case of a linear time-invariant system $\dot{x} = Ax$, the above condition becomes $x^T PAx \leq 0$, for all $x \in \partial\mathcal{E}(P,1)$, which is equivalent to

$$A^T P + PA \preceq 0,$$

the well-known Lyapunov inequality. If the ellipsoid is contractive, then for every x on the boundary (i.e., such that $\sqrt{x^T Px} = 1$) the following condition

$$D^+ \|x\|_P = D^+ \sqrt{x^T Px} = \frac{1}{\sqrt{x^T Px}} x^T PAx \; \leq -\beta$$

must be satisfied. By scaling, we have that

$$D^+ \|x\|_P = \frac{1}{\sqrt{x^T Px}} x^T PAx \leq -\beta\sqrt{x^T Px} = -\beta\|x\|_P$$

must hold for all x. Thus contractivity condition assures exponential β-convergence with the transient estimate

$$\|x(t)\|_P \leq \|x(0)\|_P \, e^{-\beta t} \tag{4.17}$$

If one considers the quadratic function $\|x\|_P^2$, the contractivity conditions becomes

$$D^+[x^T Px] \leq -2\beta x^T Px$$

so achieving

$$A^T P + PA \prec -2\beta P \tag{4.18}$$

which is an LMI necessary and sufficient condition for an ellipsoid to be β-contractive for the linear system. LMIs have had a great success in the literature since they involve efficient numerical tools for their manipulation. The strong property that characterizes this kind of conditions is the convexity of the admissibility domain. In fact the set of all the symmetric matrices $P \succ 0$ such that (4.18) is satisfied is convex, precisely if P_1 and P_2 satisfy (4.18) so does $\alpha P_1 + (1-\alpha)P_2$, for all $0 \leq \alpha \leq 1$. As it will be shown in several parts of the book, many significant problems can be reduced to the solution of a set of LMIs.

Going to synthesis problems, the conditions for an ellipsoid $\mathcal{E}(P,1)$ to be controlled-invariant (or contractive) for a simple linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bu(t),$$

require the existence a control function Φ, which has to be Lipschitz continuous on the boundary of $\mathcal{E}(P,1)$, such that the following inequality

$$x^T PAx + x^T PB\Phi(x) < 0$$

is satisfied for $x^T Px = 1$. One possible choice for $\Phi(x)$ is the gradient-based controller (2.54), which in this case is indeed linear and precisely

$$u_\gamma(x) = -\gamma B^T Px \tag{4.19}$$

for $\gamma > 0$ large enough. Therefore, it is always possible to associate a linear controller with a contractive ellipsoid (this property holds even if $A(w)$ is an uncertain matrix as long as B is known and constant) [Meı74, BPF83]. If such a control law is substituted in the expression of the derivative on the boundary, it turns out that the level of contractivity β is assured if

$$\dot{\Psi}(x) = 2x^T PAx - \gamma 2x^T PBB^T Px = x^T (A^T P + PA - 2\gamma PBB^T P)x < -2\beta x^T Px$$

namely if the inequality

$$\left(A^T P + PA - 2\gamma PBB^T P + 2\beta P\right) \preceq 0$$

holds. The condition provided here is not linear in P. By defining $Q \doteq P^{-1}$, and pre and post multiplying the expression between brackets by Q one gets

$$\left(QA^T + AQ - 2\gamma BB^T + 2\beta Q\right) \preceq 0$$

which is an LMI. More generally, it is possible to consider a linear feedback matrix K leading to the inequality

$$A^T P + PA + K^T B^T P + PBK \prec 0 \tag{4.20}$$

(which is not necessarily a gradient-based control). Though this condition is nonlinear (since P and K are unknown). Still it can be re-parameterized into a linear condition by setting $Q = P^{-1}$, $R = KQ$, and by pre and post multiplying (4.20) by Q. The above multiplication indeed transforms (4.20) into

$$QA^T + AQ + R^T B^T + BR \prec 0, \quad Q \succ 0 \tag{4.21}$$

which is again a nice LMI condition. If a solution Q does exist, then

$$K = RP \tag{4.22}$$

is a suitable feedback matrix. Condition (4.21) characterizes the set of all P such that $\mathcal{E}(P,1)$ is a contractive ellipsoid and (4.22) characterizes the set of all linear stabilizing controls $u = Kx$ (which are not necessarily of the form (4.19)).

There are invariance conditions for systems with persistent additive disturbances, namely systems of the form

$$\dot{x}(t) = Ax(t) + Ed(t)$$

Assume that the bound for the disturbance is $d^T d \leq 1$. Then in [USGW82] it has been shown that the ellipsoid $\mathcal{E}(P,1)$ is positively invariant if $Q = P^{-1}$ satisfies the condition

$$QA^T + AQ + \alpha Q + \frac{1}{\alpha} EE^T \preceq 0, \quad \text{for some} \quad \alpha > 0 \tag{4.23}$$

The reader is referred to [PHSG88] for an interesting application of this condition to the synthesis of a control of a power plant. Quite surprisingly, the above condition has been later proved to be also necessary if (A, E) is a reachable pair [BC98].

It is worth mentioning that, in some problems, degenerate forms of ellipsoids $\mathcal{E}(P,\mu)$, with P positive semi-definite, can arise[8]. For instance, these sets are useful in cases in which one is interested just in the stability of part of the state variables. This condition, known as partial stability, arises in many contexts [Vor98]. The positive invariance conditions are exactly the same already proposed, with the only understanding that P is not necessarily positive definite.

A fundamental issue is the output feedback case. Unfortunately, as it is well known, this is a hard problem. Indeed, if one wishes to consider a control of the form

$$u = Ky = KCx$$

where $y = Cx$, the gain parameterization previously introduced has to be reconsidered. Such parameterization now becomes

$$RP = KC$$

or

$$R = KCQ,$$

where $Q = P^{-1}$. Unfortunately, linearity is lost by this new constraint. Indeed, the known methods for static output feedback synthesis do not have strong properties such as convexity. Needless to say, static output feedback is known to be one of the open problems in control theory.

[8] Or even regions of the form $\mathcal{N}(x^T Px, \mu)$, deriving from quadratic functions which are not sign definite.

4.4.2 Ellipsoidal invariant sets for discrete-time systems

Let us now analyze the case of a discrete-time linear system. The positive invariance of $\mathcal{E}(P,1)$ is equivalent to the fact that, given the vector norm $\|x\|_P = \sqrt{x^T P x}$, the corresponding induced norm of matrix A is less than 1, precisely

$$\|A\|_P = \sup_{\sqrt{x^T P x}=1} \sqrt{(Ax)^T P (Ax)} < 1$$

for all x. The above leads to the discrete-time Lyapunov inequality

$$A^T P A - P \prec 0. \tag{4.24}$$

Let us consider now the problem of checking controlled invariance. Consider the linear discrete-time system

$$x(t+1) = Ax(t) + Bu(t)$$

The ellipsoid $\mathcal{E}(P,1)$ is contractive if and only if there exists $\lambda < 1$ such that, for all x such that $x^T P x \leq 1$ (say, $\|x\|_P \leq 1$), there exists $u(x)$ such that $\|Ax+Bu(x)\|_P \leq \lambda$. For the moment being, let us assume that such a control does exist. Then it can be easily derived by the well-known minimization problem

$$u(x) = \arg\min_{u \in \mathbb{R}^n} \|Ax + Bu\|_P$$

which, solved by the least square formula, yields

$$u = -(B^T P B)^{-1} B^T P A x$$

Denote $\hat{B} = P^{\frac{1}{2}} B$, $\hat{A} = P^{\frac{1}{2}} A P^{-\frac{1}{2}}$ and $\hat{x} = P^{\frac{1}{2}} x$ so that $x \in \mathcal{E}(P,1)$ results in $\|\hat{x}\|_2 \leq 1$.

By substitution, the following is achieved

$$\|Ax + Bu\|_P^2 = (Ax + Bu)^T P (Ax + Bu) = (\hat{A}\hat{x} + \hat{B}u)^T (\hat{A}\hat{x} + \hat{B}u),$$

say

$$\|Ax + Bu\|_P^2 = \|\hat{A}\hat{x} + \hat{B}u\|_2^2 = \left\| \left(I - \hat{B} \left(\hat{B}^T \hat{B} \right)^{-1} \hat{B}^T \right) \hat{A}\hat{x} \right\|_2^2$$

Since the above must hold for every $\|\hat{x}\|_2 \leq 1$, the controlled invariance of the ellipsoid $\mathcal{E}(P,1)$ is equivalent to the fact that the induced norm of the rightmost term of last expression is less than 1, that is

$$\left\| \left(I - \hat{B} \left(\hat{B}^T \hat{B} \right)^{-1} \hat{B}^T \right) \hat{A} \right\|_2 =$$
$$\left\| \left(I - P^{\frac{1}{2}} B \left(B^T P B \right)^{-1} B^T P^{\frac{1}{2}} \right) P^{\frac{1}{2}} A P^{-\frac{1}{2}} \right\|_2 < 1$$

The previous condition can be used to check controlled invariance of $\mathcal{E}(P, 1)$, but, as it is written, it is not suitable to determine P.

There is a mechanism to determine $P \succ 0$ along with a linear controller $u = Kx$. Closing the loop, from (4.24), one has

$$(A + BK)^T P (A + BK) - P \prec 0.$$

Again, pre and post multiplying by $Q = P^{-1}$ and setting $KQ = R$ one gets

$$(QA^T + R^T B^T) Q^{-1} (AQ + BR) - Q \prec 0,$$

which, along with $Q \succ 0$, is known to be equivalent to [BEGFB04]

$$\begin{bmatrix} Q & QA^T + R^T B^T \\ AQ + BR & Q \end{bmatrix} \succ 0$$

This is a "nice" convex condition with respect Q and R. In the sequel of the book, positively invariant ellipsoids will be investigated in connection with the control of constrained systems, the control of LPV systems and system performance evaluation.

4.5 Polyhedral invariant sets

Polyhedral sets and the associated polyhedral functions, although less popular than ellipsoids, have been widely accepted as good candidate invariant sets. They present several theoretical and practical advantages over the ellipsoids, but they suffer from the problem of complexity of their representation. Here, some basic invariance conditions for polyhedral sets are provided. To this aim, we remind that a polyhedral set can be represented as in (3.22)

$$\mathcal{P}(F) = \{x : \ Fx \leq \bar{1}\} \tag{4.25}$$

or in its dual form (3.23)

$$\mathcal{V}(X) = \{x = Xz, \ \bar{1}^T z \leq 1, \ z \geq 0\} \tag{4.26}$$

where $\bar{1} = [1 \ 1 \ldots 1]^T$. We remind that the notation $P \leq Q$ between matrices or vectors has to be intended component-wise.

If a polyhedral C-set is considered, the Minkowski (gauge) functions deriving from representations (4.25) and (4.26) with $\mathcal{P}(F)$ and $\mathcal{V}(X)$

$$\Psi_{\mathcal{P}(F)}(x) = \max\{Fx\} \doteq \max_i\{F_i\, x\} \tag{4.27}$$

and

$$\Psi_{\mathcal{V}(X)}(x) = \min\{\bar{1}^T\, w : \;\; x = Xw, \;\; w \geq 0\}. \tag{4.28}$$

Note that the first expression is valid for unbounded sets, including 0 in the interior (we have seen that the Minkowski function can be defined for unbounded sets) provided that we add the 0 constraint

$$\Psi_{\mathcal{P}(F)}(x) = \max\{0, F_1 x, \; F_2 x, \dots\}$$

The expression (4.27) is useful to characterize the Lyapunov derivative along the trajectory of the dynamic system

$$\dot{x}(t) = G(x(t), w(t)), \quad w(t) \in \mathcal{W}.$$

The Lyapunov derivative, whose expression was given in (2.30), is

$$D^+\Psi(x, w) = \max_{i \in \mathcal{I}(x)} F_i\, G(x, w), \tag{4.29}$$

where $\mathcal{I}(x)$ is the maximizer subset

$$\mathcal{I}(x) = \{i : F_i(x) = \Psi_{\mathcal{P}(F)}(x)\}.$$

Fortunately, in the case of interest, this terrifying expression will be seldom used and, even when used, it will just be for theoretical purposes.

4.5.1 Contractive polyhedral sets for continuous-time systems

In the case of a linear system, necessary and sufficient conditions for a C-set of the polyhedral type to be positively invariant can be derived. To this aim, we introduce the following definition

Definition 4.32 (Metzler matrix). Matrix M is a Metzler matrix if $M_{ij} \geq 0$ for $i \neq j$.

The importance of Metzler matrices will be seen in the next result. We remind that $\bar{1} = [1\; 1 \dots 1]^T$.

Theorem 4.33. *Consider the linear system*

$$\dot{x}(t) = Ax(t)$$

and let $\mathcal{S}$ be a polyhedral C-set of the form (4.25), with $F \in \mathbb{R}^{s\times n}$, or of the form (4.26) $X \in \mathbb{R}^{n\times r}$. Then the next statements are equivalent.

i) $\mathcal{S} = \mathcal{P}(F) = \mathcal{V}(X)$ *is* β*-contractive.*
ii) There exists a Metzler matrix H such that

$$HF = FA, \quad H\,\bar{1} \leq -\beta\bar{1} \tag{4.30}$$

iii) There exists an Metzler matrix H such that

$$AX = XH, \quad \bar{1}^T H \leq -\beta\bar{1}^T$$

Proof. The equivalence between i) and ii) is shown here, whereas the equivalence between i) and iii) will be proved later (see Remark 4.38 after Theorem 4.37) in a more general case.

ii) $\Longrightarrow$ i). Proving contractivity of $\mathcal{S}$ is equivalent to proving that for every $x(t_0)$ on the boundary, $D^+\Psi(x(t_0)) \leq -\beta$. From equation (4.29) this is in turn equivalent to the requirement $\frac{d}{dt}[F_i x(t_0)] \leq -\beta$, for all $i \in \mathcal{I}(x)$.

Consider then $x(t_0) \in \partial\mathcal{S}$. Since $x(t_0)$ belongs to the boundary, by definition of polyhedral C-set, $F_i x(t_0) = 1$ for every $i \in \mathcal{I}(x(t_0))$ and $F_j x(t_0) < 1$, for $j \notin \mathcal{I}(x(t_0))$. From (4.30), for every $i \in \mathcal{I}(x(t_0))$

$$\frac{d}{dt}F_i x(t)\bigg|_{t=t_0} = F_i\dot{x}(t_0) = F_i A x(t_0) = H_i F x(t_0)$$

where H_i is the ith row of H. Then, for $x = x(t_0)$,

$$\frac{d}{dt}F_i x(t_0) = \sum_{j=1}^{s} H_{ij}F_j x = \sum_{j\neq i} \underbrace{H_{ij}}_{\geq 0}\underbrace{F_j x}_{\leq 1} + H_{ii}\underbrace{F_i x}_{=1} \leq$$

$$\sum_{j\neq i} H_{ij} + H_{ii} = H_i\bar{1} \leq -\beta$$

which is what was to be shown. i) $\Longrightarrow$ ii). This part of the proof is more involved, because matrix H has to be constructed. The key idea is provided in [VB89]. The contractivity of the set implies that for all x on the boundary, $\frac{d}{dt}[F_i x] \leq -\beta$, for all $i \in \mathcal{I}(x)$. To construct matrix H, the boundary of $\mathcal{S}$ is considered face by face and it is shown how to construct the row of H corresponding to a specific face. Let us then see how to compute the first row of H starting from the first face of $\mathcal{S}$, namely the set

$$\{x:\ F_1 x = 1,\ \ F_j x \leq 1, j \neq 1\}$$

On this set, in view of the β-contractivity, $\frac{d}{dt}[F_1 x] = F_1 Ax \leq -\beta$. Consider the following linear programming problem

$$\begin{aligned} \mu &= \max F_1 Ax \\ &s.t. \\ F_1 x &= 1 \\ \tilde{F} x &\leq \bar{1} \end{aligned}$$

where $\tilde{F}$ is the $(s-1 \times n)$ matrix achieved by removing the first row of F (therefore $\bar{1}$ in the last inequality has $s-1$ ones). The optimal value must be such that $\mu \leq -\beta$. The optimal value of the dual [Pad99] linear programming problem

$$\begin{aligned} \mu &= \min \left[w_1 \ \tilde{w}^T \right] \bar{1} \\ &\text{s.t.} \\ \left[w_1 \ \tilde{w}^T \right] F &= F_1 A \\ w_1 \quad &\in \mathbb{R} \\ \tilde{w}^T &\geq \bar{0} \end{aligned}$$

is also $\mu \leq -\beta$, and therefore, there exists a feasible dual solution $[w_1 \ \tilde{w}^T]$ which has the same cost μ. Take H_1, the 1st row of H, equal to this solution so that

$$\mu = H_1 \bar{1} \leq -\beta, \ \ H_1 F = F_1 A, \quad \text{and} \quad H_{1j} \geq 0, \ \ \text{for } j \neq 1.$$

The ith row H_i of H, $i = 2, 3, \ldots, s$ can be determined exactly in the same fashion. Then we achieve

$$\begin{aligned} H_i \bar{1} &\leq -\beta \\ H_i F &= F_i A \\ H_{ij} &\geq 0, \ \ j \neq i \end{aligned}$$

Therefore, the so composed matrix H satisfies condition ii).

Remark 4.34. The previous theorem provides, as a special case for $\beta = 0$, necessary and sufficient conditions for the invariance of a polytope. In this case the condition $F_i Ax \leq 0$, $i \in \mathcal{I}(x)$ on the boundary has a clear interpretation in terms of Nagumo's condition.

The following dual property, immediate consequence of Theorem 4.33, holds.

Proposition 4.35. *The set $\mathcal{P}(F)$ (or $\bar{\mathcal{P}}(F)$) is β-contractive for the system*

$$\dot{x}(t) = Ax(t)$$

if and only if the dual set $\mathcal{V}(F^T)$ (or $\bar{\mathcal{V}}(F^T)$)) is β-contractive for the dual system

$$\dot{x}(t) = A^T x(t).$$

Theorem 4.33 admits a "symmetric" version. Precisely if one considers a symmetric polyhedral C-set of the form (3.24) and (3.25)

$$\bar{\mathcal{P}}(F) = \{x : \ \|Fx\|_\infty \leq 1\} \tag{4.31}$$

and its dual

$$\bar{\mathcal{V}}(X) = \{x = Xz, \ \ \|z\|_1 \leq 1\} \tag{4.32}$$

then $\mathcal{S} = \bar{\mathcal{P}}(F)$ is β-contractive if and only if there exists a matrix H such that $HF = FA$ and

$$\begin{bmatrix} \bar{H} & \underline{\mathrm{H}} \\ \underline{\mathrm{H}} & \bar{H} \end{bmatrix} \ \begin{bmatrix} \bar{1} \\ \bar{1} \end{bmatrix} \leq -\beta \begin{bmatrix} \bar{1} \\ \bar{1} \end{bmatrix}$$

where $\bar{H}$ and $\underline{\mathrm{H}}$ are defined as follows: $\bar{H}_{ij} = \max\{H_{ij}, 0\}$, for $i \neq j$ and $\bar{H}_{ii} = H_{ii}$, $\underline{\mathrm{H}}_{ij} = \max\{-H_{ij}, 0\}$, for $i \neq j$ and $\underline{\mathrm{H}}_{ii} = 0$.

Similarly, $\mathcal{S} = \bar{\mathcal{V}}(X)$ is β-contractive if and only if there exists a matrix H such that $AX = XH$ and

$$\begin{bmatrix} \bar{1}^T & \bar{1}^T \end{bmatrix} \ \begin{bmatrix} \bar{H} & \underline{\mathrm{H}} \\ \underline{\mathrm{H}} & \bar{H} \end{bmatrix} \ \leq -\beta \begin{bmatrix} \bar{1}^T & \bar{1}^T \end{bmatrix}$$

The proof of the former statement can be found in [Bit91]. The proof of the latter statement follows by duality, in view of Proposition 4.35.

Remark 4.36. Note that the previous inequalities for H are equivalent to the following "diagonal-dominance conditions"

$$H_{ii} + \sum_{j \neq i} |H_{ij}| \leq -\beta$$

and

$$H_{jj} + \sum_{i \neq j} |H_{ij}| \leq -\beta$$

(the diagonal terms H_{jj} have to be negative). We will comment on this aspect later on in Subsection 4.5.5.

Let us now consider the problem of controlled invariance. It is possible to provide a characterization of the contractivity of a set for a controlled system by means of the vertex representation (4.26).

Theorem 4.37. *Consider the linear system*

$$\dot{x}(t) = Ax(t) + Bu(t)$$

where $u(t) \in \mathcal{U} \subseteq \mathbb{R}^m$, with $\mathcal{U}$ a convex set. Let $\mathcal{S}$ be a polyhedral C-set of the form (4.26), with $X \in \mathbb{R}^{n\times r}$. Then $\mathcal{S} = \mathcal{V}(X)$ is β-contractive if and only if there exist a Metzler matrix H and a matrix $U \in \mathbb{R}^{m\times r}$) such that

$$AX + BU = XH, \tag{4.33}$$

$$\bar{1}^T H \leq -\beta \bar{1}^T \tag{4.34}$$

$$u_k \in \mathcal{U}, \tag{4.35}$$

where u_k is the kth column of U.

Proof. A simple way to prove the theorem is based on Lemma 4.30. Consider the modified system

$$\dot{x}(t) = [\beta I + A]x(t) + Bu(t) \doteq \hat{A}x(t) + Bu(t) \tag{4.36}$$

To prove necessity, we first show that controlled invariance of $\mathcal{S}$ for this system implies conditions (4.33)–(4.35) above with $\beta = 0$. If the system is controlled-invariant, there exists a control action $u = \Phi(x) \in \mathcal{U}$, such that Nagumo's condition holds. Let x_i be the ith vertex of $\mathcal{V}(x)$, i.e. the ith column of X. Set $u_i \doteq \Phi(x_i)$, the control at the vertices. The tangent cone in the ith vertex is given by the cone generated by all the vectors $x_j - x_i$ namely:

$$\mathcal{T}_{\mathcal{S}}(x_i) = \{y = \sum_{j\neq i} \nu_j (x_j - x_i), \quad \nu_j \geq 0\}$$

Nagumo's condition implies that

$$\dot{x}_i = \hat{A}x_i + Bu_i \in \mathcal{T}_{\mathcal{S}}(x_i),$$

say there exist $\nu_j \geq 0$ such that

$$\hat{A}x_i + Bu_i = \sum_{j\neq i} \nu_j (x_j - x_i) = \sum_{j=1}^{r} \hat{h}_{ij}\, x_j \tag{4.37}$$

where $\hat{h}_{ij} = \nu_j$ if $j \neq i$ and $\hat{h}_{ii} = -\sum_{j \neq i} \nu_j$. Let $\hat{H}$ be the Metzler matrix whose coefficients are the $\hat{h}_{ij}$ and U the matrix formed by the control at the vertices, $U = [u_1\ u_2\ \dots\ u_r]$. Then

$$\hat{A}X + BU = X\hat{H}, \quad \bar{1}^T\hat{H} \leq \bar{0}, \quad u_k \in \mathcal{U} \tag{4.38}$$

Going back to the original system, by defining the Metzler matrix

$$H = \hat{H} - \beta I$$

one gets (4.33)–(4.35).

To prove sufficiency note that (4.33)–(4.35) imply (4.38) which, in turn, implies, by reverse reasoning, that Nagumo's subtangentiality conditions are verified at the vertices. To prove controlled-invariance of $\mathcal{S}$ for the system (4.36) hence contractivity for the original system, we need to prove the existence of a suitable controller.

As a first step in this direction, consider the concept of control at the vertices [GC86a], precisely a piecewise-linear control $u = \Phi(x)$ that interpolates the control value at the vertices as follows:

- for any pair (x_i, u_i) of columns of X and U, respectively, $\Phi(x_i) = u_i$;
- for $x \in \mathcal{S}$, $\Phi(x) = Uz$ where vector $z \geq 0$ is such that $x = Xz$, $\bar{1}^T z = \Psi(x)$, existing in view of (4.26);
- $\Phi(x)$ is Lipschitz.

We show that such a controller renders a set positively invariant for the system (4.36). First of all note that if the constraints are satisfied at the vertices, then, by convexity, $\Phi(x) = Uz \in \mathcal{U}$. Consider any point $x \in \mathcal{S}$. The tangent cone is

$$\mathcal{T}_{\mathcal{S}}(x) = \{y = \sum_{j=1}^{r} \nu_j(x_j - x), \quad \nu_j \geq 0\}$$

To derive a simple proof we first consider the following (simple) characterization of $\mathcal{T}_{\mathcal{S}}(x)$, valid for any polytope $\mathcal{S}$,

$$\mathcal{T}_{\mathcal{S}}(x) = \{y : \ \exists\, \tau \geq 0 : \ x + \tau y \in \mathcal{S}\}$$

Then we show that the derivative vector in $x = Xz$, with $u = Uz$, as above, is $\hat{A}x + Bu \in \mathcal{T}_{\mathcal{S}}(x)$ for x on the boundary of $\mathcal{S}$. Given x and z, take $\tau > 0$ such that

$$0 \leq 1 + \tau\bar{1}^T\hat{H}z \leq 1$$

which is possible because $\bar{1}^T\hat{H} \leq 0$ and $z \geq 0$. Then

$$x + \tau(\hat{A}x + Bu) = X[I + \tau\hat{H}]z = Xz^* \in \mathcal{S}$$

where the last inclusion holds because $0 \leq \bar{1}^T z^* \leq 1$ with the chosen τ.

The last step is to show that the controller with the mentioned properties does exist. Consider the following control: given x find z such that $x = Xz$ and take, among all such z one of the minimizers of the expression $\bar{1}^T z$, so that, $\bar{1}^T z = \Psi(x)$, namely one of the elements of the set

$$\mathcal{Z}(x) = \{z : x = Xz,\ \ z \geq 0,\ \ \bar{1}^T z = \Psi(x)\}.$$

A control function can be associated with such a set-valued map as follows:

$$u = \Phi(x) = Uz, \quad \text{with } z \in \mathcal{Z}(x).$$

For each x the set $\mathcal{Z}(x)$ is a polytope and it can be shown that this set-valued map is Lipschitz continuous (see Exercise 10) and then it admits a Lipschitz selection $z(x) \in \mathcal{Z}(x)$ [AC84]. Therefore the so derived control $\Phi(x)$ is also Lipschitz[9].

In the authors' knowledge, there is no obvious and direct extension of Theorem 4.37 for the plane representation (4.25). An alternative condition can be derived by means of the conic representation of projections [DDS96]. See also [DH99] for further results on this topic.

Remark 4.38. Note that the theorem holds with no assumptions on B and so, for $B = 0$, the equivalence i) $\Longleftrightarrow$ iii) in Theorem 4.33 remains proved.

A special Lipschitz control "at the vertices" is the Gutman and Cwikel control [GC86a][10]. Any polytope $\mathcal{S}$ which includes the origin as an interior point can be partitioned into simplices, $\mathcal{S}^{(k)}$, each formed by n vertices and the origin. These simplices have zero measure intersections and their union is $\mathcal{S}$.

$$\mathcal{S}^{(k)} = \{x = \alpha_1 x_1^{(k)} + \alpha_2 x_2^{(k)} + \cdots + \alpha_n x_n^{(k)},\ \ x_i^{(k)} \in vert\{\mathcal{S}\}, \alpha_i \geq 0,\ \ \sum_i \alpha_i \leq 1\ \}$$

A convex polyhedral cone (henceforth a sector) having the origin as vertex is generated by each simplex $\mathcal{S}^{(k)}$ as follows (see Fig. 4.5):

$$\mathcal{C}^{(k)} = \{x = \gamma_1 x_1^{(k)} + \gamma_2 x_2^{(k)} + \cdots + \gamma_n x_n^{(k)},\ \ x_i^{(k)} \in vert\{\mathcal{S}\}, \gamma_1 \geq 0\}$$

Note that $\mathcal{S}^{(k)} = \mathcal{C}^{(k)} \bigcap \mathcal{S}$. The simplices $\mathcal{S}^{(k)}$ and the associated cones $\mathcal{C}^{(k)}$ can be selected in such a way that

- $\mathcal{S}^{(k)}$ and $\mathcal{C}^{(k)}$ have non-empty interior;
- $\mathcal{S}^{(k)} \bigcap \mathcal{S}^{(h)}$ and $\mathcal{C}^{(k)} \bigcap \mathcal{C}^{(h)}$ have an empty interior if $k \neq h$
- $\bigcup_k \mathcal{S}^{(k)} = \mathcal{S}$ and $\bigcup_k \mathcal{C}^{(k)} = \mathbb{R}^n$

[9] A concrete way to derive Φ will be proposed soon.

[10] Developments of such a control have been proposed in [NGOH13].

Denote by

$$X^{(h)} = [x_1^{(h)}\ x_2^{(h)}\ \dots\ x_n^{(h)}]$$

the square sub-matrix of X formed by the vertices generating $\mathcal{S}^{(k)}$ and the corresponding cone $\mathcal{C}^{(k)}$ and let

$$U^{(h)} = [u_1^{(h)}\ u_2^{(h)}\ \dots\ u_n^{(h)}]$$

be the matrix achieved by the controllers associated with these vertices. Since $\mathcal{S}^{(h)}$ has a non-empty interior, $X^{(h)}$ is invertible. In each of these sectors consider a linear gain

$$K^{(h)} \doteq U^{(h)}[X^{(h)}]^{-1},$$

resulting in the following piecewise linear control

$$u = K^{(h)}x, \quad \text{for } x \in \mathcal{C}^{(h)} \tag{4.39}$$

This control is a special case of control at the vertices in the sense that it satisfies the requirements mentioned in the proof of Theorem 4.37. This can be immediately seen as follows. Assume that sector 1 (for the remaining sectors the same considerations apply) $\mathcal{C}^{(1)}$ is generated by the first n columns of $X = [X^{(1)}\ \tilde{X}]$. Then, if $x \in \mathcal{S}^{(1)}$,

$$x = [X^{(1)}\ \tilde{X}] \begin{bmatrix} g \\ 0 \end{bmatrix},$$

where $g \in \mathbb{R}^n$ is a non-negative vector. Therefore

$$u = K^{(1)}x = U^{(1)}[X^{(1)}]^{-1}[X^{(1)}]g = U^{(1)}g$$

When $x = x_i^{(1)}$ is one of the vertices of $\mathcal{S}^{(1)}$, $g_i = 1$ and $g_j = 0$, for $j \neq i$, so that

$$u = K^{(1)}x_i^{(1)} = u_i^{(1)}$$

It is left as an exercise to show that inside $\mathcal{S}$ the control law satisfies the constraints if and only if the control values at the vertices do (obviously, outsides the constraints can be violated). The global Lipschitz continuity of this control has been proved in [Bla95].

Example. Consider the system $\dot{x} = Ax + Bu$ with

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and the constraint $u \in \mathcal{U} = \{u : \ |u| \leq 8\}$. The set $\mathcal{S} = \mathcal{V}(X)$, where

$$X = \begin{bmatrix} 3 & 0 & -4 & -3 & 0 & 4 \\ 0 & 4 & 4 & 0 & -4 & -4 \end{bmatrix},$$

together with the control matrix

$$U = [\ -8\ \ -8\ \ -4\ \ 8\ \ 8\ \ 4\ ],$$

satisfies the conditions in Theorem 4.37 with

$$H = \frac{1}{3} \begin{bmatrix} -8 & 4 & 0 & 0 & 0 & 0 \\ 0 & -6 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 6 & 0 & 0 \\ 0 & 0 & 0 & -8 & 4 & 0 \\ 0 & 0 & 0 & 0 & -6 & 0 \\ 6 & 0 & 0 & 0 & 0 & -3 \end{bmatrix}.$$

Since

$$\bar{1}^T H = \left[-\tfrac{2}{3}\ -\tfrac{2}{3}\ -1\ -\tfrac{2}{3}\ -\tfrac{2}{3}\ -1 \right] \leq -\frac{2}{3}\bar{1}^T$$

the set $\mathcal{S} = \mathcal{V}(X)$ turns out to be contractive with a speed $\beta = 2/3$. The sectors individuated by this region are depicted in Figure 4.5. Sector $\mathcal{C}^{(1)}$ is individuated by vertices x_1 and x_2, sector $\mathcal{C}^{(2)}$ by the vertices x_2 and x_3, and so on in the counterclockwise sense until $\mathcal{C}^{(6)}$ which is individuated by vertices x_6 and x_1. The control is then readily computed. For instance, the control in sector $\mathcal{C}^{(2)}$ (formed by vertices $x^{(2)}$ and $x^{(3)}$ associated with controls $u^{(2)}$ and $u^{(3)}$) is

$$u = \begin{bmatrix} -8 & -4 \end{bmatrix} \begin{bmatrix} 0 & -4 \\ 4 & 4 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The whole set of control gains is reported in the next table.

sector number	control gain
1	$[-8/3\ -2]$
2	$[-1\ -2]$
3	$[-8/3\ -11/3]$
4	$[-8/3\ -2]$
5	$[-1\ -2]$
6	$[-8/3\ -11/3]$

The symmetric version of Theorem 4.37 is the following.

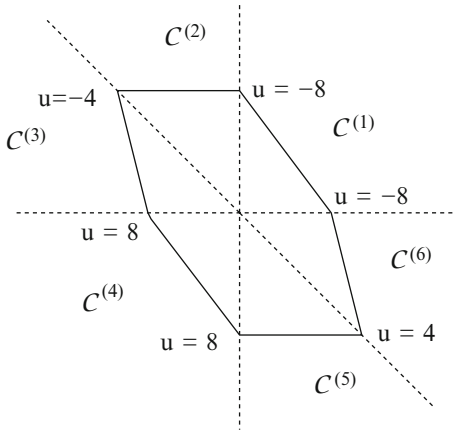


Fig. 4.5 The sector partition

Corollary 4.39. *The symmetric set $\mathcal{S} = \bar{\mathcal{V}}(X)$ as in (4.32) is β-contractive for system $\dot{x} = Ax + Bu$, $u(t) \in \mathcal{U}$, a convex set, if and only if there exist a matrix $H \in \mathbb{R}^{r\times r}$ and a matrix $U \in \mathbb{R}^{m\times r}$, having columns in $\mathcal{U}$, such that*

$$AX + BU = XH$$

and

$$\begin{bmatrix} \bar{1}^T & \bar{1}^T \end{bmatrix} \begin{bmatrix} \bar{H} & \underline{H} \\ \underline{H} & \bar{H} \end{bmatrix} \leq -\beta \begin{bmatrix} \bar{1}^T & \bar{1}^T \end{bmatrix}$$

where $\bar{H}$ and $\underline{H}$ are defined as before: $\bar{H}_{ij} = \max\{H_{ij}, 0\}$, for $i \neq j$ and $\bar{H}_{ii} = H_{ii}$, $\underline{H}_{ij} = \max\{-H_{ij}, 0\}$, for $i \neq j$ and $\underline{H}_{ii} = 0$.

The proof of the corollary is not reported, since the statement can be easily derived from the unsymmetrical version of the theorem. A more compact version of this corollary can be derived by introducing the next definition.

Definition 4.40 ($\mathcal{H}_1$-matrix). The matrix H is an $\mathcal{H}_1$-matrix if there exists τ such that

$$\|I + \tau H\|_1 < 1$$

Corollary 4.41. *The symmetric set $\mathcal{S} = \bar{\mathcal{V}}(X)$ as in (4.32) is contractive (for some $\beta > 0$) for the system $\dot{x} = Ax + Bu$, $u(t) \in \mathcal{U}$, a convex set, if and only if there exist a $\mathcal{H}_1$-matrix $H \in \mathbb{R}^{r\times r}$ and a matrix $U \in \mathbb{R}^{m\times r}$, with columns in $\mathcal{U}$, such that*

$$AX + BU = XH \tag{4.40}$$

The proof of the corollary can be immediately derived by considering the EAS and is left as an exercise to the reader (a proof can be found in [Bla00]).

We point out that, although so far much of the results have been given for polyhedral C-sets, the same properties hold, with proper assumptions, for polyhedral sets in general, which are unbounded or with empty interior. Indeed contractivity can be defined for convex and closed sets including the origin in the interior, since the Minkowski function can be defined for this class of systems. Contractivity can be defined for a convex and closed set $\mathcal{P}$ including 0, but with empty interior, if we restrict the space to the largest subspace $\mathcal{S}$ included in $\mathcal{P}$. It is not hard to see that the controlled-invariance of a polyhedron implies the controlled invariance ((A,B)-invariance) [BM92] of the smallest subspace that includes it.

For instance, the following corollary holds.

Corollary 4.42. *Consider the linear system $\dot{x}(t) = Ax(t)$. The polyhedral set $\mathcal{S} = \mathcal{P}(F)$ with $F \in \mathbb{R}^{s\times n}$ including the origin (but not necessarily bounded) is β-contractive if and only if there exists a Metzler matrix H such that*

$$HF = FA, \quad H\,\bar{1} \leq -\beta\bar{1}.$$

The polyhedral set $\mathcal{S} = \mathcal{V}(X)$ including the origin (not necessarily with a non-empty interior) is β-contractive if and only if there exists a Metzler matrix H such that

$$AX = XH, \quad \bar{1}^T H \leq -\beta\bar{1}^T$$

Finally, if we consider subspaces, these are just polyhedra with both pathologies, since they are unbounded and have empty interior. If X is the matrix that generates a subspace $\mathcal{X}$, controlled invariance is equivalent to the equation

$$AX + BU = XH$$

which is identical to (4.33). No conditions have to be imposed on H, unless we need to impose stability of the motion on $\mathcal{X}$, which in turn requires that the eigenvalues of H have negative real part.

4.5.2 Contractive sets for discrete-time systems

In this section, necessary and sufficient conditions for positive invariance of polyhedral sets for discrete-time systems are provided. The case in which there is no control action is considered first. The next theorem [Bit88] is the natural counterpart of Theorem 4.33.

Theorem 4.43. *Consider the linear system*

$$x(t+1) = Ax(t)$$

and let $\mathcal{S}$ be a polyhedral C-set of the form (4.25), with $F \in \mathbb{R}^{s\times n}$, or of the form (4.26) $X \in \mathbb{R}^{n\times r}$. The next statements are equivalent.

i) $\mathcal{S} = \mathcal{P}(F) = \mathcal{V}(X)$ *is* λ*-contractive.*

ii) There exists a matrix $P \geq 0$ *such that*

$$PF = FA, \quad P\,\bar{1} \leq \lambda\bar{1}$$

iii) There exists a matrix $P \geq 0$ *such that*

$$AX = XP, \quad \bar{1}^T P \leq \lambda\bar{1}^T$$

Proof. As for the continuous-time case, we prove just the equivalence between i and ii), since the equivalence between i) and iii) is stated later along with the generalized version of the Theorem (see remark 4.45 after Theorem 4.44).

ii) $\Longrightarrow$ i). Take $x \in \mathcal{S}$, say $Fx \leq \bar{1}$. λ-contractivity of the polyhedral set is equivalent to showing that $y = Ax$ is such that $Fy \leq \lambda\bar{1}$, say $y \in \lambda\mathcal{S}$. The above condition can be immediately derived since

$$FAx = PFx \leq P\bar{1} \leq \lambda\bar{1}$$

(the first inequality is due to the fact that P is non-negative).

i) $\Longrightarrow$ ii). Assume that $\mathcal{S}$ is contractive. This means that if $Fx \leq \bar{1}$, then $FAx \leq \bar{\lambda}1$. Consider the following linear programming problems

$$\begin{aligned} &\mu = \max F_i Ax \\ &\text{s.t.} \\ &Fx \leq \bar{1} \end{aligned}$$

$i = 1, 2, \ldots, s$. These problems are all such that $\mu \leq \lambda$, by the contractivity assumption. The ith dual problem is

$$\begin{aligned} &\mu = \min w^T\bar{1} \\ &\text{s.t.} \\ &w^T F = F_i A \\ &w^T \geq 0 \end{aligned}$$

and has the same optimal value $\mu \leq \lambda$. Define P as the matrix whose ith row P_i is a feasible solution of the ith dual problem. Clearly, P is non-negative and $PF = FA$. Since $\mu \leq \lambda$, the sum of the elements of w, hence of P_i, is less than λ, thus $P\bar{1} \leq \lambda\bar{1}$, as requested.

A simple proof of the equivalence between i) and ii) can be given by means of the inclusion property of Proposition 3.31 as shown in [DH99]. Indeed contractivity implies $A\mathcal{S} \subseteq \lambda\mathcal{S}$, thus the condition can be derived by applying that proposition.

The discrete-time counterpart of Theorem 4.37 is the following.

Theorem 4.44. *Consider the linear system*

$$x(t+1) = Ax(t) + Bu(t)$$

where $u(t) \in \mathcal{U} \subseteq \mathbb{R}^m$, with $\mathcal{U}$ a convex set. Let $\mathcal{S}$ be a polyhedral C-set of the form (4.26), with $X \in \mathbb{R}^{n\times r}$. Then $\mathcal{S} = \mathcal{V}(X)$ is λ-contractive if and only if there exist a matrix $P \geq 0$ and a matrix $U \in \mathbb{R}^{m\times r}$ such that

$$AX + BU = XP, \tag{4.41}$$

$$\bar{1}^T P \leq \lambda \bar{1}^T, \tag{4.42}$$

$$u_k \in \mathcal{U}, \tag{4.43}$$

where u_k is the kth column of U.

Proof. To prove necessity of conditions (4.41)–(4.43) assume that $\mathcal{S}$ is λ-contractive. Then for all $x_k \in vert\{\mathcal{S}\}$, there exists $u_k = \Phi(x_k)$ such that $Ax_k + Bu_k \in \lambda\mathcal{S}$, namely

$$Ax_k + Bu_k = Xz_k, \quad \bar{1}^T z_k \leq \lambda, \quad z_k \geq 0$$

Denoting by P the square matrix whose kth column is $P_k = z_k$, conditions (4.41)–(4.43) follow immediately.

To prove sufficiency, assume that (4.41)–(4.43) hold. Any $x \in \mathcal{S}$ can be written as $x = Xz$, with $\bar{1}^T z \leq 1$ and $z \geq 0$. Consider a control at the vertices as in (4.39), namely a control such that $u = Uz$, so that, denoting by $x' = Ax + Bu$,

$$x' = AXz + BUz = XPz = Xz'$$

Since P and z are non-negative and $\bar{1}^T P \leq \lambda\bar{1}^T$, then $\bar{1}^T z' = \bar{1}Pz \leq \lambda$, say $\Psi_{\mathcal{S}}(x') \leq \lambda$.

Remark 4.45. Note that the above theorem holds even when $B = 0$ which proves the equivalence between i) and iii) in Theorem 4.43.

From the proof of the theorem it is straightforward to derive the following corollary, which provides a vertex interpretation of the result.

Corollary 4.46. *The polyhedral C-set $\mathcal{S}$ is λ-contractive if and only if for each of its vertexes x there exists a control u such that $Ax + Bu \in \lambda\mathcal{S}$.*

We stress that, although in the discrete-time case there is no theoretical necessity of a continuous control $u = \Phi(x)$ [11], a Lipschitz continuous control can be derived as in (4.39).

As in the continuous-time case, the presented results can be extended to polyhedral sets which are not C-sets, in that they are unbounded or include 0 but they have empty interior. For instance, the following corollary holds.

Corollary 4.47. *Consider the linear system $x(t+1) = Ax(t)$. The polyhedral set $\mathcal{S} = \mathcal{P}(F)$ with $F \in \mathbb{R}^{s\times n}$ including the origin (but not necessarily compact) is λ-contractive if and only if there exists a matrix $P \geq 0$ such that*

$$PF = FA, \quad P\,\bar{1} \leq \lambda\bar{1}.$$

The polyhedral set $\mathcal{S} = \mathcal{V}(X)$ including the origin (not necessarily with a non-empty interior) is λ-contractive if and only if there exists a non-negative P such that

$$AX = XP, \quad \bar{1}^T P \leq \lambda\bar{1}^T$$

In the discrete-time case, there exists a version of the theorem for symmetric sets which is reported next. Let us again consider a symmetric polyhedral C-sets of the form (3.24) and (3.25)

$$\bar{\mathcal{P}}(F) = \{x : \ \|Fx\|_\infty \leq 1\} \tag{4.44}$$

and its dual

$$\bar{\mathcal{V}}(X) = \{x = Xz, \ \ \|z\|_1 \leq 1\} \tag{4.45}$$

Then $\mathcal{S} = \bar{\mathcal{P}}(F)$ is λ-contractive for the uncontrolled system $x(t+1) = Ax(t)$ if and only if there exists a matrix P such that

$$PF = FA, \quad \text{with } \|P\|_\infty \leq \lambda;$$

if and only if there exists a matrix P such that

$$AX = XP, \quad \text{with } \|P\|_1 \leq \lambda.$$

Furthermore, if we consider the system $x(t+1) = Ax(t) + Bu(t)$, contractivity is equivalent to the existence of X and U (whose columns are in $\mathcal{U}$ in the constrained case) such that

$$AX + BU = XP, \quad \text{with } \|P\|_1 \leq \lambda.$$

[11] Although discontinuous controls may cause chattering and thus can be undesirable.

4.5.3 Associating a control with a polyhedral control Lyapunov function and smoothing

The conditions presented for polyhedral Lyapunov functions are necessary and sufficient, however they do not explicitly provide the control action. A possible control has been proposed in (4.39) which has the trouble that, as it will be seen later, its complexity in terms of number of gains, can be even greater than that of the generating polyhedral function.

In the discrete-time case, complexity can be somehow reduced by considering a different type of control which is based on an on-line-optimization. Assume one is given a discrete-time control Lyapunov function $\Psi(x)$ which must be such that

$$\Psi(Ax + Bu) \leq \lambda \Psi(x),$$

for some appropriate control $u = \Phi(x) \in \mathcal{U}$. If Ψ is the Minkowski function of $\mathcal{P}(F) = \mathcal{N}[\Psi, 1]$, then

$$\Psi(x) = \min\{\xi \geq 0 : Fx \leq \xi \bar{1}\}$$

and therefore the previous condition can be equivalently stated by saying that u must be taken inside the regulation map (see Subsection 2.4.1).

$$u \in \Omega(x) = \{v : F(Ax + Bv) \leq \lambda \Psi(x) \bar{1}, \quad v \in \mathcal{U}\} \tag{4.46}$$

Therefore, the problem reduces to the on-line choice of a proper control u. It is immediate to see that one can optimize the contractivity by adopting as control $u(x)$ the maximizer $\hat{u}$ of the following optimization problem

$$(\hat{u}, \hat{\xi}) = \arg\min \quad \{\xi \geq 0 : F(Ax + Bu) \leq \xi \bar{1}, \ \ u \in \mathcal{U}\}. \tag{4.47}$$

By construction, the optimal value is upper bounded by $\lambda \Psi(x)$. Other possible optimization criteria, such as the minimum-effort criterion, are also clearly possible, as evidenced next:

$$u = \arg\min_{u} \{\|u\| : \ \ u \in \Omega(x)\}$$

It can be shown that this control, named minimal selection [AC84], is Lipschitz continuous. This construction is particularly convenient for single input systems [BMM95]. Indeed the set $\Omega(x)$

$$\Omega(x) := \{u \in \mathbb{R} : F_k B u \leq -F_k A x + \Psi(x)\lambda, \ \ k = 1, 2, \ldots, r\}$$

turns out to be the interval

$$\Omega(x) = \{u : \ \alpha(x) \leq u \leq \beta(x)\}$$

where

$$\alpha(x) \doteq \max_{k:F_kB<0} \frac{-F_kAx + \psi(x)\lambda}{F_kB}, \quad \text{and} \quad \beta(x) \doteq \min_{k:F_kB>0} \frac{-F_kAx + \psi(x)\lambda}{F_kB}$$

The minimum effort control is

$$\Phi_{ME} = \begin{cases} \alpha(x) & \text{if } \alpha(x) > 0, \\ \beta(x) & \text{if } \beta(x) < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Unfortunately, the continuous-time case is a different story, since we must take care about regularity of the control. The piecewise linear control (4.39) is globally Lipschitz but, as noticed, it has high complexity. It works reasonably well for systems with low dimensions.

One possibility is the following. We have seen that if a set is contractive for the EAS (that can be used to compute the function as we will see) it is contractive for the continuous-time system. Then one should

- compute a Lipschitz control (for instance, $\Phi_{ME}(x)$) using the EAS;
- apply it continually (in practice using a sampling time: $T << \tau$!).

The mentioned control is appropriate, since it is Lipschitz, and for sufficiently small T assures convergence.

The problem is that, being the polyhedral functions non-smooth, the gradient based controller is not suitable. A way to proceed is to use the smoothing procedure already considered in Chapter 3, Section 3.4. Consider, for simplicity, the symmetric case. The polyhedral function $\|Fx\|_\infty$ can be approximated as $\|Fx\|_{2p}$, thus achieving a function which is smooth away from 0. For $p \to \infty$ such a function converges uniformly to the original one $\|Fx\|_\infty$ on each compact set. Now, each of the functions $\|Fx\|_{2p}$ has a gradient that can be expressed as

$$\begin{aligned} \nabla \|Fx\|_{2p} &= \left(\sum_{i=1}^{s} (F_i x)^{2p} \right)^{\frac{1}{2p}-1} \sum_{i=1}^{s} (F_i x)^{2p-1} F_i^T \\ &= \|Fx\|_{2p}^{1-2p} \, F^T G_p(x), \ x \neq 0 \end{aligned}$$

where the vector $G_p(x)$ is

$$G_p(x) = [(F_1 x)^{2p-1} \ (F_2 x)^{2p-1} \ \ldots \ (F_s x)^{2p-1}]^T.$$

The explicit expression of the gradient allows the adoption of the gradient based control

$$u(t) = -\gamma(x)\ \left(\|Fx\|_{2p}\right)^{1-2p}\ B^T F^T G_p(x), \tag{4.48}$$

which works for $\gamma(x)$ positive and large enough. In practice, the resulting control must be positively homogeneous of order 1 and then the function $\gamma(\xi x)$ must grow linearly with respect to $\xi > 0$. This means that a possible (and typical) choice of γ is

$$\gamma(x) = \gamma_0 \|Fx\|_{2p}$$

with $\gamma_0 > 0$ large enough. We will consider this control later.

As mentioned in Chapter 3, Section 3.4, this smoothing procedure can be extended to non-symmetric sets. Consider, for p integer, the function

$$\sigma_p(\xi) = \begin{cases} 0 & \text{if} \quad \xi \leq 0, \\ \xi^p & \text{if} \quad \xi > 0. \end{cases}$$

Then the approximating smoothing function for the Minkowski function of a polyhedral C-set of the form (3.17)

$$\Psi(x) = \max_i\ F_i x$$

is given by (it is not necessary to have even numbers now)

$$\Psi_p(x) \doteq \sqrt[p]{\sum_{i=1}^{s}\ \sigma_p(F_i x)}$$

The gradient of such a function can be computed as

$$\begin{aligned} \nabla \Psi_p(x) &= \left(\sum_{i=1}^{s}\ \sigma_p(F_i x)\right)^{\frac{1}{p}-1}\ \sum_{i=1}^{s}\ \sigma_{p-1}(F_i x) F_i^T \\ &= \Psi_p(x)^{1-p}\ F^T \tilde{G}_p(x) \end{aligned}$$

where

$$\tilde{G}_p(x) \doteq [\sigma_{p-1}(F_1 x)\ \sigma_{p-1}(F_2 x) \ldots \sigma_p(F_s x)]^T$$

and the gradient-based control can be applied.

In the case of ellipsoidal controlled invariant sets, we have seen that they can always be associated with a linear feedback. A natural question is whether we can

associate a linear feedback with a controlled-invariant or contractive polytope. The answer is negative and this turns out to be one of the problems encountered when dealing with polytopes.

To examine this problem, assume that a candidate controlled-invariant polyhedral C-set is given in its vertex representation $\mathcal{P} = \mathcal{V}(X)$. We can check if it is actually contractive by solving the problem

$$\min\{\lambda \geq 0 : \ AX + BU = XP, \ \bar{1}^T P \leq \bar{1}^T \lambda, \ P \geq 0\}$$

in the discrete-time case or

$$\max\{\beta \geq 0 : \ AX + BU = XH, \ \bar{1}^T P \leq -\bar{1}^T \beta, \ H_{ij} \geq 0\}$$

in the continuous-time case. Since X is given, both problems are linear programming problems in the variables U, P (or H), λ (or β), thus easily solvable. The given set $\mathcal{P}$ is contractive if the optimal value is $\lambda_{opt} < 1$ (or $\beta_{opt} > 0$).

If we wish to know if a *linear* feedback can be adopted, then a minor modification can be introduced. Precisely, the new linear constraint

$$U = KX$$

must be added with the new variable K. The problem remains an LP one. Note also that there is no restriction to the state feedback case since, if $y = Cx$, we can add the constraint

$$U = KCX$$

Unfortunately, if the LP has no useful solution (i.e., $\lambda_{opt} < 1$ or $\beta_{opt} < 0$), then one must replace X. Then the problem becomes nonlinear and hard to solve. We can conclude with the next theorem (see, for instance, [VB89, VHB88, Bla90b, Bla91, Szn93]).

Theorem 4.48. *Consider the system*

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

or the corresponding discrete-time version

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

with the output

$$y(t) = Cx(t)$$

where $u \in \mathcal{U}$ and $d(t) \in \mathcal{D}$ with $\mathcal{U}$ closed and convex including 0 and $\mathcal{D}$ a C-set. Assume that the C-set $\mathcal{P}$ is contractive. Then the set $\mathcal{K}$ of all linear gains K such that $\mathcal{P}$ can be associated with the linear control

$$u(t) = Ky(t)$$

is a convex set. Moreover, if $\mathcal{P}$ and $\mathcal{D}$ are polytopes and $\mathcal{U}$ is polyhedral and if $A(w)$ and $B(w)$ are polytopic matrices, then the set $\mathcal{K}$ of gains K is a polyhedron.

4.5.4 Existence of positively invariant polyhedral C-sets

A fundamental result in basic Lyapunov theory is that a linear time-invariant system is asymptotically stable if and only if it admits a quadratic Lyapunov function, hence a contractive ellipsoid. Therefore asymptotic stability is a necessary and sufficient condition for the existence of contractive ellipsoidal C-sets (in the uncertain system case, the condition is necessary only). As we will see later, asymptotic stability is necessary and sufficient for the existence of contractive polyhedra (even for uncertain systems). An interesting question is what can we say about invariance.

The existence of invariant ellipsoids for linear systems is equivalent to marginal stability. However, for polyhedral sets this is not the case. Indeed we have the following [Bla92]

Theorem 4.49. *The system $\dot{x}(t) = Ax(t)$ admits a polyhedral invariant C-set if and only if*

i) it is at least marginally stable;
ii) the eigenvalues on the imaginary axis are all equal to 0.

The discrete-time counterpart sounds differently

Theorem 4.50. *The system $x(t+1) = Ax(t)$ admits a polyhedral invariant C-set if and only if*

i) it is at least marginally stable;
ii) the eigenvalues on the unit circle have phases that are rational multiples of π, precisely

$$|\lambda| = 1 \Rightarrow \quad \lambda = e^{j\theta}, \quad \textit{with } \theta = \frac{p}{q} 2\pi, \textit{ for some integer } q, p$$

We sketch the proofs of the above theorems. Consider the discrete-time case first. To prove sufficiency, without restriction, apply a transformation such that

$$T^{-1}AT = \text{blockdiag}\{\Theta_1, \Theta_2, \dots \Theta_N, A_S\}$$

where Θ_i are blocks of order 1 or 2 associated with eigenvalues with unitary modulus, while A_S is a stable matrix having eigenvalues in the open unit disk. We show we can find polyhedral invariant C-sets for all the sub-systems, so that the Cartesian product is invariant. We can associate an invariant set with the stable sub-system as follows. Consider any full row rank matrix X_0 and consider the symmetric polyhedron $\bar{\mathcal{V}}(X_0) = conv\{X_0, -X_0\}$. Compute recursively the images

$$X_k = A_S X_{k-1}$$

and consider the convex hull of all the computed vectors

$$\bar{\mathcal{V}}([X_0\ X_1\ \ldots\ X_{k+1}]) = \text{conv}\{\pm X_0\ \pm X_1\ \ldots\ \pm X_k\}$$

In view of the stability X_k converges to 0 and therefore, in a finite number of steps we get

$$\bar{\mathcal{V}}([X_0\ X_1\ \ldots\ X_{k+1}]) = \bar{\mathcal{V}}([X_0\ X_1\ \ldots\ X_{k+1}])$$

which is positively invariant.

To associate a polyhedral C-set with all the blocks Θ_i, note that if they are of the first order, then they are either 1 or -1 so any symmetric interval is invariant. In the second-order case, we can assume that the block is of the form

$$\Theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

and that the rational phase condition is satisfied. Then for some k $\Theta^k = I$. Then the same set iteration proposed above starting from any $\bar{\mathcal{V}}(X_0)$ in two-dimensions will eventually produce a positively invariant polyhedron.

To prove necessity, note that for the existence of invariant polyhedral C-sets it is necessary that the system is marginally stable. Assume that a polyhedral invariant C-set exists. The intersection of such a set with the eigenspace associated with any second order block Θ_i is a polygon. But it is not difficult to see that, if the rational phase condition is not satisfied, the only invariant C-sets for Θ_i are circles, since for any vector $x \in \mathbb{R}^2$, the points $\Theta_i^k x$ are dense on a circle.

The continuous-time case can be proved along the same lines. To form an invariant set for the stable part we can use the EAS, since $I + \tau A_S$ is stable in the discrete-time sense for $\tau > 0$ small. If there is a marginally stable part, this is associated with the 0 eigenvalue, for which any interval is invariant. Necessity can be proved by noticing that marginal stability is necessary. If the marginally stable eigenvalues are imaginary, we can associate any pair of them with a block of the form

$$\Theta = \begin{bmatrix} 0 & \omega \\ -\omega & 0 \end{bmatrix}.$$

Again, the intersection of any invariant polyhedron with the corresponding subspace is a polygon. However, the only invariant C-sets for this sub-system are circles. So no polyhedral C-set can exist.

Remark 4.51. In the discrete-time case, the proposed iterative procedure applied to an asymptotically stable A provides in a finite number of steps an invariant polyhedron which is the smallest invariant set including the original set $\bar{\mathcal{V}}(X_0) = conv\{X_0, -X_0\}$.

It is not to difficult to see that these results can be immediately extended as follows.

Corollary 4.52. *For a continuous-time (resp. discrete-time) linear time-invariant system there exists a polyhedral C-set which assures a level of contractivity β (resp. λ) if and only if all the eigenvalues of A have real part less or equal to $-\beta$ (resp. modulus less or equal to λ) and all the eigenvalues for which the equality holds have degree one and null imaginary part (resp. have phases that are rational multiple of π).*

If one wishes to consider special classes of polyhedra, it turns out that some other restriction on the eigenvalues has to be made. For instance, for polyhedra which are affine transformations of the symmetric unit cube,

$$\mathcal{S} = \{x : -g \leq Fx \leq g\},$$

with g positive and F square invertible, a sufficient condition is that the eigenvalues are in the damping region $Re(\lambda) \leq -|Im(\lambda)|$, in the continuous-time case, or in the square $|Re(\lambda)| + |Im(\lambda)| \leq 1$ in the discrete-time case [Bit91, Bit88]. It is not difficult to see that this kind of conditions hold if one seeks for a region which is the linear transformation of the unit diamond

$$\mathcal{S} = \{x = Xy, \ \ \|y\|_1 \leq 1\}$$

4.5.5 Diagonal dominance and diagonal invariance

We analyze in this section a special class of linear systems which have a strong stability property: they admit the scaled 1 or ∞ norm as Lyapunov functions. This line of research is due to the work of [PV04]. Let us first introduce the following definition.

Definition 4.53 (Diagonal dominance). The square matrix A is weakly row-diagonally dominant if

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$$

for all i. It is column-diagonally dominant if its transpose is row-diagonally dominant: $|a_{jj}| \geq \sum_{i \neq j} |a_{ij}|$, for all j. Diagonal dominance is strong if the inequality is strict for all rows (columns).

Since the main focus of the present work is stability, we will be mostly interested in systems with negative diagonal entries, for which the previous inequality becomes

$$-a_{ii} \geq \sum_{j \neq i} |a_{ij}|$$

We now dualize Definition 4.40 and we introduce the $\mathcal{H}_\infty$**-matrix**

Definition 4.54 ($\mathcal{H}_\infty$-matrix). The matrix H is an $\mathcal{H}_\infty$-matrix if there exists τ such that

$$\|I + \tau H\|_\infty < 1$$

The previous definition is related to the concept of matrix measure.

Definition 4.55. Given any vector norm $\|\cdot\|$, the quantity

$$\mu(A) \doteq \lim_{h \to 0} \frac{\|I + hA\| - 1}{h}$$

is referred to as matrix measure.

The following proposition is immediate.

Proposition 4.56. *A is in the class $\mathcal{H}_\infty$ ($\mathcal{H}_1$) iff the corresponding matrix measure is negative.*

Note that in principle we may define, for any norm $*$, the $\mathcal{H}_*$ class as the class of all matrices with negative matrix measure with respect to that norm. It is also worth noticing that a 0-symmetric C-set is contractive for a linear system $\dot{x} = Ax$ if and only if the matrix measure associated with its induced norm (Minkowsi functional) is negative [KAS92].

The following proposition holds.

Proposition 4.57. *The following conditions are equivalent.*

- *A has negative diagonal entries and it is row (column) diagonally dominant.*
- *$V(x) = \|x\|_\infty$ (respectively $V(x) = \|x\|_1$) is a Lyapunov function for $\dot{x} = Ax$.*
- *A is a $\mathcal{H}_\infty$-matrix (respectively $\mathcal{H}_1$-matrix).*

Proof. Since in the case of the ∞-norm $F = I$, the equation $FA = HF$ gives $A = H$. Then the ∞-norm is Lyapunov function in view of remark 4.36. The dual property is immediate as well because in the case of the 1-norm $X = I$ and then $A = H$. Finally, the equivalence of the third property with the second one follows from Lemma 4.26.

A less restrictive requirement is that the system admits a weighted norm, or, more precisely, the weighted ∞ or 1 norm, as a Lyapunov function. Let D be a square $n \times n$ matrix whose diagonal elements are the positive coefficients d_i, say

$$D = diag\{d_1, d_2, \ldots, d_n\}, \quad d_i > 0 \tag{4.49}$$

and consider the weighted norm

$$\|x\|_{\infty,D} = \|Dx\|_\infty$$

or

$$\|x\|_{1,D} = \|Dx\|_1$$

so that the diagonal matrix $X = D^{-1}$ or $F = D$ are the matrices describing the unit ball. By resorting to the weighted norms, it is possible to prove the next proposition, whose proof is left as an exercise:

Proposition 4.58. *The weighted norm $\|x\|_{\infty,D}$ (resp. $\|x\|_{1,D}$) is a Lyapunov functions if and only if D as in (4.49) is such that DAD^{-1} is diagonally row (resp. column) dominant with negative diagonal entries.*

A more intriguing fact is that in some cases we cannot rely on strict diagonal dominance, namely some of the inequalities in definition 4.53 are not strict. Clearly in this case it is not possible in general to claim that the system admits a Lyapunov function in the strong sense since it may happen that the contractivity factor β is zero. For instance, the following system

$$\begin{bmatrix} -\alpha & \alpha \\ \delta & -(\delta + \gamma) \end{bmatrix}$$

with positive α and δ and non-negative γ is just weakly row diagonally dominant and if we take $\gamma = 0$ the system is only marginally stable.

A question then arises: Under which conditions weak dominance implies asymptotic stability?

To reply to this question we need a definition.

Definition 4.59 (Irreducible matrix). A matrix A is irreducible if there does not exist a variable permutation such that the new system matrix has the form

$$\tilde{A} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

with A_{11} and A_{22} square submatrices.

The following classical result holds.

Theorem 4.60. *Assume that matrix A has negative diagonal coefficients and that it is row (column) weakly diagonally dominant. Assume that matrix A has at least one strictly diagonally dominant row (column) and that it is irreducible. Then the system $\dot{x} = Ax$ is asymptotically stable and admits a weighted $\infty(1)$-norm as a Lyapunov function.*

Proof. We prove that under these assumptions, a diagonal transformation exists which renders the systems strictly diagonally dominant. Assume that there are $m-1$, $m > 1$ strictly row-dominant diagonal coefficients and assume that these are the first $m-1$. We proceed by induction to show that it is possible to find a diagonal transformation such that there will be m strictly dominant diagonal coefficients and, by iterating the reasoning, we will arrive to a fully dominant matrix. The matrix can be written as follows:

$$\left[\begin{array}{cccc|ccc} \hat{a}_{11} & a_{12} & \dots & a_{1,m-1} & a_{1,m} & \dots & a_{1,n} \\ a_{21} & \hat{a}_{22} & \dots & a_{2,m-1} & a_{2,m} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ a_{m-1,1} & a_{m-1,2} & \dots & \hat{a}_{m-1,m-1} & a_{m-1,m} & \dots & a_{m-1,n} \\ \hline a_{m1} & a_{m2} & \dots & a_{m,m-1} & a_{m,m} & \dots & a_{m,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & \dots & a_{n,m-1} & a_{n,m} & \dots & a_{n,n} \end{array}\right]$$

(where a "hat" has been put onto all the strictly dominant coefficients). Since the matrix is irreducible, then there will be at least one of the coefficients in the lower-left part which is non-zero. Let us assume that this is the element $a_{m,1}$ and consider the transformation

$$\Lambda = diag\{\lambda, 1, 1, \dots, 1\}$$

with

$$\max_{i \in 1,\dots,m-1} \frac{\sum_j |a_{ij}|}{|a_{ii}|} < \lambda < 1 \tag{4.50}$$

Then (we underline the non-zero element $\underline{a}_{m1}$)

$$\Lambda^{-1}A\Lambda = \left[\begin{array}{cccc|ccc} \hat{a}_{11} & a_{12}/\lambda & \dots & a_{1,m-1}\lambda & a_{1,m}\lambda & \dots & a_{1,n}\lambda \\ \lambda a_{21} & \hat{a}_{22} & \dots & a_{2,m-1} & a_{2,m} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \lambda a_{m-1,1} & a_{m-1,2} & \dots & \hat{a}_{m-1,m-1} & a_{m-1,m} & \dots & a_{m-1,n} \\ \hline \lambda \underline{a}_{m1} & a_{m2} & \dots & a_{m,m-1} & \hat{a}_{m,m} & \dots & a_{m,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \lambda a_{n1} & a_{n2} & \dots & a_{n,m-1} & a_{n,m} & \dots & a_{n,n} \end{array}\right]$$

In view of (4.50), the first coefficient $\hat{a}_{11}$ maintains its strict dominance as well as all the other row-dominant diagonal entries since $\lambda < 1$. Since $\underline{a}_{m,1}$ is non-zero, the coefficient $\hat{a}_{m,m}$ becomes dominant[12], thus we obtain m strictly dominant entries. Repeating the above, it is possible to obtain n strictly diagonally dominant entries and thus, in view of Proposition 4.57, the transformed system has the ∞-norm as a Lyapunov function. Since the applied transformations are all diagonal positive and so is the overall transformation, then the original system admits a weighted ∞-norm as a Lyapunov function.

We conclude the subsection with a consideration about weak diagonal dominance: for a weakly diagonally dominant matrix, non-singularity is equivalent to Hurwitz stability, according to the following proposition (see [Alt13] and the references therein). An application of the result to the stability analysis of biochemical systems is found in [BG14].

Proposition 4.61. *Assume that matrix A satisfies, for any row i,*

$$-a_{ii} \geq \sum_{j \neq i} |a_{ij}|$$

(respectively, for any column j, $-a_{jj} \geq \sum_{i \neq j} |a_{ij}|$). Then A is Hurwitz if and only if it is non-singular.

The proof follows from the fact that diagonal dominance implies that the unit ball of the 1-norm (resp. of the ∞ norm) is positively invariant. In view of Theorem 4.49, the system has to be at least marginally stable. On the other hand, the only marginally unstable eigenvalues (if any) must be null.

4.5.6 Observability invariance and duality

A limit of the set-theoretic approach in control is that most of the results are well suited for state-feedback rather than for output feedback.

[12] Hence it is awarded a "hat."

Still a question can be immediately addressed to see if the previous results admit some dual version and which are the implication with the state estimation problem. We briefly address this issue here. It will be reconsidered later in the sections devoted to LPV and switching systems.

We remind that, given a linear system with output $y \in \mathbb{R}^p$.

$$\begin{array}{l} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{array} \qquad \text{(primal)},$$

its dual is

$$\begin{array}{l} \dot{z}(t) = A^T x(t) + C^T u(t) \\ y(t) = B^T x(t) + D^T u(t) \end{array} \qquad \text{(dual)}.$$

In a compact form, the dual is represented by matrices (A^*, B^*, C^*, D^*) as follows:

$$\left[\begin{array}{c|c} A^* & B^* \\ \hline C^* & D^* \end{array}\right] = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]^T = \left[\begin{array}{c|c} A^T & C^T \\ \hline B^T & D^T \end{array}\right]$$

We know, from the standard theory of quadratic functions, that the quadratic stabilizability conditions can be immediately dualized.

Consider, for instance, the dual conditions of (4.21)

$$A^T Q + QA + SC + C^T S^T \prec 0, \quad Q \succ 0 \tag{4.51}$$

If a solution Q exits, then denoting by

$$L = Q^{-1} S \tag{4.52}$$

one gets

$$Q(A + LC) + (A + LC)^T Q \prec 0,$$

which means that $x^T Qx$ is a Lyapunov function for $A + LC$. If we consider the observer

$$\dot{z}(t) = Az(t) + Bu(t) + L\left(Cz(t) + Du(t) - y(t)\right)$$

denoting by $e(t) = z(t) - x(t)$ the estimation error, we get

$$\dot{e}(t) = (A + LC)e(t)$$

meaning that $\mathcal{E}(Q, 1)$ is a contractive ellipsoid.

The situation is more involved in the case of polyhedral sets. To introduce this subject, with some purely mathematical caprice, we start by transposing ("dualizing") the controlled invariance conditions. For brevity we consider the symmetric continuous-time version of Corollary 4.41 (eq. (4.40)). We remind that the vertex representation $\bar{\mathcal{V}}(X)$ considered there is the dual of the plane representation $\bar{\mathcal{P}}(F)$

$$\bar{\mathcal{P}}(F) = \{x : \|Fx\|_\infty \leq 1\}$$

We need again Definition 4.54 of $\mathcal{H}_\infty$-matrix, the dual of Definition 4.40. Consider the symmetric polyhedron represented via the plane representation $\mathcal{P} = \bar{\mathcal{P}}(F)$ as in (4.31) and let us assume that

$$FA + LC = HF \tag{4.53}$$

for some $\mathcal{H}_\infty$-matrix $H \in \mathbb{R}^{s\times s}$ and some matrix $L \in \mathbb{R}^{s\times p}$. The condition is obviously the dual of that in Corollary 4.41.

Without lack of generality, let us assume from now on that the input-to-output matrix D is null[13] and consider the following system

$$\dot{z}(t) = Hz(t) - Ly(t) + FBu(t) \tag{4.54}$$

$$\hat{x}(t) = Pz(t) + Qy(t) \tag{4.55}$$

where the matrices P and Q will be defined later. It is easy to see that the variable $z(t)$ is an estimate of $Fx(t)$. To this aim define

$$\xi(t) \doteq Fx(t) - z(t)$$

so that

$$\dot{\xi} = F\dot{x} - \dot{z} = FAx + FBu - Hz + Ly - FBu = H\xi$$

Hence the system with variable $\xi(t)$ admits the $\|\cdot\|_\infty$ as a Lyapunov norm which implies that

$$\|Fx(t) - z(t)\|_\infty \to 0$$

monotonically as $t \to \infty$. Hence (4.54) looks like a state estimator. If we assume that $z(0) = 0$ and that the initial system state $x(0)$ is in a set of the form $\|Fx(0)\| \leq \mu$, then we will have $\|Fx(t) - z(t)\|_\infty \leq \mu$.

[13] The presence of a $D \neq 0$ matrix can be easily dealt with by adding some extra terms in the estimator.

We let the reader note that this property in true in general, say for any $x(0)$ and $z(0)$ such that $\|Fx(0) - z(0)\|_\infty \leq \mu$, this inequality will be satisfied in the future:

$$\|Fx(0) - z(0)\|_\infty \leq \mu \quad \Rightarrow \quad \|Fx(t) - z(t)\|_\infty \leq \mu, \; t \geq 0. \tag{4.56}$$

We call the family of sets

$$\bar{\mathcal{P}}[F, z]$$

observability-invariant as long as there exists an "observer" as in (4.54) which satisfies (4.56).

This family of sets provides a set-theoretic estimation a subject we will reconsider later in a dedicated chapter. To have a state estimation it is possible to consider (4.55). Assume that P and Q are such that

$$I - QC = PF.$$

Note that this equation is always solvable if F has full column rank. Then from (4.55) one gets

$$x(t) - \hat{x}(t) = [I - QC]x(t) - Pz(t) = P[Fx(t) - z(t)] \to 0$$

meaning that (4.55) is an asymptotic observer. This basically means that to construct an asymptotic (polyhedral) observer it is sufficient to determine a polyhedral controlled invariant set for the dual system, similarly to what is normally done in the quadratic case.

The previous equations and their dual have an interesting connection for the characterization of stabilizing compensators. For linear time-invariant systems this is mainly an academic exercise, but it will soon be shown to be of fundamental importance for the stabilization of switching and LPV systems.

Assume the pair of inequalities

$$PA^T + AP + BR + R^T B^T \prec 0, \quad P \succ 0 \tag{4.57}$$

$$A^T Q + QA + SC + C^T S^T \prec 0, \quad Q \succ 0 \tag{4.58}$$

is satisfied, so that $A + BJ$ and $A + LC$ are stable matrices with $L = Q^{-1}S$ and $J = RP^{-1}$. The well-known theory of linear systems tells us that we can achieve an observer-based compensator out of these matrices:

$$\dot{z}(t) = [A + BJ + LC]z(t) - Ly(t), \quad u(t) = +Jz(t)$$

Conversely, it is possible to show that, given a stabilizable system, the two quadratic inequalities (4.57)–(4.58) have a solution. Although this property can be deduced, quite obviously, by the detectability and stabilizability property of the system,

we wish to derive them directly (again, this will be of great help when switching and LPV systems will be dealt with).

Let

$$\dot{z}(t) = Fz(t) + Gy(t)$$
$$u(t) = Hz(t) + Ky(t)$$

be a stabilizing compensator. The closed-loop system matrix satisfies the Lyapunov inequality

$$\begin{bmatrix} A + BKC & BH \\ GC & F \end{bmatrix} \begin{bmatrix} P_1 & P_{12} \\ P_{12}^T & P_2 \end{bmatrix} + \begin{bmatrix} P_1 & P_{12} \\ P_{12}^T & P_2 \end{bmatrix} \begin{bmatrix} A + BKC & BH \\ GC & F \end{bmatrix}^T \prec 0$$

for some

$$\begin{bmatrix} P_1 & P_{12} \\ P_{12}^T & P_2 \end{bmatrix} \succ 0.$$

Since a block partitioned matrix is positive/negative definite only if its upper left block is so, it must hold that $P_1 \succ 0$, and

$$(A + BKC)P_1 + BHP_{12}^T + P_1(A + BKC)^T + P_{12}H^TB^T =$$
$$= AP_1 + P_1A^T + B(KCP_1 + HP_{12}^T) + (KCP_1 + HP_{12}^T)^TB^T \prec 0$$

Denoting by $R \doteq KCP_1 + HP_{12}^T$, the latter inequality results in

$$AP_1 + P_1A^T + BR + R^TB^T \prec 0$$

which is exactly (4.57). Condition (4.58) can be obtained dually.

At this point one may ask whether the same property holds with the polyhedral equations presented so far, say whether the existence of a stabilizing compensator is equivalent to the satisfaction of (4.40) and (4.53).

The answer is affirmative, although in this case the observer is of the generalized form and the state feedback is dynamic. The only problem is that there are no bounds on the compensator complexity, which is basically the same as that of the polyhedral invariant set which we can associate with the stable closed-loop.

Assume that (4.53) holds, for some $\mathcal{H}_\infty$-matrix $H \in \mathbb{R}^{s\times s}$ and matrix $L \in \mathbb{R}^{m\times p}$ so that a "polyhedral" observer can be found. If also the primal condition as in Corollary 4.41 is satisfied

$$AX + BU = XP$$

for some $\mathcal{H}_1$-matrix $P \in \mathbb{R}^{s\times s}$, and some full row rank X, then we can derive a linear state-feedback dynamic compensator. Let the matrix Z be such that $[X^T Z^T]^T$ is invertible and set $V \doteq ZP$. If we let

$$\left[\begin{array}{c|c} \tilde{K} & \tilde{H} \\ \hline \tilde{G} & \tilde{F} \end{array}\right] = \begin{bmatrix} U \\ V \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix}^{-1}$$

then it is readily seen that the system $\dot{x} = Ax + Bu$ equipped with the dynamic state feedback

$$\begin{aligned} \dot{z}(t) &= \tilde{F}z(t) + \tilde{G}x(t) \\ u(t) &= \tilde{H}z(t) + \tilde{K}x(t) \end{aligned}$$

leads to the closed-loop matrix

$$A_{cl} = \left[\begin{array}{c|c} A + B\tilde{K} & B\tilde{H} \\ \hline \tilde{G} & \tilde{F} \end{array}\right]$$

which satisfies the equation

$$\left[\begin{array}{c|c} A + B\tilde{K} & B\tilde{H} \\ \hline \tilde{G} & \tilde{F} \end{array}\right] \begin{bmatrix} X \\ Z \end{bmatrix} = \left[\begin{array}{c|c} A & 0 \\ \hline 0 & 0 \end{array}\right] + \left[\begin{array}{c|c} B & 0 \\ \hline 0 & I \end{array}\right] \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} X \\ Z \end{bmatrix} P$$

This is in turn equivalent to say that the closed-loop matrix A_{cl} is similar to P and hence asymptotically stable.

We leave to the reader as an exercise to prove that, if we associate this state feedback with the generalized "polyhedral" observer (i.e., by replacing x by $\hat{x}$, with $\hat{x}$ given by the generalized observer (4.54)–(4.55)) the overall compensator is stabilizing.

4.5.7 *Positive linear systems*

In this section the special case of positive linear systems is considered, since there are interesting connections with the polyhedral invariance. We limit our analysis to autonomous systems without taking into account inputs or outputs. A dynamic system is positive if $x(0) \geq 0$ implies $x(t) \geq 0$, for all $t \geq 0$ namely if it admits the positive orthant as a positively invariant set. Here, just basic facts are recalled: the interested reader is referred to specialized literature, for instance [FR00], for a comprehensive treatment of the subject and its use in many applications.

A continuous-time positive linear system is represented by the equation

$$\dot{x}(t) = Ax(t), \qquad (\text{respectively } x(t+1) = Ax(t)), \tag{4.59}$$

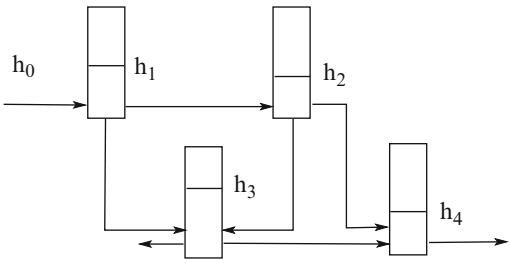


Fig. 4.6 A fluid network.

where A is a Metzler matrix in the continuous-time case and a non-negative matrix in the discrete-time case. These systems are clearly a special case of linear systems and as such enjoy all the properties presented in the previous sections. Positive systems are sometimes equipped with an input so that the model is

$$\dot{x}(t) = Ax(t) + Bu(t)$$

with B a non-negative matrix and $u(t)$ a non-negative input.

Example 4.62 (A fluid network model). Consider the fluid network depicted in Fig. 4.6. The reservoirs are connected by pipes and the flow in the pipes is caused by gravity. For each reservoir, if a linear approximation close to an equilibrium value is considered, it is possible to write an equation of the form

$$\dot{h}_i(t) = \sum_{j \in \mathcal{C}_i} \left[-\alpha_{ij}(h_i(t) - h_j(t)) - \beta_{ji} h_i(t) + \beta_{ij} h_j(t) \right]$$

where α_{ij}, β_{ij} are non-negative coefficients. The terms in the equation have the following meaning

- $\alpha_{ij}(h_i(t) - h_j(t))$ is the outgoing flow from reservoir i to reservoir j, or vice versa and it depends on the level difference (so that $\alpha_{ij} = \alpha_{ji}$);
- $\beta_{ji} h_i(t)$ is the outgoing flow from reservoir i to reservoir j and it depends only on the level $h_i(t)$;
- $\beta_{ij} h_j(t)$ is the incoming flow to reservoir i from reservoir j and it depends only on the level $h_j(t)$.

Note that we can always model an external contribution by fixing an "external" reservoir, conventionally numbered as 0 which has a constant fixed level and provides the incoming flow $\gamma_{j0} h_0(t) \geq 0$. The addition of the "external" term turns the dynamic equations into

$$\begin{aligned}
\dot{h}_1 &= -\alpha_{12}(h_1 - h_2) - \beta_{31} h_1 + \beta_{10} h_0 \\
\dot{h}_2 &= -\alpha_{21}(h_2 - h_1) - \alpha_{23}(h_2 - h_3) - \beta_{42} h_2 \\
\dot{h}_3 &= -\alpha_{34}(h_3 - h_4) + \beta_{31} h_1 - \alpha_{32}(h_3 - h_2) - \beta_{03} h_3 \\
\dot{h}_4 &= -\alpha_{43}(h_4 - h_3) + \beta_{42} h_2 - \beta_{04} h_4
\end{aligned}$$

If we assume that the network is lossless, i.e. for any link connecting two reservoirs the flow entering one is the same leaving the other, the same terms appear in pairs of equations associated with this link. Therefore, $\alpha_{21} = \alpha_{12}$, $\alpha_{23} = \alpha_{32}$, and $\alpha_{34} = \alpha_{43}$. The system is linear, with state and input matrices

$$A = \begin{bmatrix} -(\alpha_{12} + \beta_{31}) & \alpha_{12} & 0 & 0 \\ \alpha_{21} & -(\alpha_{21} + \alpha_{23} + \beta_{42}) & \alpha_{23} & 0 \\ \beta_{31} & \alpha_{32} & -(\alpha_{23} + \alpha_{34} + \beta_{03}) & \alpha_{34} \\ 0 & \beta_{42} & \alpha_{43} & -(\alpha_{43} + \beta_{04}) \end{bmatrix}$$

and

$$B = \begin{bmatrix} \beta_{10} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The next definition, which is consistent with Definition 4.59, plays a key role in the theory of positive systems.

Definition 4.63 (Irreducible positive system). A positive system of the form (4.59) is irreducible if its state matrix A is irreducible.

For example, the previous system is reducible if $\alpha_{23} = 0$. Physically this condition represents the fact that the dynamics of the third and fourth tanks has no effect on the other two.

A key aspect in the theory of positive systems is the existence of a dominant eigenvalue, both in the continuous and in the discrete-time case.

Definition 4.64 (Dominant eigenvalue). An eigenvalue λ_1 of A is said to be **dominant** in the continuous-time (resp. discrete-time) case if there are no other eigenvalues of A of greater real part (resp. of greater magnitude).

The following theorem is the famous Perron–Frobenius theorem, for which a "set-invariance" proof (under simplifying assumptions) is provided.

Theorem 4.65. *A discrete-time (resp. continuous-time) positive system has a real dominant eigenvalue, known as Perron–Frobenius eigenvalue, that is non-negative in the discrete-time case. Corresponding to the Perron–Frobenius eigenvalue, there is always an eigenvector with non-negative components, known as Perron–Frobenius eigenvector. Moreover, if the system is irreducible, the Perron–Frobenius eigenvalue is simple and the Perron–Frobenius eigenvector has positive components.*

Proof (Sketch). Let us give the proof in the discrete-time case under the assumption that matrix A is non-singular and irreducible (see [FR00] for further details). Consider the auxiliary discrete-time nonlinear system

$$z(t+1) = \varphi(z(t)) = \frac{Az(t)}{\|Az(t)\|_1}$$

It is immediately seen that this is a positive system. It is also immediate that the set of non-negative vectors with components summing up to 1

$$\mathcal{S} = \{z \geq 0 : \|z\|_1 = 1\}$$

is positively invariant: $z \in \mathcal{S} \Longrightarrow \varphi(z) \in \mathcal{S}$.

Since the function φ is continuous and $\mathcal{S}$ is convex and compact, in view of the fixed-point theorem, it admits a fixed point $\bar{z}$ which is such that

$$A\bar{z} = \|A\bar{z}\|_1 \bar{z},$$

hence $\lambda_1 = \|A\bar{z}\|_1$ is a real positive eigenvalue of A and $\bar{z}$ is a corresponding eigenvector.

The eigenvector $\bar{z}$ has clearly non-negative components since it belongs to $\mathcal{S}$. Moreover, if the matrix is irreducible, the components are positive. Indeed, assume by contradiction that there are zero components and, without restrictions, assume that they are in the last positions, namely $\bar{z}^T = [\bar{z}_1^T 0]$ where $\bar{z}_1^T > 0$. Since $\bar{z}$ is an eigenvector, then

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \bar{z}_1 \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} \bar{z}_1 \\ 0 \end{bmatrix}$$

which in turn implies that $A_{21} = 0$ (since the components of A_{21} are non-negative and those of $\bar{z}_1$ are positive), say the matrix is reducible.

We need now to prove that there are no eigenvalues with magnitude greater than $\lambda_1 = \|A\bar{z}\|_1$.

Let us introduce a diagonal state transformation $y = D^{-1}z$, characterized by the matrix

$$D = diag\{1/\bar{z}_1, 1/\bar{z}_2 \ldots, 1/\bar{z}_n\},$$

and the transformed auxiliary (linear) system

$$y(t+1) = \frac{D^{-1}AD}{\lambda_1} y(t) = By(t) \tag{4.60}$$

The vector $\bar{y} = D^{-1}\bar{z} = \bar{1}$ is an eigenvector associated with the eigenvalue $\lambda = 1$ for (4.60), since

$$D^{-1}AD\bar{y} = D^{-1}ADD^{-1}\bar{z} = D^{-1}A\bar{z} = D^{-1}\lambda_1\bar{z} = \lambda_1\bar{y}$$

Showing that the dominant eigenvalue of A is λ_1 is equivalent to showing that the dominant eigenvalue of B is 1, which is in turn a consequence of the fact that the unit ball of the infinity norm $\mathcal{B} = \mathcal{N}[\|x\|_\infty, 1]$ is positively invariant. This is immediate, since for $\|y\|_\infty \leq 1$ it holds that

$$\|By\|_\infty = \max_i \left\{ \left| \sum_j B_{ij} y_j \right| \right\} \leq \max_i \left\{ \sum_j |B_{ij} y_j| \right\} =$$
$$\max_i \left\{ \sum_j B_{ij} |y_j| \right\} \leq \max_i \left\{ \sum_j B_{ij} \right\} = 1$$

where the last equality comes from $B\bar{1} = \bar{1}$.

The continuous-time proof can be derived by considering the sampled data discrete-time system

$$x(t+1) = e^{AT} x(t)$$

with positive T and the fact that A is a Metzler matrix if and only if e^{AT} is non-negative. Hence e^{AT} has a real positive dominant eigenvalue let say $\sigma_1 = e^{\lambda_1 T}$. Since

$$|e^{\lambda_1 T}| = e^{Re(\lambda_1) T},$$

then λ_1 must have the largest real part.

In the proof we have not shown that if the matrix is irreducible then the Frobenius eigenvalue is simple. The opposite is not true: the reducible matrix $A = \text{diag}\{1; 0\}$, is a counterexample since the Frobenius eigenvalue $\lambda = 1$ is simple. Note that the Frobenius eigenvector $v = [1\ 0]^T$ does not have positive components. The reader is referred to specialized literature, e.g., [FR00].

Note that there can be more than one dominant eigenvalue. For instance, the discrete-time matrix

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

has the two dominant eigenvalues 1 and -1.

The previous theorem admits a corollary.

Corollary 4.66. *The following conditions are equivalent for a positive system.*

- *it is asymptotically stable;*
- *it admits a polyhedral Lyapunov function of the form $\|D^{-1}x\|_\infty$ with diagonal positive D;*
- *it admits a polyhedral Lyapunov function of the form $\|D^{-1}x\|_1$ with diagonal positive D.*

Proof. In discrete-time the equivalence of the first two statements follows from the proof of Theorem 4.65. In detail, the second statement implies the first. Conversely, assume A asymptotically stable and consider the modified system

$$x(t+1) = \frac{A}{\lambda}x(t)$$

with $0 < \lambda < 1$ and with λ greater than Perron–Frobenius eigenvalue. Then the modified system (which is asymptotically stable) admits the unit ball of $\|D^{-1}x\|_\infty$, for some diagonal and positive D, as an invariant set, hence the original system admits such ball as a λ-contractive set.

The equivalence of the second and third statements is due to duality.

The proof of the continuous-time case can be obtained by using the exponential matrix.

The following well-known result in the theory of positive systems admits a set-theoretic proof.

Proposition 4.67. *Assume that A is Metzler asymptotically stable matrix. Then its inverse is a non-positive matrix:* $A^{-1} \leq 0$.

Proof. Consider the system

$$\dot{x}(t) = Ax(t) + v$$

with $v \geq 0$ and A Metzler asymptotically stable. Then the solution of the system with $x(0) \geq 0$ has the property $x(t) \geq 0$. Indeed Nagumo's conditions are met with $v = 0$, because the system is positive. If we add a non-negative vector v, these are still satisfied on the boundary of the positive orthant. In fact, if $x_i = 0$ and $x_j \geq 0$, we have

$$\dot{x}_i = \sum_{j \neq i} A_{ij}x_j + v_i \geq 0.$$

Since the system is asymptotically stable, from any initial condition, $x(t)$ converges to some vector $\bar{x}(v)$. Such a vector is the solution of $A\bar{x} + v = 0$, which is unique because A is invertible

$$\bar{x}(v) = -A^{-1}v.$$

On the other hand, for any initial condition $x(0) \geq 0$, $x(t) \geq 0$ and so $x(t) \to \bar{x}(v)$ implies $\bar{x}(v) = -A^{-1}v \geq 0$. Since this condition holds for an arbitrary non-negative v, A^{-1} has to be non-positive.

There is a strong connection between the theory of positive invariance of polyhedral C-sets and the theory of positive systems, which is briefly illustrated next for discrete-time systems. Assume that the system

$$x(t+1) = Ax(t)$$

admits an invariant polytope

$$\mathcal{P} = \mathcal{V}(X) = \{x = Xp,\ \ \bar{1}^T p \leq 1,\ \ p \geq 0\}$$

Then any state $x(t)$ inside $\mathcal{P}$ can be represented as

$$x(t) = Xp(t), \quad p(t) \geq 0,\ \ \bar{1}^T p(t) \leq 1 \tag{4.61}$$

Since $\mathcal{P}$ is invariant, from iii) in Theorem 4.43 (assuming $\lambda = 1$) there must exist a non-negative P such that

$$AX = XP, \quad \bar{1}^T P \leq \bar{1}^T$$

Then for any initial condition

$$x(0) = Xp(0) \in \mathcal{P},$$

where $p(0) \geq 0$ is a (non-unique) vector such that $\bar{1}^T p(0) \leq 1$, the evolution can be described by

$$x(t) = A^t x(0) = XP^t p(0) = Xp(t),$$

where the non-negative vector $p(t)$ is such that

$$p(t+1) = Pp(t)$$

It is immediate to see that $\bar{1}^T p(t) \leq 1$, and therefore the evolution of the state inside $\mathcal{P}$ can be represented as the output of the positive system

$$\begin{aligned} p(t+1) &= Pp(t) \\ x(t) &= Xp(t) \end{aligned}$$

It is easy to see that exactly the same property holds for continuous-time systems. Clearly the positive representation is of the form

$$\dot{p}(t) = Hp(t)$$

where H comes from condition iii) in Theorem 4.33, assuming $\beta = 0$.

Note also that for a single-input single-output system (A, B, C) which admits a positively invariant polyhedral C-set $\mathcal{V}(X)$ for the free dynamics, it is always possible to find an equivalent "semi-positive" (non-minimal) realization (P, Q, R), where $P \geq 0$ is the matrix derived above, $Q \geq 0$ is a non-negative matrix such that $B = XQ$ and $R = CX$. We finally stress that the idea described here is different from the positive realization of systems with positive impulse response (see [FB04] for details).

4.6 Other classes of invariant sets and historical notes

Ellipsoids and polyhedral sets have been, undoubtedly, the most successful classes of candidate invariant sets. In particular, the ellipsoidal ones along with their quadratic functions are often the most natural choice since they arise in several contexts such as linear quadratic optimal control theory [KS72] and Lyapunov theory for linear systems [Lya66].

Quadratic functions have been used to approximate the domain of attraction of nonlinear systems [LL61, Lya66]. Indeed the theory of ellipsoidal invariant sets and quadratic functions is classical and well established. Perhaps the book [BEGFB04] is one of the best choices to provide a broad view in the modern context of LMIs. In the book [HL01] ellipsoidal sets (although not only ellipsoidal sets) are considered in the context of constrained control. In the classical book [Sch73] ellipsoids are used as a confinement region for uncertain systems. A more recent book about the properties of ellipsoidal contractive set is [PPA14]. In the context of stabilization of uncertain systems, quadratic stability has played a fundamental role. Pioneering work on this topic can be found in [HB76, Gut79, BPF83, BCL83] and several references will be given in the proper sections of the book.

Polyhedral Lyapunov functions have attracted a certain interest more recently. A motivation of their consideration is that in many control problems constraints are expressed by means of linear inequalities. The main point is that adopting polyhedra instead of ellipsoids allows a reduction of conservativeness which is paid by a higher complexity.

Besides ellipsoids and polyhedral sets, other families of sets have been considered together the associated Lyapunov functions. For instance, polynomial Lyapunov functions which have been exploited for the computation of the domain of attraction for nonlinear systems [GTV85] and for robust stability analysis [Zel94, CGTV03]. The so-called semi-ellipsoidal controlled invariant sets form an interesting class studied in [OM02].

It has been shown that a Lyapunov-type of equation, named the Minkowski–Lyapunov equation, holds for the general class of set-induced Lyapunov functions, which include polyhedral and quadratic norms as a special case [RL14].

Piecewise quadratic functions for the stability analysis of hybrid systems have been considered in [RJ98]. This class of functions generalizes the piecewise linear ones and offers a higher level of flexibility.

A generalization of quadratic functions, the so-called composite quadratic Lyapunov functions, has been recently introduced in the context [HL03] of constrained control (see also [HL01]).

The main difficulty in providing a complete review of the literature of invariant sets is that this concept arises in very different contexts. In a recent review [Bla99], to which the reader is referred to for a specific dissertation, a list of references of basic contributions concerning quadratic and polytopic invariant sets is given. The mathematical literature considered the concept of positive invariance as a basic tool for the qualitative study of dynamical systems [Bre70, Gar80, Yor67, FZ87,

Zan87, Yos75, FH76a]. The concept of positive invariance has also been recently exploited in the context of differential inclusions [AC84]. In particular, the concept of contractivity and controlled invariance considered here is related to the concept of viability [Aub91]. We point out that the theory of differential inclusions (after its digestion not so easy for an engineer) provides elegant tools to deal with some problems that unavoidably arise in control theory, such as discontinuity in the control action.

Among the first contributions concerning the condition of invariance of polyhedral sets for linear systems, we point out those due to [Bit88, VB89, BB89] for discrete-time systems and [Bit91, CH93, Bla90b, Rac91] for continuous-time systems. Other relevant references are [BG95, BG99, BT94, DH01]. Extensions of the invariance conditions considered in this chapter to systems with time delays have been given in [HT98]. Surprisingly enough, some of the positive invariance conditions were already known, although presented in the complete different context of uncertain systems dealt with polyhedral Lyapunov functions [MP86a]–[MP86c] (in particular the latest). The properties of norms, therefore including the polyhedral ones, as candidate Lyapunov functions for linear system have been analyzed in [KAS92] and [Szn93]. The proposed theory of invariance of polyhedral sets is related to the theory of componentwise stability (see [Voi84, PV04, PV06] for details). Polyhedral invariant sets will be reconsidered in several context later, where further references will be provided.

4.7 Exercises

1. Try to find analytic expressions for the sets in Fig. 4.2 (hint: the right figure is achieved by circle boundaries and, for the figure of the left, see Figure 2.9 ...). Explain why these are non-practical.
2. Disprove the assertion: the intersection of two robustly controlled invariant sets is controlled invariant (Hint: take $x(t+1) = u(t) + d(t)$, $|d| \leq 1$ and two intervals ...)
3. Show that the assumptions of compactness, convexity, and boundedness of $\mathcal{S}$ are all fundamental in Theorems 4.20 and 4.21. Precisely, show examples of invariant sets for which one of the three properties fails (but the remaining hold) and which do not include fixed or stationary points. Note that convexity can be actually replaced by the condition of being "isomorphic to a convex set."
4. Consider the system

$$\begin{aligned}
\dot{x}_1(t) &= -\alpha_1 x_1(t) + \phi_{12}(x_1(t), x_2(t)) + \cdots + \phi_{1n}(x_1(t), x_n(t)) \\
\dot{x}_2(t) &= \phi_{21}(x_2(t), x_1(t)) - \alpha_2 x_2(t) + + \cdots + \phi_{2n}(x_n(t), x_2(t)) \\
&\vdots \\
\dot{x}_n(t) &= \phi_{n1}(x_1(t), x_n(t)) + \phi_{n2}(x_2(t), x_n(t)) + \cdots + -\alpha_n x_n(t)
\end{aligned}$$

where functions ϕ_{ij} are bounded as $|\phi_{ij}(x_i, x_j)| \leq \mu_{ij}$ and $\phi_{ij}(0, x_j) = 0$. Show that the set $\mathcal{S}$ defined in Example 4.14 is invariant for the system provided that ρ is large enough.

5. Consider the competition model. Show that the ith-population-absence set $\{x : \ x_i = 0\}$ is positively invariant (so that if we set $x_i(0) = 0$ the problem is two-dimensional). Analyze the simple competition model achieved for $x_3 = 0$

$$\dot{x}_1(t) = x_1(t)\,[1 - x_1(t) - \alpha x_2(t)]$$
$$\dot{x}_2(t) = x_2(t)\,[1 - \beta x_1(t) - x_2(t)]$$

Draw the two lines $1 = x_1 - \alpha x_2$ and $1 = \beta x_1 - x_2$ and analyze the invariance of the sets delimited by these lines and the circle $\|x\| = \rho$ with $\rho > 0$ large enough, in the positive quadrant.

6. Prove the claim that if $\dot{x}(t) = Ax(t) + Bu(t)$ admits a quadratic control-Lyapunov function associated with the Lipschitz controller $u = \Phi(x)$, then it can be associated with the controller (4.19). Try to prove the same property if A is uncertain ($A(w)$ continuous and $w \in \mathcal{W}$ compact; the proof is in [BPF83]).
7. Consider a system of the form

$$\dot{x}(t) = f(x(t), w(t)), \quad w(t) \in \mathcal{W}$$

Show that if Ψ_1 and Ψ_2 are Lyapunov functions, then $\alpha\Psi_1 + \beta\Psi_2$, with $\alpha,\ \beta > 0$ is a Lyapunov function. Show that the same property does not hold for Control Lyapunov functions.

8. Prove Lemma 4.31.
9. Prove formally Proposition 4.35.
10. A non-empty set-valued map $\mathcal{Z}(x)$ with convex and compact values is Lipschitz continuous if there exists $L > 0$ such that for all x and x'

$$\mathcal{Z}(x) \subseteq \mathcal{Z}(x') + L\|x - x'\|\mathcal{B}_1$$

($\mathcal{B}_1$ is the norm unit ball so $\mathcal{Z}(x)$ is included in $\mathcal{Z}(x')$ enlarged by the ball of radius $L\|x - x'\|$). Theorem 1 in [AC84] says that any Lipschitz continuous set-valued map admits a Lipschitz continuous selection, namely a Lipschitz continuous function $z(x) \in \mathcal{Z}(x)$. Consider the following set-valued polytopic map

$$\mathcal{Z}(x) = \{z : \ Mz = v(x), \quad z \geq 0\}$$

and assume that it is not empty for each x and that $v(x)$ is a Lipschitz function. Show that it is Lipschitz and show how to find a Lipschitz selection. (Hint: the candidate vertices of the polytope $\mathcal{Z}(x)$ are linear functions of $v(x)$ ... the barycenter of the vertices is in $\mathcal{Z}(x)$ and ...)

11. Find a counterexample to the wrong claim: "$f(x) \in \mathcal{S}$ for all $x \in \partial\mathcal{S}$ implies that the C-set $\mathcal{S}$ is positively invariant for $x(t+1) = f(x(t))$" (a simple one is given in [Bla99]).
12. Prove Proposition 4.13.
13. Show that $f(x) \in \lambda\mathcal{S}$, where $\mathcal{S}$ is a C-set, for every $x \in \partial\mathcal{S}$ and $f(x)$ positively homogeneous of order one imply $\Psi_{\mathcal{S}}(x(t)) \leq \lambda^t \Psi_{\mathcal{S}}(x(0))$.
14. Show that if $\Psi_{\mathcal{S}}(x)$ is the Minkowski function of a convex C-set, then $\partial\Psi_{\mathcal{S}}(\xi x) = \partial\Psi_{\mathcal{S}}(x)$ for any $\xi > 0$ (hint: use Lemma 4.29).
15. Prove that if the C-set $\mathcal{S}$ is β-contractive for the linear system $\dot{x} = Ax + Ed$, with $d \in \mathcal{D}$ and $\mathcal{D}$ is a C-set, then the following implications hold. i) A has all eigenvalues with negative real part; ii) $-\beta$ is greater than the largest real part of the eigenvalues; iii) $\lambda\mathcal{S}$ is β-contractive for $d \in \lambda\mathcal{D}$.
16. Give an example of a control $u = Kx$ such that $\dot{x} = (A + BK)x$ is stable but K is not the gain matrix of a gradient-based control.

Chapter 5
Dynamic programming

In this section a jump back in the history of control is made and we consider problems which had been theoretically faced in the early 70s, and thereafter almost abandoned. The main reason is that the computational effort necessary to practically implement these techniques was not suitable for the computer technology of the time. Today, the situation is different, and many authors are reconsidering the approach. In this section, the main focus will be put on discrete-time systems, although it will also be shown, given the existing relation between continuous- and discrete-time invariant sets presented in Lemma 4.26, how the proposed algorithms can be used to deal with continuous-time systems as well.

5.1 Infinite-time reachability set

For an appropriate historical journey, the problem one should start with, is the one named *the state in a tube*. This is a special case of the more general category of the pursuit-evasion dynamic games [BGL69].

Problem 5.1. Consider the discrete-time system

$$x(t+1) = f(x(t), u(t), w(t))$$

where $u \in \mathcal{U}$ and $w \in \mathcal{W}$, with $\mathcal{U}$ and $\mathcal{W}$ assigned compact subsets of $\mathbb{R}^m$ and $\mathbb{R}^q$, respectively, and assume that a set $\mathcal{X} \subseteq \mathbb{R}^n$ in the state space is given. Imagine u and w are actions of two players: the "good one," P_u who chooses $u(t)$, and the "bad one" P_w, who can choose $w(t)$, and such choices have to be made at each time instant t. The goal of the good player is to assure that the following condition

$$x(t) \in \mathcal{X} \tag{5.1}$$

is satisfied for all $t \geq 0$ or for all t inside a proper interval $0, 1, \ldots T$ (as usual 0 is taken as the initial time), no matter what the "moves" of the bad player are.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_5

The problem is thought by assuming that P_w has the attitude of acting in the "worst possible way" from the point of view of player P_u. In brief, the problem is that of deciding if there exists a strategy for P_u that assures her[1] to win against any possible action of her opponent P_w. Player P_w is granted the "full information advantage," in particular he is aware of the strategy adopted by his opponent P_u. There are basically three types of strategies

Open-loop: The whole sequence $u(t)$ is decided a priori.
State feedback: $u(t) = \Phi(x(t))$ is decided based on the knowledge of $x(t)$.
Full information: $u(t) = \Phi(x(t), w(t))$ is decided based on the knowledge of $x(t)$ and $w(t)$.

Since both authors work in the control area, the open-loop strategy is immediately disregarded (we will comment on it later); only the second and third ones will be considered.

It is apparent that the condition $x(0) \in \mathcal{X}$ is not in general sufficient to assure $x(t) \in \mathcal{X}\ t > 0$. A natural way of trying to solve the above problem could be that of thinking about the natural evolution of the game. It is a well-known principle in Dynamic programming that in this way it is rather difficult to come out with a solution [Ber00]. Indeed, the right way to face the problem is not that of considering all possible future evolutions, but to analyze it backward in time. To illustrate the idea, consider the problem of controlling the motion of two vehicles to avoid collisions (see Fig. 5.1).

It is rather clear that if the vehicles are too close and/or their relative velocity is too high if the front vehicle starts braking suddenly, then an accident is most likely to occur. Without getting into the details of the second driver's response time and driving capability, let us see how it is possible to prevent an accident from occurring. Clearly we can imagine two different scenarios: the one in which there is cooperation between the two drivers and the one in which the two act independently.

Denote by y_1 and y_2 the relative positions and by $\dot{y}_1$ and $\dot{y}_2$ the corresponding velocities. For the two-vehicles system we can consider equations of the form

$$\ddot{y}_1(t) = f_1(y_1, \dot{y}_1) + u_1(t)$$
$$\ddot{y}_2(t) = f_2(y_2, \dot{y}_2) + u_2(t)$$

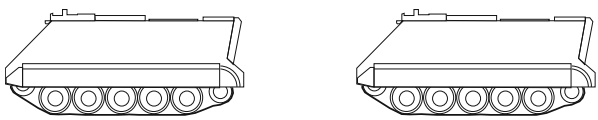


Fig. 5.1 The simplified accident setting

[1]The sexist connotation of the game is not a new idea of the authors.

where u_i are the normalized forces, subject to $|u_i| \leq \bar{u}_i$. It is possible to determine in the state space a "forbidden" region $\mathcal{F}$ which corresponds to the values of $(y_1, y_2, \dot{y}_1, \dot{y}_2)$ which could eventually lead to an accident. This region will be derived soon under some simplifications.

The first step is to define a "collision region" which is rather intuitive, in this case. Such a region $\mathcal{C}$ may be reasonably assumed to be represented by $|y_1 - y_2| < \mu$, thus the non-collision set is just the complement

$$\tilde{\mathcal{C}} = \{(y_1, y_2, \dot{y}_1, \dot{y}_2) : \ |y_1 - y_2| \geq \mu\},$$

where $\mu > 0$ is a function of the vehicle dimensions[2].

Clearly being outside the "collision region" is far from assuring that no collision will occur. Collision avoidance depends on the decision of the drivers and the initial conditions. Here u_1 and u_2 are thought as "agents" or "players of the game." Now there are two different ways to formulate the problem, precisely

- the two agents u_1 and u_2 cooperate in achieving the goal;
- the two agents u_1 and u_2 do not cooperate.

In the cooperative case there are other options. For instance, the decision can be "centralized" or "decentralized" (i.e., any agent makes her/his own decision). Different choices for the system output are possible: the positions of both vehicles, their relative position, the positions and the speeds possibly including measurement errors.

If the agents cooperate, we are dealing with a constrained control problem. We will deal with this kind of problems later on. In the non-cooperative case the situation is more difficult to handle because, perhaps in an artificial way, one has to attribute "a nature" to the two agents. In the mentioned game-theoretic framework, as already mentioned, the idea is to consider the game from the perspective of one player let us say $u = u_1$ "the good player" (the following driver) and to consider $w = u_2$ as the "opponent" (the front driver). In plain words u_2 wishes the crash to occur while u_1 would like to avoid it. This could seem unrealistic since in true life (quite often) u_2 has by no means the intention to provoke a crash. However, this setup is quite reasonable because if we find a winning strategy for u_1, we are on the safe side even in the presence of unexpected decisions of u_2, since any action of u_2 cannot have fatal consequences.

To keep things simple, let us assume $f_1 = f_2 = 0$ and let us introduce variables $x_1 = y_1 - y_2$, $\dot{x}_1 = \dot{y}_1 - \dot{y}_2$. After discretization (we used the exponential approximation) we derive a very simple discrete-time model describing the above situation:

$$\begin{aligned} x_1(t+1) &= x_1(t) + Tx_2(t) + \tfrac{1}{2}T^2(u_1(t) - u_2(t)) \\ x_2(t+1) &= x_2(t) + (u_1(t) - u_2(t)) \end{aligned}$$

[2]Including the necessary margins, an aspect that in Italy should be more seriously taken into account.

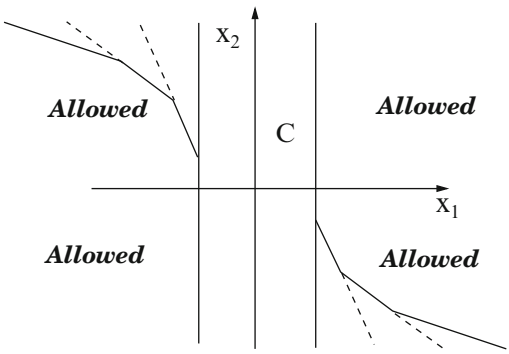


Fig. 5.2 Allowed set for simplified accident setting

Note that we could take into account a reaction delay by considering a delayed control $u_2(t-\tau_D)$. Given the above (simplified) model, it is possible to define a no-crash policy based on an "allowed region" (Fig. 5.2). Such a region is computed starting from the no-crash set $\tilde{\mathcal{C}}$, i.e. the set defined by $|x_1| \geq \mu$.

We already know that, if we do not want an accident to occur, the relative speed and velocity, at any time instant t, must not lie in the set $\mathcal{C}$, say $(x_1(t), x_2(t)) \in \tilde{\mathcal{C}}$. If one goes one step backward in time, at time $t-1$, the condition to be met to avoid a collision are a) $(x_1(t-1), x_2(t-1)) \in \tilde{\mathcal{C}}$ b) there exists $|u_1(t-1)| \leq \bar{u}_1$ such that $(x_1(t), x_2(t)) \in \tilde{\mathcal{C}}$ no matter what $|u_2(t-1)| \leq \bar{u}_2$ is. In view of the dynamic equation and the limitation on the input values, for the latter to be satisfied, the values $x_1(t-1), x_2(t-1)$ must be such that there exists a control value $u_1(t-1)$ such that

$$\begin{bmatrix} x_1(t-1) + Tx_2(t-1) + .5T^2(u_1(t-1) - u_2(t-1)) \\ x_2(t-1) + T(u_1(t-1) - u_2(t-1) \end{bmatrix} \in \tilde{\mathcal{C}}$$

for each $|u_2(t-1)| \leq \bar{u}_2$. In other words, the allowed set induces a new set of admissible states at time $t-1$. The intersection of these states and the no-crash region $\tilde{\mathcal{C}}$ is the set of *two-steps-admissible-states*, precisely states which are admissible at time $t-1$ and remain such in t. If we desire admissibility at all times, then we can iterate the algorithm backward in time. We determine the set of the states at time $t-2$ which are in the no-crash region $\tilde{\mathcal{C}}$ and will be in the two-steps-admissible-states at time $t-1$: this forms the three-steps-admissible-states. If we iterate this procedure we derive the set of all allowable states. In this case, the no-crash region is formed by the union of two disjoint convex sets represented in Fig. 5.2. The dashed lines are introduced by the algorithm that "cuts" at each step part of the k-steps-admissible-set to form the $(k+1)$-steps-admissible-set. How to construct these lines will be shown soon. It can be seen (and it is intuitive) that the allowable set is non-empty if and only if $\bar{u}_2 \leq \bar{u}_1$.

So far an "avoidance" scenario has been briefly sketched. The general idea is basically the following: consider a generic instant t and the *set of all states at the previous time* $t-1$ for which, by means of a suitable choice of $u(t-1) \in \mathcal{U}$, the condition $x(t) \in \mathcal{X}$ is satisfied for all $w(t) \in \mathcal{W}$. This set, named preimage set, is

$$Pre_{SF}(\mathcal{X}) = \{x : \text{ there exists } u \in \mathcal{U}, \text{ s.t.} f(x,u,w) \in \mathcal{X}, \text{ for all } w \in \mathcal{W}\}. \tag{5.2}$$

The notation Pre_{SF} means "state feedback" if the choice of $u \in \mathcal{U}$ is based on the knowledge of the state x alone, and thus the inclusion must be satisfied for all $w \in \mathcal{W}$. Since (5.1) has to be satisfied also at time $t-1$, also the inclusion $x(t-1) \in \mathcal{X}$ must be satisfied, say $x(t-1)$ must be such that

$$x(t-1) \in \mathcal{X}_{-1} = Pre_{SF}(\mathcal{X}) \bigcap \mathcal{X}$$

From this simple reasoning it is apparent that to solve the problem the constraint set $\mathcal{X}$ must be replaced by $Pre_{SF}(\mathcal{X}) \bigcap \mathcal{X}$, because this condition is necessary and sufficient to assure that the state remains in $\mathcal{X}$ (for a proper choice of the control value at time $t-1$) for two consecutive steps. The procedure can be obviously reiterated. If it is desired to have $x(t-2) \in \mathcal{X}$, $x(t-1) \in \mathcal{X}$ and $x(t) \in \mathcal{X}$, then it is readily seen that the following must hold true:

$$x(t-2) \in Pre_{SF}\left(\mathcal{X}_{-1}\right) \bigcap \mathcal{X}$$

and so on. The procedure, indefinitely iterated backward in time, is the following.

Procedure. Backward construction of the admissible sets.

1. Set $k = 0$ and $\mathcal{X}_0 \doteq \mathcal{X}$.
2. Define

$$\mathcal{X}_{-k-1} \doteq Pre_{SF}(\mathcal{X}_{-k}) \bigcap \mathcal{X} \tag{5.3}$$

3. Set $k = k+1$ and go to the previous step.

We point out that in the sequel we will replace the notation of

$$\mathcal{X}^{(k)} \doteq \mathcal{X}_{-k},$$

which appears more clear. Indeed the $-k$ index is introduced now just to explain the main idea of the backward construction.

In the general case, the previous procedure is not guaranteed to stop and its implementation requires the adoption of stopping criteria that will be examined later. Even when the procedure does not stop, it is possible to define the following set:

$$\bar{\mathcal{X}}_{-\infty} = \bigcap_{k=0}^{\infty} \mathcal{X}_{-k} \tag{5.4}$$

The following properties are consequences.

- The sets $\mathcal{X}_{-k}$ are ordered by inclusion

$$\mathcal{X}_{-k-1} \subseteq \mathcal{X}_{-k} \tag{5.5}$$

- The set $\mathcal{X}_{-k}$ has the property that $x(-k) \in \mathcal{X}_{-k}$ implies $x(t) \in \mathcal{X}$ for $t = -k+1, -k+2, \ldots, 0$ for all admissible sequences $w(t) \in \mathcal{W}$ provided that a proper strategy is applied and, conversely, $x(t) \notin \mathcal{X}$ implies $x(t) \notin \mathcal{X}$ for some $t = -k+1, -k+2, \ldots, 0$ and for some admissible sequence $w(t) \in \mathcal{W}$.
- If, for some k,

$$\mathcal{X}_{-k} = \mathcal{X}_{-k-1} \tag{5.6}$$

then

$$\bar{\mathcal{X}}_{-\infty} = \mathcal{X}_{-k}$$

As it will be shown later, the last property is fundamental since it allows to stop the recursive computation. Unfortunately, in general, the condition is not satisfied in a finite number of steps and stopping criteria based on numerical tolerances will be considered in specific cases.

The following results [Wit68a, BR71a, GS71, Ber72] hold true.

Theorem 5.2. *Assume that f is a continuous function, and that $\mathcal{X}$, $\mathcal{U}$, and $\mathcal{W}$ are compact sets. Then, there exists a strategy $u = \Phi(x)$ and a set of initial conditions $\mathcal{X}_{ini}$ such that, for any $x(0) \in \mathcal{X}_{ini}$, $x(t) \in \mathcal{X}$ for all $t \geq 0$ and all $w \in \mathcal{W}$ if and only if $\bar{\mathcal{X}}_{-\infty}$ is non-empty. Moreover, any initial conditions set $\mathcal{X}_{ini}$ of initial of conditions must be a subset of $\bar{\mathcal{X}}_{-\infty}$, namely*

$$\mathcal{X}_{ini} \subseteq \bar{\mathcal{X}}_{-\infty}. \tag{5.7}$$

Proof. The proof of the theorem is in [Ber72] but it is given here for completeness.

Necessity is obvious. If the intersection is empty then, for any $x(0) \in \mathcal{X}$, $x(0) \notin \mathcal{X}_{-k}$ for some k, say in at most k steps the state $x(t)$ will be outside $\mathcal{X}$ for a proper sequence $w(t) \in \mathcal{W}$.

Sufficiency It is first shown that if $\mathcal{X}$ is a closed set, then $Pre_{SF}(\mathcal{X})$ is also closed. Proving that it is closed amounts to showing that, if a sequence of vectors $x_i \in Pre_{SF}(\mathcal{X})$ converges to $\bar{x}$, then $\bar{x} \in Pre_{SF}(\mathcal{X})$.

Take one of such sequences $x_i \in Pre_{SF}(\mathcal{X})$. By definition, for all i there exists $u_i \in \mathcal{U}$ such that

$$f(x_i, u_i, w) \in \mathcal{X}$$

for all $w \in \mathcal{W}$. Since $\mathcal{U}$ is compact, it is always possible to extract from $\{u_i\}$ a subsequence $u_i^* \in \mathcal{U}$ that converges to an element $\bar{u} \in \mathcal{U}$ (clearly the corresponding

subsequence $x_i^* \to \bar{x}$). Therefore, without restriction, it can be assumed that also the original sequence converges to $\bar{u}$. For any $w \in \mathcal{W}$, by continuity,

$$y_i = f(x_i, u_i, w) \to f(\bar{x}, \bar{u}, w)$$

and, since $y_i \in \mathcal{X}$ (which is closed), $f(\bar{x}, \bar{u}, w) \in \mathcal{X}$, and then $\bar{x} \in Pre_{SF}(\mathcal{X})$ which is then closed.

Recursively, it can be readily seen that the elements of the set sequence $\mathcal{X}_{-k}$ are all closed because the intersection of closed sets is a closed set. Then $\bar{\mathcal{X}}_{-\infty}$ is closed by (5.4) (this is always true with the understanding that the empty set is closed).

The second step of the proof requires showing that, if $x \in \bar{\mathcal{X}}_{-\infty}$, then there exists $u = \Phi(x)$ such that

$$f(x, u, w) \in \bar{\mathcal{X}}_{-\infty}.$$

Since $x \in \bar{\mathcal{X}}_{-\infty}$, $x \in \mathcal{X}_{-k-1}$ for all $k \geq 0$. Then for any k there exists $u_k \in \mathcal{U}$ such that

$$f(x, u_k, w) \in \mathcal{X}_{-k},$$

for all $w \in \mathcal{W}$. Again, it is possible to extract a converging subsequence of $\{u_k\}$ $u_k \to \bar{u} \in \mathcal{U}$. Since $\mathcal{X}_{-k}$ are nested, for any arbitrary h, there exists k' such that

$$f(x, u_k, w) \in \mathcal{X}_{-h}.$$

for all $w \in \mathcal{W}$, and for $k \geq k'$. Since $\mathcal{X}_{-h}$ is closed and since, by continuity, $f(x, u_k, w) \to f(x, \bar{u}, w)$, one gets

$$f(x, \bar{u}, w) \in \mathcal{X}_{-h}.$$

for all $w \in \mathcal{W}$ and for arbitrary h, and therefore $f(x, \bar{u}, w)$ belongs to the intersection

$$f(x, \bar{u}, w) \in \bar{\mathcal{X}}_{-\infty}, \quad \text{for all } w \in \mathcal{W},$$

which is what was to be shown.

The so defined vector $\bar{u}$ is chosen as a function of $x \in \bar{\mathcal{X}}_{-\infty}$ and therefore implicitly defines a function $u(x)$ such that if $x(t) \in \bar{\mathcal{X}}_{-\infty}$ then $x(t+1) \in \bar{\mathcal{X}}_{-\infty}$, for all $w \in \mathcal{W}$. The function $u(x)$ can be formally defined as any selection $u(x) \in \Omega(x)$ where Ω is the regulation map

$$\Omega(x) = \left\{u : f(x, u, w) \in \bar{\mathcal{X}}_{-\infty}, \quad \text{for all } w \in \mathcal{W}\right\} \tag{5.8}$$

The set $x \in \bar{\mathcal{X}}_{-\infty}$ is called the infinite-time reachability tube (or infinite-time reachability set) and is the set of all the states in which the state evolution can

be confined indefinitely. This set has the property of being the largest controlled-invariant set included in $\mathcal{X}$.

The assumption of continuity of f and compactness of $\mathcal{U}$ and $\mathcal{X}$ play a fundamental role in the theorem. Indeed, if any of these is dropped, condition $\bar{\mathcal{X}}_{-\infty}$ remains necessary, but in general not sufficient (see Exercise 2).

The state feedback control case theory just reported (including procedure and theorem) works without modifications if one considers the full information control case $\Phi(x,w)$, with the only exception that the preimage set of a set $\mathcal{X}$ must be defined in a different way

$$Pre_{FI}(\mathcal{X}) = \{x : \text{ for all } w \in \mathcal{W}, \text{ there exists } u \in \mathcal{U}, \text{ s.t.} f(x,u,w) \in \mathcal{X}\} \tag{5.9}$$

From the game-theoretic point of view the new construction is based on the idea that the opponent w plays first. The infinite-time reachability set $\mathcal{X}_{-\infty}^{FI}$ can be computed in the same way explained before, as the intersection of the backward computed preimage sets. The control $u = \Phi(x,w)$ must be selected in the following regulation map

$$\Omega(x,w) = \left\{u : f(x,u,w) \in \bar{\mathcal{X}}_{-\infty}^{FI}\right\} \tag{5.10}$$

5.1.1 Linear systems with linear constraints

The backward construction procedure presented in the previous section can be efficiently particularized and used to compute the largest invariant set for linear discrete-time systems of the form

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

where $A(w)$ and $B(w)$ are as in (2.64)

$$A(w) = \sum_{i=1}^{s} A_i w_i, \quad B(w) = \sum_{i=1}^{s} B_i w_i$$

with

$$w_i \geq 0, \quad \sum_{i=1}^{s} w_i = 1$$

and where

$$d(t) \in \mathcal{D}$$

with $\mathcal{D}$ a polyhedral C-set. It is assumed that both $\mathcal{X}$ and $\mathcal{U}$ are assigned polyhedral set including the origin as an interior point:

$$\mathcal{X} = \mathcal{P}(F, g) = \{x : \ Fx \leq \bar{g}\}$$
$$\mathcal{U} = \mathcal{P}(H) = \{u : \ Hu \leq \bar{1}\}$$

Moreover, $\mathcal{X}$ is assumed to be a C-set. The procedure to generate the sequence $\mathcal{X}_{-k}$ is described next. For convenience, the notation

$$\mathcal{X}^{(k)} = \mathcal{X}_{-k}$$

is now used.

Procedure. Backward construction of controlled-invariant sets for polytopic systems.

Set $k = 0$, $F^{(k)} \doteq F$, $g^{(k)} \doteq g$, set $\mathcal{X}^{(k)} = \mathcal{P}(F^{(k)}, g^{(k)})$, fix a tolerance $\epsilon > 0$ and a maximum number of steps k_{max}.

1. Compute the erosion of the set $\mathcal{X}^{(k)} = \mathcal{P}(F^{(k)}, g^{(k)})$ w.r.t. $E\mathcal{D}$:

$$\tilde{\mathcal{P}}_{E\mathcal{D}}(F^{(k)}, \tilde{g}^{(k)}) = \{x : \ F^{(k)}(x + Ed) \leq g^{(k)}, \ \text{ for all } d \in \mathcal{D}\}$$

   This set is given by the set of inequalities $F_i^{(k)} x \leq \tilde{g}_i^{(k)}$, where

$$\tilde{g}_i^{(k)} = g_i^{(k)} - max_{d \in \mathcal{D}} \, F_i^{(k)} Ed = g_i^{(k)} - max_{d \in vert\mathcal{D}} \, F_i^{(k)} d$$

2. Expand the set $\tilde{\mathcal{P}}_{E\mathcal{D}}(F^{(k)}, \tilde{g}^{(k)})$ in the extended state-control space $\mathbb{R}^{n+m}$ as follows:

$$\mathcal{M}^{(k)} = \{(x, u) \in \mathbb{R}^{n+m} : \ u \in \mathcal{U}, \ A(w)x + B(w)u \in \mathcal{P}(F^{(k)}, \tilde{g}^{(k)}),$$
$$\text{for all } \ w \in \mathcal{W}\}$$

   This set is given by the following inequalities for (x, u)

$$F^{(k)}[A_i x + B_i u] \leq \tilde{g}^{(k)}, \quad 1 = 1, 2, \ldots, s,$$
$$Hu \leq \bar{1}$$

3. Compute set $\mathcal{R}^{(k)}$, as the preimage of $\mathcal{M}^{(k)}$, which turns out to be the projection of the set $\mathcal{M}^{(k)} \subset \mathbb{R}^{n+m}$ onto the state subspace

$$\mathcal{R}^{(k)} = Pr\left(\mathcal{M}^{(k)}\right) = \left\{x : \ \text{ there exists } u, \text{ s. t. } (x, u) \in \mathcal{M}^{(k)}\right\}. \tag{5.11}$$

4. Set

$$\mathcal{X}^{(k+1)} = \mathcal{R}^{(k)} \bigcap \mathcal{X} = \mathcal{R}^{(k)} \bigcap \mathcal{X}^{(k)}$$

5. If

$$\mathcal{X}^{(k)} \subseteq (1+\epsilon)\mathcal{X}^{(k+1)} \tag{5.12}$$

then STOP successfully;
6. If $\mathcal{X}^{(k)} = \emptyset$, then STOP unsuccessfully.
7. If $k > k_{max}$, then STOP indeterminately.
8. Set $k := k+1$ and go to Step 1.

We remind the reader that all the operations between polyhedral sets carried out in the above procedure, say erosion, intersection and projection, produce polyhedra, unfortunately of increasing complexity. Given the above, the following proposition is an obvious consequence.

Proposition 5.3. *If $\mathcal{X}$ and $\mathcal{U}$ are polyhedral sets (which is indeed our case), then the sets $\mathcal{X}^{(k)}$ are polyhedra. Also, since it has been assumed that $\mathcal{X}$ is compact, they are also compact and therefore $\mathcal{X}^{(k)}$ is a sequence of nested polytopes.*

The intersection of infinitely many polytopes is a convex set, but not a polyhedron in general. Finding a controlled-invariant polyhedron is related to the finite stopping of the procedure which is considered next.

We remind that, according to (5.6) if $\mathcal{X}^{(k)} = \mathcal{X}^{(k+1)}$ for some k, then $\mathcal{X}^{(k)} = \bar{\mathcal{X}}^{(\infty)}$. The condition (5.12) basically is the equivalent "up to a numerical tolerance." Even if this condition is met for a finite k it is not clear how to estimate such a value. Therefore the procedure must consider a maximum iteration number for practical implementation. However it is conceptually interesting to see what would happen in an ideal computation if no limits on the maximum number of iterations k_{max} were imposed (and, of course, there were no limits on the computer time and memory). The following proposition clarifies the relation between the ideal infinite recursion and the real one.

Proposition 5.4.

i) If $k_{max} = \infty$ and

$$\bar{\mathcal{X}}^{(\infty)} = \bigcap_{k=0}^{\infty} \mathcal{X}^{(k)}$$

is a C-set, then, for any tolerance $\epsilon > 0$, the procedure stops successfully in a finite number of steps.

ii) Conversely assume that $\bar{\mathcal{X}}^{(\infty)} = \emptyset$, then the procedure stops unsuccessfully in a finite number of steps.

Proof. The proof of the first statement is based on the following fact: if $\bar{\mathcal{X}}^{(\infty)}$ is a C-set then, for finite $\bar{k}$, the following condition holds:

$$\mathcal{X}^{(k)} \subseteq (1+\epsilon)\bar{\mathcal{X}}^{(\infty)}, \quad k \geq \bar{k}, \tag{5.13}$$

which implies (5.12). To prove (5.13) it is sufficient to show that it holds for some $k > 0$, since the sets $\mathcal{X}^{(k)}$ are nested. By contradiction, assume that $\mathcal{X}^{(k)} \not\subset (1+\epsilon)\bar{\mathcal{X}}^{(\infty)}$. Then there exists $x_k \in \mathcal{X}^{(k)}$, but $x_k \notin (1+\epsilon)\bar{\mathcal{X}}^{(\infty)}$. Since $x_k \in \mathcal{X}^{(k)}$, a compact set, then it is possible to extract a subsequence x_k^* that converges to a point $\bar{x}$. Since the sets are nested, then for k large enough, $x_k^* \in \mathcal{X}^{(h)}$ and, since any set $\mathcal{X}^{(h)}$ is compact, $\bar{x} \in \mathcal{X}^{(h)}$ for any h. This means that

$$\bar{x} \in \bar{\mathcal{X}}^{(\infty)}. \tag{5.14}$$

This also implies that $\bar{x} \in (1+\epsilon)\bar{\mathcal{X}}^{(\infty)}$. On the other hand, $\bar{x}$ cannot belong to the interior because $x_k^* \to \bar{x}$ and $x_k^* \notin (1+\epsilon)\bar{\mathcal{X}}^{(\infty)}$. Then $\bar{x}$ must be on the boundary

$$\bar{x} \in \partial\left\{(1+\epsilon)\bar{\mathcal{X}}^{(\infty)}\right\}. \tag{5.15}$$

Now, conditions (5.14) and (5.15) imply that $\bar{\mathcal{X}}^{(\infty)}$ is not a C-set, so the first assertion is proved.

The second assertion is immediate because the intersection of an infinite number of closed nested set $\bigcap_{i=0}^{\infty} \mathcal{X}^{(k)}$ is empty if and only if $\bigcap_{i=0}^{r} \mathcal{X}^{(k)}$ is empty for a finite r.

Remark 5.5. The procedure can be formulated in a more general framework (although only the "polytopic formulation" appears numerically reasonable) by assuming that $\mathcal{X}$, $\mathcal{U}$, and $\mathcal{D}$ are convex and closed sets and that $A(w)$ and $B(w)$ are continuous function of $w \in \mathcal{W}$, with $\mathcal{W}$ compact. Then the sequence $\mathcal{X}^{(k)}$ is formed by convex sets, which are necessarily compact if $\mathcal{X}$ is compact. Note that Proposition 5.4 is valid in the more general framework (see [Bla94] for details).

Once the final set is achieved, the regulation map is given by

$$\Omega(x) = \{u : A_i x + B_i u \in \tilde{\mathcal{X}}^{(\infty)}\} \tag{5.16}$$

where $\tilde{\mathcal{X}}^{(\infty)}$ is the erosion of $\bar{\mathcal{X}}^{(\infty)}$ with respect to $E\mathcal{D}$.

There are cases in which the procedure does not stop in a finite number of steps unless a finite k_{max} is fixed. According to the proposition, this may happen if $\bar{\mathcal{X}}^{(\infty)}$ has empty interior, but it is not empty. A very simple example is the system

$$x(k+1) = 2x(k)$$

with $\mathcal{X} = \{x : |x| \le 1\}$. The sequence of sets is

$$\mathcal{X}^{(k)} = \left\{x : |x| \le \left(\frac{1}{2}\right)^k\right\},$$

whose intersection is the origin. Thus, fixing a maximum number of steps is fundamental, besides being reasonable for obvious practical reasons.

It is also to say that, in general, the set $\bar{\mathcal{X}}^{(\infty)}$ does not necessarily include the origin even if $\mathcal{X}$, $\mathcal{U}$ and $\mathcal{D}$ do so (see Exercise 3). For these cases the stopping criterion (5.12) is not suitable. A different criterion for this case is the set distance

$$\mathcal{X}^{(k)} \subseteq \mathcal{X}^{(k+1)} + \epsilon\mathcal{B} \tag{5.17}$$

where $\mathcal{B}$ is the unit ball of any adopted norm.

In the sequel of the chapter (and of the book) other stopping criteria will be proposed. For the moment being, note that if the problem is 0-symmetric, namely all the sets $\mathcal{X}$, $\mathcal{U}$, and $\mathcal{D}$ are 0-symmetric, then all the polyhedral sets are 0-symmetric and the set $\bar{\mathcal{X}}^{(\infty)}$ is non-empty if and only if it includes the origin.

Example 5.6. Consider the problem of tracking a point moving in the space with a video-camera. The problem consists in moving the camera in such a way that the object remains visible. If we consider a planar motion for both the object and the camera (which is reasonable if the object is far enough) then we can write the following set of equations

$$\begin{aligned} \dot{y} &= v - s \\ \dot{v} &= u - f \end{aligned}$$

where $y \in \mathbb{R}^2$ is the position of the camera with respect to the object (see Fig. 5.3) s is the speed of the object, v is the speed of the camera, f is a disturbance force acting on the camera, and input u is the force which controls the camera. In this model we

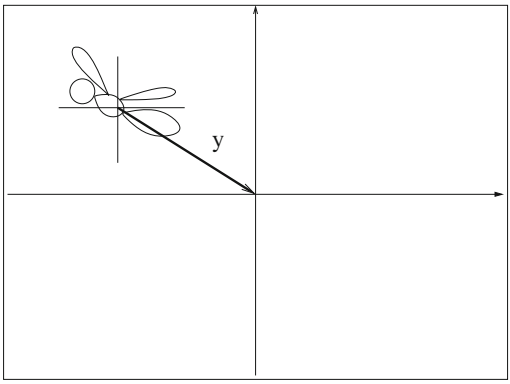


Fig. 5.3 The image tracking problem

assume that the flying object has "no inertia" so that it can vary its speed at each time without any continuity requirement[3]. In practice, the speed as well as the force acting on the camera can be bounded as follows:

$$\|s\| \leq s^{max} \quad \text{and} \quad \|f\| \leq f^{max},$$

with assigned s^{max} and f^{max}. The control action is also bounded in practice, say it can be assumed that

$$\|u\| \leq u^{max},$$

with assigned u^{max}. If the image is a rectangle with dimensions $2y_1^{max}$ and $2y_2^{max}$, respectively, then the tracking problem can be reduced to the following one.

Find a control strategy such that the center of the image y satisfies the conditions

$$|y_1| \leq y_1^{max} \quad \text{and} \quad |y_2| \leq y_2^{max},$$

respectively. Note that, by their nature, the equations are decoupled in the y_1 and the y_2 directions and moreover, if all ∞-type norms are considered, so that the components of s, f, and u are independently bounded, the problem can be split into two independent problems.

To simplify the exposition, we now consider a discrete version of the problem in which horizontal speed and acceleration are evaluated as

$$\frac{y(t+1) - y(t)}{T} = v(t) - s(t)$$
$$\frac{v(t+1) - v(t)}{T} = u(t) - f(t)$$

(where the index has been dropped). Assume $T = 1, x_1 = y, x_2 = v$ and $d = -[s\, f]^T$ so as to get the discrete-time model

$$x(t+1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} d(t)$$

Let us assume the following constraint sets

$$\mathcal{X} = \bar{\mathcal{P}}\,(I, 4) = \{x : \ \|x\|_\infty \leq 4\}, \ \ \mathcal{U} = \{u : \ |u| \leq 4\}$$

and

$$\mathcal{D} = \bar{\mathcal{P}}\,(I, 1) = \{d : \ \|d\|_\infty \leq 1\}.$$

[3] Fairly reasonable in the case of an insect.

Now the described set-theoretic procedure is able to reply with an yes–no answer to the next question

- does there exist a control strategy that solves this tracking problem?

In the case of an "yes" answer, the procedure provides the set of initial condition for which the problem can be solved.

Let us describe all the steps of the procedure in detail for this example. As the first step, the erosion of the set

$$\mathcal{X}_{E\mathcal{D}}^{(0)} = \{x: \ -3 \leq x_1 \leq 3, \quad -3 \leq x_2 \leq 3\}$$

is computed. Then, the expansion of the set $\mathcal{M}^{(1)}$ is characterized by the inequalities

$$-3 \leq x_1 + x_2 \leq 3, \quad -3 \leq x_2 + u \leq 3, \quad -4 \leq u \leq 4$$

The projection of this set on the $x_1 - x_2$ space and the subsequent intersection with $\mathcal{X}^{(0)}$ produces the first set of the sequence

$$\mathcal{X}^{(1)} = \{x: \ -4 \leq x_1 \leq 4, \quad -4 \leq x_2 \leq 4, \quad -3 \leq x_1 + x_2 \leq 3\}$$

The second step requires to compute the erosion of this set

$$\mathcal{X}_{E\mathcal{D}}^{(1)} = \{x: \ -3 \leq x_1 \leq 3, \quad -3 \leq x_2 \leq 3, \quad -1 \leq x_1 + x_2 \leq 1\}$$

The expanded set $\mathcal{M}^{(2)}$ is characterized by the next inequalities

$$-3 \leq x_1 + x_2 \leq 3, \quad -3 \leq x_2 + u \leq 3, \quad -4 \leq u \leq 4, \quad -1 \leq x_1 + 2x_2 + u \leq 1$$

Projecting and intersecting with $\mathcal{X}^{(1)}$ we achieve

$$\mathcal{X}^{(2)} = \{x: \ -4 \leq x_1 \leq 4, \ -4 \leq x_2 \leq 4, \ -3 \leq x_1 + x_2 \leq 3, \ -5 \leq x_1 + 2x_2 \leq 5\}$$

This set remains unchanged in the subsequent iteration and therefore

$$\mathcal{X}^{(3)} = \mathcal{X}^{(2)} = \bar{\mathcal{X}}^{(\infty)}$$

The computed sequence of sets is represented in Figure 5.4. Note that the new plane is a position-speed plane different from that represented in Figure 5.3, which is the image plane, therefore the original problem will have two of these "pictures," associated with the horizontal and the vertical motion, respectively, as long as the problem can be decoupled.

The regulation map for this system is derived from (5.16) and precisely

$$\Omega(x) = \{u: -\tilde{g} \leq F[Ax + Bu] \leq \tilde{g}\}$$

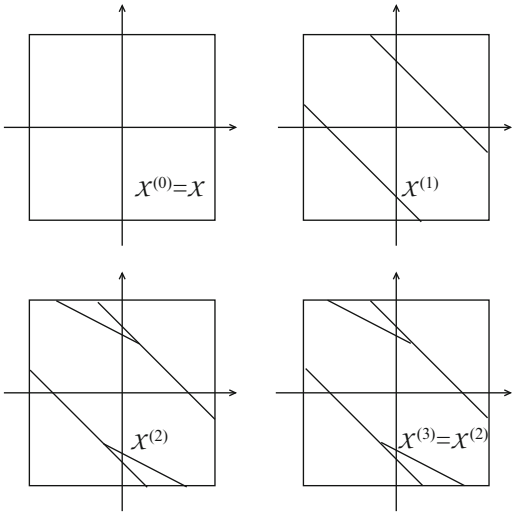


Fig. 5.4 The sequence of sets $\mathcal{X}^{(k)}$

where

$$F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \tilde{g} = \begin{bmatrix} 4 \\ 4 \\ 3 \\ 5 \end{bmatrix}$$

This is clearly a lucky case. Very simple examples of systems can be provided for which the complexity of the found sets grows rather rapidly with time. Furthermore, unless in very special cases, the set $\bar{\mathcal{X}}^{(\infty)}$ *is not finitely determined.*

Example 5.7. Consider the θ-rotation system $\dot{x} = A_\theta x$ with

$$A_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

and $\mathcal{X} = \bar{\mathcal{P}}(I) = \{x : \|x\|_\infty \leq 1\}$. The set $\mathcal{X}^{(k)}$ is given by the intersection of all the rotations of the square $\mathcal{X}$

$$\bar{\mathcal{X}}^{(\infty)} = \bigcap_{k=0}^{\infty} \bar{\mathcal{P}}\left(A_\theta^k\right) = \bigcap_{k=0}^{\infty} \bar{\mathcal{P}}\left(A_{\theta k}\right),$$

precisely the set defined by the inequalities

$$\begin{aligned} -1 \leq \cos(\theta k)x_1 - \sin(\theta k)x_2 \leq 1 \\ -1 \leq \sin(\theta k)x_1 + \cos(\theta k)x_2 \leq 1, \end{aligned}$$

for $k = 0, 1, 2, \ldots$. Simple geometrical reasoning leads to the conclusion that this set is

- a polytope if $\theta = \frac{p}{q}\pi$ is a rational multiple of π;
- the unit circle in the opposite case.

Therefore, in the latter case, even for this trivial two-dimensional example, the procedure would not stop without the adoption of the tolerance ϵ. Note also that a discontinuity phenomenon shows up: arbitrarily small variations of the parameters of the matrix (θ in the previous example) produce completely different sets (e.g., a circle instead of a polytope).

It can be shown that, even under stabilizability assumptions, the largest controlled-invariant set is not always finitely determined for a discrete-time system (see Exercise 13 of Chapter 8).

We already mentioned that the fundamental motivation for the adoption of polyhedral sets instead of sets of simpler shape, e.g., the ellipsoids, lies in the fact that the considered set operations, such as intersection or erosion, do not preserve the ellipsoidal shape. This is why (at least at the time of writing the book) the only reasonable way to compute the infinite-time reachability set is to approximate it by means of a polyhedron.

5.1.2 State in a tube: time-varying and periodic case

The results of the previous section can be easily generalized to the case of time-varying systems and target tubes. Actually, the pioneering work on this topic [BR71a, GS71] considered time-varying system. In general, one might have systems of the form

$$x(t+1) = A(t)x(t) + B(t)u(t) + E(t)d(t)$$

with time-varying constraints of the form

$$x(t) \in \hat{\mathcal{X}}_t, \quad u(t) \in \hat{\mathcal{U}}_t, \quad d(t) \in \hat{\mathcal{D}}_t$$

The sequence of set $\hat{\mathcal{X}}_t$ is called the target tube. Given the initial condition $x(\tau) \in \hat{\mathcal{X}}_\tau$, the control must assure that $x(t) \in \hat{\mathcal{X}}_t$ for $t > \tau$. To assure the condition in the future, one must assure that the state is included in a proper subset $x(\tau) \in \mathcal{X}_\tau \subseteq \hat{\mathcal{X}}_\tau$. The sequence of these subsets is called the reduced target tube.

In this framework it is practically reasonable to consider two cases:

- the finite horizon case $t = -T, -T+1, \cdots -1, 0$[4];
- the infinite-time periodic case.

[4]For convenience, negative values of time are assumed.

The backward set construction is the natural extension of that previously proposed. Fix $\mathcal{X}_0 = \hat{\mathcal{X}}_0$ and $k = 0$ and consider the following set (we go back to the original time-reversed notation):

$$\begin{aligned}\tilde{\mathcal{X}}_{-1-k} = \{x : \exists u \in \mathcal{U}_{-1-k} :\\ A(-1-k)x(-1-k) + B(-1-k)u + E(-1-k)d \in \mathcal{X}_0,\\ \forall d \in \mathcal{D}_{-1-k}\}\end{aligned}$$

Define

$$\mathcal{X}_{-1-k} = \tilde{\mathcal{X}}_{-1-k} \bigcap \hat{\mathcal{X}}_{-1-k}$$

and iterate the procedure in the same way for $k = -1, -2, \ldots$ till the end of the horizon. The sets $\mathcal{X}_{-\tau} \subseteq \hat{\mathcal{X}}_{-\tau}$ are the sets of states at time τ for which the problem of keeping the state inside $\hat{\mathcal{X}}_t$ is solvable for $t = -\tau + 1, -\tau + 2, \ldots, 0$.

Even though this procedure can, in principle, be iterated over an infinite horizon, it must be said that, for a generic sequence of time-varying data (i.e., $A(t)$, $B(t)$, $E(t)$, $\hat{\mathcal{X}}_t$, $\mathcal{U}_t$ and $\mathcal{D}_t$), it is not clear how to establish convergence criteria. There are anyhow special cases of interest for which some results can be derived.

One special case worth mentioning is the periodic one, when all the data (i.e., $A(t)$, $B(t)$, $E(t)$, $\mathcal{X}_t$, $\mathcal{U}_t$ and $\mathcal{D}_t$) are periodic with a common period T, and in this event it can be shown that, for any starting set, the sequence either collapses to an empty set or converges (unfortunately, not necessarily in finite time) to a periodic sequence of sets, the periodic target tube. For details on this problem, the interested reader is referred to [Pic93] and [BU93].

Example 5.8. Let us consider the example reported in [BU93] of a distribution system, already considered in Example 3.9 (see Fig. 3.7), subject to a periodic demand with uncertainties:

$$\begin{aligned}x_1(t+1) &= x_1(t) - \bar{f}_1(t) - d_1(t) + u_1(t) - u_2(t)\\ x_2(t+1) &= x_2(t) - \bar{f}_2(t) - d_2(t) + u_2(t)\end{aligned}$$

The period is $T = 8$. The signal $\bar{f}_1$ and $\bar{f}_2$ represent the periodic nominal values of the demand and their values are reported in table 5.1.The uncertainties are bounded as $|d_1| \leq 1$ and $|d_2| \leq 1$. We assume the bounds

$$0 \leq x_1 \leq 8, \;\; 0 \leq x_2 \leq 8, \;\; 0 \leq u_1 \leq 5, \;\; 0 \leq u_2 \leq 5.$$

Table 5.1 The periodic demand for Example 5.8

t	0	1	2	3	4	5	6	7
f_1	1	1	2	3	2	3	1	1
f_2	0	1	1	2	2	1	1	0

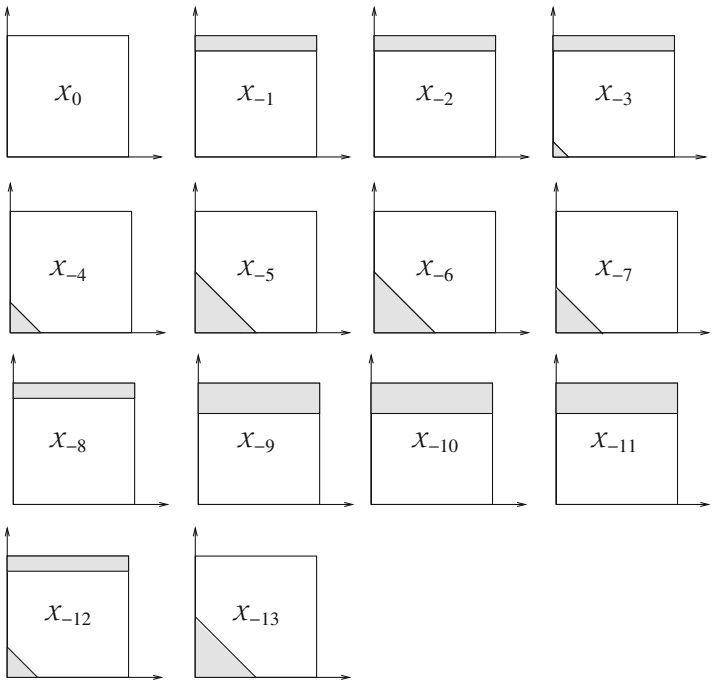


Fig. 5.5 The sequence of set converging to the periodic target tube

Although the data of the plant are all constant, the periodic nature of f makes this problem periodic. The evolution of the target tube computed backward in time is depicted in Figure 5.5. In this case the condition $\mathcal{X}_{-13} = \mathcal{X}_{-5}$ is satisfied and this implies that the periodic target tube is achieved by extending the family of sets from $\mathcal{X}_{-13}$ to $\mathcal{X}_{-6}$ by periodicity.

According to the works in [BR71a, GS71], the heavy computation of the sequence of sets can be rendered "practically" possible[5] by replacing the sequence of actual sets by a sequence of sets of a "simpler shape," typically ellipsoids. However, the consequence is that the procedure becomes conservative. To briefly examine this problem, consider the case of an ellipsoidal constraint set $\mathcal{X}^{(0)} = \mathcal{X}$. The first problem arises with the erosion: indeed the erosion of an ellipsoid is not in general an ellipsoid. Therefore, the erosion must be replaced by an approximating ellipsoidal subset:

$$\mathcal{E}^{(0)} \subseteq [\mathcal{X}_0]_{E\mathcal{D}}$$

[5] Although the extent of possible computation changed in the meanwhile.

Then the expansion of this set has to be computed (to keep the notation simple, A and B do not contain any index)

$$\{(x,u):\ Ax+Bu\in\mathcal{E}^{(0)},\ \text{ for some }\ u\in\mathcal{U}\}$$

Again, this set in not an ellipsoid (even if $\mathcal{U}$ is such), its projection on the x space is not an ellipsoid, and the subsequent intersection with $\mathcal{X}^{(0)}=\mathcal{X}$ is not an ellipsoid. Therefore, again, we must replace the computed set by an ellipsoidal subset $\mathcal{X}_{-1}$. The procedure repeats exactly in the same way. The obvious drawback is that the sequence of ellipsoids may end, due to an empty element, even if the true tube is not empty.

It is to say that this kind of procedures were originally presented in a slightly different form. The main difference with respect to the one proposed here was the computation of the preimage set, which was performed differently from (5.11). Given a set $\mathcal{X}$ the following set was preliminarily computed: $A^{-1}\mathcal{X}_{E\mathcal{D}}=\{x:Ax\in\mathcal{X}\}$. In some sense this is a "lucky" operation because, if A is invertible, the preimage of an ellipsoid is an ellipsoid, thus if the erosion is an ellipsoid (it happens, for instance, when $\mathcal{D}$ is a singleton) $A^{-1}\mathcal{X}_{E\mathcal{D}}$ is also an ellipsoid.

Then the preimage can be computed as the following sum of sets

$$Pre_{SF}(\mathcal{X})=A^{-1}\mathcal{X}-B\mathcal{U} \tag{5.18}$$

which is anyway not an ellipsoid even if $A^{-1}\mathcal{X}_{E\mathcal{D}}$ and $\mathcal{U}$ are such. There are formulas which allow the replacement of the results of these operations by included subsets of the ellipsoidal type. Specific formulas are provided in the work by [GS71, Sch73] to which the reader is referred for more details. Clearly the sum and the preimage $A^{-1}\mathcal{X}_{E\mathcal{D}}$ of polyhedra produce polyhedra and so, also in this different formulation, adopting polyhedra does not cause (but numerical) problems.

5.1.3 Historical notes and comments

As already mentioned, the ideas presented here of the backward construction of target tubes trace back to the early 70s and are due to the work by Witsenhausen, by Glover and Schweppe and by Bertsekas and Rhodes [Wit68a, BR71a, GS71, Ber72]. Many works related to the above were written in the following years, often considering sub-problems, precisely dealing only with constraints, or with disturbance rejection problems. For instance, a subsequent work [MB76] basically considered the specific case of the infinite-time reachability set for constrained linear systems. The construction of Lyapunov functions for systems proposed in [Bla94] is based on a backward mechanism as the considered one. The disturbance rejection problem has been faced in [Sha96b] where the connections with the l_1 problem ([DP87, DDB95]) have been evidenced. Related work along this lines is in [DDS96, BP98, DSDBB04, AR08, RB10]. The periodic case was considered

later by [Pic93, BU93]. The spirit of the procedure is typical of the dynamic programming. The reader is referred to the excellent work by D.P. Bertsekas [Ber00] for a deeper exposition on this matter. Many other problems can be faced with the same spirit. For instance, the well-known path selection on an oriented graph is a typical problem that can be dealt with by means of backward construction. Suppose you want to reach a target node, say node k. Then, consider the set of all nodes that are connected to node $\{k\}$ and denote $\mathcal{R}(1,\{k\})$. Then go one step backward and compute the set of all nodes which are connected to one of the elements of this set (we assume that each node is connected to itself so the new set includes the former one). These are nodes from which, in two transitions, you can reach node $\{k\}$, precisely

$$\mathcal{R}(1,\mathcal{R}(1,\{k\})) = \mathcal{R}(2,\{k\})$$

By iterating this procedure, the recursive relation

$$\mathcal{R}(k+1,\{k\}) = \mathcal{R}(1,\mathcal{R}(k,\{k\}))$$

is achieved. It is more than obvious that this procedure stops in finite time (since the set of nodes is finite). Now assign to each node $\{h\}$ the number

$$dist(h,k) = \min\{i : \{h\} \in \mathcal{R}(i,\{k\})\}.$$

Then the feedback procedure is the following: given the state in node $\{h(t)\}$, go to the node $h(t+1)$ which is at the minimal distance from $\{k\}$ among the connected nodes. It is obvious that if $\{k\}$ is the target node, the distance $dist(h(t),k)$ decreases of one unit at each transition. The function $dist(h,k)$ can be considered as a "discrete event" Lyapunov function.

5.2 Backward computation of Lyapunov functions

In this section, it is shown how to compute (control) Lyapunov functions by means of a procedure inspired by the concepts presented in the previous section. In simple words, it is shown how to construct a Lyapunov functions proceeding backward in time. In practice, the determination of a Lyapunov function proving stability or stabilizability of the system under consideration requires to find a contractive set rather than an invariant domain. Take $0 < \lambda < 1$[6] and consider a system of the form

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

with A and B continuous functions of $w \in \mathcal{W}$, with $\mathcal{W}$ compact, $d \in \mathcal{D}$, with $\mathcal{D}$ a C-set (possibly under constraints $u \in \mathcal{U}$, a closed and convex set).

[6] To avoid both singularities and trivialities it is always assumed that $\lambda \neq 0$.

According to the results in Section 4.3, a convex C-set $\mathcal{S}$ is λ-contractive for the considered system if and only if it is controlled invariant for the modified system

$$x(t+1) = \frac{A(w(t))}{\lambda}x(t) + \frac{B(w(t))}{\lambda}u(t) + \frac{E}{\lambda}d(t) \tag{5.19}$$

Therefore, to achieve a contractive set, instead of a controlled-invariant one, one can just use the backward procedure described in Subsection 5.1.1.

To this aim, a new stopping criterion for the procedure is introduced. Such a criterion can be derived by the next proposition, which basically states that it is possible to determine a contractive set in finite time if the contractivity factor λ is relaxed by an arbitrarily small amount.

Proposition 5.9. *Consider the modified system (5.19), where $u \in \mathcal{U}$ includes the origin as an interior point, and with $A(w)$ and $B(w)$ continuous functions of $w \in \mathcal{W}$, a compact set. Consider the backward construction procedure defined in Subsection 5.1.1, initialized with a (polyhedral) C-set $\mathcal{X}$. Then, if there exists a λ-contractive C-set included in $\mathcal{X}$, for any $\epsilon > 0$ there exists $\bar{k}$ such that, for $k \geq \bar{k}$, the set $\mathcal{X}^{(k)}$ (computed by means of the procedure for the modified system (5.19)) is λ^*-contractive with contractivity factor $\lambda^* = (1+\epsilon)\lambda$*

Proof. It follows from Proposition 5.4 (assuming $k_{max} = \infty$) since in a finite number of steps $\bar{k}$ condition (5.12) occurs

$$\mathcal{X}^{(\bar{k})} \subseteq (1+\epsilon)\mathcal{X}^{(\bar{k}+1)}$$

By definition, for all $x \in \mathcal{X}^{(\bar{k}+1)}$ there exists $u(x) \in \mathcal{U}$ such that

$$\frac{A(w)}{\lambda}x + \frac{B(w)}{\lambda}u(x) + \frac{E}{\lambda}d \in \mathcal{X}^{(\bar{k})} \subseteq (1+\epsilon)\mathcal{X}^{(\bar{k}+1)}$$

for all $d \in \mathcal{D}$ and $w \in \mathcal{W}$. By multiplying the leftmost and rightmost terms in the above formula by λ, one gets

$$A(w)x + B(w)u(x) + Ed \in (1+\epsilon)\lambda\mathcal{X}^{(\bar{k}+1)},$$

say $\mathcal{X}^{(\bar{k}+1)}$ is $(1+\epsilon)\lambda$-contractive.

The fact that this condition holds for all $k \geq \bar{k}$, follows from (5.13).

The sequence of sets $\mathcal{X}^{(k)}$ computed for the modified system (5.19) converges to $\mathcal{S}_\lambda$, the largest λ-contractive set included in $\mathcal{X}$. This means that if any other λ-contractive set included in $\mathcal{X}$ exists, then it is included in $\mathcal{S}_\lambda$. This follows from the fact that $\mathcal{X}^{(k)}$ converges to the largest controlled-invariant set for the modified system.

Note that the concept of "largest contractive" is well defined as it is easy to see. Indeed, given two λ-contractive sets $\mathcal{S}_\lambda^1$ and $\mathcal{S}_\lambda^2$, the convex hull of their union

$$conv\{\mathcal{S}_\lambda^2 \bigcup \mathcal{S}_\lambda^2\}$$

is λ-contractive. Therefore the λ-contractive sets included in $\mathcal{X}$ are partially ordered, thus forming what it is called a joint-semi-lattice.

The interest for the above result lies in the fact that, once the procedure has produced a λ-contractive polyhedral C-set, the corresponding Minkowski function turns out to be

- in the case $\mathcal{U} = \mathbb{R}^m$, a Lyapunov function outside $\mathcal{X}$;
- in the case $E = 0$, a Lyapunov function inside $\mathcal{X}$;
- in the case $E = 0$ and $\mathcal{U} = \mathbb{R}^m$, a global Lyapunov function.

In general nonlinear control laws can be associated with this control Lyapunov function. The class of all such controllers can be derived as $\Phi(x) \in \Omega(x)$, where Ω is the regulation map (5.16). A possible controller is the piecewise linear controlled (4.39) that can be constructed from the polyhedral set $\mathcal{S}_\lambda$.

Although discrete-time systems only have been considered so far, it is to say that the proposed procedure can also be used to compute contractive sets for continuous-time systems, in view of the existing relation between a continuous-time system and its Euler Auxiliary System (EAS). Indeed, according to Lemma 4.26, a β-contractive polyhedron for the continuous-time system

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

can be computed by determining a λ-contractive set for the Euler Auxiliary System (EAS)

$$x(t+1) = [I + \tau A(w(t))]x(t) + \tau B(w(t))u(t) + \tau Ed(t)$$

where τ is a "small" positive number. The following proposition holds true.

Proposition 5.10. *If the procedure produces a polyhedral λ-contractive set $\mathcal{S} \subseteq \mathcal{X}$ for the EAS, then $\mathcal{S}$ is a β-contractive C-set for the continuous-time system with*

$$\beta = \frac{1-\lambda}{\tau}.$$

Conversely assume that the continuous-time system admits a β-contractive C-set $\mathcal{S} \subseteq \mathcal{X}$. Then for all $0 < \beta^* < \beta$ there exists a λ^*-contractive polyhedral C-set for the EAS with

$$\lambda^* = 1 - \tau\beta^*,$$

which can be determined by the procedure, by choosing τ small enough, in a finite number of steps.

Proof. The first claim follows from Lemma 4.26. The second statement is more involved. The reader is referred to [Bla95, BM96a] for a proof.

The importance of this proposition is in that it says that whenever there exists a contractive C-set for the continuous-time system, it can be found by the procedure. The weakness is that it provides no recipe on how to choose the parameters τ and λ.

In general, the application of the procedure turns out to be the following. One chooses β and fixes τ and λ such that $\lambda = 1 - \tau\beta$, applies the procedure with a certain ϵ to find the largest λ-contractive set and waits long enough (fixing the maximum number of steps) to see if it converges. If it converges, then it produces a λ^*-contractive set with $\lambda^* = \lambda(1+\epsilon)$ which is a β^* contractive set for the continuous-time system with $\beta^* = \frac{1-\lambda^*}{\tau}$. If the procedure does not converge, then one should reduce τ or/and the convergence requirement β and restart the computation. An adaptation criterion suggested in [Bla95] is to assume $\lambda(\tau) = 1 - \rho\tau^2$, with ρ a positive fixed parameter which allows only to iterate only over τ. Note that in this way, by reducing τ, the corresponding convergence requirements decrease roughly as $\beta = (1 - \lambda(\tau))/\tau = \rho\tau$ and this is the reason of assuming a square dependence of $\lambda(\tau)$ on τ.

Example 5.11. Consider the two-dimensional dynamic system with matrices

$$A_1 = \begin{bmatrix} 0.3 & 0.5 \\ -0.6 & 0.3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.4 & 0.4 \\ -0.7 & 0.2 \end{bmatrix}, \quad B_1 = B_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and with the following constraint sets, $\mathcal{X} = \bar{\mathcal{P}}(I, 3)$, $\mathcal{U} = \bar{\mathcal{P}}(I, 1)$ and $\mathcal{D} = \bar{\mathcal{P}}(I, 1.5)$. The largest 0.9-contractive set included in $\mathcal{X}$ resulted in $\bar{\mathcal{P}}(F, 1)$, with

$$F = \begin{bmatrix} 0 & 0.3333 \\ 0.2727 & -0.1364 \\ 0.3182 & -0.0909 \\ 0.3492 & 0.0635 \end{bmatrix}$$

and is depicted in Figure 5.6. The vertices (just half of them are reported) are

$$X = \begin{bmatrix} 2.3182 & 2.9722 & 2.4444 & 2.1667 \\ 3.0000 & -0.5972 & -2.4444 & -3.0000 \end{bmatrix}.$$

5.3 The largest controlled invariant set

In this section, several topics concerning the computation of the largest controlled-invariant set (or the largest contractive sets) are discussed. The first issue is how to manage the computation of contractive sets in the presence of joint control-state

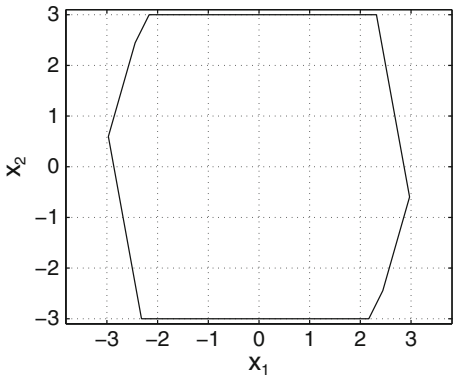


Fig. 5.6 The .9 contractive set for example 5.11

constraints. Consider the system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

and a set of the form

$$\mathcal{S}(\mu) = \{x : \|Cx + Du\| \leq \mu\} \tag{5.20}$$

For brevity we do not consider uncertainties. The above set represents a typical set of mixed constraints on both state and control action. Clearly, as special cases, the above constraints include the cases in which the constraints are separately given to input and state. To motivate this generality, a couple of cases in which this kind of constraints naturally arises is presented.

A typical case in which the joint state-control constraints have to be considered is when one wishes to consider a pre-stabilizing compensator. This means that the control is parameterized as $u = Kx + v$ where K is a given stabilizing gain and v is a new term which is typically used for performance, as we will see later. The constraint $u \in \mathcal{U}$ then becomes

$$Kx + v \in \mathcal{U}$$

which is of the considered form if, for instance $\mathcal{U}$ is the unit ball of some norm.

Another case is the presence of rate bounds. Assume that one wishes to impose constraints on $\dot{y} = C\dot{x}$. If a bound of the form $\|\dot{y}\| \leq \mu$ is assumed, then the above translates into

$$\|CAx + CBu\| \leq \mu.$$

More in general, without difficulties, it is possible to consider systems defined by constraints of the form

$$C_i x + D_i u \leq \mu_i.$$

To simplify the exposition, symmetric constraints derived by bounds on the ∞-norm are considered, which are special polyhedral sets. Furthermore, we limit the discussion to certain systems with no disturbances as inputs (the uncertain case can be easily inferred from this by using, once again, some of the previously presented results).

Definition 5.12 (Compatible set). A contractive set $\mathcal{P}$ is said to be compatible with the constraints (5.20) if one of the possible control laws $u = \Phi(x)$ is such that

$$\|Cx + D\Phi(x)\| \leq \mu$$

for all $x \in \mathcal{P}$

Note that, in general, it is not possible to talk about "inclusions" of $\mathcal{P}$ inside $\mathcal{S}(\mu)$ in the extended state-control space since the set $\mathcal{P}$ is unbounded in the "u-direction" whereas $\mathcal{S}(\mu)$ could be bounded. Clearly, if the constraints are separable in $\|\tilde{C}x\| \leq \mu$ and $\|\tilde{D}u\| \leq \mu$, then we are just talking of the largest contractive set inside the polyhedron $\{x : \|\tilde{C}x\| \leq \mu\}$ under constrained control (here we do not use $\bar{\mathcal{P}}[C, \mu]$ to avoid conflicts of notations). Consider, for instance, the scalar system

$$x(t+1) = 2x(t) + u(t) \tag{5.21}$$

with $|u| \leq 4$. For this system one can easily realize that the largest set of state which can be driven to 0 (in a sufficient number of steps) is the open interval $(-4, 4)$. For sure the closed interval $[-3, 3]$ is λ-contractive with $\lambda = 2/3$[7].

Let us consider the control $u = -2x + v$, so that

$$x(t+1) = v(t)$$

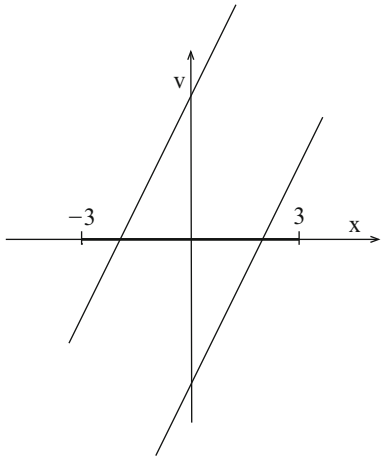
The new constraints are

$$|-2x(t) + v(t)| \leq 4.$$

In the $x - v$ plane this set is depicted as in Figure 5.7. As previously mentioned, the interval $[-3, 3]$ (represented by the thick line in Figure 5.7) is contractive and still remains such, as it can be easily checked. Indeed any state inside it can be driven to

[7]Note that, for fixed $\lambda \leq 1$, the largest contractive set is closed while the set of states that can be driven to 0 is open.

Fig. 5.7 The new constraints



the interval $[-2, 2]$ with the considered constraint. But no inclusion can be claimed with respect to the new constraint set in the joint $x - v$ space.

Compatible sets have an important property, reported in the next proposition.

Proposition 5.13. *Given two λ-contractive sets $\mathcal{P}_1$ and $\mathcal{P}_2$, compatible with $\mathcal{S}(\mu)$, the convex hull of their union is a contractive compatible with $\mathcal{S}(\mu)$.*

Proof. Take $x_1 \in \mathcal{P}_1$ and $x_2 \in \mathcal{P}_2$. By definition there exist u_1 and u_2 such that

$$Ax_1 + Bu_1 \in \lambda\mathcal{P}_1, \;\; Ax_2 + Bu_2 \in \lambda\mathcal{P}_2.$$

Take any convex combination $x = \alpha x_1 + \beta x_2$ and $u = \alpha u_1 + \beta u_2$, $\alpha + \beta = 1$, $\alpha, \beta \geq 0$, to get

$$A(\alpha x_1 + \beta x_2) + B(\alpha u_1 + \beta u_2) \in \lambda conv\{\mathcal{P}_1 \bigcup \mathcal{P}_2\}$$

and

$$\|C(\alpha x_1 + \beta x_2) + D(\alpha u_1 + \beta u_2)\| \leq \alpha\|Cx_1 + Du_1\| + \beta\|Cx_2 + Du_2\| \leq \mu.$$

Therefore, a proper control function $\Phi(x)$ can be defined by taking

$$\Phi(x) = \alpha(x)\Phi_1(x) + \beta(x)\Phi_2(x),$$

where $\Phi_1(x)$ and $\Phi_2(x)$ are the control laws associated with $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively.

Therefore it is reasonable to seek for the largest λ-contractive set compatible with the constraints (5.20). The procedure which performs such a task is essentially the same previously proposed. Consider an initial constraint set of the following form

$$\Sigma = \{(x, u) : \; F_x x + F_u u \leq g\}$$

along with the system

$$x(t+1) = Ax(t) + Bu(t) + Ed(t)$$

and $d(t) \in \mathcal{D}$, the latter being a C-set. Parametric uncertainties are not considered for brevity but they can be included without essential changes. We will comment on this later on. The next procedure extends the one previously described for the computation of the largest λ-contractive set.

Procedure. Backward construction for polytopic systems with mixed state-input constraints.

Set $k = 0$, $[F_x^{(k)}\ F_u^{(k)}] \doteq [F_x\ F_u]$. Fix a tolerance $\epsilon > 0$ and a maximum number of steps k_{max}.

1. Consider the projection of the set Σ

$$\mathcal{X}^{(k)} = Pr(\Sigma) = \{x :\ \exists u,\ \text{such that } (x,u) \in \Sigma\}$$

2. Compute $\tilde{\mathcal{X}}^{(k)}$, the erosion of the set $\mathcal{X}^{(k)}$ w.r.t. $E\mathcal{D}$:

$$\tilde{\mathcal{X}}^{(k)} = \{x :\ x + Ed \in \lambda\mathcal{X}^{(k)},\ \forall d \in \mathcal{D}\}.$$

3. Expand the set in the extended state-control space $\mathbb{R}^{n+m}$ as follows:

$$\mathcal{M}^{(k)} = \{(x,u) \in \Sigma :\ \ Ax + Bu \in \tilde{\mathcal{X}}^{(k)}\}$$

4. Compute the preimage set $Pr\left(\mathcal{M}^{(k)}\right)$, that is the projection of the set $\mathcal{M}^{(k)} \subset \mathbb{R}^{n+m}$ on the state subspace

$$\mathcal{R}^{(k)} = Pr\left(\mathcal{M}^{(k)}\right) = \left\{x :\ \ \text{there exists } u,\ \text{s. t. } (x,u) \in \mathcal{M}^{(k)}\right\}. \tag{5.22}$$

5. Set

$$\mathcal{X}^{(k+1)} = \mathcal{R}^{(k)} \bigcap \mathcal{X}^{(k)}$$

6. If

$$\mathcal{X}^{(k)} \subseteq (1+\epsilon)\mathcal{X}^{(k+1)} \tag{5.23}$$

then STOP successfully.

7. If $\mathcal{X}^{(k)} = \emptyset$, then STOP unsuccessfully.
8. If $k > k_{max}$, then STOP indeterminately.
9. Set $k := k+1$ and go to Step 1.

Note that there is no guarantee that the resulting set is compact, unless additional assumptions are introduced (for instance, that the initial projection of Σ is compact).

There is an extra condition which assures the boundedness of the resulting set. To this aim, the next well-known definition is introduced

Definition 5.14 (Transmission zeros). Given the system (A, B, C, D) if there exist a complex value σ and appropriate vectors $\bar{x}$ and $\bar{u}$, not both null, for which

$$\begin{bmatrix} A - \sigma I & B \\ C & D \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

then σ is said to be (finite) right transmission zero.

This condition is equivalent to the existence of an input and an appropriate initial condition for which the output is identically zero. It is a trivial exercise to show that this input is $\bar{u}e^{\sigma t}$ ($\bar{u}\sigma^t$ in the discrete-time case) and the initial condition is $\bar{x}$.

Proposition 5.15. *The largest controlled invariant set (or λ-contractive set) compatible with $\mathcal{S}(\mu) = \{(x, u) : \|Cx + Du\|_\infty \leq \mu\}$ is bounded if and only if the system (A, B, C, D) has no finite transmission zeros.*

Proof (If). Let $\mathcal{P}$ be a closed and convex contractive set which is 0-symmetric and compatible with $\mathcal{S}(\mu)$. This set can be represented as the sum of a compact 0-symmetric set $\bar{\mathcal{P}}$ and a subspace $\mathcal{T}$:

$$\mathcal{P} = \bar{\mathcal{P}} + \mathcal{T}$$

The "if" assertion is proved if we show that $\mathcal{T} = \{0\}$, the null subspace, in the absence of finite transmission zeros. Indeed we can see that if $\mathcal{T} \neq \{0\}$ the system admits finite transmission zeros.

The subspace $\mathcal{T}$ is necessarily (A, B) invariant. By contradiction, assume that there exists $x \in \mathcal{T}$ for which $Ax + Bu \notin \mathcal{T}$, for all u. Define

$$\xi(x) = \min_{z \in \mathcal{T}, u \in \mathbb{R}^m} \|Ax + Bu - z\| > 0$$

the minimum Euclidean distance of $Ax + Bu$ from $\mathcal{T}$. Denote by $\hat{u}$ and $\hat{z}$ the vectors which realize the minimum distance. Now consider the scaled vector ρx. Then necessarily

$$\xi(\rho x) = \min_{z \in \mathcal{T}, u \in \mathbb{R}^m} \|A\rho x + Bu - z\| = \|A\rho x + B\rho\hat{u} - \rho\hat{z}\| = |\rho|\xi$$

that is, the minimum distance $\xi(\rho x)$ of $A\rho x + Bu$ from $\mathcal{T}$ grows arbitrarily with ρ. On the other hand, if $x \in \mathcal{T}$, then $\rho x \in \mathcal{T} \subseteq \mathcal{P}$, so that $A\rho x + Bv \in \mathcal{P}$ for some v, since $\mathcal{P}$ is contractive. Then

$$A\rho x + Bv = z + w, \quad z \in \mathcal{T}, \quad w \in \bar{\mathcal{P}},$$

where w is bounded: a contradiction with the statement that $\xi(\rho x)$ grows arbitrarily.

Once we have established that $\mathcal{T}$ is (A,B)-invariant, consider a pair of vectors $\bar{x} \in \mathcal{T}$ and $\bar{u}$ such that

$$A\bar{x} + B\bar{u} \in \mathcal{T},$$

which implies

$$A\rho\bar{x} + B\rho\bar{u} \in \mathcal{T}$$

Consider now the case in which $D = 0$. In this case, we must have $\|C\bar{x}\| \leq \mu$ for any $\bar{x} \in \mathcal{T}$. Since $\mathcal{T}$ is a proper subspace, then for any non-zero element $\bar{x} \in \mathcal{T}$ $\|C\rho\bar{x}\| \leq \mu$, for any arbitrary scaling factor ρ.

Then, necessarily,

$$C\bar{x} = 0$$

To complete the "if"-proof, consider a basis matrix T for $\mathcal{T}$, $T = [T_1 \dots T_s]$ and let $U = [U_1 \dots U_s]$ be a set of vectors such that $AT_k + BU_k \in \mathcal{T}$. Then [BM92], for some square P, the following equality holds

$$AT + BU = TP$$

Take σ as any eigenvalue of P and let r be a corresponding eigenvector. Then $ATr + BUr = T\sigma r$ and, since $Tr \in \mathcal{T}$,

$$[A - \sigma I]Tr + BUr = 0, \quad CTr = 0$$

which means that the system has finite transmission zeros.

A simple way to consider the case $D \neq 0$ it is to consider the delay-augmented system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ u(t+1) &= v(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

namely the system with matrices

$$A_{aug} = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}, \; B_{aug} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \; C_{aug} = \begin{bmatrix} C & D \end{bmatrix}, \; D_{aug} = \begin{bmatrix} 0 \end{bmatrix}$$

Then it is immediate that, if we set

$$v(t) = \Phi(x(t+1)) = \Phi(Ax(t) + Bu(t)),$$

any trajectory of the original system can be achieved by a proper initialization of $u(0)$ and therefore $\mathcal{P}$ is λ-contractive for the original system if and only if it is such for the augmented system. The fact that the set is unbounded implies that there are finite transmission zeros for the augmented system. But it is no difficult to see that these are those of the original system.

(**Only if**) To prove the opposite implication, we have that if the system has finite transmission zeros, then there exists a controlled-invariant subspace $\mathcal{T}$ [BM92], namely such that for all $x \in \mathcal{T}$ there exists u assuring that $Ax + Bu(x) \in \mathcal{T}$ and $Cx + Du(x) = 0$. This means that the proper subspace $\mathcal{T}$ is a contractive set compatible with the constraint[8]

For further details on the issue of (possibly unbounded) controlled invariant polyhedral sets, the reader is referred to [DH99].

The main reason for which it is important to have a bounded contractive set is stability. In the case of an unbounded contractive set stability might not be achieved (only partial stability can [Vor98]) as in the following example.

Example 5.16. Consider the system whose matrices are

$$A = \begin{bmatrix} \nu & -\kappa^2 \\ 1 & \nu \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad D$$

Consider the constraints $|y| = |x_1 + x_2 + Du| \leq 1$. If $D \neq 0$, the largest contractive set compatible with the constraints is clearly the whole space $\mathbb{R}^2$. This is true no matter how C is chosen. In the case $D = 0$ the whole constraint-admissible state

$$\Sigma(1) = \{(x_1, x_2) : \ |x_1 + x_2| \leq 1\}$$

is λ-contractive. Indeed no matter how $(x_1(t), x_2(t))$ are taken (inside this strip), there is a choice of u such that $y(t+1) = x_1(t+1) + x_2(t+1) = 0$ (i.e., $\Sigma(1)$ is 0-contractive). However, the fact that this strip is contractive does not imply that the system is stable. The main role is played by the system zero which turns out to be $\sigma = \nu + k^2$ (while the poles are $\nu \pm jk$). If and only if the zero is stable it is possible to find a stabilizing control which renders the strip invariant. This can be seen by considering the transformation $z_1 = x_1$ and $z_2 = x_1 + x_2$, so as to achieve

$$\hat{A} = \begin{bmatrix} \nu + \kappa^2 & -\kappa^2 \\ 1 + \kappa^2 & \nu - \kappa^2 \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \hat{C} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \hat{D} = 0$$

Then we can indeed assure the condition $|y| = |z_2| \leq 1$, however the first equation becomes

$$z_1(t+1) = (\nu + k^2) z_1(t) - \kappa^2 z_2(t)$$

[8] Since in the case of subspaces $\lambda\mathcal{T} = \mathcal{T}$ any controlled-invariant subspace is contractive, but the motion on this subspace is not necessarily stable. If we wish to have convergence we must impose conditions on the eigenvalues of P.

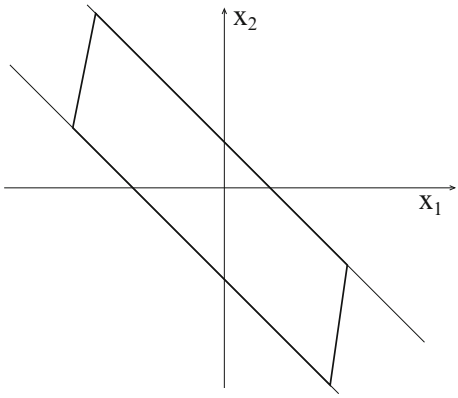


Fig. 5.8 The largest 0.98 contractive set inside the strip

Then if the zero is unstable (for instance, $(\nu + k^2) > 1$), there is no way to drive the state to zero for any $z_1(0)$, since z_1 is "controlled through z_2" which is bounded as $|z_2| < 1$. Conversely, if $|(\nu + k^2)| < 1$, stability is assured, for instance by considering the control

$$u(t) = -(\nu + \kappa^2)z_1(t) - \nu z_2(t),$$

which drives z_2 to zero in finite time. To solve the problem in the case of an unstable zero we must add fictitious constraints. For instance, for $\nu = 0.9$ and $\kappa = 0.4$, the system is stable, $|\nu + jk| = \sqrt{0.97}$, but non minimum-phase, since $\nu + \kappa^2 = 1.06$. There is no hope to achieve stability and the whole strip as an invariant set. Then we can add the "bounding constraints" $|x_1| \leq 1000$ and run the program. The largest λ-contractive set included in the strip with $\lambda = 0.98$ is represented in Figure 5.8. This set has the vertex and plane representations $\bar{\mathcal{V}}[X]$ and $\bar{\mathcal{P}}[F]$, respectively, where

$$X = \begin{bmatrix} 2.002 & 1.7 \\ -1.002 & -2.7 \end{bmatrix}, \quad F = \begin{bmatrix} 0.4586 & -0.0815 \\ 1 & 1 \end{bmatrix}.$$

Note that the fictitious constraints introduced to achieve a bounded set are (by far) not active in this set. The control which can be associated with this set, which is affine to a diamond, is linear and precisely

$$u = -1.451x_1 - 0.8199x_2.$$

The presented results are valid, without changes, for continuous-time systems. They can be extended also to uncertain systems, with the understanding that the maximal contractive set is bounded if the system has no zeros finite transmission zeros for any value of the uncertainties.

5.4 The uncontrolled case: the largest invariant set

A special important case is that in which no control action is present, precisely $B = 0$ or, equivalently, a given linear control law has been chosen. Assume initially that $E = 0$ and consider the following problem.

Problem. Given a C-set $\mathcal{X}$ and the system

$$x(t+1) = A(w(t))x(t), \tag{5.24}$$

find the largest λ-contractive set $\mathcal{S}_\lambda^{max}$ (here we will consider also the largest positively invariant set corresponding to $\lambda = 1$) included in $\mathcal{X}$.

There are two basic motivations for this problem. The first is that, if such a largest set includes the origin as an interior point, then the system is stable and a speed of convergence λ is assured (a special case of the results in the previous section). The second is that, in this way, it is possible to determine a set of initial conditions such that $x(0) \in \mathcal{S} \Longrightarrow x(t) \in \mathcal{X}$ for $t > 0$. If one considers $\lambda = 1$, then $\mathcal{S}$ is the set of *all the initial conditions* for which the property holds. However, if $\lambda = 1$, it is clearly not possible to deduce asymptotic stability.

We face the problem in the case of a polytopic system, i.e.

$$A(w) = \sum_{i=1}^{s} A_i w_i,$$

$\sum_{i=1}^{s} w_i = 1, \;\; w_i \geq 0$, and we assume that

$$\mathcal{X} = \mathcal{P}(F, \bar{1}).$$

Denote by

$$\mathcal{I}_k = \{(i_1, i_2, \ldots, i_k), \;\; i_j = 1, 2, \ldots, s\}$$

the set of all possible choices of k indices among the numbers $1, 2, \ldots, s$ and let, for $C_k = (i_1, i_2, \ldots, i_k) \in \mathcal{I}_k$,

$$\Pi_{C_k} = A_{i_1} A_{i_2} \ldots A_{i_k},$$

the product of the k generating matrices with indices in C_k, where we assume $\Pi_{C_0} = I$. Then the largest invariant set is given by

$$\mathcal{S}^{max} = \mathcal{S}^{(\infty)} = \{x : F\Pi_{C_h} x \leq 1, \;\; C_h \in \mathcal{I}_h \;\; h = 0, 1, 2, \ldots\}$$

Briefly this means that $\mathcal{S}$ is represented by all the inequalities of the form

$$FA_{i_1}A_{i_2}\dots A_{i_k}x \le \bar{1},$$

in principle, an infinite number. Define the set produced at the k iteration as

$$\mathcal{S}^{(k)} = \{x : F\Pi_{C_h}x \le 1,\ \ C_h \in \mathcal{I}_h\ \ h = 0,1,2,\dots,k\} \tag{5.25}$$

The natural question is that whether the set is actually finitely determined, namely whether there exists a $\bar{k}$ such that

$$\mathcal{S} = \mathcal{S}^{(\bar{k})} \tag{5.26}$$

We will provide a complete answer to this question later when we will relate the property of finite determination with the spectral radius of the set of matrices. For the moment being let us state the following preliminary results.

Theorem 5.17. *Assume that system (5.24) is asymptotically stable. Then condition (5.26) holds true for a finite $\bar{k}$, i.e. $\mathcal{S}^{max}$ is finitely determined.*

Proof. We sketch the proof (the reader is referred to [BM96b] for more details and stopping criteria for the procedure). We show that in finite time $\bar{k}$ the condition

$$\mathcal{S}^{(\bar{k}+1)} = \mathcal{S}^{(\bar{k})}$$

is satisfied. Let μ_{int} be the diameter of the largest ball centered in 0 and included in $\mathcal{S}^{(0)}$ and μ_{ext} the diameter of the smallest ball including $\mathcal{S}^{(0)}$. By the asymptotic stability assumption, for k large enough we have $\|A_{i_1}A_{i_2}\dots A_{i_{k+1}}\| \le \mu_{int}/\mu_{ext}$. Then for any $x \in \mathcal{S}^{(0)}$, $\|A_{i_1}A_{i_2}\dots A_{i_{k+1}}x\| \le \mu_{int}$ which means that

$$FA_{i_1}A_{i_2}\dots A_{i_{k+1}}x \le \bar{1}.$$

This means that at a certain value $\bar{k}+1$ these new inequalities are satisfied by all the points $x \in \mathcal{S}^{(0)}$, hence by all the points in its subset $\mathcal{S}^{(\bar{k})}$.

In the case of a single matrix, namely of the system $x(t+1) = Ax(t)$, the procedure stops in a finite number of steps if $\Sigma(A) < 1$ for any C-set $\mathcal{X}$ (equivalently, the procedure for the modified system stops in finite time for $\lambda > \Sigma(A)$). This was shown in [GT91]. Actually in [GT91] it is shown that, under some observability assumptions, the maximal invariant set included in a symmetric polyhedron (i.e., non-necessarily bounded) is compact. We propose this result here in a slightly more general form.

Proposition 5.18. *The largest invariant set $\mathcal{S}^{max}$ for the asymptotically stable system $x(t+1) = A(w)x(t)$ (respectively $\dot{x}(t) = A(w(t))x(t)$) included in the polyhedral set $\mathcal{P}[F,g]$, $g > 0$ is compact provided that there exists $w \in \mathcal{W}$ such*

that $(A(w), F)$ is an observable pair. In the case of a certain system the condition is necessary and sufficient, precisely $\mathcal{S}^{max}$ is compact iff (A, F) is observable.

Proof. The largest invariant set for a system with certain A set is given by the intersection of polyhedra

$$\mathcal{S} = \bar{\mathcal{P}}[F, g] \bigcap \bar{\mathcal{P}}[FA, g] \bigcap \bar{\mathcal{P}}[FA^2, g] \bigcap \cdots \bigcap \bar{\mathcal{P}}[FA^k, g] \bigcap \ldots \tag{5.27}$$

If the system is observable, then the observability matrix

$$O = [F^T \ (FA)^T \ (FA^2)^T \ \ldots \ (FA^{n-1})^T]^T$$

has full column rank. Then, denoting by $\hat{g} = [\underbrace{g^T \ g^T \ g^T \ \ldots g^T}_{n \text{ times}}]^T$, we have

$$\mathcal{S} \subseteq \bar{\mathcal{P}}[O, \hat{g}]$$

Since O has full column rank (see Section 3.3) the rightmost set is compact, hence so is $\mathcal{S}$.

Conversely if (A, F) is not observable, then the system can be reduced in the observability form by a proper transformation $T\tilde{x} = x$, so that the polyhedron becomes $\bar{\mathcal{P}}[\tilde{F}, g] = \bar{\mathcal{P}}[FT, g]$

$$\tilde{A} = \begin{bmatrix} A_{no} & A_{no,o} \\ 0 & A_o \end{bmatrix} \quad \tilde{F} = \begin{bmatrix} 0 & H_o \end{bmatrix}$$

Then it is immediate that if we consider expression (5.27) for the set $\mathcal{S}$, we get only inequalities of the form

$$-g \leq \begin{bmatrix} 0 & H_o A_o^k \end{bmatrix} \begin{bmatrix} \tilde{x}_{no} \\ \tilde{x}_o \end{bmatrix} \leq g.$$

where $\tilde{x}_o$ and $\tilde{x}_{no}$ are the observable and unobservable components of $\tilde{x}$. So $\mathcal{S}$ is unbounded in the unobservable direction $\tilde{x}_{no}$.

To consider the uncertain case, let $w' \in \mathcal{W}$ a value for which $(A(w'), F)$ is observable. The largest invariant set for the system $x(t+1) = A(w)x(t)$ is included in the largest invariant set for the (certain) system $x(t+1) = A(w')x(t)$, henceforth $\mathcal{S}$ is compact.

To prove the results in the continuous-time case $\dot{x}(t) = Ax(t)$, one needs just to consider the exponential approximation

$$x(t+1) = e^{A\tau}x(t) \tag{5.28}$$

and notice that (A, F) is observable iff $(e^{A\tau}, F)$ is observable for $\tau > 0$ small enough. Then the proof is immediate. Indeed, if A is stable, then $e^{A\tau}$ is (discrete-time) stable. Any invariant set for the continuous-time system is invariant for the discrete-time exponential approximation then the largest invariant set is included in the largest (w.r.t. $\bar{\mathcal{P}}[F, g]$) invariant set for (5.28), thus it is compact.

The proof of the previous theorem is inspiring about the following issue: *exponential or EAS*?. We remind that the EAS is the system

$$x(t+1) = [I + \tau A]x(t). \tag{5.29}$$

Indeed we know that, for $\tau > 0$ small, the two tend to coincide since

$$e^{A\tau} \approx I + \tau A.$$

So it is reasonable that in both cases, for $\tau > 0$ small, we can achieve approximation of invariant sets for the continuous-time case. This is certainly the case, but the interesting point arises as far as we are concerned with the maximal invariant set $\mathcal{S}_{ct}$ for the continuous-time system included in $\bar{\mathcal{P}}[F, g]$:

- by means of the EAS (5.29) it is possible to derive internal approximations of $\mathcal{S}_{ct}$;
- by means of the exponential approximating system (5.28) it is possible to derive external approximations of $\mathcal{S}_{ct}$.

This fact, which has also been pointed out in [LO05], is a consequence of the following proposition.

Proposition 5.19.

i) If the C-set $\mathcal{S}$ is positively invariant for the continuous-time asymptotically stable system $\dot{x}(t) = Ax(t)$, then it is positively invariant for the "exponential system" (5.28). The converse is not in general true[9].

ii) If the C-set $\mathcal{S}$ is positively invariant for the EAS (5.29), assuming that it is asymptotically stable, then it is positively invariant for the continuous-time system, which is also asymptotically stable. The opposite is not in general true.

Example 5.20. As a simple example, consider the continuous-time system with

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \tag{5.30}$$

and consider the strip $\bar{\mathcal{P}}[F, 1]$ with $F = [1 \ -1]$. For $\tau = 0.2$, we computed the largest invariant set for the exponential and the EAS approximations, which are

[9]The converse is true in the case of polytopes with the understanding that τ has to be "small enough" [Bla90b].

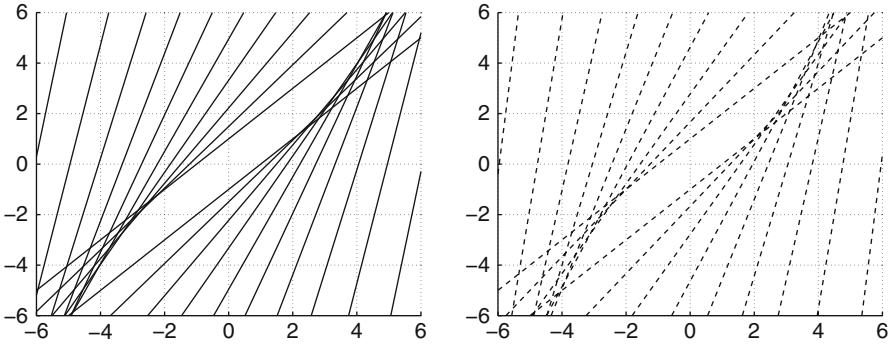


Fig. 5.9 The largest invariant set for the exponential (left) and the EAS (right)

reported in Figure 5.9. As expected, the former includes the latter. In this case the maximal invariant set for the continuous-time system can be computed exactly. This is delimited by the strip $\bar{\mathcal{P}}[F,1]$ and two arcs. The first is that generated by the solution of the system which is identified by the initial condition

$$x_0 = [2(1+\sqrt{2}) \quad 3+2\sqrt{2}]^T,$$

which ends at time $t_f = \log(1+\sqrt{2})$ in the point

$$x_f = [2 \;\; 1]^T.$$

Note that this arc connects the two lines delimiting the strip. The second arc is the opposite (i.e., that generating from $-x_0$ and terminating in $-x_f$ at time t_f). Note that in the final point the two arcs are tangent to the arrival lines. It is easy to see that the so achieved figures satisfies Nagumo's conditions. It is also easy to see that the derived figure is the maximal one.

Let us now consider the case of a system with additive uncertainties

$$x(t+1) = A(w(t))x(t) + Ed(t) \tag{5.31}$$

where $d(t) \in \mathcal{D}$ and $\mathcal{D}$ is a C-set. The next theorem provides a criterion to ensure that the iteration stops in a finite number of steps successfully/unsuccessfully. Although not already formally defined (we will do it soon), let $\mathcal{R}$ be the set of all states reachable from the initial condition $x(0) = 0$ in some (arbitrary) time T with constrained input $d(t) \in \mathcal{D}$. Let $\bar{\mathcal{R}}$ be the closure of $\mathcal{R}$. It is quite easy to see that both $\mathcal{R}$ and $\bar{\mathcal{R}}$ are robustly positively invariant.

Theorem 5.21. *Assume that system (5.31) is asymptotically stable. Then condition (5.26) holds true for a finite $\bar{k}$, i.e. $\mathcal{S}$ is finitely determined, if $\mathcal{X}$ includes a*

closed robustly positively invariant set in its interior (in particular if it encloses $\bar{\mathcal{R}}$). Conversely if $\mathcal{X}$ does not include $\bar{\mathcal{R}}$, then $\mathcal{S}^{(k)} = \emptyset$ for a finite $\bar{k}$

Proof. See [BMS97]

Example 5.22. As an example of application of the largest invariant set, we show how it is possible to determine, given a linear compensator, the set of all initial states for which no input or output violations occur. The size of this set can be considered as an important parameter to evaluate the efficacy of the compensator. Consider the discrete-time system

$$A = \begin{bmatrix} 1.2 & 0.5 \\ -0.5 & 0.8 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

equipped with a control of the form $u = Kx$. The input and output values are constrained as $|u| \leq 1$ and $|y| \leq 1$, respectively. So, given a stabilizing gain K, the problem is that of determining the largest invariant set included in the polyhedron $\bar{\mathcal{P}}[F, \bar{1}]$, where $\bar{1} = [1\ 1]^T$ and

$$F = \begin{bmatrix} K \\ C \end{bmatrix}$$

In the sequel four solutions are considered and compared: the case of dead-beat control and the cases of a closed-loop system with: real distinct, real coincident, and complex eigenvalues. The results are reported in Figures 5.10–5.13, in each of which are depicted:

- the largest invariant set compatible with the input constraint $|Kx| \leq 1$ (plain lines, left);
- the largest invariant set compatible with the output constraint $|Cx| \leq 1$ (dashed lines, center);
- the largest invariant set compatible with both input and output constraints, namely the intersection of the previous two (plain and dashed lines, right).

  The first one is the dead-beat compensator $K = [-2.38\ \ -2.00]$ which places both poles at the origin. The results are shown in Fig. 5.10.

  As a second solution, two distinct real poles, $\lambda_1 = 0.9317$ and $\lambda_2 = 0.2683$, are assigned. The reader could puzzle why these strange numbers. Actually, the gain $K = [0.00\ \ -0.80]$, which results in these poles, was chosen in order to "insert a damping" in the 2–2 coefficient. The results are in Fig. 5.11.

  If two real coincident poles $\lambda_1 = \lambda_2 = 0.9$ are assigned with $K = [+0.3200\ \ -0.2000]$, the situation is quite different w.r.t the previous one. The results are in Fig. 5.12.

  Finally, two complex poles $\lambda_{12} = 0.8 \pm j0.4$ were assigned, resulting in $K = [0.14\ 0.40]$. The results are in Fig. 5.13.

  A comment on the derived pictures is useful. First of all, note that all the determined sets are compact. This is due to the fact that (A_{cl}, K) and (A_{cl}, C) are

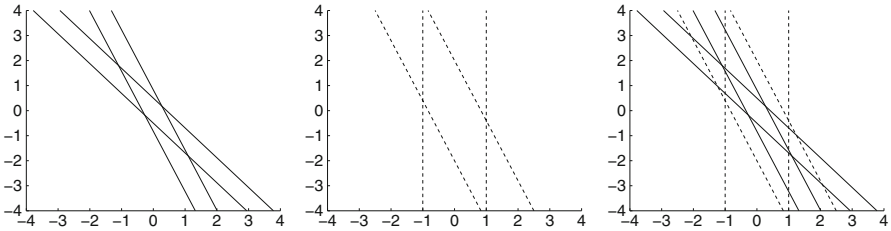


Fig. 5.10 The cases of the dead-beat compensator

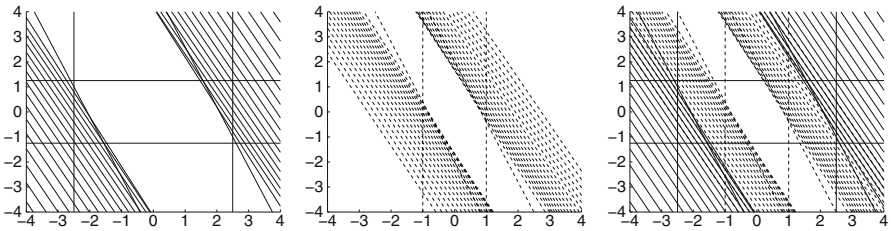


Fig. 5.11 The cases of compensator assigning two distinct poles

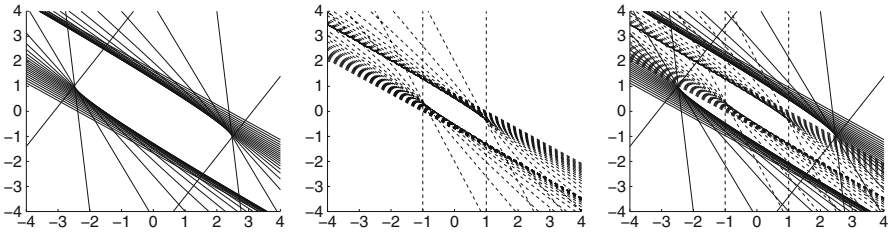


Fig. 5.12 The cases of compensator assigning two coincident poles

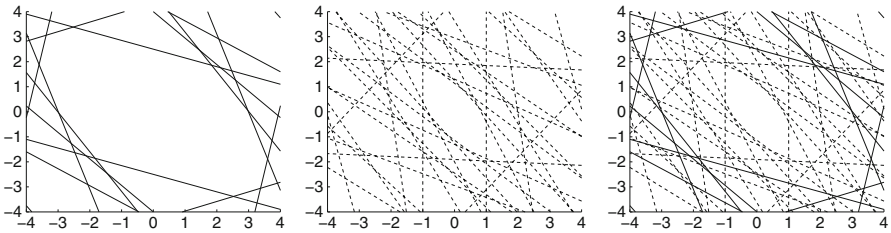


Fig. 5.13 The cases of compensator assigning two complex poles

always observable. We did not remove the redundant constraints. We computed all the constraints $-1 \leq KA_{cl}^i x \leq 1$ and $-1 \leq CA_{cl}^i x \leq 1$ up to $i = 20$. In this way it is apparent that, for some choices of the closed-loop eigenvalues, the convergence is faster than that achieved with other choices (for instance, for the dead-beat controller there are only 8 constraints, since $A_{cl}^i = 0$ for $i \geq 2$).

A circumstance which is apparent is that some controllers, such as the dead-beat one, produce large output-constraint compatible sets while other controllers produce large input-constraint compatible sets and this is certainly the case of the compensator assigning complex poles. This is not surprising. Basically, the dead-beat compensator requires a big control effort which is paid with control exploitation, while the "complex-pole-assigning" one requires a much smaller effort. It is not surprising that the situation is reversed between these two compensators if the output-constraint admissible set, which is quite bigger for the dead-beat one, is considered.

5.4.1 Comments on the results

A drawback of the presented procedures (that will be evidenced throughout the book) is that the derived algorithms do no work with fixed complexity. It turns out that the number of planes which describes the polyhedral Lyapunov function may be huge even for simple problems as in the next example. Computational experience shows that troubles arise when the system is pushed close to its limits. A further source of troubles is the presence of control constraints.

Example 5.23. Consider the 2-dimensional continuous-time system

$$\dot{x}(t) = [w(t)A_1 + (1 - w(t))\, A_2]\, x(t) + Bu(t)$$

with

$$A_1 = \begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix}, \; A_2 = \begin{bmatrix} 0 & 1 \\ -(2+k) & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{5.32}$$

and $0 \leq w(t) \leq k$ for all $t \geq 0$. This system is obtained by the example in [Zel94] (subsequently reconsidered in [CGTV03]), by adding an input matrix. For the above system it can be shown that the maximum value of k that assures open-loop stability is 7.

So, for $k = 7.5$ the system is open-loop unstable. Let us then consider an unconstrained stabilization problem. Let us set $\tau = 0.1$ and run the algorithm. Starting from $\mathcal{X}^{(0)} = \{x : \|x\|_\infty \leq 1\}$ with $\lambda = 0.9$, it turns out that the largest contractive set is simply the symmetric diamond $\mathcal{V}(X)$ with

$$X = \begin{bmatrix} 0.4042 & -0.4042 & 0.5911 & -0.5911 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

It turns out that this system can be associated with the linear stabilizing gain

$$u = -0.8599\, x_1(t) - 1.069\, x_2(t).$$

To see the effect of control constraints, set $k = 7.5$ and let the input bound be $|u| \leq 1$. The same value of $\tau = 0.1$ is used and $\lambda = 0.96$, which correspond to a contractive coefficient

$$\beta = (1 - \lambda)/\tau = 0.4.$$

Starting from the same initial region, a set represented by 40 delimiting planes and vertices was found. This set is a guaranteed domain of attraction (although not the largest one) with the considered control constraints.

Let us now consider the problem of approximating, as closely as possible, the largest contractive set under the considered control constraints. To this aim, a very small discretization parameter $\tau = 2.5 \times 10^{-3}$ is used to derive the discrete-time EAS. The contractivity coefficient for the EAS was chosen as $\lambda = 0.999$, in order to keep the continuous-time contractivity coefficient as before ($\beta = (1 - \lambda)/\tau = 0.4$) and a tolerance $\epsilon = 9 \times 10^{-4}$ was adopted. The procedure was started with the initial set $\mathcal{X}^{(0)} = \{x : \|x\|_\infty \leq 1 \times 10^4\}$ (that simulates unbounded state-space variables) in order to take into account only control constraints. The resulting polyhedral set is represented by 960 planes. Such a set is the region internal to the thick line depicted in Figure 5.14.

It is rather clear that the situation is completely different in terms of complexity. The same trouble occurs if one wishes to push the system close to "its limits." Let us now consider a stability analysis problem. Precisely let us check the system stability when $k = 6.97$. Starting from the initial set $\mathcal{X}^{(0)} = \{x : \|x\|_\infty \leq 1\}$

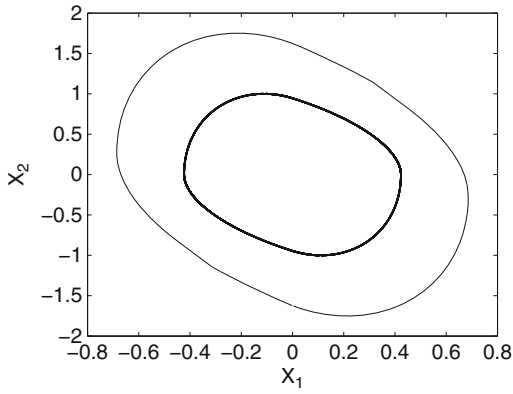


Fig. 5.14 Thin line: a $7.6e^{-6}$-contractive set for the autonomous system with $k = 6.97$; thick line: the constrained-control convergence domain

and by using the discrete-time EAS with $\tau = 2.5 \times 10^{-4}$ and the constants $\lambda = 1 - 1 \times 10^{-9}$ and $\epsilon = 9 \times 10^{-10}$ (corresponding to a guaranteed contractivity $\beta = 7.6e^{-6}$), a "polyhedral" Lyapunov function was constructed. In this case, the contractivity is pushed almost to its limit (since for $k = 7$ the system does not admit a contractive region). The quotes in the polyhedral are easily justified by the shape of the level set of the Lyapunov function, depicted in Fig. 5.14 (thin line), which counts 12842 planes (though the region is symmetric, thus half of this number is a measure of the complexity of the region). The computational time is 39 s on a x86 750 MHz calculator. The reader is referred to [RKMK07] for a promising technique of construction of controlled-invariant sets of a reasonable complexity. A possible way to construct polyhedral Lyapunov functions is facing the bilinear equations. Unfortunately, in this case the solution needs non-convex methods, which are in general hard to solve. A promising approach is using randomized algorithms (see, for instance, [TCD04]).

As we will see later, if one chooses ellipsoidal sets, then the algorithms have a fixed complexity, a big advantage. On the other hand, ellipsoidal sets are conservative, because it is known that a system may admit a contractive set but no ellipsoidal contractive sets. Furthermore ellipsoidal sets have no maximality properties. For instance, the maximal λ-contractive set included in an ellipsoid is not an ellipsoid in general. To overcome the conservativity deriving from the use of ellipsoidal sets while maintaining the advantages of their computation, in recent years some authors (see [CGTV03] and the references therein) have started to consider homogeneous polynomial Lyapunov functions and, more precisely, those which can be written as sum of squares (SOS) (and treated by standard LMI techniques). We will come back to this point in the next chapter, when we will deal with polynomial Lyapunov functions.

5.5 Exercises

1. Show that the chain of inclusions (5.5) holds (hint: show that if $\mathcal{X}_1 \subset \mathcal{X}_2$ then $Pre_{SF}(\mathcal{X}_1) \subseteq Pre_{SF}(\mathcal{X}_2)$)
2. Consider the backward construction of $\bar{\mathcal{X}}_{-\infty}$ for the scalar system

$$x(t+1) = 2x(t) + u(t)w(t),$$

where $\mathcal{W} = \{w : |w| \leq 1\}$, $\mathcal{X} = \{x : 0 \leq x \leq 1\}$, and $\mathcal{U} = \{x : 0 < u \leq 1\}$ which is not closed. Compute the sequence of (non-closed sets) $\mathcal{X}_{-k}$, show that they have non-empty intersection, but the problem of keeping the state in $\mathcal{X}$ has no solution. (Hint: given the set $\{x : \ 0 \leq x < k\}$ or the set $\{x : \ 0 \leq x \leq k\}$ the

one-step-controllability set to this set is $\{x:\ 0 \le x < k/2\}$ thus the sequence of sets is $\mathcal{X}_{-k} = \{x:\ 0 \le x < (1/2)^k\}$[10])

3. Show that the set $\bar{\mathcal{X}}^{(\infty)}$ for the system $x(t+1) = 2x(t) + u(t) + d(t)$ with $-1 \le x \le 10$, $-20 \le u \le 1$, $-2 \le d \le 2$, is non-empty but it does not include the origin.
4. Consider the reachability problem formulated on a graph as in subsection 5.1.3 with the following modifications. Assume that to each arch a transition weight is provided. Determine a "Lyapunov" node-function such that $V(\{k\}) = 0$ and defined in such a way that a feedback strategy to reach node $\{k\}$ can be achieved by moving at each time to the node (among the connected ones) having the minimum value of V and such that the achieved path from each starting node $\{j\}$ is of minimum cost, namely the sum of the transition costs is minimal among all the paths connecting $\{j\}$ to $\{k\}$.
5. Prove the statement of Remark 5.5.
6. Show, by means of a counterexample, that the maximal λ-contractive set included in an ellipsoid is not necessarily an ellipsoid. How does this set look like for $x(t+1) = A(w(t))x(t)$, with polytopic A?
7. Find the set of states that can be driven to 0 (in an arbitrary number of steps) for system (5.21) after the transformation $u = -2x + v$ is applied. (... of course you know it: it is unchanged, just $(-4, 4)$ ... but prove it!)
8. Consider again the continuous-time system whose matrix is given in (5.30). Prove that the largest invariant set is actually that described in the example.
9. Prove/disprove the following: system (A, B, C) has finite transmission zeros iff there exists a controlled-invariant subspace included in $Ker(C)$ (see [BM92] for details).
10. Show that the largest λ-contractive set for the system $x(t+1) = A(w(t))x(t) + B(w(t))u(t)$ $y(t) = Cx(t)$, included in the strip $\|Cx\| \le \mu$ is bounded, provided that the system $(A(w), B(w), C)$ has no finite transmission zeros for all w.
11. Write the regulation map $\Omega(x)$ for the Example 5.11.
12. Rephrase example 5.16 for continuous-time systems.
13. Show that the delayed-extended system considered in Proposition 5.15 has the same zeros of the original one.

[10] This hint is rather a solution ....

Chapter 6
Set-theoretic analysis of dynamic systems

In this section, several applications of set-theoretic methods to the performance analysis of dynamic systems will be presented. Although, in principle, the proposed techniques are valid for general systems, their application is computationally viable in the case of (uncertain) linear systems and thus we restrict the attention to this case.

6.1 Set propagation

6.1.1 Reachable and controllable sets

Consider a dynamic system of the form

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

or of the form

$$x(t+1) = f(x(t), u(t))$$

where $u(t) \in \mathcal{U}$. The following classical definitions of reachability and controllability sets are reported.

Definition 6.1 (Reachability set). Given the set $\mathcal{P}$, the reachability set $\mathcal{R}_T(\mathcal{P})$ from $\mathcal{P}$ in time $T < +\infty$ is the set of all vectors x for which there exists $x(0) \in \mathcal{P}$ and $u(\cdot) \in \mathcal{U}$ such that $x(T) = x$.

Definition 6.2 (Controllability set). Given the set $\mathcal{S}$, the controllability set $\mathcal{C}_T(\mathcal{S})$ to $\mathcal{S}$ in time $T < +\infty$ is the set of all vectors x for which there exists $u(\cdot) \in \mathcal{U}$ such that if $x(0) = x$ then $x(T) \in \mathcal{S}$.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_6

More in general, $\mathcal{S}$ is said to be reachable from $\mathcal{P}$ in time T if for all $x \in \mathcal{S}$ there exists $x(0) \in \mathcal{P}$ and $u(\cdot)$ such that $x(T) = x$. Similarly, $\mathcal{P}$ is said to be controllable to $\mathcal{S}$ in time T if, for all $x(0) \in \mathcal{P}$, there exists $u(\cdot)$ such that $x(T) \in \mathcal{S}$. Unless for very specific cases, the fact that $\mathcal{P}$ is reachable from $\mathcal{S}$ does not imply that $\mathcal{S}$ is controllable to $\mathcal{P}$ and vice versa.

However, if backward systems are considered, namely systems that evolve backward in time of the form

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

or of the form

$$x(t+1) = f^{-1}(x(t), u(t))$$

where f^{-1} is the inverse of f with respect to x (if it exists at all), precisely the map such that $f^{-1}(f(x,u),u) = x$ for all x and $u \in \mathcal{U}$, then the set $\mathcal{P}$ is reachable from (controllable to) $\mathcal{S}$ if and only if $\mathcal{S}$ is controllable to (reachable from) $\mathcal{P}$ for the backward system.

Controllable sets have the following composition property[1]. If $\mathcal{S}_0$ is controllable in time T_1 to $\mathcal{S}_1$ and $\mathcal{S}_1$ is controllable in time T_2 to $\mathcal{S}_2$, then $\mathcal{S}_0$ is controllable in time $T_1 + T_2$ to $\mathcal{S}_2$. The analogous composition property holds for reachability.

Reachable sets are useful to describe the effects of a bounded disturbance on a dynamical system or to describe the range of effectiveness of a bounded control. Unfortunately, the computation of reachable sets is, in general, very hard even in the discrete-time case, although effort is currently put in this direction [RKML06]. For simple systems, typically planar ones, they can be computed (approximately) by simulation and the approximated reachable and controllable sets can be visualized by appealing computer graphics. Unfortunately, as the dimension grows, our mind gets somehow lost, besides the inherent intractability of reachable set computation.

From the theoretical point of view, some results that characterize the closedness or compactness of controllability/reachability sets which are available in the mathematical literature. For instance, in the discrete-time case, if the map f is assumed to be continuous and $\mathcal{U}$ compact, then the expression of the one-step reachability set of a compact set $\mathcal{P}$, precisely

$$f(\mathcal{P}, \mathcal{U})$$

is compact. Therefore the reachable set in k steps, that can be recursively computed by setting $\mathcal{R}_0 := \mathcal{P}$ and

$$\mathcal{R}_{k+1} = f(\mathcal{R}_k, \mathcal{U})$$

[1] A semi-group property.

is compact. Some compactness results for the controllability sets can be easily inferred under some assumptions, such as the system reversibility (i.e., that $f^{-1}(x,u)$ is defined for all u and continuous). This kind of closedness–compactness results are valid also for continuous-time systems under suitable regularity assumptions.

The next theorem shows that, at least for the case of linear systems (which are those we will be mostly interested in), some reasonable procedures can be devised.

Theorem 6.3. *Consider the system*

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (or\ \ x(t+1) = Ax(t) + Bu(t))$$

where $u \in \mathcal{U}$, with $\mathcal{U}$ a convex and compact set and, in the discrete-time case, A is assumed to be invertible. Let $\mathcal{P}$ be a convex and compact set. Then, for all $T < +\infty$,

- *the controllability set in time T to $\mathcal{P}$ is convex and compact.*
- *the reachability set in time T from $\mathcal{P}$ is convex and compact.*

In the discrete-time case, if matrix A is singular, the reachability set is still convex and compact whereas the controllability set is convex, closed but not necessarily bounded.

Proof. The proof of compactness will be reported in the continuous-time case only, whereas the proof of convexity and the discrete-time case are left as an exercise.

The reachability set is given by the set of all vectors

$$x = e^{AT}\bar{x} + \int_0^T e^{A(T-\sigma)}Bu(\sigma)d\sigma, \tag{6.1}$$

(with $u(\cdot)$ a measurable function) namely the sum of the image of $\mathcal{P}$ with respect to e^{AT} (which is compact) and the set of all vectors reachable from 0 in time T:

$$\mathcal{R}_T(\mathcal{P}) = e^{AT}\mathcal{P} + \mathcal{R}_T(\{0\})$$

The set of states reachable from 0 in a finite time is compact as shown in [PN71], and, since the sum of two compact sets is compact, $\mathcal{R}_T(\mathcal{P})$ is compact. The analogous proof for controllable sets is derived by noticing that the controllable set is the set of all $\bar{x}$ for which (6.1) holds with x in $\mathcal{P}$. By multiplying both sizes by e^{-AT} one immediately derives

$$\mathcal{C}_T(\mathcal{P}) = e^{-AT}\mathcal{P} + \mathcal{R}_T^-(\{0\})$$

where we denoted by $\mathcal{R}_T^-(\{0\})$ the set of 0-reachable states of the time-reversed sub-system $(-A, -B)$, hence the claim.

Convexity of reachability and controllability sets in the case of linear systems is a strong property which allows to obtain practical results, as it will be seen later.

An important problem that can be solved, in principle, in a set-theoretic framework is the analysis of uncertainty effects via set propagation. The literature on this kind of investigation is spread in different areas. A classical approach to the problem is that based on the concept of differential inclusion. As we have seen, a system of the form $\dot{x}(t) = f(x(t), w(t)),\ w(t) \in \mathcal{W}$, is a special case of differential inclusion. Given an initial condition $x(0)$, if one is able to determine the reachable set $\mathcal{R}_t(\{x(0)\})$, then one can actually have an idea of the uncertainty effect. The literature presents some effort in this sense, however, most of the work is effective only for special classes of systems, typically of low dimensions. A survey of numerical methods for differential inclusions can be found in [DL92].

6.1.2 Computation of set propagation under polytopic uncertainty

Let us now consider the discrete-time system

$$x(t+1) = A(w(t))x(t) + E(w(t))d(t) \tag{6.2}$$

with

$$A(w) = \sum_{i=1}^{s} A_i w_i(t), \quad E(w) = \sum_{i=1}^{s} E_i w_i(t)$$

$$w \in \mathcal{W} = \{w : w_i \geq 0,\ \sum_{i=1}^{s} w_i = 1\}$$

and $d \in \mathcal{D}$, also a polytope. Here, no control action is considered and w and d are both external uncontrollable signals.

Consider the problem of computing the propagation of the uncertainty for this system, starting from a set $\mathcal{X}_0$ of initial conditions which is a polytope. This set can be propagated forward in time, keeping into account the effect of the uncertainty and disturbance, by considering the set:

$$\mathcal{X}_1 = \mathcal{R}_1(\mathcal{X}_0) = \{A(w)x + E(w)d :\ w \in \mathcal{W},\ d \in \mathcal{D}\} \tag{6.3}$$

Even from the first step it is not difficult to see that the one step reachable set $\mathcal{X}_1$ is not convex (then it cannot be a polytope). The lack of convexity is shown in the next example.

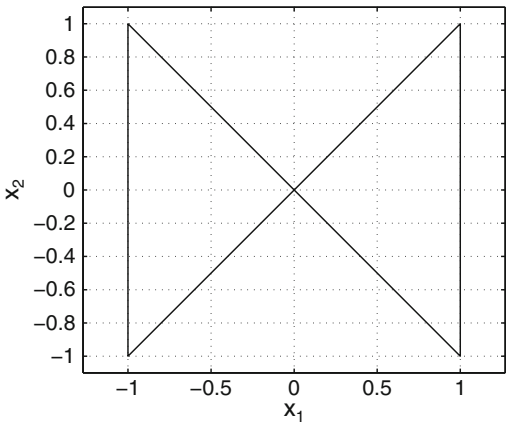


Fig. 6.1 Butterfly shaped non-convex one step reachable set for example 6.4

Example 6.4. Consider the autonomous system whose matrices are

$$A = \begin{bmatrix} 1 & w \\ 0 & 1 \end{bmatrix}, \quad E = 0,$$

and $|w| \leq 1$. Let $\mathcal{X}_0 = \{(x_1, x_2) : \quad x_1 = 0, \ x_2 \in [-1, \ 1]\}$. The reachable set is the set of all the possible images of the considered segment with respect to matrix $A(w)$, which turns out to be the union of two triangles with center in the origin (the one having vertices $[0\ 0]^T$, $[1\ 1]^T$ $[1\ -1]^T$ and its opposite), depicted in Figure 6.1 and clearly non-convex.

Though the reachable set is non-convex, Barmish and Sankaran [BS79] showed that the convex hull of the reachable sets can be propagated recursively, as per the next result.

Proposition 6.5. *Let $\mathcal{X}_0$ be a polytope and let $\mathcal{X}_k$ be the k-step reachability set of (6.2) from $\mathcal{X}_0$. Let $\hat{\mathcal{X}}_k = conv\{\mathcal{X}_k\}$ be its convex hull. Then the sequence of convex hulls can be generated recursively as*

$$\hat{\mathcal{X}}_{k+1} = conv\left\{\mathcal{R}_1\left(\hat{\mathcal{X}}_k\right)\right\},$$

roughly, as the convex hulls of image sets of convex hulls.

The remarkable property evidenced by the previous theorem is that one can compute the convex hulls of the reachability sets by just propagating the vertices. Precisely, assume a vertex representation of the polytope $\hat{\mathcal{X}} = \mathcal{V}(x_1, x_2, \ldots, x_s)$ is known. Let A_i and E_i, $i = 1, 2, \ldots, r$, be the matrices generating $A(w)$ and $E(w)$ and let $\mathcal{D} = \mathcal{V}(D)$, where $D = [d_1, d_2, \ldots, d_h]$. Then the convex hull of the one-step reachability set (which might be non-convex) is given by the convex hull of all the points of the form $A_i x_k + E_i d_j$, say its expression is

$$conv\left\{\mathcal{R}_1\left(\hat{\mathcal{X}}\right)\right\} =$$
$$= conv\left\{A_i x_k + E_i d_j,\ i = 1, 2, \ldots, r,\ k = 1, 2, \ldots, s,\ j = 1, 2, \ldots, h\right\} \tag{6.4}$$

Therefore, a routine which propagates the vertices of the sets $\hat{\mathcal{X}}_k$ can be easily constructed. Its application is a different story. Indeed, the complexity of the problem is enormous, since the number of candidate vertices grows exponentially. One can apply the mentioned methods to remove internal points, but still the number of true vertices might explode in few steps even for small dimensional problems. The reader is referred to [RKKM05b, RKK+05] for more recent results on this construction.

Example 6.6. As previously mentioned, the one-step forward reachability set of a convex set is in general non-convex [BS79]. Here, by means of a simple two-dimensional system, another graphical representation of such lack of convexity is reported. Consider the two-dimensional autonomous uncertain system

$$x(k+1) = A(w(k))x(k)$$

with $|w(k)| \leq 1$ and

$$A(w) = \begin{bmatrix} 1/3 & -2/3\, w \\ -2/3 + 2/3\, w & 1/2 \end{bmatrix}$$

and the set

$$\mathcal{X} = \{x : \|x\|_\infty \leq 1\}$$

The one step forward reachability set (computed on a grid of points) is depicted in figure 6.2

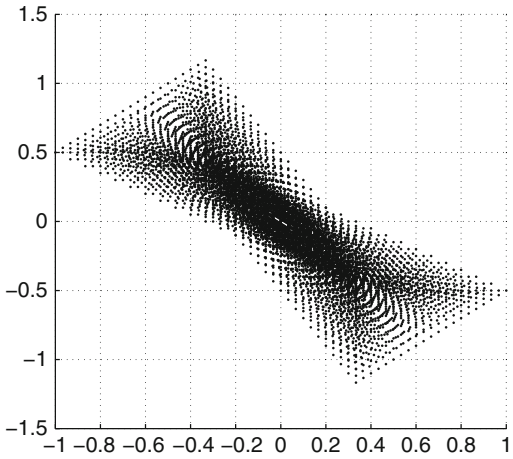


Fig. 6.2 Image (computed on a grid) of the one step reachable set for example 6.6

It is immediately seen that this set is non-convex (to double check such sentence one can try to determine whether there exist $x \in \mathcal{X}$ and a value $-1 \leq w \leq 1$ such that $A(w)x = [1/2,\ 0]^T$).

Conversely the preimage set is convex and precisely

$$\mathcal{S} = \{x : \|A(1)x\|_\infty \leq 1,\ \|A(-1)x\|_\infty \leq 1\}$$

If A and E are certain matrices (i.e., the family of matrices are singletons), then the following result, reported without proof, holds:

Proposition 6.7. *Let $\mathcal{X}_0$ and $\mathcal{D}$ be polytopes. Consider the system*

$$x(t+1) = Ax(t) + Ed(t), \quad \textit{with } d \in \mathcal{D}$$

Then $\mathcal{X}_k$, the k-step reachability set from $\mathcal{X}_0$, is a polytope.

Again, if $\mathcal{X}$ and $\mathcal{D}$ are known, then (6.4), as a special case, provides the expression of the one-step reachability set. It is worth mentioning that the propagation of the uncertainty effect cannot be achieved by considering ellipsoidal sets. Indeed, even in the case in which no parametric uncertainty is present, the one step reachable set from an ellipsoid is convex, but it is not an ellipsoid.

We have seen that the attempt of propagating the disturbance effect forward in time can be frustrating even in the case of linear systems, if parameter uncertainties are to be considered. Thus, working with reachable sets forward in time, unless for the special case of linear systems with no parameter uncertainties, is very hard. The reader is referred to [RKKM05a, RK07, LO05] for recent results on the topic. Luckily enough, there is another bullet to shoot, the controllability one. It will soon be shown that, by working backward in time, it is possible to keep convexity, a property which allows to derive efficient numerical algorithms. We will consider this aspect later when the concept of worst case-controllability will be considered.

6.1.3 Propagation of uncertainties via ellipsoids

A known method to investigate the effect of uncertainty is the adoption of ellipsoids. However, as already mentioned, the methods based on ellipsoids are conservative, since they are usually unfit to describe the true reachability set. However, they typically require less computational effort. We remind the an ellipsoid with center c, radius 1 and characterizing matrix $G^{-1} \succ 0$ is denoted by

$$\mathcal{E}(c, G^{-1}, 1) = \{x : (x-c)^T G^{-1} (x-c) \leq 1\}$$

Let us consider the case of the following linear system

$$\dot{x}(t) = Ax(t) + Bu(t) + Ed(t)$$

where $u(t)$ is a known input and $d(t)$ is an uncertain input, bounded as

$$d \in \mathcal{E}(0, G^{-1}, 1)$$

(i.e., $d^T G^{-1} d \leq 1$). Let us assume that the state is initially confined in the following ellipsoid with center c_0

$$x(0) \in \mathcal{E}(c_0, Q^{-1}, 1)$$

Then the state of the system is confined at each time in the ellipsoid[2] (see [Sch73], section 4.3.3)

$$x(t) \in \mathcal{E}(c(t), Q^{-1}(t), 1) \tag{6.5}$$

where the center $c(t)$ and the matrix $Q^{-1}(t)$ describing the ellipsoid satisfy the following equations

$$\dot{c}(t) = Ac(t) + Bu(t) \tag{6.6}$$

$$\dot{Q}(t) = AQ(t) + Q(t)A^T + \beta(t)Q(t) + \beta(t)^{-1}EGE^T \tag{6.7}$$

where $\beta(t)$ is an arbitrary positive function. A discussion on how to choose the free function β to achieve some optimality conditions for the ellipsoid $\mathcal{E}(c(t), Q^{-1}(t), 1)$ is proposed in [Che81, CO04, Sch73]. The reader is referred to the recent survey books [Che94, KV97] for a more complete overview. It is worth noticing that, in the case of a stable A, assuming $u = 0$ and a constant function β, the asymptotic value of Q is achieved by setting $\dot{Q} = 0$, thus achieving, as a particular case, equation (4.23). Note also that, by setting $Q(0) = 0$ (in this case the expression $\mathcal{E}(0, Q^{-1}(t), 1)$ has no significance for $t = 0$), the initial state is set to 0. Then the corresponding set $\mathcal{E}(0, Q^{-1}(t), 1)$ (defined for $t > 0$) includes the set of states reachable in time t from the origin. It will be seen how to compute, at least approximately, the reachability set from 0.

There is a corresponding equation for discrete-time ellipsoidal confinement. The reader is referred to specialized literature (see, for instance, [Sch73], Section 4.3.2)

[2] In general the inclusion is quite conservative.

6.2 0-Reachable sets with bounded inputs

In this section, a specific problem, and precisely that of estimating the 0-reachable set of a linear time invariant system, will be considered. We consider the reachable sets with pointwise bounded inputs for both discrete and continuous-time systems. We will consider also the problem of determining reachable sets with energy-bounded inputs (a problem elegantly solvable via ellipsoidal sets) although such a class of signals have not been considered in this book so far.

6.2.1 Reachable sets with pointwise-bounded noise

Consider initially the discrete-time system

$$x(t+1) = Ax(t) + Ed(t).$$

Assume that $d \in \mathcal{D}$ is a convex and closed set including the origin. Denote by $\mathcal{R}_T$ the set of all reachable states in T steps. It is not difficult to see that, since $0 \in \mathcal{D}$,

$$\mathcal{R}_T \subseteq \mathcal{R}_{T+1}$$

namely the sets $\mathcal{R}_T$ are nested. The T step reachability set is given by

$$\mathcal{R}_T = \sum_{k=0}^{T-1} A^k E\mathcal{D}.$$

and it can be recursively computed as follows:

$$\mathcal{R}_{T+1} = A\mathcal{R}_T + E\mathcal{D}$$

This involves known operations amongst sets, such as computing the sum and the image of a set (see Section 3.1.1, page 96). These operations can be done, in principle, in the environment of convex sets. However, for computational purposes, sticking to polyhedral sets is of great help. Let us assume that $\mathcal{D}$ is a polytope. Then, assuming the following vertex representation,

$$\mathcal{R}_T = \mathcal{V}[x_1^{(T)}, x_2^{(T)}, \ldots, x_{r_T}^{(T)}], \quad \mathcal{D} = \mathcal{V}[d_1, d_2, \ldots, d_s]$$

the set $\mathcal{R}_{T+1}$ has the points $Ax_j^{(T)} + Ed_k$ as candidate vertices, precisely

$$\mathcal{R}_T = conv\left\{Ax_j^{(T)} + Ed_k,\ j = 1, \ldots, r_T,\ k = 1, \ldots, s\right\}$$

Again the number of candidate vertices grows exponentially. Therefore the algorithm may result difficult to apply to systems of high dimension. It is worth noticing that the generation of 0-reachability sets for polytopic systems $x(t+1) = A(w(t))x(t) + E(w(t))d(t)$, in view of the previous consideration, leads to sets that are non-convex. However, if one is satisfied with the convex hulls $conv\{\mathcal{R}_T\}$, these computations can be done in an exact way according to Proposition 6.5 and the operations in (6.4).

A different approach that may be used for evaluating reachable sets is based on the hyperplane representation of the set. Since $\mathcal{R}_T$ is convex and, in general, closed, it can be described by its support functional as

$$\mathcal{R}_T = \{x : z^T x \leq \phi_T(z), \ \ \forall z\}$$

The support functional $\phi_T(z)$ can be computed as follows. Denote by $\mathcal{D}_T = \mathcal{D} \times \mathcal{D} \times \cdots \times \mathcal{D}$, ($T$ times), the convex and compact set of finite sequences of T vectors in $\mathcal{D}$. The T-step reachability set is given by

$$\mathcal{R}_T = \left\{x = [\, E \; AE \; A^2E \; \ldots \; A^{T-1}E \,]d_T, \quad d_T \in \mathcal{D}_T\right\}.$$

Therefore

$$\begin{aligned}\phi_T(z) &= \sup_{d_T \in \mathcal{D}^T} \left\{z^T [\, E \; AE \; A^2E \; \ldots \; A^{T-1}E \,]d_T\right\} = \\ &= \sum_{i=0}^{T-1} \sup_{d \in \mathcal{D}} z^T A^i E d \\ &= \sum_{i=0}^{T-1} \phi_{\mathcal{D}}(z^T A^i E),\end{aligned}$$

where $\phi_{\mathcal{D}}(\cdot)$ is the support functional of $\mathcal{D}$. Therefore the evaluation of $\phi_T(z)$ at a point z requires the solution of the programming problem $\sup_{d \in \mathcal{D}} z^T A^i E d$. If $\mathcal{D}$ is a C-set, then "sup" is actually a "max." Remarkable cases are those in which $\mathcal{D}$ is the unit box of the p norm, with $1 \leq p \leq \infty$

$$\mathcal{D} = \{d : \|d\|_p \leq 1\}$$

For the above,

$$\max_{d \in \mathcal{D}} z^T A^i E d = \|z^T A^i E\|_q,$$

where q is such that $1/p + 1/q = 1$. In particular, if $\mathcal{D}$ is assumed to be a hyperbox, the components of d are all bounded as $|d_i| \leq \bar{d}_i$. Without restrictions, we can assume $|d_i| \leq 1$ a condition always achievable by scaling the columns of matrix E.

Then

$$\phi_T(z) = \sum_{i=0}^{T-1} \|z^T A^i E\|_1$$

Example 6.8. Let us consider a very simple example in which the computation can be carried out by hand (perhaps an isolated case in this book). Consider the matrices

$$A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and let the disturbance set be $\mathcal{D} = [-1, 1]$. The reachable sets for the first three steps, also depicted in Fig. 6.3, are

$$\mathcal{R}_1 = \bar{\mathcal{V}}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right),$$

$$\mathcal{R}_2 = \bar{\mathcal{V}}\left(\begin{bmatrix} 3/2 & -1/2 \\ -1/2 & -1/2 \end{bmatrix}\right),$$

$$\mathcal{R}_3 = \bar{\mathcal{V}}\left(\begin{bmatrix} 3/2 & -1/2 & -3/2 \\ -1 & -1 & 0 \end{bmatrix}\right).$$

For the above system, consider the problem of determining the largest absolute value of the output $y(t) = x_2(t)$. This problem may be recast as follows: consider constraints of the form $z^T x \leq \mu$ and $-z^T x \leq \mu$, where $z = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ and determine the smallest value of μ such that the reachable set is inside a proper μ-thick strip (see Fig. 6.3).

$$\bar{\mathcal{P}}[z, \mu] = \{x : |z^T x| \leq \mu\}$$

It is immediately seen that such a value is the support functional of $\mathcal{R}_t$ computed in z. In this example the smallest value in three steps is $\mu_{min} = 1$.

This value can be computed by considering the expression

$$\mu_{min} = \phi_3(z) = \|z^T E\|_1 + \|z^T AE\|_1 + \|z^T A^2 E\|_1 = 0 + \frac{1}{2} + \frac{1}{2}$$

It is clear that, in principle, one could compute in an approximate way the infinite-time reachability set

$$\mathcal{R}_\infty = \bigcup_{k=0}^{\infty} \mathcal{R}_k$$

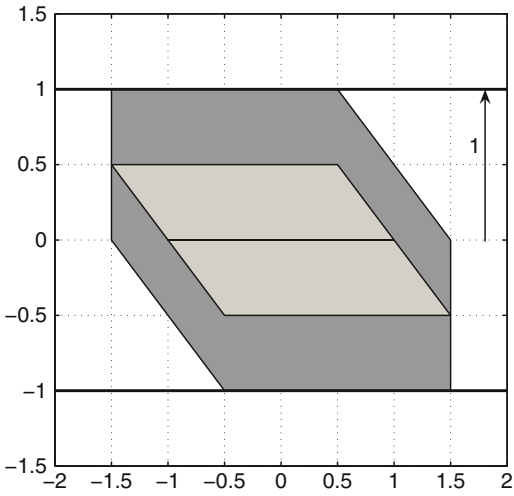


Fig. 6.3 The reachability sets for example 6.8

by computing $\mathcal{R}_k$ with k large. Clearly, this set would be an internal approximation. The following convergence property holds.

Proposition 6.9. *Assume that $\mathcal{D}$ is a C-set, that (A, E) is a reachable pair and that A is asymptotically stable. Then the set $\mathcal{R}_\infty$ is bounded and convex and its support functional is given by*

$$\phi_\infty(z) = \sum_{i=0}^{\infty} \phi_{\mathcal{D}}(z^T A^i)$$

Furthermore, $\mathcal{R}_k \to \mathcal{R}_\infty$ in the sense that for all $\epsilon > 0$ there exists $\bar{k}$ such that for $k \geq \bar{k}$

$$\mathcal{R}_k \subseteq \mathcal{R}_\infty \subseteq (1+\epsilon)\mathcal{R}_k$$

Proof. There are several proofs of the previous result in the literature, for instance [GC86b]. We just sketch the proof. The support functional is

$$\phi_\infty(z) = \sup_{d(\cdot)\in\mathcal{D}} z^T \sum_{h=0}^{\infty} A^h E d(h) =$$

$$= \sum_{h=0}^{\infty} \sup_{d(\cdot)\in\mathcal{D}} z^T A^h E d(h) = \sum_{h=0}^{\infty} \phi_{\mathcal{D}}(z^T A^h E)$$

This value is finite because $\|z^T A^i E\| \leq \|z^T\| \|A\|^i \|E\|$ converges to 0 exponentially, and then $\mathcal{R}_\infty$ is bounded. As far as the inclusions are concerned, the first one is obvious. The second can be proved as follows. Fix $\bar{k} > 0$. Any reachable state in

$k \geq \bar{k}$ steps can be written as

$$x = \sum_{h=0}^{\bar{k}-1} A^h E d(h) + A^{\bar{k}} \sum_{h=0}^{k-\bar{k}} A^h E d(h+\bar{k}) =$$
$$= \sum_{h=0}^{\bar{k}-1} A^h E d(h) + A^{\bar{k}} s_k \in \mathcal{R}_{\bar{k}} + A^{\bar{k}} \mathcal{R}_{k-\bar{k}}$$

where $s_k \in \mathcal{R}_{k-\bar{k}}$ (it is understood that if $k \leq \bar{k}$ then $s_k = 0$), so $\|s_k\| \leq \mu$ for some positive μ. Then

$$\nu_{\bar{k}} \doteq \|A^{\bar{k}}\|\mu \to 0 \quad \text{as } \bar{k} \to \infty$$

and, denoting by $\mathcal{B} = \mathcal{N}[\|\cdot\|, 1]$ the unit ball of $\|\cdot\|$ we have that the k-step reachable state is

$$\mathcal{R}_k \subseteq \mathcal{R}_{\bar{k}} + \nu_{\bar{k}} \mathcal{B} \subseteq (1+\epsilon)\mathcal{R}_{\bar{k}}$$

This, in turn, implies that any reachable state is in $(1+\epsilon)\mathcal{R}_{\bar{k}}$.

By means of the just reported property one can compute an internal approximation of the set $\mathcal{R}_\infty$ by computing $\mathcal{R}_k$. By the way, it turns out that each of the sets $\mathcal{R}_k$, under the assumption of the theorem, is a C-set as long as $\mathcal{D}$ is a C-set. The same property does not hold for the set $\mathcal{R}_\infty$, which is convex and bounded, but in general is not closed. This assertion is easily proved by the next scalar counterexample.

Example 6.10. The infinite time reachability set for the system

$$x(t+1) = \frac{1}{2}x(t) + d(t), \quad |d(t)| \leq 1$$

is clearly the open interval $\mathcal{R}_\infty = (-2, 2)$. We stress that we defined the reachability set as the set of all states that can be reached in *finite time* $0 < T < \infty$, although for T arbitrary. This is why the extrema are not included.

The situation is different if one considers the set $\bar{\mathcal{R}}_\infty = \sum_{k=0}^{\infty} A^k E \mathcal{D}$ which is closed, indeed the closure of $\mathcal{R}_\infty$ [RK07].

To achieve an external approximation one can use several tricks. The first one is that of "enlarging" $\mathcal{D}$. Indeed, if the reachability set $\mathcal{R}_t^\epsilon$ with disturbances $d \in (1+\epsilon)\mathcal{D}$ is considered, by linearity the condition

$$\mathcal{R}_t^\epsilon = (1+\epsilon)\mathcal{R}_t$$

is obtained, thus achieving an external approximation. A different trick is that of computing the reachable set for the modified system

$$x(t+1) = \frac{A}{\lambda}x(t) + \frac{E}{\lambda}d(t) \tag{6.8}$$

Denoting by $\bar{\lambda}_{max} = \max\{|\lambda_i|, \quad \lambda_i \in eig(A)\}$, for $\bar{\lambda}_{max} < \lambda \leq 1$, the system remains stable, so that the reachability sets of the modified system $\mathcal{R}_k^\lambda$ are bounded. For any $0 < \lambda < 1, A^k E\mathcal{D} \subset (A/\lambda)^k (E/\lambda)\mathcal{D}$, and then

$$\mathcal{R}_T \subset \mathcal{R}_T^\lambda = \sum_{k=0}^{T-1} \left(\frac{A}{\lambda}\right)^k \frac{E}{\lambda}\mathcal{D}$$

For λ approaching 1, $\mathcal{R}_k^\lambda$ approaches $\mathcal{R}_k$ from outside. An interesting property is the following.

Proposition 6.11. *Assume that $\mathcal{D}$ is a C-set, that (A, E) is a reachable pair and that A is asymptotically stable. Let $\lambda < 1$ be such that A/λ is stable. Then there exists $\bar{k}$ such that, for $k \geq \bar{k}$, the set $\mathcal{R}_k^\lambda$, computed for the modified system (6.8), is robustly positively invariant for the original system and*

$$\mathcal{R}_\infty \subset \mathcal{R}_k^\lambda$$

The proof of the above proposition can be deduced by the fact that:

- $\mathcal{R}_\infty^\lambda$ is positively invariant for the modified system (this fact will be reconsidered later) and then, in view of Lemma 4.31, it is contractive for the original system;
- $\mathcal{R}_k^\lambda \to \mathcal{R}_\infty^\lambda$, which has been shown in Proposition 6.9.

We refer the reader to [RKKM05a, RKK+05] for further details on this kind of approximations.

Let us now consider the problem of evaluating the reachability set for continuous-time system

$$\dot{x}(t) = Ax(t) + Ed(t).$$

It is at this point rather clear that the problem cannot be solved by considering a sequence $\mathcal{R}_t$, because such a set is not polyhedral even if $\mathcal{D}$ is such. Therefore the hyperplane method, previously considered for discrete-time systems and based on the support functional, seems the most appropriate. Let $\mathcal{R}_t$ be the set of all the states reachable in time t from the origin, with $\mathcal{D}$ a C-set. Let us consider the support functional $\phi_t(z)$ of $\mathcal{R}_t$. Then, in view of the following chain of equalities

$$\phi_t(z) = \sup_{d\in\mathcal{D}} z^T \int_0^t e^{A\sigma}Ed(\sigma)d\sigma = \int_0^t \sup_{d(\sigma)\in\mathcal{D}} z^T e^{A\sigma}Ed(\sigma)d\sigma =$$

$$\int_0^t \phi_{\mathcal{D}}(z^T e^{A\sigma}E)d\sigma$$

the reachability set in time t turns out to be the convex set characterized in terms of support functional as

$$\mathcal{R}_t = \left\{x : z^T x \leq \phi_t(z),\ \forall z\right\}$$

The reader can enjoy her/himself in investigating special cases of $\mathcal{D}$ given by her/his preferred norms. Let us consider the single input case and the set $\mathcal{D} = [-1, 1]$. In this case the support functional of $\mathcal{D}$ is $\phi_{\mathcal{D}}(\delta) = |\delta|$ and then

$$\phi_t(z) = \int_0^t |z^T e^{A\sigma} E|\, d\sigma \tag{6.9}$$

By (possibly numerical) integration, as done and reported graphically in the next example, it is possible to determine $\phi_t(z)$ and $\phi_\infty(z)$, the support functional of $\mathcal{R}_\infty$, at least approximately.

Example 6.12. By using Eq. (6.9), the reachable sets $\mathcal{R}_t$ for the continuous-time dynamic system

$$\dot{x}(t) = \begin{bmatrix} -0.3 & 1 \\ -1 & -0.3 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} d(t)$$

when the disturbance is bounded as $|d(t)| \leq 1$, were computed for $t = 1,\ 2,\ 4,\ 100$. Such sets are depicted in Figure 6.4.

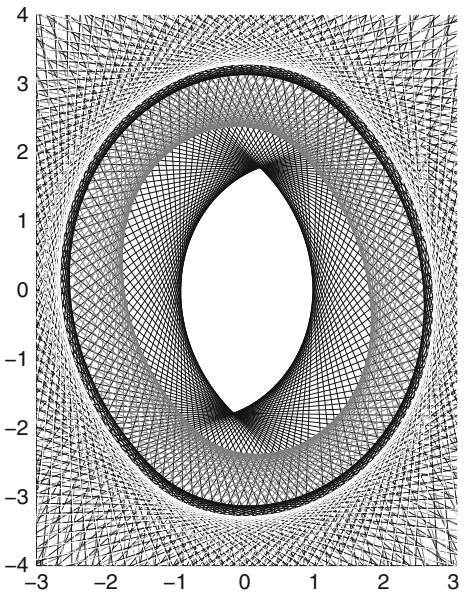


Fig. 6.4 Reachable sets with pointwise-bounded noise for Example 6.12, computed for different values of t

Henceforth we establish some properties that concern the reachability sets in both continuous and discrete-time case. We start with the following basic fact.

Proposition 6.13. *Assume that $\mathcal{D}$ is a C-set, (A, E) is a reachable pair and that A is asymptotically stable. Then $\mathcal{R}_\infty$ is the smallest robustly-positively-invariant set for the system, precisely any robustly-positively-invariant set includes $\mathcal{R}_\infty$.*

Proof. The discrete-time case only is considered (the continuous-time case proof is in practice the same). The fact that $\mathcal{R}_\infty$ is robustly-positively-invariant is obvious.

Proving minimality is equivalent to showing that any invariant set $\mathcal{P}$ contains $\mathcal{R}_\infty$, say $x \in \mathcal{R}_\infty$ implies $x \in \mathcal{P}$. Assume then that $\mathcal{P}$ is invariant, let k be arbitrary and let $\bar{x} \in \mathcal{R}_k$ be an arbitrary vector. Let $\mathcal{B}$ be any neighborhood of the origin such that any state of $\mathcal{B}$ can be driven to 0 in finite time with a signal $d(t) \in \mathcal{D}$. Consider any $x(0) \in \mathcal{P}$ and let $d(t) = 0$ until the state $x(t_1) \in \mathcal{B}$. For $t \geq t_1$ take the sequence $d(t) \in \mathcal{D}$ which drives the state to zero at t_2, $x(t_2) = 0$. Then consider the sequence of further k inputs $d(t) \in \mathcal{D}$ which drive the state $x(t_3) = \bar{x}$. Since $\mathcal{P}$ is robustly invariant, $\bar{x} \in \mathcal{P}$. Since $\bar{x}$ is arbitrary, $\mathcal{P}$ contains any point of $\mathcal{R}_k$ and since k is also arbitrary, $\mathcal{P}$ contains any point of $\mathcal{R}_\infty$.

The set $\mathcal{R}_\infty$ is the limit set of the stable system (A, E). In other words, for any $x(0)$, $x(t) \to \mathcal{R}_\infty$[3]. This fact is important because it allows to characterize the asymptotic behavior of a system. As an example of application, let us consider the problem of characterizing the worst case state estimation error.

Example 6.14 (Observer asymptotic error). Let us consider an observer-based control for the system

$$\begin{aligned}(x(t+1)) \quad \dot{x}(t) &= Ax(t) + Bu(t) + Ed(t)\\ y(t) &= Cx(t) + v(t)\end{aligned}$$

in which d and v are external inputs subject to $d(t) \in \mathcal{D}$ and $v \in \mathcal{V}$. In most cases these inputs represent noise and cannot be measured. If a standard linear observer is designed,

$$\begin{aligned}(z(t+1)) \quad \dot{z}(t) &= (A + LC)z(t) + Bu(t) - Ly(t)\\ e(t) &= z(t) - x(t)\end{aligned}$$

where $e(t)$ is the error, the error equation results in

$$(e(t+1)) \quad \dot{e}(t) = (A + LC)e(t) - Lv(t) - Ed(t) \tag{6.10}$$

It is apparent that, under persistent noises d and v, the observer error does not vanish asymptotically. The asymptotic effect of the noise can be clearly evidenced

[3] In the sense that $\delta(x(t), \mathcal{R}_\infty)$, the distance from $x(t)$ to $\mathcal{R}_\infty$ converges to 0.

by computing the reachability set of the error system (6.10). If an invariant set $\mathcal{E}$ for this system is computed, then it is possible to assure that, whenever $e(0) \in \mathcal{E}$,

$$x(t) \in z(t) - \mathcal{E}$$

$t > 0$. We will come back on this problem in Chapter 11.

So far the problem of determining the reachability set has been considered under the following assumptions: reachability of (A, E), asymptotic stability of A, and $\mathcal{D}$ a C-set. The assumption that $\mathcal{D}$ is a C-set is reasonable. If $\mathcal{D}$ has an empty interior, but 0 is inside the relative interior[4] then it is possible to reconsider the problem by redefining E. to this subspace, namely by involving a new matrix ED, where D is any basis of such a subspace.

Conversely there are cases in which the constraint set does not include 0 as an interior point. In this case the problem has to be managed in a different way. For instance, one can decompose $d \in \mathcal{D}$ by choosing a constant $d_0 \in int\mathcal{D}$. Then $d = d_0 + d_1$, where $d_1 \in \mathcal{D}_1 = \mathcal{D} - d_0$. Now, the translated disturbance d_1 is in a C-set $\mathcal{D}_1$. The effect of d_0 and d_1 can be investigated separately. An interesting case is that of systems with positive controls. We do not analyze this case but we refer the reader to specialized literature, such as [FB97].

Then let us still assume that $\mathcal{D}$ is a C-set, but let us remove the stability or the reachability assumption.

Proposition 6.15. *For the 0-reachability sets $\mathcal{R}_k$ the following properties hold:*

- *$\mathcal{R}_k \subseteq reach(A, E)$, the reachable space of (A, E).*
- *$\mathcal{R}_\infty$ is bounded if and only if the reachable sub-system of (A, E) is asymptotically stable.*
- *Assume that (A, E) is reachable, and denote by X_{sta} and X_{uns} the eigen-spaces of A associated with the stable and the unstable modes. Then the reachable set is given by*

$$\mathcal{R}_\infty = \mathcal{R}_\infty^{sta} + X_{uns}$$

where $\mathcal{R}_\infty^{sta}$ denotes the set of reachable states in the subspace X_{sta}.

Proof. The first statement is obvious. The second statement is obvious in its sufficient part because, by the previous statement we can consider the reachable sub-system and conclude that the reachability set is bounded. As far as necessity is concerned, assume that the reachable sub-system is unstable. Then, by means of a bounded input, it is possible to reach from $x(0) = 0$ an eigenvector $\bar{v}$ associated with an unstable eigenvalue λ (in general an unstable subspace) in time $[0, \bar{t}]$ and, assuming $d(t) = 0$ for $t > \bar{t}$ so that $x(t) = e^{\lambda(t-\bar{t})}\bar{v}$, it is immediate to see that

[4] 0 is in the interior relatively to the smallest subspace including $\mathcal{D}$.

$x(t)$ cannot be bounded. The third statement requires more work and its proof is postponed to the problem of controllability of systems with bounded-control.

6.2.2 *Infinite-time reachability and l_1-norm*

We now investigate an important connection between the infinite-time reachability set $\mathcal{R}_\infty$ and the l_1-norm, often referred to as ∞ to ∞ induced norm or peak-to-peak norm of a system. Consider the SISO stable system (A, E, H)

$$\begin{aligned} x(t+1) &= Ax(t) + Ed(t) \\ y(t) &= Hx(t) \end{aligned}$$

The ∞ to ∞ induced norm is defined as

$$\|H(zI - A)^{-1}E\|_{\infty,\infty} \doteq \sup_{t\geq 0, x(0)=0, |d(t)|\leq 1} |y(t)|$$

The reason why this norm is referred to as l_1-norm is that it turns out to be the l_1-norm [DP87] of the sequence of Markov parameters

$$\|H(zI - A)^{-1}E\|_{\infty,\infty} = \|H(zI - A)^{-1}E\|_{l_1} \doteq \sum_{k=0}^{\infty} |HA^kE|$$

In the general case of a MIMO (possibly not strictly proper) system the l_1-norm can be defined by replacing $|\cdot|$ by $\|\cdot\|$, precisely

$$\|H(zI - A)^{-1}E + D\|_{\infty,\infty} \doteq \sup_{t\geq 0, x(0)=0, \|d(t)\|\leq 1} \|y(t)\|$$

Such a norm can be evaluated as the sum of a series [DP87]

$$\begin{aligned} &\|H(zI - A)^{-1}E + D\|_{\infty,\infty} = \\ &\|H(zI - A)^{-1}E + D\|_{l_1} \doteq \max_i \{\|D_i\|_1 + \textstyle\sum_{k=0}^{\infty} \|[HA^kE]_i\|_1\} \end{aligned} \tag{6.11}$$

where D_i and $[HA^kE]_i$ denote the ith row of the matrices D and HA^kE, respectively. A set-theoretic equivalent condition is given in the next proposition.

Proposition 6.16. *Consider the asymptotically stable system (A, E, H) (i.e., assume $D = 0$). Then the smallest value μ_{inf} of μ such that $\mathcal{R}_\infty$ is included in the strip*

$$\bar{\mathcal{P}}[H, \mu\bar{1}] = \{x : \|Hx\| \leq \mu\}$$

is equal to the l_1-norm of the system:

$$\mu_{inf} = \inf\{\mu : \mathcal{R}_\infty \subset \bar{\mathcal{P}}[H, \mu\bar{1}]\} = \|H(zI - A)^{-1}E\|_{\infty,\infty} = \|H(zI - A)^{-1}E\|_{l_1}$$

In the single output case this is the support functional of $\mathcal{R}_\infty$ evaluated in H, i.e. $\phi_\infty(H) = \phi_\infty(-H)$ (by symmetry).

When D is non-zero, the following holds:

Proposition 6.17. *Consider the asymptotically stable system (A, E, H, D), with p outputs. Then the l_1-norm of $\|(A, E, H, D)\|_{l_1}$ is the smallest value of μ for which the 0-reachability set $\mathcal{R}_\infty$ is included in the set*

$$\bar{\mathcal{P}}[H, \tilde{\mu}\bar{1}]\,\{x : \|H_i x\| \leq \mu - \|D\|_1,\ \ i = 1, 2, \ldots, p\}$$

where $\tilde{\mu} = \mu - \|D\|_1$

Proof. It is known [DP87] hat the l_1-norm condition $\|H(zI - A)^{-1}E + D\|_{l_1} = \mu$ is equivalent to the fact that for $x(0) = 0$, the condition $\|y(t)\|_\infty \leq \mu$ holds for all $\|d(t)\|_\infty \leq 1$, namely,

$$-\mu \leq y_i(t) \leq \mu,$$

which is, in turn, equivalent to

$$-\mu \leq H_i x(t) + D_i d(t) \leq \mu$$

for all $\|d(t)\|_\infty \leq 1$. Since the current value of $d(t)$ does not depend on $x(t)$ and can be any arbitrary vector with ∞-norm not greater than 1, it is possible to write

$$-\min_{\|d\|_\infty \leq 1} D_i d - \mu \leq H_i x(t) \leq \mu - \max_{\|d\|_\infty \leq 1} D_i d$$

Then the proof is completed since

$$-\min_{\|d\|_\infty \leq 1} D_i d = \max_{\|d\|_\infty \leq 1} D_i d = \|D_i\|_1$$

The previous proposition represents an interesting interpretation of the $\|\cdot\|_{l_1}$ norm of a system in terms of reachability. In practice, the $\|\cdot\|_{l_1}$ norm less than μ is equivalent to the inclusion of $\mathcal{R}$ in $\bar{\mathcal{P}}[H, \mu\bar{1}]$. It will be soon shown that this interpretation is very useful to compute the peak-to-peak induced norm in those cases (i.e., polytopic systems) in which the computation via Markov parameters is not possible.

6.2.3 Reachable sets with energy-bounded noise

In this section, a characterization of the disturbances which is unusual in the book is analyzed. Precisely, the focus of the present section are linear dynamic systems of the form

$$\dot{x}(t) = Ax(t) + Ed(t).$$

with disturbances bounded as follows:

$$\int_0^\infty d^T(t)Rd(t)\ dt \leq 1, \quad \text{with} \quad R \succ 0$$

To avoid unnecessary complications, it is assumed, without lack of generality, that $R = I$, since if this is not the case one can replace the matrix E by $ER^{-1/2}$ and consider the input $\hat{d} = R^{1/2}d$. Let us then assume

$$\int_0^\infty d^T(t)d(t)\ dt \leq 1. \tag{6.12}$$

Denote by $\mathcal{B}(t)$ the set of all the functions having energy bounded by 1 on the interval $[0, t]$, precisely such that

$$\mathcal{B}(t) = \left\{ d(t) : \int_0^t d^T(t)d(t)\ dt \leq 1 \right\}$$

Note that the set of reachable states with inputs $d \in \mathcal{B}(t)$ is non-decreasing with t, precisely, $\mathcal{B}(t')$ includes $\mathcal{B}(t)$ for $t' > t$. Let us consider the set of all 0-reachable states with inputs bounded as above. It turns out that this set is an ellipsoid according to the following theorem. We remind that an ellipsoid $\mathcal{D}(Q) = \mathcal{D}(Q, 1)$ can be described as in (3.14)

$$\mathcal{D}(Q) = \left\{ x : z^T x \leq \sqrt{z^T Qz}, \ \text{ for all } z \right\}$$

where $\sqrt{z^T Qz}$ is its support functional [5]

Theorem 6.18. *Let A be a stable matrix and let (A, E) be a reachable pair. The closure of the set of all the states reachable from $x(0) = 0$ with inputs bounded as in (6.12) is given by the ellipsoid $\mathcal{D}(Q)$, where Q is the reachability Gramian, i.e. the unique solution of*

$$QA^T + AQ = -EE^T$$

[5] Note that $\mathcal{D}$ has not the same meaning of the previous subsection, but represents now the ellipsoid.

Proof. Consider any state $x(t)$ reachable at time t with energy bounded as

$$\int_0^t d^T(\sigma)d(\sigma)\,d\sigma \leq 1.$$

Take any vector z and consider the following optimization problem

$$\mu_t = \sup_{d\in\mathcal{B}(t)} z^T x(t) = \sup_{d\in\mathcal{B}(t)} \int_0^t z^T e^{A\sigma} E d(t-\sigma)\,d\sigma = \sup_{d\in\mathcal{B}(t)} \left(z^T e^{A(\cdot)}E, d(\cdot)\right)$$

where $(\cdot,\cdot)$ is a scalar product in the Hilbert space of the square measurable functions defined on the time interval $[0,t]$ with values in $\mathbb{R}^p$ [Lue69]. Such an optimization problem has solution

$$\mu_t = \|z^T e^{A(\cdot)}E\|_2 = \sqrt{\int_0^t z^T e^{A\sigma} E E^T e^{A^T\sigma} z d\sigma} = \sqrt{z^T Q(t) z}$$

where

$$Q(t) \doteq \int_0^t e^{A\sigma} E E^T e^{A^T\sigma} d\sigma$$

Therefore the set of all reachable states in time t is the ellipsoid $\mathcal{D}(Q(t))$. By obvious mathematical speculations, such an ellipsoid is non-decreasing with t, precisely, $z^T Q(t) z \leq z^T Q(t') z$ for $t \leq t'$. Now consider the identity

$$\int_0^t \frac{d}{d\sigma}\left[e^{A\sigma} E E^T e^{A^T\sigma}\right] d\sigma = \int_0^t \left[A e^{A\sigma} E E^T e^{A^T\sigma} + e^{A\sigma} E E^T e^{A^T\sigma} A^T\right] d\sigma$$

$$= A\left[\int_0^t e^{A\sigma} E E^T e^{A^T\sigma} d\sigma\right] + \left[\int_0^t e^{A\sigma} E E^T e^{A^T\sigma} d\sigma\right] A^T = AQ(t) + Q(t)A^T$$

On the other hand, we can write the same quantity as

$$\int_0^t \frac{d}{d\sigma}\left[e^{A\sigma} E E^T e^{A^T\sigma}\right] d\sigma = e^{At} E E^T e^{A^T t} - E E^T =$$

$$= \frac{d}{dt}\int_0^t e^{A\sigma} E E^T e^{A^T\sigma}\,d\sigma - E E^T = \dot{Q}(t) - E E^T$$

and notice that $Q(t)$ is solution of the following equation

$$\dot{Q}(t) = AQ(t) + Q(t)A^T + E E^T \tag{6.13}$$

Now, since A is stable, a finite limit value $Q = \lim_{t\to\infty} Q(t)$ exists and is achievable by setting $\dot{Q} = 0$. Then the theorem is proved, if we remind that $\mathcal{D}(Q(t))$ is non-decreasing and then included in the limit value $\mathcal{D}(Q)$. Moreover, for any $\bar{x}$ in the boundary of $\mathcal{D}(Q)$, we can find points in $\mathcal{D}(Q(t))$ arbitrarily close to $\bar{x}$, so $\mathcal{D}(Q)$ is the closure of all $\mathcal{D}(Q(t))$.

The discrete-time version of the theorem is the following:

Theorem 6.19. *Consider the system*

$$x(t+1) = Ax(t) + Ed(t)$$

with A stable and (A, E) reachable. The closure of the set of all the states reachable from $x(0) = 0$ with inputs bounded as

$$\sum_{t=0}^{\infty} d(t)^T d(t) \leq 1$$

is given by the ellipsoid $\mathcal{D}(Q)$, where Q is the discrete-time reachability Gramian which is the unique solution of

$$AQA^T - Q = -EE^T$$

Proof. (Sketch). The proof of the theorem is basically the same as the previous one. Let

$$Q(t) = \sum_{k=0}^{t-1} A^k EE^T (A^T)^k$$

so that

$$z^T [Ed(t-1)\, AEd(t-2)\, A^2 Ed(t-3) \ldots A^{t-1} Ed(0)] = \sqrt{z^T Q(t) z},$$

say the ellipsoid $\mathcal{D}(Q(t))$ is the t-step reachability set with bounded energy. The matrix $Q(t)$ clearly satisfies the equation

$$Q(t+1) = AQ(t)A^T + EE^T.$$

and its limit value is the solution of the Lyapunov equation in the theorem statement.

Remark 6.20. The same results might have been obtained, both in the discrete and the continuous-time case, by resorting to the adjoint operator theory. We have skipped that powerful and elegant approach, since the main focus here has been put on set-theoretic aspects.

6.2.4 Historical notes and comments

The history of set-propagation is wide, especially as far as the computation of reachable sets is concerned. The first contribution are due to the seminal works of Bertesekas and Rhodes [BR71a] and Glower and Schweppe [GS71, Sch73], followed by several further contributions of which only a portion is mentioned in this book. We have mentioned the work by Chernousko and Kurzhanski[Che81, KV97, Che94], which provided techniques for ellipsoidal approximation. Considering the problem of the computation of the reachability sets, the available literature is enormous and providing a survey is a major challenge. Among the first contributions, it has to be mentioned [PN71], where several types of input bounds have been considered and [HR71], where a numerical algorithm for the determination of reachable sets via amplitude bounded inputs is provided. In [GC86b] and [GC87] it has been exploited the fact that the 0-reachable sets are the 0-controllable set for the inverse system and an algorithm based on polyhedral sets has been proposed. The hyperplane method idea is due to [SS90b] and [GK91b]. See also [Gay86, Las87, SS90a, Las93] for further results on the topic.

For further references, we refer to the survey [Gay91] or to [RKKM05a] for some more recent contributions concerning the computation and approximation of the minimal invariant set [RKKM05a].

6.3 Stability and convergence analysis of polytopic systems

Stability analysis is a fundamental problem in system theory. For linear systems this trivial task requires the computation of the eigenvalues of a matrix. This method cannot be applied when dealing with an uncertain system. Let us consider again a system of the form

$$\begin{array}{c} x(t+1) = A(w(t))x(t), \quad (\text{respectively, } \dot{x}(t) = A(w(t))x(t)) \\ A(w) = \sum_{i=1}^{s} A_i w_i, \quad \sum_{i=1}^{s} w_i = 1, \quad w_i \geq 0 \end{array} \tag{6.14}$$

with the basic questions:

- is the system stable?
- assumed that it is, how fast does it converge?

These questions will be faced next by means of quadratic and non-quadratic Lyapunov functions.

6.3.1 Quadratic stability

One approach to the problem is inspired by the well-known fact that any stable linear time-invariant system admits a quadratic Lyapunov function, leading to the following criterion.

Theorem 6.21. *The discrete-time (resp. continuous-time) system (6.14) is stable if all the systems share a common quadratic Lyapunov function, equivalently, if there exists a positive definite matrix P such that*

$$A_i^T P A_i - P \prec 0, \quad (\textit{respectively } A_i^T P + P A_i \prec 0)$$

for $i = 1, 2, \ldots, s$.

Corollary 6.22. *The condition of the theorem is equivalent to the existence of* $\epsilon > 0$, $\beta > 0$ *or* $0 \leq \lambda < 1$ *such that, for all i*

$$A_i^T P A_i - \lambda^2 P \preceq 0, \quad (\textit{respectively } A_i^T P + P A_i + 2\beta P \preceq 0), \quad P \succ \epsilon I$$

The easy proof of this theorem (the corollary follows obviously) is not reported here. We will come back on it later, when we will show that the provided condition is sufficient, but not necessary at all. To provide necessary and sufficient conditions one might think about resorting to another family of Lyapunov functions. The class of polyhedral Lyapunov functions is an appropriate one as we will show soon.

6.3.2 Joint spectral radius

To provide non-conservative and constructive solutions to the stability analysis of a Linear Parameter-Varying (LPV) system one can consider the procedure for the construction of the largest invariant and the basic finite determination of Theorem 5.17. To investigate on this matter, a connection with the joint spectral radius is established.

Given a square matrix A its spectral radius is defined as the largest modulus of its eigenvalues $\Sigma(A) = \max\{|\lambda| : \lambda \in eig(A)\}$. For a set of matrices the joint spectral radius of the set is defined as the supremum of the spectral radius of all possible products of generating matrices.

Definition 6.23 (Joint spectral radius). Given a finite set of square matrices $[A_1, A_2, \ldots, A_s]$, the quantity

$$\Sigma(A_1, A_2, \ldots, A_s) \doteq \limsup_{k \geq 0} \max_{C_k \in \mathcal{I}_k} \Sigma(\Pi_{C_k})^{\frac{1}{k}} \tag{6.15}$$

is said the joint spectral radius of the family [RS60].

We remind that C_k is a string of k elements of $\{1, 2, \ldots, s\}$ and Π_{C_k} is the product of the matrices A_i indexed by the corresponding elements. The above quantity can be equivalently defined as

$$\Sigma(A_1, A_2, \ldots, A_s) = \limsup_{k\to\infty} \max_{C_k \in \mathcal{I}_k} \|\Pi_{C_k}\|^{1/k}.$$

(the quantity does not depend on the adopted norm) and it is related to the notion of Lyapunov exponent [Bar88a, Bar88b, Bar88c]. The following property is well known:

Proposition 6.24. *The robust exponential stability of the discrete-time system $x(t+1) = A(w(t))x(t)$ as in (6.14) is equivalent to $\Sigma(A_1, A_2, \ldots, A_s) < 1$.*

Proof. It is obvious that $x(t+1) = A(w(t))x(t)$ stable implies that the switching

$$x(t+1) = A(k)x(t), \quad A(k) \in \mathcal{A} = \{A_1, A_2, \ldots, A_s\}$$

is stable hence $\Sigma(A_1, A_2, \ldots, A_s) < 1$. The converse statement can be proved by using Proposition 6.5. Consider any initial polytopic set $\mathcal{X}_0$. The T-steps reachable set of the discrete inclusion is included in the convex hull of the points

$$A_{i_{T-1}} A_{i_{T-2}} \ldots A_{i_0} v_j, \quad A_{i_t} \in \mathcal{A},\ v_j \in vert\{\mathcal{X}_0\}$$

Thus, if $\Sigma(A_1, A_2, \ldots, A_s) < 1$, these points converge to 0 as $T \to \infty$.

The following theorem holds.

Theorem 6.25. *Assume that the matrices in the set have no common proper non-trivial invariant subspaces*[6]*. Then the following implications hold.*

i) If the spectral radius $\Sigma(A_1, A_2, \ldots, A_s) < 1$, then for any initial polyhedral C-set $\mathcal{X}$, the largest invariant set $\mathcal{S}$ included in $\mathcal{X}$ is represented by a finite number of inequalities.

ii) Conversely, if $\Sigma(A_1, A_2, \ldots, A_s) > 1$, then there exists $\bar{k}$ such that

$$\mathcal{S}^{(\bar{k})} \subset int\{\mathcal{X}\}$$

Proof. See [BM96b].

It has to be stressed that claim i) of the theorem holds even in the case in which the A_i share a common invariant subspace, which is the case of a single matrix A [GT91]. Statement ii) requires the assumption (see Exercise 11).

As previously mentioned, this implies that the procedure for computing $\mathcal{S}$ can be used to check the stability of a system. The following theorem formalizes this fact.

[6] Say there is no proper subspace $\mathcal{G}$, $\{0\} \neq \mathcal{G} \subset \mathbb{R}^n$ such that $A_i\mathcal{G} \subset \mathcal{G}$, for all i.

Let us consider the sequence (5.25) of sets $\mathcal{S}^{(k)}$ computed for the modified system

$$z(t+1) = \frac{A(w(t))}{\lambda} z(t) \tag{6.16}$$

(note that $x(t) = z(t)\lambda^t$ if $x(0) = z(0)$), which turns out to be

$$\mathcal{S}_\lambda^{(k)} = \{x : F\frac{\Pi_{C_h}}{\lambda^h} x \leq 1, \;\; C_h \in \mathcal{I}_h \;\; h = 0, 1, 2, \ldots, k\}$$

The next theorem formalizes some of the properties concerning the spectral radius.

Theorem 6.26. *Define the following numbers:*

$$\lambda_1 = \inf\{\lambda > 0 : \;\; \textit{the modified system (6.16) is stable}\}$$
$$\lambda_2 = \inf\{\lambda > 0 : \|x(t)\| \leq C\|x(0)\|\lambda^t, \;\; \textit{for some} \;\; C > 0\}$$
$$\lambda_3 = \inf\{\lambda > 0 : \;\; \mathcal{S}_\lambda^\infty \;\; \textit{is a C-set}\}$$
$$\lambda_4 = \inf\{\lambda > 0 : \;\; \mathcal{S}_\lambda^\infty = \mathcal{S}_\lambda^{(k)} \;\; \textit{for a finite } k\}$$

Then

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \Sigma(A_1, A_2, \ldots, A_s)$$

Proof. The fact that $\lambda_1 = \lambda_2 = \Sigma(A_1, A_2, \ldots, A_k)$ is a well-known result, see, for instance, [Bar88a, Bar88b, Bar88c]. The remaining equalities are immediate consequence of the previous theorem.

It follows immediately from Theorems 6.25 and 6.26 that, in the case of a single linear time-invariant system $x(t) = Ax(t)$, the procedure stops in a finite number of steps if $\Sigma(A) < \lambda$ for any C-set $\mathcal{X}$ or determines a set $\mathcal{S}^{(k)}$ which is in the interior of $\mathcal{X}$ if $\Sigma(A) > \lambda$ (this is in perfect agreement with the earlier result in [GT91]).

A remarkable consequence which can be drawn from Theorem 6.26 is that, in principle, the joint spectral radius can be approximately computed by bisection, by increasing (resp. decreasing) λ if the numeric procedures, applied to the modified system, stops unsuccessfully (resp. successfully). As previously pointed out, the procedure produces a number of constraints which increases enormously when $\lambda \simeq \Sigma$. This is in agreement with the work presented in [TB97, BT00] which analyzes the complexity of computing or approximating, the joint spectral radius of matrices and which can provide an explanation of this phenomenon (although there are particular interesting cases in which the complexity can be reduced, see [BNT05]) The reader is referred to [BN05] for more details and references on this topic. We will show later also that the considered type of procedures can be used to compute, beside the spectral radius, other performance indices for uncertain systems.

6.3.3 Polyhedral stability

To face the problem of robust stability analysis we can exploit the fact that polyhedral stability is equivalent to stability for an LPV system, as stated next.

Theorem 6.27. *The following statements are equivalent.*

1. *The discrete-time (continuous-time) system (6.14) is asymptotically stable.*
2. *The discrete-time (continuous-time) system (6.14) is exponentially stable, namely there exists* (C, λ), $0 \leq \lambda < 1$ *(resp.* (C, β), $\beta > 0$*) such that*

$$\|x(t)\| \leq C\|x(0)\| \, \lambda^t \tag{6.17}$$

(respectively

$$\|x(t)\| \leq C\|x(0)\| \, e^{-\beta t}) \tag{6.18}$$

3. *The system admits a polyhedral norm* $\|Fx\|_\infty$ *as a Lyapunov function. Precisely, there exists* $0 \leq \lambda < 1$ *(resp.* $\beta > 0$*) such that*

$$\|FA(w)x\|_\infty \leq \lambda \|Fx\|_\infty, \quad (\textit{resp. } D^+\|FA(w)x\|_\infty \leq -\beta\|Fx\|_\infty,) \ \forall w$$

4. *All the vertex systems* $x(t+1) = A_i x(t)$ *share a common polyhedral Lyapunov function* $\|Fx\|_\infty$.
5. *For any signal* $v(t)$*, with* $\|v(t)\| \leq \nu$*, there exist* β*,* C_1 *and* C_2 *such that the solution of the system* $x(t+1) = A(w(t))x(t) + v(t)$ *(resp.* $\dot{x}(t) = A(w(t))x(t) + v(t)$*) is bounded as*

$$\|x(t)\|_* \leq C_1\|x(0)\| \, \lambda^t + C_2, \quad (\textit{resp. } \|x(t)\|_* \leq C_1\|x(0)\| \, e^{-\beta t} + C_2)$$

Proof. The proof of the equivalence of the first three statements is reported in [MP86a, MP86b, MP86c] and [Bar88a, Bar88b, Bar88c]. See also the work in [BT80]. The equivalence 3–4 is easy, while the equivalence of statement 5 to the other ones is a tedious exercise (suggested but not required to the reader).

The theorem, as stated, is non-constructive. To check stability of an assigned discrete-time polytopic system (along with the determination of a proper polyhedral Lyapunov function, whenever stable) it is possible to proceed iteratively as previously mentioned. Indeed it is possible to use the procedure described in Section 5.4, starting from an *arbitrary* polyhedral set $\mathcal{X}^{(0)}$. Precisely, given the initial set $\mathcal{X}^{(0)} = \{x : \ \|F^{(0)}x\|_\infty \leq 1\}$ it is possible to recursively compute the sets

$$\begin{aligned}
\mathcal{X}^{(k+1)} &= \{x :\in \mathcal{X}^{(k)} : \ A_i \, x \in \mathcal{X}^{(k)}, \ \ i = 1, 2, \ldots, s\} \\
&= \{x : \|F^{(k)}x\|_\infty \leq 1, \ \ \|F^{(k)}A_i \, x\|_\infty \leq 1, \ \ i = 1, 2, \ldots, s\} \\
&\doteq \{x : \ \ \|F^{(k+1)}x\|_\infty \leq 1\}
\end{aligned}$$

Theorem 6.25 assures that if the system is stable then no matter how the polyhedral C-set $\mathcal{X}$ is chosen the largest invariant set included in it is also a polyhedral C-set and can be determined by a recursive procedure in a finite number of steps.

This theorem can be used "the other way around." Precisely, one can try to compute $\Sigma(A_1, A_2, \ldots, A_s)$ by computing the largest invariant set for the system

$$x(t+1) = \frac{A(w(t))}{\lambda} x(t)$$

and to reduce/increase λ if the procedure stops successfully/unsuccessfully. In detail, given a tentative λ one runs the procedure and

- decreases λ if for some $\bar{k}$

$$\mathcal{S}^{(\bar{k})} = int\mathcal{S}^{(\bar{k}-1)} (= \mathcal{S})$$

- increases λ if for some $\bar{k}$

$$\mathcal{S}^{(\bar{k})} \subset int\{\mathcal{X}^{(0)}\}$$

According to Theorem 6.25, under the assumption that the matrices do not admit a common proper invariant subspace (unless for the critical value $\lambda = \Sigma$), both conditions are detected in a finite number of steps. We will come back on this later, when we will deal with the more general problem of computing the best transient estimate.

For the continuous-time case one can, once again, resort to the EAS

$$x(t+1) = [I + \tau A(w)]x(t)$$

supported by the next proposition.

Proposition 6.28. *The following two statements are equivalent.*

- *The continuous-time system is stable and admits the Lyapunov function* $\|Fx\|_\infty$.
- *There exists* $\tau > 0$ *such that the EAS is stable and admits the Lyapunov function* $\|Fx\|_\infty$.

Proof. See [BM96a].

Therefore, the stability of a continuous-time polytopic system can be established by applying the previously described bisection algorithm to the EAS. In this case, there are two parameters on which it is necessary to iterate: λ and τ. One possibility to avoid this double iteration is that of iterating over the parameter τ only by assuming $\lambda(\tau) = 1 - \rho\tau^2$, as already mentioned in Section 5.2.

A possibility of reducing the complexity of the computation of the Lyapunov function is based on the following Proposition, which basically states that the stability of a differential inclusion is unchanged if we multiply it by a positive function.

Proposition 6.29. *Consider the differential inclusion*

$$\dot{x}(t) = \rho(t)A(w(t))x(t), \quad 0 < \rho^- \leq \rho(t) \leq \rho^+ \tag{6.19}$$

Then its stability does not depend on the bounds $0 < \rho^- \leq \rho^+$. *In particular it is stable iff* $\dot{x}(t) = A(w(t))x(t)$ *is stable.*

Proof. We prove sufficiency, since necessity is obvious. If $\dot{x}(t) = A(w(t))x(t)$ is stable, then it admits a polyhedral Lyapunov function $\Psi(x)$ such that $D^+\Psi(x, A(w)x) \leq -\beta\Psi(x)$. If we consider this function for (6.19) we get, denoting by $h' = h\rho$,

$$D^+\Psi(x, \rho A(w)x) = \lim_{h\to 0^+} \sup \frac{\Psi(x + h\rho A(w)x) - \Psi(x)}{h}$$

$$= \lim_{h'\to 0^+} \sup \frac{\Psi(x + h'A(w)x) - \Psi(x)}{h'} \rho = \rho D^+\Psi(x, A(w)x) \leq -\beta\rho\Psi(x)$$

Note that multiplication by $\rho > 0$ is equivalent to a time scaling: it changes the speed of convergence, but cannot compromise stability.

As a simple corollary, in the case of polytopic systems we can replace the generating matrices by scaled matrices

$$A(w) = \sum_{i=1}^{s} \rho_i A_i \, w_i$$

with positive scalars $\rho_i > 0$ without affecting the stability/instability properties. As an immediate consequence, when we consider the EAS for the computation of a polyhedral function, we can adopt different τ_i for different matrices. Precisely stability of the continuous-time system can be proven by computing a polyhedral function for the "EAS".

$$x(t+1) = [I + \sum_{i=1}^{s} \tau_i A_i w_i]x(t)$$

This property can be applied as follows. Given a single stable A, the eigenvalues of the EAS are $1 + \tau\lambda_i$, where λ_i are the eigenvalues of A. If τ is small enough, then $I + \tau A$ is stable, but if τ is too small, then the discrete-time eigenvalues are squeezed to 1, so that the discrete-time contractivity is very low. In general, different matrices A might suggest different values of τ. We can take advantage of this fact in computing a Lyapunov function, reducing both the computation time and the function complexity.

Example 6.30. Consider the polytopic system generated by the two matrices

$$A_1 = \begin{bmatrix} 0 & 1 \\ -\frac{1}{4} & -1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 1 \\ -\frac{7}{4} & -1 \end{bmatrix}.$$

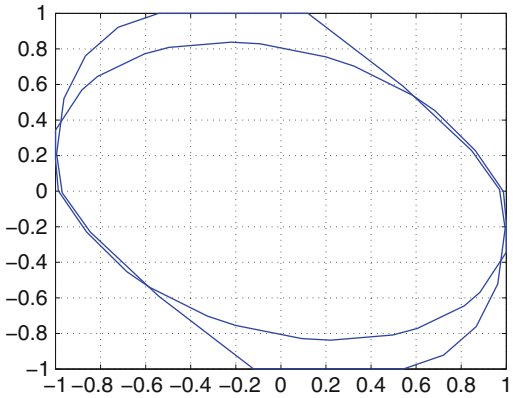


Fig. 6.5 The computed regions with a single $\tau = 0.2$ (external) and with two different values $\tau_1 = 0.5$ and $\tau_2 = 0.2$ (internal)

The polyhedral Lyapunov function computed with $\tau = 0.2$ considering the unit square (unit ball of $\|\cdot\|_\infty$) as initial set, produced a unit ball with 28 delimiting planes. The eigenvalues of A_1 are $-0.5, -0.5$ while those of A_2 are $-0.5 \pm 1.225j$. Those of the EAS are $0.9, 0.9$, and $0.9 \pm 0.24495j$. If we notice that it is reasonable to take a smaller τ for A_1 than for A_2, then we can take $\tau_1 = 0.5$, So the EAS has eigenvalue $0.75, 0.75$, and $\tau_2 = 0.2$. The resulting function is represented by a unit ball of 14 delimiting planes (Fig. 6.5). Clearly by no means the stability of the two discrete-time matrices assures convergence and continuous-time stability. In general, we will have to reduce all the τ_i when the procedure stops unsuccessfully.

6.3.4 The robust stability radius

Let us now consider the problem of computing the "robustness measure." Consider the system $\dot{x}(t) = A(w(t))x(t)$ (or $x(t+1) = A(w(t))x(t)$), with

$$A(w) = [A_0 + \Delta(w(t))], \quad \Delta(w) \in \rho\mathcal{W}$$

where $\mathcal{W}$ is compact and $\Delta(w)$ is continuous. The robustness measure we are thinking about is reported in the next definition.

Definition 6.31. Assuming A_0 a stable matrix

$$\rho_{ST} = \sup\{\rho : \text{ the system is robustly stable}\}$$

In the discrete-time case it is possible to apply the bisection procedure, precisely by starting with a tentative ρ, and to increase/reduce it if the computed set includes/does-not-include the origin in its interior. Thus, by applying the proposed procedure and by iterating by bisection on ρ, it is possible to derive an upper and lower bound on ρ_{ST}.

This algorithm may be directly applied to polytopic discrete-time systems in which

$$A(w) = A_0 + \rho[\sum_{i=1}^{s} \Delta_i w_i], \quad \sum_{i=1}^{s} w_i = 1, \quad w_i \geq 0$$

with Δ_i assigned. As mentioned above, to consider continuous-time systems, one can use the EAS and iterate over τ.

6.3.5 *Best transient estimate*

Detecting stability only can be a non-sufficient task. One could be interested in evaluating the transient quality. To this aim, one can evaluate the evolution with respect to a given norm $\|\cdot\|_*$ by computing a transient estimate.

Definition 6.32. A transient estimate is a pair (C, λ) (respectively (C, β)) for which (6.17) (respectively (6.18)) holds for the solution.

Note that, in principle, λ may be any non-negative number and β any number. In other words, it is possible to estimate the transient of an unstable system (thus determining the "speed of divergence").

Let us consider the problem of computing a transient estimate with respect to the ∞-norm $\|x\|_\infty$ (the procedure immediately generalizes to any polyhedral norm of the form $\|Fx\|_\infty$). This can be done, in the discrete-time case, by performing the following steps.

Procedure. Computation of a transient estimate, given a contraction factor λ.

1. Fix a positive $\lambda < 1$.
2. Compute the largest invariant set $\mathcal{P}_\lambda$ inside the unit ball of the ∞-norm $\mathcal{X} = \mathcal{N}[\|\cdot\|_\infty, 1]$, for the modified system $x(t+1) = (A(w)/\lambda)x(t)$. Note that $\mathcal{P}_\lambda$ is the largest λ-contractive set for the considered system.
3. If $\mathcal{P}_\lambda$ has empty interior, then the transient estimate does not exist for the given λ (then one can increase λ and go back to Step 2).
4. Determine $C_\lambda > 0$ as the inverse of the largest factor μ such that $\mu\mathcal{X}$ is included inside $\mathcal{P}_\lambda$

$$C_\lambda^{-1} = \max_{\mu > 0} \ \text{s.t.} \ \mu\mathcal{X} \subseteq \mathcal{P}_\lambda$$

It can be shown that C_λ is the smallest constant such that (C_λ, λ) is a transient estimate. It is then clear that, by iterating over λ, it is possible to determine the "best transient estimate" (see [BM96a] for details). It turns out that if the system

converges (or diverges) with speed $\lambda_0 < \lambda$, then the set $\mathcal{P}_\lambda$ is a polyhedral set and, as we have already seen, the procedure for its generation converges in a finite number of steps.

The same procedure can be used for continuous-time systems as follows. We can fix $\beta > 0$ and consider a small τ such that $\lambda(\tau) = 1 - \tau\beta < 1$. Then apply the procedure with such a λ to the EAS. It turns out that if the system converges with speed of convergence $\lambda_0 > \beta$ then, for sufficiently small τ, it is possible to compute a λ-contractive polyhedral set for the EAS and then a β-contractive polyhedral set with $\beta = (1 - \lambda)/\tau$.

Note that in principle, the transient estimate could be computed by means of any Lyapunov function, possibly quadratic, as shown later on. However the results are conservative.

Example 6.33. We report as an example the continuous-time system considered in [Zel94], with $m = 2$

$$A_1 = \begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 1 \\ -2 - \Delta & -1 \end{bmatrix}.$$

$\Delta = \Delta(t) \geq 0$. Quadratic stability is assured for this system if and only if $0 \leq \Delta < \Delta_Q \approx 3.82$ [Zel94] (this bound can be also obtained via standard continuous-time $\mathcal{H}_\infty$ analysis, as it will shown later). Zelentsowsky [Zel94] found the stability limit $\Delta_Z = 5.73$, say a 50% improvement. By using homogeneous polynomial Lyapunov functions and LMI techniques, in [CGTV03] it was shown that stability is assured for $\Delta_S = 6.7962$. Though not explicitly dealing with transient estimates, it is worth recalling that those techniques can be applied to the problem as well. Using the (EAS) with $\tau = 2.5 \times 10^{-4}$, $\lambda = 1 - 1 \times 10^{-9}$ and the polyhedral Lyapunov function construction, we were able to determine a polyhedral function for $\Delta_P = 6.97$. The computed transient estimate corresponding to Δ_Q, Δ_Z, and Δ_P are $(C_Q, \beta_Q) = (2.5439, 0.14)$, $(C_Z, \beta_Z) = (2.7068, 0.02)$ $(C_P, \beta_P) = (2.7805, 4.0 \times 10^{-6})$. The unit ball $\{x : \|Fx\|_\infty \leq 1\}$ of the Lyapunov function corresponding to Δ_P is reported in Fig. 6.6.

As it has been underlined several times, polyhedral Lyapunov functions are non-conservative. However, they generally require algorithms for the generation of their unit ball that are extremely heavy from the computational standpoint. The number of planes necessary for the description of such sets can drive out-of-range the most powerful machines, even for trivial instances. Clearly a transient estimate can be computed by means of quadratic function. If a positive definite matrix P such that

$$A_i^T P + P A_i + 2\beta I \prec 0$$

is found, then the corresponding family of ellipsoids $\mathcal{E}(P, \nu)$ is β-contractive. This in turn implies that one can take a β-contractive ellipsoid $\mathcal{E}(P, \nu)$ included in the box $\mathcal{X}$ and including $\mu\mathcal{X}$ for a proper $\mu \leq 1$. Then (C, β) with $C = 1/\mu$ is a transient estimate. Clearly such a transient estimate is, in general, conservative, not only because β is smaller, but also because C is quite greater than the best

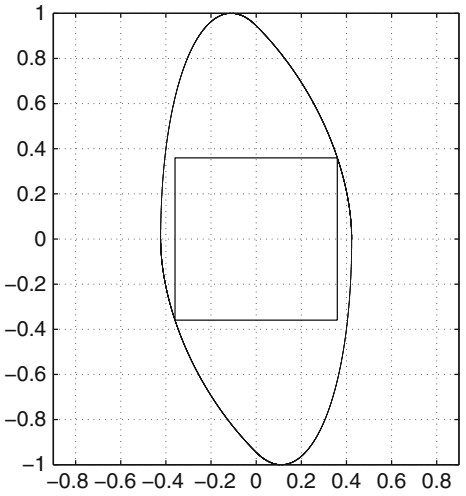


Fig. 6.6 The computed region for $\beta = 4.0 \times 10^{-6}$ and the inscribed region $1/C_P\mathcal{X}$

transient estimate (see Exercise 12). On the other hand, the computation is much easier. Indeed in the inclusion and containment constraints

$$\mu\mathcal{X} \subset \mathcal{E}(P, \nu) \subset \mathcal{X} \tag{6.20}$$

only the variables μ, ν, and P come into play. Besides, there are several problems, such as finding the smallest invariant set including a polytope or the largest invariant set in a polytope (in the sense of volume), that have the further important property of being convex in their variable and therefore very efficient algorithms are available. The reader is referred to [BEGFB04] for further details. As a final remark, it should be mentioned that the proposed analysis does not take into account variation of speed limits in the parameter. Taking into account these limits makes the problem harder (see, for instance, [Ran95, ACMS97]).

6.3.6 Comments about complexity and conservativity

Polyhedral functions are non-conservative, but computationally demanding[7]. Thus considering polyhedral functions instead of quadratic ones can be dramatic since the former might be extremely complex. A legitimate question is whether this is always the case. We show by means of a simple example that there are systems which are not quadratically stabilizable, but they admit a polyhedral function whose representation is not more complex than the representation of a quadratic function.

[7]Perhaps the reader will find this a tedious repetition in the book, still this conservativeness issue was not well known in the control literature for a long period [Ola92, Bla95].

Example 6.34 (A low complexity polyhedral function). Consider the four matrices

$$A_1 = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1 & 0 \\ -1 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & -1 \\ 0 & -1 \end{bmatrix},$$

and the system

$$\dot{x}(t) = \sum_{i=1}^{4} w_i[-\varepsilon I + A_i]x(t), \quad \sum_{i=1}^{4} w_i = 1, \ \ w_i \geq 0 \tag{6.21}$$

with $\varepsilon > 0$ sufficiently small. The system admits $V(x) = \|x\|_1$ as a common Lyapunov function. Indeed any of the generating matrices $-\varepsilon I + A_i$ has $\|\cdot\|_1$ as an LF because it is strictly diagonally dominant, with negative diagonal entries.

We show that there are no common quadratic positive definite Lyapunov functions. To prove this, we first note that the set of matrices $\{A_k, \ k = 1, 2, 3, 4\}$ is invariant with respect to the following transformations

$$T_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad T_2 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad T_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

namely changes of signs or reflections along the bisectors, since, for every choice of $T_i, i = 1, 2, 3$, we have $\{T_i^{-1} A_k T_i, \ k = 1, 2, 3, 4\} = \{A_k, \ k = 1, 2, 3, 4\}$. This amounts to saying that for every $i = 1, 2, 3$ and every $k = 1, 2, 3, 4$ there exists $j = 1, 2, 3, 4$ such that $T_i^{-1} A_k T_i = A_j$. This same property applies to the matrices $A_k - \varepsilon I, k = 1, 2, 3, 4$. Consequently, if the positive definite matrix

$$P_1 = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

defines a common quadratic Lyapunov function for the matrices $A_k - \varepsilon I, k = 1, 2, 3, 4$, so does

$$P_2 = \begin{bmatrix} a & -b \\ -b & c \end{bmatrix} = T_1^{-1} P_1 T_1.$$

Since the set of common Lyapunov matrices for $\{A_k - \varepsilon I, \ k = 1, 2, 3, 4\}$ is a convex cone, then

$$P_3 = \frac{P_1 + P_2}{2} = \begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}$$

defines a common quadratic Lyapunov function for the matrices $A_k - \varepsilon I, k = 1, 2, 3, 4$. But since the set $\{A_k, \ k = 1, 2, 3, 4\}$ is also invariant over bisector reflections, the positive definite matrix

$$P_4 = T_3^{-1} P_3 T_3 = \begin{bmatrix} c & 0 \\ 0 & a \end{bmatrix}$$

and hence the scalar matrix

$$\frac{P_4 + P_3}{2} = \begin{bmatrix} a+c & 0 \\ 0 & c+a \end{bmatrix} \frac{1}{2}$$

obtained as the average of P_3 and P_4, both define common quadratic Lyapunov functions. This implies that $P = I_2$ defines a common Lyapunov matrix, in other words that $\hat{A}_k^T + \hat{A}_k < 0$ for every $k = 1, 2, 3, 4$. To verify that such condition is not true it is sufficient to compute

$$\det\left(-\hat{A}_1^T - \hat{A}_1\right) = \det\left(\begin{bmatrix} 2(1+\varepsilon) & -1 \\ -1 & 2\varepsilon \end{bmatrix}\right) = 4\varepsilon^2 + 4\varepsilon - 1$$

which is clearly negative for $0 < \varepsilon < \frac{-1+\sqrt{2}}{2}$.

6.3.7 Robust stability/contractivity analysis via system augmentation

A possibility to investigate stability contractivity in a less conservative way than using quadratic Lyapunov functions is based on system augmentation. Let us consider the discrete-time first. Assume we wish to establish the stability of the system

$$x(t+1) = [\sum_{i}^{s} A_{i=1} w_i(t)] x(t), \qquad \sum_{i=1} w_i = 1, \;\; w_i \geq 0$$

or equivalently, we wish to check if $\Sigma(A_1, A_2, \ldots, A_s) < 1$.

We can consider the T step system defined as follows:

$$x(k+T) = [A_{i_{T-1}} A_{i_{T-2}} \ldots A_{i_0}] x(t) = \Phi_t x(t) \tag{6.22}$$

where

$$\Phi_t \in \bar{\mathcal{A}}_T \doteq [A_{i_{T-1}} A_{i_{T-2}} \ldots A_{i_0}]$$

are the matrices formed by all possible T-products of the given A_i. The following proposition holds.

Proposition 6.35. *The difference inclusion is stable (or $\Sigma(A_1, A_2, \ldots, A_k) < 1$) if and only if for T large enough (6.22) is quadratically stable. Moreover, for T large enough, any quadratic positive definite function $x^T P x,\ P \succ 0$ is a suitable quadratic Lyapunov function for system (6.22).*

Proof. The proof is similar to that of Proposition 6.24 and left to the reader as an exercise.

It should be noticed that the previous proposition is essentially a re-statement of old theory, for instance [GY93].

To apply the criterion we have two possibilities. One possibility is to fix an horizon T and check if all the matrices $\Phi \in \bar{\mathcal{A}}_T$ share a common Lyapunov function. The second one is to fix an horizon (usually much larger) to check if $\Phi \in \bar{\mathcal{A}}_T$ have all norms $\|\Phi\| < 1$. The shortcoming of the approach is that the number of matrices $\Phi \in \bar{\mathcal{A}}_T$ grows exponentially.

In the continuous-time case we can use a different system augmentation. This technique was used for the first time in [Zel94] and deeply investigated later [CGTV03, Che10, CGTV09, CCG+12]. The idea, explained in brief, sounds as follows. Instead of $x(t)$ we introduce the variable $x^{(m)}(t)$ formed by all the monomials of order m. For instance, for $x(t) \in \mathbb{R}^2$

$$x^{(3)}(t) = [x_1^3(t), x_1^2(t)x_2(t), x_1(t)x_2^2(t), x_2^3(t)]^T$$

Consider the linear system

$$\dot{x}(t) = Ax(t)$$

Then the system in the new variable $x^{(m)}$ is described by the following "expanded" dynamic system

$$\dot{x}^{(m)}(t) = A^{(m)}x^{(m)}(t)$$

where the matrices $A^{(m)}$ can be computed as shown in [CGTV09]. Now it is obvious that the stability of the original system and of the expanded one are equivalent.

If we consider a quadratic candidate Lyapunov function for the new system

$$\Psi(x^{(m)}) = (x^{(m)})^T P x^{(m)}$$

this function turns out to be a polynomial Lyapunov function for the original system [CGTV09]. We know that the class of positive polynomials are universal, hence non-conservative for the robust stability problem [MP86a, MP86b, MP86a, BM99c]. It has recently been proved that the stability of the original system is equivalent to the quadratic stability of the extended system for m large enough [Che11b, CCG+12, Che13].

6.4 Performance analysis of dynamical systems

In this section, it is shown how several problems related to the performance analysis of dynamic systems may be solved via a set-theoretic approach. We start by considering the fact that, as we have seen, for Linear Time-Invariant (LTI) systems, basic properties such as the effect of bounded inputs on the system can be in practice solved without set-computing. For instance, the evaluation of the l_1 norm of a system (i.e., the worst-case output peak for all possible peak-bounded inputs) requires the computation of the sum of a series. We have given a set-theoretic interpretation which has its own interest but it does not provide practical or theoretical advantages. Here it is shown how the set-theoretic formulation can be used to solve some analysis problems for uncertain systems for which the formulas known for LTI systems are no longer useful.

6.4.1 Peak-to-peak norm evaluation

Let us consider the problem of evaluating the largest output value achievable by the constrained inputs with 0 initial conditions for the discrete-time polytopic system

$$\begin{aligned} x(t+1) &= A(w(t))x(t) + Ed(t) \\ y(t) &= Hx(t) \end{aligned}$$

where, again, $A(w) = \sum_{i=1}^{s} A_i w_i$, with $w \in \mathcal{W}$, namely, $\sum_{i=1}^{s} w_i = 1$, $w_i \geq 0$ and d belongs to the compact set $\mathcal{D}$.

The paradigm consists in the following question: assume $x(0) = 0$ and let $d(t) \in \mathcal{D}$. Is the constraint

$$\|y(t)\|_* \leq \mu,$$

(with $\|\cdot\|_*$ a given norm) satisfied for all $t \geq 0$?

In the case in which also $\mathcal{D}$ is the unit ball of the same norm $\|\cdot\|_*$, we are evaluating the system induced norm. Formally the question is

- Q0:

$$\|(A(w), E, H)\|_{*,*} = \sup_{\substack{w(t) \in \mathcal{W} \\ x(0) = 0 \\ \|d(t)\|_* \leq 1,\ t \geq 0}} \sup_{t>0} \|y(t)\|_* \leq \mu\,?$$

The actual system norm can be estimated by iterating over μ. One way to proceed is that of computing the convex hulls of the 0-reachable sets. From the results

previously presented it is indeed apparent that, denoting by $\mathcal{R}_t$ the 0 reachable set in t steps, the sequence of the convex hulls $conv\{\mathcal{R}_t\}$ can be computed as shown in Proposition 6.5. In view of the above consideration, an "yes" answer is equivalent to checking that

$$conv\{\mathcal{R}_t\} \in \mathcal{Y}(\mu) \doteq \mathcal{N}[\|Hx\|_*, \mu],$$

(roughly the set of all x such that $\|Hx\|_* \leq \mu$), for all t. This way of proceeding has the drawback that if the previous condition is satisfied till a certain $\bar{t}$, there is no guarantee that the same condition will be satisfied in the future. As it often happens, inverting the reasoning can be helpful. This is equivalent to reverting time, in this case. The problem can be solved in two steps as follows.

- Compute the largest robustly invariant set $\mathcal{P}_\mu$ for system $x(t+1) = A(w(t))x(t) + Ed(t)$ inside $\mathcal{Y}(\mu)$.
- If $0 \in \mathcal{Y}(\mu)$ then the answer to Q0 is "yes", otherwise it is "no".

In principle we should assume that the system has passed the stability test. Under some assumptions, such as the existence of an observable pair$(A(w), H)$ the stability test is actually included in the procedure according to following theorem.

Theorem 6.36. *Assume that there exists $w' \in \mathcal{W}$ such that $(A(w'), H)$ is an observable pair and that there exists $w'' \in \mathcal{W}$ such that $(A(w''), E)$ is reachable. The following statements are equivalent.*

- *All the reachable sets (equivalently, their convex hulls) are inside $\mathcal{Y}(\mu)$, say $\mathcal{R}_t \subset \mathcal{Y}(\mu)$, for all $t > 0$.*
- *The largest robustly invariant set $\mathcal{P}_\mu$ included in $\mathcal{Y}(\mu)$ is a C-set.*
- *The system is stable and question Q0 has answer "yes" (in the case of the induced norm $\|(A(w), E, H)\|_{*,*} \leq \mu$).*

Proof. The set $\mathcal{P}_\mu$ is the region of initial states starting from which the condition $x(t) \in \mathcal{Y}(\mu)$ is guaranteed for all $t \geq 0$. Therefore, the first two statements are obviously equivalent to the third statement, with the exception of the "stability claim." To include stability, we need to consider the observability and reachability assumption. Indeed, if we assume that $(A(w'), H)$ is observable, then the closed and convex set $\mathcal{P}_\mu$ is necessarily bounded [GT91]. Furthermore, if $(A(w), E)$ is reachable, then the reachable set $\mathcal{R}_T$ includes the origin as an interior point for all $T > 0$ (for T large enough in the discrete-time case) and then $\mathcal{P}_\mu$ is a $C - set$. Then we are in the position of proving stability.

Take any initial condition x_0 on the boundary of the C-set $\mathcal{P}_\mu$. The corresponding solution is given by $x(t) = x_f(t) + x_d(t)$ where $x_f(t)$ is the free response (i.e., such that $x_f(t+1) = A(w(t))x_f(t)$ and $x_f(0) = x_0$) and $x_d(t)$ is the response driven by d (precisely $x_d(0) = 0$ and $x_d(t+1) = A(w(t))x_d(t) + Ed(t)$. Since $x_d(T) \in \mathcal{R}_T$, then

$$x(T) \in \{x_f(T)\} + \mathcal{R}_T \subseteq \mathcal{P}_\mu$$

Denote by $\mathcal{S}_T = conv\{\mathcal{R}_T\}$ the convex hull of $\mathcal{R}_T$. Being $\mathcal{P}_\mu$ convex the last inclusion can be replaced by

$$x(T) \in \{x_f(T)\} + \mathcal{S}_T \subseteq \mathcal{P}_\mu$$

therefore $\{x_f(t)\}$ is in the erosion, $[\tilde{\mathcal{P}}_\mu]_{\mathcal{S}_T}$, of (see Definition 3.8) $\mathcal{P}_\mu$ with respect to $\mathcal{S}_T$

$$x_f(T) \in [\mathcal{P}_\mu]_{\mathcal{S}_T}$$

Since $\mathcal{S}_T$ is a C-set there exists $\lambda < 1$ such that $x_f(T) \in \lambda\mathcal{P}_\mu$.

We have proved that, for all $x_0 \in \partial\mathcal{P}_\mu$, we have that in T steps $x_f(T) \in \lambda\mathcal{P}_\mu$. Consider the T-step forward system

$$x_f(t+T) = [A(w(t+T-1))A(w(t+T-2))\dots A(w(t))]\, x_f(t)$$

which is linear, hence homogeneous. By applying Theorem 4.18 and Lemma 4.31 we can see $\mathcal{P}_\mu$ (which is invariant for $x_f(t+1) = A(w(t))x_f(t)$) is λ-contractive for such a system which implies stability.

We sketch now the algorithm proposed in [FG95] and [BMS97] that can be used for the $\|\cdot\|_\infty$ norm. Precisely we assume that $\|d(t)\|_\infty \leq 1$ and we seek for the largest possible $\|y(t)\|_\infty$.

1. Fix an initial guess $\mu > 0$ and a tolerance $\epsilon > 0$.
2. Set $F^{(0)} = H$, $g^{(0)} = \mu\bar{1}$, $k = 0$, $\mu^+ = +\infty$ and $\mu^- = 0$.
3. If $\mu^+ - \mu^- \leq \epsilon$ STOP. Else
4. Given the set $\mathcal{S}_k = \{x : |F_i^{(k)}x| \leq g_i^{(k)},\ \ i = 1, 2, \dots, r^{(k)}\}$, where $F_i^{(k)}$ is the ith row of matrix $F^{(k)}$ and $g_i^{(k)}$ is the ith component of vector $g^{(k)}$, compute the pre-image set $\mathcal{P}_{k+1}$ as

$$\mathcal{P}_{k+1} = \{x : |F_i^{(k)}A_j x| \leq \mu^{(k)} - \|F_i^{(k)}E\|_1,\ \ j = 1, 2, \dots, s,\ \ i = 1, 2, \dots, r^{(k)}\}$$

5. Compute the intersection

$$\mathcal{S}_{k+1} \doteq \mathcal{P}_{k+1} \bigcap \mathcal{S}_k$$

   to form the matrix $F_i^{(k+1)}$ and the vector $g^{(k+1)}$.
6. If $0 \notin \mathcal{S}_{k+1}$, then set $\mu^- = \mu$, increase μ and GOTO step 3.
7. If $\mathcal{S}_k = \mathcal{S}_{k+1}$, then set $\mu^+ = \mu$, reduce μ and GOTO step 3,

The previous results can be applied to continuous-time systems by means of the EAS. It can be shown that the ∞-to-∞ induced norm of the EAS system is always an upper bound for the corresponding induced norm of the continuous-time system

$$\|(A, E, H)\|_{\infty,\infty} \leq \|((I + \tau A), \tau E, H)\|_{\infty,\infty}$$

This fact can be inferred from the property that $\|((I+\tau A),\tau E,H)\|_{\infty,\infty} \leq \mu$ implies the existence of an invariant set for the EAS included in $\mathcal{Y}(\mu)$. In view of Lemma 4.26 and Proposition 5.10, such an invariant set is positively invariant for the continuous-time system.

The computation of the norm of system

$$(A(w_A),E(w_E),H(w_H)) = \left(\sum A_i w_{A,i}, \sum E_i w_{E,i}, \sum H_i w_{H,i}\right)$$

with polytopic structure can be handled as follows. The input $E(w_E)d$ is replaced by $v \in \mathcal{V}$ the convex hull of all possible points of the form $E_k d_h$, with E_k and d_h on their vertices. It is quite easy to see that the convex hulls of the reachability sets of $x(t+1) = A(w_A)x(t) + v(t)$ are the same as those of the original system. As far as the output uncertainty is concerned, the condition to be faced is

$$\|y\|_\infty = \|H(w_H)x\|_\infty \leq \mu \Leftrightarrow \|y^{(k)}\|_\infty = \|H_j x\|_\infty \leq \mu, \quad \forall j$$

Therefore the problem requires repeating the iteration for all matrices H_j and retaining the minimum value. Note that, in this extension, it has been assumed that the uncertainties affecting $(A(w_A),E(w_E),H(w_H))$ are independent.

We remind the reader that the induced norm for the time-varying uncertain system we are considering here, say $\|(A(w),E,H)\|_{\infty,\infty}$, is quite different from the time-invariant norm, namely $\|(A(\bar{w}),E,H)\|_{\infty,\infty}$, the norm computed for the time-invariant system achieved by fixing $w = \bar{w}$. Clearly the time-invariant norm is not greater than the time-varying worst case norm:

$$\|(A(\bar{w}),E,H)\|_{\infty,\infty} \leq \|(A(w),E,H)\|_{\infty,\infty}$$

Example 6.37. Let us consider the following system

$$A(w) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -(3+w) & -2 & -3 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1\ 0\ 0 \end{bmatrix},$$

$w \in [0,1]$, and let us consider the corresponding EAS $(I+\tau A,\tau E,H)$. The algorithm provided the following limits for the norm

$$13.5 = \mu^- \leq \|((I+\tau A(w)),\tau E,H)\|_{\infty,\infty} \leq \mu^+ = 13.6$$

Note that the upper bound is actually an upper bound for the continuous-time system, while the lower bound is not. The algorithm required 3638 iterations to detect that the origin was not included in the largest invariant set for $\mu = \mu^- = 13.5$ and required 144 iterations to find an invariant set $\mu = \mu^+ = 13.6$. Such an invariant set is generated by 1292 constraints (by symmetry these correspond to

646 rows for the describing matrix F, not reported for paper-saving reasons). In the next table we report the number of iterations to detect the "yes/no" answer and the number of residual constraints (the actual constraints forming $\mathcal{P}_\mu$ in the "YES" answer case) as a function of the "guess" μ. We point out that the norms

μ	8	10	12	13.5	13.6	14	16	18
$\|\| \leq \mu$?	NO	NO	NO	NO	YES	YES	YES	YES
iterations	646	969	1554	3638	144	53	44	43
constraints	226	276	270	288	1292	586	530	524

of the extreme systems are quite smaller, $\|((I + \tau(A(0)), \tau E, H)\|_{\infty,\infty} \approx 1.98$ $\|((I + \tau(A(1)), \tau E, H)\|_{\infty,\infty} \approx 5.95$, and this means that high values of the norm are not due to a special "critical" value of w, but mainly to its variation inside $\mathcal{W}$.

Clearly the performance of the system might be estimated via ellipsoids. Let us consider the following problem. Consider the system

$$\dot{x}(t) = A(w)x(t) + Ed(t), \quad y(t) = Hx(t)$$

with

$$\|d(t)\| \leq \frac{1}{\mu}.$$

Now we assume that the norm is the Euclidean one (in the single input case it does not matter). Then we can consider the condition (4.23) (see [Sch73, USGW82]) to state that the ellipsoid $\mathcal{E}(P, 1)$ is positively invariant if, denoting by $Q = P^{-1}$, we have, for all i:

$$QA_i^T + A_iQ + \alpha Q + \frac{1}{\alpha}EE^T\frac{1}{\mu^2} \preceq 0, \quad \text{for some} \quad \alpha > 0 \tag{6.23}$$

The condition (4.23) has been stated for a single system $\dot{x} = Ax + Ed$, but the generalization above is obvious. The problem is that of including the ellipsoid $\mathcal{E}(Q^{-1}, 1)$ inside the strip $\mathcal{Y}(1)$. Note that for convenience we are iterating over μ by scaling the control disturbance rather than changing the size of $\mathcal{Y}$ which is obviously equivalent.

The condition $\mathcal{E}(Q^{-1}, 1) \subset \mathcal{Y}(1)$ can be easily expressed. Let us consider the single-input case for brevity. Then $\mathcal{Y}(1) = \{x : |Hx| \leq 1\}$, so that $\mathcal{E}(Q^{-1}, 1) \subset \mathcal{Y}(1)$ iff

$$HQH^T \leq 1. \tag{6.24}$$

Then, if we find a matrix $Q \succ 0$ such that conditions (6.23) and (6.24) are satisfied, then we are sure that the induced norm of the system is less than μ. If such an ellipsoid does not exist, however, we cannot conclude that the induced norm of the system is greater than μ.

Example 6.38. To show that the previous condition can be conservative, consider the example in [USGW82], namely the system $\dot{x}(t) = Ax(t) + Bu(t) + Ed(t)$ with matrices

$$A = \begin{bmatrix} -0.0075 & -0.0075 & 0 \\ 0.1086 & -0.149 & 0 \\ 0 & 0.1415 & -0.1887 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ -0.0538 \\ 0.1187 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0037 \\ 0 \\ 0 \end{bmatrix}$$

to which the linear feedback control

$$u = Kx = -37.85x_1 - 4.639x_2 + 0.475x_3$$

is applied. Four outputs were considered: the state components and the control input. On all these variables, constraints are imposed as follows:

$$|x_1| \leq 0.1, \quad |x_2| \leq 0.01, \quad |x_3| \leq 0.1, \quad |u| \leq 0.25$$

which can be written as $\|Hx\|_\infty \leq 1$, where

$$H = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 10 \\ -151.40 & -18.55 & 1.90 \end{bmatrix}.$$

The disturbance input is bounded as $|d| \leq \alpha$. The ellipsoidal method provides the bound

$$\alpha_{ell} = 1.27$$

which implies the bound for the induced norm equal to $\|(A, E, H)\|_{\infty,\infty} \leq (1.27)^{-1} = 0.787$ By considering the EAS with $\tau = 1$, we achieved the bound

$$\alpha_{EAS} = 1.45$$

which implies $\|(A, E, H)\|_{\infty,\infty} \leq (1.45)^{-1} = 0.685$. Clearly, by reducing τ, tighter bounds can be achieved. We will reconsider this example later as a synthesis benchmark. We remind that the condition is also necessary [BC98] for positive invariance (if there exists a reachable pair $(A(w), E)$), therefore conservativeness is not due to condition (6.23), but to the adoption of ellipsoids.

6.4.2 Step response evaluation

We consider now the problem of computing the peak of the step response of the system

$$x(t+1) = A(w(t))x(t) + Ed(t), \qquad y(t) = Hx(t) + Gd(t)$$

where it is assumed that $d(t) \equiv 1$ and $x(0) = 0$. Basically the values one would like to evaluate for this system are the worst case peak and the asymptotic error and precisely, for given positive μ and ν, the questions now are:

Q1 : is the largest peak bound less than μ, $\sup_{t\geq 0} \|y(t)\| \leq \mu$?
Q2 is the largest asymptotic value less than ν, $\limsup_{t\to\infty} y(t) \leq \nu$?

We can answer this questions as follows. We assume that $A(w)$ is stable. Let us consider the following sets

$$\mathcal{Y}(\xi) = \{x : \|Hx + G\| \leq \xi\}$$

(remind that $d \equiv 1$). Then we can claim the following.

Proposition 6.39.

- *The answer to question Q1 is yes if and only if the largest invariant set included in $\mathcal{Y}(\mu)$ includes the origin.*
- *The answer to question Q2 is yes if and only if the largest invariant set included in $\mathcal{Y}(\nu)$ is non-empty.*

The proof of this proposition can be found in [BMS97] where a more general case with both disturbances and constant inputs is considered.

Again, in terms of ellipsoids, a bound can be given as suggested in [BEGFB04] (see notes and references of Chapter 6). Indeed, the unit step is a particular case of norm-bounded input. However, as pointed out in [BEGFB04], the method is conservative (see Exercise 8).

Example 6.40. Consider the system

$$\dot{y}(t) = -[1 + w(t)/2]y(t) + u(t)$$

with the integral control

$$\dot{u}(t) = -\kappa(y(t) - r(t)).$$

Assume $r(t) \equiv \bar{r} = 1$, $\kappa = 5$ and $u(0) = y(0) = 0$. We use the EAS with $\tau = 0.1$, so achieving the system

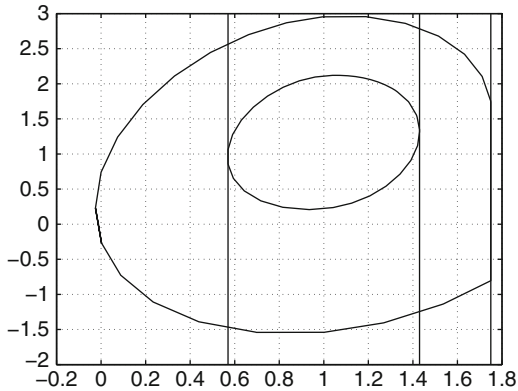


Fig. 6.7 The maximal invariant set (external) and the limit set (internal)

$$A = \begin{bmatrix} [0.85, 0.9] & 1 \\ -0.5 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix},$$

We first compute the output peak with respect to output y, asking

Q1: is condition $\sup_{t\geq 0} |y(t)| \leq \mu$ true for all $w(t)$?

It turns out that $\mu^+ = 1.75$, $\mu^- = 1.74$ are the upper and lower limits for the "yes" answer. In Figure 6.7 the largest invariant set included in $\mathcal{Y}_{max}(1.75)$ is depicted. Its margin is the rightmost vertical line, which includes the origin and certifies that $\mu^+ = 1.75$ is actually an upper limit. Then we compute the asymptotic behavior of the error $y - r$, the question now is

Q2: is the condition $\limsup_{t\to\infty} |y(t) - \bar{r}| \leq \nu$, true for all $w(t)$?

It turns out that the limit for the "yes" answer is between $\nu^+ = 0.430$ and $\nu^- = 0.429$. The smaller set in Fig. 6.7 is the largest invariant set included in the set $\mathcal{Y}_{lim}(0.430)$, the strip included between the two leftmost lines which certifies that the asymptotic behavior of y is between $\bar{r} + \nu^+ = 1.430$ and $\bar{r} - \nu^+ = 0.570$ (we remind that $\bar{r} = 1$)

The conclusions that can be drawn are the following. The step response of the system with the considered control does not exceed 1.75 as a peak, no matter how $0 \leq w(t) \leq 1$ changes. The asymptotic error is clearly not constant unless w has a limit value $0 \leq \bar{w} \leq 1$ (in which case the integrator assures $e(t) \to 0$). For persistent fluctuating values of w, in agreement with the considerations in Subsection 2.1.3. the error fluctuates and the worst case (for the EAS) is 0.430, which assures that the worst case for the continuous-time system does not exceed 0.430. It is intuitively clear that by taking τ smaller and smaller one converges to the actual value for the continuous-time system. Such intuition is supported by the results in [BS94], where such an assertion is proved. In Figure 6.8 a simulated step response is proposed. The value of $w(t)$ is alternatively taken equal to 0 and 1 starting with $w(0) = 0$ and by switching at $t = 3,\ 13,\ 15,\ 17,\ 19,\ 21$. It appears that the estimated values are sensibly larger than the actual ones. These are essentially due to two reasons. First, the realization of $w(t)$ considered in the simulation is not necessarily the "worst

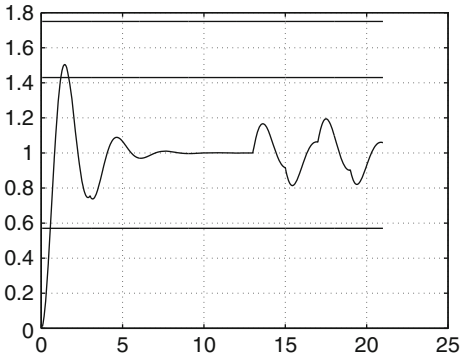


Fig. 6.8 The proposed simulation

case." Second, the provided bounds are non-conservative as $\tau \to 0$. Thus, we could reduce the value of τ (the considered one is 0.1), at the price of a noticeable increase in the number of planes delimiting the set.

6.4.3 Impulse and frequency response evaluation

It is possible to analyze impulse responses in the set-theoretic framework. Consider the SISO system

$$x(t+1) = A(w(t))x(t) + Ed(t), \quad y(t) = Hx(t)$$

with $w(t) \in \mathcal{W}$, $x(0) = 0$, $d(t) = \delta_0(t) = \{1, 0, 0, \dots\}$, and assume that $(A(\tilde{w}), H)$ is observable for some $\tilde{w} \in \mathcal{W}$. The question is to find

$$\sup_{t \geq 0} |y(t)|_\infty$$

The problem can be reformulated as by fixing a $\mu > 0$ and checking if $\sup_{t\geq 0} |y(t)|_\infty \leq \mu$. By iterating over μ we can solve the problem up to a numerical approximation. We have the following.

Proposition 6.41. *Assume that the system is asymptotically stable. Then the impulse response y is such that* $\sup_{t\geq 0} |y(t)|_\infty \leq \mu$ *if and only if the (finitely determined) largest invariant set in the strip* $\{x : |Hx| \leq \mu\}$ *for the system includes the vector E.*

Note that, in principle the step response analysis proposed in the previous subsection could be carried out by augmenting the (stable) system

$$x(t+1) = A(w(t))x(t) + Ed(t), \tag{6.25}$$

by adding a fictitious equation

$$d(t+1) = d(t), \quad d(0) = 1$$

and testing the impulse response for the resulting system with output $y(t) = Hx(t) + 0d(t)$. The only problem is that, in this way the augmented system is not stable anymore and then there is no way to assure that the algorithm which computes the largest invariant set converges in finite time. To fix the problem, we can decide to accept the approximation achieved by replacing the equation $d(t+1) = d(t)$ by a slow decay

$$d(t+1) = \lambda d(t)$$

with $0 < \lambda < 1$ and $\lambda \approx 1$. With this kind of tricks we can manage other kind of problems. For instance, we can augment system (6.25) by adding the second order system

$$z(t+1) = R(\theta)z(t) + Pr(t), \quad d(t) = z_1(t)$$

where $R(\theta)$ is the θ-rotation matrix

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

and $P = [\ 1 \quad 0\ ]^T$. By means of the impulse response of this plant we can test the frequency response amplitude of the original plant (at frequency θ). Again the convergence of the algorithm is an open issue, since the augmented system is not asymptotically stable.

6.4.4 Norm evaluation via LMIs

We briefly now discuss some problems that can be solved by means of methods which are in some sense related to the set-theoretic approach we are dealing with. An important performance index for a system is the so-called $\mathcal{L}_2$-to-$\mathcal{L}_2$ induced gain which can be defined as follows:

$$\|(A, E, H, G)\|_{2,2} = \sup_{w \neq 0} \frac{\|y\|_2}{\|d\|_2}$$

where the $\mathcal{L}_2$ norm is defined as

$$\|u\|_2 = \sqrt{\int_0^\infty u(\sigma)^T u(\sigma)\, d\sigma}$$

It is well known that, if A is stable, such a norm has the fundamental frequency response characterization

$$\|(A,E,H,G)\|_{2,2} = \sup_{\omega \geq 0} \sqrt{\max \text{eig}\left[W(j\omega)^T\ W(-j\omega)\right]}$$

where $W(s) = H(SI-A)^{-1}E + G$ is the transfer function matrix and $\text{eig}(W^TW)$ is the set of the eigenvalues of W^TW (which are real non-negative) This norm is also referred to as $\mathcal{H}_\infty$-norm. It is known that the property $\|(A,E,H,G)\|_{2,2} < 1$ has an LMI characterization [SPS98]. Let us assume now that (A,E,H,G) has a polytopic structure

$$[A,E,H,G] = \sum_{i=1}^{s}\ w_i\ [A_i,E_i,H_i,G_i]$$

$\sum_{i=1}^{s}\ w_i = 1$, $w_i \geq 0$, Then the induced norm condition $\|(A,E,H,G)\|_{2,2} < 1$ is assured if there exists a positive definite P such that

$$\begin{bmatrix} A_i^TP + PA_i & PE_i & H_i^T \\ E_i^TP & -I & G_i^T \\ H_i & G_i & -I \end{bmatrix} \prec 0$$

Again this condition is a complete characterization (i.e., it provides a necessary and sufficient condition) for a single system (A,E,H,G), but for polytopic systems it is only sufficient when the condition is far to be necessary. Indeed as it will be seen later, there exist stable systems, therefore with finite induced gains, which are not quadratically stable (a condition which is implied by the previous LMI condition). Clearly, the discrete-time version of the problem has also an LMI characterization and the reader is referred to specialized literature.

Similar considerations can be done for the computation of the impulse response energy. Precisely, one might be interested in the evaluation of $\|(A,E,H)\|_2$, the $\mathcal{L}_2$ system norm, defined as the $\mathcal{L}_2$ norm of the system impulse response (for simplicity, the SISO case only is considered). Such norm is then equal to

$$\|(A,E,H)\|_2 = \|y_{imp}\|_2 = \sqrt{\int_0^\infty (He^{At}E)^T He^{At}E\ dt} = \sqrt{E^TPE}$$

where $P \succ 0$ is the unique (assuming (A,H) observable) solution to the equation

$$A^TP + PA + H^TH = 0$$

Let us consider a polytopic system. Assume that there exists $P \succ 0$ such that

$$A_i^TP + PA_i + H^TH \prec 0$$

and define the function $\Psi(x) = x^T P x$. Then, for any initial condition $x(0)$, the free response is such that

$$\dot{\Psi}(x) = x^T[A(w)^T P + PA(w)]x \leq -x^T H^T H x = -y^T y$$

say $y^T y \leq -\dot{\Psi}(x)$. By integrating we have

$$\int_0^t y(t)^T y(t)\, dt \leq \Psi(x(0)) - \Psi(x(t))$$

Consider the system impulse response y, namely the free evolution with initial condition $x(0) = E$. Then, in view of the assumed asymptotic stability, $\Psi(x(t)) \to 0$ as $t \to \infty$, then

$$\|y_{imp}\|_2^2 \leq E^T P E$$

Again this is not a tight bound since the condition implies quadratic stability, which is stronger than stability.

6.4.5 Norm evaluation via non-quadratic functions

It is clear that if we consider bounds based on quadratic functions, then the system has to be quadratically stable. So the criterion is conservative for polytopic systems.

In general, given a stable system of the form

$$\dot{x}(t) = A(w(t))x(t) + Ed(t), \tag{6.26}$$

$$y(t) = Hx(t) \tag{6.27}$$

and a positive definite positively homogeneous function of the second order $\Psi(x)$, from a condition of the form

$$D^+\Psi(x) \leq -y^2(t) + \gamma d^2(t)$$

by integration we get, assuming $d(t) \to 0$ and $x(t) \to 0$

$$\int_0^\infty y^2(t)dt \leq \gamma \int_0^\infty d^2(t)dt + \Psi(x_0)$$

where x_0 is the initial state. The function $\Psi(x)$ is not necessarily quadratic and we can derive a polytopic bound on the output energy of the impulse response as follows. For brevity we consider the SISO case and $d \equiv 0$.

Consider the set

$$\mathcal{Y} = \{x : |Hx| \leq 1\}$$

and compute a β-contractive (possibly the largest) set $\mathcal{S}$ inside $\mathcal{Y}$. Consider the corresponding Minkowski functional $\Psi(x)$, for which, if $d = 0$, we get $D^+\Psi(x) \leq -\beta\Psi(x)$. Let $\psi(x) = \Psi^2(x)$. We get

$$D^+\psi(x) \leq -\frac{1}{\mu}\psi(x)$$

where $\mu \doteq (2\beta)^{-1}$. On the other hand, by construction, $\mathcal{S} \subset \mathcal{Y}$, so $\psi(x) \geq y^2$ (because the 1-level surface of ψ, $\mathcal{S} = \mathcal{N}[\psi, 1]$ is included in the 1-level surface of y^2, namely $\mathcal{Y}$. Hence

$$D^+\psi(x) \leq -\psi(x)/\mu \leq -y^2/\mu$$

By integrating we get for $d = 0$

$$\int_0^\infty y^2(t)dt \leq \mu\Psi(x_0)$$

which provides a bound for the output energy with initial condition x_0, so $\Psi(E)$ is a bound for the energy of the impulse response (say, when $d(t) = \delta(t)$).

The computation of the set $\mathcal{S}$ can be performed as previously described.

6.5 Periodic system analysis

We briefly consider the analysis problem of periodic systems. It is a known problem in the mathematical literature and we sketch some basic results. Consider the system

$$\dot{x}(t) = f(x(t), w(t))$$

and assume that f is Lipschitz and that $w(t)$ is a periodic signal of period T. A basic question considered in the literature is the existence of periodic trajectories. Clearly, the periodicity of $w(t)$ does not imply the existence of a periodic trajectory. However, there are some sufficient condition. Assume that there exist a C-set $\mathcal{X}$, t_0 and a period $T > 0$ such that, for all $x(t_0) \in \mathcal{X}$, $x(t_0 + T) \in \mathcal{X}$. Then, there exists a periodic trajectory. This fact can be shown by considering the Brouwer fixed-point theorem. Consider the map $F : \mathcal{X} \to \mathcal{X}$ which associates to $x \in \mathcal{X}$ the solution of the equation with initial condition $x(t_0)$ at time $t_0 + T$ $x(t_0 + T) = F(x(t_0))$. The map

F is continuous in view of the continuous dependence on the initial condition. Therefore there exists $\bar{x} \in \mathcal{X}$ such that $F(\bar{x}) = \bar{x}$, which implies that the solutions which starts from $\bar{x}$ at t_0 is T-periodic.

However, this basic result does not characterize the behavior of the periodic solution, for instance as far as its stability is concerned. Here we propose some results for systems of the form

$$\dot{x}(t) = A(t, w(t))x(t), \qquad \text{or, as usual } \; x(t+1) = A(t, w(t))x(t) \tag{6.28}$$

with $A(t, w)$ periodic in t. For this class of systems, stability implies exponential stability, as proved below.

Theorem 6.42. *Assume that in Eq. (6.28) $A(t, w)$ is continuous and periodic of period T, for any fixed $w \in \mathcal{W}$, with $\mathcal{W}$ compact. Assume that (6.28) is globally uniformly asymptotically stable (GUAS), according to Definition 2.16. Then it is exponentially stable.*

Proof. If the system is GUAS, for all $\mu > 0$ and $\epsilon > 0$, there exists an integer $\kappa = \kappa(\mu, \epsilon) > 0$ such that if $\|x(0)\| \leq \mu$ then $\|x(t)\| \leq \epsilon$, for all $t \geq \kappa T$ and it is bounded as $\|x(t)\| \leq \nu \; 0 \leq t \leq \kappa T$, for some $\nu > 0$. Take $\mu = 1$ and $\epsilon = \mu/2 = 1/2$. Then $\|x(\kappa T)\| \leq 1/2$. Consider the modified system

$$\dot{z}(t) = [\beta I + A(t, w(t))]z(t),$$

and recall that if $x(0) = z(0)$ then, $z(t) = e^{\beta t}x(t)$ is the solution of the modified system, since

$$\frac{d}{dt}(xe^{\beta t}) = \beta e^{\beta t}x + e^{\beta t}\dot{x} = [\beta I + A(t, w)](xe^{\beta t}).$$

Take $\beta > 0$ small enough to assure that $\|z(\kappa T)\| \leq 1$. Then, since $\|z(0)\| \leq 1$ implies $\|z(\kappa T)\| \leq 1$ and since $z(t)$ is bounded for $0 \leq t \leq \kappa T$, by the assumed periodicity we have that $\|z(r\kappa T)\| \leq 1$ for all integer r and that z is bounded, say $\|z(t)\| \leq \rho$ for some $\rho > 0$. Therefore

$$\|x(t)\| = \|e^{-\beta t}z(t)\| = e^{-\beta t}\|z(t)\| \leq e^{-\beta t}\rho$$

for all $\|x(0)\| \leq 1$, and thus also for $\|x(0)\| = 1$. In view of the linearity, in general we have

$$\|x(t)\| \leq e^{-\beta t}\rho\|x(0)\|$$

thus exponential stability.

The previous result, as a particular case, proves that an LPV system is stable if and only if it is exponentially stable, precisely the equivalence of the first two items of Theorem 6.27.

In the case of discrete-time periodic systems, stability can be checked by algorithms, which are based on the approach previously described. Indeed, one can start the backward construction (see Section 5.1.2) of the sets

$$\mathcal{X}_{-k-1} = \left\{ x : \frac{A(t,w)}{\lambda} x \in \mathcal{X}_{-k} \right\} \bigcap \mathcal{X}_0$$

starting from any arbitrary C-set $\mathcal{X}_0$. It can be shown that the sequence of sets, which is nested in T steps

$$\mathcal{X}_{-k-T} \subseteq \mathcal{X}_{-k}$$

either collapses to the origin or converges to a periodic sequence. The occurrence of the latter proves stability of the system under consideration. Precisely, assume that

$$\mathcal{X}_{-k-T} = \mathcal{X}_{-k}$$

(a condition which is typically met within a certain tolerance). The above states the fact that $x(t) \in \mathcal{X}_{-k}$ implies $x(t+T) \in \lambda^T \mathcal{X}_{-k}$, where λ is the contractivity factor.

The provided set-theoretic approach to performance evaluation can be easily extended to non-autonomous periodic systems. Consider, for instance, the system

$$x(t+1) = A(t,w(t))x(t) + Ed(t), \quad y(t) = Hx(t)$$

with $A(t,w)$ periodic in t with period T and d belonging to the C-set $\mathcal{D}$. Assume that one wishes to check if the worst case magnitude is $\|y(t)\|_\infty \leq \mu$. Then, setting $\mathcal{X}_0 = \mathcal{Y}(\mu) = \{x : \|Hx\| \leq \mu\}$, it is possible to start a similar backward construction:

$$\mathcal{X}_{-k-1} = \{x : A(t,w)x + Ed \in \mathcal{X}_{-k}\} \bigcap \mathcal{X}_0.$$

Again the sequence of sets is nested in T steps, say $\mathcal{X}_{-k-T} \subseteq \mathcal{X}_{-k}$. The sequence either stops due to an empty element, $\mathcal{X}_{-k} = \emptyset$, or converges to a periodic sequence [BU93].

6.6 Exercises

1. Show an example of a set $\mathcal{S}$ controllable to $\mathcal{P}$ such that $\mathcal{P}$ is not reachable form $\mathcal{S}$ and vice versa.
2. Show that if f is continuous, and if $\mathcal{P}$ and $\mathcal{U}$ are compact, then

$$f(\mathcal{P},\mathcal{U})$$

   is compact (too easy?).
3. Assume that $\mathcal{P}$ is controlled-invariant. Show that, for $T_1 \leq T_2$, $\mathcal{C}_{T_1}(\mathcal{P}) \subseteq \mathcal{C}_{T_2}(\mathcal{P})$ where $\mathcal{C}_T(\mathcal{P})$ is the controllability set in time T. Show that the implication $\mathcal{R}_{T_1}(\mathcal{P}) \subseteq \mathcal{R}_{T_2}(\mathcal{P})$ is not true in general.
4. Explain why $\mathcal{C}_T(\mathcal{P})$ is not compact, even for a compact $\mathcal{P}$, in the case of discrete-time linear systems (Hint: take A singular ...).
5. Prove Proposition 6.7.
6. Show, by means of an example, that the one step reachable set from an ellipsoid $\mathcal{E}$ for the system $x(t+1) = Ax(t) + Ed(t)$, $d \in \mathcal{D}$ is convex, but in general it is not an ellipsoid, no matter if $\mathcal{D}$ is an ellipsoid or a polytope.
7. The set $\mathcal{R}_\infty$ with bounded input $d \in \mathcal{D}$ is robustly positively invariant. Is the set $\mathcal{R}_\infty(\bar{x})$ of all states reachable from $\bar{x} \neq 0$, for some arbitrary $T > 0$, positively invariant? Is the set$\mathcal{C}_\infty(\bar{x})$ of all states controllable to $\bar{x} \neq 0$ for arbitrary $T > 0$, positively invariant?
8. Given a stable system, the ratio between a) the maximum (worst case) output peak persistent disturbance inputs $|d(t)| \leq 1$, and b) unit step output, may be arbitrarily large. Can you show a sequence of LTI systems for which this ratio grows to infinity?
9. The l_1-norm of a MIMO system (A,E,H) is defined as follows. Denote by $Q^{(1)}, Q^{(2)}, \ldots, Q^{(k)}, \ldots$ the sequence of Markov parameters ($p \times m$ matrices). Then the l_1 norm is defined as

$$\|H(zI-A)^{-1}E\|_{l_1} = \sup_i \sum_{k=1}^{\infty} \sum_{j=1}^{m} |Q_{ij}^{(k)}|$$

   This norm is known to be equal to

$$\|H(zI-A)^{-1}E\|_{\infty,\infty} \doteq \sup_{t\geq 0, x(0)=0, \|d(k)\|_\infty \leq 1} \|y(t)\|_\infty$$

   Provide the "reachability set" characterization of this norm which is the MIMO version of Proposition 6.16.
10. Formulate a "convex" optimization problem to find P, μ, and ν which satisfy (6.20).

11. The statement ii) of Theorem 6.25 does not hold, in general, if the matrices share a proper invariant subspace. Show this by considering the single matrix $A = diag\{2, 1/2\}$ and $\mathcal{X}$ the unit square.
12. Consider the system $x(t+1) = Ax(t)$ with

$$A = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

and the norm $\|\cdot\|_\infty$ and $1/\sqrt{2} \leq \lambda < 1$. Find the best transient estimate (the largest λ-contractive set is delimited by 8 planes). What about the transient estimate evaluated with the Lyapunov norm $\|\cdot\|_2$?

Chapter 7
Control of parameter-varying systems

In this chapter the problem of feedback control of uncertain systems is considered, with a special attention to the control of polytopic systems. To properly introduce the results, let us reconsider the stability analysis problem for an uncertain system of the form

$$\dot{x}(t) = A(w(t))x(t)$$

where $w(t) \in \mathcal{W}$, with $\mathcal{W}$ compact, and $A(\cdot)$ is continuous. In the case of a single stable linear system, stability is equivalent to the fact that the eigenvalues of A have negative real part (modulus less than one in the discrete-time case). The speed of convergence associated with the maximum real part (modulus) of the eigenvalues, precisely the maximum value of $\beta > 0$ such that (4.17) holds, is $\max\{Re(\lambda), \ \lambda \in eig(A)\}$, where $eig(A)$ is the set of the eigenvalues of A. In the case of an uncertain system, there is no analogous concept of eigenvalues. The eigenvalues of $A(w)$, intended as functions of w, do not play a substantial role anymore, since they may all have negative real part bounded away from 0 (i.e. $\max\{Re(\lambda), \ \lambda \in eig(A(w))\} \leq \beta < 0$), and still the time-varying system be unstable. Indeed, the condition becomes necessary only.

Most of us[1] experienced how parameter variations can affect the stability of a system which is stable for any fixed value of the parameters, for example when sitting on a swing. A swing can be destabilized (among the different strategies) by changing the distance of the mass barycenter from the hinge. A slightly different system, similar in spirit, consists in a pendulum which is hinged on a non-inertial frame (see Fig. 7.1), and subject to a vertical acceleration $a(t)$. This system is represented by the model

$$J\ddot{\theta}(t) = -(g - a(t))\sin(\theta(t)) - \hat{\alpha}\dot{\theta}(t)$$

[1] Before knowing the pleasure of studying applied mathematics.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_7

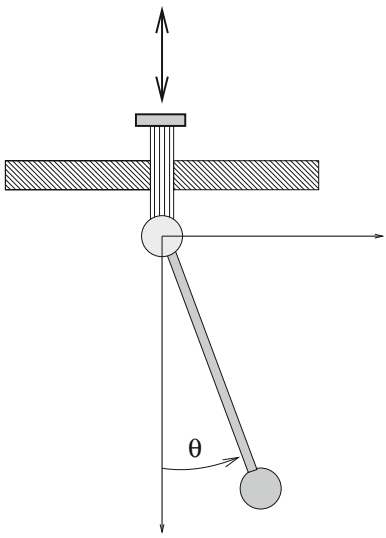


Fig. 7.1 The experimental device

If one assumes small angles, so that $\sin(\theta) \approx \theta$, the following linear system is obtained:

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -\rho(t) & -\alpha \end{bmatrix} x(t) \tag{7.1}$$

where $x = [\theta \ \dot{\theta}]^T$ and $\alpha = \hat{\alpha}/J$. If $\rho(t)$ is bounded as $0 < \rho^- \leq \rho(t) \leq \rho^+$ and $\alpha > 0$, then this system is stable for any frozen value of the parameters. However, if the friction coefficient $\alpha > 0$ is assumed small enough, then the system becomes unstable for a suitable variation of $\rho(t)$. We do not report here a formal proof since it can be found on most books (see, for instance, [Lib03]), and since this issue will be reconsidered later in Section 9.1.

In practice, to assure stability of a system with time-varying parameters (or to assure a proper convergence speed by means of a control action), Lyapunov theory is fundamental: a proper Lyapunov function is needed which has to be common to all the members of the family of systems represented by $A(w)$.

As already pointed out, the most commonly used Lyapunov functions are the quadratic ones. They are the most famous in the control community and they are easy to deal with. However, they are known to be conservative. Conversely, polyhedral functions have stronger properties as far as the conservativity is concerned, but they may require computationally heavy procedures. In this section, both kinds of functions will be considered, with the aim of enlightening their advantages and disadvantages in the problem of stabilizing a system, possibly by assigning a convergence speed.

7.0.1 Control of a flexible mechanical system

To motivate the chapter, we introduce a preliminary example.

Example 7.1. Consider the system depicted in Fig. 7.2, whose dynamics is

$$\ddot{\theta} = -\alpha(\theta - \varphi) + \tau$$
$$\ddot{\varphi} = -\alpha(\varphi - \theta) + \beta \sin(\varphi)$$

Let $\bar{\varphi}$ be a reference angle for φ and let $\bar{\theta}$ be the corresponding equilibrium value for θ, namely such that

$$\alpha(\bar{\varphi} - \bar{\theta}) + \beta \sin(\bar{\varphi}) = 0$$

Define $x_1 = \varphi - \bar{\varphi}$, $x_2 = \theta - \bar{\theta}$, $x_3 = \dot{x}_1 = \dot{\varphi}$ and $x_4 = \dot{x}_2 = \dot{\theta}$. Let $\bar{\tau}$ be the equilibrium value for the control, which satisfies

$$\bar{\tau} + \beta \sin(\bar{\varphi}) = 0$$

and let $u = \tau - \bar{\tau}$. If we write the equations for $\dot{x}_i$ we see that one of them is nonlinear:

$$\ddot{x}_1 = -\alpha(x_1 - x_3) + \beta \left[\sin(x_1 + \bar{\varphi}) - \sin(\bar{\varphi})\right]$$

To eliminate the nonlinearity, let

$$\omega \doteq \frac{\sin(\varphi) - \sin(\bar{\varphi})}{\varphi - \bar{\varphi}}$$

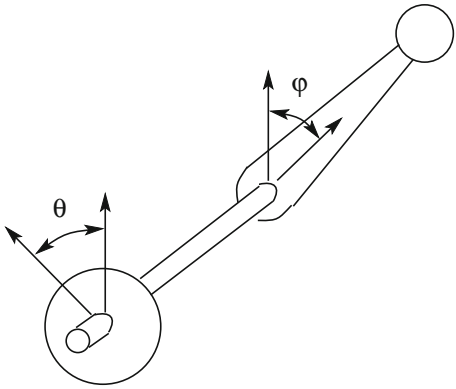


Fig. 7.2 The flexible mechanical system

This function is bounded as $|\omega| \leq 1$. Indeed write[2]

$$\omega = \frac{\sin(\bar{\varphi} + x) - \sin(\bar{\varphi})}{x},$$

Candidate minima and maxima are achieved by computing the derivative and equating it to zero:

$$\frac{d}{dx}\omega = \frac{\cos(x + \bar{\varphi})x - \sin(x + \bar{\varphi}) + \sin(\bar{\varphi})}{x^2} = 0$$

Let x^* be a maximum or a minimum point, then

$$\sin(\bar{\varphi} + x^*) - \sin(\bar{\varphi}) = \cos(\bar{\varphi} + x^*)x^*$$

By replacing this term in ω one can see that the corresponding minimum or maximum ω^* is such that

$$\omega^* = \frac{\cos(\bar{\varphi} + x^*)x^*}{x^*} = \cos(\bar{\varphi} + x^*)$$

hence $-1 \leq \omega^* \leq 1$.

The nonlinear equation becomes $\ddot{x}_1 = -\alpha(x_1 - x_3) + \beta\omega x_1$ and then the system can be absorbed in the following linear differential inclusion

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha + \beta\omega & \alpha & 0 & 0 \\ \alpha & -\alpha & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} u$$

If we assume that the measured variables are the angles $x_1 = \varphi - \bar{\varphi}$, $x_2 = \theta - \bar{\theta}$, we have to consider the output vector $y \in \mathbb{R}^2$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

The resulting model has the form

$$\begin{aligned} \dot{x}(t) &= A(\omega(t))x(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned}$$

Note that parameter $\omega(t)$ is available for control purpose as long as $x_1 = \varphi - \bar{\varphi}$ is measured.

[2] We assume that the function is extended by continuity, for $x = 0$, $\omega = \cos(\bar{\varphi})$.

It is clear that if we succeed in stabilizing this system, for any possible function $|\omega(t)| \leq 1$, then the same control law stabilizes the nonlinear system.

The previous example motivates the gain-scheduling control technique for LPV systems which has been deeply investigated in the control literature [SA90, SA91, SB92, AG95, Sha96a, Hel98, SEG98, LR95].

7.1 Robust and Gain-scheduling control

First, a "vertex" result for polytopic systems is introduced. The systems considered here are of the form

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t) + Ed(t) \tag{7.2}$$

in the continuous-time case, whereas they have the form

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t) \tag{7.3}$$

in the discrete-time case. The uncertain matrices are assumed to be polytopic

$$A(w(t)) = \sum_{i=1}^{s} A_i w_i, \qquad B(w(t)) = \sum_{i=1}^{s} B_i w_i, \tag{7.4}$$

where $w(t) \in \mathcal{W}$,

$$\mathcal{W} = \left\{ w \in \mathbb{R}^s : \; w_i \geq 0, \quad \sum_{i=1}^{s} w_i = 1 \right\} \tag{7.5}$$

and

$$d(t) \in \mathcal{D}$$

where $\mathcal{D} = \mathcal{V}(D)$ is a polytope having vertices grouped in the matrix $D = [d_1 \; d_2 \ldots d_r]$. Denote by

$$w^{(i)} = \left[\begin{array}{ccc} 0 \ldots 0 & \underbrace{1}_{i\text{th position}} & 0 \ldots 0 \end{array} \right]^T$$

the ith vertex of $\mathcal{W}$.

The next theorem states a fundamental extreme point result (basically a special case of Proposition 2.35).

Theorem 7.2. *Let $\mathcal{S}$ be a C-set and $\Psi_{\mathcal{S}}$ its Minkowski functional. Then $\mathcal{S}$ is β-contractive (λ-contractive) for system (7.2) (system (7.3)) under state feedback (possibly under input constraints of the form $u \in \mathcal{U}$) if and only if there exists a Lipschitz control function $\Phi(x)$ such that*[3]

$$\begin{aligned} D^+\Psi_{\mathcal{S}}(x, A_i x + B_i\Phi(x) + Ed_k) &\le -\beta, \quad \textit{(continuous-time case)} \\ \Psi_{\mathcal{S}}(A_i x + B_i\Phi(x) + Ed_k) &\le \lambda, \quad \textit{(discrete-time case)} \end{aligned} \tag{7.6}$$

(and $\Phi(x) \in \mathcal{U}$), for all $x \in \partial\mathcal{S}$, all $i = 1, 2, \ldots, s$ and all $k = 1, 2, \ldots, r$.

Proof. The condition is obviously necessary, because $w(t) \equiv w^{(i)}$ and $d(t) \equiv d_k$ are possible realizations of w and d.

We prove sufficiency in the case $E = 0$ and in the continuous-time case only. Since $\sum_i w_i = 1$ and $\Psi_{\mathcal{S}}$ is convex, then

$$\begin{aligned} & D^+\Psi_{\mathcal{S}}(x, A(w)x + B(w)\Phi(x)) \\ &= \limsup_{h\to 0} \frac{\Psi_{\mathcal{S}}(x + h(\sum[A_i w_i x + B_i w_i \Phi(x)]) - \Psi_{\mathcal{S}}(x)}{h} = \\ &= \limsup_{h\to 0} \frac{\Psi_{\mathcal{S}}(\sum\ w_i[x + h(A_i x + B_i\Phi(x))] - \sum\ w_i\Psi_{\mathcal{S}}(x)}{h} \\ &\le \limsup_{h\to 0} \sum\ w_i \frac{\Psi_{\mathcal{S}}(x + h(A_i x + B_i\Phi(x))) - \Psi_{\mathcal{S}}(x)}{h} \le -\beta \end{aligned}$$

The proof for the discrete-time case is quite similar and it is omitted.

The next important corollary, which is a consequence of Theorem 7.2 and Theorem 4.24, enlightens the relation between contractive sets and Lyapunov functions.

Corollary 7.3. *Assume that $\Psi_{\mathcal{S}}(x)$ is a gauge function, say the Minkowski functional of a C-set $\mathcal{S}$. Then $\Psi_{\mathcal{S}}(x)$ is a global control Lyapunov function (inside $\mathcal{S}$, outside $\mathcal{S}$) if and only if there exists an admissible control $\Phi(x) \in \mathcal{U}$ such that the condition*

$$D^+\Psi_{\mathcal{S}}(x, A_i x + B_i\Phi(x) + Ed_k) \le -\beta\Psi_{\mathcal{S}}(x)$$

or, in the discrete-time case, the condition

$$\Psi_{\mathcal{S}}(A_i x + B_i\Phi(x) + Ed_k) \le \lambda\Psi_{\mathcal{S}}(x))$$

is satisfied for all x (for all $x \in \mathcal{S}$, for all $x \notin \mathcal{S}$) and for all $i = 1, 2, \ldots, s$ and all $k = 1, 2, \ldots, r$.

[3]Note that $\Psi_{\mathcal{S}} = 1$ on the boundary, so the inequalities could be written in the familiar way $D^+\Psi_{\mathcal{S}} \le -\beta\Psi_{\mathcal{S}}$.

We stress that the control function Φ must be common for all k and i. The existence of several control functions, each "working" for some k, is not sufficient, as exemplified next.

Example 7.4. Extreme system stabilization with the same Lyapunov function, but with different controls, is not sufficient: a counterexample.

The system

$$\dot{x}(t) = x(t) + \delta(t)u(t), \quad |\delta| \leq 1 \tag{7.7}$$

is clearly not stabilizable since $\delta(t) = 0$ is a possible realization. However, both the "extremal" systems $\dot{x} = x + u$ and $\dot{x} = x - u$ admit the control Lyapunov function $|x|$ (or x^2), as it is easy to check, although associated with different control actions (e.g., $u = -2x$ and $u = 2x$, respectively).

The corollary could be stated in a quite more general way according to Proposition 2.35.

Proposition 7.5. *Consider a system of the form*

$$\dot{x}(t) = F(x(t), u(t), w(t), d(t))$$

where

$$F(x,u,w,d) = \sum_{i=1}^{s} w_i f_i(x,u) + Ed, \quad \textit{with} \sum_{i=1}^{s} w_i = 1, \quad w_i \geq 0$$

and $d \in \mathcal{D}$ is a polytope with r vertices $d_1, \ldots, d_r$. Then, for a smooth positive definite function $\Psi(x)$, the following two conditions are equivalent:

1. *for every $w \in \mathcal{W}$ and any $d \in \mathcal{D}$*

$$\dot{\Psi}(x, \Phi(x), w, d) = \nabla^T \Psi(x) F(x, \Phi(x), w, d) \leq -\phi(\|x\|)$$

2. *for any $i = 1, \ldots, s$ and any d_k, $k = 1, \ldots, r$,*

$$\nabla^T \Psi(x)[f_i(x, \Phi(x)) + Ed_k)] \leq -\phi(\|x\|)$$

Clearly, if $\phi(x)$ is a κ-function, $\Psi(x)$ is a Lyapunov function for the closed-loop system. The mentioned property has been stated in a general form, although in the case of LPV systems we will consider, without restrictions, Minkowski functions as candidate Lyapunov functions. Note also that the same property holds for a convex (non-differentiable) function Ψ, if the classical derivative $\dot{\Psi}$ is replaced by $D^+\Psi$.

It is worth stressing that, for discrete-time systems, smoothness alone is not enough for this "extreme point property" and the convexity assumption on Ψ is essential.

Proposition 7.6. *In the discrete-time case, the following two conditions are equivalent*

1. for every $w \in \mathcal{W}$ and any $d \in \mathcal{D}$

$$\Psi\left(F(x, \Phi(x), w, d)\right) - \Psi(x) \leq -\phi(\|x\|)$$

2. for any $i = 1, \ldots, s$ and any d_k, $k = 1, \ldots, r$,

$$\Psi\left(f_i(x, \Phi(x)) + Ed_k\right) - \Psi(x) \leq -\phi(\|x\|)$$

provided that Ψ is convex.

So far we considered state feedback. Let us now consider the full-information case $u = \Phi(x, w)$. Here we assume B certain (we will comment on this assumption later).

Theorem 7.7. *Consider and LPV system with matrices as in (7.4) under the assumption that B is certain (i.e., $B_i = B$ for all i). Let $\mathcal{S}$ be a C-set and $\Psi_{\mathcal{S}}$ be its Minkowski functional. Then $\mathcal{S}$ is β-contractive (λ-contractive) under full information control*

$$\Phi(x, w) = \sum_{i=1}^{s} w_i \Phi(x, w^{(i)})$$

(possibly under convex constraints $u \in \mathcal{U}$) if and only if Φ is such that

$$\begin{aligned} D^+\Psi_{\mathcal{S}}(x, A_i x + B\Phi(x, w^{(i)}) + Ed_k) &\leq -\beta \quad \textit{(continuous-time)} \\ \Psi_{\mathcal{S}}(A_i x + B\Phi(x, w^{(i)}) + Ed_k) &\leq \lambda \quad \textit{(discrete-time)} \end{aligned}$$

(and $\Phi(x, w^{(i)}) \in \mathcal{U}$), for all $x \in \partial\mathcal{S}$, for every $i = 1, 2, \ldots, s$ and every $k = 1, 2, \ldots, r$.

Proof. The condition is obviously necessary, because $w(t) \equiv w^{(i)}$ and $d(t) \equiv d_k$ are possible realizations of w and d. Again, sufficiency is proved in the case $E = 0$ and in the continuous-time case only. Since $\sum_i w_i = 1$ and $\Psi_{\mathcal{S}}$ is convex, the following chain of conditions holds true.

$$\begin{aligned} & D^+\Psi_{\mathcal{S}}(x, A(w)x + B(w)\Phi(x, w)) \\ &= \limsup_{h \to 0} \frac{\Psi_{\mathcal{S}}(x + h(\sum [A_i w_i x + B w_i \Phi(x, w_i)]) - \Psi_{\mathcal{S}}(x)}{h} = \\ &= \limsup_{h \to 0} \frac{\Psi_{\mathcal{S}}(\sum\ w_i [x + h(A_i x + B\Phi(x, w_i))] - \sum\ w_i \Psi_{\mathcal{S}}(x)}{h} \\ &\leq \limsup_{h \to 0} \sum\ w_i \frac{\Psi_{\mathcal{S}}(x + h(A_i x + B\Phi(x, w_i))) - \Psi_{\mathcal{S}}(x)}{h} \leq -\beta \end{aligned}$$

The theorem suggests the idea that we can associate a control with each "vertex i". Indeed this is the case, as we can see from the next corollary.

Corollary 7.8. *Let $\Psi(x)$ be a gauge function, say the Minkowski functional of a C-set $\mathcal{S}$, and let $B(w) = B$. Then $\Psi(x)$ is a global control Lyapunov function (inside $\mathcal{S}$, outside $\mathcal{S}$) under full-information control there exist admissible controls $\Phi_i(x)$, $i = 1, 2, \ldots, s$ such that the condition*

$$D^+\Psi(x, A_i x + B\Phi_i(x) + Ed_k) \leq -\beta\Psi(x)$$

or, in the discrete-time case, the condition

$$\Psi(A_i x + B\Phi_i(x) + Ed_k) \leq \lambda\Psi(x)$$

is satisfied with $\beta > 0$, or in the discrete-time case $0 \leq \lambda < 1$, along with $\Phi_i(x) \in \mathcal{U}$, for all x (for all $x \in \mathcal{S}$, for all $x \notin \mathcal{S}$), for every $i = 1, 2, \ldots, s$ and every $k = 1, 2, \ldots, r$.

Note that, once the Φ_i are known, the control is given by $\Phi(x, w) = \sum w_i \Phi_i(x)$. Note also that the assumption of a certain B cannot be dropped. This can immediately be seen from Example 7.4.

We remind now Theorem 2.54 and its original and more general formulation [LSW96]. Consider an LPV system with matrices as in (7.4), equipped by a continuous control $u = \Phi(x, w)$ which is locally Lipschitz in x uniformly with respect to w. In view of the results in [LSW96], an LPV system is Globally Uniformly Asymptotically Stable (GUAS) if and only if there exists a smooth positive definite function which is a global Lyapunov function and satisfies

$$D^+\Psi(x, w, d) = \nabla\Psi(x)^T[A(w)x + B(w)\Phi(x, w) + Ed] \leq -\phi(\|x\|)$$

where $\phi(\|x\|)$ is a positive definite function. If the system is uniformly ultimately bounded within a C-set $\mathcal{X}$, then there exists a smooth positive definite function which is a Lyapunov function outside $\mathcal{X}$. If the system is uniformly locally stable including a C-set $\mathcal{X}$ in its domain of attraction, then there exists a Lyapunov function inside $\mathcal{X}$. Similar results hold in the discrete-time case [JW02].

We admit that the reader could be surprised at this point. We asked her/him to make a major effort to consider functions that can be non-smooth and now we let her/him know that there is no restriction in considering smooth ones for, basically, the most general problem considered in the book. The reason is quite simple. The previous result is fundamental from a theoretical standpoint, but it is non-constructive. We will indeed present constructive methods to generate polyhedral Lyapunov functions which are, by their nature, non-smooth.

7.2 Stabilization of LPV systems via quadratic Lyapunov functions

In this section, stability and stabilizability problems will be considered and the conditions to check stability (stabilizability) by means of a single quadratic Lyapunov function will be reported.

Definition 7.9 (Quadratic stabilizability). The system

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t)$$

is said to be quadratically stabilizable (quadratically stable if $B = 0$) if it admits a quadratic control Lyapunov function (Lyapunov function if $B = 0$). The system is said to be quadratically stabilizable via linear control (linearly quadratically stabilizable) if the controller associated with the control Lyapunov function is of the form $u = Kx$.

7.2.1 Quadratic stability

When $B = 0$, the existence of a common quadratic Lyapunov function, namely of a symmetric solution P to the parametrized LMI

$$A(w)^T P + PA(w) \prec -2\beta P, \quad P \succ 0, \quad \text{for all } w \in \mathcal{W} \tag{7.8}$$

assures the condition $x^T(t)Px(t) \leq x^T(0)Px(0)\, e^{-2\beta t}$. Equivalently, we can write

$$\|x(t)\|_P \leq \|x(0)\|_P\, e^{-\beta t}$$

(we remind that $\|x\|_P = \sqrt{x^T Px}$) and therefore exponential stability is guaranteed.

In the case of polytopic systems the previous condition reduces to a finite number of inequalities. Indeed, to check whether the system converges with speed β, according to Theorem 7.2 and Corollary 7.3 (for $B = 0$ and $E = 0$), one can consider the following LMI problem

Problem 7.10. Find a matrix P such that

$$\begin{aligned} A_i^T P + PA_i &\preceq -2\beta P, \quad i = 1, 2, \ldots, r \\ P &\succeq \epsilon I \end{aligned}$$

One nice property of the above problem (as we have already seen) is that the set of all its solutions $\mathcal{F}(A_i, \epsilon, \beta)$, also known as feasible set, is convex. Indeed, if P_1 and P_2 are feasible solutions, then $P = \alpha P_1 + (1 - \alpha)P_2$ is a feasible solution

for all $0 \leq \alpha \leq 1$. This implies that the search of feasible solutions may rely on very efficient methods [BEGFB04, BV04]. Unfortunately, the existence of a matrix P satisfying condition (7.8) is sufficient but not necessary to assure a speed of convergence β (this trade-off between computational efficiency and conservativity is a recurring argument).

7.2.2 Quadratic stabilizability

Let us now proceed with our investigation by considering the stabilization problem for an LPV. Inspired by the certain case, for which assuming a linear control is not a restriction, a control law of the form $u = Kx$ is sought. The stability condition (7.8) previously stated becomes then (with $\beta = 0$)

$$(A(w) + B(w)K)^T P + P(A(w) + B(w)K) \prec 0.$$

If a polytopic structure in the matrices $A(w)$ and $B(w)$ is assumed, then one gets a finite number of inequalities

$$A_i^T P + PA_i + K^T B_i^T P + PB_i K \prec 0, \quad i = 1, 2, \ldots, s, \quad P \succ 0$$

which are nonlinear in K and P. Interestingly, the problem can be re-parametrized by assuming $Q = P^{-1}$ and $R = KQ$. Indeed, by pre and post multiplying both sides by Q, the following extension of condition (4.21) is obtained:

$$QA_i^T + A_i Q + R^T B_i^T + B_i R \prec 0, \quad i = 1, 2, \ldots, s, \quad Q \succ 0 \tag{7.9}$$

which is again a set of LMIs. If a solution pair (Q, R) to the above problem exists, then

$$K = RP$$

is the desired linear control gain and condition (7.9) characterizes the set of all Q such that $\mathcal{E}(Q^{-1}, 1)$ is a contractive ellipsoid that can be associated with a linear gain K.

If a specific level of contractivity $\beta > 0$ is desired, then the above set of inequalities has simply to be changed into

$$QA_i^T + A_i Q + R^T B_i^T + B_i R + 2\beta Q \prec 0, \quad i = 1, 2, \ldots, r, \quad Q \succ 0 \tag{7.10}$$

Again, the existence of $\beta > 0$ along with R and a positive definite Q is a sufficient condition only for the system stabilizability.

To conclude the present section, some results concerning the stabilizability problem via linear control are reported. The first one concerns general polytopic systems, whereas the second one is an interesting particularization to systems, with no uncertainties affecting the input matrix.

Theorem 7.11. *Condition (7.10) is necessary and sufficient for a polytopic system to be quadratically stabilizable via linear control.*

The above condition is quite strong, but it is worth stressing that there exist quadratically stabilizable systems, henceforth admitting contractive ellipsoids, for which linear stabilizing controllers do not exist. Put in other words: quadratic stabilizability is not equivalent to linear quadratic stabilizability [Pet85].

An exception to the rule just recalled is the case in which matrix B is certain, when the condition reported in the previous theorem is necessary and sufficient for quadratic stabilizability, being the latter always achievable via linear compensators when B is known. This fact, which has already been mentioned in Section 4.4 (see also Exercise 12 in Chapter 2) is reported in the next theorem.

Theorem 7.12 ([Mei74, BPF83]). *The system $\dot{x} = A(w)x + Bu$ with $w \in \mathcal{W}$, where $\mathcal{W}$ is a compact set, with $A(w)$ continuous, is quadratically stabilizable if and only if it is quadratically stabilizable via linear control.*

These results can be extended under some additional assumptions to uncertain B [BCL83] (see Exercise 7)

We now consider the case in which the control can be a function of the parameter w. Let us first analyze the case in which B is known. In this case it is possible to consider the following set of inequalities (see Corollary 7.8)

$$QA_i^T + A_iQ + R_i^TB^T + BR_i + 2\beta Q \prec 0, \quad i = 1, 2, \ldots, s, \quad Q \succ 0 \tag{7.11}$$

If the above set of LMIs admits a solution, then the system

$$\dot{x}(t) = [A(w) + BK(w)]x(t)$$

is asymptotically stable with the gain scheduled linear control law $u = K(w)x$, where

$$K(w) = \sum_{i=1}^{s} K_i w_i$$

where

$$K_i = R_iP = R_iQ^{-1}. \tag{7.12}$$

We stress once again that, if B was also a function of w, the expression (7.11) could not be extended in general since a pathology might occur as in system (7.7). We also have to stress the following fact.

Remark 7.13. If the system $\dot{x}(t) = A(w(t))x(t) + Bu(t)$, $w \in \mathcal{W}$, where $A(w)$ is continuous and $\mathcal{W}$ is compact can be quadratically stabilized by a control of the

form $u = \Phi(w, x)$, then it can be stabilized by a pure state-feedback control of the form $u = \Phi(x)$ and precisely by a gradient-based controller [Mei79]

$$u = -\gamma B^T P x.$$

However, potential advantages in terms of performances can be obtained when the control uses information on w (see Exercise 1).

7.2.3 *Quadratic Lyapunov functions: the discrete-time case*

Given a discrete-time polytopic system of the form

$$x(t+1) = A(w(t))x(t) = \left[\sum_{i=1}^{s} A_i \, w_i(t)\right] x(t) \tag{7.13}$$

the set of LMI conditions to be satisfied to check quadratic stability is the following

$$A_i^T P A_i - P \prec 0, \quad i = 1, 2, \ldots, s, \tag{7.14}$$

which is the natural counterpart of (4.24). To prove this claim note that, for any x, the set of all vectors $y(w) = A(w)x$ is a polytope and, since the norm is a convex function, the expression

$$\sqrt{x^T A(w)^T P A(w) x} = \|A(w)x\|_P,$$

thought as a function of w, reaches the maximum on one of the vertices (see Exercise 17 in Section 3).

Let us now consider a mechanism to determine P in the case of an uncertain system with a polytopic matrix and a linear controller $u = Kx$. From (7.14) we get

$$(A_i + B_i K)^T P (A_i + B_i K) - P \prec 0, \quad i = 1, 2, \ldots, s$$

Similarly to what has been done in the continuous-time case, by pre and post multiplying both sides by $Q = P^{-1}$ and by defining $KQ = R$ one gets

$$(QA_i^T + R^T B_i^T) Q^{-1} (A_i Q + B_i R) - Q \prec 0, \quad i = 1, 2, \ldots, s$$

which, along with $Q \succ 0$ is known to be equivalent to the set of LMIs [BEGFB04]

$$\begin{bmatrix} Q & QA_i^T + R^T B_i^T \\ A_i Q + B_i R & Q \end{bmatrix} \succ 0, \quad i = 1, 2, \ldots, s$$

Again, assuming that the controller is linear is in general a restriction.

7.2.4 Quadratic stability and $\mathcal{H}_\infty$ norm

As we have seen, for linear systems, positive invariance of an ellipsoid is equivalent to quadratic stability. Although the subject is not strictly related with the theme of the book, we point out a fundamental connection between the quadratic stability of an uncertain system with non-parametric uncertainties and the $\mathcal{H}_\infty$ norm of an associated transfer function. Given a stable strictly proper rational transfer function $W(s)$, its $\mathcal{H}_\infty$ norm is the value

$$\|W(s)\|_\infty = \sup_{Re(s)\geq 0} \sqrt{\sigma^+[W^T(s^*)W(s)]} \tag{7.15}$$

where s^* is the complex conjugate of s and $\sigma^+[N] = \max_{\lambda\in \mathrm{eig}(N)} |\lambda|$ is the maximum modulus of the eigenvalues of matrix N, which is related to the induced 2-norm of a matrix as follows:

$$\|M\|_2 \doteq \sup_{x\neq 0} \frac{\|Mx\|_2}{\|x\|_2} = \sqrt{\sigma^+[M^TM]}$$

The following property holds.

Theorem 7.14. *Given the system $\dot{x}(t) = A(\Delta(t))x(t)$ with*

$$A(\Delta) = A_0 + D\Delta E, \quad \|\Delta\|_2 \leq \rho,$$

there exists a positive definite matrix P such that

$$x^T PA(\Delta)x < 0, \quad \textit{for all} \quad \|\Delta\|_2 \leq \rho,$$

if and only if A_0 is asymptotically stable and

$$\|E(sI - A_0)^{-1}D\|_\infty < \frac{1}{\rho}$$

For the stabilization problem, this theorem admits the following extension:

Theorem 7.15. *Consider the system*

$$\begin{aligned}\dot{x}(t) &= [A_0 + D\Delta E]x(t) + [B_0 + D\Delta F]u(t)\\ y(t) &= C_0x(t), \quad ||\Delta(t)|| \leq \rho.\end{aligned}$$

Then the control $u(s) = K(s)y(s)$ is quadratically stabilizing if and only if the closed-loop system

$$\begin{aligned}sx(s) &= A_0x(s) + Dd(s) + B_0u(s)\\ z(s) &= Ex(s) + Fu(s)\\ y(s) &= C_0x(s)\\ u(s) &= K(s)y(s)\end{aligned}$$

is stable and the corresponding d-to-z transfer function

$$\|W_{zd}(s)\|_\infty \leq \frac{1}{\rho}$$

The proof of both theorems can be found in [KPZ90]. In our context these results show an important connection between properties that can be naturally defined in the state space and the frequency domain characterization of a system.

This connection goes much further beyond its noteworthy theoretical interest. It turns out that $\mathcal{H}_\infty$ control theory provides several tools (including LMI and Riccati equations) for efficient control synthesis and therefore robustness can be faced in that framework. For details on this matter, the reader is referred to specialized literature [SPS98, ZDG96].

7.2.5 *Limits of quadratic functions and linear controllers*

It has been recalled that finding a quadratic function along with a linear compensator is a "nice" convex problem. Also, it has been shown how, in the case of a polytopic system, the solution of such a problem is equivalent to a set of LMIs, for which efficient tools are available. The only potential trouble is the number of LMIs involved. To explain this fact, let us consider the case of an interval matrix A for which the uncertainty specification is given in the form

$$A_{ij}^- \leq A_{ij} \leq A_{ij}^+$$

with $A_{ij}^- \leq A_{ij}^+$, for each matrix entry. This means that each A_{ij} is, potentially, an uncertain parameter (unless $A_{ij}^- = A_{ij}^+$). If a polytopic representation of the form (7.4) is adopted, in the worst case all the vertices have to be considered. The total number of vertices is

$$n_v = 2^{n^2}$$

This means that, if we do not limit the number of uncertain entries of A, then the number of vertices grows exponentially and, as a consequence, the "forget it" conclusion is the only possible. Clearly this is a trouble of the problem, which is intrinsically difficult, and it is not due to the quadratic approach. It is interesting to point out that the number of vertices that actually have to be checked can be (strongly) reduced to $n_v = 2^{2n}$, as shown in [ATRC07]. However, the computation remains of exponential complexity.

As a matter of fact, if the number of uncertain entries is limited, LMI tools can face problems of high dimensions with no particular difficulties.

The drawbacks of LMI based techniques are well known:

1. quadratic stability for any fixed value of the parameter (i.e., the existence of a family of matrices $P(w)$ such that $A(w)^T P(w) + P(w)A(w) \prec 0$) does not imply robust time-varying stability;
2. robust stability does not imply quadratic robust stability;
3. robust stabilizability does not imply quadratic robust stabilizability;
4. quadratic stabilizability does not imply quadratic stabilizability via linear controllers;
5. robust stabilizability does not imply robust stabilizability via linear controllers.

The first two claims can be supported by very simple counterexamples. For instance, the system governed by matrix

$$A(w(t)) = \begin{bmatrix} 0 & 1 \\ 1+w & -\epsilon \end{bmatrix}$$

with $\epsilon > 0$, is stable for any fixed $0 \leq w \leq 1$, hence it admits a quadratic Lyapunov function for each fixed w. However, for ϵ small enough, the system can switch between the extrema, $w(t) \in \{0, 1\}$, thus exhibiting the well-known unstable behavior (see, for instance, [Lib03], Part II). Examples which support the second claim are known (see [BM96b, Lib03]). We will give a very simple example soon. The third claim has been proved in [BM99b]. The fourth claim is much more difficult and the reader is referred to the famous Petersen counterexample [Pet85]. Actually, it was later pointed out that the Petersen counterexample is indeed linearly stabilizable (but not via a quadratic Lyapunov function [Sta95b, Sta95a]). The final claim has been proved in [BM99b].

A simple system which is robustly stable, but not quadratically stable, is presented next.

Example 7.16. Consider the following A matrix

$$A(\delta(t)) = \begin{bmatrix} 0 & 1 \\ -1+\delta(t) & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \delta(t) \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad |\delta| \leq \rho \tag{7.16}$$

The system is stable iff (see Exercise 5)

$$\rho < \rho_{ST} = 1, \quad \text{(robust stability radius)}$$

However the time-varying system is quadratically stable iff

$$\rho < \rho_Q = \frac{\sqrt{3}}{2}, \quad \text{(quadratic stability radius)}$$

To prove the last assertion one can fix the candidate Lyapunov matrix

$$P = \begin{bmatrix} x & 1 \\ 1 & y \end{bmatrix}$$

which is positive definite if $x > 0$, $y > 0$, $xy > 1$. The derivative is negative if

$$Q = PA(\delta) + A(\delta)^T P = \begin{bmatrix} -2(1-\delta) & x-1-(1-\delta)y \\ x-1-(1-\delta)y & 2(1-y) \end{bmatrix} \prec 0$$

The matrix Q is negative definite for the values of x, y, and δ such that: $\delta < 1$, $y > 1$, and $-4(1-\delta)(1-y) - (x-1-(1-\delta)y)^2 > 0$. The last condition requires that the intersection of all the δ parametrized regions

$$\{(x,y): \quad -4(1-\delta)(1-y) - (x-1-(1-\delta)y)^2 > 0\}$$

is non-empty (it can be shown that these regions are the internal parts of parabolas in the x-y plane, all placed above the line $y = 1$, which is their envelope). This condition is true if and only if the extremal sets (i.e., those achieved for $\delta = -\rho$ and $\delta = \rho$) have non-empty intersection and therefore the limit condition is achieved by the parameter $0 \leq \rho < 1$ such that the curves represented by the two equations

$$\begin{cases} 4(1+\rho)(y-1) = (x-1-(1+\rho)y)^2, \\ 4(1-\rho)(y-1) = (x-1-(1-\rho)y)^2, \end{cases}$$

are tangent. Denoting by $z = x - 1 - y$ we achieve the equivalent system

$$\begin{cases} 4(1+\rho)(y-1) = (z-\rho y)^2, \\ 4(1-\rho)(y-1) = (z+\rho y)^2. \end{cases} \tag{7.17}$$

so the mentioned tangency conditions are equivalent to the fact that system (7.17) has a double solution (y, z). We show that such a tangency condition occurs for $\rho < 1$. The two equations in (7.17) imply

$$(z-\rho y)^2(1-\rho) = (z+\rho y)^2(1+\rho)$$

that yields

$$z^2 + 2zy + \rho^2 y^2 = 0$$

Since $y > 1$, it is possible to divide by y^2 so that

$$\left(\frac{z}{y}\right)^2 + 2\left(\frac{z}{y}\right) + \rho^2 = 0$$

which means that the any solutions (y, z) of system (7.17) must satisfy condition

$$\frac{z}{y} = -1 \pm \sqrt{1-\rho^2}$$

Let us derive z from the previous equation and replace it in any of equations (7.17), for instance the second, to derive

$$\left(\left(-1 \pm \sqrt{1-\rho^2}\right) y + \rho y\right)^2 - 4(1-\rho)y + 4(1-\rho) = 0$$

This equation admits a real root iff

$$(1-\rho) - \left((\rho-1) \pm \sqrt{1-\rho^2}\right)^2 \geq 0$$

The inequality associated with "$-$" is never satisfied, thus we consider only the "$+$" version. Since we are seeking the limit condition, we consider the equality so that we need to solve

$$(1-\rho) - \left((\rho-1) + \sqrt{1-\rho^2}\right)^2 = 0$$

which has solutions $\rho = \pm\sqrt{3}/2$ and $\rho = 1$. The values $\rho = -\sqrt{3}/2$ and $\rho = 1$ must be discharged, so we can conclude that the limit value for quadratic stability is

$$\rho_Q = \frac{\sqrt{3}}{2}.$$

As we have seen, by means of a (not so immediate) chain of algebraic manipulations, the quadratic limit could be found. One would have easier life by using the result in Theorem 7.14. Indeed, even in this case, the computation of the $\mathcal{H}_\infty$ norm of the transfer function $F(s) = D(sI - A_0)^{-1}E$ where

$$A_0 = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

is rather straightforward and results in $\|D(sI - A_0)^{-1}E\|_\infty = \frac{2}{\sqrt{3}}$.

As far as stabilizability is concerned, the next counterexample shows that the conservativity of the methods based on quadratic Lyapunov functions can be arbitrarily high. To this aim, let us consider again the system

$$\dot{x}(t) = A(p(t))x(t) + B(p(t))u(t)$$

where

$$p \in \rho\mathcal{P}$$

and $\rho \geq 0$ is a measure of the uncertainty. If we assume a linear dependence, we get

$$A(p(t)) = A_0 + pA, \quad B(p(t)) = B_0 + pB, \qquad p \in \rho\mathcal{P}$$

with (A_0, B_0) stabilizable and $p \in \mathcal{P}$, a C-set. Then we can define the following two stabilizability margins

$$\rho_{ST} = \sup\{\rho : \text{(S) is stabilizable}\};$$
$$\rho_Q = \sup\{\rho : \text{(S) is quadratically stabilizable}\}.$$

The next counterexample shows that there are systems for which

$$\frac{\rho_{ST}}{\rho_Q} = \infty$$

Example 7.17 (Stabilizability and quadratic stabilizability can be far away). In [BM99b] it has been shown that, given $\mathcal{P} = [-1, 1]$, for the dynamic system:

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} p(t) \\ 1 \end{bmatrix}, \qquad p \in \rho\mathcal{P} = [-\rho, \rho],$$

the two parameters previously introduced are

$$\rho_{ST} = \infty, \qquad \rho_Q = 1.$$

Thus this system is stabilizable for any arbitrary ρ, but no quadratic Lyapunov function exists for $\rho \geq 1$. Also, it is possible to show that, for ρ large enough, this system is not stabilizable by means of any linear static state feedback control law of the form

$$u = k_1 x_1 + k_2 x_2$$

(in which k_1 and k_2 do not depend on w). It is worth saying that there are examples of stabilizable systems which cannot be stabilized via linear (even dynamic) compensators. The interested reader is again referred to [BM99b].

The example evidences once more that seeking quadratic Lyapunov functions and/or linear compensators is in general a conservative choice. Nevertheless, quadratic stabilizability and linear quadratic stabilizability are notions frequently adopted for convenience.

Actually, there are special cases in which the quadratic functions are not conservative. One of such cases is that of systems with uncertainties satisfying the so-called matching conditions (we already found in Subsection 2.4.4), as in the next theorem.

Theorem 7.18. *Consider a system of the form*

$$\dot{x}(t) = A_0 x(t) + B_0 u(t) + B_0 \Delta(x(t), w(t))$$

where $\Delta(x(t), w(t))$ is a Lipschitz function bounded as

$$\|\Delta(x(t), w(t))\| \leq K\|x(t)\|$$

and assume that (A_0, B_0) is stabilizable. Then the system is quadratically stabilizable via linear control.

Proof. See [BCL83].

There are several generalizations of the previous theorem. Perhaps one of the most important is the case of system satisfying the so-called generalized matching conditions [Bar85].

7.2.6 Notes about quadratic stabilizability

As we have mentioned, the quadratic stabilizability approach has an established history which started since the late 70s [Gut79, Lei79]. Subsequently, the connections with Riccati equation [PH86] and with $\mathcal{H}_\infty$ theory [KPZ90] were established. The role of convex optimization in the synthesis of quadratic Lyapunov functions for uncertain systems was pointed out in [GPB91] and it has become very popular (see the book [BEGFB04] and its first version).

Several other studies originated from the quadratic theory. To reduce the level of conservatism, some authors [DCAF94, FAG96] have considered parameter-dependent quadratic Lyapunov functions which obviously offer more degrees of freedom than quadratic function. If the parametrization is simple enough, it is possible to maintain the advantages of an LMI approach [DB01, dOBG99]. A particular interesting case is that of affine parametrization, having the following form

$$V(x) = x^T \left(\sum_i^s P_i w_i(t) \right) x \tag{7.18}$$

where the parameters w entering the function definition are exactly those appearing in the system dynamics (7.13). We refer the reader to the cited works for more details, stressing once again that even by using the above-mentioned functions it is just possible to get sufficient conditions for robust stability. Interesting extensions of quadratic functions are the composite quadratic functions, introduced in [HL03], that can be successfully applied to robust control problems [Hu07].

7.3 Polyhedral Lyapunov functions

In this section, we show that polyhedral functions do not have the same theoretical limitations of quadratic functions.

7.3.1 Polyhedral stabilizability

From Theorem 4.33 and Theorem 7.2, the following result can be derived. We remind that M is a Metzler matrix if $M_{ij} \geq 0$ for $i \neq j$.

Proposition 7.19. *Consider the continuous-time linear LPV system $\dot{x}(t) = A(w(t))x(t)$, with the polytopic structure (7.4), and let $\mathcal{S}$ be a polyhedral C-set of the form (4.25), with $F \in \mathbb{R}^{q\times n}$, or of the form (4.26) with $X \in \mathbb{R}^{n\times r}$. Then, the next statements are equivalent*

i) $\mathcal{S} = \mathcal{P}(F) = \mathcal{V}(X)$ is β-contractive.
ii) There exist s Metzler matrices $H^{(i)}$ such that, for any $i = 1 \ldots, s$,

$$H^{(i)}F = FA_i, \quad H^{(i)}\,\bar{1} \leq -\beta\bar{1}$$

iii) There exist s Metzler matrices $H^{(i)}$ such that, for any $i = 1 \ldots, s$,

$$A_iX = XH^{(i)}, \quad \bar{1}^T H^{(i)} \leq -\beta\bar{1}^T$$

The next proposition generalizes Theorem 4.37 (see also [Bla00] for further details).

Proposition 7.20. *Consider the polytopic system (7.2) with the constraint $u \in \mathcal{U}$, with $\mathcal{U}$ a convex set. There exists a state-feedback control law $u = \Phi(x)$ satisfying the constraints and which makes the polyhedral C-set $\mathcal{S} = \mathcal{V}(X)$ β-contractive for the closed-loop system if and only if there exist s Metzler -matrices $H^{(i)}$ and a single matrix $U \in \mathbb{R}^{m\times r}$ such that, for any $i = 1, \ldots, s$,*

$$A_iX + B_iU = XH^{(i)}, \tag{7.19}$$

$$\bar{1}^T H^{(i)} \leq -\beta\bar{1}^T \tag{7.20}$$

$$u_k \in \mathcal{U} \tag{7.21}$$

where u_k is the kth column of U.

Proof. Necessity follows immediately from Theorem 4.37, since the set is β-contractive for the polytopic system only if it is such for the extreme systems $\dot{x} = A_ix + B_iu$, therefore the conditions must hold.

Conversely if the conditions hold, then there exists a linear variable structure control $u = \Phi(x)$ of the form (4.39) which can be constructed from U and X as shown in Subsection 4.5.1. Such a control does not depend on w and thus satisfies the conditions of Theorem 7.2 and Corollary 7.3. Therefore, sufficiency follows.

Example 7.21. Consider the system presented in Subsection 7.0.1 with $\alpha = 1$ and $|w| \leq 0.2$. Using the EAS with $\tau = 0.4$, a contractive factor $\lambda = 0.98$, and a tolerance $\epsilon = 0.005$, by means of the recursive procedure presented in the previous chapter, a polyhedral Lyapunov function was computed. The unit ball of such a function has 106 delimiting planes.

In the case of a polytopic system, with known B, we have the following.

Corollary 7.22. *Consider the polytopic system (7.2) with the constraint $u \in \mathcal{U}$, with $\mathcal{U}$ a convex set. There exists a gain scheduled state-feedback control law $u = \Phi(x, w)$ satisfying the constraints and which makes the polyhedral C-set $\mathcal{S} = \mathcal{V}(X)$ β-contractive for the closed-loop system if and only if there exist s Metzler matrices $H^{(i)}$ and s matrices $U^{(i)} \in \mathbb{R}^{m\times r}$ such that*

$$A_i X + BU^{(i)} = XH^{(i)}, \tag{7.22}$$

$$\bar{1}^T H^{(i)} \leq -\beta \bar{1}^T \tag{7.23}$$

$$u_k^{(i)} \in \mathcal{U}, \quad k = 1, \ldots, r \tag{7.24}$$

for any $i = 1, \ldots, s$, where $u_k^{(i)}$ is the kth column of $U^{(i)}$.

Proof. Necessity follows, as in the previous proposition, by the fact that if $u = \Phi(x, w)$ is a control which assures contractivity for the polytopic system, then it assures contractivity for fixed $w = w_i$, say the conditions must be satisfied.

Sufficiency: if the conditions hold, for each $w = w_i$, in view of Theorem 4.37 there exists a control $\Phi_i(x)$ (which can be, for instance, the piecewise linear controller (4.39)) associated with a contractive set for the vertex system $\dot{x} = A_i x + B\Phi_i(x)$. In view of Theorem 7.7 and Theorem 7.8, the control

$$u = \Phi(x, w) = \sum_{i=1}^{s} w_i \, \Phi_i(x)$$

is a suitable controller.

We remark that a consequence of Proposition 7.20 and Corollary 7.22 is that, when X is given (therefore the function is fixed) checking contractivity is a linear programming problem. It is very easy to see that checking contractivity via linear gains of the form

$$u = Kx(t),$$

or, in the case of a certain B, of the form

$$u(t) = K(w(t))x(t) = \left[\sum_{i=1}^{s} w_i(t) K_i\right] x(t)$$

is a linear programming problem.

It should be noticed that there exist examples of systems for which a certain function $\Psi(x)$ is a Lyapunov function that can be associated with a gain scheduling control only (say, a state feedback controller alone would not do the job).

Example 7.23. Consider the system $\dot{x} = A(w)x + Bu$ with

$$A = \begin{bmatrix} -1 & 0 \\ w & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

with $|w| \leq 2$ and the candidate Lyapunov function $\|x\|_1$. This function becomes a Lyapunov function with the control $u = -x_2 - wx_1$. However, for any control $u = \Phi(x_1, x_2)$ applied to the vertex $[1, 0]^T$ of the unit ball of $\|x\|_1$, if $w = 2sgn(u)$, the derivative vector does not satisfy Nagumo's condition. This pathological behavior does not show up if we take Ψ differentiable.

The following property establishes a connection between the gain scheduling and the robust stabilization problems.

Theorem 7.24 ([Bla00]). *The system $\dot{x} = A(w)x + B(w)u$, where the set $\{[A(w), B(w)], \quad w \in W\}$ is a convex and compact set, is robustly stabilizable if (and obviously only if) it is gain-scheduling stabilizable.*

The previous theorem is in some way valid for a more general class of nonlinear systems: that of convex processes, namely systems of the form

$$\dot{x}(t) = A(x, w) + B(x, w)u(t)$$

where the set $\{[A(x, w), B(x, w)], \quad w \in W\}$ is compact and convex for any x. A previous version of the theorem in the case of certain $B(x)$ was introduced in [Mei79].

As a consequence, it is not necessary to seek for controllers of the form $\Phi(x, w)$ for an LPV system since, if the system can be stabilized, then the same goal can be achieved by a controller of the form $u = \Phi(x)$ which does not require on-line measurement of the parameter. Furthermore, we saw that the conditions concerning the gain scheduling stabilization are limited by the fact that uncertainties in the matrix B cannot be admitted. In the continuous-time case, this is not a problem.

Let us now consider the discrete-time case, namely systems of the form (7.3). From Theorem 4.43 and Theorem 7.2, the following can be derived.

Proposition 7.25. *Consider the polytopic LPV system $x(t+1) = A(w(t))x(t)$, with s vertices, and let $\mathcal{S}$ be a polyhedral C-set of the form (4.25), with $F \in \mathbb{R}^{q \times n}$, or of the form (4.26) with $X \in \mathbb{R}^{n \times r}$. Then the next two statements are equivalent.*

i) $\mathcal{S} = \mathcal{P}(F) = \mathcal{V}(X)$ *is λ-contractive.*

ii) There exist s non-negative matrices $P^{(i)} \geq 0$ such that

$$P^{(i)}F = FA_i, \quad P^{(i)}\,\bar{1} \leq \lambda\bar{1}, \quad i = 1, \ldots, s.$$

iii) There exist s non-negative matrices $P^{(i)} \geq 0$ such that

$$A_iX = XP^{(i)}, \quad \bar{1}^T\,P^{(i)} \leq \lambda\bar{1}^T, \quad i = 1, \ldots, s.$$

The next proposition, reported without proof (similar to that of Theorem 7.19 and left as an exercise), generalizes Theorem 4.44.

Proposition 7.26. *Consider the polytopic system (7.3) with the constraint $u \in \mathcal{U}$, with $\mathcal{U}$ a convex set. There exists a state-feedback control law $u = \Phi(x)$ satisfying the constraints and which makes the polyhedral C-set $\mathcal{S} = \mathcal{V}(X)$ λ-contractive for the closed-loop system if and only if there exist s non-negative matrices $P^{(i)}$ and a single matrix $U \in \mathbb{R}^{m \times r}$ such that, for any $i = 1, \ldots, s$,*

$$A_i X + B_i U = X P^{(i)} \tag{7.25}$$

$$\bar{1}^T P^{(i)} \leq \lambda \bar{1}^T \tag{7.26}$$

$$u_k \in \mathcal{U}, \quad k = 1, \ldots, r, \tag{7.27}$$

where u_k is the kth column of U.

Parallel to the continuous-time case, when B is known, the following corollary holds (the proof is almost identical to the continuous-time case one):

Corollary 7.27. *Consider the polytopic system (7.3) with the constraint $u \in \mathcal{U}$, with $\mathcal{U}$ a convex set. There exists a gain scheduled state-feedback control law $u = \Phi(x, w)$ satisfying the constraints making the polyhedral C-set $\mathcal{S} = \mathcal{V}(X)$ λ-contractive for the closed-loop system if and only if there exist s non-negative matrices $P^{(i)}$ and s matrices $U^{(i)} \in \mathbb{R}^{m \times r}$ such that, for any $i = 1, \ldots, s$,*

$$A_i X + B U^{(i)} = X P^{(i)} \tag{7.28}$$

$$\bar{1}^T P^{(i)} \leq \lambda \bar{1}^T \tag{7.29}$$

$$u_k^{(i)} \in \mathcal{U}, \quad k = 1, \ldots, r, \tag{7.30}$$

where $u_k^{(i)}$ is the kth column of $U^{(i)}$.

Remark 7.28. There are conditions similar to those proposed for systems with additive uncertainties in a polytope, $d(t) \in \mathcal{D} = \mathcal{V}(D)$. Basically the idea is that matrices $P^{(ij)}$ (resp. $H^{(ij)}$ in the continuous-time case) have to be found, for all (A_i, B_i) and all $d_j \in vert\{\mathcal{D}\}$, such that

$$A_i X + B_i U + E d_j \bar{1}^T = X P^{(ij)} \ \ (= X H^{(ij)})$$

where, for instance, in discrete-time, $P^{(ij)} \geq 0$ and $\bar{1}^T P^{(ij)} \leq \lambda \bar{1}^T$. The reader is referred to [Sav07] for details.

The relation between gain scheduling and robust control is different from that we have seen for the continuous-time case. Indeed in general the two concepts of gain-scheduling stabilizability and robust stabilizability are not equivalent, as shown in the next example.

Example 7.29. Consider the system

$$x(t+1) = w(t)x(t) + u(t)$$

with $|w| \leq 2$. This system is stabilized by the control $u = -wx$. However no controller which ignores the current value of $w(t)$ is stabilizing. Indeed, for any $x \neq 0$ and for each u (no matter which oracle provides it) which does not depend on w, either $|2x+u| > |x|$ or $|-2x+u| > |x|$. Since both $w = 2$ and $w = -2$ are possible, $|x(t)|$ may increase at each time. Then the system cannot be stabilized for arbitrary $w(t)$. Therefore, in the context of discrete-time systems, seeking controllers of the form $u = \Phi(x, w)$ or $u = \Phi(x)$ for LPV systems are different problems.

7.3.2 *Universality of polyhedral Lyapunov functions (and their drawbacks)*

An important property of polyhedral Lyapunov functions relies on the fact that their existence is a necessary and sufficient condition for the system stabilizability. We name this property universality. Precisely the following theorem holds.

Theorem 7.30. *Consider the system*

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t)$$

with $A(w)$ and $B(w)$ continuous functions of $w \in \mathcal{W}$, with $\mathcal{W}$ a compact set. Then the following statements are equivalent.

i) There exists a control $u = \Phi(x)$, locally Lipschitz in x uniformly with respect to w, such that the closed-loop system is GUAS.

ii) There exists a polyhedral control Lyapunov function $\Psi(x)$ associated with a piecewise linear controller $u = \Phi(x)$ of the form (4.39), which assures the condition

$$D^+\Psi(x) \leq -\beta\Psi(x)$$

for some positive β.

Furthermore, if $A(w)$ and $B(w)$ have a polytopic structure, then the previous statements are equivalent to the next one.

iii) There exist matrices X, $H^{(i)}$ and U such that (7.19) and (7.20) are satisfied.

Proof. Clearly ii) implies i). The proof that i) implies ii) is in [Bla95]. We have already seen that ii) is equivalent to condition iii) for polytopic systems.

The previous theorem is essentially a converse Lyapunov theorem for linear uncertain systems. Note that no assumptions have been made on $\mathcal{W}$, beside its compactness to claim the equivalence between i) and ii) which applies to the important case in which $\mathcal{W}$ is a discrete set, precisely to switching systems [Bla95, SG05].

Remark 7.31. If we strengthen our assumptions by requiring that the set of matrices $\{[A(w), B(w)], \quad w \in \mathcal{W}\}$ is convex, we have, in view of Theorem 7.24, that the statements i) and ii) (and iii) for polytopic systems) are equivalent to the following one: there exists a control $u = \Phi(x, w)$, continuous and locally Lipschitz with respect to x, uniformly with respect to w which assures GUAS.

The theorem admits the following corollaries. The first concerns the constrained control case.

Corollary 7.32. *Under the same assumptions of Theorem 7.30, assume $u \in \mathcal{U}$, with $\mathcal{U}$ a C-set. Let $\mathcal{X}$ be a C-set in the state-space. The following statements are equivalent:*

i) there exists a locally Lipschitz control $u = \Phi(x)$ that satisfies the constraints and a control Lyapunov function inside $\mathcal{X}$;

ii) there exists a polyhedral control Lyapunov function $\Psi(x)$ (inside $\mathcal{X}$) such that $\mathcal{X} \subseteq \mathcal{N}[\Psi, 1]$ and the associated piecewise linear controller $u = \Phi(x)$ of the form (4.39) assures that $D^+\Psi(x) \leq -\beta\Psi(x)$ for some positive β and that $u = \Phi(x) \in \mathcal{U}$ for all $x \in \mathcal{N}[\Psi, 1]$.

Again, for polytopic systems, the following statement is equivalent to the previous ones.

iii) There exist matrices X, $H^{(i)}$, and U which satisfy conditions (7.19)–(7.21) where the columns of U $u_k \in \mathcal{U}$.

In the case of systems with additive uncertainties, the next corollary holds.

Corollary 7.33. *Consider the system*

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

where $A(w)$ and $B(w)$ are continuous functions of $w \in \mathcal{W}$ and $d \in \mathcal{D}$, with $\mathcal{W}$ and $\mathcal{D}$ compact sets. Let $\mathcal{X}$ be a C-set in the state-space. The following statements are equivalent.

i) There exists a locally Lipschitz control $u = \Phi(x)$ and a control Lyapunov function outside $\mathcal{X}$.

ii) There exists a polyhedral control Lyapunov function $\Psi(x)$ (outside $\mathcal{X}$) such that $\mathcal{N}[\Psi, 1] \subseteq \mathcal{X}$ and which is associated with a piecewise linear controller $u = \Phi(x)$ of the form (4.39) and such that condition

$$D^+\Psi(x) \leq -\beta\Psi(x)$$

holds for some positive β and for $x \notin \mathcal{N}[\Psi, 1]$.

The proof can be immediately derived from [Bla95].

Remark 7.34. Note that no assumptions have been made on B which can be zero. Therefore the previous corollary tells us that if the system $\dot{x}(t) = A(w(t))x(t)+Ed(t)$ is ultimately bounded with a proper Lyapunov function outside a C-set $\mathcal{X}$, then there exists a polyhedral Lyapunov function outside $\mathcal{X}$.

The previous theorem and corollaries admit a discrete-time version which is readily written. Here a distinction between controllers of the form $u = \Phi(x, w)$ and $u = \Phi(x)$, even in the case of convex process, has to be made, which is not a deal in the continuous-time case (see Remark 7.31).

Theorem 7.35. *Consider the system*

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t)$$

with $A(w)$ and $B(w)$ continuous functions of $w \in \mathcal{W}$, with $\mathcal{W}$ a compact set. Then the following statements are equivalent.

i) There exists a locally Lipschitz control $u = \Phi(x)$ such that the system is GUAS.
ii) There exists a polyhedral control Lyapunov function $\Psi(x)$ associated with a piecewise linear controller $u = \Phi(x)$ of the form (4.39) such that

$$\Psi(A(w)x + B(w)\Phi(x)) \leq \lambda\Psi(x)$$

for some positive $\lambda < 1$.
iii) Conditions (7.25) and (7.26) are satisfied by proper matrices X, $P^{(i)}$ and U.

Proof. It is obvious that iii) $\Rightarrow$ ii) $\Rightarrow$ i). We provide a simplified proof of i) $\Rightarrow$ ii), assuming that the system has a polytopic structure.

Assume that the system can be stabilized by a certain control $\Phi(x)$. Let $\mathcal{X}_0$ be an arbitrary polytopic C-set. By assumption, there exists T such that for, $t \geq T$, all the solutions originating in $\mathcal{X}_0$ are included in $\mathcal{X}_0/2$. Take a positive $\lambda < 1$ close enough to one such that the solution $z(t)$ of the modified system

$$z(t+1) = \frac{A(w(t))}{\lambda}z(t) + \frac{B(w(t))}{\lambda}\Phi(z(t))$$

is included in $\mathcal{X}_0$ at time $t = T$ (note that this implies that the solution $z(t)$ will be in such set for all $t > T$). Consider all the possible trajectories $z(t)$ of the modified system having initial condition on the vertices of $\mathcal{X}_0$ and corresponding to all possible sequences with values $w(t) \in vert\{\mathcal{W}\}$ (then $A(w(t)) = A_i$ for all t). The (enormous) set of all the points forming these trajectory is finite. Let $\mathcal{S}$ be their convex hull and Z the matrix including its vertices (so that $\mathcal{S} = \mathcal{V}(X)$). By construction, for each vertex $x^{(j)}$, we have that[4]

[4] x^+ means "the state at the next step."

$$x^+ = \frac{A_i}{\lambda}x^{(j)} + \frac{B_i}{\lambda}\Phi(x^{(j)}) \in \mathcal{S}$$

namely

$$A_i x^{(j)} + B_i \Phi(x^{(j)}) \in \lambda\mathcal{S}$$

This means that x^+ can be expressed as a convex combination of the vertices of $\mathcal{S}$

$$A_k x^{(j)} + B_k \Phi(x^{(j)}) = \sum_i x^{(i)} P_{ij}^{(k)}$$

with

$$P_{ij}^{(k)} \geq 0, \quad \sum_i P_{ij}^{(k)} \leq \lambda$$

Repeating the same reasoning for all the vertices $x^{(t)}$ and defining the matrix $P^{(k)} \doteq [P_{ik}^{(k)}]$ and

$$U = [\Phi(x^{(1)})\ \Phi(x^{(2)}) \dots \Phi(x^{(s)})]$$

one can see that (7.25) and (7.26) are satisfied, say the system can be robustly stabilized by a control of the form (4.39).

The previous theorem admits the following corollaries whose proofs are omitted for brevity, but can be easily inferred. The first concerns the constrained control case.

Corollary 7.36. *Under the same assumptions of the previous theorem, assume that $u \in \mathcal{U}$, a C-set. Let $\mathcal{X}$ be a C-set in the state space. The following statements are equivalent.*

i) There exists a locally Lipschitz control $u = \Phi(x)$ which satisfies the constraints and a control Lyapunov function inside $\mathcal{X}$.
ii) There exists a polyhedral control Lyapunov function $\Psi(x)$ (inside $\mathcal{X}$) such that $\mathcal{X} \subseteq \mathcal{N}[\Psi, 1]$ and the associated piecewise linear controller $u = \Phi(x)$ of the form (4.39) is such that $\Phi(x) \in \mathcal{U}$ and $\Psi(A(w)x + B(w)\Phi(x)) \leq \lambda\Psi(x)$ for some positive $\lambda < 1$, for all $x \in \mathcal{N}[\Psi, 1]$.
iii) Conditions (7.25) (7.26) are satisfied and the columns u_k of U are such that $u_k \in \mathcal{U}$.

In the case of systems with additive uncertainties, we have the next corollary.

Corollary 7.37. *Consider the system*

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

with $A(w)$ and $B(w)$ continuous functions of $w \in \mathcal{W}$, and $d \in \mathcal{D}$, where $\mathcal{W}$ and $\mathcal{D}$ are compact sets. Let $\mathcal{X}$ be a C-set in the state-space. The following facts are equivalent:

i) There exists a continuous control $u = \Phi(x)$ and a control Lyapunov function outside $\mathcal{X}$.

ii) There exists a polyhedral control Lyapunov function $\Psi(x)$ (outside $\mathcal{X}$) such that $\mathcal{N}[\Psi, 1] \subseteq \mathcal{X}$ which is associated with a piecewise affine controller $u = \Phi(x)$ such that condition

$$\Psi(A(w)x + B(w)\Phi(x) + Ed) \leq \lambda\Psi(x)$$

holds for some positive $\lambda < 1$ and for $x \notin \mathcal{N}[\Psi, 1]$.

Remark 7.38. The piecewise-affine control mentioned in the corollary can actually be the piecewise linear one (4.39) if we assume 0 symmetric sets $\mathcal{X}$ and $\mathcal{D}$. If the symmetry fails, we might be forced to use a "drift" term [Sav07]. For instance, consider the system

$$x(t+1) = 2x(t) + u(t) + d(t)$$

with $-2.9 \leq d(t) \leq 0.9$. The state $x(t)$ can be driven to the set $\mathcal{X} = [-2, 2]$ (for instance in one step by the control $u = -2x + 1$). However, no piecewise linear control, even non-symmetrical, can keep the state inside.

Similar converse Lyapunov results for discrete-time systems, with gain scheduling control, have been proposed in [BM03]. Note that in all the mentioned properties no rank assumptions on matrix B have been made and thus they are valid if there is no control action, i.e. $B = 0$. When $B = 0$ results coming from the Soviet Union literature about robust stability (often referred to as absolute stability) [MP86a, MP86b, MP86c, Bar88a, Bar88c, Bar88c] are recovered. Independent results were previously known in the western literature where polyhedral functions had already been introduced, although not so deeply investigated [BT80, MNV84, OIGH93] (see also [BM94] for interesting connections between eastern and western literature). The next result summarizes several established properties.

Proposition 7.39. *Consider the continuous-time system*

$$\dot{x}(t) = A(w(t))x(t)$$

or the discrete-time system

$$x(t+1) = A(w(t))x(t)$$

with $A(w)$ continuous and $w \in \mathcal{W}$ a compact set. Then the following statements are equivalent.

i) The system is asymptotically stable.
ii) The system is exponentially stable.
iii) The system admits a polyhedral Lyapunov function.

If the case of polytopic systems is considered, the next statements are equivalent to the previous ones.

iv) There exist matrices which satisfy the conditions of Proposition 7.19 (respectively Proposition 7.25).

The conclusions that can be drawn from the previous theorem are that polyhedral Lyapunov functions are an appropriate class as long as non-conservative stabilizability conditions are desired. The comparison with quadratic functions shows that, apparently, the "ellipsoidal shape" is not sufficiently general to confine the dynamics of an uncertain system. There are clearly special cases for which this is not the case, as it was pointed out in the previous sections.

The drawbacks are evident at this point. The complexity of the representation of a polyhedral function depends on the number of delimiting planes or on the number of vertices of its unit ball. Therefore the claimed existence of polyhedral functions for stabilizable system can be frustrated in the attempt of computing them numerically.

There is a further drawback that has to be considered. This is due to the fact that the provided necessary and sufficient conditions for a polyhedral function to be a control Lyapunov function do not characterize the control in an explicit form. Indeed, the control law which has been repeatedly invoked in the "polyhedral" theorems is the piecewise-linear control (4.39), whose complexity can be greater than that of the generating polyhedral function (which is already complex enough in most examples!).

In the discrete-time case, this problem can be avoided by considering a different type of feedback, precisely the on-line-optimization-based control presented next.

Assume a discrete-time control Lyapunov function Ψ is given. Clearly for such function, for any value of x we have that

$$\Psi\left(A(w)x + B(w)u\right) \leq \lambda\Psi\left(x\right)$$

no matter which is $w \in \mathcal{W}$, for some appropriate control $u = \Phi(x) \in \mathcal{U}$ (possibly with $\mathcal{U} = \mathbb{R}^m$). If Ψ is polyhedral and assigned as the Minkowski function of a polyhedral C-set $\mathcal{P}(F) = \{x : \ Fx \leq \bar{1}\}$, this condition can be equivalently stated by saying that u must be taken in such a way that

$$u \in \Omega(x) = \{v : F[A(w)x + B(w)v] \leq \lambda\Psi(x)\bar{1}, \ \ \forall w \in \mathcal{W}, \ \ v \in \mathcal{U}\}\,.$$

In the case of a polytopic system the set $\Omega(x)$, (the regulation map) becomes

$$\Omega(x) = \{v : F(A_i x + B_i v) \leq \lambda\Psi(x)\bar{1}, \ \ i = 1, 2, \ldots, r, \ \ v \in \mathcal{U}\}$$

which is a polyhedron for each x. As mentioned in Subsection 4.5.3, this expression is convenient for single input systems since the set $\Omega(x)$ is an interval for each x. Needless to say, in the case of the gain-scheduling control problem we just have to move the dependence on w in the arguments of the regulation map Ω

$$u \in \Omega(x,w) = \{v : F(A(w)x + B(w)v) \leq \lambda\Psi(x)\bar{1}, \quad v \in \mathcal{U}\}$$

7.3.3 *Smoothed Lyapunov functions*

The previous discrete-time control can be applied to continuous-time systems by means of the Euler Auxiliary System (EAS)

$$x(t+1) = [I + \tau A(w(t))]x(t) + \tau B(w(t))u(t)$$

Indeed, if the EAS has been used to determine the polyhedral control Lyapunov function $\Psi(x) = \max Fx = \max_i F_i x$, then, by construction, the associated regulation map Ω_{EAS} is such that for all x there exists $u(x) \in \mathcal{U}$ ensuring

$$\tau F_k[A_i x + B_i u(x)] \leq \lambda\Psi(x) - F_k x \leq \lambda\Psi(x) - \max_k F_k x = (\lambda - 1)\Psi(x)$$

for all k and i. This, in turn, means that

$$F_k[A_i x + B_i u(x)] \leq -\frac{1-\lambda}{\tau}\Psi(x) = -\beta\Psi(x)$$

for all i and for all k, and in particular for all k in the maximizer set $k \in I(x) = \{h : F_h x = \Psi(x)\}$. Then, for all i

$$D^+\Psi(x) = \max_{k \in I(x)} F_k[A_i x + B_i u(x)] \leq -\beta\Psi(x)$$

This means that the control can be computed as a selection in $\Omega_{EAS}(x)$:

$$u \in \Omega_{EAS}(x) \tag{7.31}$$

One can consider, for instance, the minimal selection [AC84] which is continuous. As we have already pointed out, the sampling time of the implementation should be $T << \tau$.

Though mathematically sound, the selection problem is a relevant one, since no general tools are in general available to derive a closed form expression from (7.31). The main reason why this problem shows up is the lack of differentiability of the considered functions, which does not in general allow to derive a gradient-based expression for the control law. When B is known, one might fix the problem

by properly smoothing the polyhedral function, which is what is about to be discussed next. Indeed we can apply the smoothing procedure already discussed in Subsection 4.5.3. Let us consider the symmetric case. The computed polyhedral function $\|Fx\|_\infty$ is replaced by the function $\|Fx\|_{2p}$, which is expressed as

$$\nabla \|Fx\|_{2p}^T = (\|Fx\|_{2p})^{1-2p} \; G_p^T(x)F, \; x \neq 0$$

where

$$G_p(x) = [(F_1x)^{2p-1} \; (F_2x)^{2p-1} \; \dots \; (F_sx)^{2p-1}]^T.$$

The gradient based controller can then be computed according to (4.48). In particular in this case $\gamma(x)$ can be taken as $\gamma(x) = \gamma_0\|Fx\|_{2p}$, resulting in the control law

$$u(t) = -\gamma_0 \; (\|Fx\|_{2p})^{(2-2p)} \; B^TF^TG_p(x) \tag{7.32}$$

which, for γ_0 large enough, guarantees stability of the closed-loop system, as stated in the next proposition [BM99c].

Proposition 7.40. *Assume the polyhedral function $\|Fx\|_\infty$ is a control Lyapunov function which assures the level of contractivity $\hat{\beta} > 0$. Then, for any positive $\beta < \hat{\beta}$, there exists p and γ_0 such that the system with the control (7.32) assures the condition $D^+\|Fx\|_{2p} \leq -\beta\|Fx\|_{2p}$.*

The reader is referred to [BM99c] for further details, in particular for the computation of γ_0.

Example 7.41. Consider the following uncertain system:

$$\begin{bmatrix} \dot{x_1}(t) \\ \dot{x_2}(t) \end{bmatrix} = \begin{bmatrix} 0 & -1.5 + \delta(t) \\ -2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 10 \end{bmatrix} u$$

where $|\delta(t)| \leq 0.5$. A polyhedral Lyapunov function $\hat{\Psi}(x) = \|Fx\|_\infty$ with a decreasing rate $\hat{\lambda} = 0.8$ for the corresponding EAS with $\tau = 0.05$ was computed. This function is characterized by the plane matrix

$$F = \begin{bmatrix} 0.000 & 1.000 \\ 4.997 & -.4997 \\ 4.997 & -.2498 \end{bmatrix}$$

corresponding to the vertex matrix

$$X = \begin{bmatrix} 0.2501 & 0.2001 & 0.1001 \\ 1.0000 & 0 & -1.0000 \end{bmatrix}$$

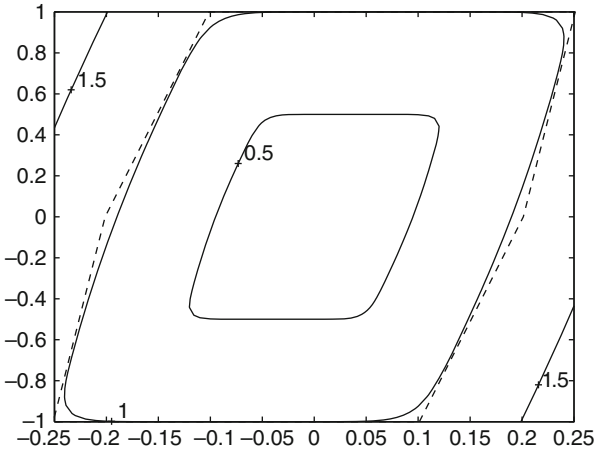


Fig. 7.3 The smoothed function level surfaces (plain lines) and the original polyhedral function level surfaces (dashed lines)

and the associated input values

$$U = \begin{bmatrix} -0.2500 & 1.6400 & 3.5200 \end{bmatrix}$$

The contractivity achievable by this polyhedral function is $\hat{\beta} = \frac{1-\lambda}{\tau} = 4$. By means of the proposed smoothing procedure, relaxing the speed of convergence to $\beta = 2$, a polynomial Lyapunov function with $p = 6$ and $\gamma_0 = 3.027$ can be found, thus obtaining the smooth control

$$u(x) = -\frac{3.027}{\Psi_6(x)^{11}} \begin{bmatrix} 10 & -4.970 & -2.485 \end{bmatrix} G_6(x).$$

The level surfaces of the original polyhedral function $\|Fx\|_\infty$ and of the polynomial one $\|Fx\|_{12}$ are depicted in Figure 7.3.

7.4 Gain scheduling linear controllers and duality

In the previous section it has been shown that for continuous-time systems, gain-scheduling (full information) stabilizability and robust stabilizability are equivalent, whereas such property does not hold for discrete-time systems. Also, it was shown that stabilizability of an LPV system is not equivalent to robust stabilizability via linear (not even dynamic) controllers. However, it turns out that for the gain scheduling problem the situation is quite different.

Again, we consider dynamic systems of the form

$$\dot{x}(t) = A(w(t))x(t) + B(w(t))u(t)$$

or

$$x(t+1) = A(w(t))x(t) + Bu(t)$$

(note that in the discrete-time case B is assumed constant).

Definition 7.42 (Linear gain scheduling stabilization). The previous system is said to be stabilizable via linear gain scheduling controller if there exists a control

$$z(t+1) = F(w(t))z(t) + G(w(t))x(t) \tag{7.33}$$

$$u(t) = H(w(t))z(t) + K(w(t))x(t) \tag{7.34}$$

such that the closed-loop system is GUAS.

The following theorem holds.

Theorem 7.43. *Assume the continuous-time system is stabilizable*[5]*. Then it is stabilizable via a linear gain scheduling controller.*

Proof. If the system is stabilizable, it admits a polyhedral control Lyapunov function. Then equations (7.19) (7.20) hold for some full row rank X (eq. (7.21) does not play any role since no control bounds are assumed). Let us now augment equation (7.19) by adding a matrix Z such that the matrix

$$X_{AUG} \doteq \begin{bmatrix} X \\ Z \end{bmatrix}$$

is invertible and let $V^{(i)} = ZH^{(i)}$, so that

$$A_i X + B_i U = XH^{(i)}$$
$$V^{(i)} = ZH^{(i)}$$

namely

$$\begin{bmatrix} A_i & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix} + \begin{bmatrix} B_i & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} U \\ V^{(i)} \end{bmatrix} = \begin{bmatrix} X \\ Z \end{bmatrix} H^{(i)} \tag{7.35}$$

[5]Either robustly or in the gain-scheduling sense, which are equivalent.

Consider now the matrices F_i, G_i, H_i, and K_i such that

$$\begin{bmatrix} U \\ V^{(i)} \end{bmatrix} = \begin{bmatrix} F_i & G_i \\ J_i & K_i \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix}$$

and the linear gain scheduling controller (7.33) with

$$\begin{bmatrix} F(w) & G(w) \\ J(w) & K(w) \end{bmatrix} = \sum_{i=1}^{s} w_i \begin{bmatrix} F_i & G_i \\ J_i & K_i \end{bmatrix}$$

(similarly for the other matrices).

By simple calculations it can be seen that the closed-loop system vertex matrices result in

$$A_i^{CL} = \begin{bmatrix} A_i & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} B_i & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} F_i & G_i \\ J_i & K_i \end{bmatrix}$$

which satisfies the condition

$$A_i^{CL} \begin{bmatrix} X \\ Z \end{bmatrix} = \begin{bmatrix} X \\ Z \end{bmatrix} H^{(i)}$$

which, together with (7.20), implies that the polyhedral function generated in the extended space by X_{AUG} is indeed a polyhedral Lyapunov function in view of item iii) of Proposition 7.19.

The reader is referred to [Bla00] for more details. The discrete-time version of the above, reported without proof (which is almost identical to the one reported for the continuous-time case) is the following.

Theorem 7.44. *Assume the discrete-time system is stabilizable in the gain-scheduling sense. Then it can be stabilized via a linear gain scheduling controller.*

The reader is referred to [BM03] for more details. The previous results admit some extension by duality which concern the output feedback case. Consider a discrete-time system of the form

$$x(t+1) = A(w(t))x(t) + Bu(t) \tag{7.36}$$

$$y(t) = Cx(t) \tag{7.37}$$

for which an observer is to be determined. Motivated by the purely linear case, generalized observer of the form

$$\begin{aligned} z(t+1) &= P(w(t))z(t) - L(w(t))y(t) + T(w(t))Bu(t) \\ \hat{x}(t) &= Q(w(t))z(t) + R(w(t))y(t) \end{aligned} \tag{7.38}$$

with $z(t) \in \mathbb{R}^s$, is analyzed. Such an observer, for a given *constant* $\bar{w}$, represents the most general form of a linear observer [O'R83].

Also, it is known that, for a given constant $\bar{w}$, for (7.38) to be an observer, the following necessary and sufficient conditions must hold

$$P(\bar{w})T(\bar{w}) - T(\bar{w})A(\bar{w}) = L(\bar{w})C \tag{7.39}$$

$$Q(\bar{w})T(\bar{w}) + R(\bar{w})C = I \tag{7.40}$$

where $P(\bar{w})$ is a stable matrix (i.e., its eigenvalues are inside the open unit disk).

The main problem is now to see what happens when $w(t)$ is time-varying, as in our case. According to [BM03], there is no restriction in assuming (7.38) of the form

$$\begin{aligned} z(t+1) &= P(w(t))z(t) - L(w(t))y(t) + TBu(t) \\ \hat{x}(t) &= Q(w(t))z(t) + R(w(t))y(t), \end{aligned} \tag{7.41}$$

namely with T *constant and of full column rank* with the conditions

$$P(\bar{w})T - TA(\bar{w}) = L(\bar{w})C \tag{7.42}$$

$$Q(\bar{w})T + R(\bar{w})C = I \tag{7.43}$$

It is quite easy to see how this observer works. Consider the variables

$$r(t) = Tx(t) - z(t), \quad e(t) = \hat{x}(t) - x(t).$$

Simple computations yield

$$r(t+1) = P(w(t))r(t) \tag{7.44}$$

$$e(t) = -Q(w(t))r(t) \tag{7.45}$$

Thus, if matrix $P(w)$ is stable, the variable $r(t)$ converges to 0, so that $e(t) \to 0$, which in turn implies that $\hat{x}(t) \to x(t)$.

The problem is then that of assuring the stability of $P(w(t))$. We have seen that for this property the existence of a polyhedral Lyapunov function is a necessary and sufficient condition. This result can be achieved by duality.

To state the result in its generality, let us now formally define the primal system as

$$\begin{aligned} x(t+1) &= A(w(t))x(t) + Bu(t) \\ y(t) &= x(t) \end{aligned} \tag{7.46}$$

and its dual as

$$\begin{aligned} x(t+1) &= A^T(w(t))x(t) + u(t) \\ y(t) &= B^T x(t) \end{aligned} \tag{7.47}$$

Remark 7.45. It must be said that, in the standard literature, the dual of a linear time-varying system is normally obtained by transposition and time reversing (e.g., $A^T(-t)$)[6]. We refer the reader to the seminal work [AM81] for a clear statement of the problem, as well as for a quadratic based solution.

Since we have established that the generalized observer design and the state feedback design are dual problems, we can focus on the stabilization via state feedback only.

Theorem 7.46. *The following statements are equivalent.*

i) There exist a static control law $u(t) = \Phi(x(t), w(t))$ and a polyhedral function which is a Lyapunov function for system (7.46).
ii) There exist a full row rank matrix $X \in \mathbb{R}^{n\times l}$, s matrices $P^{(h)} \in \mathbb{R}^{l\times l}$ and $U^{(h)} \in \mathbb{R}^{m\times l}$, such that for every $h = 1, \ldots, s$

$$A_h X + BU^{(h)} = XP^{(h)}, \quad \text{with } \|P^{(h)}\|_1 < 1. \tag{7.48}$$

iii) System (7.46) is stabilizable via linear gain scheduled controllers.
iv) System (7.46) is exponentially stabilizable via linear gain scheduled controllers.

Proof. See [BM03].

The next theorem introduces a duality result.

Theorem 7.47. *The following statements are equivalent.*

i) There exists a linear gain scheduling observer for system (7.46) of the form (7.41) with T full column rank.
ii) There exist a full column rank matrix $F \in \mathbb{R}^{l\times n}$, s matrices $H^{(h)} \in \mathbb{R}^{l\times l}$ and s matrices $Y^{(h)}$ such that, for every $h = 1, \ldots, s$, the dual equation of (7.48) holds:

$$FA_h + Y^{(h)}C = H^{(h)}F, \quad \text{with } \|H^{(h)}\|_\infty < 1. \tag{7.49}$$

iii) The dual system

$$x(t+1) = A^T(w(t))x(t) + C^T u(t)$$

is gain-scheduling stabilizable.

Proof. See [BM03].

[6] Roughly, in our case it is not necessary since if $A(t)$ is admissible also $A(-t)$ is such.

7.4.1 Duality in a quadratic framework

An important duality relation also holds in the framework of quadratic stability for gain scheduling linear design. If the parameters are available on-line one can always use the observer

$$\dot{z}(t) = A(w(t))z(t) + L(w(t))Cz(t) - L(w(t))y(t) + Bu(t) \tag{7.50}$$

As usual, denoting by $e(t) \doteq z(t) - x(t)$, we derive the error equation

$$\dot{e}(t) = [A(w(t)) + L(w(t))C]e(t)$$

This system is quadratically stable if and only if its dual

$$\dot{x}(t) = [A(w(t)) + L(w(t))C]^T x(t) = [A(w(t))^T + C^T L(w(t))^T]x(t)$$

is such. Then we can infer the following.

Proposition 7.48. *System $(A(w(t)), C)$ (no matter which is B) can be detected by means of the linear gain-scheduled observer (7.50) if and only if its dual $(A(w(t))^T, C^T)$ (no matter which is the "output" matrix B^T) is quadratically gain-scheduling stabilizable via linear controller $u = K(w)x$.*

Some kind of mixed relation holds if we assume that L and, respectively, K must be independent of w.

Proposition 7.49. *System $(A(w(t)), C)$ can be detected by means of the linear gain-scheduled observer (7.50), with L constant, if and only if its dual $(A(w(t))^T, C^T)$ is quadratically robustly stabilizable via constant linear controller $u = Kx$.*

It is apparent that, even in the quadratic framework, the existence of a robustly stabilizing constant linear gain does not imply the existence of some kind of "robust" observer. Indeed the structure of the Luenberger observer must replicate the dynamic of the system, which is impossible if the parameter w is unavailable on-line.

For further detail on the problem of gain scheduling control and its solution via quadratic functions the reader is referred to specialized surveys, for instance [SR00].

7.4.2 Stable LPV realization and its application

We anticipate a subject that will be reconsidered in more detail in Subsection 9.6.3. The issue concerns the parametrization of stabilizing compensators and its application to LPV design.

For brevity, we consider the case of an LPV system

$$\begin{aligned} \tfrac{d}{dt}x(t) &= A(w)x(t) + B(w)u(t) \\ y(t) &= C(w)x(t) \end{aligned}$$

which is quadratically stabilizable. Precisely, let us consider the case in which the two inequalities

$$\begin{aligned} PA(w)^T + A(w)P + B(w)U(w) + U(w)^T B(w)^T &\prec 0 \\ A(w)^T Q + QA(w) + Y(w)C(w) + C(w)^T Y(w)^T &\prec 0 \end{aligned}$$

are satisfied for some positive definite symmetric $n \times n$ matrices P and Q, and matrices $U(w) \in \mathbb{R}^{m\times n}$, $Y(w) \in \mathbb{R}^{n\times p}$. The previous ones are standard quadratic stabilizability conditions which involve LMIs [BP94, AG95, BEGFB04]. If they are satisfied, then the following observer-based compensator turns out to be stabilizing

$$\begin{aligned} \tfrac{d}{dt}\hat{x}(t) &= (A(w) + L(w)C(w) + B(w)J(w))\hat{x}(t) - L(w)y(t) + B(w)v(t) \\ u(t) &= J(w)\hat{x}(t) + v(t) \end{aligned} \tag{7.51}$$

where

$$J(w) = U(w)P^{-1} \qquad L(w) = Q^{-1}Y(w)$$

The input $v(t)$ will be used later. For the moment being let $v = 0$. To prove that the control is quadratically stabilizing denote, as usual, the error as $e(t) = \hat{x}(t) - x(t)$, so that

$$\frac{d}{dt}\begin{bmatrix} x(t) \\ e(t) \end{bmatrix} = \begin{bmatrix} A(w) + B(w)J(w) & B(w)J(w) \\ 0 & A(w) + L(w)C(w) \end{bmatrix} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix}$$

This is a triangular system for which the block matrices on the diagonal $A(w) + B(w)J(w)$ and $A(w) + L(w)C(w)$ satisfy the conditions

$$P^{-1}(A(w) + B(w)J(w)) + (A(w) + B(w)J(w))^T P^{-1} \prec 0$$

and

$$(A(w) + L(w)C(w))Q^{-1} + Q^{-1}(A(w) + L(w)C(w))^T \prec 0$$

hence it is exponentially robustly stable.

We show now how the signal $v(t)$ comes into play. Assume that in the previous machinery the signal $v(t)$ is generated as the output of the following system

$$v(t) = T(w)C(\hat{x} - x) = T(w)Ce(t)$$

where $T(w)$ is a "stable operator." Plugging such a T modifies the compensator, but it cannot destabilize the plant as long as we can assure that

$$e(t) \to 0 \quad \text{implies} \quad v(t) \to 0$$

Although in general any input–output stable operator T would fit, we limit ourselves to operators T which can be described by finite-dimensional LPV systems.

By the Youla–Kucera parametrization theory [ZDG96, SPS98], the following well-known proposition (which will be reconsidered later in Proposition 9.40) holds:

Proposition 7.50. *Let w be* constant *and let $W(w,s)$ be the transfer function of any compensator stabilizing the plant. Then there exists a Youla–Kucera parameter $T(w,s)$ such that the resulting compensator is stabilizing and has transfer function $W(w,s)$.*

Therefore one can realize any compensator scheduled as a function of w which is stabilizing *for constant* w. For instance, one could consider a family of compensators $W_{opt}(w,s)$ optimal with respect to w and achieve optimality for any w, as long as w is constant (or time-varying with a very slow variation rate). However $w(t)$ can vary, hence the following question arises.

Question: if one chooses a parametrized compensator transfer function

$$W_{opt}(w,s)$$

and implements it by means of some parametrized realization, is stability assured *even under variations of $w(t)$?*

The answer to the previous question is no. However, there always exists a *suitable realization* such that:

- for constant w, the compensator transfer function is the desired one $W_{opt}(w,s)$;
- stability[7] is assured under arbitrary variations $w(t) \in \mathcal{W}$.

Consider, for instance, the system in Example 7.0.1. One could synthesize an optimal local controller which is "good" for small variations from the reference. This will provide a transfer function $W_{opt}(w,s)$. On the other hand, it is desirable that the system remains stable even in the case of large transients among reference values.

We do not prove the previous claim here since it will be discussed later in the context of switching systems (Theorem 9.45). We only anticipate the method. Assume that the compensator transfer function $W_{opt}(w,s)$ corresponds to a Youla–Kucera parameter $T_{opt}(w,s)$. The problem is solved if one realizes such a parameter as

$$\begin{aligned} \dot{z}(t) &= F_T(w)z(t) + G_T(w)o(t) \\ v(t) &= H_T(w)z(t) + K_T(w)o(t) \end{aligned} \tag{7.52}$$

[7] Not optimality.

in a stable way. Precisely we must have

$$H_T(w)(sI - F_T(w))^{-1}G_T(w) + K_T(w) = T_{opt}(w,s)$$

with the additional condition that the LPV system (7.52) is LPV stable.

Example 7.51. Assume that the following transfer function is assigned

$$W(w,s) = \frac{\kappa}{s^2 + 2\xi s^2 + \xi^2 + w}$$

The realization

$$A = \begin{bmatrix} 0 & 1 \\ -(\xi^2 + w) & -2\xi \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} \kappa & 0 \end{bmatrix}$$

is not necessarily LPV stable if $\xi > 0$ is small and $0 < w^- \leq w(t) \leq w^+$ is time-varying. This is indeed a system of the form (7.1) (describing the physical system in Fig. 7.1). Conversely

$$A = \begin{bmatrix} -\xi & -\sqrt{w} \\ \sqrt{w} & -\xi \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & \kappa/\sqrt{w} \end{bmatrix}$$

is the realization of the same transfer function for constant w and it is an LPV stable system. If w is constant, then the two realizations are clearly equivalent. This is absolutely not true when $w(t)$ is time-varying.

If we realize the compensator by means of a pre-stabilizer as in (7.51) equipped by a proper operator $v = To$ as in (7.52), then any transfer function can be matched. Moreover, if we realize the Youla–Kucera parameter (7.52) in such a way that it is quadratically stable when w is time-varying, then our goal is achieved.

Remark 7.52. The word "realization" is formally correct when w is constant. With an abuse of word we say that (7.52) is a realization as long as its transfer function is the desired one for any constant w. The additional requirement is that it remains stable, for $w(t)$ time-varying, although (7.52) is not the realization of a transfer function anymore.

In Subsection 9.6.3 we will prove the following (Lemma 9.41): Given a stable square matrix F, there exists an invertible T such that $\hat{F} = T^{-1}FT$ has $P = I$ as Lyapunov matrix. Then the LPV stable operator T described by (7.52) can be achieved as follows.

1. Given the current w, compute any realization $\hat{F}_T(w), \hat{G}_T(w), \hat{H}_T(w), \hat{K}_T(w)$ such that

$$\hat{H}_T(w)(sI - \hat{F}_T(w))^{-1}\hat{G}_T(w) + \hat{K}_T(w) = T_{opt}(w,s).$$

2. Take the matrices $(F_T(w), G_T(w), H_T(w), K_T(w))$ which implement $T_{opt}(w,s)$ as in (7.52) given by

$$F_T(w) = T^{-1}(w)\hat{F}_T(w)T(w), \quad G_T(w) = T(w)^{-1}\hat{G}_T(w),$$
$$H_T(w) = \hat{H}_T(w)T(w), \qquad K_T(w) = \hat{K}_T(w)$$

where $T(w)$ has the property that $T^{-1}(w)\hat{F}_T(w)T(w)$ has the identity I as Lyapunov matrix.

A problem of the previous machinery to implement the compensator is that these computations have to be performed on-line for the current w, i.e. within the sampling time.

7.4.3 Separation principle in gain-scheduling and robust LPV control

We have seen that there are interesting duality properties in LPV control theory. The following result establishes a separation principle for gain-scheduling design [BCMV10]. Consider the LPV system

$$\begin{aligned} \dot{x}(t) &= A(w(t))x(t) + B(w(t))u(t) \\ y(t) &= C(w(t))x(t) \end{aligned} \tag{7.53}$$

and the class of LPV compensators of the form

$$\begin{aligned} \dot{z}(t) &= F(w(t))z(t) + G(w(t))y(t) \\ u(t) &= H(w(t))z(t) + K(w(t))y(t) \end{aligned} \tag{7.54}$$

Theorem 7.53. *System (7.53) can be stabilized by a control of the form (7.54) if and only if the two following conditions are satisfied:*

- *the system is stabilizable via state feedback by a compensator of the form*

$$\dot{z}(t) = F_{SF}(w(t))z(t) + G_{SF}(w(t))x(t), \tag{7.55}$$
$$u(t) = H_{SF}(w(t))z(t) + K_{SF}(w(t))x(t), \tag{7.56}$$

- *there exists a generalized observer of the form*

$$\dot{r}(t) = Q(w(t))r(t) - L(w(t))y(t) + RB(w(t))u(t), \tag{7.57}$$
$$\hat{x}(t) = Mr(t), \tag{7.58}$$

which produces an asymptotic estimation of the state: $\hat{x}(t) - x(t) \to 0$ for all $u(\cdot)$

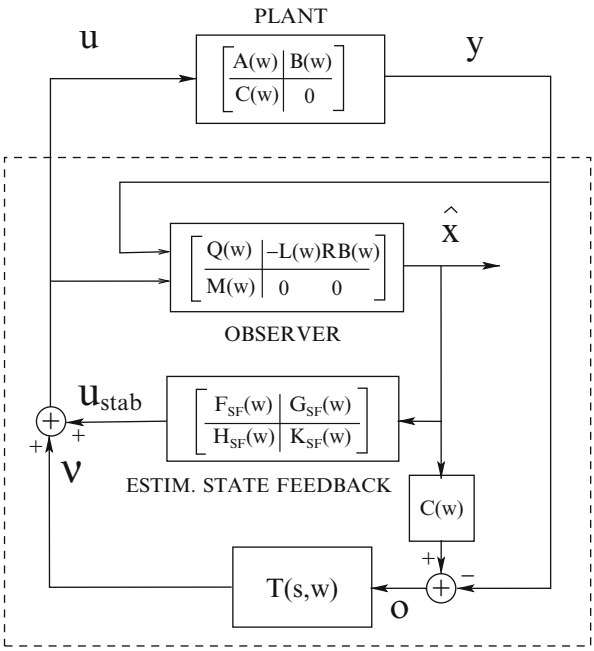


Fig. 7.4 Observer-based controller

If the conditions of the theorem are satisfied, then an observer-based compensator is achieved as in Fig. 7.4. It is also known that state feedback control and observer design are dual problems [BM03, BCMV10].

Theorem 7.54. *System (7.53) can be stabilized by a gain scheduling state feedback if and only if its dual admits a gain scheduling generalized observer.*

Conversely in robust control neither separation principles nor duality properties can be established, at least under the standard definitions. Roughly, we say that the stabilization problem for a class of systems enjoys the *separation principle* if whenever we know that a system in this class is stabilizable, then a possible stabilizer is achieved by means of a state observer, which reconstruct asymptotically the state *for all u and w*, and an estimated-state-feedback [BG86].

We show that in robust control of LPV systems there is no hope to establish a general separation principle. This can be done by means of a very simple counterexample (which is taken and suitably adapted from [BM99b]). Consider the LPV system

$$\begin{aligned} \dot{x}_1 &= -\omega x_2 \\ \dot{x}_2 &= \omega x_1 + \beta u \\ y &= \gamma x_2 \end{aligned}$$

with $\mu \leq \omega(t) \leq 1/\mu$, $\rho \leq \beta(t) \leq 1/\rho$ and $\xi \leq \gamma(t) \leq 1/\xi$, and with $0 < \mu, \rho, \xi < 1$. This system can be stabilized by the control $u = -\kappa y$, as it can be shown by means of the Lyapunov function

$$\Psi(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$$

whose derivative, along the system trajectory, is

$$\dot{\Psi}(x_1, x_2) = -\kappa\gamma\beta x_2^2$$

Although the derivative is negative semi-definite, stabilizability can be proved by Krasowskii arguments.

However, no device can asymptotically reconstruct the state with arbitrary $u(\cdot)$. For instance, take $u \equiv 0$, and consider the periodic trajectory which is on the unit circle. Whenever x_2 is in the strip $\xi \leq |x_2| \leq 1/\xi$, one might assume that $\gamma(t) = 1/|x_2(t)|$. The value of γ is ignored by the compensator. The output remains constant: $y(t) = 1$ in the upper strip and $y(t) = -1$ in the lower strip, therefore no information about the state is available. On the other hand, the state sweeps the circle with angular speed $\omega(t)$ and therefore, when the state enters the "blind strip" of Figure 7.5, no detection is possible and a persistent error is unavoidable if $\omega(t)$ changes and the compensator has no information about ω and γ. It is easy to modify this example by perturbing the second equation as $\dot{x}_2 = \omega x_1 + \epsilon x_2 + \beta u$, with a small term $\epsilon > 0$ to show that the error might even diverge.

Even duality fails in robust control. Here we show a continuous-time example by adapting that proposed in [BM99b]. The system is

$$\begin{aligned} \dot{x}_1 &= \omega x_2 + u \\ \dot{x}_2 &= -\omega x_1 + \beta u \end{aligned} \tag{7.59}$$

with $|\beta| \leq \rho$ and $\mu \leq \omega \leq 1/\mu$. This system can be stabilized by the gain scheduling feedback

$$u(t) = -\kappa[x_1 + \beta(t)x_2]$$

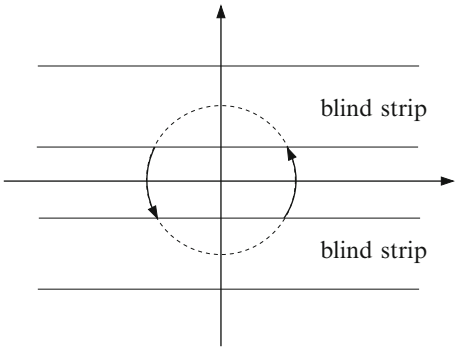


Fig. 7.5 The blind strip

Consider the same Lyapunov function as before $\Psi(x_1, x-2) = (x_1^2 + x_2^2)/2$ to get

$$\dot{\Psi}(x_1, x_2) = -\kappa[x_1 + \beta x_2]^2$$

Again, the derivative is only non-positive for arbitrary $\kappa > 0$. However, if κ is small the system has a "rotating behavior," so that the state periodically reaches the region in which

$$[x_1 + \beta x_2]^2 < 0$$

and the Lyapunov derivative is necessarily negative so that $x(t) \to 0$.

In view of Theorem 7.46, the existence of a gain scheduling stabilizing feedback implies that the system is also robustly stabilizable. However, the dual system

$$\begin{aligned} \dot{x}_1 &= -\omega x_2 \\ \dot{x}_2 &= \omega x_1 \\ y &= x_1 + \gamma x_2 \end{aligned}$$

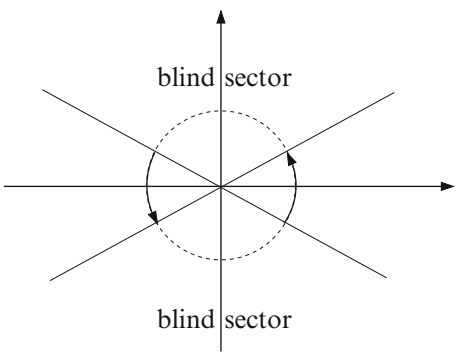
$|\gamma| \leq \rho$ and $\mu \leq \omega \leq 1/\mu$, admits a GS observer but not a robust observer. Indeed, the following gain scheduling Luenberger observer

$$\dot{\hat{x}} = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix} \hat{x} + \begin{bmatrix} \kappa \\ \kappa\gamma(t) \end{bmatrix} \left(y - \begin{bmatrix} 1 & \gamma(t) \end{bmatrix} \hat{x} \right)$$

provides an asymptotic estimate of the true state x when $\kappa > 0$ (this can be seen by using the same Lyapunov arguments used for the gain scheduled state feedback). If no parameter measurements are available, say in the robust case, no asymptotic estimator can be found. Indeed, in the "blind" sector (see Fig. 7.6)

$$\mathcal{S}_\rho = \{(x_1, x_2) : \ |x_1| \leq \rho |x_2|\}$$

Fig. 7.6 The blind sector



$\gamma(t) = -x_1(t)/x_2(t)$ is admissible and yields $y(t) = x_1(t) + \gamma x_2(t) = 0$. For such $\gamma(t)$ whenever $x(t) \in \mathcal{S}_\rho$, no output information is available. On the other hand, $x(t)$ rotates at a variable speed $\omega(t)$, so that, again, no state detection is possible.

Obviously, there are no LPV systems which are robustly detectable whose dual is not state feedback stabilizable. Indeed robust detection implies gain scheduling detection which implies gain scheduling state-feedback stabilizability, hence robust stabilizability, of the dual.

To conclude this subsection we stress that, as long as we are considering gain-scheduling design, we may always count on duality and separation principles no matter if we are considering quadratic or polyhedral Lyapunov functions. Conversely, in the case of robust control, there is no "obvious" way to "dualize" all the procedures we have presented for state feedback. In the section concerning robust set-theoretic estimation we will see that some set-theoretic procedures are possible for robust state detection but these are essentially different from the procedure proposed for state feedback design.

7.5 Exercises

1. Consider a single-input system of the form $\dot{x} = A(w)x + Bu$ in which the goal is that of assuring a contractivity level β. Show that for some realizations of $w(t)$ a gain-scheduling control is preferable to a state feedback controller from the actuator effort point of view, in the sense that, on the average, the control effort is greatly reduced. What about the worst case?
2. Show that the system of equation (7.1) for $0 < \rho^- < \rho^+$ can be destabilized via a proper time-varying $\rho(t)$, provides that $\alpha > 0$ is small enough. (Hint: consider the case in which $\alpha = 0$ and show that, for $\rho(t) = \rho^-$ and $\rho(t) = \rho^+$, the trajectories are on ellipses. Show how to jump from a trajectory to another in order to go arbitrarily far from $0 \ldots$)
3. Show that the system with equation (7.1) for $0 < \rho^- < \rho^+$ and $\alpha < 0$ can be stabilized by choosing a proper time-varying $\rho(t)$, provided that $\alpha > 0$ is small enough. (Hint: read the hint in Exercise 2)
4. Equation (7.11) is valid for a constant B only. Why?
5. Consider the equation (7.16) and show that $\rho_Q = \sqrt{3}/2$, by computing the inverse of the $\mathcal{H}_\infty$ norm of the system.
6. Solve the previous exercise directly, by developing all the computation in (7.16) in detail.
7. Quadratic stabilizability is equivalent to linear quadratic stabilizability if $B = B_0(I + \Delta)$ with $\|\Delta\|_2 \leq \nu < 1$ [BCL83]. Prove it (Hint: take a gradient based control $u = -\gamma B_0^T P x \ldots$).

8. Show why convexity is essential in Proposition 7.6. Hint: take the linear system $x(t+1) = [A_1 w_1(t) + A_2 w_2(t)]x$, $w_1 + w_2 = 1$, $w_1, w_2 \geq 0$, with

$$A_1 = \rho \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad A_2 = \rho \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

   $\rho = 0.99$, and the candidate non-convex function

$$\Psi(x_1, x_2) = \min\{|x_1|/2 + |x_2|; |x_1| + |x_2|/2\}$$

   Show that $\Psi(x_1, x_2)$ is a Lyapunov function if‘ the system is of the switching type, namely only $(1, 0)$ and $(0, 1)$ are admitted for (w_1, w_2).
9. In the previous exercise, is

$$\hat{\Psi}(x_1, x_2) = \max\{|x_1|/2 + |x_2|, |x_1| + |x_2|/2\}$$

   a suitable Lyapunov function to prove stability? (consider both polytopic and switching case).

Chapter 8
Control with time-domain constraints

Constraints are encountered practically in every real control problem. It is a tradition (although questionable) that in many textbooks constraints are mentioned but, with several exceptions, the design of control systems which take into account constraints is frequently disregarded.

Basically, there are two ways to cope with constraints and precisely

- the control design is performed disregarding the constraints and the effect of the constraints is considered a posteriori, by simulation analysis or by computing the domain of attraction;
- the control design is performed directly by taking constraints into account.

Clearly the first approach is more suitable when the problem primary aspect is that of optimizing local performances, while the constrains have a secondary role. Conversely, when the constraints are critical, the second type of design is often preferable. A typical example is the control problem of a cart-pendulum system. It is quite reasonable that, if the system is to be controlled in the lower (stable) position, the presence of constraints is not a critical issue. Conversely, if the system is to be controlled in the more challenging and not purely academic (see [GM04] for a relevant application) upper position, constraints have necessarily to be kept into account in the design stage, since disregarding them might lead to far-from-ideal behavior of the control loop. Indeed, if limitations in the control action are present, for large values of the angle the system might not even be brought to the upper position. Things get even worse if state constraints (e.g., the cart position being limited) are also present.

In this section the problem will be analyzed by considering input and output constraints by means of a set-theoretic approach. The relevance of such a set-theoretic approach can be appreciated in view of the following basic result. Consider the system

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_8

$$\begin{array}{ll} \dot{x}(t) = f(x(t), u(t)) & (f(x(t), u(t), w(t))) \\ y(t) = g(x(t)) & (g(x(t), w(t))) \end{array} \tag{8.1}$$

where the control $u(t)$ is constrained as $u(t) \in \mathcal{U}$ and the state as $x(t) \in \mathcal{X}$ (and $w(t) \in \mathcal{W}$ is the external output, if any[1]), and the (stabilizing, if possible) controller

$$\begin{array}{ll} \dot{z}(t) = h(z(t), y(t)) & (h(z(t), y(t), w(t))) \\ u(t) = k(z(t), y(t)) & (k(z(t), y(t), w(t))) \end{array} \tag{8.2}$$

Definition 8.1 (Admissible set). A set $\mathcal{P}$ in the extended state space is said to be admissible if, for all $[x^T \; z^T]^T \in \mathcal{P}$,

$$u = k(z, g(x)) \in \mathcal{U} \;\text{ and }\; x \in \mathcal{X}$$

(for all $w(t) \in \mathcal{W}$).

The following result holds.

Theorem 8.2. *Given the dynamic system (8.1) and the controller (8.2), then $x(t) \in \mathcal{X}$ and $u(t) \in \mathcal{U}$ for all $t \geq 0$ and for all $w(t) \in \mathcal{W}$ if and only if the initial state $[x(0)^T \; z(0)^T]^T$ is included in a set $\mathcal{P}$ which is admissible and (robustly) positively invariant for the closed-loop system.*

Proof. It is rather obvious that the condition is sufficient. Indeed, if the initial condition belongs to a positively invariant set $\mathcal{P}$, then the (extended) state evolution will belong to $\mathcal{P}$ and since this set is also admissible no constraint violation will occur. The fact that it is necessary is simply derived by noticing that, if one considers the set $\mathcal{P}_{max}$ of all initial state $[x(0)^T \; z(0)^T]^T$ for which the constraints are not violated, by definition any future state $[x(\tau)^T \; z(\tau)^T]^T$ is also in such a set (because $[x(\tau)^T \; z(\tau)^T]^T$ is still an initial state for the future transient, i.e. $t \geq \tau$, then $[x(\tau)^T \; z(\tau)^T]^T \in \mathcal{P}_{max}$).

Even in its simplicity, the extent of the previous result is fundamental: to solve a constrained control problem, a controlled-invariant set, say an invariant admissible set which can be associated with a control law for which there is no constraints violation, must exist.

This "if it is fine now it will be fine forever" concept, naturally associated with the invariance concept, is of course involved and it is crucial because it assures that violations will not occur in the future.

Though such a concept is always present in constrained control problems, there are techniques to handle constraints in which the word invariance is not even mentioned, but it hidden, as in the next example. Consider the feedback loop in Fig. 8.1, where the scalar control is constrained as $|u(t)| \leq \bar{u}$ and the system has to satisfy such a constraint for all the scalar "tracking signals" $|r(t)| \leq \bar{r}$. Assuming

[1] Though in this chapter the main focus will be put on systems without uncertainties, the signal w is considered here for the sake of generality.

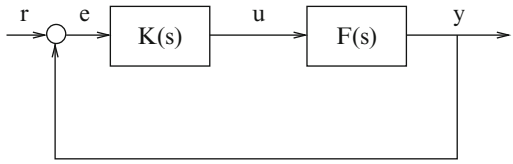


Fig. 8.1 The feedback loop for tracking

closed-loop stability, this no-constraint-violation condition can be analyzed as a "worst case" output analysis by considering r as input and the control action u as output. More precisely, it can be stated as an ∞-to-∞ norm condition under the assumption that the initial condition is 0.

To provide a set-theoretic interpretation of the above problem along the lines of Subsection 6.4.1, consider any state space realization of the closed-loop system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Er(t) \\ u(t) &= Cx(t) + Hr(t) \end{aligned}$$

Proposition 8.3. *The next statements are equivalent*

- *for $x(0) = 0$ the stable loop of Figure 8.1 is such that the control constraints are satisfied for all reference signals such that $|r(t)| \leq \bar{r}$;*
- *the induced ∞-to-∞ norm[2] is such that*

$$\mu = \sup_{r \neq 0} \frac{\|u\|_\infty}{\|r\|_\infty} \leq \frac{\bar{u}}{\bar{r}};$$

- *the largest robustly invariant set $\mathcal{S}_\infty$ of the system included in the strip*

$$\mathcal{P}[C, \bar{u} - |H|\bar{r}] = \{x : \ |Cx| \leq \bar{u} - |H|\bar{r}\}$$

is not empty (in particular, by symmetry, it includes the origin).

In many books and papers it is often assumed that the initial condition is zero. From this point of view, the last statement of the above proposition might appear to be useless since the computation of $\mathcal{S}_\infty$ is normally much harder than the computation of the ∞-to-∞ norm (the so-called $\mathcal{L}_1$ norm) which is given by

$$\mu = \int_0^\infty |Ce^{At}E|dt + |H|$$

Though the above assertion is clearly true, a set-theoretic investigation of the problem has some advantages:

- $\mathcal{S}_\infty$ is the set of *all the initial conditions* (not just the 0 one) for which there is no violation;

[2] We remind that the infinity norm of a signal is defined as $\|u\|_\infty \doteq \sup_{t \geq 0} |u(t)|$.

- the results can be extended to uncertain systems of the form $(A(w), E, C, H)$.
- the analysis can be extended, without difficulties, to asymmetrical constraints of the form

$$u^- \leq u \leq u^+, \quad r^- \leq r \leq r^+,$$

- the $\mathcal{L}_1$ control optimization is undoubtedly a very nice theory [BG84, DP87], but it is not simple to deal with. Furthermore, a set-theoretic interpretation gives the designer a complementary reading.

With the above motivations in mind, in this section the constrained input problem will be faced by means of invariant sets. We refer the reader to the nice book [HL01] for a less set-invariant alternative.

8.1 Input constraints

In this section, the constrained input control problem is specifically considered. Assume that the control is constrained as

$$u(t) \in \mathcal{U}v$$

where $\mathcal{U}$ is a C-set.

As it is well known, controllability and reachability problems in the presence of constraints are completely different from the same problems in the absence of constraints.

Consider a linear time-invariant system of the form

$$\dot{x}(t) = Ax(t) + Bu(t)$$

or its corresponding discrete-time version

$$x(t+1) = Ax(t) + Bu(t)$$

Let

$$\mathcal{R}_T \ \ (\text{respectively } \ \mathcal{C}_T)$$

be the reachability (controllability) set from (to) the origin in time T, according to Definition 6.1 (Definition 6.2), and let

$$\mathcal{R}_\infty \ \ (\text{respectively } \ \mathcal{C}_\infty)$$

be the set of all the reachable (controllable) states in a *finite, but arbitrary, time* from (to) the origin.

Controllable and reachable sets are strictly related and precisely, according to the concepts expressed in Subsection 6.1.1, the constrained reachability problem can be equivalently stated as a controllability problem for the "reverse-time" system:

$$\dot{x}(t) = -Ax(t) - Bu(t), \quad \big(x(t+1) = A^{-1}x(t) - A^{-1}Bu(t)\big)$$

(in the discrete-time case, it is assumed for brevity that A is invertible). The main properties and relations of reachable and controllable sets for linear systems are summarized in the next proposition (part of them are reported for the sake of completeness, since they have already been stated in Section 6.1.1)

Proposition 8.4.

i) $\bar{x} \in \mathbb{R}^n$ *is reachable under constraints if and only if it is controllable under constraints for the reverse system (and vice versa).*

ii) Reachability (controllability) sets are monotonically increasing (decreasing), i.e., for $T_1 \leq T_2 \leq \infty$,

$$\mathcal{R}_{T_1} \subseteq \mathcal{R}_{T_2} \subseteq \mathcal{R}_\infty, \qquad \mathcal{C}_{T_1} \subseteq \mathcal{C}_{T_2} \subseteq \mathcal{C}_\infty$$

iii) The sets $\mathcal{R}_T$ *and* $\mathcal{C}_T$*, as well as* $\mathcal{R}_\infty$ *and* $\mathcal{C}_\infty$*, are convex sets.*

The next proposition is a preliminary step for the reachability and controllability analysis.

Proposition 8.5. *Denote by*

$$\mathcal{R}(A,B) = range[B|AB|A^2B|\dots|A^{n-1}B]$$

the reachable subspace (without constraints) of the pair (A,B)*. Then*

$$\mathcal{R}_T \subseteq \mathcal{R}(A,B) \quad \big(respectively\ \ \mathcal{C}_T \subseteq \mathcal{R}(A^{-1},A^{-1}B) = \mathcal{C}(A,B)\big)$$

The proof of the proposition can be immediately derived by the Kalman reachability decomposition and is left to the reader. In view of this property, it is henceforth possible to work under the assumption that the pair (A,B) is reachable, precisely

$$\mathcal{R}(A,B) = range[B|AB|A^2B|\dots|A^{n-1}B] = \mathbb{R}^n$$

For the sake of completeness, we state the next lemma which tells us that, as long as the system state is sufficiently close to the origin, constraints are not a major issue.

Lemma 8.6. *If* (A,B) *is reachable, then*

i) there exists a neighborhood $\mathcal{B}_1$ *of the origin such that for all* $x(0) \in \mathcal{B}_1$ *the state can be driven to* 0 *in finite time without constraints violation;*

ii) given any stabilizing linear feedback control $u = Kx$ *and any* $\eta > 0$*, there exists a neighborhood* $\mathcal{B}$ *of the origin and* $\epsilon > 0$ *such that for all* $x(0) \in \mathcal{B}$ *and for all* $\|w\| \leq \epsilon$ *the state of the system*

$$\dot{x}(t) = (A + BK)x(t) + w(t), \quad (\textit{respectively } x(t+1) = (A + BK)x(t) + w(t))$$

remains bounded and $\|u(t)\| \leq \eta$*,*

iii) for any stabilizing linear feedback control there exist μ *and* ν *such that, if* $x(0) = 0$ *and* $\|w(t)\| \leq \epsilon$*, then* $\|x(t)\| \leq \epsilon\nu$ *and* $\|u(t)\| \leq \epsilon\mu$*.*

With the above in mind, the next lemma, which relates the controllable (reachable) set to the eigenspaces of the system, can be introduced.

Lemma 8.7. *If* (A, B) *is reachable, then*

i) if A has all its eigenvalues in the closed left-half-plane (resp. in the closed unit disk), then

$$\mathcal{C}_\infty = \mathbb{R}^n;$$

ii) if A has all its eigenvalues in the closed right-half-plane (resp. in the complement of the open unit disk), then

$$\mathcal{R}_\infty = \mathbb{R}^n;$$

The proof of this lemma can be found in [BS80] and in [Son84] (see also [HL01]). An important consequence of this result is the following. For a generic system, one can consider two subspaces in the state space: the first, $\mathcal{T}^{\bar{L}}$, related to the closed left-half-plane eigenvalues and the second $\mathcal{T}^R$ associated with the other eigenvalues. In other terms, one can apply a state transformation

$$\hat{A} = [T^{\bar{L}}|T^R]^{-1}A[T^{\bar{L}}|T^R], \quad \hat{B} = [T^{\bar{L}}|T^R]^{-1}B$$

where $T^{\bar{L}}$ and T^R denote two matrices whose columns form a basis for $\mathcal{T}^{\bar{L}}$ and $\mathcal{T}^R$, so that the system is reduced to

$$\hat{A} = \begin{bmatrix} A^{\bar{L}} & 0 \\ 0 & A^R \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} B^{\bar{L}} \\ B^R \end{bmatrix} \tag{8.3}$$

By exploiting the above, the following result concerning controllable sets can be presented.

Theorem 8.8. *Consider the system decomposition (8.3) and let* $\mathcal{C}_\infty^R$ *be the controllable set under constraints for the system* (A^R, B^R)*. Then*

$$\mathcal{C}_\infty = \mathcal{T}^{\bar{L}} + \mathcal{C}_\infty^R$$

Proof. A proof of this statement can be found in [Haj91]; here, an alternative proof based on Lemma 8.7 is reported. To show that any state in the set $\mathcal{T}^{\bar{L}} + \mathcal{C}_\infty^R$ can be driven to 0, consider initially the anti-stable system A^R and a control action $u_1(T) \in \mathcal{U}$, defined on the interval $[0, T_1]$, which drives $x_R(0) \in \mathcal{C}_\infty^R$ to 0 in time T_1, $x_R(T_1) = 0$. Such a control action and such a T_1 exist by definition of $\mathcal{C}_\infty^R$.

The problem now is to show that $x_{\bar{L}}(t)$ can be driven to zero without letting $x_R(t)$ "escape" from $\mathcal{C}_\infty^R$. Consider a control action of the form

$$u = K^R x_R + v, \quad \|v\| \leq \epsilon$$

where K^R is such that $A_2 = A^R + B^R K^R$ is asymptotically stable and has no common eigenvalues with $A^{\bar{L}}$. With the above control law, the controlled system becomes

$$\begin{bmatrix} \dot{x}^{\bar{L}} \\ \dot{x}^R \end{bmatrix} = \begin{bmatrix} A^{\bar{L}} & B^{\bar{L}} K^R \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} x^{\bar{L}} \\ x^R \end{bmatrix} + \begin{bmatrix} B^{\bar{L}} \\ B^R \end{bmatrix} v(t)$$

Consider the transformation

$$\begin{bmatrix} x^{\bar{L}} \\ x^R \end{bmatrix} = \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \begin{bmatrix} x^1 \\ x^R \end{bmatrix}$$

where X satisfies the Sylvester equation

$$A^{\bar{L}} X - X A_2 = -B^{\bar{L}} K^R$$

which admits a solution since $A^{\bar{L}}$ and A_2 have distinct eigenvalues. By applying the above transformation, the new equation

$$\begin{bmatrix} \dot{x}^1 \\ \dot{x}^R \end{bmatrix} = \begin{bmatrix} A^{\bar{L}} & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} x^1 \\ x^R \end{bmatrix} + \begin{bmatrix} B^1 \\ B^R \end{bmatrix} v(t)$$

is obtained. Note in passing that the second sub-system remains unchanged. In view of statement iii) of Lemma 8.6, since $x_R(T_1) = 0$, for any $\epsilon > 0$ we have that, if we limit v as $\|v(t)\| \leq \epsilon$, then $\|u(t)\| \leq \mu\epsilon$ and $\|x_R(t)\| \leq \nu\epsilon$, $t \geq T_1$, for some $\mu, \nu > 0$. Then, for ϵ small enough, we have $u(t) \in \mathcal{U}$.

Now, according to Lemma 8.7, it is possible to drive, in a proper interval of time $[T_1, T_2]$, $x^1(t)$ to $x^1(T_2) = 0$ by means of a bounded control $\|v(t)\| \leq \epsilon$, thus simultaneously assuring $u(t) \in \mathcal{U}$, for ϵ small enough. Once $x^1(T_2) = 0$, take $v(t) = 0$ so that $x^R(t) \to 0$, so that in time T_3 large enough the condition $\|x^R(T_3)\| \leq \delta$ can be met with δ arbitrarily small. Since in the meanwhile x^1 remains zero, in time T_3 large enough a neighborhood of the origin can be reached which is arbitrarily small. Statement i) of Lemma 8.6 finally assures that the state can be driven to zero in a finite interval $[T_3, T_4]$.

It is clear that, if the x_R-sub-system is not present, then the domain of attraction is not needed since it is a priori known that the system state can be driven to 0 for all initial conditions. In this case the problem can be faced in terms of the so-called global [Son84] or semi-global stabilization approach [LSS96, SSS00, SHS02].

As far as the reachability analysis is concerned, a similar decomposition has to be applied. In this case one has indeed to consider two subspaces: the first, $\mathcal{T}^L$, related to the open left-half-plane eigenvalues and the second, $\mathcal{T}^{\bar{R}}$, associated with the closed right-half-plane eigenvalues. Again, by applying a state transformation

$$\hat{A} = [T^L|T^{\bar{R}}]^{-1}A[T^L|T^{\bar{R}}], \quad \hat{B} = [T^L|T^{\bar{R}}]^{-1}B$$

where T^L and $T^{\bar{R}}$ denote two matrices whose columns form a basis for $\mathcal{T}^L$ and $\mathcal{T}^{\bar{R}}$, the system is reduced to

$$\hat{A} = \begin{bmatrix} A^L & 0 \\ 0 & A^{\bar{R}} \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} B^L \\ B^{\bar{R}} \end{bmatrix}. \tag{8.4}$$

The above decomposition allows us to derive the following reachability result.

Theorem 8.9. *Consider the system decomposition (8.4). Let $\mathcal{R}^L_\infty$ be the reachable set under constraints for the system (A^L, B^L). Then*

$$\mathcal{R}_\infty = \mathcal{T}^{\bar{R}} + \mathcal{R}^L_\infty$$

Proof. The proof follows immediately from Proposition 8.4 and the fact that the reverse system spectrum is symmetric, with respect to the imaginary axis, to that of the "direct system."

8.1.1 Construction of a constrained control law and its associated domain of attraction

The previous results are essentially theoretical. Our intention is that of providing a constructive way to drive the state asymptotically to zero by means of a feedback control law. Precisely, the following two problems are addressed:

- that of finding a set $\mathcal{S}$ such that, for all initial states $x(0) \in \mathcal{S}$, $x(t) \to 0$ with $u(t) \in \mathcal{U}$;
- that of finding a proper stabilizing feedback law $u = \Phi(x) \in \mathcal{U}$;

As already mentioned, it is possible to restrict our attention to sets $\mathcal{S}$ which are controlled-invariant since the problem, under stabilizability assumptions, is solvable if and only if such a set exists.

We also consider the special but important case in which the set $\mathcal{U}$ is a polytope

$$\mathcal{U} = \mathcal{P}[R, p]$$

There are many different approaches to compute a region of attraction. Again, one has to choose between conservative but efficient approaches on one side and efficient but computationally intensive techniques on the other.

The starting point is the next theorem, which basically shows that it is possible to approximate any domain of attraction by a polyhedral domain of attraction. This fact is well known in the literature and due to previous work [Las93, GC86b, GC87, KG87, BM98, BM96a] (see also [Bla99, HL01] for surveys).

In view of the fact that the reachability analysis boils down to the controllability analysis of the reverse system, in the following the null controllability case only will be dealt with.

Theorem 8.10. *Consider a reachable linear system and let $\mathcal{C}_\infty$ be the controllable set (which includes the origin as an interior point). Consider any finite number of points $\bar{x}_k \in int\{\mathcal{C}_\infty\}$, the interior of $\mathcal{C}_\infty$, $k = 1, 2, \ldots, r$. Then there exists a polyhedron $\mathcal{P}$ such that*

i) $\bar{x}_k \in \mathcal{P}$ for all k;
ii) $\mathcal{P}$ is controlled invariant;
iii) there exist a feedback controller $u = \Phi(x)$, a polyhedral Lyapunov function $\Psi(x)$ (the Minkowski function of $\mathcal{P}$) and $\lambda < 1$ (respectively, $\beta > 0$) such that

$$\Psi(x(t+1)) \leq \lambda \Psi(x(t)) \quad (\textit{respectively,} \quad D^+\Psi(x(t)) \leq -\beta \Psi(x(t))$$

and $u(t) \in \mathcal{U}$ for all initial conditions in $\mathcal{P}$.

Proof. The proof, inspired by that in [BMM95], is reported in the discrete-time case only, whereas the continuous-time one will just be sketched.

Assume without restriction that the convex hull of the assigned points $\mathcal{V}[\bar{x}_1 \bar{x}_2 \ldots \bar{x}_r]$ includes the origin as an interior point (if this is not the case, then new ones can be added). By definition, there exist control sequences $u_k(t) \in \mathcal{U}$ such that, if $x(0) = \bar{x}_k$, then $x(t) \to 0$. Consider the modified system

$$x(t+1) = \frac{A}{\lambda} x(t) + \frac{B}{\lambda} u(t)$$

and denote by $x_k^\lambda(t)$ the trajectory associated with the initial condition $\bar{x}_k$ and input $u_k(t)$. Note that the trajectories computed for $\lambda = 1$, $x_k^1(t)$, are those of the original system. Since all the $x_k^1(t)$ converge to zero, for all ϵ such that the ball of radius ϵ, $\mathcal{B}_\epsilon$, is in the interior of $\mathcal{V}[\bar{x}_1 \bar{x}_2 \ldots \bar{x}_r]$, there exists T such that $\|x_k^1(T)\| \leq \epsilon/2$. For λ sufficiently close to 1, by continuity arguments one has that $\|x_k^\lambda(T)\| \leq \epsilon$. Let $\mathcal{P}$ be the polytope achieved as convex hull of all the points of the form $x_k^\lambda(t)$

$$\mathcal{P} = \mathcal{V}\left[x_1^\lambda(0)\ x_2^\lambda(0) \ldots x_r^\lambda(0)\ x_1^\lambda(1)\ x_2^\lambda(1) \ldots\right]$$

Since $\mathcal{P}$ includes the points $\bar{x}_k$, condition i) is satisfied.

Showing that this set is also controlled invariant for the modified system is straightforward, since for each vertex $x_k^\lambda(t)$ the next state $x_k^\lambda(t+1)$ is either a vertex or a point inside $\mathcal{P}$. More precisely,

$$x_k^\lambda(t+1) = \frac{A}{\lambda}x_k^\lambda(t) + \frac{B}{\lambda}u_k(t) \in \mathcal{P}$$

that is

$$Ax_k^\lambda(t) + Bu_k(t) \in \lambda\mathcal{P}$$

Therefore any vertex can be associated with an admissible control $u_k(t) \in \mathcal{U}$ and, according to Theorem 4.44 and Corollary 4.46, $\mathcal{P}$ is λ-contractive.

Finally, statement iii) follows by the fact that it is always possible to associate a control of the form (4.39) with $\mathcal{P}$.

Let us consider the continuous-time case. Denote by $x_k(t)$ the solution of the system when the initial state is $\bar{x}_k$ associated with the nominal control $u_k(t)$ (which is assumed to be continuous). Since $u(t) \in \mathcal{U}$, $\dot{x}$ is bounded and therefore it is possible to approximate the continuous-time solution with the solution of the EAS

$$x^\tau(k\tau+\tau) = [I+\tau A]x^\tau(k\tau) + \tau Bu(k\tau)$$

For T large enough, the solutions $x_k^\tau(T)$ are inside the interior of the ball $\mathcal{B}_{\epsilon/2}$ of radius $\epsilon/2$. Take ϵ such that $\mathcal{B}_\epsilon$ is included in the interior of the set $\mathcal{V}[\bar{x}_1\bar{x}_2\ldots\bar{x}_k]$. Then the proof proceeds as the previous one by considering the modified Euler system

$$x^\tau(k\tau+\tau) = \frac{[I+\tau A]}{\lambda}x^\tau(k\tau) + \frac{\tau}{\lambda}Bu(k\tau)$$

and by determining a polyhedral contractive set for the EAS. This set results to be β contractive for the continuous-time system with $\beta = (1-\lambda)/\tau$. The rest of the proof is identical to the discrete-time case.

The previous theorem can be constructively used to approximate the largest invariant set. Indeed, it basically says that, whenever there exists a family of points in the interior of the domain of attraction, it is always possible to "capture" them (intended as initial states) and drive the state evolution to the origin. In other words, the procedure implicitly suggested by the theorem is the following.

Procedure: Construction of a constrained feedback with a specified initial condition set.

1. Given the points $\bar{x}_k$, check if there exists an open-loop control law which drives the state to the origin in T_k steps under constraints (later, it will be shown how this problem can be solved).

2. If such open-loop control does not exist for some initial point $\bar{x}_k$, and T_k large enough, this means that $\bar{x}_k$ is not included in the controllability region.
3. Once the initial states $\bar{x}_k$ and the corresponding open-loop control laws have been determined, pick a λ close enough to 1 and compute the convex hull of all the possible trajectories of the modified system.

It must be said that the last task can be very hard. Indeed the number of vertices of the region $\mathcal{P}$ grows enormously as the number of points $\bar{x}_k$ is increased and as the points get closer to the boundary of $\mathcal{C}_\infty$, since the closer to the border, the greater the horizon to reach the origin (a problem which has evident bad side effects on the controller realization, which depends on the number of computed vertices). Still, in some problems in which candidate initial conditions are naturally specified, the suggested procedure might turn out to be very useful, since the number of involved vertices is not so high. We will dwell on this later when relatively optimal control laws will be considered in Section 10.4.

So far it has been shown, or at least we have tried to convince the reader, that the choice of a finite number of vertices is reasonable and can lead to the determination of a domain of attraction. Though the above approach has some advantages, it must be said that in principle one might be interested in determining a set which is in some sense the maximal or it approaches the maximal domain of attraction, without fixing a candidate family of initial conditions. This problem can be solved, by means of the procedure provided in Section 5.1 and more specifically in Section 5.1.1, by proceeding as follows:

- fix a "large" set $\mathcal{X}$, typically a polyhedral one (although it is not explicitly required by the problem) and compute (or approximate) the largest invariant (or contractive) set $\mathcal{P}_{max}$ inside $\mathcal{X}$.

Clearly, if $\mathcal{X}$ is big enough, the problem is not practically affected, because the state variables have to be bounded anyway. Note that, in this way, it is possible to solve in a more general form the problem of driving to zero the state from a set of initial conditions, since any state in the derived set can be kept inside the constraints or driven to zero if we may relay on a contractive set.

Example 8.11. As an example of the possibility of determining a good approximation of the set $\mathcal{C}_\infty$, consider the continuous-time system whose dynamic and input matrices are

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 9 & 0 & -1 \\ 0 & 16 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

with constraints $|u| \leq \bar{u} = 1$. This system has eigenvalues $1.4809, -1.7404 \pm 3.0209j$. Consider initial conditions of the form $\bar{x} = [\ \mu\ 0\ 0\ ]^T$ and the following *question*: which is μ_{max}, the largest possible value for $|\mu|$?

In principle, one could solve the open-loop problem of determining a constrained trajectory for an assigned μ and then iterate over $\mu > 0$ by increasing/decreasing the value of μ if the problem is unfeasible/feasible.

To solve the problem in a more general way, a domain of attraction was determined. The fictitious constraints $\|x\|_\infty \leq 1000$ were added and, by using the EAS with $\tau = 0.1$, a contractive region for $\lambda = 0.9$ was computed. Such a discrete-time contractivity amounts to a continuous-time contraction coefficient equal to $\beta = (1-\lambda)/\tau = 1$. The set can be represented as $\|Fx\|_\infty \leq 1$, where F is the matrix

$$\begin{bmatrix} 0 & 0 & 0.0010 \\ 0 & 0.0018 & 0.0009 \\ 0.0018 & 0.0036 & 0.0006 \\ 0.0055 & 0.0052 & 0.0001 \\ 0.0114 & 0.0067 & -0.0005 \\ 0.0193 & 0.0078 & -0.0012 \\ 0.0294 & 0.0088 & -0.0019 \\ 0.0415 & 0.0097 & -0.0027 \\ 58.4180 & 9.6123 & -2.7614 \end{bmatrix},$$

and is depicted in Figure 8.2. A linear-variable structure control can be derived as previously explained (see Section 4.5.1), though it will be shown in a while how this example can be more conveniently managed.

The reply to our initial question is then $\mu_{max} = 1/58.4180$.

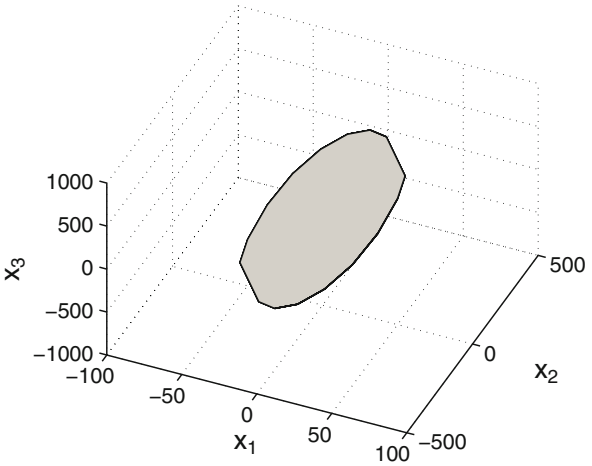


Fig. 8.2 The "slice of ham" contractive set for the third-order system in Example 8.11

The previous example indeed has some peculiarities, basically its controllability set is a sort of slice. Such a behavior is clearly due to the (single) unstable eigenvalue which, as it is rather well known, has a bad role in constrained controllability. This aspect is worth a further investigation and in the next section it will be shown how to determine a domain of attraction by working only in the unstable subspace.

8.1.2 The stable–unstable decomposition

To provide constructive results in a simple way it is better to use a decomposition which is obtained by incorporating the marginally stable sub-system in the unstable part. By facing the unstable part only we achieve, in most cases, for a significant reduction of complexity.

Let us consider the system decomposed in its asymptotically stable part (i.e., the dynamics associated with the eigenvalues with strictly negative real part) and the marginally stable and unstable part, thus grouping the imaginary axis eigenvalues with the open right-half-plane ones:

$$\begin{bmatrix} \dot{x}^L \\ \dot{x}^{\bar{R}} \end{bmatrix} = \begin{bmatrix} A^L & \\ 0 & A^{\bar{R}} \end{bmatrix} \begin{bmatrix} x^L \\ x^{\bar{R}} \end{bmatrix} + \begin{bmatrix} B^L \\ B^{\bar{R}} \end{bmatrix} v(t), \tag{8.5}$$

The following corollary of Theorem 8.8 holds.

Corollary 8.12. *Consider the system decomposition (8.5). Let $\mathcal{C}_\infty^R$ be the controllable set under constraints for the system $(A^{\bar{R}}, B^{\bar{R}})$ and let $\mathcal{T}^L$ be the stable subspace. Then*

$$\mathcal{C}_\infty = \mathcal{T}^L + \mathcal{C}_\infty^{\bar{R}}$$

Proof. The proof is similar, though actually much simpler, to that of Theorem 8.8, and it is just sketched. To drive any state of $\mathcal{C}_\infty$ to 0 one needs just to drive $x^{\bar{R}}(t)$ to 0 in an interval $[0, T_1]$ by means of a proper control action and then sit-and-wait during an interval$[T_1, T_2]$ with $u(t) = 0$ so that $x^L(t)$ comes sufficiently close to 0, from where it can be brought to the origin in finite time.

The statement of the corollary is quite important: when controllability sets have to be computed, one can simply ignore the asymptotically stable sub-system and face the unstable one alone. In other words, one can construct a control $u = \Phi(x^{\bar{R}}(t))$ along with a proper controlled invariant set $\mathcal{S}^{\bar{R}}$ for the unstable sub-system so as to achieve

$$\dot{x}^{\bar{R}}(t) = A^{\bar{R}} x^{\bar{R}}(t) + B^{\bar{R}} \Phi(x^{\bar{R}}(t))$$

This control can be continuous and, as it has been shown, such that $u(t) \to 0$ as $x^{\bar{R}}(t) \to 0$. By applying this feedback, the problem is solved because $x^L(t) \to 0$ spontaneously.

Thus a simplified way to solve the problem is the following:

- decompose the system in the form (8.5);
- construct a controller and a domain of attraction for the unstable sub-system.

This procedure is typically useful when, as it is often the case, the system admits few unstable eigenvalues and the work space can then be reduced. It is worth pointing out that the procedure can be extended by decomposing a system in a fast and a slow sub-system as follows:

$$\begin{bmatrix} \dot{x}^F \\ \dot{x}^S \end{bmatrix} = \begin{bmatrix} A^F & \\ 0 & A^S \end{bmatrix} \begin{bmatrix} x^F \\ x^S \end{bmatrix} + \begin{bmatrix} B^F \\ B^S \end{bmatrix} u(t), \tag{8.6}$$

where the eigenvalues of A^F are on the left of the vertical line $Re(s) = -\beta$ (resp. inside the λ-disk) and those of A^S on the right (resp. outside the λ-disk), where β (resp. λ) is a given convergence speed requirement.

8.1.3 Systems with one or two unstable eigenvalues

An interesting case worthy of investigation is that of systems with one or two unstable eigenvalues. The case of a single-input system with a single unstable eigenvalue will be considered first and, to keep things as general as possible, the constrained stabilizability problem will be replaced by the problem of driving the state to the origin with a certain "speed of convergence."

Proposition 8.13. *Let (A, B) be a reachable continuous-time (discrete-time) system with scalar control ($m = 1$) and let the input u be symmetrically constrained as $|u| \leq \bar{u}$. Let $\beta > 0$ ($0 < \lambda < 1$) be fixed. If there exists a single eigenvalue whose real part is larger or equal to β (a single eigenvalue whose modulus is larger or equal to λ), then the largest β-contractive (λ-contractive) set is a strip of the form*

$$\mathcal{P} = \{x : |fx| \leq \bar{u}\}$$

which can be associated with a linear controller[3].

Proof. The proof is provided in the continuous-time case only, since the discrete-time one is almost identical. Consider the decomposition (8.5)

$$\begin{bmatrix} \dot{z} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} A_z & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix} + \begin{bmatrix} B_z \\ \eta \end{bmatrix} u(t), \tag{8.7}$$

[3] With a slight abuse of notation, since we are not dealing here with bounded sets; we remind that the Minkowski function can be defined for closed and convex sets including 0 as an interior point; so $\mathcal{P}$ is contractive if $x(0) \in \mathcal{P}$ implies $x(t) \in e^{-\beta t}\mathcal{P}$.

and extract the last equation

$$\dot{y}(t) = \rho y(t) + \eta u(t)$$

By symmetry, a contractive set for this system is an interval of the form

$$|y| \leq \bar{y}$$

associated with the Minkowski function

$$\Psi(x) = \left|\frac{y}{\bar{y}}\right|$$

For positive values of y, its derivative is

$$D^{+}\Psi(y) = \frac{\rho y + \eta u}{\bar{y}}$$

which achieves its maximum, w.r.t. to the state variable, at the boundary ($y = \bar{y}$) where it is equal to $D^{+}\Psi_{max} = \frac{\rho\bar{y}+\eta u}{\bar{y}}$. Such a quantity is minimized by $u = -sign(\eta)\bar{u}$ and thus, imposing the derivative to be smaller than $-\beta$,

$$\frac{\rho\bar{y} - |\eta|\bar{u}}{\bar{y}} \leq -\beta$$

it turns out that the maximal set of y which can be driven to 0 with speed at least β is

$$-\bar{y} \leq y \leq \bar{y}$$

where

$$\bar{y} = \frac{|\eta|}{\rho + \beta}\bar{u}$$

It is immediate that the above set is β-contractive with the linear controller

$$u = -\frac{\bar{u}}{\bar{y}}y$$

and, since y is a linear function of the system original state variable, say $y = v^{T}x$ for some $v \in \mathbb{R}^{n\times 1}$, the strip

$$-\bar{u} \leq \underbrace{\frac{\rho + \beta}{|\eta|}v^{T}}_{f^{T}} x \leq \bar{u}$$

is contractive and can be associated with the linear control law

$$u = -f^T x$$

Remark 8.14. It is not difficult to see that convergence can be improved by adopting a saturated control law instead of a linear one. Assume $\eta > 0$ and $\bar{u} = 1$ (without restriction, because one can always discharge the value of $\bar{u} \neq 0$ on B as follows: $Bu = (B\bar{u})(u/\bar{u}) = (B\bar{u})u'$) and consider a control of the form

$$u(t) = -\text{sat}(\kappa y)$$

with $\kappa \geq \frac{1}{\bar{y}}$. By using the above, convergence is still assured for all $|y| \leq \bar{y}$ since $\text{sat}(\kappa y) \geq \frac{1}{\bar{y}} y$ as long as $y \leq |\bar{y}|$. For positive values of y, the derivative is

$$\dot{y} = \rho y - \eta \text{sat}(\kappa y) \leq \rho y - \eta \frac{1}{\bar{y}} y$$

therefore smaller than the derivative achieved by the linear controller. Once again the remaining dynamics, that associated with $z(t)$, is not affected by the proposed control law, say it does not necessarily converge faster.

Example 8.15. Reconsider the dynamic system in Example 8.11 to which, since there is only one unstable eigenvalue, it is possible to apply the transformation

$$T = \begin{bmatrix} 0.0438 & -0.0305 & 0.1421 \\ 0.0159 & 0.1852 & 0.2104 \\ 0.9811 & 0 & 0.9672 \end{bmatrix} \quad T^{-1} = \begin{bmatrix} -7.5257 & -1.2383 & 1.3750 \\ -8.0251 & 4.0779 & 0.2918 \\ 7.6338 & 1.2561 & -0.3609 \end{bmatrix}$$

to get the new representation

$$\hat{A} = T^{-1}AT = \begin{bmatrix} -1.7404 & 3.0209 & 0 \\ -3.0209 & -1.7404 & 0 \\ 0 & 0 & 1.4809 \end{bmatrix}, \quad \hat{B} = T^{-1}B = \begin{bmatrix} 1.3750 \\ 0.2918 \\ -0.3609 \end{bmatrix}$$

The unstable system scalar equation is

$$\dot{z}_3 = 1.4809 z_3 - 0.3609 u$$

and then, according to the just introduced results, the largest β-contractive set with $\beta = 1$ is $[-\bar{z}_3, \ \bar{z}_3]$, where

$$\bar{z}_3 = \frac{|\eta|}{\rho + \beta} \bar{u} = \frac{0.3609}{1.4809 + 1} = 0.1455$$

The linear control law associated with such set is

$$u = -\frac{\bar{u}}{\bar{z}_3} z_3 = \hat{\kappa} z_3 = -6.874 z_3$$

(note once again that this control law is a maximal effort one, say $u = 1$ when $z_3 = \bar{z}_3$). Note also that $u = \begin{bmatrix} 0 & 0 & \hat{\kappa} \end{bmatrix} \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}^T = \hat{\kappa} z_3$ does not affect the remaining eigenvalues which are stable. Indeed, the closed-loop system matrix is

$$\hat{A}_{cl} = \begin{bmatrix} -1.7404 & 3.0209 & 9.4532 \\ -3.0209 & -1.7404 & 2.0060 \\ 0.0000 & -0.0000 & -1 \end{bmatrix}$$

In the original coordinates the control law is

$$u = f^T x = \begin{bmatrix} 0 & 0 & \hat{\kappa} \end{bmatrix} T^{-1} x = 52.4829 x_1 + 8.6357 x_2 - 2.4809 x_3$$

The associated contractive region $|f^T x| \leq 1$ is an unbounded strip oriented as the slice of ham in Figure 8.2 (which is actually a part of such a strip).

If there is more than one input, the extension of the solution is straightforward since it is possible to consider a decomposition as the one previously introduced so that the last equation becomes

$$\dot{y}(t) = \rho y(t) + \sum_{j=1}^{m} \eta_j u_j(t) = \rho y(t) + v(t)$$

If symmetrical constraints (the non-symmetrical case is a trivial extension) are assumed, say

$$|u_j| \leq \bar{u}_j$$

the "new" control variable

$$v(t) \doteq \sum_{j=1}^{m} \eta_j u_j(t) \tag{8.8}$$

is subject to the constraint

$$|v| \leq \sum_{j=1}^{m} |\eta_j| \bar{u}_j$$

so the previous arguments apply without modifications.

A further interesting case is the one in which there are two unstable eigenvalues. This situation has been accurately discussed in [HL01] and here a different presentation, in the case of single-input systems only, will be provided.

If the reachable system has only two "β-unstable" (say has two eigenvalues with real part greater than $-\beta$) modes, one can always compute a decomposition of the form

$$\begin{aligned}\dot{z}(t) &= A_z z(t) + B_z u(t)\\ \dot{y}(t) &= A_y y(t) + B_y u(t)\end{aligned}$$

where the eigenvalues of A_z have real part smaller than $-\beta$, $A_y \in \mathbb{R}^{2\times 2}$ has the two eigenvalues (λ_1 and λ_2) with real part greater than $-\beta$, and the input u is assumed to be bounded as $|u| \leq \bar{u}$. In view of the results of the previous section, it is possible to disregard the $z(t)$ variable and focus on the "y part" alone, thus, without restriction, assume $A_y = diag\{\lambda_1, \lambda_2\}$ and $B_y = [\,1 \;\; 1\,]^T$.

Assume that one is interested in assuring a convergence speed β for the overall system and consider first the case in which the eigenvalues of A_y, λ_1 and λ_2, are real and distinct. Finding a β-contractive set is equivalent to the determination of a controlled invariant set for the y sub-system

$$\dot{y}(t) = [\beta I + A_y]\, y(t) + B_y u(t)$$

say

$$\begin{aligned}\dot{y}_1(t) &= (\beta + \lambda_1) y_1(t) + u(t)\\ \dot{y}_2(t) &= (\beta + \lambda_2) y_2(t) + u(t)\end{aligned}$$

with $|u| \leq \bar{u}$. For such a system, the largest controlled-invariant set can be computed as the 0 reachable set of the reverse system

$$\dot{y}(t) = -A_y y(t) - B_y u(t)$$

which is the convex set confined between the extremal trajectories achieved by keeping the control at its extreme values and originating from the two extreme equilibrium points:

$$\bar{y}^+ = \begin{bmatrix} \frac{\bar{u}}{\beta+\lambda_1} \\ \frac{\bar{u}}{\beta+\lambda_2} \end{bmatrix} \quad \text{and} \quad \bar{y}^- = \begin{bmatrix} -\frac{\bar{u}}{\beta+\lambda_1} \\ -\frac{\bar{u}}{\beta+\lambda_1} \end{bmatrix}$$

Such trajectories admit analytic expressions; more precisely, the upper one is

$$y(t) = e^{-A_y t}\bar{y}^- + \bar{u}\int_0^t e^{-A_y t} B_y \, dt = \begin{bmatrix} \left[-2e^{-(\beta+\lambda_1)t} + 1\right] \frac{\bar{u}}{\beta+\lambda_1} \\ \left[-2e^{-(\beta+\lambda_2)t} + 1\right] \frac{\bar{u}}{\beta+\lambda_2} \end{bmatrix}$$

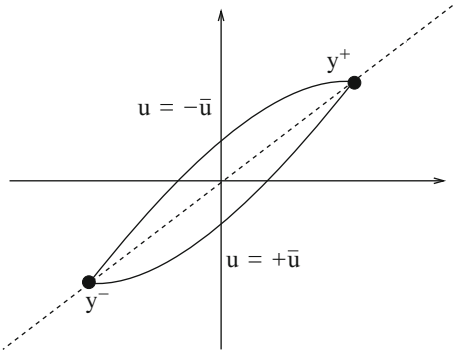


Fig. 8.3 The contractive set for the second-order system (real eigenvalues)

whereas the lower one is symmetric. Note that the found set has non-empty interior because (Fig. 8.3), by the assumed reachability, $\lambda_1 \neq \lambda_2$. Note also that the set is divided into two parts by the line passing through the points $\bar{y}^-$ and $\bar{y}^+$. Below and above this line the control action is $u = \bar{u}$ and $u = -\bar{u}$, respectively. If one applies this control to the modified system, it turns out that the identified set satisfies Nagumo's condition.

These heuristic considerations show that the considered set is controlled- invariant. To make these arguments more precise, one can observe that, if the system with the mentioned control strategy is initialized at any point of the boundary, it remains on the boundary by construction, thus $D^+\Psi(y) = 0$, where $\Psi(y)$ is the Minkowski function. Clearly, if the same strategy is applied to the original system, the following inequality is obtained on the boundary

$$D^+\Psi(y) \leq -\beta\Psi(y)$$

say the set is β-contractive with the proposed switching strategy. If one is not satisfied with a discontinuous switching strategy, then it can be shown that it is possible to derive locally Lipschitz controllers by considering an internal approximating polytope (possibly "smoothed") [BM96a, BM98], but losing (an arbitrarily small) part the region. The reader is referred to [HL01] for a more detailed exposition.

In the case of complex eigenvalues the situation is similar and, even in this case, it is possible to derive a set as a proper limit cycle.

Rather than following the above route, here a different argument will be proposed, and precisely the approach presented in [BM98].

It is known that the domain of attraction of the planar system is convex and can be arbitrarily closely approximated by a polyhedral set of the form $\{y : \|Fy\|_\infty \leq 1\}$ and hence by a "smooth" set of the form

$$\{y : \|Fy\|_{2p} \leq 1\}$$

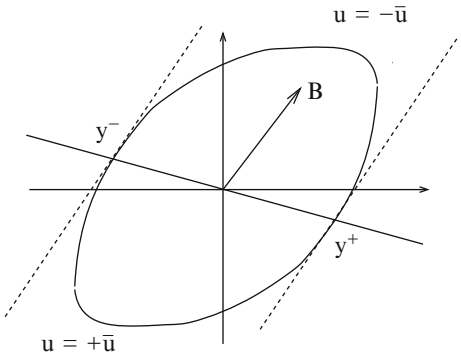


Fig. 8.4 The contractive set for the second-order system (complex eigenvalues)

Then, up to an arbitrarily good approximation, one can assume that the domain of attraction is a smooth set of the form represented in Figure 8.4. According to the results presented in Subsection 4.5.3, the gradient of the function $\|Fy\|_{2p}$ is

$$\begin{aligned}\nabla\|Fy\|_{2p} &= \left(\sum_{i=1}^{s}(F_i y)^{2p}\right)^{\frac{1}{2p}-1}\sum_{i=1}^{s}(F_i y)^{2p-1}F_i^T \\ &= \left(\|Fy\|_{2p}\right)^{1-2p}F^T G_p(y),\ y\neq 0\end{aligned}$$

where $G_p(y)$ denotes the vector

$$G_p(y) = [(F_1 y)^{2p-1}\ (F_2 y)^{2p-1}\ \ldots\ (F_s y)^{2p-1}]^T.$$

The explicit expression of the gradient allows the adoption of a gradient based control

$$u(y) = -\gamma(y)\ \left(\|Fy\|_{2p}\right)^{1-2p}\ B^T F^T G_p(y)$$

which works for $\gamma(y)$ positive and large enough. Now since the control is bounded, the largest value of $\gamma(y)$ can be chosen in such a way that u is on the boundary. In the scalar case, the resulting control is a bang–bang one:

$$u_{bb}(y) = -\bar{u}\ \mathrm{sgn}[B^T F^T G_p(y)]$$

It is immediate that this control is the one which minimizes the derivative of $\|Fy\|_{2p}$, say is the minimizer of the following problem

$$\begin{aligned}\min_{|u|\leq\bar{u}} D^+\|Fy\|_{2p} &= \min_{|u|\leq\bar{u}}\ \left(\|Fy\|_{2p}\right)^{1-2p}\ G_p^T(y)F\left(Ay+Bu\right) \\ &= \left(\|Fy\|_{2p}\right)^{1-2p}\ G_p^T(y)F\left(Ay+Bu_{bb}(y)\right)\end{aligned}$$

Note now that in the two-dimensional case there are only two points on the boundary of the set $\mathcal{N}[\|Fy\|_{2p}, 1]$ in which the quantity $B^T F^T G_p(y)$ vanishes. These two points have been denoted by y^- and y^+ in Figure 8.4. It is clear that the line passing through these points is a discriminating line. On one side of this line one has to apply $+\bar{u}$, while $-\bar{u}$ has to be applied on the other side. If one assumes that this line has equation $fy = 0$, then this control can be equivalently written as

$$u_{bb}(y) = -\bar{u}\,\mathrm{sgn}[fy]$$

(this equivalence does not hold for systems of greater dimensions in general). Finally, note that this control can be approximated by a saturated controller

$$u_{\mathrm{sat}}(y) = -\bar{u}\,\mathrm{sat}\,[\kappa fy]$$

for $\kappa > 0$ large enough (see [BM98]).

To recap, in the case of two unstable eigenvalues only (or two eigenvalues slower than β), it turns out that it is always possible to use a linear-saturated control law, virtually without loss in the size of the domain of attraction. Such a result is in perfect agreement with the limit cycle approach presented in [HL01], though the arguments used here provide an alternative way to determine the line $fy = 0$.

The discrete-time version of the problem has a straightforward extension in the single-unstable-pole case. Indeed, one can apply the same transformation and reduce the problem to the scalar case. It is not clear how to extend the exposed theory in the case of two unstable poles. Indeed the bang–bang control does not work for discrete-time systems (not even for scalar input systems). However the control at the vertices, namely the piecewise-linear controller described in Subsection 4.5.2, is reasonably simple in the two-dimensional case.

8.1.4 *Region with bounded complexity for constrained input control*

As repeatedly mentioned, an alternative and simpler way to compute a domain of attraction is to resort to ellipsoids along with an associated linear controller. This is certainly a simplified solution since the resulting region is not maximal. However the complexity of the compensator is low, no matter how many unstable poles are present. The present discussion is a brief summary of what is extensively treated in [BEGFB04].

Consider the problem of finding an ellipsoid $\mathcal{E}[P, 1] = \mathcal{E}[Q^{-1}, 1]$ (i.e., of all x such that $x^T Px = x^T Q^{-1} x \leq 1$) which

- is β-contractive and associated with a linear control $u = Kx$;
- guarantees that for all $x \in \mathcal{E}$ the control is bounded as $\|u\|_2 \leq \mu$;
- includes a certain number of assigned vectors $\bar{x}_k$.

Note that removing any of the three requirements would render the problem trivial. The contractivity requirement reduces to the conditions derived in Section 4.4.1 for the positive definite matrix Q. The corresponding controller is $u = Kx = RQ^{-1}x$. The constraint $\|u\|_2 \leq \mu$ can be dealt with by defining $z = Q^{-1/2}x$ and by requiring that

$$\begin{aligned}\max_{x\in\mathcal{E}} \|u\|_2 &= \max_{x^TQ^{-1}x\leq 1} \|RQ^{-1}x\|_2 = \\ &= \max_{z^Tz\leq 1} \|RQ^{-1/2}z\|_2 = \|RQ^{-1/2}\|_2 \leq \mu\end{aligned}$$

The above constraint can be written as

$$Q^{-1/2}R^TRQ^{-1/2} \preceq \mu^2 I \tag{8.9}$$

Finally, requiring the states $\bar{x}_k$ to be included in the set $\mathcal{E}$ can be stated as

$$\bar{x}_k^T Q^{-1} \bar{x}_k \leq 1 \tag{8.10}$$

Putting the condition for the β-contractivity of $\mathcal{E}$ along with (8.9) and (8.10) results in the following set of LMI conditions

$$\begin{gathered} AQ + QA^T + BR + R^TB^T + 2\beta Q \preceq 0, \\ \begin{bmatrix} Q & R^T \\ R & \mu^2 I \end{bmatrix} \succeq 0 \\ \begin{bmatrix} 1 & \bar{x}_k \\ x_k^T & Q \end{bmatrix} \succeq 0 \end{gathered} \tag{8.11}$$

which can be solved by existing efficient software tools [BEGFB04]. The last set of conditions in (8.11) can be replaced by

$$Q \succeq I$$

if one wishes to impose that $\mathcal{E}$ includes the unit ball $x^Tx \leq 1$.

Similar conditions can be derived if one considers constraints of the form $|u_i| \leq \mu$ (mind that this is not restrictive since it is always possible to achieve the same bound for all components by possibly scaling the entries in B), so the constraints are expressed in terms of the ∞-norm $\|u\|_\infty \leq \mu$ (see [BEGFB04] for details). In this case one needs to introduce a new variable X and consider new constraints. Precisely, the condition of invariance and admissibility for the ellipsoid is now

$$\begin{bmatrix} X & R \\ R^T & Q \end{bmatrix} \succeq 0, \quad X_{ii} \leq \mu^2.$$

The discrete-time case admits an almost identical expressions with the understanding that the basic contractivity condition is now different, as shown in Subsection 4.4.2.

A quite different approach for constrained control proposed in the late 80s and intensively studied in the literature is based on fixed-complexity polytopes and linear control laws. The problem can be formulated as follows. Given a linear system with a control subject to the constraint $u \in \mathcal{U}$, find a control $u = Kx$ such that

- $A + BK$ is asymptotically stable;
- the admissibility set

$$\mathcal{A}(K) = \{x : \ Kx \in \mathcal{U}\}$$

is positively invariant.

Consider the case in which $\|u\|_\infty \leq 1$, say $-1 \leq u_k \leq 1$, for all k, so that the candidate admissible set is the polyhedron

$$\mathcal{A}(K) = \bar{\mathcal{P}}[K, \bar{1}]$$

The following result holds.

Theorem 8.16. *Consider the n dimensional linear system*

$$\dot{x} = Ax + Bu$$

and assume the input matrix $B \in \mathbb{R}^{n\times m}$ has full column rank. Then there exists a solution to the mentioned problem if and only if (A, B) is stabilizable and A has no more than m (the number of inputs) unstable eigenvalues.

Proof. The proof of the theorem can be found in [BV90, BV95] based on an eigenstructure assignment. Here a sketch of the proof (actually an "explanation of the result") in an algebraic form is provided.

It is first shown that the condition is generically necessary. Assume a solution K exists and define the variable $y = Kx \in \mathbb{R}^m$. Assume also that $KB \doteq B_y$ is square and invertible (so that K must have full row rank: this is a condition which can always be met by possibly applying an input transformation and discharging some columns of B and K). Let D be an $n \times (n - m)$ orthonormal matrix whose columns span the null space of the matrix B^T, say

$$D^T D = I_{n-m}, \quad B^T D = 0_{m\times(n-m)}$$

Set $J = D^T$ and consider the matrix

$$\begin{bmatrix} K \\ J \end{bmatrix}$$

which is invertible since

$$\begin{bmatrix} K \\ J \end{bmatrix} \begin{bmatrix} B & D \end{bmatrix} = \begin{bmatrix} B_y & KD \\ 0 & I_{n-m} \end{bmatrix}$$

has rank n. Define the variable $z = Jx \in \mathbb{R}^{n-m}$ and consider the transformation

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} K \\ J \end{bmatrix} x$$

to get

$$\begin{bmatrix} \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} A_{yy} & A_{yz} \\ A_{zy} & A_{zz} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} + \begin{bmatrix} B_y \\ 0 \end{bmatrix} u(t)$$

The feedback gain matrix in the new representation is related to K as

$$K = [K_y \; K_z] \begin{bmatrix} K \\ J \end{bmatrix} = [I \; 0] \begin{bmatrix} K \\ J \end{bmatrix}$$

This means that the transformed closed-loop system is

$$\hat{A}_{cl} = \begin{bmatrix} A_{yy} & A_{yz} \\ A_{zy} & A_{zz} \end{bmatrix} + \begin{bmatrix} B_y \\ 0 \end{bmatrix} \begin{bmatrix} I & 0 \end{bmatrix}$$

The positive invariance of the set $\bar{\mathcal{P}}[K, \bar{1}]$ can then be stated in terms of the new variables as the invariance of the set

$$\{(y, z) : \; \|y\|_\infty \leq 1\}$$

The only possibility for the above to hold for every z is that

$$A_{yz} = 0$$

namely, the z-subspace is A-invariant. Indeed, since for $\|y\|_\infty = 1$ we must have $D^+\|y\| \leq 0$ for all z, $\dot{y}$ must point inside. If $y_k = 1$ is an active constraint we must have $\dot{y}_k \leq 0$. On the other hand, the invariance condition imposes $\dot{y}_k = [A_{yy}y + A_{yz}z]_k \leq 0$, even for arbitrary (large) $z \neq 0$, which necessarily implies that the kth column of A_{yz} is zero. Repeating the argument for all k, we see that A_{yz} is zero. Since this sub-matrix is not affected by the control gain, the open-loop-system state matrix in the new representation must then be

$$\begin{bmatrix} A_{yy} & 0 \\ A_{zy} & A_{zz} \end{bmatrix}$$

Therefore, if the problem can be solved, then the matrix $A_{zz} \in \mathbb{R}^{(n-m)\times(n-m)}$ whose eigenvalues are a subset of the original ones must be stable. Hence, there can be at most m unstable eigenvalues. To see why the condition is also sufficient, consider the system decomposed in its stable part, associated with variable z and its "possibly unstable part" associated with the variable y, as follows:

$$\begin{bmatrix} \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} A_{yy} & 0 \\ A_{zy} & A_{zz} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} + \begin{bmatrix} B_y \\ B_z \end{bmatrix} u(t)$$

where $y \in \mathbb{R}^m$ and where $A_{zz} \in \mathbb{R}^{(n-m)\times(n-m)}$ includes only stable eigenvalues. Consider a feedback of the form $[K_y \; 0]$ to achieve the closed-loop system

$$\begin{bmatrix} \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} A_{yy} + B_y K_y & 0 \\ A_{zy} + B_z K_y & A_{zz} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \tag{8.12}$$

Then the problem can be solved if one assures the positive invariance of the set $\|K_y y\|_\infty \leq 1$ by means of a proper control action. We prove that this is possible under the simplifying assumption that B_y is invertible. Indeed one can take K_y is such a way that

- $A_{yy} + B_y K_y = -\beta I$, $\beta > 0$, in the continuous-time case;
- $A_{yy} + B_y K_y = \lambda I$, $0 \leq \lambda < 1$, in the discrete-time case.

The closed-loop "y-system" becomes

$$\dot{y}(t) = -\beta y(t), \quad (\text{respectively } \; y(t+1) = \lambda y(t))$$

and the set $\|K_y y\|_\infty$ is invariant (indeed β or λ contractive) with respect to the new dynamics.

It is worth stressing that the proposed "sketch of proof" of the results requires an assumption, basically the matrix B_y being invertible, which is not necessary and which is not present in the original work [BV90, BV95]. Other contributions in this sense can be found in [BH93, Ben94, HB91] These references exploit the property that the invariance of the set $\bar{\mathcal{P}}[K, \bar{1}]$ is equivalent to the following condition

$$KA + KBK = HK \tag{8.13}$$

where H must be such that

$$\|H\|_\infty = \lambda \leq 1$$

in the discrete-time case, and such that

$$\begin{bmatrix} \bar{H} & \underline{H} \\ \underline{H} & \bar{H} \end{bmatrix} \begin{bmatrix} \bar{1} \\ \bar{1} \end{bmatrix} \leq -\beta \begin{bmatrix} \bar{1} \\ \bar{1} \end{bmatrix}$$

in the continuous-time case (see Subsection 4.5.1 for details). Equation (8.13) is well-known in the eigenstructure assignment framework [Ben94, HB91]. Basically, the result can be interpreted in the following way: to enforce the invariance of the set $\bar{\mathcal{P}}[K, \bar{1}]$ one must assure that a certain number of the closed-loop eigenvectors lie in the null space $ker(K)$ and are associated with stable eigenvalues, as it is apparent from Equation (8.12). Interesting extensions to the case of descriptor systems can be found in [GK91a] and [CT95]

8.2 Domain of attraction for input-saturated systems

In the control of input-saturated systems, it is assumed that each input component $u_k(t)$ follows exactly the ideal input signal $v_k(t)$, computed by the controller, as long as this signal is between its lower and upper bounds u_k^- and u_k^+. Below or upper this bounds, the signal is truncated as

$$u = \text{sat}_{[u^-, u^+]}(v)$$

precisely

$$u_k(t) = \begin{cases} u_k^+ \text{ if } v_k > u_k^+ \\ v_k \text{ if } u_k^- \leq v_k \leq u_k^+ \\ u_k^- \text{ if } v_k < u_k^- \end{cases}$$

This is equivalent to the insertion in the loop of the nonlinear block represented in Figure 8.5. The basic difference with respect to the approach previously considered is that the satisfaction of the constraints is not assured by the nominal control law: the bounds are naturally imposed by actuator saturation. We stress that, in many cases, letting the actuators reach their physical bounds could be inappropriate[4]. An intervention of the control software should limit the input signal preventing the reaching of such bounds. In other words, the "saturation" block in Fig. 8.5 should be added to the control. The typical approach is the following:

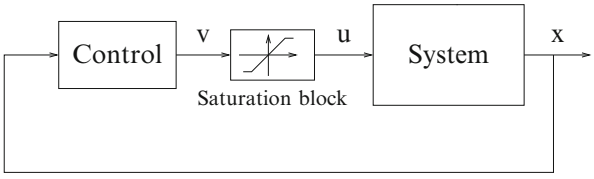


Fig. 8.5 The saturation function

[4] "Criminal" according to some experts.

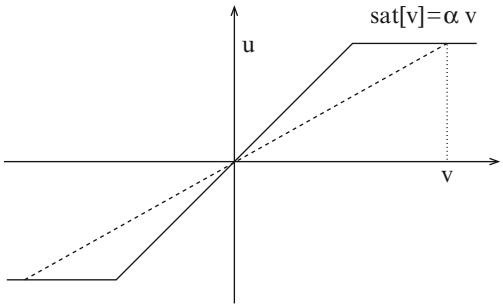


Fig. 8.6 The saturation function

- the control is computed based on local optimality criteria;
- the effect of the saturations is then taken into account.

As we will see, the latter step might require a redesign of the control if the results in terms of domain of attraction are not satisfactory. To simplify the exposition, consider the case in which upper and lower bounds are opposite and their absolute value is equal to one, precisely $-u^- = u^+ = 1$ and write

$$u = \text{sat}[Kx]$$

(the theory can be easily extended to the general unsymmetrical case without special difficulties).

A possible way to evaluate the domain of attraction [GdST01, GdST99, HL01] is to consider the saturation function as a variable gain (see Fig. 8.6). More precisely, in the scalar case, one has

$$u = \text{sat}[v] = \alpha(v)v$$

where $0 \leq \alpha(v) \leq 1$. Clearly, $\alpha(v) = 1$ for v sufficiently close to 0 (the non-saturation region) and $\alpha(v) \to 0$ as v increases. Thus, in the single input case, when a linear control is applied we have

$$\text{sat}[Kx] = \alpha(Kx)Kx$$

If the state excursion is limited by assuming as a constraint a strip of the form

$$\bar{\mathcal{P}}[F, \kappa] = \{x : |Fx| \leq \kappa\}$$

then it is possible to determine a bound for α valid inside $\bar{\mathcal{P}}[F, \kappa]$, precisely, if $\kappa > 1$, then

$$\bar{\alpha} = \frac{1}{\kappa} \leq \alpha(Kx) \leq 1$$

while, if $\kappa \leq 1$, there is no saturation effect as long as x is in the strip $\bar{\mathcal{P}}[F, \kappa]$. This means that the system with a saturated control can be seen as an LPV system

$$\dot{x}(t) = Ax(t) + B\text{sat}[Kx] = Ax(t) + B(w)Kx$$

where $B(w)$ is a segment of matrices

$$B(w) = wB + (1 - w)\bar{\alpha}B, \;\; 0 \leq w \leq 1. \tag{8.14}$$

This argument can be extended in an immediate way to the case of multiple input systems, by considering a polyhedral cylinder of the form

$$\bar{\mathcal{P}}[F, \kappa] = \{x : \|Fx\|_\infty \leq \kappa\}$$

and by assuming the polytopic structure

$$B(w) = \sum_{k=1}^{2^m} B_k w_k, \quad \sum_{k=1}^{2^m} w_k = 1, \; w_k \geq 0, \tag{8.15}$$

where the matrices B_k are all the possible combinations of

$$B[\alpha], \;\; [\alpha] = diag\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$$

by assuming each of the α_k at its extrema

$$\alpha_k \in \{\bar{\alpha}_k, 1\}$$

The problem can then be managed as follows.

Procedure: Computation of a guaranteed region of attraction for a saturated system.

Given the gain K and the amplitude $\kappa > 1$, perform the following steps

1. For each input channel determine the lower bounds for α_k, $\bar{\alpha}_k = \frac{1}{\kappa}$.
2. Check the stability of the system

   $$\dot{x}(t) = [A + B(w(t))K]x(t)$$

   where $B(w)$ is given as in (8.14), by constructing a Lyapunov function $\Psi(x)$. If the system is unstable, then reduce κ (keeping it greater than 1) and restart the procedure.
3. Determine the largest value ρ such that $\mathcal{N}[\Psi, \rho] \in \bar{\mathcal{P}}[F, \kappa]$. Then $\mathcal{N}[\Psi, \rho]$ is an invariant set which is a guaranteed domain of attraction for the input saturated system.

The considered procedure can be used also for synthesis, if one assumes that K is an unknown matrix. In this case, $\kappa > 1$ is fixed and the inequality

$$AQ + QA^T + B_k R + R^T B_k^T \prec 0, \quad Q \succ 0, \quad k = 1, 2, \ldots, m$$

is exploited to provide the controller

$$u = -RPx = -RQ^{-1}x$$

together with the associated domain of attraction, which results to be the largest ellipsoid $\mathcal{E}[P, \rho]$ (in terms of ρ) which is included in the strip $\|RPx\|_\infty \leq \kappa$.

Clearly, instead of considering ellipsoids, one could consider polyhedral sets and norms. One method is to compute the largest invariant set included in the strip $\mathcal{P}[F, \kappa]$ for the associated system using the technique suggested in [BM96b].

Several procedures have been proposed in the literature to face the problem of systems with saturating actuators. The most traditional ones are based on ellipsoidal invariant sets, starting with the seminal work [GH85]. Several works have been done more recently [HL02, HL01, GdST99, NJ99, NJ00, GdST01]. Other classes of functions which perform better than ellipsoids are piecewise-affine functions [Mil02b, Mil02a], composite quadratic functions [HL03, HTZ06], saturation dependent Lyapunov functions [CL03], or function associated with the so-called SNS domain of attraction [ACLC06]. An universal formula to associate a control with a control Lyapunov function has been suggested in [LS95].

The provided approach and all those based on convex Lyapunov functions are computationally effective in providing domains of attraction for saturated systems. Unfortunately they are not always very accurate, as it can be seen in the next example which enlightens the fact that

- a saturated system has a domain of attraction which is, in general, non-convex;
- a convex region has no hope, in general, to be a faithful representation of such a domain.

Example 8.17. Consider the simple system

$$A = \begin{bmatrix} 0 & 1 \\ 1 & -0.1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and the following locally stabilizing control

$$u = -\text{sat}[2x_1]$$

Let us now see what can be found by means of the different techniques, say let us compare ellipsoidal versus polyhedral domains of attraction. Concerning ellipsoids, it turns out that $\kappa_Q = 1.105$, corresponding to $\bar{\alpha}_Q = 0.9050$, is the largest value for which the polytopic system $\dot{x} = [A + B(w)K]x$ with $B(w) = wB + (1 - w)\bar{\alpha}_Q B$ is quadratically stable (such a value can be found by means of the results presented in Theorem 7.14).

For such a value of κ_Q the quadratically stable polytopic system is characterized by the two matrices:

$$A + BK = \begin{bmatrix} 0 & 1 \\ -1 & -0.1000 \end{bmatrix} \text{ and } A + \bar{\alpha}_Q BK = \begin{bmatrix} 0 & 1 \\ -0.8100 & -0.1000 \end{bmatrix}$$

for which the Lyapunov function $x^T P x$ with

$$P = \begin{bmatrix} 85.1833 & 4.7063 \\ 4.7063 & 94.1241 \end{bmatrix}$$

could be found. By considering polyhedral functions, as expected, κ could be increased to $\kappa_P = 1.1628$ (corresponding to $\bar{\alpha}_P = 0.86$). The resulting polytopic system is characterized by the next two pair of matrices

$$A + BK = \begin{bmatrix} 0 & 1 \\ -1 & -0.1 \end{bmatrix} \text{ and } A + \bar{\alpha}_P BK = \begin{bmatrix} 0 & 1 \\ -0.7200 & -0.1000 \end{bmatrix}$$

By discretizing the system via EAS with $\tau = 0.001$, a polyhedral Lyapunov function (with 6628 vertices) was computed imposing $\lambda = 0.99999$ and a tolerance $\epsilon = 0.00001$ for the discrete-time system. In Figure 8.7 the ellipsoidal (thick dashed line) and polyhedral (solid line) domain of attraction are reported, together with the respective strips $Kx \leq \kappa_Q$ and $Kx \leq \kappa_P$. Unfortunately these sets are very conservative, as it can be seen from Fig. 8.8, which pictorially depicts the domain of attraction determined by simulation of the backward system. The main point is that, as it can be seen in Figure 8.8, the actual domain of attraction is non-convex.

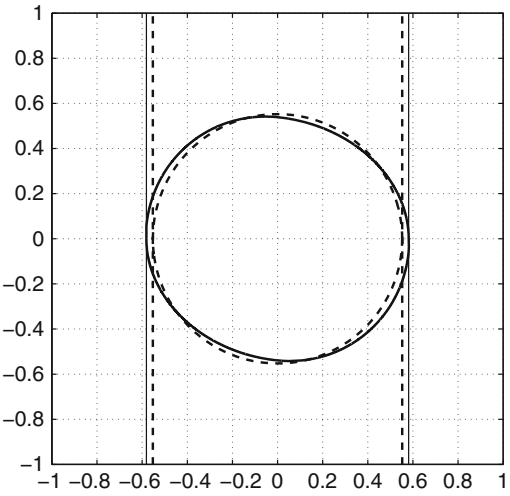


Fig. 8.7 The ellipsoidal (thick dashed line) and polyhedral (solid line) domains of attraction

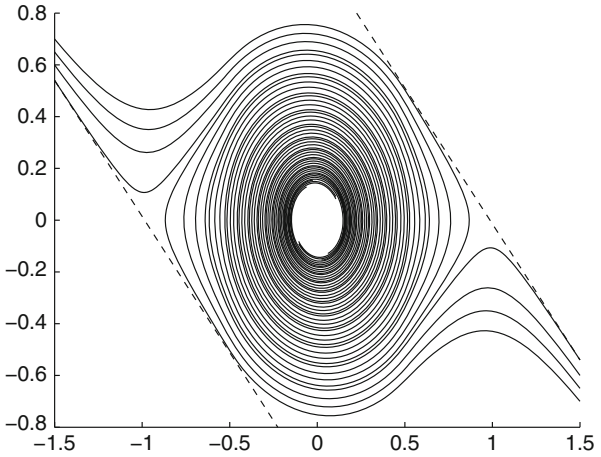


Fig. 8.8 The saturated and the absolute domain of attraction computed via simulation

Therefore by means of the proposed method we cannot aspect accurate results, since we are enforcing convexity in order to have computability.

Note also that the choice of the saturated control $u = \text{sat}[Kx] = \text{sat}[2x_1]$ can lead to a smaller domain of attraction than the largest achievable. In this case the system has a single unstable pole and therefore it is straightforward to compute, according to the results of Subsection 8.1.3, the largest domain of attraction which turns out to be a set of the form

$$|K_* x| = |-1.0140x_1 - 0.9646x_2| \leq 1$$

and can be associated with the linear controller $u = K_* x$ This domain is the one included between the two dashed lines.

8.3 State constraints

In several control problems it is fundamental to include in the design state constraints, possibly together with control constraints. Essentially the same approach described in the case of control constraint can be applied, although the techniques are different when it comes to the details. Given the system

$$\dot{x}(t) = Ax(t) + Bu(t)$$

and the constraints

$$x(t) \in \mathcal{X}$$

where $\mathcal{X}$ is a C-set, the problem consists in finding a stabilizing compensator such that the system state is driven to 0 without constraint violation. The next theorem is an obvious preliminary results.

Theorem 8.18. *The constraint are not violated by means of a proper control action if and only if the initial state belongs to a controlled-invariant set included in $\mathcal{X}$.*

Therefore, no matter which technique is adopted, the existence of a controlled invariant set is a fundamental requirement for the problem to be solved. Basically the computation of a controlled invariant set can be achieved by the methods already presented. Therefore one can simultaneously consider state and control constraints (and even uncertainties) without changing the exposed theory, as shown in the next subsection.

8.3.1 A two-tank hydraulic system

Consider the system shown in the Figure 8.9 and represented with the scheme in Figure 8.10 It is formed by the electric pump EP which supplies water to the two

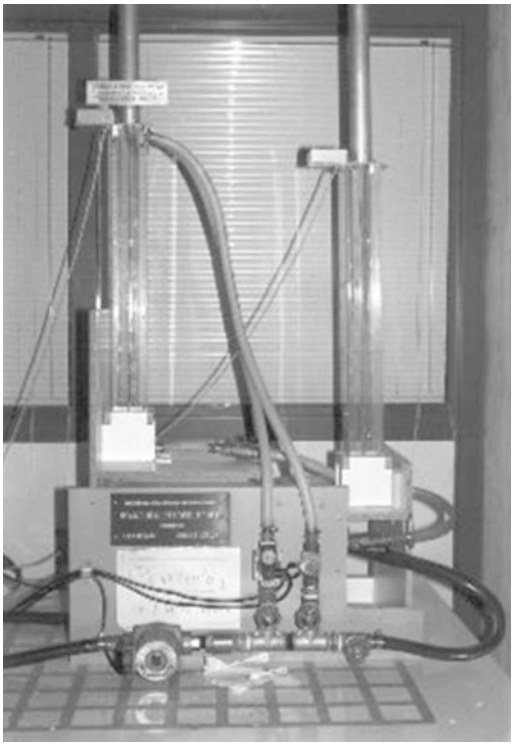


Fig. 8.9 The two tank system

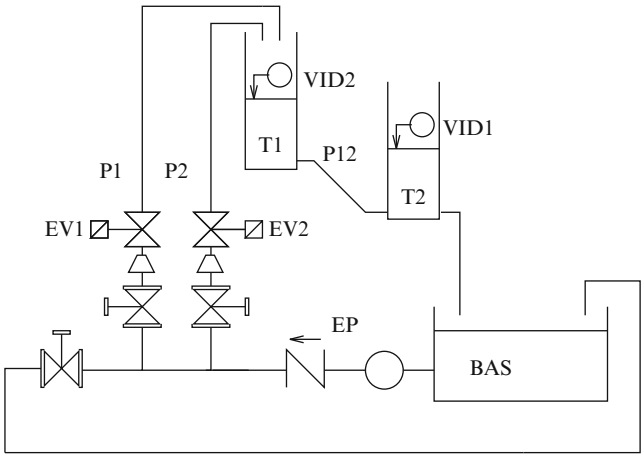


Fig. 8.10 The two tank system scheme

parallel pipes P1 and P2, whose flow can be either 0 or U_{max} and is regulated by two on–off electro-valves EV1 and EV2 which are commanded by the signals coming from the digital board BRD1 (not reported in Figure 8.10). The two parallel pipes bring water to the first tank T1 which is connected, through P12, to an identical tank T2 positioned at a lower level. From T2 the water flows out to the recirculation basin BA. The two identical variable inductance devices VID1 and VID2, together with a demodulating circuit in BRD1, allow the computer to acquire the water levels inside the two tanks. These levels, denoted by h_1 and h_2, are the state variables of the system.

Choosing as linearization point the steady state value $[h_{10} \;\; h_{20}]^T$ corresponding to the constant input $u_0 = U_{max}$ and setting $x_1(t) = h_1(t) - h_{10}(t)$ and $x_2(t) = h_2(t) - h_{20}(t)$, one gets the linearized time-invariant system, whose state and input matrix A and B are

$$A = \begin{bmatrix} -\frac{\alpha}{2\sqrt{h_{10}-h_{20}}} & \frac{\alpha}{2\sqrt{h_{10}-h_{20}}} \\ \frac{\alpha}{2\sqrt{h_{10}-h_{20}}} & -\frac{\alpha}{2\sqrt{h_{10}-h_{20}}} - \frac{\beta}{2\sqrt{h_{20}}} \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The parameters entering the above matrix are $\alpha = 0.08409$, $\beta = 0.04711$, $h_{10} = 0.5274$, $h_{20} = 0.4014$, and $u_0 = 0.02985$. To keep into account the effects due to the nonlinear part of the system, the uncertain model described by

$$A(\xi,\eta) = \begin{bmatrix} -\xi & \xi \\ \xi & -(\xi+\eta) \end{bmatrix} \qquad B(\xi,\eta) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

with $\xi = 0.118 \pm .05$ and $\eta = 0.038 \pm 0.01$ was considered. The state and control constraint sets are, respectively, given by $X = \{[x_1 \; x_2]^T : |x_1| \leq 0.1, |x_2| \leq 0.1\}$ and $U = \{-u_{max}, u_{max}\}$.

A 0.2-contractive region inside X was computed by using the corresponding EAS with $\tau = 1$ and $\lambda = 0.8$. Such a region (the maximal 0.8-contractive for the EAS) turns out to be $P = \{x : \|Fx\|_\infty \leq 1\}$, where

$$F = \begin{bmatrix} 1.000 & 0.000 \\ -0.1299 & -1.727 \\ -0.2842 & -1.871 \\ -0.4429 & -1.932 \\ -0.5833 & -1.905 \\ -0.6903 & -1.806 \\ -0.8258 & -1.671 \\ -0.8716 & -1.557 \\ -0.9236 & -1.414 \\ -0.9295 & -1.317 \end{bmatrix}$$

Matrix F has been ordered in a way such that the ith row of F is associated with the ith sector and, by symmetry, its opposite in Figure 8.11. This region is formed by 20 symmetric sectors and, as it is always the case in two dimensions, these are all simplicial. Hence the piecewise-linear control is characterized by 10 different gains, which are reported in matrix K

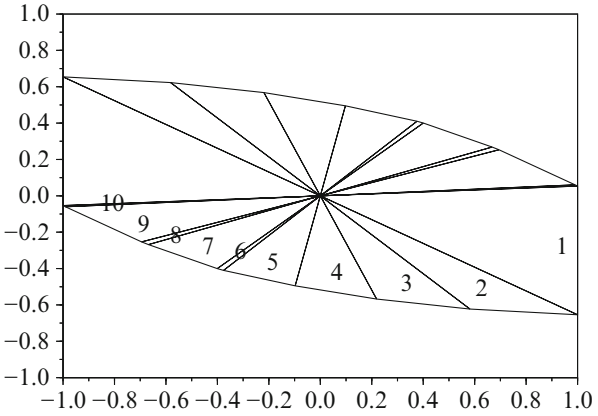


Fig. 8.11 The maximal β-contractive region, with β=0.2

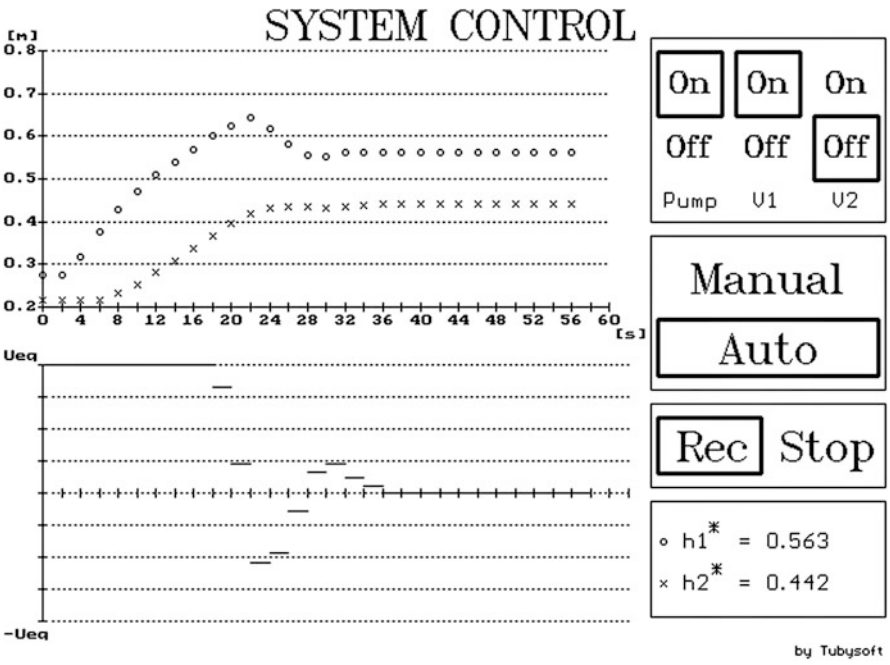


Fig. 8.12 Variable structure control: experimental test

$$K = \begin{bmatrix} -0.2839 & -0.3003 \\ -1.035 & -1.449 \\ -0.0855 & -0.5613 \\ -0.1329 & -0.5796 \\ -0.1750 & -0.5713 \\ -0.2071 & -0.5419 \\ -0.2477 & -0.5012 \\ -0.2614 & -0.4672 \\ -0.2771 & -0.4243 \\ -0.2788 & -0.3964 \end{bmatrix}$$

which, again, is ordered in a way such that the i-th row of K corresponds to the i-th sector of P (i-th row of F). The result of the implementation of the variable structure control law $u(x) = K^{I(x)}x$ is reported in figure 8.12.

We let the reader note that in this simple experiment we didn't force the initial state to belong to the set P. This can be immediately seen from the fact that the control saturates for the first 20 seconds. After this period the system enters and remains inside the region and converges asymptotically to the steady state value (the origin of the linearized system) with the assigned contractivity speed.

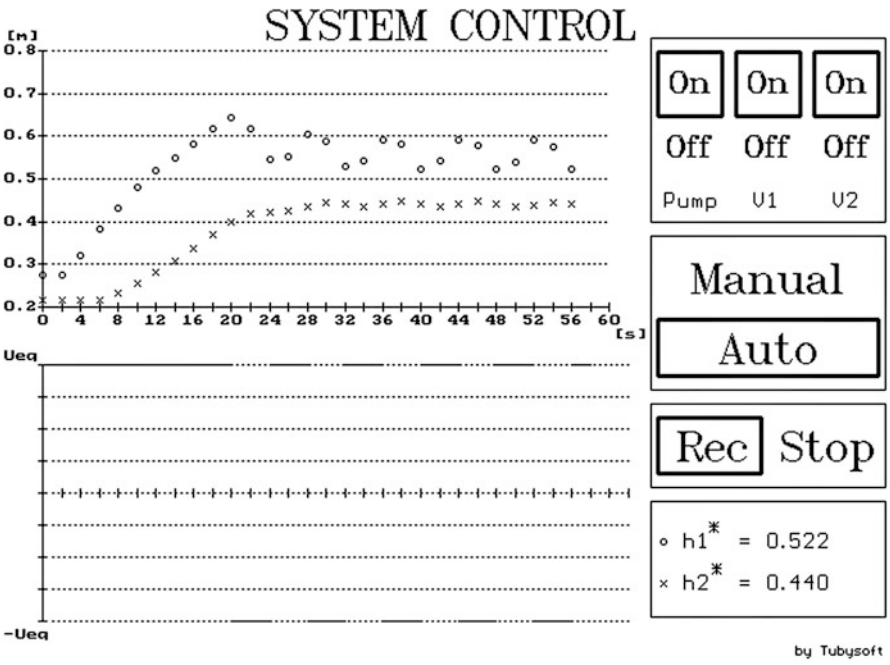


Fig. 8.13 Bang–bang control system evolution

We remind that the piecewise-linear control is just one of the suitable controls. For comparison, a discontinuous control law was also implemented [BM96a]:

$$u = \arg \max_{|u| \le u_{max}, i \in \mathcal{I}(x)} (F_i(Ax + Bu))$$

where $\mathcal{I}(x) = \{i : F_i x = \max_k F_k x\}$ is the maximizer subset of indices. This control is piecewise constant and possibly discontinuous on the sector boundaries. Although this control is stabilizing in theory, one can see from the experimental results in Figure 8.13 that, due to the extremely rough sampling frequency (0.5 Hz), the system exhibits a limit cycle, thus it is not converging to the origin (a typical behavior for sampled systems under bang–bang control).

An interesting heuristic procedure which the authors have seen to produce good results is that of considering a simple linear control law whose gain is obtained by averaging the gains of the just reported variable structure control law. In our case the average gain is given by $k = [\ -0.2984 \quad -0.5792\ ]$ and the maximal β-contractive region, with $\beta = 0.2$, of the closed-loop system included in the non-saturation set $X \bigcap X_U$, where $X_U = \{x : |kx| \le .3\}$, resulted in the internal region in Figure 8.14. This set is smaller than the original one in Figure 8.14; however, its existence assures a speed of convergence $\beta = 0.2$ for the closed-loop system with the obtained linear control. In fact the domain of attraction is greater, being the largest invariant (not contractive) set, and it is the external region in Figure 8.14.

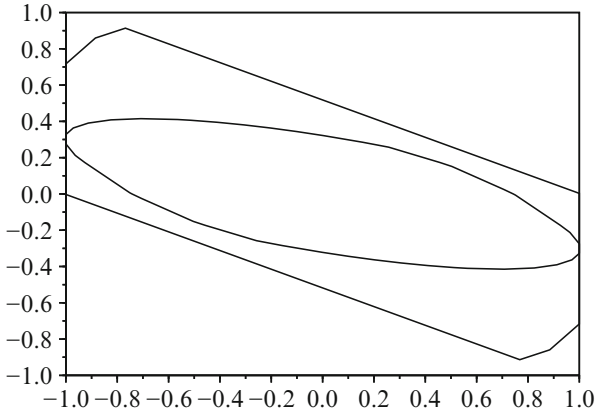


Fig. 8.14 The largest β-contractive and invariant sets with $u = kx$

8.3.2 The boiler model revisited

Let us reconsider the model described in Subsection 6.4.1 and proposed in [USGW82]. There we considered a linear controller and a legitimate question is whether we can do better by a nonlinear control. The system equations are reported here for convenience

$$\dot{x}(t) = Ax(t) + Bu(t) + Ed(t)$$

with

$$A = \begin{bmatrix} -0.0075 & -0.0075 & 0 \\ 0.1086 & -0.149 & 0 \\ 0 & 0.1415 & -0.1887 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0037 \\ 0 \\ 0 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ -0.0538 \\ 0.1187 \end{bmatrix}$$

We have seen that assuming the constraints $|x_1| \leq 0.1$, $|x_2| \leq 0.01$, $|x_3| \leq 0.1$, $|u| \leq 0.25$, we obtain the linear feedback control

$$u = Kx = -37.85x_1 - 4.639x_2 + 0.475x_3$$

The ellipsoidal method provides $|d| \leq \alpha_{ell} = 1.27$ as the bound of the disturbance amplitude which assures the possibility of keeping the constraints satisfied. We have also seen that a less conservative bound can indeed be found for this controller, $\alpha_{EAS} = 1.45$, by computing a robustly invariant set for the EAS.

Question 1: can we do better?

To answer this question, we can compute the largest invariant (contractive) set under disturbance without forcing the control to be linear. We can again use the EAS with $\tau > 0$ small. With $\tau = 0.3$, $\lambda = 0.995$ and $\epsilon = 0.01$, we obtain

$$\alpha^* = 1.47$$

Further reducing τ does not give essential improvements. The price of using a nonlinear control is complexity. For instance, the piecewise-linear control can be computationally hard to be implemented. In the performed experiments, the number of simplicial sectors, hence the number of required gains is around 50–100. Therefore, these numerical tests show that, to increment the performance of a small percentage, we have to adopt a much more complex control in terms of description of the resulting invariant set. If we try to push the performance close to the limits, we get either failures or empty sets or extremely complex controllers.

Then we have another legitimate question:

Question 2: so what?

The reply is that paradoxically, by adopting the set-theoretic approach, we do not always find realistic compensators but we can evaluate the performance of *any* other control. The computed value $\alpha^* = 1.47$ is virtually the "best we can do" up to numerical factors. So, given any control, no matter how designed, we can compare it with the "magic number" arising from the set-theoretic approach.

8.3.3 Assigning an invariant (and admissible) set

We have seen that controlled invariant sets have a fundamental role in the solution of state constrained problems. A possible way to approach a state constrained problem is then to fix a certain region and seek for a controller which renders this region positively invariant. This problem has an elegant solution in the case of an ellipsoid $\mathcal{E}[P, 1]$. Indeed such set is positively invariant for the system with a linear controller of the form $u = Kx$ if and only if

$$(A + BK)^T P + P(A + BK)^T \prec 0$$

which is an LMI in the unknown K. Note that this result applies as well to output feedback (in this case we have $u = KCx$, where $y = Cx$). As we have seen, a controlled invariant ellipsoid can always be associated with a linear controller and therefore the previous condition is necessary and sufficient for the controlled invariance of $\mathcal{E}[P, 1]$. The corresponding discrete time version can be written as follows. Consider the vector norm $\|x\|_P = \sqrt{x^T P x}$ and, for a matrix M, let $\|M\|_P$ be the corresponding induced norm. The set $\mathcal{E}[P, 1]$ is controlled invariant if and only if, for some K,

$$\|(A + BK)\|_P \leq 1$$

Again, this is a nice convex condition, suitable to be solved by means of efficient optimization tools.

Let us now consider the case of polytopes. Given a polytope in its vertex representation $\mathcal{V}[X]$, it is quite understood that its controlled invariance is equivalent to the existence of a Metzler-matrix H such that

$$\begin{aligned} AX + BU &= XH \\ \bar{1}^T H &\leq -\beta \bar{1}^T \\ \beta &\geq 0 \end{aligned}$$

For a fixed X, the above is a linear programming problem in the unknowns H, U, and β. The corresponding discrete-time condition is

$$\begin{aligned} AX + BU &= XP \\ \bar{1}^T P &\leq \lambda \bar{1}^T \\ 0 \leq \lambda &\leq 1 \end{aligned}$$

As already stated, the difference is that a controlled-invariant polytope cannot be, in general, associated with a linear controller. If one wishes to impose linearity, then there is only a further condition to add and precisely

$$U = KX$$

with the new unknown K. There are important exceptions of polyhedral regions, however, in which the control law turns out to be of reduced complexity. Such exceptions are diamond-shaped sets

$$\mathcal{D} = \{x = Xp, \ \ \|p\|_1 \leq 1\} \tag{8.16}$$

with X a square matrix, and simplices

$$\mathcal{S} = \{x = Xp, \ \ \bar{1}^T p = 1, \ \ p \geq 0\} \tag{8.17}$$

with X an $n \times n+1$ matrix, for which the following holds.

Proposition 8.19. *A controlled-invariant diamond of the form in (8.16) for a linear system can be associated with a linear feedback control.*

Proof. The discrete-time version of the proposition only is proved, since the continuous-time case is essentially the same. Let x_i be a column of X (then x_i and its opposite are vertices of $\mathcal{D}$). By assumption there exists u_i such that

$$Ax_i + Bu_i = Xp_i^*, \quad \|p_i^*\|_1 \leq 1, \ \ k = 1, \ldots, n$$

then any vector x_i is driven by u_i in $\mathcal{D}$. By symmetry, $-x_i$ is driven by $-u_i$ in $\mathcal{D}$. Consider the unique matrix K such that

$$U = KX$$

(uniqueness follows by the fact that X is square invertible). Then

$$Ax_i + BKx_i = Xp_k^*, \quad \|p_k^*\|_1 \leq 1$$

so that

$$(A + BK)X = X[p_1^* \; p_2^* \; p_n^*] \doteq XP^*$$

where $\|P^*\|_1 \leq 1$. Thus, according to the results presented in Subsection 4.5.2, $\mathcal{D}$ is positively invariant.

In the case of simplices, the following result holds

Proposition 8.20. *A controlled-invariant simplex of the form (8.17) for a linear systems can be associated with an affine feedback control.*

Proof. See [HvS04]

Actually, in the simplex case, more general problems can be solved via affine control such as that of reaching a specified face [HvS04, BR06].

There are other special classes of regions which have been considered in the literature. One of such cases is that in which the region is associated with the output variable $y = Cx$, which is constrained. Let us assume symmetric bounds and then, without restriction, consider the set

$$\mathcal{S} = \{\|Cx\|_\infty \leq 1\}$$

This set is unbounded as long as C has no more rows than columns (more in general when C has a non-trivial right kernel). Henceforth, it is assumed that C has full row rank. The problem is now that of checking whether there exists a linear control which renders this set positively invariant. This is, again, a linear programming problem. Indeed, in the discrete-time case, such a controller exists if there exist matrices H and K such that

$$\begin{aligned} C(A + BK) &= HC \\ \|H\|_\infty &\leq 1 \end{aligned} \tag{8.18}$$

In the continuous-time case the matrices H and K must be such that

$$\begin{aligned} C(A + BK) &= HC \\ H &\in \mathcal{H} \end{aligned} \tag{8.19}$$

where $\mathcal{H}$ is the class introduced in Definition 4.40.

We investigate now this problem more in detail to establish a connection with the theory of (A, B)-invariant subspaces [BM92]. Consider a matrix complementary to C, precisely a matrix $\tilde{C}$ such that

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

is invertible. Then, if the equality in (8.18) holds, then

$$\begin{bmatrix} C \\ \tilde{C} \end{bmatrix} (A + BK) = \begin{bmatrix} H & 0 \\ P & Q \end{bmatrix} \begin{bmatrix} C \\ \tilde{C} \end{bmatrix}$$

where P and Q are appropriate matrices. Consider the transformation

$$\begin{bmatrix} y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} C \\ \tilde{C} \end{bmatrix} x(t)$$

so that the closed-loop system is transformed into

$$\begin{bmatrix} \dot{y}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} H & 0 \\ P & Q \end{bmatrix} \begin{bmatrix} y(t) \\ z(t) \end{bmatrix}$$

which means that, in the new reference frame, the z-subspace (i.e., the subspace of vectors for which $y = 0$) is an invariant subspace. In the original variables, the set of all vectors such that

$$Cx = 0$$

is invariant. The above reasoning can be formalized in the following proposition.

Proposition 8.21. *The problem of finding a stabilizing control which makes the set* $\mathcal{S} = \{x : \|Cx\|_\infty \leq 1\}$ *positively invariant can be solved only if* $ker\{C\}$ *is an* (A, B)*-invariant subspace* [5].

The problem with the unbounded set $\mathcal{S}$ is that its contractivity does not assure stability. It can be shown that the condition implies partial stability, namely "asymptotic stability" of the variable y only.

A natural question is how to assure stability and precisely: under which structural condition does there exist a control $u = Kx$ such that

- condition (8.18) (or (8.19)) is satisfied so that $\mathcal{S} = \{x : \|Cx\|_\infty \leq 1\}$ is positively invariant;
- the closed-loop system is stable.

[5] We remind that a subspace is said (A, B)-invariant if it is controlled-invariant[BM92].

This problem was considered and solved in [CH92] where it is shown that the problem can be faced by means of an eigenstructure assignment approach.

We consider the problem in the special case of square systems (i.e., as many inputs as outputs) and with relative degree 1, precisely *CB* invertible. The following proposition holds.

Proposition 8.22. *Given a square system with relative degree* 1*, the two requirements can be met if and only if the system has no unstable zeros.*

Proof. If the system is square with relative degree 1, then there exists a state transformation such that the system is in the form

$$\begin{bmatrix} \dot{y}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} y(t) \\ z(t) \end{bmatrix} + \begin{bmatrix} CB \\ 0 \end{bmatrix} u(t) \tag{8.20}$$

Note that the output matrix is $[\, I \;\; 0 \,]$. The zeros of this system are the eigenvalues of A_{11}. Necessity follows by Proposition 8.21. Indeed we have seen that the kernel of C must become positively invariant. This means that the closed-loop-system matrix must be of the form

$$\begin{bmatrix} A_{11} + CBK_1 & A_{12} + CBK_2 \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} + CBK_1 & 0 \\ A_{21} & A_{22} \end{bmatrix} \tag{8.21}$$

Then, to have closed-loop stability, A_{22} must be stable. Conversely, assume that A_{22} is stable. Take K_2 such that $A_{12} + CBK_2 = 0$, so the system is reduced as in (8.21). The problem of assuring the invariance of the set $\mathcal{S} = \{x : \|Cx\|_\infty \leq 1\}$ reduces, in the new reference frame, to assure that the set $\{(y, z) : \|y\|_\infty \leq 1\}$ is positively invariant. Since the variable y is governed by the equation

$$\dot{y}(t) = [A_{11} + CBK_1]y(t) = Hy(t)$$

this reduces to the choice of H such that the unit ball $\{y : \|y\|_\infty \leq 1\}$ is invariant for this sub-system. This is an LP problem. This problem as feasible solution: for instance, take K_1 such that

$$H = A_{11} + CBK_1 = -\beta I$$

$\beta > 0$ (Exercise 11). Note that in the discrete-time case the condition can be changed by seeking for H such that $\|H\|_\infty \leq 1$. Thus the condition is sufficient.

The reader is referred to the paper [CH92], where this stable-zero condition has been introduced. The considered condition can be viewed as a minimum-phase condition for the constraints, as explained in [SHS02]. We remind that determining a convex set, including 0 as an interior point, which is positively invariant (or contractive) does not assure the system stability as long as this set is unbounded. Therefore the

assumption of stable zeros is essential. If we remove it, only partial stability can be assured, roughly the stability of the y variable (see [Vor98] for details). An example of application is reported next.

Example 8.23 (Stabilization with monotonic output). As an example we consider the following problem. Given a SISO system, we wish to stabilize it with the additional condition that for all initial conditions the output converges to zero monotonically. This implies that any set of the form $\{x : |Cx| \leq \mu\}$ is positively invariant (or contractive). This problem can be relevant in all the circumstances in which the primary goal is to drive the output to zero without overshooting or damped oscillations. This property can be a desirable achievement in fluid control.

Consider the two tank system previously considered and its linearization which turns out to be of the form

$$\begin{bmatrix} \dot{z}_1(t) \\ \dot{z}_2(t) \end{bmatrix} = \begin{bmatrix} -\rho & \rho \\ \rho & -(\rho+\delta) \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t)$$

where z_1 and z_2 are water levels (with respect to the steady state) and u is the input flow. Then consider the following problem.

Find a stabilizing control such that, for any initial condition $z_1(0)$ and $z_2(0)$, the error on the first (upper) tank decreases monotonically.

This problem can be solved if the output is taken to be the level of the first tank, $y(t) = z_1(t)$, since the system is of degree 1 and has stable zeros, as in the theorem assumptions. The control law which solves the problem is

$$u(t) = -k_1 z_1(t) - \rho z_2(t)$$

with k_1 taken such that $k_1 + \rho = \alpha > 0$ which yields the closed-loop matrix

$$\begin{bmatrix} -\alpha & 0 \\ \rho & -(\rho+\delta) \end{bmatrix}$$

It is immediate that any stripe

$$\Sigma(\mu) = \{(z_1, z_2) : \ |z_1| \leq \mu\}$$

is positively invariant (actually it is α-contractive) for the closed-loop system. Note also that only the z_1 dynamics can be affected, since the other closed-loop pole turns out to be equal to the system zero $-(\rho+\delta)$, which is clearly invariant.

A further question concerns the possibility of achieving the same goal when the output is the second tank level. This is not possible. This basically means that for any control there exists a proper initial condition for which the error on the second tank level has necessarily a "rest" or a "trend inversion" in the convergence. Indeed there is no control which can render positively invariant the set $\{(z_1, z_2) : \ |z_2| \leq \mu\}$.

8.4 Control with rate constraints

There are several cases in which it is necessary not only to deal with amplitude bounds but also with rate bounds on the control input. Clearly, under such a requirement, in the continuous-time case, the control action $u(t)$ must be differentiable, say the following constraints hold [GdSTG03, MTB04, BMT96]:

$$\begin{aligned} u(t) &\in \mathcal{U} \\ \dot{u}(t) &\in \mathcal{V} \end{aligned} \tag{8.22}$$

In the discrete-time case the similar requirements are

$$\begin{aligned} u(t) &\in \mathcal{U} \\ u(t+1) - u(t) &\in \mathcal{V} \end{aligned} \tag{8.23}$$

The conceptually simplest way to cope with the above limitations is to introduce a new "control variable" $v = \dot{u}$ (respectively $v(t) = u(t+1) - u(t)$) and consider the following extended system:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{u}(t) \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} v(t) \tag{8.24}$$

or

$$\begin{bmatrix} x(t+1) \\ u(t+1) \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & I \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} v(t) \tag{8.25}$$

The extension is obtained by putting an integrator on the input of the original system. Therefore, u becomes a new state variable and a new input vector v shows up. By doing so, the original constraint $u \in \mathcal{U}$ has become a state constraint for the new system and the slew-rate constraint for the original system, $\dot{u} \in \mathcal{V}$, has now become an input constraint, say $v \in \mathcal{V}$. This means that, if an integral control of the form

$$\begin{aligned} &\dot{u}(t) = v(t), \quad (\text{respectively } \; u(t+1) = u(t) + v(t)) \\ &v(t) = \phi(x, u) \end{aligned} \tag{8.26}$$

is applied, we obtain the same closed-loop system which could be achieved by applying the static controller

$$v(t) = \phi(x, u) \tag{8.27}$$

to the augmented system. The following result holds.

Theorem 8.24. *Assume the extended system (8.24) (or (8.25)) can be stabilized by means of a constrained control law $v(t) = \phi(x(t), u(t))$ computed for the extended*

system. Then, the original system can be stabilized by a control law complying with the constraints (8.22) (or (8.23)). The set of all the initial control-states for which constrained convergence is assured is $\mathcal{S}$, the domain of attraction associated with the extended controlled system.

It is worth focusing the reader's attention on the initial condition problem. The previous theorem (whose proof is omitted) does not distinguish, as far as the domain of attraction is concerned, between control and state initial value. Precisely, in applying the previous results one has that convergence is guaranteed under bounds for all states of the extended systems, namely for all pairs $[x^T \; u^T]^T \in \mathcal{S}$. Therefore the convergence to the origin with constraint satisfaction does not depend only on the initial state, but also on the initial control value. Two cases have to be distinguished.

i) The control value can be initialized. In this case the set of initial states which can be driven to the origin without constraint violation is the projection of $\mathcal{S}$ on the original state space

$$Pr(\mathcal{S}) = \{x : \; [x^T \; u^T]^T \in \mathcal{S}, \;\; \text{for some } u\}$$

ii) The control value is initially assigned $u(0) = u_0$. In this case the set of initial states which can be driven to the origin without constraint violation is the intersection of $\mathcal{S}$ with the $u = u_0$ affine manifold

$$\{x : \; [x^T \; u_0^T]^T \in \mathcal{S}\}$$

Example 8.25. Consider the continuous-time scalar system

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

with $|u| \leq 1$. For such a system it has been shown that the set

$$\mathcal{P} = \{x : |x| \leq a\}$$

with $a < 1$ is a domain of attraction. Let us now investigate what happens when the input derivative is also constrained as $|\dot{u}| \leq \bar{v}$. By adding the equation

$$\dot{u} = v$$

the extended system

$$\begin{bmatrix} \dot{x}(t) \\ \dot{u}(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v(t)$$

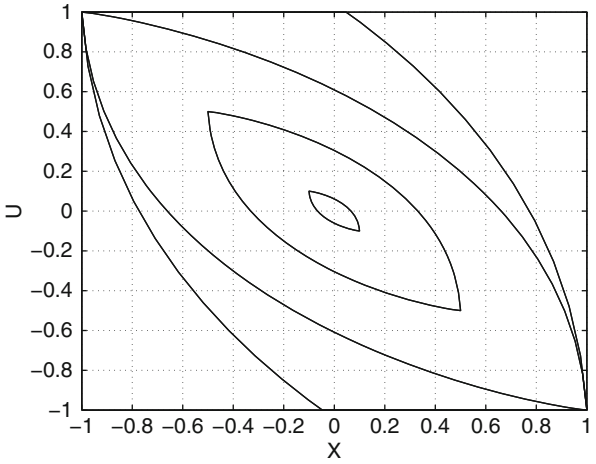


Fig. 8.15 Domains of attraction for the extended system with the given constraints

is obtained (with $|v| \leq \bar{v}$). The domains of attraction of the extended system for $\bar{v} = 2, 1, 0.5, 0.1$ are depicted in Figure 8.15. Such sets were computed by choosing $\tau = 0.1$ and setting $\lambda = 0.9999$ and $\epsilon = 0.00001$. The lines resemble continuous curves since the vertices of the polytopes are really close (when $\bar{v} = 0.1$ there are more than 350 vertices). Had there been no constraints on $v(t)$ ($\bar{v} = \infty$) we would have expected a squared domain of attraction $\{(x, u) : |x| \leq 1, |u| \leq 1\}$. By choosing a finite value of $\bar{v}$ and decreasing it is possible to observe how the domain of attraction shrinks.

8.4.1 The rate bounding operator

A typical way to face the bounded variation problem is that of inserting in the input channel a suitable operator (Fig. 8.16). We have seen that in the case of a bounded amplitude system, the typical operator is the saturation one. Conversely if the problem is rate-bounded, a more sophisticated model is necessary and in the literature different operators have been proposed [SSS00, BHSB96].

Essentially, also the rate bound turns out to be dynamic. If ω is the desired rate-bounded signal and u is the actual signal, assuming $|\dot{u}| \leq \bar{v}$, the operator can be expressed as

$$\dot{u}(t) = \text{sat}_{\bar{v}} \left[\frac{\omega - u}{\tau} \right]$$

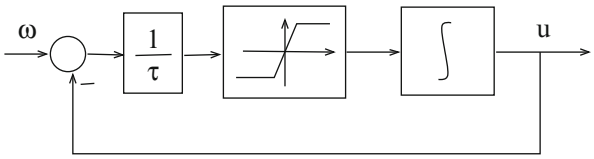


Fig. 8.16 The saturation operator.

Given these operators, the problem normally consists in the determination of a controller which produces the "desired control value" ω in such a way that there are no bad consequences once this is saturated. Note that for $\tau \to 0$ the above operator converges to

$$\dot{u}(t) = \bar{v}\,\mathrm{sgn}\,[\omega - u]$$

In the previous section a different approach was pursued, and more precisely a control law which satisfies both amplitude and rate constraints without considering any saturation operator was derived. We will reconsider the above saturation operator later, in Subsection 9.7.1.

8.5 Output feedback with constraints

So far, state feedback control problem has been considered. Clearly, as it is known in most practical problems, output feedback is the general standard case. This fact is, as already mentioned, a source of difficulties in our approach.

We might try to convince the reader that, in principle, in the continuous-time case, the domain of attraction is essentially the same for state and output feedback provided that (living in a ideal world)

- the model is known exactly;
- the output is disturbance free.

Indeed, under such conditions one can always apply a "fast observer," which recovers the state, and then apply an estimated state feedback. We just give a sketch of the idea. Given the reachable and observable system

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t)$$

it is possible to design a control which, starting at time $t = 0$, performs the following operations:

1. keeps the control $u = 0$ for an "estimation" period $t \in [0,\ \tau]$ in such a way that the estimated state $\hat{x}(\tau)$ is sufficiently close to $x(\tau)$;
2. applies a proper estimated-state feedback control $u = \Phi(\hat{x})$ for $t \geq \tau$.

An easy way to perform the operation is to consider the sampled data system during the period $t \in [0,\ \tau]$

$$x(t+T) = e^{AT}x(t) = A_D x(t), \qquad y(t) = Cx(t)$$

with $T = \tau/k$. For $k \geq n$ it is immediate to derive the expression which provides $x(t)$ as a function of the first t measurements

$$\underbrace{\begin{bmatrix} C \\ A_D C \\ \vdots \\ A_D^{k-1} C \end{bmatrix} A_D^{-k}}_{M_t} x(\tau) = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(k-1) \end{bmatrix} \doteq Y_\tau$$

Under ideal condition this system is consistent and its solution, the estimate of x, can be derived as

$$\hat{x}(t) = {M_\tau}^\dagger\, Y_\tau \tag{8.28}$$

where ${M_\tau}^\dagger$ is the pseudoinverse. It is well known that this procedure can be applied even in the presence of noise, and in this case the expression (8.28) provides the mean square solution.

Assume that the ideal condition $\hat{x}(\tau) = x(\tau)$ holds. Then, for $t \geq \tau$ one can apply an observer-based control

$$\dot{z}(t) = (A - LC)z(t) + Ly(t) + Bu, \quad u = \Phi(\hat{x})$$

where $\Phi(\hat{x})$ is computed by means of the preferred "state feedback recipe".

The problem is that, during the interval $[0,\ \tau]$, the state evolves open-loop, say $x(\tau) = e^{A\tau}x(0)$. If we wish to have $x(\tau) \in \mathcal{P}$, then we must require that $x(0) = e^{-A\tau}x(\tau)$ is in the set

$$\tilde{\mathcal{P}} = [e^{A\tau}]^{-1}\mathcal{P} = e^{-A\tau}\mathcal{P}$$

which is a feasible set of initial conditions. If $\mathcal{P} = \mathcal{P}[F, g]$ is a polyhedral set, one gets the following explicit expression

$$\tilde{\mathcal{P}} = \mathcal{P}[Fe^{A\tau}, g]$$

while, in the case of an ellipsoid $\mathcal{P} = \{x : x^T P x \leq 1\}$, one gets

$$\tilde{\mathcal{P}} = \{x : x^T (e^{A\tau})^T P e^{A\tau} x \leq 1\}$$

It is immediate that, as the "observation time" is reduced,

$$\tilde{\mathcal{P}} \to \mathcal{P}$$

so that the same domain of attraction of the state-feedback compensator is recovered.

A similar strategy can be applied in the discrete-time case, though in this case it is not possible to reduce the "observation interval" below a certain quantity. For single-output systems, the lower limit is precisely n steps (see Exercise 12).

Unfortunately, the above ideal arguments are in general far from the real situation. Indeed, there are two main sources of problems. The matrix M_t becomes very ill conditioned as $\tau \to 0$. Disturbances have to be taken into account and thus, as it is well known, it is not possible to reduce the estimation time too much, because in this case there is no "filtering" action on the disturbances.

Some approaches in the literature have been proposed to solve output feedback constrained control problems by means of observers (see, for instance, [MM96]). An explicit method to take observer errors into account will be considered later in Section 11.2.

8.6 The tracking problem

In this section, the special problem of tracking under constraints is considered. The problem in its generality is the following: assume we are given the plant

$$\begin{aligned}
\dot{x}(t) &= Ax(t) + Er(t) + Bu(t) \\
h(t) &= C_h x(t) + D_{hr} r(t) + D_{hu} u(t) \\
e(t) &= C_e x(t) + D_{er} r(t) + D_{eu} u(t) \\
y(t) &= C_y
\end{aligned}$$

where y is the measured output, r is the reference, e is a performance output, and h is an output on which constraints are imposed (see Figure 8.17)

$$h(t) \in \mathcal{H}$$

It is worth noting that a typical assumption for tracking problems is that the initial conditions are known (in practice, $x(0) = 0$) and in this case the output feedback problem is equivalent to the state feedback problem (provided that the

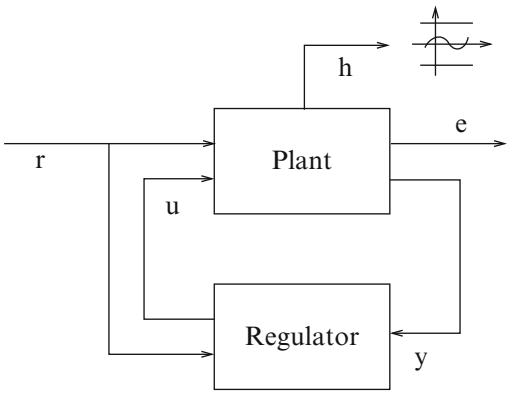


Fig. 8.17 Tracking scheme

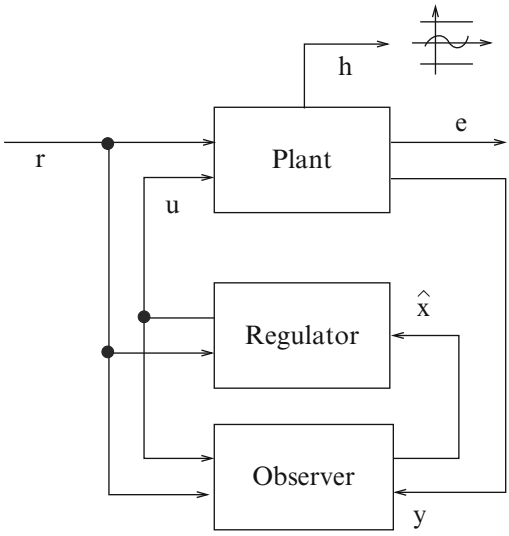


Fig. 8.18 Tracking scheme with observer

system is detectable) since r is a known signal and, by introducing an observer (see Figure 8.18), one can recover the state of the system. Therefore, as long as the system dynamics is known (the matrices are known exactly), the problem is equivalent to the state feedback problem, say to that of determining controllers of the form

$$u(t) = \Phi(x(t), r(t))$$

It is worth evidencing that the tracking problem is quite different with respect to constrained stabilization, since the reference is an additive input which has to be taken into account. Precisely, due to the dynamics, the tracking signal may produce constraint violation. This may well happen even in those cases in which the reference does not produce violation at steady state.

To face the problem there are basically three possible approaches.

- **(A posteriori)** The control is designed without taking into account the constraints. The effects w.r.t. to reference tracking are then analyzed and a class of reference signals for which the system performs well are determined.
- **(A posteriori, on line)** The control is designed without taking into account the existing constraints. In its on-line implementation, the reference r is suitably modified to prevent the system from violating the imposed constraints.
- **(A priori)** The control is designed by directly taking into account the existing constraints.

The first technique basically leads to a trial-and-error approach. Once the controller is designed, one can just check if under "reasonable reference signals" the system works appropriately. If one considers bounded references

$$r(t) \in \mathcal{R}$$

then the problem basically reduces to the computation of the reachability sets, to check if these are admissible with respect to constraints. Note that $r(t) \in \mathcal{R}$ includes constant references as the special case in which $\mathcal{R}$ is a singleton.

Conversely, if one wishes to check the system with respect to all the constant references inside $\mathcal{R}$, then, assuming $\mathcal{R}$ a polytope, the following property comes into play as long as linear compensators are adopted.

Proposition 8.26. *Assume the constraint set $\mathcal{H}$ is convex. Then, for $x(0) = 0$, there is no constraint violation for all constant references $\bar{r} \in \mathcal{R}$ if and only if each of the trajectories of the system corresponding to constant references chosen on the vertices $r(t) \equiv \bar{r}^{(k)} \in vert\,\{\mathcal{R}\}$ does not violate the constraints.*

Note that it is very easy, by scaling the problem, to compute the "largest size" of admissible signals, precisely

$$\sigma_{max} = \sup\{\lambda : \text{ for } r \in \lambda\,\mathcal{R} \text{ no constraint violations occur}\}$$

If $\mathcal{H}$ is the unit ball of a norm, one has

$$\sigma_{max}^{-1} = \mu = \sup_{t \geq 0} \|y(t)\|$$

More general results can be achieved by means of reachability sets. However this way of dealing with constraints is not always efficient, since an unsuitable choice of the compensator might produce very unsatisfactory results. This is especially true in the case of linear compensators, since linearity, which is always seen as a desirable property, introduces big restrictions on the class of control functions.

As we will be seen in the next section, by introducing appropriate non-linearities, considerable advantages can be achieved.

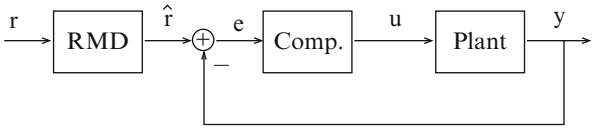


Fig. 8.19 Open-loop reference management

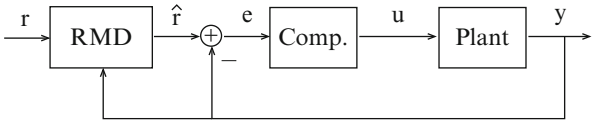


Fig. 8.20 Closed-loop reference management

8.6.1 Reference management device

An idea to face a constrained tracking problem is that of reference managing. With the term Reference Management Device (RMD) we generically refer to a device which modifies the reference value in an appropriate way. As a special case, these RMD include the reference supervisor [KAS89, KAS88], the reference governor [GT91, GKT95, GK02], and the command governor [BCM97, CMA00].

In general, the main idea of an RMD is the following. The control is computed based on some optimality criteria and, in order to avoid constraint violation, the reference signal is suitably modified. Basically, this job can be performed in two ways.

- **Open-loop reference management**: basically a pre-filter is adopted to smooth the reference (see Fig. 8.19).
- **Closed-loop reference management**: the reference is modified based on the current system (possibly estimated) state (see Fig. 8.20).

A typical choice of the open-loop reference management device or governor is a low-pass filter. Let us explain this concept with basic considerations in the case of system overshoot. It is well known that in the control design practice one can decide to allow a system step response overshoot in order to achieve a short raising time, and then a fast response. However, the overshoot can cause constraint violations if the reference is too close to the limit. Therefore if the system response exhibits an overshoot due to a too "sharp" reference, the reference can be smoothed in order to achieve a softer response. However, by proceeding in this way, the speed of convergence is reduced even for small reference signals which might not produce violations. The closed-loop reference management device works in a different way since the reference signal is modified only in case of "danger."

A typical and simple way to construct this reference management device is the following. Assume $r(t)$ is the reference signal and $\hat{r}(t)$ is the modified reference. Then

$$\hat{r}(t) = \sigma(r(t), x(t))\, r(t), \quad 0 \leq \sigma(r(t), x(t))) \leq 1 \tag{8.29}$$

where $\sigma(r(t), x(t)))$ is a "reduction coefficient" which is thought to be equal to its upper limit 1 under normal circumstances and it is smaller whenever necessary. A reasonable idea to exploit $\sigma(r(t), x(t))$ is to reduce it whenever there is danger of constraint violation. Consider again the discrete-time system

$$\begin{aligned} x(t+1) &= Ax(t) + E\hat{r}(t) + Bu(t) \\ u(t) &= Kx(t) + H\hat{r}(t) \in \mathcal{U} \\ h(t) &= C_h x(t) \in \mathcal{H} \end{aligned}$$

and assume that $A + BK$ is Schur stable and that $\mathcal{U}$ and $\mathcal{H}$ are C-sets. Assume that $\mathcal{P}$ is a contractive C-set for the system when $r = 0$ and

$$K\mathcal{P} \subseteq \mathcal{U}, \quad C_h\mathcal{P} \subseteq \mathcal{H}$$

Then one can design the RMD as follows:

$$\begin{aligned} \sigma(r, x) = \max_{0 \le \sigma \le 1} \sigma, \ \text{ s.t.} \\ Ax + E\sigma r + B(Kx + H\sigma r) \in \mathcal{P} \\ Kx + H\sigma r \in \mathcal{U} \end{aligned} \tag{8.30}$$

The next property, reported without proof, holds.

Proposition 8.27. *The RMD in (8.30) is such that, for all $x(0) \in \mathcal{P}$ and for all $r(t)$, the constraints are not violated.*

Example 8.28. Consider the simple discrete-time plant

$$y(z) = \frac{1}{z - \eta} u(z)$$

equipped with the integral compensator

$$u(z) = \frac{k}{z - 1}(r(z) - y(z))$$

and with the constraint $|y| \le 1$. Let $\eta = 0.8$ and $k = 0.08$ and assume zero initial conditions for both the integrator and the plant variables and let the reference be $r(t) \equiv 0.8$. Setting $x_1(k) = y(k)$ and $x_2(k) = u(k)$, the following state realization is found

$$\begin{aligned} \begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} &= \begin{bmatrix} \eta & 1 \\ -k & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ k \end{bmatrix} r(t), \\ y(t) &= x_1(t) \end{aligned}$$

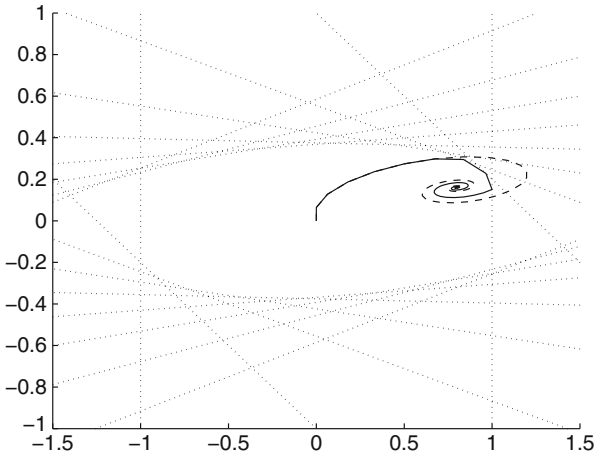


Fig. 8.21 The trajectory with RMD (plain) and without RMD (dashed)

The largest invariant set included in the constraint set $|x_1| \leq 1$ was computed, resulting in the polyhedron

$$\mathcal{P} = \{\|Fx\|_\infty \leq 1\}$$

where

$$F = \begin{bmatrix} 1.0000 & 0 \\ 0.8000 & 1.0000 \\ 0.5400 & 1.8000 \\ 0.2520 & 2.3400 \\ -0.0324 & 2.5920 \\ -0.2851 & 2.5596 \\ -0.4841 & 2.2745 \\ -0.6147 & 1.7904 \\ -0.6708 & 1.1757 \end{bmatrix}$$

By computing the natural evolution, one can see from both the phase diagram and the time evolution (Figures 8.21 and 8.22, dashed lines) that there are constraint violations. Conversely, if one computes the modified reference as

$$\hat{r}(t) = \max\{\rho : \|F(Ax + B\rho)\|_\infty \leq 1, \;\; 0 \leq \rho \leq r\}$$

then the plain trajectories which satisfy the constraints can be achieved. It is apparent that the RMD is active only when the system output $y = x_1$ approaches the constraints. The open-loop RMD was also considered, which basically consists in adopting the filter

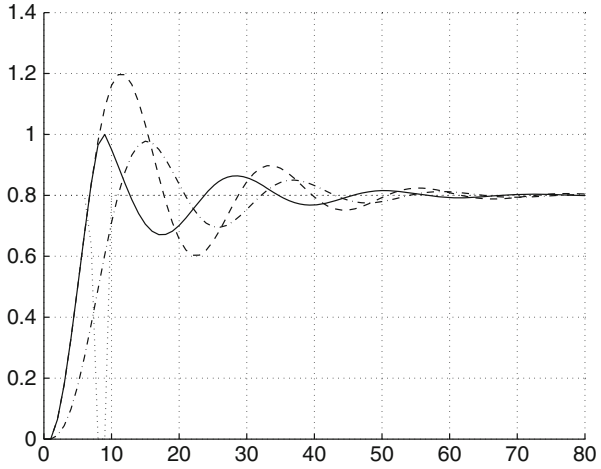


Fig. 8.22 The time evolution of y with RMD (plain), without RMD (dot-dashed), and with open-loop RMD and the modified reference $\hat{r}$, for $r(t) \equiv 0.8$

$$\hat{r}(z) = \frac{1-\xi}{z-\xi}\, r(z)$$

where $\xi = 0.81$ is the smallest value, determined by iteration, such that there are no violations of the bound $|y| \leq 1$. The corresponding (dot-dashed) time evolution in Fig. 8.22 actually remains feasible, though its convergence is quite slower than that achieved by the nonlinear reference governor.

It is quite interesting to notice that the device modifies the reference before the variable y reaches its bounds. It is quite easy to see that if the device waited until the bound (in this case $y = 1$) is reached, there would be no possibility of avoiding constraints. Indeed, the system reaches the bound in a point for which the next output would be $y(k) > 1$ and, since the vector B has a zero first entry, there would be no input choices $\hat{r}$ at that point[6]. Finally we analyze the behavior of the system for a reference value which is very close to the limit, and precisely when $r(t) \equiv 0.98$. In this case, both the closed-loop system without RMD and with open-loop RMD violate the constraints, while that achieved by means of the closed-loop RMD still exhibits an admissible transient (see Fig. 8.23).

The previous example shows the main idea at the basis of the reference governor technique. The reader is referred to specialized literature [GT91, GKT95, GK02, BCM97, Bem98, CMA00, GTC11] for further details. Applications of RMD can be found in [KGC97, CMP04].

[6] Too late!

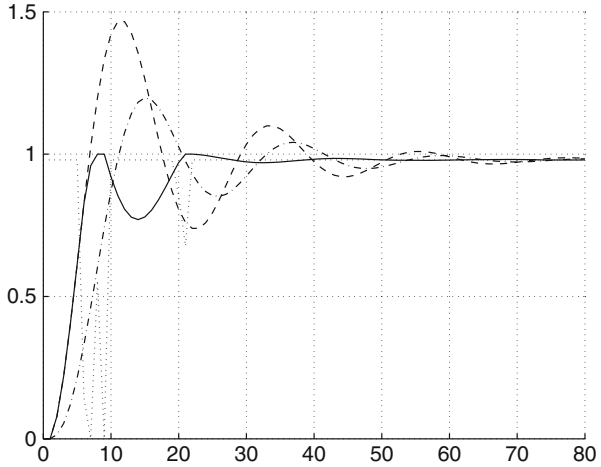


Fig. 8.23 The time evolution of y with RMD (plain), without RMD (dot-dashed), and with open-loop RMD and the modified reference $\hat{r}$, for $r(t) \equiv 0.8$.

It has to be pointed out that there are other techniques to deal with the tracking problem in the presence of control saturation, amongst which it is definitely worth mentioning the anti-windup approach (see, for instance, [GS83, HKH87]) and the override control (see [GS88]). A quite different approach, based on the reachable set computation, has been proposed in [GK92], where a technique was presented which produces suitable bounds on the reference signal (and its derivatives) in such a way that no constraint violation can occur.

8.6.2 The tracking domain of attraction

In this section we present a contribution by the authors [BM00], concerning the tracking domain of attraction, namely the set of all states starting from which it is possible to track a constant reference. In the following, both discrete and continuous-time square (i.e., $p = m$) systems

$$\begin{aligned} \delta x(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{8.31}$$

are considered, where δ represents the shift operator $\delta x(t) = x(t+1)$ in the discrete-time case and the derivative operator in the continuous-time case. The vector $y(t) \in \mathbb{R}^q$ is the system output, the vector $x(t) \in \mathbb{R}^n$ is the system state, and $u(t) \in \mathbb{R}^q$ is the control input. The couple (A, B) is assumed to be stabilizable and x and u are subject to the constraints

$$x(t) \in \mathcal{X} \tag{8.32}$$

$$u(t) \in \mathcal{U} \tag{8.33}$$

where $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{U} \subset \mathbb{R}^q$ are assigned C-sets. The tracking problem is more general than the stabilization problem as long as 0 is a possible reference. Since in practice we are not only interested in stability, but also in assuring a certain speed of convergence, we assume that a certain contractive set is given. This set will be referred to in the sequel as domain of attraction with speed of convergence β (or λ). This assumption is not restrictive since, as we have seen in the previous sections, the existence of a domain of attraction $\mathcal{P} \subset \mathcal{X}$ is essential in the stabilization problem under constraints. Once a domain of attraction $\mathcal{P}$ has been computed to solve the problem by means of one of the available techniques (for instance [MB76, GC86a, GC87, KG87, BM96a, BM98]), it is possible to replace the constraint $x(t) \in \mathcal{X}$ by the new constraint

$$x(t) \in \mathcal{P},$$

which is what will be done hereafter.

A quite natural assumption in dealing with tracking problems is that the system is free from zeros at one (zeros at the origin in continuous-time) and, to keep things simple, this assumption will be adopted at the beginning. In terms of state space representation, this amounts to imposing that the square matrix

$$M_d = \begin{bmatrix} A - I & B \\ C & D \end{bmatrix} \left(\text{resp. } M_c = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right) \tag{8.34}$$

is invertible (the general case in which M_d (M_c) is not invertible or even non-square will be commented later). Under condition (8.34), the system has the property that for any constant reference $r \in \mathbb{R}^q$ there is a unique state-input equilibrium pair $(\bar{x}, \bar{u})$ such that the corresponding equilibrium output is r. Such a pair is the unique solution of the equation

$$M_d \begin{bmatrix} \bar{x}(r) \\ \bar{u}(r) \end{bmatrix} = \begin{bmatrix} 0 \\ r \end{bmatrix} \left(\text{resp. } M_c \begin{bmatrix} \bar{x}(r) \\ \bar{u}(r) \end{bmatrix} = \begin{bmatrix} 0 \\ r \end{bmatrix} \right)$$

Thus it is possible to define the set of admissible constant reference vectors

$$\mathcal{R} = \{ r \in \mathbb{R}^q : \bar{u}(r) \in \mathcal{U},\ \bar{x}(r) \in \mathcal{P} \}.$$

Being $\mathcal{P}$ and $\mathcal{U}$ both bounded, $\mathcal{R}$ is also bounded. The set $\mathcal{R}$ is of fundamental importance. It is the set of all reference vectors for which the corresponding input and state steady-state pairs do not violate the constraints $\bar{u} \in \mathcal{U}$ and $\bar{x} \in \mathcal{P}$. While the reason for imposing the former is obvious, the second deserves some explanations because we originally imposed $x \in \mathcal{X}$. If $\bar{x}(r)$ were not included in

a domain of attraction then, since asymptotic tracking requires $x(t) \to \bar{x}(r)$, this would automatically cause constraints violation due to the fact that $x(t)$ doesn't belong to a domain of attraction.

We are now going to introduce the set of all the admissible signals to be tracked, formed by the signals $r(t)$ having a finite limit r_∞, with the condition that r_∞ has some admissibility condition with respect to the constraints.

Definition 8.29 (Admissible reference signal). Assume that a small $0 \leq \epsilon < 1$ is given. A reference signal $r(t)$ is admissible if it is continuous[7] and such that

$$\lim_{t\to\infty} r(t) = r_\infty \in (1-\epsilon)\mathcal{R} \doteq \mathcal{R}_\epsilon.$$

The parameter ϵ, as we will see later, is introduced to avoid singularities in the control. Such an ϵ may be arbitrarily small and thus it does not practically affect the problem. We stress that an admissible reference signal *does not need to assume its values in* $\mathcal{R}_\epsilon$, but only its limit r_∞ needs to do this. Now we can state the following basic definition.

Definition 8.30 (Tracking domain of attraction). The set $\mathcal{P} \subset \mathcal{X}$ is said a tracking domain of attraction if there exists a (possibly nonlinear) feedback control

$$u(t) = \Phi(x(t), r(t))$$

such that, for any $x(0) \in \mathcal{P}$ and for every admissible reference signal $r(t)$,

i) $x(t) \in \mathcal{P}$ and $u(t) \in \mathcal{U}$,
ii) $y(t) \to r_\infty$ as $t \to \infty$.

It is worth stressing once more that, since $r(t) = 0$ is an admissible reference signal, any tracking domain of attraction is also a domain of attraction. It will soon be shown that the opposite is also true, say *every domain of attraction* $\mathcal{P}$ *is a tracking domain of attraction*. The importance of this assertion lies in the fact that the tracking problem can be solved once one has found a domain of attraction by any of the described techniques.

Remark 8.31. Note that, since the matrices in (8.34) are assumed invertible, the condition $y(t) \to r_\infty$ as $t \to \infty$ is equivalent to the two conditions $x(t) \to \bar{x}_\infty \doteq \bar{x}(r_\infty)$ and $u(t) \to \bar{u}_\infty \doteq \bar{u}(r_\infty)$.

We recall that, by choosing a parameter $\beta > 0$ large or ($\lambda < 1$ small, in the discrete-time case), it is always possible to impose fast convergence. However, under reachability assumption, for β (λ) approaching 0 (1), the set $\mathcal{P}$ approaches the largest invariant set. Therefore there is a trade-off between the convergence speed and the size of $\mathcal{P}$.

[7] The continuity requirement, obviously referred to the continuous-time case, is not essential, but avoids unnecessary complications.

For a given fixed $\bar{r}$, the corresponding state input pair is

$$\begin{bmatrix} \bar{x} \\ \bar{u} \end{bmatrix} = \begin{bmatrix} A - I & B \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \bar{r} \end{bmatrix}.$$

If our goal were asymptotic tracking only, the translated Lyapunov function $\psi_{\mathcal{P}}(x - \bar{x})$, where $\psi_{\mathcal{P}}$ is the Minkowski functional of $\mathcal{P}$ (see Subsection 3.1.2), could be used but no constraint satisfaction could be assured.

The basic idea pursued here is that of "deforming" the function $\psi_{\mathcal{P}}(x)$ in such a way that the surface of level one is unchanged, while the zero value is assumed for $x = \bar{x}$.

Let us consider the main points of the tracking problem. For every reference signal $r(t) \to \bar{r} \in \mathcal{R}_\epsilon$ and initial condition $x(0) \in \mathcal{P}$, our goal is having:

1. $y(t) \to \bar{r}$;
2. $x(t) \in \mathcal{P}$.

According to Remark 8.31, for the first condition to hold it is necessary and sufficient that $x(t) \to \bar{x}$ and $u(t) \to \bar{u}$, whereas for the second we need the Lyapunov function which will be introduced shortly. The Minkowski functional of $\mathcal{P}$ can be written as

$$\psi_{\mathcal{P}}(x) = \inf\{\alpha > 0 \ : \ \frac{1}{\alpha} x \in \mathcal{P}\} \tag{8.35}$$

For every $\bar{x} \in int\{\mathcal{P}\}$ and $x \in \mathcal{P}$, consider the following function:

$$\Psi_{\mathcal{P}}(x, \bar{x}) = \inf\{\alpha > 0 \ : \ \bar{x} + \frac{1}{\alpha}(x - \bar{x}) \in \mathcal{P}\} \tag{8.36}$$

It is immediate that the just introduced function $\Psi_{\mathcal{P}}$ recovers the values of $\psi_{\mathcal{P}}$ when $\bar{x} = 0$, say $\Psi_{\mathcal{P}}(x, 0) = \psi_{\mathcal{P}}(x)$. For fixed $\bar{x}$, $\Psi_{\mathcal{P}}(x, \bar{x})$ is convex. Furthermore, the function $\Psi_{\mathcal{P}}(x, \bar{x})$ for $(x, \bar{x}) \in \mathcal{P} \times int\{\mathcal{P}\}$ is such that

$$\Psi_{\mathcal{P}}(x, \bar{x}) = 0 \quad \text{iff} \quad x = \bar{x} \tag{8.37}$$

$$\Psi_{\mathcal{P}}(x, \bar{x}) \leq 1 \quad \text{iff} \quad x \in \mathcal{P} \tag{8.38}$$

$$\Psi_{\mathcal{P}}(x, \bar{x}) = 1 \quad \text{iff} \quad x \in \partial\mathcal{P} \tag{8.39}$$

A sketch of the function $\Psi_{\mathcal{P}}(x, \bar{x})$ for fixed $\bar{x}$ is in Fig. 8.24. One further relevant property of this function is that $\Psi_{\mathcal{P}}$ is Lipschitz in x and positively homogeneous of order one with respect to the variable $z = x - \bar{x} \in \mathbb{R}^n$, for $\bar{x} \in int\{\mathcal{P}\}$, i.e.

$$\Psi_{\mathcal{P}}(\xi z + \bar{x}, \bar{x}) = \xi \Psi_{\mathcal{P}}(z + \bar{x}, \bar{x}). \tag{8.40}$$

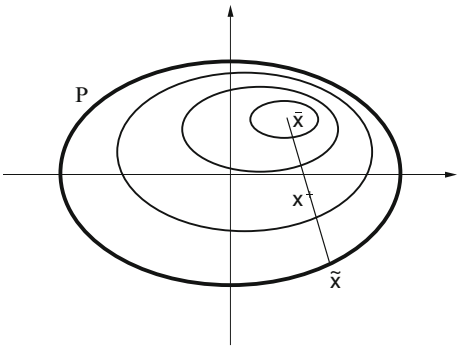


Fig. 8.24 The function $\Psi_{\mathcal{P}}(x,\bar{x})$ for fixed $\bar{x}$

In view of property (8.37), the function is a suitable Lyapunov candidate for tracking, and from (8.38–8.39), this function results to be suitable to prevent constraint violations, as it will be shown later.

Consider the function $\Psi_{\mathcal{P}}(x,\bar{x})$ and, for every $x \in \mathcal{P}$ and $\bar{x} \in int\{\mathcal{P}\}$, with $x \neq \bar{x}$, set

$$\tilde{x} \doteq \bar{x} + (x-\bar{x})\frac{1}{\Psi_{\mathcal{P}}(x,\bar{x})} \in \partial\mathcal{P}.$$

The vector $\tilde{x}$ is the intersection of $\partial\mathcal{P}$ with the half line starting from $\bar{x}$ and passing through x (see Fig. 8.24). As shown in the previous sections, it is always possible to associate with $\mathcal{P}$ a stabilizing control law $\phi(x)$ which is Lipschitz and positively homogeneous of order one, namely such that

$$\phi(\alpha x) = \alpha\phi(x), \quad \text{for } \alpha \geq 0$$

Given such a $\phi(x)$, the next step in the derivation of a feedback tracking control law $\Phi(x,r)$ is the definition of a saturation map $\Gamma : \mathbb{R}^q \to \mathcal{R}_\epsilon$ as follows:

$$\Gamma(r) = \begin{cases} r\psi_{\mathcal{R}_\epsilon}(r)^{-1} & \text{when} \quad \psi_{\mathcal{R}_\epsilon}(r) > 1. \\ r & \text{otherwise} \end{cases}$$

say

$$\Gamma(r) = r \text{ sat}\left(\frac{1}{\psi_{\mathcal{R}_\epsilon}(r)}\right)$$

where sat$(\cdot)$ is the 1-saturation function and $\psi_{\mathcal{R}_\epsilon}$ the Minkowski function of $\mathcal{R}_\epsilon$. Note that $\Gamma(r)$ is the identity if r is restricted as $r \in \mathcal{R}_\epsilon$. Conversely, for $r \notin \mathcal{R}_\epsilon$, $\Gamma(r)$ is the intersection of $\partial\mathcal{R}_\epsilon$ and the segment having extrema 0 and r.

The proposed candidate control law is the following:

$$\Phi(x,r) = \phi(\tilde{x})\Psi_{\mathcal{P}}(x,\bar{x}) + (1-\Psi_{\mathcal{P}}(x,\bar{x}))\bar{u} \tag{8.41}$$

where

$$\begin{bmatrix} \bar{x}(\bar{r}) \\ \bar{u}(\bar{r}) \end{bmatrix} \doteq M_d^{-1} \begin{bmatrix} 0 \\ \bar{r} \end{bmatrix} \left(\text{resp. } \begin{bmatrix} \bar{x}(\bar{r}) \\ \bar{u}(\bar{r}) \end{bmatrix} \doteq M_c^{-1} \begin{bmatrix} 0 \\ \bar{r} \end{bmatrix} \right) \tag{8.42}$$

and

$$\bar{r} = \Gamma(r). \tag{8.43}$$

Note that, for $r \in \mathcal{R}_\epsilon$, (8.43) does not play any role. Note also that, since $\bar{r} = \Gamma(r) \in \mathcal{R}_\epsilon$, then $\bar{x} \in int\{\mathcal{P}\}$, thus the term $\Psi_{\mathcal{P}}(x,\bar{x})$ in (8.41) is defined. However, the expression (8.41) is not defined for $x = \bar{x}$ because of the critical term $\phi(\tilde{x})\Psi_{\mathcal{P}}(x,\bar{x})$. Nevertheless, in view of the homogeneity of ϕ and from the expression of $\tilde{x}$, it turns out that

$$\phi(\tilde{x})\Psi_{\mathcal{P}}(x,\bar{x}) = \phi(\Psi_{\mathcal{P}}(x,\bar{x})\tilde{x}) = \phi\left(x + (\Psi_{\mathcal{P}}(x,\bar{x}) - 1)\bar{x}\right) \tag{8.44}$$

Then $\phi(\tilde{x})\Psi_{\mathcal{P}}(x,\bar{x}) \to 0$ as $x \to \bar{x}$, so that the function can be extended by continuity by assuming

$$\phi(\tilde{x})\Psi_{\mathcal{P}}(\bar{x},\bar{x}) = 0.$$

The introduced control law inherits most of the properties from $\phi(x)$ according to the next proposition, which assures existence and uniqueness of the solution of (8.31) when the control $\Phi(x,r)$ is used, provided that the admissible reference signal $r(t)$ is measurable [BM00].

Proposition 8.32. *Suppose $\phi(x)$ is Lipschitz and homogeneous of order 1. Then $\Phi(x,r) : \mathcal{P} \times \mathbb{R}^q \to \mathcal{U}$ defined as in (8.41)–(8.43) is continuous and it is Lipschitz w.r.t. x.*

To have an idea on how this control works, note that the following implication holds

$$0 \le \Psi_{\mathcal{P}}(x(t),\bar{x}(t)) \le 1 \Longrightarrow x(t) \in \mathcal{P} \tag{8.45}$$

Therefore, for $\bar{x} \in int\{\mathcal{P}\}$, the control is just a *convex combination* of the control $\bar{u}(\bar{r})$ and $\phi(\tilde{x})$. Since, by construction, $\bar{u}(\bar{r}) \in \mathcal{U}$ and $\phi(\tilde{x}) \in \mathcal{U}$, the above implies that $\Phi(x,r) \in \mathcal{U}$. So, everything will be fine if the proposed control law guarantees (8.45) as well as the limit condition

$$\Psi_{\mathcal{P}}(x(t),\bar{x}(r_\infty)) \to 0, \tag{8.46}$$

where $\bar{x}(r_\infty)$ is the steady state associated with $r_\infty \in \mathcal{R}_\epsilon$ (note that $\Gamma(r_\infty) = r_\infty$). Indeed such a limit condition implies $x(t) \to \bar{x}(r_\infty)$ and, from (8.41) and (8.44), $\Phi(x(t), r(t)) \to \Phi(\bar{x}(r_\infty), r_\infty) = \bar{u}(r_\infty)$. Therefore, if (8.46) holds, $y(t) \to r_\infty$.

An extended concept of speed of convergence, appropriate for the condition (8.46), will now be introduced. In the discrete-time case, given a *fixed* $\bar{x} \in int\{\mathcal{P}\}$ we say that the tracking speed of convergence is $\lambda < 1$ if

$$\Psi_{\mathcal{P}}(Ax + Bu, \bar{x}) \leq \lambda \Psi_{\mathcal{P}}(x, \bar{x}).$$

In the continuous-time case, the tracking speed of convergence is $\beta > 0$ if the Lyapunov derivative $D^+\Psi_{\mathcal{P}}(x(t), \bar{x})$ of $\Psi_{\mathcal{P}}(x(t), \bar{x}, u)$ is such that

$$D^+\Psi_{\mathcal{P}}(x, \bar{x}, u) \doteq \lim_{h \to 0^+} \frac{\Psi_{\mathcal{P}}(x + h(Ax + Bu), \bar{x}) - \Psi_{\mathcal{P}}(x, \bar{x})}{h} \leq -\beta \Psi_{\mathcal{P}}(x, \bar{x})$$

where the existence of the limit is assured by the convexity of $\Psi_{\mathcal{P}}$ with respect to x.

We start by considering the special case of a constant reference signal. In this case, it is possible to show that, if there exists a domain of attraction to the origin with a certain speed of convergence, then the tracking goal can be achieved without constraints violation for all the initial states in such domain. Furthermore, for symmetric domains, it is possible to guarantee a speed of convergence which is *independent of the reference input* and depends only on the contractivity of the domain of attraction to the origin.

Theorem 8.33. *Let $\mathcal{P}$ be a domain of attraction with speed of convergence λ for the discrete-time dynamic system (8.31) associated with the control $\phi(x)$, positively homogeneous of order 1. Then, for every admissible constant reference signal $r(t) = \bar{r}$, the control law (8.41)–(8.42) is such that, for every initial condition $x(0) \in \mathcal{P}$, $x(t) \in \mathcal{P}$ and $u(t) \in \mathcal{U}$ for every $t \geq 0$ and $lim_{t\to\infty} y(t) = \bar{r}$. Moreover, if $\mathcal{P}$ is 0-symmetric, the speed of convergence $\lambda_{TR} = \frac{\lambda+1}{2}$ is guaranteed.*

The next theorem is the continuous-time version of the above.

Theorem 8.34. *Let $\mathcal{P}$ be a domain of attraction with speed of convergence β for the continuous-time dynamic system (8.31) associated with the control $\phi(x)$, positively homogeneous of order 1. Then, for every admissible constant reference signal $r(t) = \bar{r}$, the control law (8.41)–(8.42) is such that, for every initial condition $x(0) \in \mathcal{P}$, $x(t) \in \mathcal{P}$ and $u(t) \in \mathcal{U}$ for every $t \geq 0$ and $lim_{t\to\infty} y(t) = \bar{r}$. Moreover, if $\mathcal{P}$ is 0-symmetric, the speed of convergence is at least $\beta_{TR} = \frac{\beta}{2}$.*

The proposed control law can be successfully used even when the reference $r(t)$ is allowed to vary, provided that it is asymptotically admissible according to Definition 8.29.

Theorem 8.35. *Let $r(t)$ be admissible as in Definition 8.29. Any domain of attraction $\mathcal{P}$, with speed of convergence $\beta > 0$ ($0 \leq \lambda < 1$), for system (8.31) is a tracking domain of attraction. Moreover, the control law in (8.41)–(8.43) assures the conditions i) and ii) in Definition 8.30.*

Remark 8.36. If a constant reference $r \in \mathcal{R}_\epsilon$ and the corresponding steady state vectors derived from $\bar{x}$ and $\bar{u}$ by means of (8.42) are considered, then it is immediate to see that the following state and control translation can be applied: $\delta\hat{x} = A\hat{x} + B\hat{u}$ with the new constraints $\hat{u} = u - \bar{u} \in \mathcal{U} - \bar{u} = \hat{\mathcal{U}}$ and $\hat{x} = x - \bar{x} \in \mathcal{X} - \bar{x} = \hat{\mathcal{X}}$. From this algebraic point of view, our result amounts to proving that the largest domain of attraction of the translated problem is just achieved by translating the original largest domain of attraction as $\hat{\mathcal{P}} = \mathcal{P} - \bar{x}$.

8.6.3 Examples of tracking problems

Example 8.37. As a first example, consider the following continuous-time system:

$$\dot{x}(t) = \begin{bmatrix} 1 & 0.4 \\ 0.8 & 0.5 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

$$y(t) = \begin{bmatrix} 0.2 & 0.1 \end{bmatrix} x(t)$$

A linear state feedback compensator $u = Kx$ was designed to stabilize the closed-loop system while guaranteeing that every initial state $\|x(0)\| \leq 1$ results in $\|Kx\| \leq U_{max}$ along the system trajectory. By solving the corresponding set of LMIs [BEGFB04] for $U_{max} = 5$, the following gain

$$K = \begin{bmatrix} -3.6845 & -1.5943 \end{bmatrix}$$

was obtained. The associated ellipsoidal domain of attraction (with speed of convergence $\beta = .0001$) is $\mathcal{P} = \{x : x^T Q x \leq 1\}$, where

$$Q = \begin{bmatrix} 0.7859 & 0.2950 \\ 0.2950 & 0.1172 \end{bmatrix}$$

and it is the outer region depicted in Figure 8.25. Such a set is by construction contained in the region where $|Kx| \leq U_{max}$ (also depicted in Figure 8.25), thus the constraints on the reference value $\bar{r}$ can be derived by the set $\mathcal{P}$ alone and result in $|\bar{r}| \leq 0.2403$. Such a value was slightly reduced to 0.24 to guarantee that $\bar{x}(\bar{r})$ belongs to the interior of $\mathcal{P}$. This means that $\bar{r} = \Gamma(r) = 0.24\ \mathrm{sat}(r/0.24)$. Figure 8.26 shows the zero initial state time-evolution of the output (solid-line) corresponding to the reference signal (dashed line)

$$r(t) = \begin{cases} 0.40 & \text{for } 0 \leq t \leq 100, \\ 0.20 + 0.20e^{-(t-100)/10} & \text{for } 100 < t, \end{cases}$$

whereas in Figure 8.27 the state space evolution is depicted.

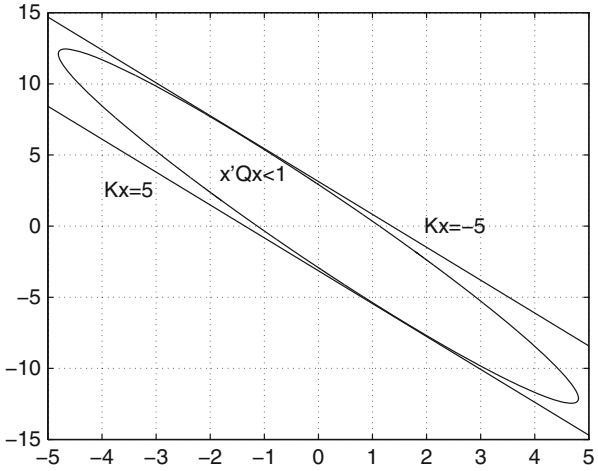


Fig. 8.25 Example 8.37: Domain of attraction

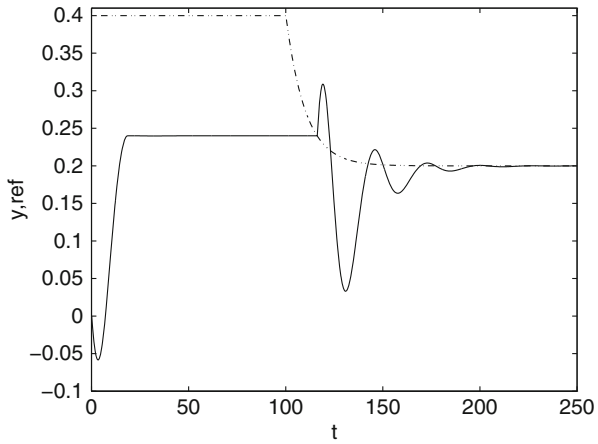


Fig. 8.26 Example 8.37: Output (solid) and reference (dashed) time evolution

Example 8.38. As a second example, consider the following discrete-time system

$$x(k+1) = \begin{bmatrix} 1 & 0.3 \\ -1 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} u(k)$$
$$y(k) = \begin{bmatrix} -0.1 & 0.3 \end{bmatrix} x(k)$$

with the state and control constraint sets, respectively, given by

$$\mathcal{X} = \{x \,:\, \|x\|_\infty \leq 1\}$$

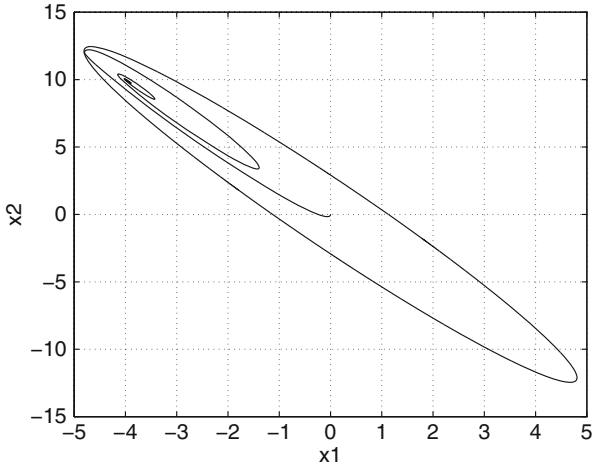


Fig. 8.27 Example 8.37: Domain of attraction and state space evolution

and

$$\mathcal{U} = \{u \ : |u| \leq 1\}.$$

A symmetric polyhedral domain of attraction $\mathcal{P} = \{x \ : \ \|Fx\|_\infty \leq 1\}$ with speed of convergence $\lambda = 0.9$ (which is the outer region in Figure 8.28) was computed and the following matrix F was determined:

$$F = \begin{bmatrix} 0 & 1 \\ 1.8160 & -0.2421 \\ 1.3140 & 0.1932 \end{bmatrix}$$

In this case, the constraints on the reference value derive from the constraint that $\bar{x} \in \mathcal{P}$ and translate in $|\bar{r}| \leq 0.27$. The linear variable structure controller associated with $\mathcal{P}$ is given by $u(x) = K_i x$, with $i = \arg\max_j |F_j x|$, where F_i and K_i are the i-th rows of F and of

$$K = \begin{bmatrix} -0.4690 & -0.7112 \\ -0.6355 & -0.7806 \\ -1.3140 & -0.1931 \end{bmatrix}$$

respectively. The control law proposed in [BM98] with $\epsilon = 0.01$ was applied, starting from zero initial state, for the reference signal

$$r(t) = 0.2 + 0.4 \sin(0.01t) e^{-0.005t}$$

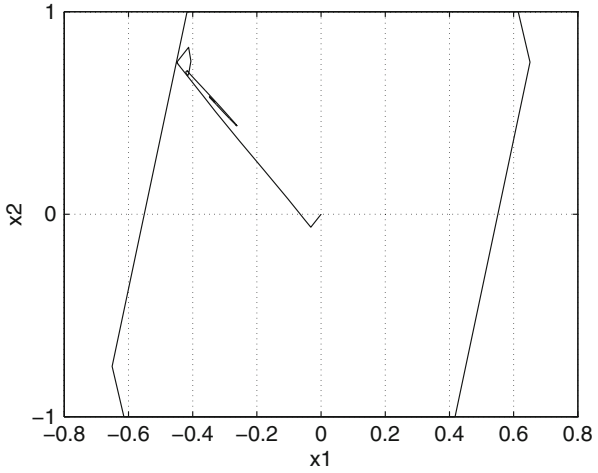


Fig. 8.28 Example 8.38: Polyhedral domain of attraction and state-space evolution

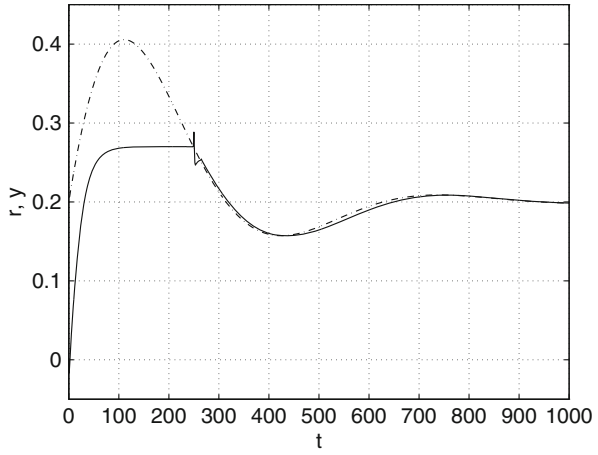


Fig. 8.29 Example 8.38: Output (solid) and reference (dotted) time-evolution

Figure 8.29 represents the time evolution of the output and the reference value, while Figure 8.28 shows the corresponding state-space trajectory.

8.7 Exercises

1. Prove the result in Proposition 8.27.
2. Derive the expression of the constraints for the control v in (8.8) assuming non-symmetric constraints $u_i^- \leq u_i \leq u_i^+$ or, more in general, $u \in \mathcal{U}$.

3. Consider Example 8.23, and prove that there is no continuous control which renders the set $|z_2| \leq \mu$ positively invariant. Provide a physical interpretation of this fact.
4. In Example 8.11, how could you find the largest cube $\mathcal{N}[\|\cdot\|_\infty, \mu]$ included in the computed domain? Hint: consider a proper norm of the rows of F.
5. The volume of the largest invariant ellipsoid for a linear system inside a polyhedral constraint set can be quite smaller than the largest invariant set. Show an example in which the volume ratio is $(4/3\pi)/8$. What about the largest invariant polyhedron included in an ellipsoidal constraint set?
6. In Example 8.11 how could you find the largest μ and ν such that $[\ 0\ \mu\ 0\ ]$ and $[\ 0\ 0\ \nu\ ]$ are included in the computed domain?
7. In Example 8.11 the most "squashing" constraint is associated with the last row of F, which is very close (up to a scaling factor) to one of the left eigenvectors. Why?
8. Look at the experimental transient in Fig. 8.12. Depict, in a qualitative way, the trajectory in the state space.
9. Consider the bang–bang control presented in the Example in Subsection 8.3.1. Find which sectors are associated with u_{max} and which ones are associated with $-u_{max}$.
10. Show that if CB is square invertible, then there exists a transformation which reduces the system as in (8.20). Hint: the transformation is of the form

$$T^{-1} = \begin{bmatrix} C \\ \tilde{C} \end{bmatrix}$$

with $\tilde{C}B = 0$.
11. In the proof of Proposition 8.22, show that determining H is an LP problem. Show that H chosen as indicated is a feasible solution. Hint: H must be a Metzler matrix, i.e. $M_{ij} \geq 0$ for $i \neq j$, and diagonally dominant $-M_{ii} \geq \sum_{i \neq j} M_{ij} \ldots$ and any C-set is invariant for $\dot{x} = [-\beta I]x \ldots$.
12. Consider the discrete-time system $x(t+1) = Ax(t) + Bu(t)$, $y(t) = Cx(t)$. Assume that we have an observer which gives the correct estimation $\hat{x}(t) = x(t)$ for $t \geq T$ and that $u(t) = 0$ for $0 \leq t < T$. Assume that $\mathcal{P} = \{x : Fx \leq g\}$ is a hard constraint set which is controlled-invariant under state feedback. Show that the set of all initial conditions for which $x(t) \in \mathcal{P}$ for all t is given by the inequalities $FA^k x \leq g, k = 0, \ldots, T$.
13. Consider the continuous-time system

$$\begin{aligned} \dot{x}_1 &= -\lambda_1 x_1 + u \\ \dot{x}_2 &= -\lambda_2 x_2 + u \end{aligned}$$

with $\lambda_2 > \lambda_1 > 0$. Assume $|u| \leq 1$ and find the infinite-time reachability set $\mathcal{R}_\infty$ from 0 (the solution is in Subsection 8.1.3). Then consider $T > 0$ and the corresponding sampled-data system

$$x_1((k+1)T) = e^{-\lambda_1 T} x_1(kT) + [1 - e^{-\lambda_1 T}]/\lambda_1 \; u$$
$$x_2((k+1)T) = e^{-\lambda_2 T} x_2(kT) + [1 - e^{-\lambda_2 T}]/\lambda_2 \; u$$

whose 0-reachability $\mathcal{R}_T$ set is a subset of $\mathcal{R}_\infty$ (why?). Show that $\mathcal{R}_T$ is a C-set but it is not a polytope. (Hint: by contradiction assume that it is a polytope, and show that this is impossible because there are infinitely many points of $\mathcal{R}_T$ which are arbitrarily close to the continuous curve delimiting $\mathcal{R}_\infty$). This shows also that the closure of the reachability set from 0 is not finitely determined.

Chapter 9
Switching and switched systems

In this chapter some basic results which concern the problem of switching and switched systems are reported. We wish to immediately tell the reader that this chapter has been written in a period of intensive research so that, quite differently from the previous sections, many other new results are expected.

There are several prestigious publications which provide complete surveys and to which the reader is referred for a comprehensive treatment of the topic. In particular, it is mandatory to mention the books [Lib03, Joh03, SG05] and the surveys [LA09, SWM+07, DBPL00, Col09]. In this chapter, we will briefly introduce the concept of hybrid systems, which are dynamical systems which include both continuous and logic variables, and then we will immediately consider the case of systems which can undergo switching, which are a special subclass of hybrid systems.

The nomenclature proposed in this chapter about switching and switched is not universal. Still, to keep the exposition simple, we will refer to "switching" systems when the commutation can be arbitrary and to "switched" systems when the commutation is controlled. There are, quite obviously, several intermediate situations. For instance, the case in which switching is state-dependent (as the bouncing ball), the case in which switching is arbitrary, but subject to dwell time, and other "mixed cases" in which some logic variables are controlled and some other are not controlled.

9.1 Hybrid and switching systems

In this section, a field in which set-theoretic considerations have a certain interest and some related problems are briefly presented. A hybrid system is a dynamic system which includes both discrete and continuous dynamics. A simple model (although not the most general one) of a hybrid system is given by

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_9

$$\dot{x}(t) = f_x(x(t), u(t), w(t), q(t)) \tag{9.1}$$

$$q(t) = f_q(x(t), u(t), w(t), q^-(t)) \tag{9.2}$$

where the variables x, u, and w have the usual meaning while $q(t)$ assumes its values in $\mathcal{Q}$, a discrete and finite set. Without restrictions it is assumed that

$$\mathcal{Q} = \{1, 2, \ldots, r\}$$

This class of systems is often encountered in many applications. The first equation (9.1) is a standard differential equation with the new input q, which can assume discrete values only. It can be interpreted as a set of r differential equations each associated with a value of q. The second equation (9.2), the novelty in this book, expresses the commutation law among the discrete values of $\mathcal{Q}$. Any possible commutation depends on the current continuous-time state, on the control, on the disturbance and on the last value $q^-(t)$ assumed by the discrete variable.

Example 9.1 (Oven in on–off mode). Let $\mathcal{Q} = \{0, 1\}$, let $\bar{x}$ be a desired temperature[1] and let $x^+ = \bar{x} + \varepsilon$ and $x^- = \bar{x} - \varepsilon$, where $\varepsilon > 0$ is a tolerance. Consider the system

$$\dot{x}(t) = -\alpha x(t) + q(t)u(t) + u_0$$

$$q(t) = \begin{cases} 0 \text{ if } x \geq x^+ \\ 1 \text{ if } x \leq x^- \\ 0 \text{ if } x^- < x(t) < x^+ \text{ and } q(t^-) = 0 \\ 1 \text{ if } x^- < x(t) < x^+ \text{ and } q(t^-) = 1 \end{cases}$$

with $\alpha > 0$, u_0 is a constant signal representing the heat introduced by the external environment, while $u(t)$ is supplied electrically. Assume for brevity that $u(t) = \bar{u}$ is constant, so that the oven works only in on–off mode. This is perhaps one of the most popular control systems. The oven works properly if the interval characterized by the extremal steady state temperatures $\bar{x}_{max} = (\bar{u} + u_0)/\alpha$ and $\bar{x}_{min} = u_0/\alpha$ includes the interval $[x^-, x^+]$ in its interior

$$\bar{x}_{min} < x^- < x^+ < \bar{x}_{max}$$

We assume that this condition is granted. We do not analyze the system, which is an elementary exercise and leads to the conclusion that the behavior is that in Figure 9.1, but we just point out some facts. If we set $\varepsilon = 0$, the temperature reaches in finite time the desired temperature. However, due to the discontinuity, we have (in theory) infinite-frequency switching between $q = 0$ and $q = 1$. This is not suitable for the application. First of all infinite frequency is not possible due to the

[1] Around 180 Celsius degrees for a Plum-Cake.

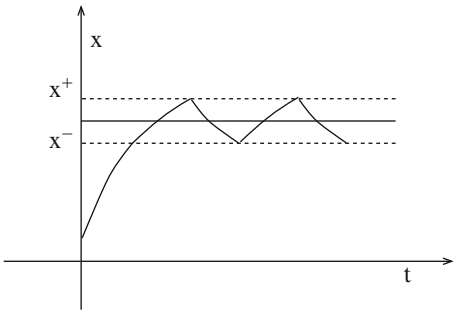


Fig. 9.1 The oven switching behavior

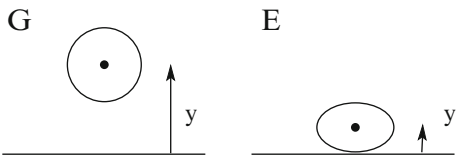


Fig. 9.2 The bouncing ball: G gravitational phase, E elastic phase.

hysteresis of real switches. Second, an hysteresis is typically introduced to avoid high frequency commutation and this is actually the reason why the region $[x^-, x^+]$ is introduced along with a discrete-dynamics for the system.

In the previous example, the logic part has its own dynamics, in the sense that in the intermediate region $[x^-, x^+]$ the logic variable $q(t)$ is not uniquely determined, but depends on its own value $q(t^-)$. We let the reader note that dynamic systems in which some logic variables depend on the state (or on an output) variable have already been encountered in the present book and are those obtained by the implementation of the Gutman and Cwikel control (see (4.39) at page 159) [GC86a, Bla99], which produces a closed-loop system which is piecewise-linear, hence characterized by a state-dependent switching.

Example 9.2 (Bouncing ball). To provide an example of purely state dependent switching, a simplified model of the bouncing ball is considered. The vertical motion of a ball bouncing on a plane has two phases (see Fig. 9.2). The gravitational one G in which the ball is subject to the gravity force and the elastic one E in which the ball is in contact with the plane. Denoting by r the radius of the ball, and denoting by $y(t)$ the level of the barycenter, it is possible to write two distinct equations:

$$\begin{array}{lll} \ddot{y}(t) = -g & \text{if} \quad y(t) > r & \text{Gravitational} \\ \ddot{y}(t) = -g + k(r - y(t)) & \text{if} \quad y(t) \leq r & \text{Elastic} \end{array} \tag{9.3}$$

This is the typical case in which the logic variable G, E is purely dependent on the output. This system is quite easy to analyze. First note that the system can be written as

$$\ddot{y}(t) = -g + k(y)(r - y(t))$$

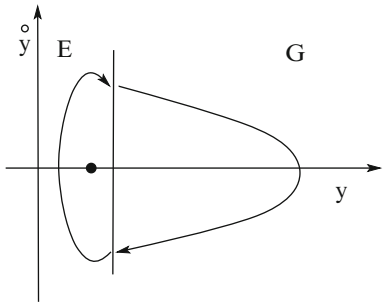


Fig. 9.3 The trajectories in the state space

where $k(y) = 0$ for $y > r$ and $k(y) = k$ for $y \leq r$. Indeed, since no dissipative terms were considered, its mechanical energy is preserved

$$\Psi(y, \dot{y}) = \frac{1}{2}\dot{y}^2 + yg + \frac{1}{2}k(y)(r - y)^2$$

(note that this is a differentiable function since $d/dy[k(y)(r - y)^2]$ is continuous). It is also immediately seen that this is a positive definite function having a unique minimum in the equilibrium $\bar{y} = r - g/k$, $\dot{y} = 0$. An example of system trajectory is reported in Fig. 9.3.

The previous example falls in the category of piecewise affine systems. A piecewise affine system has the form

$$\dot{x}(t) = A_i x(t) + b_i, \quad \text{for} \quad x \in \mathcal{S}_i$$

where the family of sets $\mathcal{S}_i$ forms a partition of the state space. An interesting case is that in which the sets $\mathcal{S}_i$ are simplices [HvS04, BR06, Bro10]. Simplices have the nice property of being quite flexible to reasonably cover non-trivial regions in the state space. Moreover, a piecewise linear function defined on a simplex is uniquely identified by the values at the vertices. So, if a region is covered by simplices having pairwise n vertices in common, it is possible to define a continuous function by choosing the value at the vertices. We will use this property later, applied to the relatively optimal control technique.

The next scholastic example aims at illustrating that even the analysis of simple hybrid systems is a tackling problem.

Example 9.3. Consider the discrete-time hybrid system

$$x(t + 1) = A_q x(t)$$

with $q \in \mathcal{Q} = \{1, 2\}$ and

$$A_1 = \begin{bmatrix} 0.8 & 0.9 \\ 0 & 0.9 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0.8 & 0 \\ 1 & 0.9 \end{bmatrix}$$

whose generating matrices are both Schur stable, having both eigenvalues inside the unit disc.

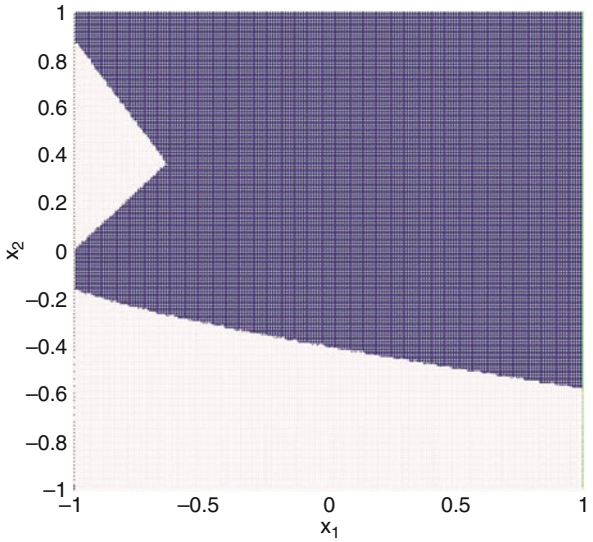


Fig. 9.4 Initial condition set for the Example 9.3

If the equation for the discrete variable q is

$$q(t) = \begin{cases} 1, & \text{if } \begin{bmatrix} 1 & -1 \end{bmatrix} x > 1 \\ 2, & \text{if } \begin{bmatrix} 1 & -1 \end{bmatrix} x \leq 1 \end{cases}$$

the set of initial conditions $\|x_0\|_1 \leq 1$ starting from which the evolution is driven to zero is the dark area depicted in Figure 9.4, whereas if the discrete variable mapping is

$$q(t) = \begin{cases} 1, & \text{if } \begin{bmatrix} 1 & -0.2 \end{bmatrix} x > 1 \\ 2, & \text{if } \begin{bmatrix} 1 & -0.2 \end{bmatrix} x \leq 1 \end{cases}$$

the evolution starting from every initial condition $\|x(0)\|_1 < 1$ converges to the origin, as one can check by running the code

```
for i=1:1000
    if [1 -1]*x0>1 ([1 -.2]*x0>1 in the second case)
        x0=A1*x0
    else
        x0=A2*x0;
    end
end
```

for different values of $x(0)$ and evaluating the final value.

Another example of linear switching systems can be found in the framework of networked control systems

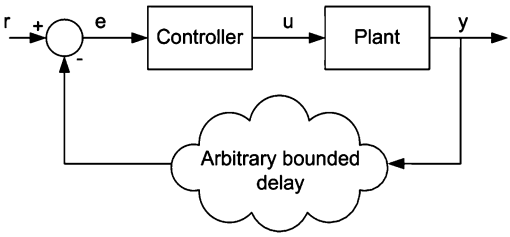


Fig. 9.5 Delay system

Example 9.4 (Networked control system). Consider the problem of controlling a strictly proper n-dimensional discrete-time linear time invariant (LTI) plant with $P = \{A, B, C\}$:

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t - q(t)) \end{cases} \tag{9.4}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t) \in \mathbb{R}^p$ and no delays or dropouts in the actuator channel, as depicted in Figure 9.5, are present. The system matrices can be thought of as obtained from a continuous-time plant controlled at a given sampling rate T_c. The controller clock is synchronized with that of the sensor and the transmitted data are time stamped, so that the sensor-to-controller delay $q(t) \in \{0, 1, \ldots, N_{max}\}$ is known. To recast such a system in a switching framework, first the system state is augmented so as to include delayed copies of the output, $y_i(t) = Cx(t - i)$, as

$$x_e(t) = \begin{bmatrix} x(t) \\ y_1(t) \\ \vdots \\ y_{N_{max}}(t) \end{bmatrix}$$

and a time-varying output matrix is introduced to get the dynamic system

$$\begin{aligned} x_e(t+1) &= \tilde{A}x_e(t) + \tilde{B}u(t) \\ \tilde{y}(t) &= \tilde{C}_{q(t)}x_e(t) \end{aligned} \tag{9.5}$$

where

$$\begin{aligned} \tilde{A} &= \begin{bmatrix} A & 0^{n\times(N_{max}-1)p} & 0^{n\times p} \\ C & 0^{p\times(N_{max}-1)p} & 0^{p\times p} \\ 0^{(N_{max}-1)p\times n} & I^{(N_{max}-1)p} & 0^{(N_{max}-1)\times p} \end{bmatrix} \\ \tilde{B} &= \begin{bmatrix} B \\ 0^{p\times m} \\ 0^{(N_{max}-1)p\times m} \end{bmatrix} \\ \tilde{C}_0 &= \begin{bmatrix} C\ 0^{p\times(N_{max}-1)p}\ 0^{p\times p} \end{bmatrix} \end{aligned} \tag{9.6}$$

and, for $1 \leq i \leq \bar{N}$

$$\tilde{C}_i = \left[0^{p\times(n+(i-1)p)} \; I^p \; 0^{p\times(N_{max}-i)p} \right]. \tag{9.7}$$

Note that when $q(t) = 0$ the augmented system output is nothing but the actual plant output, say $\tilde{y}(t) = y(t)$, whereas for $q(t) \geq 1$ the augmented system output is the $q(t)$ step delayed version of the plant output, say $\tilde{y}(t) = y(t - q(t))$.

In view of the above embedding, the problem of controlling the system in the presence of known delays in the sensing channel is recast into the problem of stabilizing the augmented switching system. This embedding is not new in this area and it has been widely used in conjunction with the theory of jump linear systems (see [ZSCH05, XHH00]).

We will come back to this problem later, in Section 9.6, when the stabilization problem of such a class of systems will be analyzed.

In the sequel, we do not consider hybrid systems in their most general form, but we rather consider special cases of systems that can switch (with controlled or uncontrolled switch) and the case in which the switching is state dependent, but the switching law is imposed by a feedback control. For a more general view, the reader is referred to specialized literature [Lib03, Joh03, SG05].

9.2 Switching and switched systems

Consider an autonomous system of the form

$$\dot{x}(t) = f(x(t), q(t)) \tag{9.8}$$

where $q(t) \in \mathcal{Q}$ and $\mathcal{Q}$ is a finite set. We consider two slightly (but substantially) different definitions.

Definition 9.5 (Switching system). The system is said to be *switching* if the signal $q(t)$ is not controlled but exogenously determined. This corresponds to the choice $q(t) = f_q(w(t)) \in \mathcal{Q}$ in (9.2).

Definition 9.6 (Switched system). The system is said to be *switched* if the signal $q(t)$ is controlled. This corresponds to the choice $q(t) = f_q(u(t)) \in \mathcal{Q}$ in (9.2).

9.3 Switching Systems

The analysis of a switching system is basically a robustness analysis problem already considered in the previous sections. In particular, a switching system is stable if and only if it admits a smooth Lyapunov function according to known

converse results [Me179, LSW96]. The special case of switching linear systems is included in the results in Subsection 7.3.2 (see Proposition 7.39), here reported in the switching framework for clarity of presentation.

Proposition 9.7. *The linear switching system*

$$\dot{x}(t) = A_{q(t)}x(t) \tag{9.9}$$

(or its discrete-time counterpart $x(t+1) = A_{q(t)}x(t)$) is asymptotically stable (equivalently exponentially stable) if and only if it admits a polyhedral norm as Lyapunov function.

The property holds also for norms of the type $\|Fx\|_{2p}$ [MP86a, MP86b, MP86c]. Note in particular that the stability of the switching system is equivalent to the stability of the corresponding polytopic system.

Theorem 9.8 ([MP86a, MP86b, MP86c]). *The stability of (9.9) under arbitrary switching is equivalent to the robust stability of the associated polytopic system*

$$\dot{x}(t) = \left[\sum_{q=1}^{r} \alpha_q(t)A_q(t)\right] x(t), \quad \sum_{q=1}^{r} \alpha_q(t) = 1, \quad \alpha_q(t) \geq 0 \tag{9.10}$$

As a consequence, the stability of each of the single systems $x(t+1) = A_q x(t)$, which is necessary for the stability of switching systems, is not sufficient [Lib03]. In view of the theorem, not even the Hurwitz stability of all the matrices (assumed constant) in the convex hull is sufficient for switching stability.

There exists a simple case in which the frozen-time Hurwitz stability assures switching, hence robust, stability.

Proposition 9.9. *Assume that all the matrices A_q are symmetric. Then the following conditions are equivalent.*

- *System (9.9) is stable under arbitrary switching.*
- *System (9.9) is robustly stable.*
- *All the matrices in conv$\{A_q,\ q = 1, \ldots, r\}$ are Hurwitz.*
- *All the matrices A_q, $q = 1, \ldots, r$, are Hurwitz.*
- *The systems is quadratically stable, i.e. A_q share a common quadratic Lyapunov function.*

Proof. See Exercise 4.

Example 9.10. Consider the two-tank system presented in Subsection 8.3.1, with state matrix

$$A(\xi, \eta) = \begin{bmatrix} -\xi & \xi \\ \xi & -(\xi + \eta) \end{bmatrix}$$

Assume that the parameters can change in an on–off mode $\xi \in \{\xi^-, \xi^+\}$ and $\eta \in \{\eta^-, \eta^+\}$ with all values strictly positive. It is immediate that for fixed values the system is asymptotically stable.

One might be interested in understanding whether this system can be destabilized by switching, say whether an inexpert[2] operator could destabilize the system by improperly changing the positions of the two valves which are on the duct between the two tanks and the duct after the second one (see Fig. 8.10).

The reply is obviously no, because the matrix is symmetric for any value of the parameters ξ and η and all the matrices are Hurwitz since the characteristic polynomial is $p(s) = s^2 + (2\xi + \eta)s + \xi\eta$.

There are quite a few cases of switching systems for which Hurwitz stability of the vertices implies the stability of all the matrices of the convex hull. One of such classes, specifically that formed by planar positive systems, will be discussed later.

9.3.1 Switching systems: switching sequences and dwell time

The concept of dwell-time, say the time interval during which no transitions of a switching system can occur (say the minimum amount of time the system "rests" in a given configuration), originates from the simple idea that a switching system composed by asymptotically stable systems exhibits an asymptotically stable behavior if the time interval between two switchings is sufficiently large. Quite clearly, the main focus of the research is directed towards the determination of the minimum dwell time. The interested reader is referred to [GC06, Col09] and the excellent survey [SWM+07].

Definition 9.11 (Dwell time). The value $\tau > 0$ is said to be the dwell time if the time instants t_k and t_{k+1} in which two consecutive switchings occur, must be such that

$$t_{k+1} - t_k \geq \tau$$

It is clear that, if all the systems are stable, τ has a stabilizing role. It is also straightforward to see that, if the system is stable with dwell time τ_1 then it is stable for any dwell time $\tau_2 > \tau_1$. The following proposition holds.

Proposition 9.12. *For any switching linear system with generating asymptotically stable matrices $A_1, A_2, \ldots A_r$ there exists a value $\bar{\tau}$ such that for any $\tau \geq \bar{\tau}$ assumed as dwell time, the switching system is stable.*

Proof. We provide a very conservative value of τ but in a constructive way (an alternative determination of the minimum dwell time by means of quadratic

[2] Evil/idiot.

functions can be found in [GC05, GCB08]). Take a 0-symmetric polytope $\bar{\mathcal{V}}[X]$ which is a C-set and for each stable system compute an invariant ellipsoid $\mathcal{E}_k \subset \lambda\bar{\mathcal{V}}[X]$, $0 \leq \lambda < 1$. For each vertex x_i compute the minimum time necessary for the k-th system to reach $\mathcal{E}_k$

$$T_{ik} = \min\{t \geq 0 : e^{A_k t} x_i \in \mathcal{E}_k\}$$

Such an operation can be easily done by iterating on $t > 0$ by bisection and by applying an LP test for any trial. Then it is immediate to see that a possible value for $\bar{\tau}$ is

$$\bar{\tau} = \max_{i,k} T_{ik}$$

This can be seen by considering the following discrete-time system

$$x(t_{k+1}) = e^{A_q(t_{k+1}-t_k)} x(t_k)$$

defined on the switching time sequence t_k. This is a linear LPV system. It is immediate that the set $\bar{\mathcal{V}}[X]$ is λ-contractive and thus the Minkowski function associated with $\bar{\mathcal{V}}[X]$ is a Lyapunov function for this system, so that $\|x(t_k)\| \to 0$ as $k \to \infty$ and the system is globally stable.

9.4 Switched systems

The situation is completely different in the case of switched systems. The stabilization problem is that of choosing a feedback law

$$q(t) = \Phi(x(t), q(t^-))$$

such that the resulting system is stabilized. In general, determining the control law $\Phi(x(t), q(t^-))$ is hard even for linear switched system. There is a sufficient condition which provides a helpful tool. The basic assumption is that there exists a Lyapunov stable system in the convex hull of the points $f(x,q)$.

Theorem 9.13. *Assume there exists a (sufficiently regular) function $\bar{f}(x)$ such that*

$$\bar{f}(x) \in conv\{f(x,q),\ q \in \mathcal{Q}\},$$

where $\mathcal{Q} = \{1, 2, \ldots, r\}$, for which the system $\dot{x} = \bar{f}(x)$ admits a smooth Lyapunov function such that

$$\nabla\Psi(x)\bar{f}(x) \leq -\phi(\|x\|)$$

where $\phi(\|x\|)$ is a κ-function. Then there exists a stabilizing strategy which has the form

$$q = \Phi(x) = \arg\min_{q \in \mathcal{Q}} \nabla\Psi(x) f(x, q)$$

Proof. The proof of the theorem is very easy, since it is immediately seen that

$$\min_{p} \ \nabla\Psi(x) f(x, p) \leq \nabla\Psi(x) \bar{f}(x) \leq -\phi(\|x\|)$$

Note that the condition on the average system implies that the "bar" system has an equilibrium point in $x = 0$. This is not necessarily true for the individual systems. Moreover, this procedure is not just useful for stabilizing the systems, but it is quite efficient to speed up the convergence rate. Some results on the equilibria of switched systems and their stability are found in [BS04].

Corollary 9.14. *Assume there exists a (sufficiently regular) function $\bar{f}(x)$ such that*

$$\bar{f}(x) \in conv\{f(x, q),\ q \in \mathcal{Q}\}$$

where $\mathcal{Q} = \{1, 2, \ldots, r\}$ and $\bar{f}(0) = 0$ (with $f(0, q)$ arbitrary such that $\bar{f}(0) \in conv\{f(0, q)\}$) and there exists a smooth Lyapunov function $\Psi(x)$ such that

$$\nabla\Psi(x) \bar{f}(x) \leq -\beta\Psi(x)$$

Then the strategy

$$q = \Phi(x) = \arg\min_{q \in \mathcal{Q}} \nabla\Psi(x) f(x, q)$$

assures

$$\Psi(x(t)) \leq \Psi(x(0)) e^{-\beta t}$$

A typical example of application of this strategy is a system with quantized control, as shown in the next example. Here the main issue is not the system stabilization (the system is already stable), but that of speeding up the convergence.

Example 9.15. Consider the two-tank hydraulic system already considered in Subsection 8.3.1 and represented in Figures 8.9 and 8.10, whose equations are[3]

$$\dot{h}_1(t) = -\alpha\sqrt{h_1(t) - h_2(t)} + q(t)$$
$$\dot{h}_2(t) = \alpha\sqrt{h_1(t) - h_2(t)} - \beta\sqrt{h_2(t)}$$

[3] In Section 8.3.1, the linearized version was considered.

where α and β are positive parameters, $h_1(t)$ and $h_2(t)$ are water levels, and $q(t)$ is the incoming flow. The device works with a couple of identical on–off valves, so the possible values of the flow are

$$q(t) \in \{0; \bar{q}; 2\bar{q}\}$$

Let us consider the steady state associated with a single open valve. Then, by defining the variables $x_1(t) = h_1(t) - \bar{h}_1$, $x_2(t) = h_2(t) - \bar{h}_2$, $u(t) = q(t) - \bar{q}$, where $\bar{h}_1$, $\bar{h}_2$, and $\bar{q}$ are the steady-state water levels and the incoming flow satisfying the conditions

$$\bar{h}_1 = \left(\frac{\bar{q}}{\alpha}\right)^2 + \left(\frac{\bar{q}}{\beta}\right)^2, \quad \bar{h}_2 = \left(\frac{\bar{q}}{\beta}\right)^2,$$

the following equations

$$\dot{x}_1(t) = -\alpha\sqrt{x_1(t) + \bar{h}_1 - x_2(t) - \bar{h}_2} + \bar{q} + u(t)$$
$$\dot{x}_2(t) = \alpha\sqrt{x_1(t) + \bar{h}_1 - x_2(t) - \bar{h}_2} - \beta\sqrt{x_2(t) + \bar{h}_2}$$

are derived.

Let us now consider the candidate control Lyapunov function (computed via linearization)

$$\Psi(x) = \frac{1}{2}\left(x_1^2 + x_2^2\right)$$

The corresponding Lyapunov derivative for $x \in \mathcal{N}[\Psi, \bar{h}_2^2/2]$ (this value is chosen in such a way that the ball is included in the positive region for the true levels $\bar{h}_i + x_i$) is

$$\begin{aligned}
&\dot{\Psi}(x,u) = \\
&= \underbrace{-(x_1 - x_2)\left(\alpha\sqrt{x_1(t) + \bar{h}_1 - x_2(t) - \bar{h}_2} - \bar{q}\right) - x_2\left(\beta\sqrt{x_2(t) + \bar{h}_2} - \bar{q}\right)}_{\doteq \dot{\Psi}_N(x_1, x_2)} \\
&+ x_1 u \leq \dot{\Psi}_N(x_1, x_2) + x_1 u
\end{aligned}$$

Note that $\dot{\Psi}_N(x_1, x_2)$, the natural derivative achieved for $u = 0$, is negative definite. Indeed the term in the left brackets has the same sign of $x_1 - x_2$ and the term in the right brackets has the same sign of x_2, thus $\dot{\Psi}_N(x_1, x_2)$ is zero only for $x_1 - x_2 = 0$ and $x_2 = 0$ and negative elsewhere. This nonlinear system is hence naturally asymptotically stable. The control input admits three admissible values

$$u(t) \in \{-\bar{q}; 0; \bar{q}\}$$

which correspond to the three cases in which none, just one or both the switching valves are open.

Let us now pretend to be able to implement a continuous controller

$$u = -\kappa x_1(t),$$

a controller which clearly cannot be implemented because the device has no continuous flow regulation. Still, in terms of convergence, it is possible to do as good as this fictitious control, since the continuous control assures a decreasing rate

$$\dot{\Psi}(x,u) \leq \dot{\Psi}_N(x_1,x_2) - \kappa x_1^2$$

and the discontinuous control

$$u = -\bar{q}\, sgn(x_1)$$

provides a better decreasing rate when $|\kappa x_1| < \bar{q}$. Indeed, the *fictitious system* with the continuous control is included in the extremal systems achieved, respectively, with $u = -\bar{q}$ and $u = +\bar{q}$, at least in the set

$$|\kappa x_1| \leq \bar{q}$$

where the closed-loop with this bang–bang control achieves better performances in terms of convergence than the continuous closed-loop plant.

From a practical point of view, the discontinuous controller has to be implemented with a threshold. We actually consider the function

$$u = \begin{cases} \bar{q} & \text{if} \quad x_1 < -\varepsilon \\ 0 & \text{if} \quad -\varepsilon \leq x_1 \leq \varepsilon \\ -\bar{q} & \text{if} \quad x_1 \geq \varepsilon \end{cases}$$

In Figures 9.6 and 9.7 the experimental behavior is shown with $\varepsilon = 0.01$ and $\varepsilon = 0.03$, being the latter less subject to ripples as expected. Ripples are due to the real implementation and they cannot be reproduced via simulation. As a final comment, we point out that in the real system α and β may vary (depending on the pipe conditions), thus producing an offset. However, the considered control basically eliminates the offset on x_2, since $\bar{h}_2$ can be fixed. Conversely, under parameter variations, an offset on the first variable $x_1 \neq 0$ is possible.

9.4.1 Switched linear systems

The problem of stabilization of switched systems is hard even in the linear case. This is apparent from the current literature which clearly shows that even in special cases, such as the case of positive switched linear systems, there are no general results which allow for algorithms of a reasonable complexity.

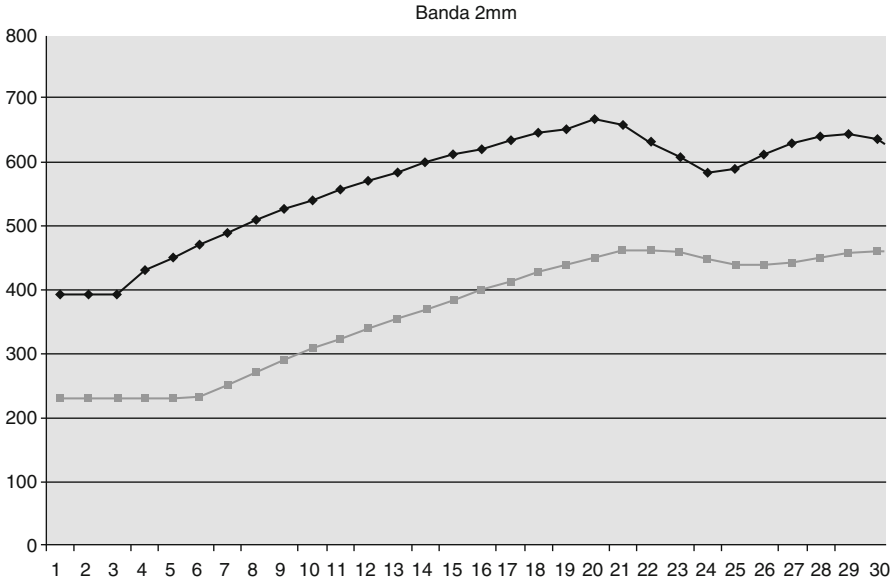


Fig. 9.6 The experimental behavior of the two-tank system with $\varepsilon = 0.01$

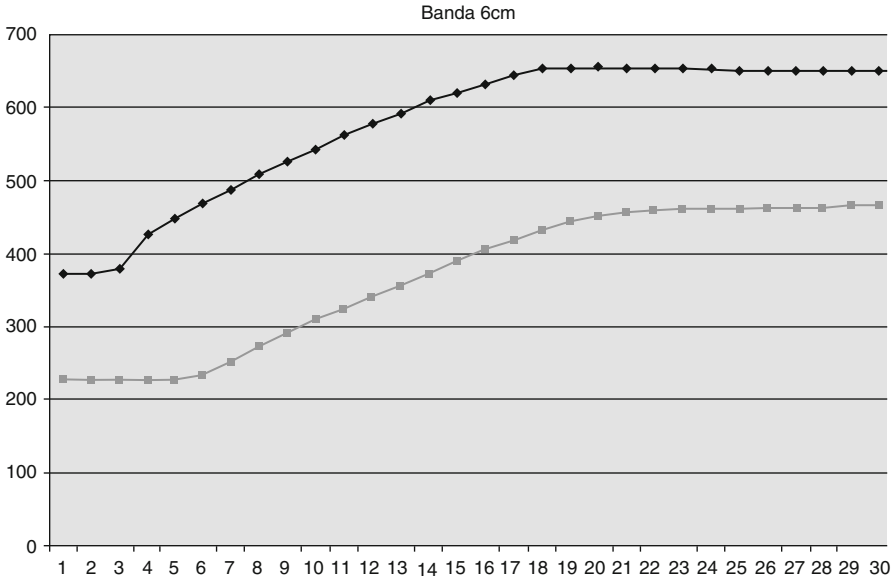


Fig. 9.7 The experimental behavior of the two-tank system with $\varepsilon = 0.03$

There are anyway important sufficient conditions which provide efficient but conservative solutions. For instance, in the case of a linear plant

$$\dot{x}(t) = A_{q(t)}x(t), \quad q \in \mathcal{Q} \tag{9.11}$$

$\mathcal{Q} = \{1, 2, \ldots, r\}$, the problem is easily solved if there exists

$$\tilde{A} \in conv\{A_i, \quad i = 1, 2, \ldots, r\}$$

which is asymptotically stable. Indeed, it is possible to consider any Lyapunov function for the system $\dot{x} = \tilde{A}x$, e.g. $\Psi(x) = x^T Px$, where P is a symmetric positive definite matrix such that

$$\tilde{A}^T P + P\tilde{A} \preceq -Q, \quad Q \succ 0$$

and choose as switched law

$$\Phi(x) = \arg\min_i \ \dot{\Psi}(x) = \arg\min_i \ x^T PA_i x$$

which insures the condition

$$\dot{\Psi}(x) \leq -x^T Qx$$

Unfortunately, the existence of such a stable element in the convex hull, besides being hard to check, is not necessary. For instance, the system given by the pair of matrices

$$A(w) = \begin{bmatrix} 0 & 1 \\ -1+w & -a \end{bmatrix}$$

where $w = \pm\bar{w}$ is a switched parameter and $a < 0$, is not stable for any value of a. However, if a is small enough, then there exists a suitable stabilizing strategy [Lib03]. Another example will be discussed later on, in the context of positive switched systems.

As far as the stabilizability of switched systems is concerned, given the designer choice of the switching rule and/or control law, three possible definitions can be considered.

Definition 9.16. System (9.11) is

- consistently stabilizable if there exists a sequence $q(t)$ such that the corresponding linear time-varying system $\dot{x}(t) = A_{q(t)}x(t)$ is asymptotically stable[4];
- open-loop stabilizable if, for any $\mu > \epsilon > 0$, there exists $T > 0$ such that, for all initial states $\|x(0)\| \leq \mu$, there exists a specific switching sequence $q(t)$ (depending on $x(0)$) assuring: $\|x(t)\| \leq \epsilon$, for $t \geq T$;
- feedback stabilizable if there exists a closed-loop strategy

$$q(t) = \Phi(x(t), t, q(t^-))$$

such that the corresponding (nonlinear discontinuous) system is globally uniformly asymptotically stable.

[4]The same $q(t)$ insures that $x(t)$ converges to 0 for all $x(0)$.

It is not difficult to see that open-loop and feedback stabilizability are equivalent. Clearly, if a system is consistently stabilizable, then it is open-loop and feedback stabilizable. The opposite is not true in general [SG11]. An example will be given in Section 9.7.2.

The difficulties in dealing with the stabilization problem for switched systems can be explained, in a Lyapunov framework, by the absence of convex and even smooth Control Lyapunov functions. Indeed, in general, the best "regularity property" which one can ask to a control Lyapunov function for a stabilizable linear system is homogeneity. The interest in such a property derives from the following fundamental result, whose proof can be found in the recent book [SG11].

Theorem 9.17. *Assume that the system $\dot{x}(t) = A_q x(t)$ (or $x(t+1) = A_q x(t)$) is closed-loop stabilizable. Then it necessarily admits a Lyapunov function which is positively homogeneous of order 1.*

To give an idea of how this function can be defined, consider the following set-theoretic considerations. First we need the following technical proposition.

Proposition 9.18. *If (9.11) is feedback (or open-loop) stabilizable, then also the following perturbed version*

$$\dot{x}(t) = [\beta I + A_{q(t)}]x(t), \quad q = 1, 2, \ldots, r \tag{9.12}$$

$q \in \mathcal{Q}$, is stabilizable for $\beta > 0$ small enough.

Proof. We have to notice that for the same initial condition $x(0)$ and the same q, the solution $x(t)$ of the unperturbed system and the solution $x_\beta(t)$ of the perturbed one (9.12) are related as

$$x_\beta(t) = e^{\beta t} x(t)$$

Indeed, for a given $q(t)$, (9.12) is just a linear time-varying system.

Assume that there exists a feedback stabilizing strategy q. Then for all $\mu > 0$ there exists $T > 0$ such that $\|x(T)\| \leq \mu/4$, for all $\|x(0)\| \leq \mu$. For $\beta > 0$ small, we also have $\|x_\beta(T)\| = \|e^{\beta T} x(T)\| \leq \mu/2$.

If we can drive all $\|x(0)\| \leq \mu$ in the ball $\|x_\beta(T)\| \leq \mu/2$, by linearity we have with similar arguments $\|x_\beta(2T)\| \leq \mu/4$, $\|x_\beta(3T)\| \leq \mu/8$, $\|x_\beta(kT)\| \leq \mu/(2^k)$.

This also means, in passing, that if the system can be stabilized, then it can be exponentially stabilized. If the system is uniformly stabilizable, then we can find a control Lyapunov function of the form

$$\Psi(x_0) = \inf_{q(\cdot)} \int_0^\infty \|x_\beta(q(t), x_0)\| \, dt < \infty$$

where we denoted by $x_\beta(q(\cdot), x_0)$ the solution of (9.12) corresponding to the initial condition x_0 and sequence $q(\cdot)$. Function Ψ is well defined if the system is

stabilizable. It is clearly positive definite. By linearity it is immediate that $\Psi(x_0)$ is positively homogeneous of order one, continuous and 0 symmetric. Such a function is non-increasing for (9.12) and therefore strictly decreasing for the original system (9.11) if a proper feedback switching law is applied. Note that $\Psi(x_0)$ is defined by taking the infimum over open-loop sequences, which proves that open-loop stabilizability implies feedback global stabilizability (the opposite is obviously true).

Unfortunately, we cannot go much further beyond homogeneity. The next negative result [BS08], which is in contrast with Proposition 9.7, shows a first headache in the stabilization of switched systems: convexity is not assured.

Proposition 9.19. *There are linear switched systems $\dot{x}(t) = A_{q(t)}x(t)$ (or $x(t+1) = A_{q(t)}x(t)$) which are stabilizable by means of a switching feedback control law but do not admit convex control Lyapunov functions.*

Example 9.20. Consider the system

$$\dot{x}(t) = A_{i(x(t))}x(t) \quad i \in \mathcal{I} = \{1, 2\}$$

where A_1 and A_2 are unstable matrices given by

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} \gamma & -1 \\ 1 & \gamma \end{bmatrix}$$

with $\gamma > 1$. Given the initial condition $x(0)$, the system trajectory is $x(t) = e^{A_i t}x(0)$, $i = i(x)$ (see Fig. 9.8), where

$$e^{A_1 t} = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \text{ and } e^{A_2 t} = e^{\gamma t} \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}$$

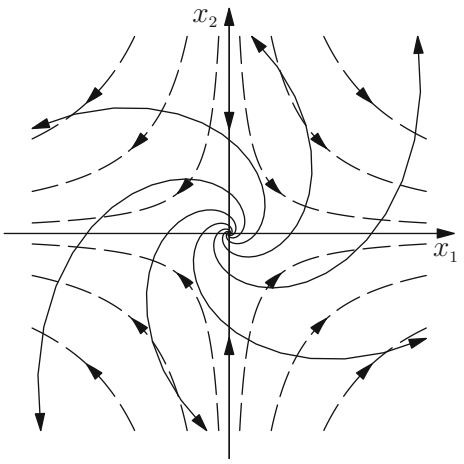


Fig. 9.8 Possible trajectories for dynamics $i = 1$ (dashed) and dynamics $i = 2$ (plain).

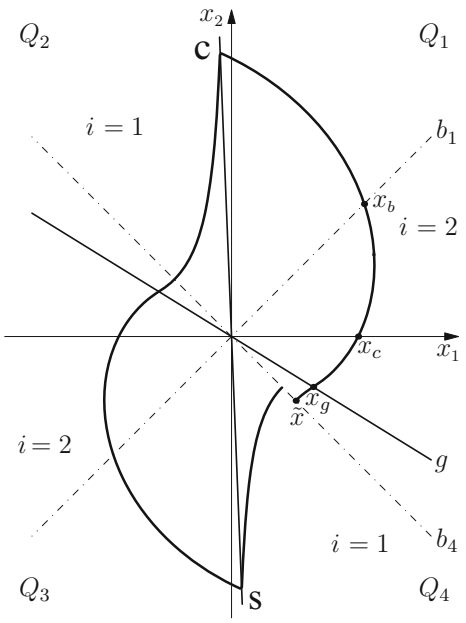


Fig. 9.9 Trajectory of the stabilized continuous-time system starting from the initial state $\tilde{x}$.

It is rather intuitive that this system is stabilizable. Indeed, when the state is of the form $(0, x_2)$ and the dynamics $i = 1$ is active, the state converges to 0. On the other hand, the state can be always driven to the x_2-axis, from every initial condition by activating the dynamics $i = 2$. Using this idea without any modification leads to the construction of a switching law which activates the dynamics $i = 1$ in a conic sector with empty interior. Such a strategy is not robust with respect to switching delays. This is not a problem, because we can take a line (line c in Fig. 9.9) originating at zero and sufficiently close to the x_2 axis. We also take a line g (see again Fig. 9.9) in which the derivatives of the two motions $A_1 x_g$ and $A_2 x_g$ are aligned.

Then a suitable strategy is $i = 1$ in sector s–0–g and $i = 2$ in sector g–0–c. Consider the point $\tilde{x}$ in sector s–0–g where $i = 1$. We let the trajectory reach line g in x_g to commute to $i = 2$, rotate counterclockwise (vector x_b) and reach line c. The we commute again to $i = 1$. Then again we reach line g, to commute to $i = 2$, we reach line s again and again we commute to $i = 1$. Then we will reach the same line aligned with the initial state $\tilde{x}$ (line b_4). It is rather intuitive that, if s is close enough to the x_2 axis, the motion with $i = 1$ reduces the norm of the state, so the state returns on line b_4 with a reduced norm. Then the system repeats the same trajectory, eventually converging to 0.

On the other hand, the following happens. Given a convex compact set $\mathcal{X}_0$ including 0, assume that a stabilizing control strategy $q(x(t))$ is given (as the one proposed before). Define as $\mathcal{R}(T)$ the set of all states $x(t)$ which are reached from $\mathcal{X}_0$ in time $0 < t \leq T$. The following result, which can be found in [BS08], holds.

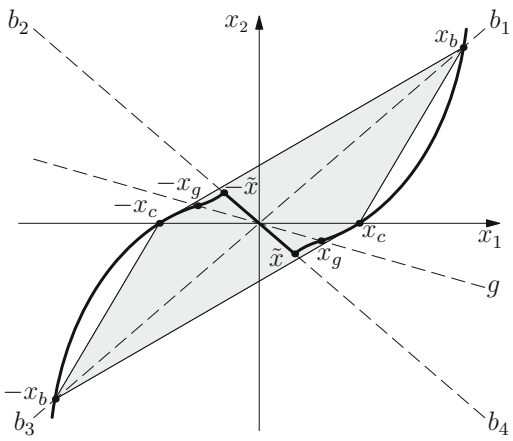


Fig. 9.10 The set $\mathcal{R}(T)$ includes $\mathcal{X}_0$.

Proposition 9.21. *If there exists a convex Lyapunov function, then the following condition is false for $T > 0$:*

$$\mathcal{X}_0 \subset conv\{\mathcal{R}(T)\}$$

where $conv\{\mathcal{R}(T)\}$ is the convex hull.

The non-existence of a convex Lyapunov function can be understood by reconsidering Example 9.20 and looking at Figure 9.10. It can indeed be seen that necessarily, if one takes as $\mathcal{X}_0$ the segment $-\tilde{x}$–$\tilde{x}$ the condition $\mathcal{X}_0 \subset conv\{\mathcal{R}(T)\}$ is satisfied for γ large for some T, no matter how the stabilizing strategy is chosen. The reader is referred to [BS08] for further details and for a discrete-time counterexample.

Absence of convexity can be a problem and[5] we have to announce another negative result (see [BCV12] and [BCV13] for details).

Proposition 9.22. *There are linear switched systems $\dot{x}(t) = A_{q(t)}x(t)$ (or $x(t+1) = A_{q(t)}x(t)$) which are stabilizable under a switching control law but do not admit smooth (away from 0) positively homogeneous control Lyapunov functions.*

We will sketch a proof of the result in Subsection 9.5.3 about positive switched systems.

So essentially the previous negative results justify the use of non-convex non-smooth functions. In particular the class of minimum-type functions such as those considered in [HL08] in the quadratic case and defined as

$$\Psi(x) = \min_i \; x^T P_i x$$

where P_i are positive definite or positive semi-definite matrices, have been proved to be especially useful.

[5] Since troubles quite often come with friends.

As a final point, it is worth recalling that in the discrete-time case the existence of a Schur convex combination is not sufficient for stabilizability (see Exercise 6).

There are several attempts in the literature to deal with the stabilization problem of switched systems. Successful techniques have been proposed for planar systems [BC06, XA00]. More general methods have been proposed based on (non-convex) piecewise linear functions [Yfo10] and on (non-convex) piecewise quadratic Lyapunov functions [HL08, GCB08, DGD11]. Necessary and sufficient conditions based on non-convex piecewise linear Lyapunov functions for the stabilizability of discrete-time switched systems have been recently successfully proposed in [FJ14], where a numerical procedure is presented along with its computational issues. Explanations of the difficulties in terms of computational complexity can be found also in [VJ14] and the references therein. Again, the reader is referred to more specialized literature [LA09, SG11].

9.5 Switching and switched positive linear systems

This section focuses on positive switching and switched linear systems, represented by the equation

$$\dot{x}(t) = A_{q(t)}x(t), \qquad (\text{respectively } x(k+1) = A_{q(t)}x(k)) \tag{9.13}$$

where $q(t) \in \{1, 2, \ldots, r\}$ and A_q, $q = 1, 2, \ldots, r$ are Metzler matrices in continuous-time and non-negative matrices in discrete-time.

Clearly, these systems are a special case of linear switching/switched systems, and therefore have all the properties presented in the previous sections. Given the extra properties enjoyed by the class of LTI positive systems (for instance, the existence of the Perron–Frobenius eigenvalue), it is legitimate to ask whether these extra properties can be somehow helpful when dealing with switching/switched positive systems.

Some examples of positive switching/switched systems are now presented. To keep things slightly more general, in the following also positive linear systems equipped with a non-negative input as follows:

$$\dot{x}(t) = A_{q(t)}x(t) + B_{q(t)}u(t) \tag{9.14}$$

where B_q are non-negative matrices and $u(t)$ is a non-negative input, will be considered.

9.5.1 The fluid network model revisited

Consider the fluid network already considered in Section 4.5.7, Example 4.62, whose equations have the form (9.14), with state and input matrices (reported here for convenience)

$$A = \begin{bmatrix} -(\alpha_{12}+\beta_{31}) & \alpha_{12} & 0 & 0 \\ \alpha_{21} & -(\alpha_{21}+\alpha_{23}+\beta_{42}) & \alpha_{23} & 0 \\ \beta_{31} & \alpha_{32} & -(\alpha_{32}+\alpha_{34}+\beta_{03}) & \alpha_{34} \\ 0 & \beta_{42} & \alpha_{43} & -(\alpha_{43}+\beta_{40}) \end{bmatrix}$$

$$B = \begin{bmatrix} \beta_{10} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

We remind the reader that $\alpha_{ij} = \alpha_{ji}$. The network is depicted in Fig. 9.11 which shows that the fluid can flow in on–off mode. This means that all the coefficients α_{ij} and β_{ij} can instantaneously change values as

$$\alpha_{ij} \in \{\alpha_{ij}^-, \alpha_{ij}^+\}$$

and

$$\beta_{ij} \in \{\beta_{ij}^-, \beta_{ij}^+\}$$

with $\alpha_{ij}^+ > \alpha_{ij}^- > 0$ and $\beta_{ij}^+ > \beta_{ij}^- > 0$. According to our distinction, if the system is "switching," we have to consider the problem of assuring stability under arbitrary switching. Conversely, if the system is "switched," then we typically wish to guarantee closed-loop stability with a prescribed convergence rate β

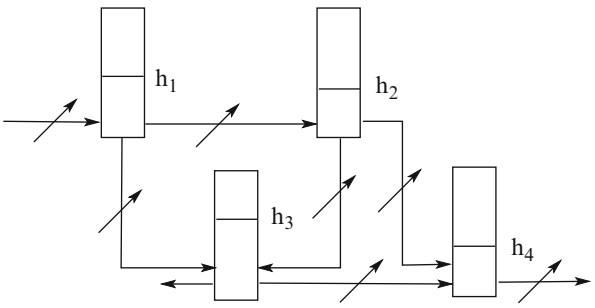


Fig. 9.11 The switched fluid network.

(or β-contractiveness). A final interesting question, at least from a practical point of view, concerns the possibility of confining the ultimate system evolution in a neighborhood of a desired equilibrium, given a certain input u.

To answer the above questions, the notion of co-positive Lyapunov functions is introduced next. The basic idea is that to solve these problems for positive systems it is possible to restrict our attention to the positive orthant only.

Definition 9.23. A function $\Psi(x)$, $x \in \mathbb{R}^n$, is co-positive if $\Psi(0) = 0$ and $\Psi(x) > 0$ for $x \geq 0$ and $x \neq 0$.

For instance, in $\mathbb{R}^2$, the function $\Psi(x_1, x_2) = x_1 + x_2^2$ is co-positive.

A co-positive function $\Psi(x)$ is a co-positive Lyapunov function for a positive systems if it is decreasing along the system trajectories. It is a weak co-positive Lyapunov function if it is non-increasing along the system trajectories. As expected, this condition is assured if the Lyapunov derivative is negative[6] in the positive orthant with the exception of 0. In the weak case, we just require the derivative to never be positive.

If we consider the fluid network example just reported, it can be immediately seen that the first problem has an immediate solution. Indeed it is apparent that the system matrix, in view of the symmetry assumption $\alpha_{ij} = \alpha_{ji}$, is weakly column diagonally dominant. Since it is irreducible (see Definition 4.59), we can render it diagonally dominant by using a diagonal state-transformation

$$D = \text{diag}\{\lambda, \lambda, 1, 1\}$$

with

$$\frac{\alpha_{23}^{+}}{\alpha_{23}^{+} + \beta_{03}^{-}} < \lambda < 1$$

(it will be shown soon that the lower bound is chosen in such a way to assure dominance) to get $\hat{A} = D^{-1}AD$ and $\hat{B} = D^{-1}B$ as follows:

$$\hat{A} = \begin{bmatrix} -(\alpha_{12} + \beta_{31}) & \alpha_{12} & 0 & 0 \\ \alpha_{21} & -(\alpha_{21} + \alpha_{23} + \beta_{42}) & \alpha_{23}/\lambda & 0 \\ \lambda\beta_{31} & \lambda\alpha_{32} & -(\alpha_{32} + \alpha_{34} + \beta_{03}) & \alpha_{34} \\ 0 & \lambda\beta_{42} & \alpha_{43} & -(\alpha_{43} + \beta_{04}) \end{bmatrix}$$

and $\hat{B} = [\, \beta_{10}\ 0\ 0\ 0\,]^T$.

Set $z = D^{-1}x$ and consider the co-positive function

$$\Psi(z) = \bar{1}^T z.$$

[6] We spare the reader the term "co-negative."

Its Lyapunov derivative is

$$\dot{\Psi}(z) = \bar{1}^T(\hat{A}z + \hat{B}u) =$$
$$- \left[(1-\lambda)\beta_{31}z_1 + (1-\lambda)(\beta_{42}+\alpha_{32})z_2 + (\beta_{03}+\alpha_{32}(1-\frac{1}{\lambda}))z_3 + \beta_{40}z_4\right] + \frac{u\beta_{10}}{\lambda} \leq$$
$$- \left[(1-\lambda)\beta_{31}^{-}z_1 + (1-\lambda)(\beta_{42}^{-}+\alpha_{32}^{-})z_2 + (\beta_{03}^{-}+\alpha_{32}^{+}(1-\frac{1}{\lambda}))z_3 + \beta_{40}^{-}z_4\right] + \frac{u\beta_{10}}{\lambda} =$$
$$= -v^T z + \frac{u\beta_{10}}{\lambda}$$

with obvious meaning of the vector v. Let $\beta_{10} = 0$. Since λ is close enough to 1 and such that the critical term $\beta_{03}^{-} + \alpha_{32}^{+}(1-\frac{1}{\lambda}) > 0$, then $v > 0$ and therefore $\dot{\Psi} < 0$ for $z > 0$. This means that the system is asymptotically stable because:

$$\dot{\Psi}(z) \leq -\min\{v_j\} \sum z_j = -\min\{v_j\}\Psi(z)$$

For $\beta_{10} > 0$ and u bounded, the system solution is bounded. Indeed, if we take the plane

$$\Psi(z) = \bar{1}^T z = \mu$$

the derivative becomes negative for $\mu > 0$ large. Note that the Lyapunov function $\bar{1}^T z$ for the modified system corresponds to the Lyapunov function $\bar{1}^T D^{-1}x$ for the original one.

Let us consider the switched stabilization problem, which is quite interesting in this case. We could use the same Lyapunov function previously derived, but we propose a different idea. We wish to find a switching strategy to control the system in the sense that we wish to try to force the system, by switching, to stay at its lowest level, given a constant incoming flow $u = \text{const} \geq 0$, or to approach 0 as quickly as possible if the external input is $u = 0$.

In this case there is a simple solution which is shown next. Indeed the "average system" is weakly diagonally dominant and irreducible, hence asymptotically stable. Then we can find a linear co-positive Lyapunov function which can be used as a control Lyapunov function for the switched systems.

Denote by $\bar{A}$ the system corresponding to the average values

$$\alpha_{ij} = \frac{\alpha_{ij}^{-} + \alpha_{ij}^{+}}{2}$$

and

$$\beta_{ij} = \frac{\beta_{ij}^{-} + \beta_{ij}^{+}}{2}$$

Let z^T be the eigenvector associated with the Frobenius eigenvalue:

$$z^T \bar{A} = \lambda_F z^T$$

with $\lambda_F < 0$. As the Frobenius eigenvalue is negative, all the others have negative real parts. Again, the average system is fictitious and "cannot be realized." However, since it would render the derivative negative when $u = 0$ and it is in the convex hull of all the matrices, then at each x there exists one of such matrices (a vertex) which assures a smaller derivative.

Denoting by A_i any of the matrices achieved by taking α_{ij} and β_{ij} in all possible ways we have, in the case of a constant input $\bar{u} > 0$,

$$\min_i \; z^T A_i x + z^T B\bar{u} \leq z^T \bar{A} x + z^T B\bar{u} = \lambda_F z^T x + z^T B\bar{u}$$

Hence, considering the strategy

$$q(x) = \arg\min_i z^T A_i x$$

the Lyapunov derivative of the co-positive Lyapunov function $\Psi(x) = z^T x$ for the system $\dot{x} = A_{q(x)}x + B\bar{u}$ would be

$$D^+\Psi(x) \leq \lambda_F z^T x + z^T B\bar{u} = \lambda_F \Psi(x) + z^T B\bar{u}$$

hence implying ultimate boundedness of the system, since $\lambda_F < 0$. Therefore $D^+\Psi(x) < 0$ for every x such that

$$\Psi(x) < -\frac{z^T B}{\lambda_F}\bar{u}$$

Remark 9.24. The "arg-min" strategy (obviously) leads to a system of differential equations with discontinuous right-hand side. This typically introduces chattering and sliding modes in the system. As we will see later, introducing chattering is not necessarily the most convenient strategy.

9.5.2 Switching positive linear systems

In the case of switching systems, namely when the sequence is arbitrary, it is a legitimate question to ask whether positivity brings something good to the analysis of switching systems, in particular whether there is any advantage from the assumption that all the matrices are Metzler.

For instance, it has been shown that (Th. 9.8) switching robust stability is equivalent to the stability of the convex differential inclusion, which is in turn equivalent to the existence of a convex (in particular polyhedral) common Lyapunov function. Therefore it is necessary that all the elements in the convex hull are Hurwitz matrices. The condition is by no means sufficient, for switching linear systems, and an example will be provided later (Example 9.50).

On the other hand, positive systems have a "dominant" eigenvalue, the Frobenius one, which rules all the others. So a stronger results might be true.

Conjecture 9.25. For positive switching systems, the Hurwitz stability of all the matrices in the convex hull is necessary *and sufficient* for switching stability.

The conjecture is true for second order systems only [GSM07, FMC09].

Proposition 9.26. *For second order continuous-time switching systems, the Hurwitz stability of all the matrices in the convex hull is necessary* ***and sufficient*** *for asymptotic stability under arbitrary switching.*

Unfortunately, we cannot go much further. Indeed, in [FMC09], a third order counterexample is provided in which it is shown that the condition is not sufficient.

What about discrete-time? Even worse. Take $x(k+1) = A_i x(k)$, $i = 1, 2$, with

$$A_1 = \begin{bmatrix} 0 & 0 \\ (2-\epsilon) & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & (2-\epsilon) \\ 0 & 0 \end{bmatrix}$$

and $\epsilon > 0$ small enough. The characteristic polynomial of the matrices in the convex hull is

$$p(z) = z^2 - \alpha(1-\alpha)(2-\epsilon)^2$$

For $0 \leq \alpha \leq 1$, $\alpha(1-\alpha)(2-\epsilon)^2 < (2-\epsilon)^2/4 < 1$. Then all the matrices are Schur. It is an exercise to see that the product $(A_1 A_2)^k$ goes to infinity, thus the system is not stable under arbitrary switching.

From the computational side, one can have some advantages in the construction of polyhedral functions, introduced in Chapters 5 and 6, since the plane generating procedure would work with positive constraints only.

Let us consider the discrete-time case. We present a procedure similar to the polyhedral Lyapunov function generation algorithm described in Sections 5.4 and 6.3.3, for the computation of the joint spectral radius. Let us introduce a definition.

Definition 9.27. We call a *P*-set a set $\mathcal{S} \subset \mathbb{R}^n_+$ which

- is closed, bounded and includes the origin;
- is star-shaped in the positive orthant: any ray originating at zero and contained in the positive orthant encounters the boundary of $\mathcal{S}$ in a single non-zero point. In other words, for any (non-negative) vector $v \geq 0$, the intersection of the positive

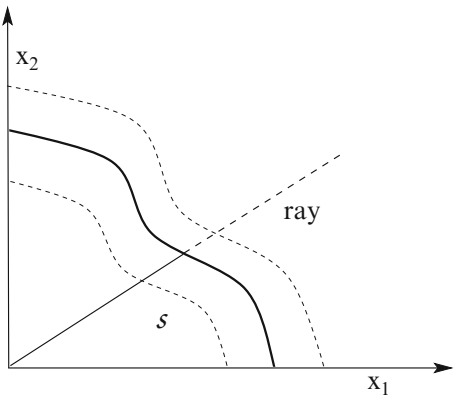


Fig. 9.12 Example of a *P*-set, and the corresponding "Minkowski-like" function

ray with direction v with $\mathcal{S}$ is a segment with one extremum in $x = 0$ and the other extremum on the boundary of $\mathcal{S}$:

$$\{\lambda : \ \lambda v \in \mathcal{S}\} = [0, \lambda_{max}(v)]$$

for some $\lambda_{max}(v) > 0$ (see Fig. 9.12).

Given any co-positive and positively homogeneous function its sub-level sets $\mathcal{N}[\Psi, \kappa]$ are *P*-sets. Conversely, given a *P*-set, we can define a co-positive function which is positively homogeneous, extending the concept of Minkowski functional

$$\Psi(x) = \inf \ \{\lambda > 0 : \ x \in \lambda \mathcal{S}\}$$

which is a co-positive function of order 1. Clearly we can generate co-positive homogeneous functions of any order by considering $\Psi(x)^p$.

Obviously a *P*-set can be convex. In this case the Minkowski-like functional would also be convex.

As a special case, we can consider polyhedral *P*-sets, which can be represented as

$$\mathcal{S} = \{x \geq 0 : Fx \leq \bar{1}\} = \mathcal{P}(F)$$

where $F \geq 0$ is a full column rank non-negative matrix.

First we remind that, to assure a certain speed of convergence λ, we have just to consider the modified system

$$x(t+1) = \frac{A_{q(t)}}{\lambda} x(t), \ \ q \in \{1, 2, \ldots r\}.$$

With the above in mind, it is possible to compute the largest invariant set starting from any arbitrary *P*-set by means of the procedure presented next.

Procedure: Computation of a co-positive Lyapunov function given a contraction factor $\lambda > 0$.

1. Take any arbitrary polyhedral P-set $\mathcal{S}$, associated with a non-negative full column rank matrix F_0. Set $\mathcal{P}_0 := \mathcal{P}(F_0) \bigcap \mathbb{R}^n_+$. Fix a tolerance $\epsilon > 0$, $k = 1$ and a maximum number of steps k_{max}.
2. Recursively compute $\mathcal{P}_k = \mathcal{P}(F_k) \bigcap \mathbb{R}^n_+$ as

$$\mathcal{P}_k = \{x : \ F_{k-1} \frac{A_i}{\lambda} x \leq \bar{1}, \ \ i = 1, 2, \ldots r\}$$

3. Remove all the redundant rows in matrix F_k to get $\mathcal{P}_k = \mathcal{P}(F_k) \bigcap \mathbb{R}^n_+$.
4. If $\mathcal{P}_k = \mathcal{P}_{k-1}$ STOP: the procedure is successful.
5. If $k = k_{max}$ or if $\epsilon \bar{1} \notin \mathcal{P}_k$ STOP: the procedure is unsuccessful for the given λ.
6. Set $k = k + 1$ and GO TO step 2.

If the procedure stops successfully, then the final set is the largest invariant set included in the initial one for the modified system (and the largest λ-contractive set for the original system).

Again, the previous considerations about the tolerance and maximum number of steps can be made. Note that, in the event that the sequence collapses, the failure can be just detected by the exclusion $\epsilon \bar{1} \notin \mathcal{S}_k$.

As a starting set, we can take a simple set, for instance $\mathcal{S} = \{x \geq 0 : \ Fx \leq 0\}$ for some row vector $F > 0$.

This method for finding polyhedral co-positive functions can be applied (needless to say) to continuous-time problems by means of the Euler Auxiliary System as described in Chapter 5. Note that, given a positive continuous-time system $\dot{x} = Ax$ with A Metzler, then $[I + \tau A]$ is a positive matrix provided that $\tau > 0$ is small enough. Similar techniques have been applied to prove structural boundedness of a class of biochemical networks [BG14].

Example 9.28 (Worst-case emptying speed). Assume that waste material is accumulated in two stock houses and it has to be eliminated. Assume that the system has two possible configurations, as in Fig. 9.13. In configuration A the waste material

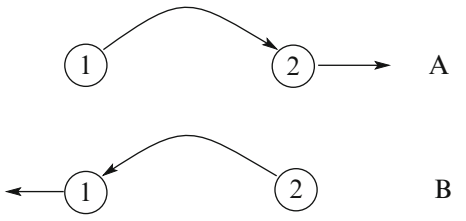


Fig. 9.13 The two-configuration emptying problem.

in node 1 is transferred to node 2 and then eliminated. Configuration B is the symmetric one. We assume a discrete-time model of the form

$$A_A = \begin{bmatrix} \beta & 0 \\ 1-\beta & \alpha \end{bmatrix}, \qquad A_B = \begin{bmatrix} \alpha & 1-\beta \\ 0 & \beta \end{bmatrix}$$

with $0 < \alpha, \beta < 1$. In any fixed configuration, the system would converge with a speed depending on the maximum eigenvalue $\max\{\alpha, \beta\}$. A legitimate question is whether switching between the two configurations can worsen the situation and how much.

If, for instance, we assume $\alpha = \beta = 0.8$, then the maximum eigenvalue is $\lambda = 0.8$, for both matrices and so, under arbitrary switching, the speed of convergence in the worst case cannot be smaller than $\lambda = 0.8$. Iterating over λ it is possible to compute numerically that the best contractivity factor, which is assured under arbitrary switching is around $\lambda^* \approx 0.92$. The reader can enjoy in Fig. 9.14 the maximal set computed for $\lambda = 0.94$ included in the region

$$\mathcal{S} = \{x_1, x_2 \geq 0 : \ x_1 + x_2 \leq 1\}$$

9.5.3 Switched positive linear systems

Perhaps the problem of switched positive systems is of more interest, because there are many situations in which this problem is encountered and positivity turns out to be an assumption under which some general interesting properties can be proved.

Consider the system

$$\dot{x}(t) = A_{q(t)} x(t) \tag{9.15}$$

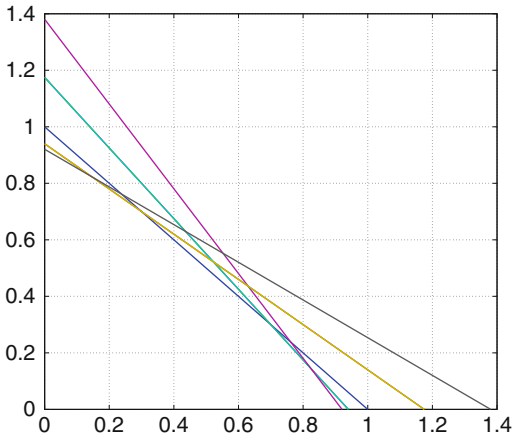


Fig. 9.14 The maximal invariant set for the modified system with $\lambda = 0.94$ (lower-left portion of the square)

where

$$q(t) \in \mathcal{Q} = \{1, 2, \dots, r\}$$

and the matrices A_q are all Metzler.

The basic question we consider is the existence of a feedback function

$$q(t) = \Phi(x, t)$$

ensuring that the state $x(t)$ converges to zero for any initial condition $x(0) \geq 0$.

Strange as it may seem, in choosing $q(t)$, the knowledge of $x(t)$ does not matter (see [FV12, BCV12] for details).

Theorem 9.29. *For a system of the form (9.15) with A_q Metzler the following conditions are equivalent.*

i) There exists $x_0 > 0$ and $q(t) \in \mathcal{Q}$ (an open-loop control depending on x_0) such that the trajectory starting from $x(0) = x_0$ converges to zero.
ii) There exists a feedback law $q(t) = \Phi(x(t), t)$ such that the trajectory starting from any $x(0) \geq 0$ converges to zero.
iii) The switched system is consistently stabilizable (i.e., there exists a single $q(t) \in \mathcal{Q}$ which drives $x(t)$ to zero from any initial condition, not necessarily positive).

Proof. ii) $\Rightarrow$ i): Clearly, if there exists a stabilizing closed-loop $q(t) = \Phi(t, x)$, given $x_0 > 0$ there exists an open-loop sequence which drives the state to 0 starting from $x(0) = x_0$.

i) $\Rightarrow$ iii): (this result is in [FV12]). Assume that, given $\bar{x}_0 > 0$, there exists a switching function $q(t)$ such that for $\bar{x}(0) = \bar{x}_0$, the solution $\bar{x}(t)$ converges to zero. Let $x(0)$ be any initial condition such that $x(0) \leq \bar{x}_0$ and let $x(t)$ be the corresponding solution. Since $q(t)$ is fixed, both $x(t)$ and $\bar{x}(t)$ are solutions and their difference $\bar{x}(t) - x(t) \doteq z(t)$ satisfies

$$\dot{z}(t) = A_{q(t)} z(t)$$

where $A_{q(t)}$ is Metzler, so the system is positive. Since by construction $z(0) = \bar{x}(0) - x(0) \geq 0$, the condition $z(t) \geq 0$ is preserved. Hence $x(t) \leq \bar{x}(t)$ for all $t > 0$. Now take the symmetric initial state $-\bar{x}_0$. For the given $q(t)$ the solution is $-\bar{x}(t)$ which goes to zero. Exactly in the same way, one can show that if $x(0) \geq -\bar{x}_0$ then $x(t) \geq -\bar{x}(t)$. Then, for $-\bar{x}_0 \leq x(0) \leq \bar{x}_0$, we have (see Fig. 9.15)

$$-\bar{x}(t) \leq x(t) \leq \bar{x}(t)$$

Therefore all the initial states in the box $\bar{\mathcal{P}}[I, \bar{x}_0]$ are driven to zero. Since the box includes zero in its interior, and the system is linear, any initial state can be included in the box $\lambda\bar{\mathcal{P}}[I, \bar{x}_0]$, for $\lambda > 0$ large enough, so $q(t)$ drives all states to zero. This proves iii).

iii) $\Rightarrow$ ii): obvious.

The previous theorem admits a corollary (see [SG05] for further details).

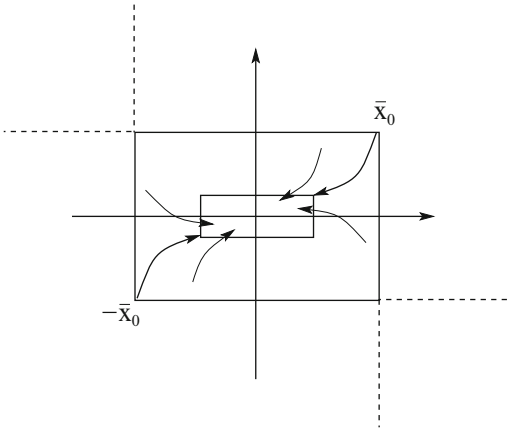


Fig. 9.15 The idea of the proof: all solutions are bounded below and above from those originating between $-\bar{x}_0$ and $\bar{x}_0$

Corollary 9.30. *If any of the equivalent conditions of Theorem 9.29 holds, then there exists a periodic sequence $q_p(t)$, with period T large enough, such that any initial state is driven to* 0.

Proof. Consider a function $q(t)$ driving the state trajectory $\bar{x}(t)$ from $\bar{x}_0$ to zero. Take a positive $\lambda < 1$ and $T > 0$ such that the solution $\bar{x}(t)$ is in the box $\lambda\bar{\mathcal{P}}[I, \bar{x}_0]$ at time $t = T$, namely $\bar{x}(T) \in \bar{\mathcal{P}}[I, \lambda\bar{x}_0]$. So for any initial state in $\bar{\mathcal{P}}[I, \bar{x}_0]$ ($x(0) \leq \bar{x}_0$) we have

$$x(T) \in \lambda\bar{\mathcal{P}}[I, \bar{x}_0]$$

Truncate function q and extend it periodically with period T. The next period we will have $x(2T) \in \lambda^2\bar{\mathcal{P}}[I, \bar{x}_0]$ and in general

$$x(kT) \in \lambda^k\bar{\mathcal{P}}[I, \bar{x}_0]$$

Since $\lambda < 1$, this means that $x(t) \to 0$.

The previous results hold, without changes, in discrete-time.

For clear reasons, it is important to find a feedback solution to the problem anyway. We have seen that in general, for switched linear systems, convex control Lyapunov function may not exist, even if the system is stabilizable. A natural question is whether there exists a class of Lyapunov functions which are universal for the problem. The following theorem provides an answer [HVCMB11] and tells us that, surprisingly, for positive linear systems, as long as we stay in the positive orthant, we can always find concave control Lyapunov functions.

Theorem 9.31. *Assume that a positive switched linear system is stabilizable. Then there exists a concave co-positive control Lyapunov function, positively homogeneous of order one.*

Proof. Consider the perturbed system

$$\dot{x}(t) = [\beta I + A_q]x(t) = A_{\beta,q}x(t)$$

for $\beta > 0$ small enough. We remind that, for the same initial condition and the same q the solutions of the unperturbed system $x(t)$ and of the perturbed one $x_\beta(t)$ are related as (see Proposition 9.18) $x_\beta(t) = e^{\beta t}x(t)$ and that for β small enough the system remains stabilizable if the original system is such.

Denote by $x_{\beta,q}(t, x_0)$ the solution with initial condition x_0, corresponding to a switching function $q(\cdot)$.

Consider for the modified system the following function:

$$\Psi_\beta(x_0) = \inf_q \int_0^\infty \bar{1}^T x_{\beta,q}(t, x_0)dt$$

which is well defined since we assumed stabilizability.

To prove concavity, consider $x_0 = \alpha_1 x_1 + \alpha_2 x_2$, $\alpha_1 + \alpha_2 = 1$, $\alpha_1, \alpha_2 \geq 0$. For any q we have

$$\int_0^\infty \bar{1}^T x_{\beta,q}(t, x_0) = \alpha_1 \int_0^\infty \bar{1}^T x_{\beta,q}(t, x_1)dt + \alpha_2 \int_0^\infty \bar{1}^T x_{\beta,q}(t, x_2)dt$$

hence

$$\begin{aligned}
\Psi_\beta(x_0) = & \inf_q \int_0^\infty \bar{1}^T x_{\beta,q}(t, x_0) = \\
& = \inf_q \left[\alpha_1 \int_0^\infty \bar{1}^T x_{\beta,q}(t, x_1)dt + \alpha_2 \int_0^\infty \bar{1}^T x_{\beta,q}(t, x_2)dt\right] \\
& \geq \alpha_1 \left[\inf_{q_1} \int_0^\infty \bar{1}^T x_{\beta,q_1}(t, x_1)dt\right] + \alpha_2 \left[\inf_{q_2} \int_0^\infty \bar{1}^T x_{\beta,q_2}(t, x_2)dt\right] \\
& = \alpha_1 \Psi_\beta(x_1) + \alpha_2 \Psi_\beta(x_2)
\end{aligned}$$

and therefore we have that $\Psi_\beta(x_0)$ is concave. It is obviously positively homogeneous of order one.

Consider the directional derivative

$$D^+\Psi(x, [\beta I + A]x) = \lim_{h \to 0^+} \frac{\Psi(x + h[\beta I + A]x) - \Psi(x)}{h}$$

and any interval $[0, \tau]$. By applying dynamic programming considerations, we have

$$\Psi_\beta(x_0) = \inf_q \int_0^\tau \bar{1}^T x_{\beta,q}(t, x_0)dt + \Psi_\beta(x_\beta(\tau))$$

Then, for all $\tau > 0$,

$$\Psi_\beta(x_0) > \Psi_\beta(x_\beta(\tau))$$

As a consequence we have

$$D^+\Psi(x, [\beta I + A]x) \leq 0$$

Let us denote by $\eta = h/(1-\beta h)$ (then $h \to 0$ implies $\eta \to 0$) and let us bear in mind that $\Psi(\lambda x) = \lambda\Psi(x)$. Now consider for the nominal system

$$\begin{aligned} D^+\Psi(x) &= \lim_{h\to 0^+} \frac{\Psi(x+hAx)-\Psi(x)}{h} = \lim_{h\to 0^+} \frac{\Psi(x-h\beta x)-\Psi(x)}{h} \\ &+ \lim_{h\to 0^+} \frac{\Psi(x+hAx+h\beta x-h\beta x)-\Psi(x-h\beta x)}{h/(1-\beta h)} \frac{1}{(1-\beta h)} \\ &= -\beta\Psi(x) + \underbrace{\lim_{\eta\to 0^+} \frac{\Psi(x+\eta[\beta I+A]x)-\Psi(x)}{\eta}}_{\leq 0} \leq -\beta\Psi(x) \end{aligned}$$

and the proof is completed.

One could at this point try a conjecture. We have seen that, in the switching system case, a convex Lyapunov function can be smoothed. Then

Conjecture: for a positive switched linear system, stabilizability implies the existence of a smooth concave positively homogeneous Lyapunov function.

The conjecture is false, unless we take $n = 2$. Precisely, we can claim the following.

Theorem 9.32. *Assume that the matrices A_i $i = 1, 2, \ldots, r$ are irreducible. Then the following statements are equivalent.*

i) The system is stabilizable and admits a co-positive and positively homogeneous smooth control Lyapunov function.
ii) There exists a matrix $\bar{A}$ in the convex hull of the A_i, $\bar{A} \in conv\{A_q,\ q = 1, \ldots, r\}$, which is Hurwitz.
iii) The system admits a linear co-positive control Lyapunov function $\Psi(x) = z^T x$, with $z > 0$.

Proof. ii) $\Rightarrow$ iii) If there exists an Hurwitz matrix in the convex hull, then we can take its Frobenius left eigenvector z and $\lambda z^T = z\bar{A}$, with $\lambda < 0$ and $z > 0$. Note that $\bar{A}$ is irreducible, if the A_i are such. Then, for all $x \geq 0$

$$\min_i \ z^T A_i x \leq \lambda z^T \bar{A} x = \lambda z^T x < 0$$

Then $\Psi(x) \doteq z^T x$ is a co-positive control Lyapunov function.

iii) $\Rightarrow$ i) Obviously, since a co-positive linear function is smooth and positively homogeneous.

i) $\Rightarrow$ ii) See [BCV12].

Proposition 9.33. *There exist Metzler matrices A_1, A_2, $\dots A_r$ for which the corresponding switched system is stabilizable but there are no positively homogeneous co-positive Lyapunov functions which are continuously differentiable.*

Proof. In view of implication i) $\Rightarrow$ ii) in Theorem 9.32 it is sufficient to show that there exist stabilizable positive switched systems which do not include Hurwitz matrices in the convex hull. This will be shown in Example 9.34.

This is not good news as previously announced. Since positive linear systems are a special case of linear systems, the existence of a positively homogeneous smooth Lyapunov function would imply the existence of a co-positive Lyapunov function if we restrict our attention to the positive orthant. Then Proposition 9.33 implies Proposition 9.22. The proof of the result is in [BCV12] and a different and more "detailed" one is in [BCV13].

The following example motivates the analysis and proves Proposition 9.33.

Example 9.34. Consider a traffic control problem in a junction. Assume that there are three main roads (A, B, and C) converging into a "triangular connection" governed by traffic lights. Three buffer variables, x_1, x_2, and x_3, represent the number of vehicles waiting at the three traffic lights inside the triangular loop. We assume that there are three symmetric configurations as far as the states of the 6 traffic lights are concerned. In the first configuration, described in Fig. 9.16, we assume that traffic lights corresponding to x_1, x_2, B and C are green, while the ones corresponding to x_3 and A are red. Accordingly,

- x_3 increases proportionally ($\beta > 0$) to x_2;
- x_2 remains approximately constant, receiving inflow from B and buffer x_1, while giving outflow to A and to buffer x_3;
- x_1 decays exponentially ($-\gamma < 0$), since the inflow from C goes all to x_2 and B.

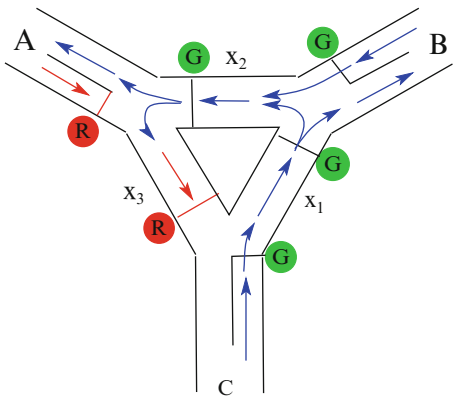


Fig. 9.16 The traffic control problem.

The exponential decay takes "approximately" into account the initial transient due to the traffic light switching. The other two configurations are obtained by a circular rotation of x_1, x_2, and x_3 (as well as of A, B, and C).

We model this problem by considering the following switched system, in which the control must select one of the three sub-systems characterized by the matrices

$$A_1 = \begin{bmatrix} -\gamma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \beta & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 0 & \beta \\ 0 & -\gamma & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad A_3 = \begin{bmatrix} 0 & 0 & 0 \\ \beta & 0 & 0 \\ 0 & 0 & -\gamma \end{bmatrix}, \tag{9.16}$$

with $\gamma = 1$ and $\beta = 1$.
First of all notice that no convex Hurwitz combination of the three matrices can be found. Indeed the characteristic polynomial of matrix $\alpha_1\hat{A}_1 + \alpha_2\hat{A}_2 + \alpha_3\hat{A}_3$ is

$$p(s, \alpha) = s^3 + (\alpha_1 + \alpha_2 + \alpha_3)s^2 + (\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1)s.$$

So $p(s, \alpha)$ is not a Hurwitz polynomial for any choice of $\alpha_i \geq 0, i = 1, 2, 3$, with $\alpha_1 + \alpha_2 + \alpha_3 = 1$, and therefore there are no Hurwitz convex combinations in the convex hull.
However, the matrix product $e^{A_1}e^{A_2}e^{A_3}$ is Schur (the dominant eigenvalue is ≈ 0.69). So, the periodic switching law

$$q(t) = \begin{cases} 3, t \in [3k, 3k + 1); \\ 2, t \in [3k + 1, 3k + 2); \\ 1, t \in [3k + 2, 3k + 3); \end{cases} \qquad k \geq 0,$$

makes the resulting system consistently (actually exponentially) stable, and hence exponentially stabilizable.

It is also worth pointing out an interesting fact. In general, the existence of a smooth control Lyapunov function would lead to the "arg-min" strategy which introduces chattering and sliding modes, as pointed out in Remark 9.24. For this problem chattering would be catastrophic, while it is obvious that to deal with this problem we must "dwell" on each configuration for a sufficiently long time. In the case of a periodic strategy with dwell time T in each mode, the product of the three exponentials $e^{A_1T}e^{A_2T}e^{A_3T}$ has to be stable. We have seen that this is the case for $T = 1$.

Note that this commutation implies that the "red" is imposed according to the circular order 3, 2, 1, 3, 2, 1 .... It is surprising to notice that, if the order is changed, not only the system performance can get worse, but the system may even become unstable. Indeed, $e^{A_3}e^{A_2}e^{A_1}$ is unstable with spectral radius ≈ 1.90, which means that the commutation order is fundamental and the order 1, 2, 3, 1, 2, 3 ... is unsuitable. A simple explanation is that switching the red light from 3 to 2 allows for a "fast recovery" from the congestion on x_3 (due to the exponential decay), while switching the red from 3 to 1 would leave such a congestion unchanged.

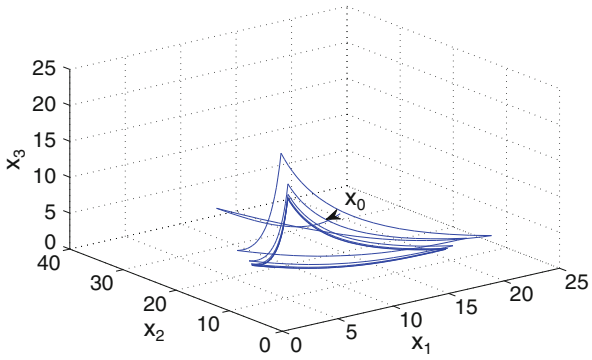


Fig. 9.17 State trajectory corresponding to $T = 2.1$ and $x(0) = [10\ 10\ 10]'$.

We complete the example by considering the effect of a constant input (the incoming traffic) and hence by introducing the system

$$\dot{x} = A_q x + b$$

with $b = \bar{1}$, $\gamma = 1$ and $\beta = 1$. It turns out that, with these values, $F := e^{A_1 T} e^{A_2 T} e^{A_3 T}$ is Schur for $T > 0.19$. This means that, under a periodic strategy with $T > 0.19$, the system converges to a periodic trajectory $\tilde{x}(t)$, as shown in Figure 9.17.

Note that it is possible to optimize T in order to achieve a strategy which reduces as much as possible the buffer levels of the periodic trajectory (see [BCV12] for details).

Remark 9.35. Note that, in principle, the matrices provided in the example do not satisfy the assumption of Theorem 9.32, because they are reducible. This is not an issue, because we can modify the system by perturbing all the coefficients with positive small numbers

$$A_i + \epsilon O$$

where O is the 1-matrix $O_{ij} = 1$, for all i, j, and $\epsilon > 0$ small. The periodic strategy would be stabilizing for ϵ small. However, no Hurwitz convex combination would exist anyway (see Exercise 5).

In the simple case of second order systems, the following result holds [BCV12].

Proposition 9.36. *A continuous-time second order positive switched system is stabilizable if and only if there exists an Hurwitz convex combination.*

The proof can be found in [BCV12]. We stress that sufficiency holds for positive switched systems of any order, since it holds for linear switched systems in continuous-time, as we have seen at the beginning of Section 9.4. Still we give a proof of the sufficiency to show that we can derive a linear co-positive

control Lyapunov function using the left Frobenius eigenvector of a Hurwitz convex combination. A similar proof works also for discrete-time switched systems. Assume that there exists

$$\bar{A} = \sum_{i=1}^{r} A_i \alpha_i, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{r} \alpha_i = 1$$

which is a Hurwitz matrix and assume, for brevity, that it is irreducible. Take the positive eigenvector $z^T > 0$ of $\bar{A}$ associated with the Frobenius eigenvalue $\lambda_F < 0$. Consider the co-positive function $\Psi(x) = z^T x$. Then, by linearity, for all $x \geq 0$

$$\min_i z^T A_i x \leq z^T \bar{A} x = \lambda_F z^T x = \lambda_F \Psi(x)$$

Therefore, the strategy $q = \arg\min_i z^T A_i x$ is stabilizing.

We know that, in general, the existence of a Schur-stable convex combination is neither necessary nor sufficient (see Exercises 6 and 7) for stabilizability of linear switched discrete-time systems. For positive discrete-time switched systems, the condition is sufficient, but not necessary (see Exercise 7) even for second order systems. The sufficiency proof can be derived exactly (mutatis mutandis) by means of the previous considerations (see Exercise 8).

We can now establish a procedure for discrete-time switched systems which leads to the generation of a polyhedral concave co-positive control Lyapunov function.

We use again some dynamic programming arguments used in Chapters 5 and 6 (see [HVCMB11]). Consider the set

$$\mathcal{X}_0 = \{x : \bar{1}^T x \geq 1\}$$

and notice that it is impossible to stabilize the system if and only if, starting from some initial value inside $\mathcal{X}_0$, the state $x(t)$ remains in this set for all possible switching sequences $q(t)$ or, equivalently, if there exists a robustly positive invariant set included in $\mathcal{X}_0$. This is also equivalent to saying that we have stabilizability if and only if there is no robustly positively invariant set included in $\mathcal{X}_0$. Since $\mathcal{X}_0$ is convex, one might try to compute the largest (convex) invariant set in $\mathcal{X}_0$ and check whether such a set is empty.

Consider the set of all vectors x in $\mathcal{X}_0$ which are driven inside $\mathcal{X}_0$ by all matrices A_i

$$\mathcal{X}_1 = \{x \geq 0 : \bar{1}^T x \geq 1, \ \bar{1}^T A_i x \geq 1, \ i = 1, \dots r\}$$

This set can be represented as

$$F^{(1)} x \geq \bar{1}$$

with

$$F^{(1)} = \begin{bmatrix} \bar{1}^T \\ \bar{1}^T A_1 \\ \vdots \\ \bar{1}^T A_r \end{bmatrix}$$

For $k = 1, 2, 3, \ldots$, recursively compute the set

$$\mathcal{X}_{k+1} = \{x \geq 0 : F^{(k)} x \geq 1, \;\; F^{(k)} A_i x \geq 1, \;\; i = 1, \ldots r\}$$

compactly represented as $\mathcal{X}_{k+1} = \{x \geq 0 : F^{(k+1)} x \geq 1\}$, and proceed in a "dynamic programming" style by generating the sets $\mathcal{X}_k$.

The set $\mathcal{X}_k$ is the set of all the states x_0 which remain in the set $\mathcal{X}_0$ ($\bar{1}^T x(t) \geq 1$) in k steps for all possible sequences $A_{q(t)} \in \{A_1, A_2, \ldots, A_r\}$. On the contrary, if $x_0 \notin \mathcal{X}_k$ there exists a sequence that brings $x(t)$ outside $\mathcal{X}_0$ hence ($\bar{1}^T x(t) < 1$).

Therefore, if for some $k > 0$ $\mathcal{X}_k$ is strictly included in $\mathcal{X}_0$ (see Fig. 9.18), then all the states $x(0)$ on its positive boundary can be driven in k steps to $x(k) \in \partial\mathcal{X}_0$, the boundary of $\mathcal{X}_0$. By construction, any point x_0 on the positive boundary of $\mathcal{X}_k$ satisfies the equality

$$\bar{1}^T A_{i_{k-1}} A_{i_{k-2}} \ldots A_{i_0} x(0) = 1$$

for a proper choice of the indices i_h, say $\bar{1}^T x(k) = 1$.

Now, if $\mathcal{X}_k$ does not intersect $\bar{1}^T x = 1$, then we can take a positive $\lambda < 1$, close enough to 1, such that the following inclusion is preserved (see Fig. 9.18)

$$\lambda \mathcal{X}_k \subset \mathcal{X}_0$$

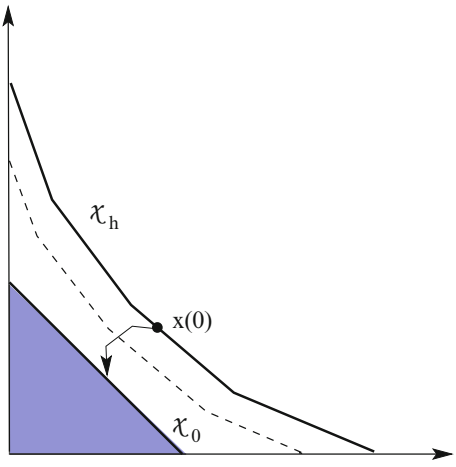


Fig. 9.18 The initial set $\mathcal{X}_0$ (complement of the dark region), the final set $\mathcal{X}_h$ (to the right of the thick curve), and the contracted version $\lambda\mathcal{X}_h$ (to the right of the dashed curve)

Let $\tilde{\mathcal{X}}_k$ be the closure of the complement of $\mathcal{X}_k$ and consider the function $\Psi(x) = \min F_i^{(k)} x$. Note that $\tilde{\mathcal{X}}_k = \mathcal{N}[\Psi(x), 1] = \{x : \Psi(x) \leq 1\}$. Given any $x(0) \in \partial\tilde{\mathcal{X}}_k$, it can be brought in k steps to $\partial\tilde{\mathcal{X}}_0$, hence to $\lambda\tilde{\mathcal{X}}_k$. Hence we have that

$$\Psi(x(k)) \leq \lambda\Psi(x(0))$$

for a proper sequence. Thus, by repeating this sequence periodically, we can assure $\Psi(x(ik)) \leq \lambda^i\Psi(x(0))$, $i = 1, 2, \ldots$, and drive the state to 0.

A possible different stopping condition for the set sequence is

$$\lambda\mathcal{X}_k \subset \mathcal{X}_{k-1}$$

for some contraction factor $\lambda > 0$. Denoting by F the matrix describing the final set $\mathcal{X}_k = \{x \geq 0 : Fx \geq 1\}$, we have that the concave co-positive piecewise-linear Lyapunov function

$$\Psi(x) = \min Fx \tag{9.17}$$

is a control Lyapunov function since

$$\min_i \Psi(A_i x) \leq \Psi(x)$$

According to the considerations in Chapters 5 and 6, we can consider the modified system

$$x(t+1) = \left(\frac{A_i}{\lambda}\right) x(t)$$

with an assigned contractivity factor $\lambda > 0$.

Procedure

1. Let $F^{(0)} = \bar{1}^T$. Fix a maximum number of steps k_{max}, a contractivity factor $\lambda > 0$ and a tolerance $\epsilon > 0$ such that $\lambda + \epsilon < 1$.
2. For $k = 1, 2, \ldots$, compute the set

$$\mathcal{X}_k = \{x \geq 0 : F^{(k-1)}x \geq \bar{1}, \quad F^{(k-1)}(A_i/\lambda)x \geq 0, \ \ i = 1, 2, \ldots, r\}$$

   This set is of the form $\mathcal{X}_k = \{x \geq 0 : F^{(k)}x \geq \bar{1}\}$.
3. Eliminate all the redundant inequalities, to achieve a minimal $F^{(k)}$ representing $\mathcal{X}_k = \{x \geq 0 : F^{(k)}x \geq \bar{1}\}$.
4. If $\mathcal{X}_k \subset (1+\epsilon)\mathcal{X}_{k-1}$ stop (successfully): the closure of the complement of $\mathcal{X}_k$ in the positive orthant is a $\lambda + \epsilon$ contractive set.

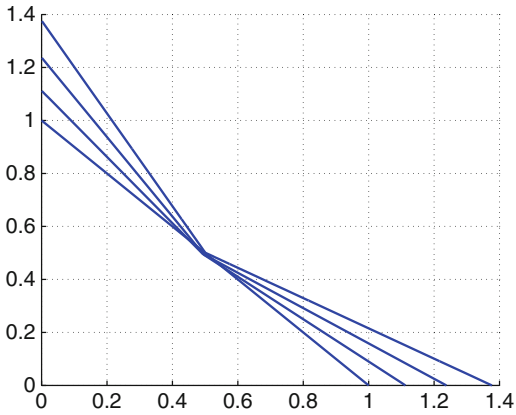


Fig. 9.19 The two-configuration emptying problem in the controlled case

5. If $\mathcal{X}_k = \mathcal{X}_{k-1}$ and the boundary of the original set $\bar{1}^T x$ is an active constraint stop (unsuccessfully): the set $\mathcal{X}_k$ is robustly invariant for the modified system. Hence, no matter which sequence $A_{q(k)}$ is chosen, for x_0 in this set we will have $\bar{1}^T x(k) \geq 1$.
6. Let $k := k + 1$. If $k \geq k_{max}$, STOP.

Example 9.37. Consider again the system of Example 9.28, but assume now that we can choose at each time the matrix (A_1 or A_2). Under arbitrary switching the system has a contractivity factor of about 0.92. If we apply the above, we see that for the controlled system the contractivity factor (obviously) reduces to 0.89. The sequence of regions is depicted in Fig 9.19. The final control Lyapunov function is

$$\Psi(x) = \min\{1.2710x_1 + 0.7263x_2, 0.7263x_1 + 1.2710x_2\}$$

9.6 Switching compensator design

The trade-off among different, often conflicting, design goals is a well-known problem in control design [LM99]. Even in simple cases, such as the servo design, it is not possible to achieve a certain performance without compromising another. For instance, a fast signal tracking in a controlled loop requires a large bandwidth which has the side-effect of rendering the system more sensible to disturbances. This problem is often thought of as an unsolvable one: trade-off is generically considered an unavoidable issue.

In this section we wish to partially contradict this sentence by presenting, in a constructing way, techniques for switching among controllers each designed for a specific goal as an efficient approach to reduce the limitations due to adopting a single controller.

9.6.1 Switching among controllers: some applications

We start with a very simple example which basically shows which are the benefits we can achieve by switching.

Example 9.38. Switching strategies can be successfully applied to the problem of semi-active damping of elastic structures. This is a problem investigated since 30 years ago [HBR83] and it is still quite popular. Here we just propose a simple example to provide an idea of what can be done by means of Lyapunov-based techniques. Consider the very simple problem of damping via feedback a single degree of freedom oscillator whose model is

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & -\mu \end{bmatrix}}_{A(\mu)} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

where μ is a damping coefficient $\mu \geq \beta > 0$. The lower bound $\beta > 0$ represents the natural damping (i.e., that achieved with no control) and in the example is assumed that $\beta = 0.1$.

Consider the problem of determining the appropriate value of μ to achieve "good convergence" from the initial state $x = -[x_0\ 0]^T$. This is an elementary problem often presented in basic control courses. The trade-off in this case is the following:

- small values of μ produce an undamped dynamics with undesirable oscillations;
- large values of μ produce an over-damped dynamics with slow convergence.

Elementary root locus considerations lead to the conclusion that the "best" coefficient to achieve the fastest transient is the critical damping, i.e. the value $\mu_{cr} = 2$ (associated with a pair of coincident eigenvalues). The situation is represented in Fig. 9.20, where we considered the initial condition $x(0) = [-2\ 0]^T$. The choice $\mu = \beta$, which produces a fast reaction of the system, results anyway in a poorly damped transient which is represented by the dotted oscillating solution. If we consider a high gain, for instance such that $\mu = 10$, then we have the opposite problem: there are no oscillations, but a slow exponential transient (represented by the exponentially decaying dotted curve in Fig. 9.20). The critical oscillation, obtained when $\mu_{cr} = 2$, is the best trade-off and is represented by the lower dashed line in Fig. 9.20.

Can we do better? We can provide a positive answer if we do no limit ourselves to considering a single value of μ. The idea is that, by switching among different values of μ we have more degrees of freedom. Note that this is equivalent to switching between derivative controllers with different gain.

We consider the case in which we can switch between two gains $\mu = \beta$ and $\bar{\mu} = 10$. The problem is clearly how to switch between the two. The first idea one might have in mind is heuristically motivated. We allow the system to go without artificial damping $\mu = \beta$ until a certain strip $|x_1| \leq \rho$ is reached, where ρ is an

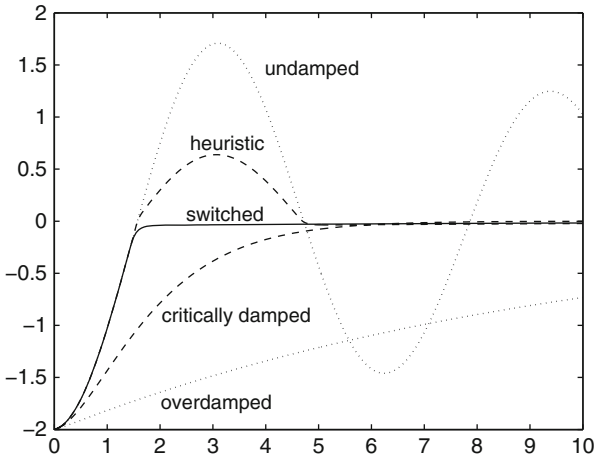


Fig. 9.20 The different solutions

assigned value, and we brake by switching to $\bar{\mu}$. We chose $\rho = 0.05$. Unfortunately, this heuristic solution is not very satisfactory and is represented by the dashed curve marked as heuristic in Fig. 9.20. As expected, it is identical to the undamped solution in the first part, and then there is a braking stage. It is however apparent that braking is not sufficient and the system has an overshoot since the state leaves the braking region in the overshoot stage and jumps out of the braking zone.

Let us consider a better solution for this problem. We activate the braking value $\bar{\mu}$ only when a proper positively invariant strip for the damped system is reached. For $\mu = 10$, which is quite larger than the critical value, the system has two real eigenvalues $\lambda_F < \lambda_S < 0$, where $\lambda_S \simeq 0$ and $\lambda_F << \lambda_S$ are the slow and the fast eigenvalue, respectively. The eigenvector associated with the fast eigenvalue is $v_F = [1 \ \lambda_F]^T$. Along the subspace corresponding the "fast" eigenvector the transient is fast. Let us consider the orthogonal unit vector

$$f^T = \frac{[-\lambda_F \ \ 1]}{\|[-\lambda_F \ \ 1]\|}$$

and a thin strip of the form

$$\mathcal{S}(\xi) = \{x : \ \ |f^T x| \leq \xi\}$$

This strip is positively invariant for the system with high damping. Indeed the vector f^T is a left eigenvector associated with λ_F and thus

$$f^T A(\bar{\mu}) = \lambda_F f^T$$

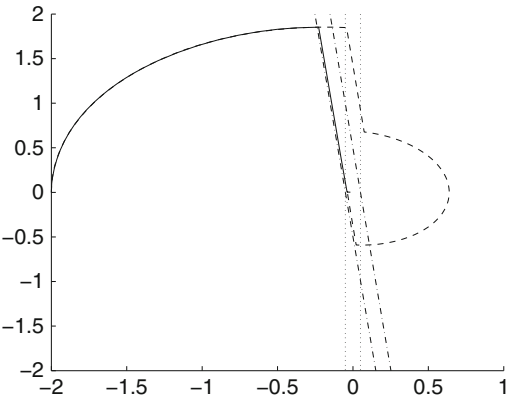


Fig. 9.21 The solution with the good (plain) and the heuristic (dashed) switching law

so that the positive invariance conditions of Corollary 4.42 are satisfied. Note also that the strip includes the subspace associated with v_F. Theoretically, the idea is to switch on the subspace associated with v_F to exploit the fast transient. However, "being on a subspace" is not practically meaningful, so we replace the above with the condition $x \in \mathcal{S}(\xi)$, which can be interpreted as "being on the subspace up to a tolerance ξ". We chose $\xi = \rho = 0.05$ for a fair comparison. It is apparent that the results are much better. The transient, the plain line in Fig. 9.20, is initially the fast undamped one and then it is close to the "fast eigenvector" motion. The situation can be visualized in the phase plane in Fig. 9.21. The heuristic solution switches, as required, in the strip $|x_1| \leq \rho$, but cannot remain inside this strip (the vertical one), since it is not invariant, and then it undergoes a further switching and the damping is improperly set to the natural β again to eventually reach the braking region. An application of the proposed technique to more general vibrating structures has been proposed in [BCGM12, BCC+14].

The idea of switching between controllers by exploiting the properties of invariant sets is not new. It was presented in [GT91, WB94, KG97, BB99]. The role played by invariant sets has been evidenced in the previous example. If switching to a new controller is subject to reaching a proper invariant set for the closed-loop system with such a controller, then we automatically assure that the new set will never be left anymore. A possible application is the transient improvement via switching. Suppose that we are given a linear system and a family of controllers with "increasing gain"

$$u = K_i x, \quad i = 1, 2, \ldots, r$$

associated with a nested family of invariant C-sets for the systems $\dot{x} = (A + BK_i)x$:

$$\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \cdots \subseteq \mathcal{S}_r$$

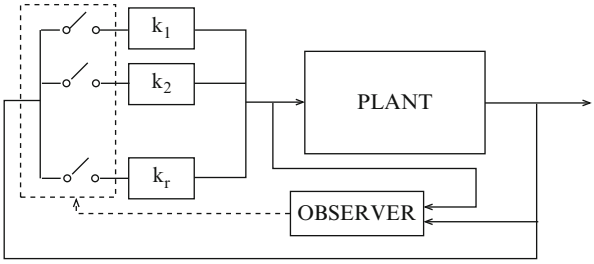


Fig. 9.22 The logic-based switching system

Choosing a nested family is always possible since any invariant set remains such if it is properly scaled. Then switching among controllers avoids control saturation when the state is far from the target and exploits the strong action of the larger gains in proximity of the origin. The idea can be extended to the case of output feedback if we introduce an observer-based supervisor [DSDB09]. Assume for brevity that a family of static output feedback gains is given. Then a possible scheme is that in Fig. 9.22. After an initial transient, in which no switching should be allowed, the estimate state is accurate (here we assume accurate modeling and negligible disturbances), say $\hat{x}(t) \simeq x(t)$. The inclusion of the estimated state $\hat{x}(t) \in \mathcal{S}_i$ in the appropriate region allows the switching to the i-th gain.

Clearly the method requires either state feedback or accurate state estimation, which is not always possible. Finally we notice that we can design compensators of different order and dynamic. In this case we have an extra degree of freedom, given by the possible initialization of the compensator state at the switching time.

A fundamental result concerning switching among controllers is the following [HM02], here reported.

Theorem 9.39. *Consider a linear plant P and a set of r linear stabilizing compensators. Then, for each compensator there exists a realization (not necessarily minimal) such that, no matter how the switching among these compensators is performed, the overall closed-loop plant is asymptotically stable.*

It is important to notice that the previous result does not imply that the property is true for any family of compensators with given realizations. In other words, it is easy to find families of compensator (even static gains) for which an unfavorable switching law can lead to instability. A simple case is given by the system

$$A = \begin{bmatrix} 0 & 1 \\ -1 & -\beta \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

If we take $u = -(1+\eta)y$ and $u = -(1-\eta)y$, with β small enough and $0 < \eta < 1$ sufficiently close to 1, there exists a destabilizing switching law [Lib03]. However, there exist equivalent non-minimal realizations of the constant gains for which stability is assured under switching.

The result [HM02], which is of fundamental importance, does not consider performances. It is apparent from Example 9.38, the friction system, that, if one wishes to assure a performance to the system, then some "logic constraints" to the switching rule have to be applied.

Further techniques of "switching among controllers" have been proposed in the literature, although of different nature. It is worth mentioning a special technique which consists in partitioning the state space in connected subsets in each of which a "local controller" is active. This idea has been pursued in [MADF00] in a context of gain-scheduling control and in [BRK99, BPV04] in a context of robot manipulator control in the presence of obstacles.

9.6.2 *Parametrization of all stabilizing controllers for LTI systems and its application to compensator switching*

One possibility to avoid the limitation of a single controller is to use more than one controller. For instance, if the system changes its working point or its configuration, one may decide to change the compensator accordingly.

In this subsection we briefly describe the essential of the idea of the mentioned Theorem 9.39 due to [HM02] and then we apply it to the case in which a switching compensator is applied to a fixed plant.

The first fundamental step towards the solution of the parametrization of a family of switching compensators is the standard Youla–Kucera parametrization of all stabilizing controllers for linear systems. Consider a stabilizable LTI system

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

and assume a stabilizing compensator is given with transfer function $W(s)$.

Then $W(s)$ can be realized as follows:

$$\dot{x}_o(t) = (A - LC)x_o(t) + Bu(t) + Ly \tag{9.18}$$

$$u(t) = -Jx_o(t) + v(t) \tag{9.19}$$

$$o(t) = Cx_o(t) - y(t) \tag{9.20}$$

$$\dot{z} = F_T z(t) + G_T o(t) \tag{9.21}$$

$$v(t) = H_T z(t) + K_T o(t) \tag{9.22}$$

where J and L are matrices such that $(A - LC)$ and $(A - BJ)$ are Hurwitz, (F_T, G_T, H_T, K_T) are suitable matrices (which depend on the plant matrices (A, B, C), as well as on $W(s)$), with F_T Hurwitz. We let the reader note that $x_o(t)$ is the state estimate.

Conversely, given L and J such that $(A - LC)$ and $(A - BJ)$ are Hurwitz, for any choice of (F_T, G_T, H_T, K_T) of suitable dimensions, with F_T Hurwitz, the corresponding compensator is stabilizing.

This double implication is equivalent to saying that (9.18)–(9.22) parametrize all the stabilizing compensators for an LTI plant.

Proposition 9.40 ([ZDG96, SPS98]). *The transfer matrix $W(s)$ is a stabilizing compensator for (A, B, C) if and only if it can be realized as in (9.18)–(9.22), with some choice of the Youla–Kucera parameter (F_T, G_T, H_T, K_T).*

Note that the realization of $W(s)$ in (9.18)–(9.22) is non-minimal in general.

The sub-system (9.21)–(9.22) represents the Youla–Kucera (*YK*) parameter. The fundamental point is that the input of this system is $o(t) = Cx_o(t) - y(t) = C(x_o(t) - x(t))$, which asymptotically converges to 0 since

$$\frac{d}{dt}(x_o(t) - x(t)) = (A - LC)(x_o(t) - x(t))$$

is an unreachable variable. This in turn means that the output $v(t)$ of the Youla–Kucera parameter is not fed back by the plant and therefore any stable choice of the *YK* parameter cannot destabilize the closed-loop. Note however that any output y of the plant is fed back by the compensator and the feedback depends on the *YK* parameter.

Going back to the general case, assume now that a family $W_1(s)$, $W_2(s)$, $\ldots, W_r(s)$, of stabilizing compensators is given: is it possible to switch among them arbitrarily while preserving closed-loop stability?

In this case Theorem 9.39 comes into play providing a positive answer: yes, stability can be preserved if the realization of each of the stabilizing compensator is done in the right way. How does such a realization look like?

The solution of this problem is quite intuitive and boils down to the YK parametrization. Since any $W_i(s)$ corresponds to some *YK* parameter $(F_T^{(i)}, G_T^{(i)}, H_T^{(i)}, K_T^{(i)})$, we can switch between compensators by fixing matrices L and J and by switching just among the *YK* parameters. However, the fact that $F_T^{(i)}$ is Hurwitz does not assure that switching between the *YK* parameters results in a stable behavior. Moreover, the $F_T^{(i)}$ may be of different dimension.

The problem of dimension is immediately solved by merging some *YK* parameters in fictitiously augmented dynamics, in order to make them all of the same size.

For stability under switching, we need the following.

Lemma 9.41. *Given a stable square matrix F, there exists an invertible T such that $\hat{F} = T^{-1}FT$ has $P = I$ as a Lyapunov matrix.*

Proof. Assume that F is stable and solves

$$F^T P + PF = -I$$

with $P \succ 0$. Let $T = P^{-1/2}$. Then, bearing in mind that $T^T P T = I$,

$$\hat{F}^T + \hat{F} = T^T F^T T^{-T} T^T P T + T^T P T T^{-1} F T = T^T (F^T P + P F) T = -T^T T < 0$$

With the above in mind, the idea is then that we may change the realization to all the *YK* parameters in such a way they share I as a Lyapunov matrix. This will in turn

- not change the compensator transfer functions;
- assure that the *YK* parameter is stable under arbitrary switching.

9.6.3 Switching compensators for switching plants

In this subsection, the case in which the compensator switching is subsequent to a plant switching is considered. More precisely, the problem is cast in the following setting: assume that the plant formed by the family of stabilizable LTI systems

$$\begin{aligned} \dot{x}(t) &= A_i x(t) + B_i u(t) \\ y(t) &= C_i x(t) \end{aligned}$$

is subject to an arbitrary switching rule

$$i = i(t) \in \mathcal{I} = \{1, 2, \ldots, r\}$$

and that the switching plant has to be controlled by means of a family of r stabilizing controllers, each (clearly) stabilizing the corresponding plant (see Fig. 9.23)

$$\begin{aligned} \dot{z}(t) &= F_i z(t) + G_i y(t) \\ u(t) &= H_i z(t) + K_i y(t) \end{aligned}$$

To check whether such a family of controllers exists, a couple of assumptions are needed.

Assumption (Non-Zenoness). The number of switching instants is finite on every finite interval.

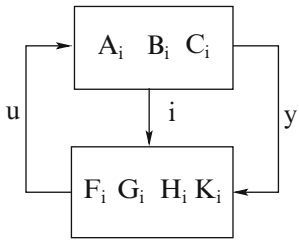


Fig. 9.23 The switching control

Assumption (No delay). There is no delay in the communication between the plant and the controller, which knows exactly the current output $y(t)$ and configuration $i(t)$.

With the above in mind, the two essential problems considered next are the following.

Problem 1. Does there exist a family of matrices (F_i, G_i, H_i, K_i), $i \in \mathcal{I}$, such that the closed-loop system is switching stable?

Problem 2. Given a set of compensators $W_i(s)$, each assuring Hurwitz stability for fixed i, does there exist a realization (F_i, G_i, H_i, K_i) such that:

1. $W_i(s) = H_i(sI - F_i)^{-1}G_i + K_i$;
2. the closed-loop system is switching stable?

Clearly Problem 1 is preliminary to Problem 2. A weaker, but computationally tractable, version is that in which, rather than requiring stability, quadratic stability of the switching closed-loop system only is requested. It will be soon shown that the solution to Problem 1 amounts to checking a set of necessary and sufficient conditions, whereas for Problem 2, assuming Problem 1 has a solution, the answer is always affirmative [BMM09].

As a first preliminary fact it must be pointed out that, when dealing with the stability of switching systems, we need to talk about their realization and not about their transfer functions. In fact, while for LTI systems the same transfer function can be realized in infinite (equivalent) ways, this is not the case for switching systems.

Example 9.42. Consider the plant

$$P(s) = \frac{1}{s+\alpha}$$

with $\alpha > 0$ and let the compensator be of the form

$$W_i(s) = \frac{k_i}{s+\beta}, \quad i = 1, 2$$

with $\beta > 0$. Using standard realization techniques, the two following closed-loop matrices

$$A_1 = \begin{bmatrix} -\alpha & k_i \\ -1 & -\beta \end{bmatrix} \quad \text{or} \quad A_2 = \begin{bmatrix} -\alpha & \sqrt{k_i} \\ -\sqrt{k_i} & -\beta \end{bmatrix}$$

can be obtained[7]. It can be seen that the first is unstable under arbitrary switching (see Section 9.7.2), whereas the second is stable under arbitrary switching.

[7] Using either $\dot{z} = -\beta z + k_i y$, $u = -z$ or $\dot{z} = -\beta z + \sqrt{k_i} y$, $u = -\sqrt{k_i} z$ as a realization.

The following theorem provides a solution to Problem 1.

Theorem 9.43. *The following two statements are equivalent.*

i) There exists a linear switching-stabilizing compensator
ii) The two equations

$$A_i X + B_i U_i = X P_i \tag{9.23}$$

$$R A_i + L_i C_i = Q_i R \tag{9.24}$$

have a solution $(P_i, Q_i, U_i, L_i, X, R)$*, with*

$$P_i \in \mathcal{H}_1 \quad \text{and} \quad Q_i \in \mathcal{H}_\infty$$

(resp. $\|P_i\|_1 < 1$ *and* $\|Q_i\|_\infty < 1$ *in the discrete-time case), and a full row-rank matrix* $X \in \mathbb{R}^{n \times \mu}$ *and a full column-rank matrix* $R \in \mathbb{R}^{\nu \times n}$*.*

If the conditions of the theorem hold, then the problem can be solved as follows. Take M: $MR = I$, $V_i = ZP_i$, where Z is any complement of X and

$$\begin{bmatrix} K_i & H_i \\ G_i & F_i \end{bmatrix} = \begin{bmatrix} U_i \\ V_i \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix}^{-1}$$

A possible compensator is the following:

$$\text{Estimated state feedback:} \quad \begin{cases} \dot{z}(t) = F_i z(t) + G_i \hat{x}(t) \\ u(t) = H_i z(t) + K_i \hat{x}(t) + v(t) \\ v(t) \equiv 0 \end{cases}$$

$$\text{Generalized state observer:} \quad \begin{cases} \dot{w}(t) = Q_i w(t) - L_i y(t) + R B_i u(t) \\ \hat{x}(t) = M w(t) \end{cases}$$

The previous compensator has a separation structure and it can be shown that $\hat{x}(t) - x(t) \to 0$ and that the first part is a dynamic state-feedback compensator. The auxiliary signal $v(t) = 0$ is a dummy signal which will be used later.

We do not report a proof here (the interested reader is referred to [BMM09]), but we just point out that in the necessity part of the theorem we generalize the results in Subsection 4.5.6. More precisely, the equations are an extension of (4.40) and (4.53). Note that the construction is identical to that proposed in Section 7.4. This is absolutely expected, since we know that an LPV system whose matrices are inside a polytope is stable if and only if its corresponding switching system is stable.

If the less stringent requirement of quadratic stability only is imposed, then the following (tractable) result holds:

Theorem 9.44. *The following two statements are equivalent.*

i) There exists a family of linear quadratically stabilizing switching compensators.
ii)

$$PA_i^T + A_iP + B_iU_i + U_i^TB_i^T < 0$$
$$A_i^TQ + QA_i + Y_iC_i + C_i^TY_i^T < 0$$

for some positive definite symmetric $n \times n$ matrices P and Q, and matrices $U_i \in \mathbb{R}^{m\times n}$, $Y_i \in \mathbb{R}^{n\times p}$.

A possible compensator is

$$\begin{aligned} \tfrac{d}{dt}\hat{x}(t) &= (A_i + L_iC_i + B_iJ_i)\hat{x}(t) - L_iy(t) + B_iv(t) \\ u(t) &= J_i\hat{x}(t) + v(t) \\ v(t) &= 0 \end{aligned}$$

where

$$J_i = U_iP^{-1}, \qquad L_i = Q^{-1}Y_i$$

In discrete-time the inequalities are

$$\begin{bmatrix} P & (A_iP + B_iU_i)^T \\ A_iP + B_iU_i & P \end{bmatrix} > 0$$
$$\begin{bmatrix} Q & (QA_i + Y_iC_i)^T \\ QA_i + Y_iC_i & Q \end{bmatrix} > 0$$

The previous ones are standard quadratic stabilizability conditions which involve LMIs [BP94, AG95, BEGFB04].

To provide an answer to Problem 2, the signal $v(t)$ comes into play. The first point is the following. Assume that in the previous machinery the signal $v(t)$ is generated as the output of the following system:

$$v(t) = T(C(\hat{x} - x))$$

where $T(\cdot)$ is a "stable operator." This in turn will change the compensator, but it will not destabilize the plant. Although in general any input–output stable operator would fit, we limit ourselves to linear finite-dimensional systems.

From Proposition 9.40 it is known that for any fixed mode there exists a Youla–Kucera parameter T_i such that the resulting compensator has transfer function W_i. The issue is only to realize the Youla–Kucera parameters in such a way that they are stable under switching as in Fig. 9.24. Note that the figure includes, as a particular case, the case of a quadratic stabilizable plant for which $R = I$, $Q = A + LC$, $M = I$, and the state feedback is static: $u = Jx + v$.

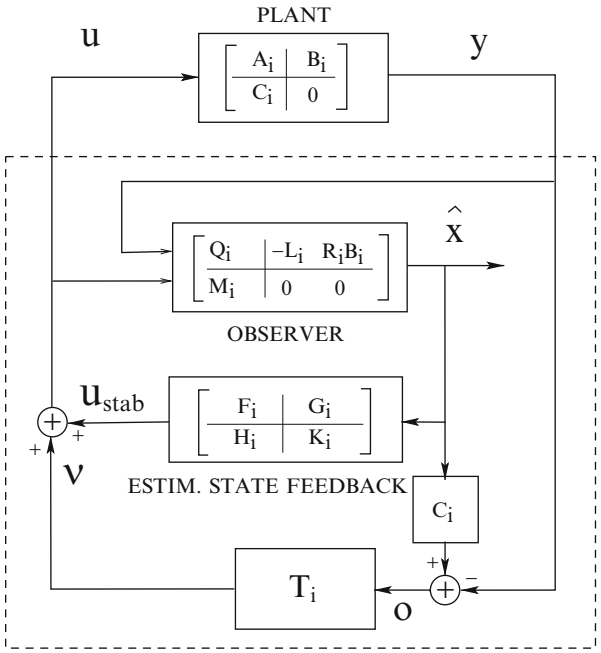


Fig. 9.24 The observer-based compensator structure

Theorem 9.45. *If the (quadratic) stabilizability conditions in Theorem 9.43 (9.44) are satisfied, then, given any* ***arbitrary*** *family of transfer functions $W_i(s)$, $i = 1, \ldots, r$ each stabilizing the i-th plant, there exists a switching compensator*

$$\begin{aligned} \delta z(t) &= F_i z(t) + G_i y(t) \\ u(t) &= H_i z(t) + K_i y(t) \end{aligned}$$

(δz is either $\dot{z}$ or $z(t+1)$) such that

1. $H_i(sI - F_i)^{-1}G_i + K_i = W_i(s)$
2. *the closed-loop system is switching stable.*

We also point out the following aspect. Assume that there are a disturbance input $\omega(t)$ and a performance output ξ

$$\begin{aligned} \delta x(t) &= A_i x(t) + B_i u(t) + B_i^{\omega}\omega(t) \\ y(t) &= C_i x(t) + D_i^{y,\omega}\omega(t) \\ \xi(t) &= E_i x(t) + D_i^{\xi,u} u(t) + D_i^{\xi,\omega}\omega(t) \end{aligned} \tag{9.25}$$

Then the i-th input–output map is of the form

$$\xi(s) = [M_i^{\xi,\omega}(s) + M_i^{\xi,v}(s)T_i(s)M_i^{o,\omega}(s)]\omega(s)$$

which is amenable for optimization.

Example 9.46 (Networked control system (continued from Example 9.4)). The system matrices of the extended switching system (9.5) for the unstable dynamic system $P(s) = \frac{s+1}{s^2-10s}$, with sampling time $T_c = 0.05s$ and $N_{max} = 5$ are:

$$A = \begin{bmatrix} 1.649 & 0 \\ 0.065 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0.130 \\ 0.003 \end{bmatrix}, \quad C = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$$

A family of quadratically stabilizing compensators is obtained by solving the conditions in Theorem 9.44, which provide the following feedback and observer gains (mind that the first set of LMIs is indeed a single inequality in view of the fact that the system update and input matrices do not switch)

$$J = \begin{bmatrix} -12.859 & -6.015 & -.035 & -.022 & -.010 & -.005 & -.002 \end{bmatrix}$$

and $L = [L_0 \dots L_{N_{max}}]$,

$$L = \begin{bmatrix} -3.04 & -5.02 & -8.25 & -13.82 & -22.79 & -27.55 \\ -0.28 & -0.46 & -0.76 & -1.29 & -2.13 & -2.60 \\ -1.00 & -1.66 & -2.72 & -4.56 & -7.53 & -9.11 \\ -0.60 & -1.00 & -1.64 & -2.76 & -4.55 & -5.51 \\ -0.36 & -0.60 & -0.99 & -1.65 & -2.73 & -3.31 \\ -0.21 & -0.36 & -0.59 & -1.01 & -1.63 & -1.98 \\ -0.13 & -0.21 & -0.35 & -0.58 & -1.07 & -1.19 \end{bmatrix}$$

The closed-loop step responses y^t and $\hat{y}^t$ for two specific delay realizations are depicted in Fig. 9.25.

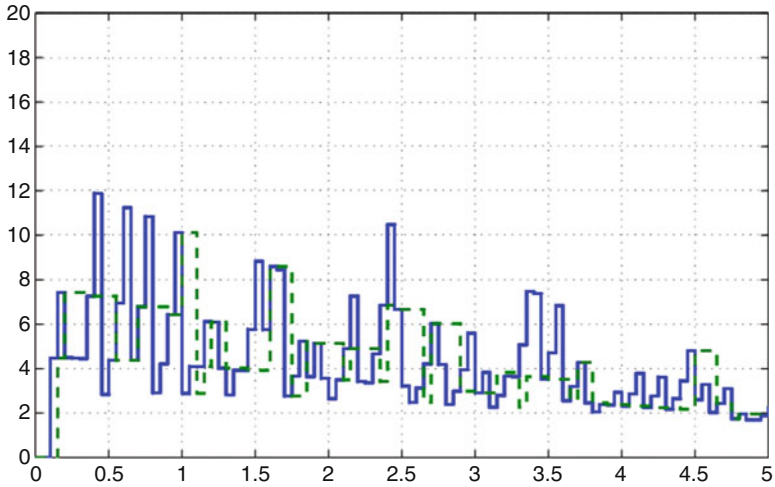


Fig. 9.25 Closed-loop step response for Example 9.46: y^t solid, $\hat{y}^t$ dashed

9.7 Special cases and examples

9.7.1 Relay systems

We now consider relay systems, which are nothing but a very special case of switching systems that can be tackled via a set-theoretic approach. Consider the scheme depicted in Figure 9.26 and representing a system with a relay feedback. We assume that the control law is

$$u = -sgn(y),$$

thus normalizing the input amplitude to 1. It is well known that feedback systems of this type have several problems, for instance that of being potential generators of limit cycles. There is a case in which one can guarantee asymptotic stability and this is the case we are going to investigate next. The following proposition holds.

Proposition 9.47. *Assume that the n-th order SISO transfer function $P(s)$ has $n-1$ zeros z_i with strictly negative real part (say it is minimum-phase with relative degree one) and that*

$$\lim_{s\to\infty} sP(s) > 0$$

Then the relay loop of Figure 9.26 is locally stable.

Proof. Since we are interested in local stability, it is assumed that $r = 0$. Let $-\beta < 0$ be greater than the largest real part of the transmission zeros. For any realization $(A, B, C, 0)$ of $P(s)$ it is possible to define the transformation matrix

$$T = \begin{bmatrix} \left(\text{null}\{B^T\}\right)^T \\ \frac{C}{CB} \end{bmatrix}^{-1}$$

(null(M) denotes a basis matrix for the kernel of M) so that

$$T^{-1}B = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}, \quad CT = \begin{bmatrix} 0 \cdots CB \end{bmatrix}$$

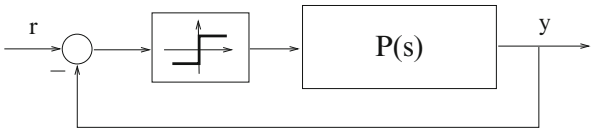


Fig. 9.26 The relay feedback loop

with $CB > 0$ in view of the condition on the limit and $T^{-1}AT$ can be partitioned as

$$T^{-1}AT = \begin{bmatrix} F & G \\ H & J \end{bmatrix}$$

where the eigenvalues of F (see Exercise 10 in Chapter 8) are the transmission zeros, thus have negative real part smaller than $-\beta$. Assume for brevity $CB = 1$. The state representation of $P(s)$ can then be written as

$$\begin{bmatrix} \dot{z}(t) \\ \dot{y}(t) \end{bmatrix} = \begin{bmatrix} F & G \\ H & J \end{bmatrix} \begin{bmatrix} z(t) \\ y(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) \tag{9.26}$$

$$y(t) = \begin{bmatrix} 0 & CB \end{bmatrix} \begin{bmatrix} z(t) \\ y(t) \end{bmatrix} \tag{9.27}$$

where y is the output and the input is $u = -sgn(y)$.

Since matrix F is asymptotically stable, if $y(t)$ is bounded, so is $z(t)$. Consider now a β-contractive C-set for the z-sub-system $\dot{z} = Fz + G\tilde{y}$, where $\tilde{y}$ is subject to $|\tilde{y}(t)| \leq 1$ and is seen as a disturbance.

Let such a set be $\mathcal{S}$ and let $\Psi_z(z)$ be its Minkowski functional. Assume $\mathcal{S}$ is 0-symmetrical, so that $\Psi_z(z)$ is a norm (for instance, a quadratic norm $\Psi_z(z) = \sqrt{z^T Q z}$ with $Q \succ 0$, associated with a contractive ellipsoid according to inequality (4.23)).

For brevity assume (although this is not necessary) that $\Psi_z(z)$ is smooth for $z \neq 0$. Since $\mathcal{S}$ is contractive (see Definition 4.15 with $u = 0$), for z on the boundary ($\Psi_z(z) = 1$) one has that

$$D^+\Psi_z(z) = \nabla\Psi_z(z)(Fz + G\tilde{y}) \leq -\beta, \tag{9.28}$$

for all $|\tilde{y}| \leq 1$. Since $\nabla\Psi_z(z) = \nabla\Psi_z(\xi z)$ for any scaling factor ξ (see Exercise 14 in Chapter 4) and since any $\hat{z}$ in $\mathbb{R}^{n-1}$ can be written as $\hat{z} = \Psi_z(\hat{z})z$ for some $z \in \partial\mathcal{S}$, then by scaling (9.28) one gets

$$D^+\Psi_z(\hat{z}) = \nabla\Psi_z(z)(F\Psi_z(\hat{z})z + Gy) \leq -\beta\Psi_z(\hat{z}), \quad \text{for } |y| \leq \Psi_z(\hat{z}) \tag{9.29}$$

say the scaled set $\xi\mathcal{S}$ is β contractive if $|y| \leq \xi$ (see Exercise 15 in Chapter 4). Now, consider the second equation and define, as a first step, the quantity

$$\mu = \max_{z \in \mathcal{S}} |Hz|,$$

which is a bound for the influence of z on y since

$$|Hz| \leq \mu\Psi_z(z)$$

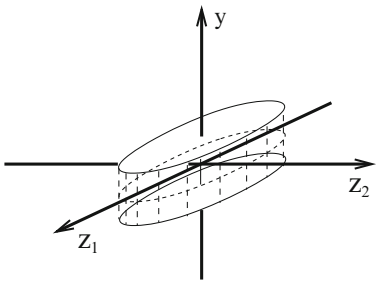


Fig. 9.27 The set $\mathcal{P}$

Then, consider the candidate Lyapunov function (referred to the y-system)

$$\Psi_y(y) = |y|,$$

which is the Minkowski function of the interval $I_1 = [-1, 1]$ and whose gradient is $\nabla\Psi_y(y) = sgn[y]$. In the (z, y) space consider the C-set which is the Cartesian product of I_1 and $\mathcal{S}$, $\mathcal{P} = \{(z, y) : z \in \mathcal{S},\ \ y \in I_1\}$ (see Fig. 9.27). The Minkowski functional of $\mathcal{P}$ is

$$\Psi(z, y) = \max\{\Psi_z(z), \Psi_y(y)\}$$

Now assume $(z, y) \in \varepsilon\mathcal{P}$ (namely $\Psi(z, y) \leq \varepsilon$), where ε is chosen as

$$\varepsilon \leq \frac{CB}{\mu + |J| + \beta}$$

so that

$$(\mu + |J|)\,\varepsilon \leq CB - \beta\varepsilon$$

For any $(z, y) \in \varepsilon\mathcal{P}$, by considering the y derivative, one gets

$$\begin{aligned} D^+\Psi_y(y) = D^+|y| &= sgn(y)[Hz + Jy + CBu] \\ &\leq \mu\Psi_z(z) + |J|\Psi_y(y) - CB \leq \mu\varepsilon + |J|\varepsilon - CB \leq -\beta\varepsilon, \end{aligned}$$

so that, for $\Psi_y(y) \leq \varepsilon$,

$$D^+\Psi_y(y) \leq -\beta\Psi_y(y) \tag{9.30}$$

Therefore, for $\Psi(z, y) \leq \varepsilon$ both (9.29) and (9.30) hold, which in turn implies that $\Psi(z(t), y(t))$ cannot increase if $\Psi(z(t), y(t)) \leq \varepsilon$ or, in other words, $\varepsilon\mathcal{P}$ is positively invariant. Since these inequalities hold inside $\varepsilon\mathcal{P}$, $y(t)$ and $z(t)$ both converge to 0 with speed of convergence β.

It can be shown, by means of arguments similar to those used for the relay system, that also almost relay systems, precisely those formed by a loop with control

$$u = -sat_{[-1,1]}(ky)$$

are locally stable for k large enough.

Example 9.48. Consider the unstable system

$$P(s) = \frac{s+1}{s^2 - s - 1}$$

represented by the equations

$$\begin{aligned} \dot{z}(t) &= -z(t) - y(t), \\ \dot{y}(t) &= z(t) + 2y(t) + u(t) \end{aligned}$$

Consider the first sub-system $\dot{z} = -z - \tilde{y}$, $|\tilde{y}| \leq 1$ and $\beta = 0.5$, which is compatible with the fact that the zero is in -1. Consider the interval $[-\zeta, \zeta]$ as candidate contractive set whose Minkowski functional is $|z|/\zeta$. The condition is that for $z = \zeta$

$$\frac{1}{\zeta}(-z - \tilde{y}) \leq -\beta$$

for all $|\tilde{y}| \leq 1$, which is satisfied for $\zeta \geq 1/(1-\beta) = 2$. The opposite condition leads to the same conclusion, so we take $\zeta = 2$. We now compute

$$\mu = \max_{|z| \leq \zeta} |Hz| = 2$$

($H = 1$) and we can finally evaluate

$$\varepsilon = \frac{1}{\mu + |J| + \beta} = \frac{2}{9}$$

($J = 2$). The derived domain of attraction is the rectangle

$$\left\{ (z, y) : \ \frac{|z|}{\zeta} \leq \varepsilon, \ \ |y| \leq \varepsilon \right\} = \left\{ (z, y) : \ |z| \leq \frac{4}{9}, \ \ |y| \leq \frac{2}{9} \right\}$$

The computed domain of attraction, the trajectories originating from the vertices and some of the system trajectories originating outside are depicted in Fig 9.28. It is apparent (and expected) that the actual domain of attraction is quite greater than the computed one.

The previously presented results can be happily (and easily) extended to the case of actuators with bounded rate. Let us now reconsider the rate-bounding

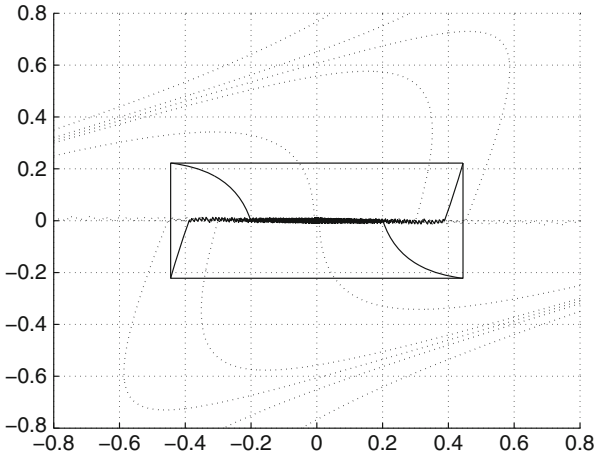


Fig. 9.28 The evaluated domain of attraction with some internal trajectories (plain lines) and some of the system trajectories originating outside (dotted curves)

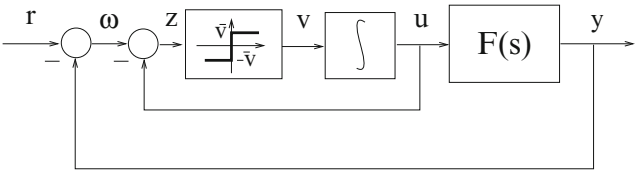


Fig. 9.29 The loop with a rate-bounding operator

operator considered in Subsection 8.4.1. In particular consider a feedback loop which includes the rate bounding operator in Figure 8.16. Here we consider the extreme case in which the loop includes the following operator

$$\dot{u}(t) = \bar{v}\ sgn[\omega - u]$$

as in Fig. 9.29. Note that this block allows a maximum increasing rate of $\bar{v}$. It is known that rate-bounds can destroy global stability of a system. However, by means of the previous results it is possible to show that a stable loop preserves local stability, achieved by a nominal control, in the presence of rate bounds, as shown in the following proposition.

Proposition 9.49. *Given the n dimensional SISO transfer function $F(s)$, if*

$$G(s) = \frac{F(s)}{1 + F(s)}$$

is asymptotically stable, then the loop represented in Fig. 9.29 is locally stable.

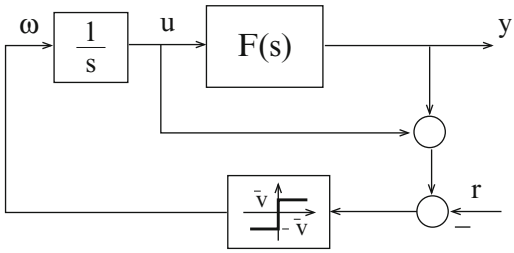


Fig. 9.30 Equivalent loop with rate-bounding operator

Proof. By "pulling out the deltas" (see [ZDG96]), say by rearranging the transfer function so as to write the linear relation between the input and the output of the nonlinear block (which in this case, plays the role of delta) it can be seen that the loop in Fig. 9.29 is equivalent to the one in Fig. 9.30 which is in turn equivalent to that achieved by feeding back the (v-to-z) transfer function

$$P(s) = \frac{1 + F(s)}{s}$$

with the *sgn* block. It is apparent that the zeros of $P(s)$ are the poles of $G(s)$ (say the poles of the closed-loop system without saturation), which are asymptotically stable by hypothesis, and that the relative degree of $P(s)$ is exactly one, since there are n zeros (the poles of $G(s)$) and $n+1$ poles (the poles of $F(s)$ plus the one in the origin). Thus, by Proposition 9.47, the system is locally stable.

It is worth stressing that the shown strict equivalence between the rate-bounding and the simply saturated problem allows to apply the construction of Proposition 9.47 to find a proper domain of attraction.

9.7.2 Planar systems

Planar systems, namely systems whose state space is two-dimensional, have several nice properties which are worth a brief presentation. The first obvious fact is that, if we limit our attention to second order systems, the major issue of the computational complexity of the considered methods almost disappears. On the other hand, limiting the investigation to this category is clearly a restriction. Still we can claim that many real-word problems are naturally represented by second order systems. Furthermore, there are many system of higher order which can be successfully approximated by second order systems. This is, for example, the case of the magnetic levitator in Fig. 2.1, which would be naturally represented by a third order system since there is an extra equation due to the current dynamics

$$L\dot{i}(t) = -Ri(t) + V(t)$$

where V is the voltage. However, since the ratio R/L is normally quite big, this equation can be replaced by the static equation $Ri(t) = V(t)$ without compromising the model (we are basically neglecting the fast dynamics).

Let us point out the main properties of second order systems. Consider the second order system $\dot{x}(t) = f(x(t))$, assume f Lipschitz, so that the problem is well-posed, and let $\bar{x}(t)$ be any trajectory corresponding to a solution of the system. Since two trajectories cannot intersect $\bar{x}(t)$ forms a barrier to any other system trajectory. This can be simply seen as follows. Consider any closed ball $\mathcal{B}$ and assume that $\bar{x}(t)$ crosses $\mathcal{B}$ side-by-side in such a way that $\bar{x}(t_1) \in \partial\mathcal{B}$, $\bar{x}(t_2) \in \partial\mathcal{B}$ and $\bar{x}(t) \in int\{\mathcal{B}\}$ for $t_1 < t < t_2$. The interior of the ball is divided in three subsets $\mathcal{B}_1$, $\mathcal{B}_2$ and $\bar{x}(t) \bigcap int\{\mathcal{B}\}$. Then no trajectory of the system originating in $\mathcal{B}_1$ can reach $\mathcal{B}_2$ without leaving $\mathcal{B}$.

Planar systems have several important properties and indeed many of the books dealing with nonlinear differential equations often have a section devoted to them. Here we propose some case studies (basically exercises) which we think are meaningful.

Example 9.50 (Stability of switching and switched systems). Consider the system $\dot{x} = A(p)x$, where

$$A = \begin{bmatrix} \alpha & 1 \\ -p(t)^2 & \alpha \end{bmatrix}, \quad p \in \{p^-, p^+\}$$

and let us consider the problems of determining:

- the supremum value of α (necessarily negative) for which the switching (p uncontrolled) system is stable;
- the supremum value of α (possibly positive) for which the switched (p controlled) system is stabilizable.

Problems of this kind can be solved by generating some extremal trajectories in view of their planar nature (see [Bos02]) and are often reported in the literature as simple examples of the following facts:

- Hurwitz stability of all elements in the convex hull of a family of matrices does not imply switching stability;
- in some particular cases it is possible to stabilize a switched system (i.e., by choosing the switching rule) whose generating matrices are all unstable and do not admit a stable convex combination;
- a linear switched system which is stabilizable via feedback is not necessarily consistently stabilizable.

As a first step we notice that, for fixed p, the solution has the following form

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = e^{\alpha(t-t_0)} \begin{bmatrix} \cos(p(t-t_0)) & \frac{1}{p}\sin(p(t-t_0)) \\ -p\sin(p(t-t_0)) & \cos(p(t-t_0)) \end{bmatrix} \begin{bmatrix} x_1(t_0) \\ x_2(t_0) \end{bmatrix}$$

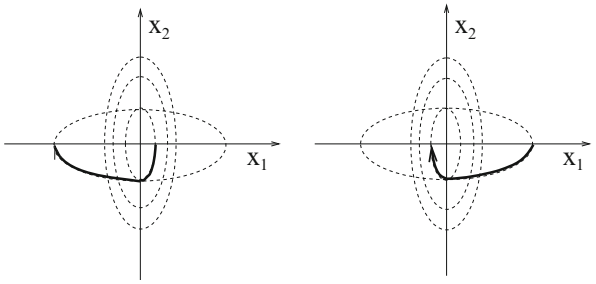


Fig. 9.31 The worst-case and the best case strategy for $\alpha = 0$

To analyze the stability of the switched system, let us start from the point $x(0) = [1\ 0]^T$ and let us assume $\alpha = 0$. It is apparent that with both values p^- and p^+, the corresponding trajectories reach the x_2 negative axis in time $\pi/(2\ p)$, but if we choose the upper value p^+ the trajectory is below (and then external to) that corresponding to the lower value. When the x_2 negative axis is reached, the situation is exactly the opposite. Then, by taking the worst case trajectory, generated by means of the following strategy

$$p = \begin{cases} p^+ & \text{if } x_1 x_2 < 0 \\ p^- & \text{if } x_1 x_2 > 0 \end{cases}$$

we achieve a trajectory which is the most external (see Fig. 9.31 left).

If we consider the time instants in which the axes are crossed, since we keep p constant inside any sector, we get the following discrete relation

$$\begin{bmatrix} x_1(t_k) \\ x_2(t_k) \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{p} \\ -p & 0 \end{bmatrix} \begin{bmatrix} x_1(t_{k-1}) \\ x_2(t_{k-1}) \end{bmatrix}$$

with $p \in \{p^-, p^+\}$.

Take the initial vector $[\ 1\ \ 0\ ]^T$. If we consider the worst-case trajectory, we encircle the origin and we reach the positive x_1 axis again, at time $t = 2\pi/p$, in a point $[\xi\ 0]^T$ where

$$\xi = \frac{(p^+)^2}{(p^-)^2} \geq 1$$

(the equality holds only if $p^- = p^+$).

By taking into account α again, we can find the largest value of $\alpha < 0$ which assures stability. This limit value is such that

$$e^{\alpha 2\pi/p}\xi = 1$$

namely

$$\alpha = -\frac{p}{2\pi}\log(\xi)$$

which is negative. To prove this, one has to take into account the fact that the solution with a generic value α is $x_\alpha(t) = e^{\alpha t}x_0(t)$, where $x_0(t)$ is that achieved with $\alpha = 0$.

To solve the "switched" problem we have just to proceed in the same way and consider the "best trajectory," which is clearly given by (see Fig. 9.31 right)

$$p = \begin{cases} p^+ & \text{if } x_1x_2 > 0, \\ p^- & \text{if } x_1x_2 < 0 \end{cases}$$

The expression is identical, with the difference that α turns out to be the opposite.

This system with $\alpha > 0$ is an example of a system which can be stabilized via state feedback, but which is not consistently stabilizable. Indeed one can show that, no matter how a fixed function $p(t)$ is taken with $p^- \le p(t) \le p^+$,[8] the system trajectory will diverge for some initial conditions.

To this aim, let us consider the case of a linear time-varying system

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

and denote by $\Phi(t, t_0)$ the state transition matrix for $A(t)$:

$$\frac{d}{dt}\Phi(t, t_0) = A(t)\Phi(t, t_0), \qquad \Phi(t_0, t_0) = I.$$

Then, by the Jacobi-Liouville formula,

$$\det(\Phi(t, t_0)) = \exp\left(\int_{t_0}^{t} \operatorname{trace}(A(\tau))d\tau\right).$$

Indeed

$$\frac{d}{dt}(\log\ \det(\Phi(t, t_0))) = \operatorname{trace}(\Phi(t, t_0)^{-1}A(t)\Phi(t, t_0)) = \operatorname{trace}(A(t)).$$

If we look back at the example we have that $\operatorname{trace}(A(p)) = 2\alpha$, hence

$$\det(\Phi(t, t_0)) = \exp\left(\int_{t_0}^{t} \operatorname{trace}(A(\tau))d\tau\right) = e^{2\alpha(t-t_0)}$$

and thus $\det(\Phi(t, t_0))$ grows arbitrarily large. Therefore some elements of $\Phi(t, t_0)$ grow arbitrarily large as $t \to +\infty$.

[8] Including the switching case in which $p(t) \in \{p^-, p^+\}$.

9.8 Exercises

1. Find the analytic expression of the function in Figure 9.1.
2. Prove the stability of the system proposed in Example 7.16 for $|\delta| \leq \rho < 1$, by using the technique of extremal trajectories adopted in Example 9.50. Hint: the extremal trajectories are as in Fig. 9.32
3. Prove that the P set $\mathcal{P}(F) \bigcap \mathbb{R}^+$, $F \geq 0$, is bounded if and only if for any j there exists k such that $F_{kj} > 0$.
4. Prove Proposition 9.9. Hint: Consider $P = I$ as a Lyapunov matrix. Then, if all the A_q are symmetric, the Lyapunov inequality which assures quadratic stability is

$$A_q^T + A_q = 2A_q < 0, \quad \text{for all } q,$$

which is assured if the A_q are Hurwitz ....
5. Let A be Metzler and not Hurwitz. Then $A + \epsilon O$ ($O_{ij} = 1$) is non-Hurwitz as well for $\epsilon > 0$. Provide a proof.
6. For discrete-time switched systems, show that the existence of a stable matrix in the convex hull is not sufficient for stabilizability. (Hint: try in dimension one).
7. For discrete-time switched systems, show that the existence of a stable matrix in the convex hull is not necessary for stabilizability. Find two non-negative matrices of order 2, A_1 and A_2, such that no Schur convex combination exists, but the corresponding switched system is stabilizable. (Hint: try two diagonal matrices).
8. Show that, for discrete-time positive switched systems, the existence of a Schur matrix in the convex hull is sufficient for stabilizability. (Hint. Assume that the Schur matrix in the convex hull is irreducible and take the left Frobenius eigenvector $z^T > 0$ and the copositive function $z^T x \ldots$).

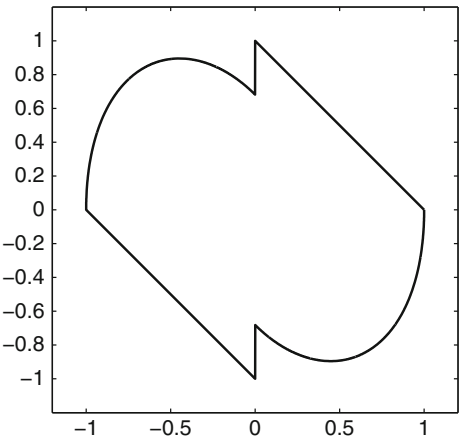


Fig. 9.32 Extremal trajectories

9. Paradox of the 0 transfer functions. Let A_i, B, C be a family of systems which are Hurwitz for fixed i, but not switching stable. Assume that each of the systems is stabilizable, for instance quadratically. By Theorem 9.45, we would have that we can choose compensator transfer functions $W_i(s) \equiv 0$, which can be implemented so that stability is assured under switching. Is this impossible, or not? (see [BMM09]).
10. Show that any minimum-phase system of relative degree 1 with transfer function $P(s)$ can be written in the form (9.26)–(9.27), where the eigenvalues of F are coincident with the zeros of $P(s)$.

Chapter 10
(Sub-)Optimal Control

There are several optimal control problems that can be naturally cast in a set-theoretic framework and for which set-theoretic techniques provide efficient tools. Although by far non-exhaustive, this section considers several cases and discusses several solutions.

10.1 Minimum-time control

In this section, minimum-time set reachability and controllability problems under disturbances are analyzed. To this aim, we reconsider and extend the previously introduced definitions of controllability and reachability to and from a set. More precisely, we deal with the concept of worst-case controllability, introduced next.

10.1.1 Worst-case controllability

In the presence of uncertainty the concept of controllability can be expressed in a dynamic game framework. Consider the dynamic system

$$x(t+1) = f(x(t), u(t), w(t))$$

(or $\dot{x}(t) = f(x(t), u(t), w(t))$) where $u(t) \in \mathcal{U}$, the control and $w(t) \in \mathcal{W}$, the disturbance (w possibly includes parameter and noise uncertainties). We say that $\mathcal{P}$ is controllable in the worst case sense to $\mathcal{S}$ in time T if for all $x(0)$ in $\mathcal{P}$ there exists a control such that $x(T) \in \mathcal{S}$ for all possible $w(t) \in \mathcal{W}$. The above problem can be thought of as a game between two opponents: a good girl u, who has the goal of driving $x(t)$ to $\mathcal{S}$, playing against a bad boy w, her opponent, who wishes her to fail.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_10

Then $\mathcal{S}$ is worst-case-controllable if she wins the game (being smart enough to apply the proper strategy). We remind the reader that a fundamental distinction about the nature of the control has to be made, since the worst-case-controllability property depends on the chosen type of control, which can fall in any of these categories:

- open-loop control: $u(t) = \Phi(x(0), t)$;
- state feedback control $u(t) = \Phi(x(t))$;
- full information feedback control $u(t) = \Phi(x(t), w(t))$;

As pointed out in the introduction, these three rules of the game are deeply different, as we can see by considering the system

$$x(t+1) = x(t) + u(t) + w(t)$$

with $w \in [-1, 1]$ and $u \in [-1, 1]$. It turns out that, in the time horizon T, the origin is controllable to $[-T, T]$ by means of the open-loop control $u \equiv 0$ (other choices produce different, but not better, results), is controllable to $[-1, 1]$ by means of a state feedback control (for instance $u = -\text{sat}(x)$) and it is controllable to $\{0\}$ by means of a full information control (write the control $u = \Phi(x, w)$ as an exercise).

For the sake of completeness (only), we just comment on the concept of reachability in the worst case sense, which can be defined by considering the backward system: $\mathcal{S}$ is worst-case reachable from $\mathcal{P}$ if $\mathcal{S}$ is controllable to $\mathcal{P}$ for the backward system (see Exercise 1).

Note that in Section 5.1 the concept of worst-case controllability has been already exploited, being intrinsic in the infinite-time reachability set construction. Indeed, the backward construction of the admissible sets is based on the recursive computation of the sets $\mathcal{X}_{-k}$ which are nothing but the k-steps worst case controllability sets to $\mathcal{X}$ under constraints (the preimage sets (5.2) are special cases, precisely they are one-step worst case controllability sets). The following theorem, concerning the controllability sets with state feedback, holds [Bla92, Bla94]

Theorem 10.1. *Consider the system*

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

$d \in \mathcal{D}$, $u \in \mathcal{U}$, where $\mathcal{D}$ and $\mathcal{U}$ are both convex and closed sets and $\mathcal{D}$ is bounded. Assume that $A(w)$ and $B(w)$ are continuous functions of $w \in \mathcal{W}$, $\mathcal{W}$ a compact set. Assume that there exists $w \in \mathcal{W}$ such that $B(w)$ has full column rank. Let $\mathcal{C}_T(\mathcal{X})$ be the T-step controllability set to $\mathcal{X}$ via state feedback control $u = \Phi(x)$. We have that

i) if $\mathcal{X}$ is a closed set, then $\mathcal{C}_T(\mathcal{X})$ is closed;
ii) if $\mathcal{X}$ is a convex set, then $\mathcal{C}_T(\mathcal{X})$ is convex;
iii) if $A(w)$ and $B(w)$ are polytopes of matrices and $\mathcal{X}$ and $\mathcal{U}$ are polyhedra, then $\mathcal{C}_T(\mathcal{X})$ is a polyhedron.

Proof. All of the above statements will be proved for the one-step controllability set only, since the rest of the proof follows by induction.

i) Assume that $\mathcal{X}$ is closed and let $z_k \in \mathcal{C}$ be a sequence converging to $\bar{z}$. Proving that $\mathcal{C}$ is closed amounts to showing that $\bar{z} \in \mathcal{C}$. By definition, there exists a sequence $u_k(z_k)$ such that

$$A(w)z_k + B(w)u_k + Ed \in \mathcal{X}$$

for all $w \in \mathcal{W}$ and $d \in \mathcal{D}$. Since A and B are continuous, $\mathcal{W}$ is compact and $\mathcal{D}$ is bounded, the sequence $\{u_k\}$ is necessarily bounded and therefore it admits a subsequence $\{u'_k\}$ which converges to $\bar{u} \in \mathcal{U}$, since $\mathcal{U}$ is closed. Let us consider the subsequence $\{z'_k\}$ corresponding to the subsequence $\{u'_k\}$. Clearly $z'_k \to \bar{z}$. We must show that for all $w \in \mathcal{W}$ and $d \in \mathcal{D}$

$$\bar{x} = A(w)\bar{z} + B(w)\bar{u} + Ed \in \mathcal{X}$$

which implies $\bar{z} \in \mathcal{C}$. By construction $x'_k = A(w)z'_k + B(w)u'_k + Ed \in \mathcal{X}$. For all $w \in \mathcal{W}$ and $d \in \mathcal{D}$

$$\begin{aligned} \|x'_k - \bar{x}\| &= [\|A(w)z'_k + B(w)u'_k + Ed] - [A(w)\bar{z} + B(w)\bar{u} + Ed]\| \leq \\ &\leq \|A(w)\|\|z'_k - \bar{z}\| + \|B(w)\|\|u'_k - \bar{u}\| \to 0, \end{aligned}$$

then $x'_k \to \bar{x}$ and, since $\mathcal{X}$ is closed, $\bar{x} \in \mathcal{X}$.

ii) To show convexity, consider again the one-step controllability set $\mathcal{C}_1(\mathcal{X})$ to a convex set $\mathcal{X}$ (convexity for the controllable set $\mathcal{C}_T(\mathcal{X})$ is achieved recursively). Let x_1 and x_2 be in $\mathcal{C}_1(\mathcal{X})$. By definition there exists two control values $u(x_1)$ and $u(x_2)$ such that $A(w)x_1 + B(w)u(x_1) + Ed \in \mathcal{X}$ and $A(w)x_2 + B(w)u(x_2) + Ed \in \mathcal{X}$ for all $w \in \mathcal{W}$ and $d \in \mathcal{D}$. Consider the vector $\alpha x_1 + (1-\alpha)x_2$ and the associated control value $u(x) = \alpha u(x_1) + (1-\alpha)u(x_2)$. Then, for all $w \in \mathcal{W}$ and $d \in \mathcal{D}$,

$$\begin{aligned} A(w)x + B(w)u(x) + Ed = \alpha[A(w)x_1 + B(w)u(x_1) + Ed] \\ + (1-\alpha)[A(w)x_2 + B(w)u(x_2) + Ed] \in \mathcal{X} \end{aligned} \tag{10.1}$$

as long as $\mathcal{X}$ and $\mathcal{U}$ are convex.

iii) The last statement is proved constructively. Let $\mathcal{X} = \mathcal{P}(F, g)$ and consider the polyhedral set

$$\mathcal{M} = \{(x,u) : [A(w)x + B(w)u + Ed] \in \mathcal{X},\ u \in \mathcal{U},\ \forall\, w \in \mathcal{W},\ \text{and}\ \forall d \in \mathcal{D}\} \subset \mathbb{R}^{n+m}$$

The one-step controllability set is given by the projection $\mathcal{P}$ of $\mathcal{M}$ on the subspace associate with the first n components:

$$\mathcal{P} = \{x :\ \text{there exists } u \ \text{ such that } \ (x,u) \in \mathcal{M}\}$$

Since

$$[A(w)|B(w)] = \sum_{i=1}^{s} [A_i|B_i]w_i, \quad \sum_{i=1}^{s} w_i = 1, \;\; w_i \geq 0,$$

the following expression is valid

$$\mathcal{M} = \{(x,u) : \; F[A_k x + B_k u + Ed] \leq g, \; u \in \mathcal{U}, \; k = 0, 1 \ldots s, \; \forall \, d \in \mathcal{D}\} \subset \mathbb{R}^{n+m}$$

Set $\mathcal{M} \subset \mathbb{R}^{n+m}$ can be expressed by considering the erosion of $\mathcal{X}$ with respect to $E\mathcal{D}$ (see Proposition 3.28) $\mathcal{P}[F, \tilde{g}]$, where

$$\tilde{g}_i = g_i - \max_{d \in \mathcal{D}} \; F_i E d = g_i - \phi_{\mathcal{D}}(F_i E),$$

where $\phi_{\mathcal{D}}(\cdot)$ is the support function of $\mathcal{D}$. Then

$$\mathcal{M} = \{(x,u) : \;\; F[A_k x + B_k u] \leq \tilde{g}, \; u \in \mathcal{U} \;\; k = 1, \ldots s\} \subset \mathbb{R}^{n+m}.$$

Thus $\mathcal{M}$ is a polyhedron and therefore, according to Proposition 3.28, its projection $\mathcal{P}$ is a polyhedron.

Note that the last claim of the previous theorem holds even if $\mathcal{D}$ is not a polytope. Actually, it can be any convex and compact set (provided that $\phi_{\mathcal{D}}$ is computable). Note also that, in the theorem, a state feedback controller was assumed. The following corollary extends the result of the previous theorem to full information controllers of the form $u = \Phi(x, w)$.

Corollary 10.2. *Under the same assumptions of Theorem 10.1, let $\mathcal{C}_T(\mathcal{X})$ be the T-step controllability set to $\mathcal{X}$ with full information control $u = \Phi(x, w)$. We have that*

- *if $\mathcal{X}$ is a closed set, then $\mathcal{C}_T(\mathcal{X})$ is closed;*
- *if $\mathcal{X}$ is a convex set and B is a certain matrix, then $\mathcal{C}_T(\mathcal{X})$ is convex;*
- *if $A(w)$ is a matrix polytope, $\mathcal{X}$ and $\mathcal{U}$ are polyhedra and $B(w) = B$ is a certain matrix, then $\mathcal{C}_T(\mathcal{X})$ is a polyhedron.*

Proof. The proofs of the first two statement are basically the same of Theorem 10.1 and are not reported.

To prove the third statement, let us characterize the one-step controllability set $\mathcal{C}(\mathcal{X})$. For each k, consider the following set

$$\mathcal{M}_k = \{(x,u) : [A_k x + Bu(x,k) + Ed] \in \mathcal{X}, \;\; u(x,k) \in \mathcal{U}, \; \forall d \in \mathcal{D}\} \subset \mathbb{R}^{n+m}$$

Consider the polyhedral projection of this set on the first n components

$$\mathcal{P}_k = \{x : \;\; \text{there exists } u \;\; \text{such that} \;\; (x,u) \in \mathcal{M}_k\}.$$

The set $\mathcal{P}_k$ is the one-step controllability set to $\mathcal{X}$ for the kth system. Clearly $\mathcal{C}(\mathcal{X}) \subseteq \mathcal{P}_k$ for all k because the uncertain matrix $A(w)$ can be equal to A_k.

Consider the polyhedral set

$$\mathcal{P} = \bigcap_k \mathcal{P}_k.$$

Then $\mathcal{C}(\mathcal{X}) \subseteq \mathcal{P}$. The inclusion is actually an equality if one shows that any state of z of $\mathcal{P}$ can be driven to $\mathcal{X}$ by a control of the form $\Phi(x, w)$. If $z \in \mathcal{P}$, then $z \in \mathcal{P}_k \forall k$, so that

$$[A_k z + Bu(x,k) + Ed] \in \mathcal{X}, \;\; k = 1, \ldots, s.$$

For any given $w \geq 0$, $\bar{1}^T w = 1$ consider the control

$$u(x,w) = \sum_{k=1}^{s} w_k \, u(x,k)$$

to get

$$A(w)z + Bu(x,w) = \sum_{k=1}^{s} w_k \, [A_k z + Bu(x,k) + Ed] \in \mathcal{X},$$

hence the claim.

The next theorem concludes this subsection.

Theorem 10.3. *Assume that $\mathcal{X}$ is a C-set. Then the following statements are equivalent.*

i) $\mathcal{C}_{k-1}(\mathcal{X}) \subseteq \mathcal{C}_k(\mathcal{X})$ *for all* $k \geq 1$.
ii) The set $\mathcal{X}$ is robustly controlled-invariant.
iii) Each element of the sequence $\mathcal{C}_k(\mathcal{X})$ is robustly controlled invariant.

Proof. Let us show the equivalence between i) and ii). Clearly if the first condition holds, for $k = 1$ we have $\mathcal{X} = \mathcal{C}_0(\mathcal{X}) \subseteq \mathcal{C}_1(\mathcal{X})$, which is equivalent to robust controlled invariance. Conversely let us assume that $\mathcal{X}$ is robustly controlled-invariant. Then $\mathcal{X} = \mathcal{C}_0(\mathcal{X}) \subseteq \mathcal{C}_1(\mathcal{X})$ holds. By construction, each state $x \in \mathcal{C}_{k-1}(\mathcal{X})$ can be driven to $\mathcal{X}$ in $k-1$ steps. If $\mathcal{X}$ is invariant, then once the condition $x(k-1) \in \mathcal{X}$ is assured there exists a control value such that $x(k) \in \mathcal{X}$. This means that x can also be driven to $\mathcal{X}$ in k steps, namely $x \in \mathcal{C}_k(\mathcal{X})$. Thus $\mathcal{C}_{k-1}(\mathcal{X}) \subseteq \mathcal{C}_k(\mathcal{X})$. The fact that iii) is equivalent to the previous ones is easy to show and is left to the reader as an exercise.

10.1.2 Time optimal controllers for linear discrete-time systems

Conceptually, the time-optimal control in discrete-time has a simple solution based on the controllability sets [GC86a, KG87]. Given the system

$$x(t+1) = Ax(t) + Bu(t),$$

with the constraint $u(t) \in \mathcal{U}$ and, possibly, $x(t) \in \mathcal{X}$, where both $\mathcal{U}$ and $\mathcal{X}$ are C-sets, one has to compute the k-step controllability sets to the origin, precisely $\mathcal{C}_k \doteq \mathcal{C}_k(0)$. Then, given an initial condition x, the minimum time to drive the state to the origin, by construction, can be expressed as

$$T_{min}(x) = \min\{k : x \in \mathcal{C}_k\}$$

The corresponding minimum-time control law is any control function $\Phi(x)$ achieved as a selection as follows:

$$\Phi(x) \in \Omega(x) = \{u \in \mathcal{U} : Ax + Bu \in \mathcal{C}_{T-1}\}, \quad \text{where } T \doteq T_{min}(x). \tag{10.2}$$

The fact that the set-valued function $\Omega(x)$ is non-empty is a simple consequence of the definition of the sets $\mathcal{C}_k$. So simple is the idea, so hard its implementation. Indeed, the controllability sets $\mathcal{C}_k$ may have an arbitrarily complex representation and therefore the method can be not realistic for high dimensional systems (for problems in which a long time-horizon is necessary, the reader is referred to the examples proposed in the original works [GC86a, KG87] to have an idea of the complexity). Here we do not report further details since, in the next section, we will consider the most general problem of controllability of a set, thus including the 0-controllability as a special case.

10.1.3 Time optimal controllers for uncertain systems

We consider now the minimum-time problem for polytopic uncertain systems of the form

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Ed(t)$$

with the polytopic constraints $u \in \mathcal{U}$, $x \in \mathcal{X}$ and with the uncertainty polytopic bounds $w \in \mathcal{W}$, $d \in \mathcal{D}$. It is quite obvious that, due to the uncertainty, the problem of reaching the origin in finite time has no solution in general (even if there is no additive uncertainty, i.e. $E = 0$, see Exercise 2). Therefore, the only reasonable way to formulate the problem is that of reaching a target set in minimum-time. This problem has two versions.

- Weak version: just reach the set in minimum time;
- Strong version: reach the set in minimum time and remain inside forever.

This problem was solved in [Bla92]. Subsequent contributions are in [MS97]. The definition of controllability under constraints is introduced next. Note that this definition is new, since we consider also the constraints on x.

Definition 10.4 (Controllable state under constraints). The state x_0 is controllable under constraints to the compact set $\mathcal{G}$ in T steps if there exists a feedback control $u = \phi(x)$ and $T \geq 0$ such that, for $x(0) = x_0$, $u(t) \in \mathcal{U}$, $x(t) \in \mathcal{X}$ for $t \geq 0$ and

$$x(T) \in \mathcal{G}$$

for any $w(t) \in \mathcal{W}$ and $d(t) \in \mathcal{D}$.

Definition 10.5 (Ultimately boundable state under constraints). The state x_0 is ultimately boundable within the compact set $\mathcal{G}$ in T steps if there exists a feedback control $u = \phi(x)$ and $T \geq 0$ such that, for $x(0) = x_0$, $u(t) \in \mathcal{U}$, $x(t) \in \mathcal{X}$, for $t \geq 0$, and

$$x(t) \in \mathcal{G}, \quad \text{for } t \geq T,$$

for any $w \in \mathcal{W}$, $d \in \mathcal{D}$.

Analogous definitions hold for controllers of the form $u = \phi(x, w)$. With the above in mind, the following two minimum-value functions can be introduced, precisely:

- the minimum reaching time $T_R(x_0) = \min \ k \geq 0$ for which x_0 is controllable to $\mathcal{G}$ in k steps
- the minimum confinement time $T_U(x_0) = \min \ k \geq 0$ for which x_0 is ultimately boundable to $\mathcal{G}$ in k steps.

In general the two values are not equal and clearly $T_R(x_0) \leq T_U(x_0)$. We remark the fact that the indices above refer to the worst case, say that a state might not be controllable (ultimately boundable) but still it might be driven to $\mathcal{G}$ under favorable disturbances d and w.

We are now able to formulate the following problems.

Problem 10.6 (Minimum-Time Controllability Problem). Find a feedback control strategy $u = \Phi(x)$ which minimizes $T_R(x_0)$ for all $x \in \mathcal{X}$.

Problem 10.7 (Minimum-Time Ultimate Boundedness Problem). Find a feedback control strategy $u = \Phi(x)$ which minimizes $T_U(x_0)$ for all $x \in \mathcal{X}$.

Henceforth we assume that the target set $\mathcal{G}$ is a C-set. According to the theory developed in the previous subsection, Problem 10.6, namely reaching the target,

can be solved by evaluating the controllable sets to $\mathcal{G}$, with the only modification that the admissible environment has to be taken into account.

In view of the constraints $x \in \mathcal{X}$, the one-step controllability set to a C-set $\mathcal{G}$ is now defined as

$$\mathcal{C}(\mathcal{G}) = \{x \in \mathcal{X} : \exists u \in \mathcal{U} \text{ s.t. } A(w)x + B(w)u + Ed \in \mathcal{G},\ \forall w \in \mathcal{W},\ \text{and } d \in \mathcal{D}\}\,.$$

The T-step controllability sets to $\mathcal{G}$ can be recursively defined as

$$\mathcal{C}_0 = \mathcal{G}, \quad \mathcal{C}_{T+1} = \mathcal{C}(\mathcal{C}_T)$$

The minimum-time reaching problem can be solved exactly as in the case of systems without uncertainty, described in Section 10.1.2.

The requirement of reaching the target set in T steps may be not suitable for applications, because what is desired in practice is to ultimately bound the state in $\mathcal{G}$. The following theorem is a preliminary step for handling this problem (for a proof see [Bla92]).

Theorem 10.8. *The state x_0 is ultimately boundable in the C-set $\mathcal{G}$ in T steps if and only if it is controllable in T steps to a controlled invariant set $\mathcal{P} \subseteq \mathcal{G}$ or, equivalently, to the largest controlled invariant set $\mathcal{P}_{max}$ in $\mathcal{G}$.*

Note that the target set $\mathcal{G}$ is controlled-invariant if and only if

$$T_R(x_0) = T_U(x_0)$$

for all $x_0 \in \mathcal{X}$. As a consequence of the theorem, to solve Problem 10.7 the following steps must be performed:

Step 1: compute the set $\mathcal{P}_{max}$
Step 2: find a control $\Phi(x)$ solving Problem 10.6 for $\mathcal{P}_{max}$, namely solving the problem of driving the state to $\mathcal{P}_{max}$ in minimum time.

We remind now that in view of Theorem 10.3, since $\mathcal{P}_{max}$ is robustly controlled-invariant, the controllability sets $\mathcal{C}_k(\mathcal{P}_{max})$ are all controlled-invariant and satisfy the inclusion

$$\mathcal{C}_{k-1}(\mathcal{P}_{max}) \subset \mathcal{C}_k(\mathcal{P}_{max})$$

Therefore we have a sequence of nested controlled invariant sets which has the following property [Bla92].

Proposition 10.9. *The sequence of sets $\mathcal{C}_k(\mathcal{X})$ converges to the domain of attraction inside $\mathcal{X}$ to $\mathcal{G}$, precisely the set of all the points in $\mathcal{X}$ that can be ultimately confined in $\mathcal{G}$.*

In the case in which all the considered sets are polyhedral and the system is polytopic

$$A(w) = \sum_{i=1}^{s} w_i A_i, \quad B(w) = \sum_{i=1}^{s} w_i B_i,$$

with $\sum_{i=1}^{s} w_i = 1$ and $w_i \geq 0$, the construction of the sets $\mathcal{C}_k(\mathcal{X})$ requires at each step the solution of the linear programming problems described in the previous sections. Once such a set has been found, the control can be computed on-line as

$$u \in \Omega(x),$$

where the control map $\Omega(x)$ is defined in eq. (10.2) and where $T_{min}(x)$ is properly replaced by $T_U(x)$.

Function Φ is a strategy that solves the minimum-time problem if and only if it fulfils condition $\Phi(x) \in \Omega(x)$. Among all the functions which fulfil such a condition, we can choose those which minimize $\|u\|$ (minimal selection [Aub91]). It is simple to show that $\Omega(x)$ is a polyhedral set. The feedback control is obtained by computing u on-line, by solving the linear programming problem of selecting a vector inside $\Omega(x)$. If the control space dimension is low, as it is often the case, such a problem may be easily solved on-line (it is extremely simple in the single-input case).

Example 10.10 (Minimum-time reachability of a prescribed level set). Let us consider the water flow system with two reservoirs represented in Figure 10.1 The upper reservoir is fed by a controlled flow while the lower one is fed by gravity through a pipeline which connects it to the first one. The water flows out by gravity from the second reservoir. We assume that in each reservoir there is an uncertain flow, which is upper and lower bounded, representing the natural contribute (rain

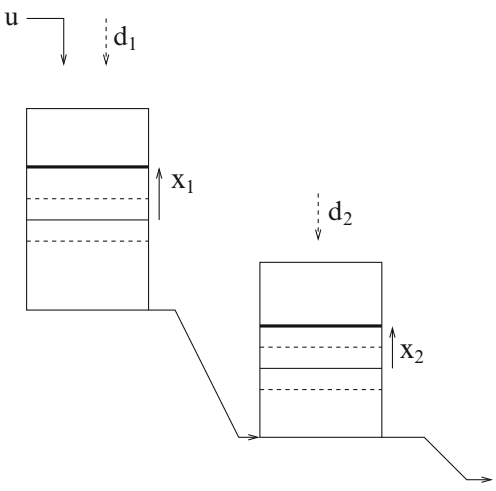


Fig. 10.1 The reservoir system

and water streams). The linearized model (which keeps into account non-linear and uncertain parameters) is

$$x_1(t+1) = x_1(t) - w_1(x_1(t) - x_2(t)) + d_1(t) + u(t)$$
$$x_2(t+1) = x_2(t) + w_1(x_1(t) - x_2(t)) - w_2 x_2(t) + d_2(t)$$

where $x_1(t)$ and $x_2(t)$ are the water levels of the upper and lower reservoirs (with respect to the nominal reference level), $u(t)$ is the controlled input flow and $d_1(t)$, $d_2(t)$ are disturbance flows. The uncertainties are assumed to be bounded as

$$w_1^- \le w_1(t) \le w_1^+, \;\; w_2^- \le w_2(t) \le w_2^+,$$
$$d_1^- \le d_1(t) \le d_1^+, \;\; d_2^- \le d_2(t) \le d_2^+,$$

while state and control values are constrained as

$$x_1^- \le x_1(t) \le x_1^+, \;\; x_2^- \le x_2(t) \le x_2^+, \;\; u^- \le u(t) \le u^+.$$

Assume the target set is

$$\mathcal{G} = \{(x_1, x_2) : g_1^- \le g_1 \le g_1^+, \;\; g_2^- \le g_2 \le g_2^+\}.$$

and the system parameters are $w_1^- = w_2^- = 1$, $w_1^+ = w_2^+ = 2$, $d_1^+ = d_2^+ = -d_1^- = -d_2^+ = 0.1$ $u^+ = -u^- = 0.1$, $x_1^+ = x_2^+ = -x_1^- = -x_2^- = 3$, $g_1^+ = g_2^+ = -g_1^- = -g_2^- = 1$. The maximal controlled invariant set in $\mathcal{G}$ turns out to be

$$\mathcal{P}_{max} = \{x_1, x_2 : -1 \le x_1 \le 1, \quad -1 \le x_2 \le 1, \quad -1 \le -0.111x_1 + x_2 \le 1\}$$

The largest controlled invariant set, i.e. the domain of attraction, was found in ten steps when the following condition

$$\mathcal{C}_{10}(\mathcal{P}_{max}) = \mathcal{C}_{11}(\mathcal{P}_{max})$$

was detected. This set turns out to be

$$\mathcal{A} = \{(x_1, x_2) : -3 \le x_1 \le 3, -3 \le x_2 \le 3, -3 \le -0.111x_1 + x_2 \le 3,$$
$$-3 \le -0.222x_1 + x_2 \le 3.333\}$$

The controllability sets to $\mathcal{P}_{max}$ are represented in Fig. 10.2. Denoting by $s(i)$ the number of constraints of $\mathcal{C}_T(\mathcal{X})$, the number of constraints defining each of the sets results in:

$$s(0) = 6, \; s(1) = 8, \; s(2) = 10, \; s(3) = 10, \; s(4) = 12, \; s(5) = 14,$$
$$s(6) = 10, \; s(7) = 12, \; s(8) = 10, \; s(9) = 8, \; s(10) = 8.$$

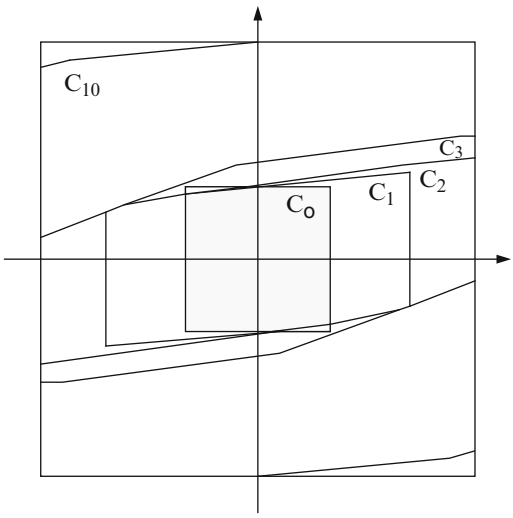


Fig. 10.2 The sequence of controllability sets

Hence, in this case, the complexity of the problem to be solved on-line is absolutely reasonable.

Recent developments of the proposed theory for non-linear systems are found in [PP14].

10.2 Optimal peak-to-peak disturbance rejection

A problem that can be solved successfully via a set-theoretic approach is the optimal disturbance rejection problem. Consider the system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Ed(t) \\ y(t) &= Cx(t) + Du(t) + Fd(t) \end{aligned}$$

and the stabilizing control $u = \Phi(x)$. Assume that the signal d is unknown but bounded as follows:

$$\|d(t)\|_\infty \leq 1,$$

(note that possible non-unit bounds can be accommodated by a proper choice of the matrices E and F). Let us consider the following definition.

Definition 10.11 (Disturbance rejection level). The control $u = \Phi(x)$ achieves a level of disturbance rejection μ if for all d as previously specified and for $x(0) = 0$, the closed-loop system output satisfies

$$\|y(t)\|_\infty \leq \mu$$

The previous expression is equivalent to $-\mu \leq y_i(t) \leq \mu$ or, more formally, $y \in \bar{\mathcal{P}}(I, \mu\bar{1})$. Note that these constraints imposed on the output y include several important cases. By means of a proper choice of matrices C, D, and F, different types of constraints assigned independently on x and u can be considered. For instance, when

$$C = \begin{bmatrix} C_1 \\ 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ D_2 \end{bmatrix}, \quad F = 0,$$

the constraints are of the polytopic form, precisely

$$x \in \{x : \ \|Cx\| \leq \mu\}, \quad u \in \{u : \ \|Du\| \leq \mu\}.$$

The peak-to-peak disturbance rejection problem is formulated next.

Problem 10.12. Find a stabilizing controller which minimizes the disturbance rejection level.

This problem can be solved with set-theoretic methods in a sub-optimal way by considering (iteratively) the following feasibility problem.

Problem 10.13 (μ-level disturbance rejection problem). Given $\mu > 0$ reply with an yes–no answer to the following question: does there exist a stabilizing controller which achieves the disturbance rejection level $\mu > 0$? If the answer is yes, find the controller.

In principle, solving the second problem is equivalent to solving the first, since one can iterate over μ, as we will show soon. Take any stabilizing controller (assume that it exists), for instance a linear gain $u = Kx$. Compute the performance of such a controller. This can be done by computing the l_1 norm of the closed-loop system

$$\begin{aligned} x(t+1) &= (A + BK)x(t) + Ed(t) \\ y(t) &= (C + DK)x(t) + Fd(t) \end{aligned}$$

as shown in Chapter 6, Subsection 6.2.2. Once this value is evaluated we have an upper bound

$$\mu_{ini} = \|(C + DK)(zI - A - BK)^{-1}E + F\|_{l_1}.$$

This is the smallest value for which the 0-reachable state of the system with the control $u = Kx$ is included in the set (see Propositions 6.16 and 6.17)

$$\{x : |[C + DK]_i x| \leq \mu - \|F_i\|_1, \quad i = 1, 2, \ldots, p\}, \tag{10.3}$$

where $[C + DK]_i$ and F_i are the ith rows of $C + DK$ and F, respectively. The so determined μ_{ini} is an initial upper bound for μ.

Before stating the main theorem, let us introduce these mild assumptions

i) the matrix E has full row rank;
ii) there exists $\nu > 0$ such that set $\mathcal{S}(\mu)$ is included in the ball $\{x : \|x\| \leq \mu\nu\}$.

Note that in practice the above assumptions do not affect the problem since it is always possible to add a fictitious disturbance $\tilde{d}$ as follows

$$x(t + 1) = Ax(t) + Bu(t) + Ed(t) + \tilde{d}(t),$$

with $\|\tilde{d}(t)\| \leq \epsilon$ to meet assumption i). Furthermore, it is absolutely reasonable to require that the system state is bounded as in ii). This can be achieved by extending C and D as follows

$$C_{aug} = \begin{bmatrix} C \\ \epsilon I \end{bmatrix}, \quad D_{aug} = \begin{bmatrix} 0 \\ D \end{bmatrix},$$

since, for ϵ small enough, the constraints are not essentially affected. The following theorem holds.

Theorem 10.14. *The largest controlled invariant set compatible with*[1] *the set (10.3) is not empty (equivalently, by symmetry, includes* 0*) if and only if there exists a control $u = \Phi(x)$ which achieves a level of disturbance rejection μ.*

The previous theorem can be inferred by set-theoretic type papers (see [BR71a, Ber72, GS71, BU93]). However the important link between these technique and the l_1 theory (see [BG84, DP87]) was established by Shamma [Sha96b] who proposed the theorem as it is formulated. Note that we can always associate a static controller with any controlled-invariant set. Therefore the next corollary holds.

Corollary 10.15. *Dynamic state feedback controllers cannot outperform static ones.*

We remark that the previous corollary is not valid in the class of linear controllers, since the optimal linear compensator can be dynamic even in the state feedback case [DBD92]. On the other hand, limiting the attention to linear compensators is a restriction, since those can be outperformed by nonlinear ones [Sto95].

[1] See Definition 5.12.

We can approach the optimal value of μ by iterating the feasibility problem, augmenting/reducing $\mu \geq 0$ if the largest controlled-invariant set $\mathcal{P}_{max}$ in the strip

$$\mathcal{S}(\mu) = \{x : |C_i x + D_i u| \leq \mu - \|F_i\|_1, \;\; i = 1, 2, \ldots, p\} \tag{10.4}$$

is empty/non-empty as follows.

Procedure: Iterative computation of the ϵ-optimal disturbance rejection level.

1. Set $\mu^+ = \mu_{ini}$ and $\mu^- = 0$. Fix a tolerance $\epsilon > 0$.
2. Compute the largest controlled invariant set $\mathcal{P}$ inside the set (10.4) with $\mu := (\mu^+ - \mu^-)/2$
3. If the answer to Problem 10.13 is "YES," then set $\mu^+ := \mu$.
4. If the answer to Problem 10.13 is "NO," then set $\mu^- := \mu$.
5. If $\mu^+ - \mu^- \leq \epsilon$, then STOP, otherwise go to Step 2;

The ideal procedure, i.e. for $\epsilon = 0$, converges to the optimal performance level

$$\hat{\mu} \doteq \inf\{\mu : \;\; \text{the system achieves the performance level } \mu\}$$

In practice, the unknown value of $\hat{\mu}$ is included in the shrinking interval

$$\mu^- \leq \hat{\mu} \leq \mu^+.$$

However the usual problem of set-theoretic computation is waiting for us, since the complexity of the representation of the sets $\mathcal{P}$ may render the problem intractable. Then we must impose a maximum number of constraints and stop the procedure if such a number is reached.

Going back to the introduced assumptions i) and ii), it has to be said that they have been introduced basically to enforce stability. Stability can be also enforced by computing the largest λ-contractive set inside the set (10.4) as proposed in the previous chapters. If we take a $\lambda < 1$, sufficiently close to 1, then the problem is not basically affected and assumption i) can be removed. The interested reader is referred to [Bla94, BMS96], and [FG97].

To remove assumption ii), one can introduce the following one.

iii) The system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

has no transmission zeros.

When $D = 0$ this means that there are no (A, B) invariant subspaces included in the kernel of C [BM92]. The following implication is an immediate consequence of Proposition 5.15.

Proposition 10.16. *If the system (A, B, C, D) has no transmission zeros, then the largest controlled invariant set (or λ-contractive set) inside $\mathcal{S}(\mu)$ is bounded.*

Therefore, if the "no-zeros" property holds, no additional fictitious constraints are needed. Note that the procedure can be extended without essential modification to the case of uncertain systems. To consider continuous-time systems, one can consider again the Euler Auxiliary System

$$\begin{aligned} x(t+1) &= [I + \tau A]x(t) + \tau B u(t) + \tau E d(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

To solve Problem 10.13 one has just to apply the procedure to the EAS by assuming a sufficiently small τ. The following proposition supports this idea.

Proposition 10.17. *Let $\hat{\mu}_{CT}$ be the continuous-time optimal disturbance rejection level as in Definition 10.11. Denote by $\hat{\mu}_\tau$ the optimal disturbance rejection level for the EAS. Then the following properties hold.*

- *If $\tau_1 \leq \tau_2$ then $\hat{\mu}_{\tau_1} \leq \hat{\mu}_{\tau_2}$.*
- *$\hat{\mu}_{CT} \leq \hat{\mu}_\tau$.*
- *If $\hat{\tau} \to 0$, then $\hat{\mu}_\tau \to \mu_{CT}$.*

Proof. The optimal disturbance rejection level is the smallest value of μ for which there exists a controlled invariant set $\mathcal{P}$ inside the set (10.4). Then the first property is based on the fact that if a C-set $\mathcal{P}$ is controlled invariant for the EAS with parameter τ_2, then it is controlled invariant for any $0 < \tau_1 \leq \tau_2$ (see the Appendix). The second property comes from the fact that a controlled-invariant C-set $\mathcal{P}$ for the EAS is controlled invariant for the continuous-time system. The third property is quite tedious to show formally (the proof is quite similar to the convergence proof proposed in [BM96a], to which the reader is referred) and it is basically due to the fact that the solution of the EAS converges to the solution of the continuous-time system as $\tau \to 0$. We skip the details here.

A remarkable difference between the continuous and the discrete-time case is the following. In general, in the discrete-time case, the full information control may assure better performances than the state feedback. Consider the system

$$\begin{aligned} x(t+1) &= x(t) + u(t) + d(t) \\ y(t) &= x(t) \end{aligned}$$

The full information control $u = -[d + x]$ assures the performance level $\mu = 0$. However no controller of the form $u_0 = \Phi(x)$ can provide a performance better than $\mu = 1$. Conversely, in the continuous-time case, the two strategies are equivalent as far as disturbance rejection is concerned, as per the following result.

Theorem 10.18. *Denote by μ_{CT} and by μ_{CTF} the optimal performance achievable in the continuous-time case by means of state and of full information stabilizing control, respectively. Then*

$$\hat{\mu}_{CT} = \hat{\mu}_{CTF}.$$

Proof. A formal proof is quite tedious and it is just sketched here. Consider the system with a full information control $u = \Phi(x, d)$ and the 0 reachability set $\mathcal{R}$ of the closed-loop-system. If a performance level $\hat{\mu}_{CTF}$ is achieved, then

$$\mathcal{R} \subset int\mathcal{S}(\hat{\mu}_{CTF} + \epsilon),$$

with $\epsilon > 0$ arbitrarily small. Then we can:

1. consider the convex hull of $\mathcal{R}$ which is controlled-invariant;
2. approximate such a set by means of a controlled invariant polytope;
3. approximate the invariant polytope, in view of the results in [BM99c], by means of a controlled invariant set which is the unit ball $\mathcal{N}[\Psi, 1]$ of a smooth gauge function $\Psi(x)$ of the type described in Subsection 4.5.3.

According to Nagumo's conditions, there exists a control $u = \Phi_1(x, d)$ which assures the condition

$$\nabla\Psi(x)[Ax + B\Phi_1(x, d) + Ed] \leq 0,$$

for all x on the boundary (i.e., such that $\Psi(x) = 1$) and then the arguments in [Me79, Bla00] can be applied to claim that there exists a control $u = \Phi_2(x)$ (possibly non-continuous in $x = 0$) for which

$$\nabla\Psi(x)[Ax + B\Phi_2(x) + Ed] \leq 0.$$

This means that the set $\mathcal{N}[\Psi, 1]$ is controlled invariant under state feedback. Since this set can fit arbitrarily well the convex hull of the original set $\mathcal{R}$ then $\epsilon > 0$ can be arbitrarily small.

Note that the price that might have to be paid to consider pure state feedback control instead of a full information one is that continuity might have to be dropped [Bla00]. Consider the simple scalar example

$$\dot{x}(t) = u(t) + d(t)$$

with output $y = x$ and $|d| \leq 1$. The full information control $u = -\alpha x - d$ assures a performance level $\hat{\mu}_{CTF} = 0$. No continuous state-feedback control can assure the same performance level. Clearly the discontinuous control

$$u = -2sgn(x)$$

assures the performance level $\hat{\mu}_{SF} = 0$, since any trajectory converges to 0, in finite time and remains null.

We conclude the section by noticing that the proposed approach can be used to solve a more general category of problems than the considered disturbance rejection problem. More precisely, a large class of pursuit-evasion games can be faced within this framework. The general idea was given in [BR71b, GS71] and [Ber72]. After being dormant for several years, the approach has been reconsidered later for the already mentioned reasons of the impressing computer facilities [RM05, CDS02].

Amongst the specific contributions concerning the application of set-theoretic methods to l_1 optimal control problems, we have to mention, [Sha96b, FG97, BS95, SSMAR99] and [BMS96]. Solutions based on ellipsoidal sets have been proposed in [ANP96, NJ99, NJ00]. Extensions to nonlinear systems have been proposed in [Lu98].

10.3 Constrained receding-horizon control

The receding-horizon approach (often referred to as Model Predictive Control (MPC)) is known to be an efficient way to solve constrained (sub-)optimal control problems. If the constrained stabilization is not the only issue, but system performances are also of interest, the invariance approach is not sufficient. However the combined receding-horizon and invariance theory is quite powerful, as it will be shown next. In this section we present only some basic properties of the Receding Horizon Control (RHC). The reader is referred to specific literature [GPM89, SD90, May01, MRRS00, KH05, Ala06] for more details.

10.3.1 Receding-horizon: the main idea

Consider the following infinite-time optimal control problem

$$\min \quad \sum_{t=0}^{\infty} g(x(t), u(t)) \tag{10.5}$$

s.t.

$$x(t+1) = f(x(t), u(t)) \tag{10.6}$$

$$u(t) \in \mathcal{U},\ x(t) \in \mathcal{X},\ t \geq 0, \tag{10.7}$$

and assume, for brevity, that the following holds.

Assumption: Function $g(x, u)$ is convex and positive definite, thus $g(x, u)$ is radially unbounded, and $\mathcal{U}$ and $\mathcal{X}$ are C-sets.

The problem, as it is formulated, is hard to solve even in the linear case. Only simple problems can be faced and, quite often, by means of brute-force computation. It is known that, according to the Hamilton Jacoby Bellman (HJB) theory, one can solve the problem by computing the cost-to-go function (that will be introduced soon). However, it is well known that this is unrealistic for most instances. The reader is referred to the book [Ber00] for more details.

The receding-horizon control is based on the optimization on-line of a finite portion of the infinite trajectory, defined on a finite horizon of length $T > 0$, which is repeated at each step of the process. This is an open-loop optimization that, being so iterated, actually provides a feedback control[2]. We state the basic sequence of an RHC computation on-line [Pro63].

RHC algorithm (on-line) At each integer time t:

1. measure (or estimate) the state $x(t)$;
2. find sequences $u(i) \in \mathcal{U}$, $x(i) \in \mathcal{X}$, $i = t, t+1, \ldots, t+T-1$, which satisfy the system equation and optimize a finite-horizon cost

$$J = \sum_{i=k}^{k+T-1} g(x(i), u(i)); \tag{10.8}$$

3. apply $u(t)$ to the system;
4. set $t = t+1$ and go to step 1.

It is intuitive that, for T large enough, this scheme provides a trajectory which reasonably close to the optimal. However, being T finite, the following issues arise:

- can recursive feasibility be assured?
- can stability be assured?
- can some sort of (sub-)optimality be assured?

The first issue boils down to the following question: if the problem is feasible at step t, will it be feasible also at step $t+1$? It is obvious that feasibility of the latter depends on how the control at the previous step has been chosen. The issue of sub-optimality is the following. At the end, what we want to do, is to minimize the infinite-horizon cost. Then a natural question is how much performance we lose in optimizing on a finite horizon only.

Let us consider the feasibility of the problem. We first show by a simple example that feasibility at time t does not imply feasibility in the future.

[2] So it has been named OLOF (= Open Loop Optimal Feedback) control, by the author of [Gut86], whose name, by chance, is Per-Olof.

Example 10.19. Consider the simple scalar problem of a linear system with a quadratic cost

$$x(t+1) = 2x(t) + u(t) \tag{10.9}$$

$$g(x,u) = x^2 + \beta u^2 \tag{10.10}$$

$$|u(t)| \leq 3 \tag{10.11}$$

$$|x(t)| \leq 100 \tag{10.12}$$

For this system the largest domain of attraction is the open set $-3 < x < 3$. If $x(t) = 2$ and we minimize over a horizon of length $T = 2$ (this example can be extended to a longer horizon) we have to minimize

$$\begin{aligned} J_2 &= x(t)^2 + \beta u(t)^2 + x(t+1)^2 + \beta u(t+1)^2 \\ &= x(t)^2 + \beta u(t)^2 + (2x(t) + u(t))^2 + \beta u(t+1)^2 \end{aligned}$$

The minimizer of this expression is the couple $u(t+1) = 0$ and

$$u(t) = -2x(t)/(\beta + 1)$$

(provided that $|2x/(\beta+1)| \leq 3$), which is the control at the current step. Consider the state $x(0) = 2$. If we apply the "optimal" strategy, then the new state will be

$$x(1) = [2 - 2/(\beta+1)]x(0) = [2 - 2/(\beta+1)]2$$

If β is large enough, the control action is weak and the next state $x(1) \geq 3$, so the state, although it remains admissible, leaves the largest domain of attraction. Then the problem will remain admissible for a certain number of steps, but the constraints will be eventually violated and the problem will become eventually unfeasible. Note that for $x(0) = 2$, the admissible control sequence $-3,\ -2,\ 0,\ 0, \ldots$ drives the system to 0, since $x = 2$ is a recoverable state. This is a typical case in which recursive feasibility is not assured. This very simple example shows that, if the constraints on x are removed then we can fix the recursive feasibility of the problem, but we cannot assure stabilizability. Finally, it can be shown that optimality can be arbitrarily violated in the sense that the global cost is arbitrarily high, even for values of β for which the considered two-steps receding horizon scheme stabilizes the system. This fact is related to the concept expressed in Subsection 2.4.7. Precisely, limiting the optimization to the first steps can be seen as a greedy strategy, which can have undesirable results.

In the unconstrained control case, recursive feasibility is not an issue and the question of stability can be solved in a simple way. As it will be apparent, the price to be paid, since we must artificially transform the cost, is once again a compromise. Consider the unconstrained minimization problem

$$\min \quad \sum_{k=t}^{t+T-1} g(x(k), u(k)) + \alpha h(x(t+T)) \tag{10.13}$$

$$s.t.$$

$$x(k+1) = f(x(k), u(k)) \tag{10.14}$$

which is identical to the previous, with the exception that the cost is modified by the addition of the function h parameterized by $\alpha > 0$. If h is a convex and positive definite function (and radially unbounded function) it is quite intuitive that, for T large enough and α large enough, the modified cost has a stabilizing effect [May01, MRRS00]. We will propose some examples of final cost later.

10.3.2 Recursive feasibility and stability

Before starting with the description of the technique we introduce the cost-to-go function as

$$J_{opt}(x) \doteq \text{optimal value of (10.5)–(10.7) with initial condition } x(0) = x$$

Assume that

$$f(0,0) = 0$$

If the problem with initial condition is unfeasible, then the cost is set to

$$J_{opt}(x) = +\infty$$

Let us define the finite-horizon optimal cost function as follows

$$J_{opt}(x,T) \doteq \text{optimal value of (10.13)–(10.14) with initial condition } x(0) = x$$

with the understanding that, again, $J_{opt}(x,T) = +\infty$ if the problem is unfeasible.

Theorem 10.20. *Assume that $g(x,u)$ is positive definite with respect to x. The following basic properties hold.*

i) If $g(x,u)$ is positive definite and radially unbounded, then both $J_{opt}(x,T)$ and $J_{opt}(x)$ are positive definite and radially unbounded.

ii) If $g(x,u)$ and the constraints are convex, then both $J_{opt}(x,T)$ and $J_{opt}(x)$ are positive definite and convex.[3]

iii) $J_{opt}(x)$ is a local control Lyapunov function.

Proof. Property i) is straightforward since $J_{opt}(x) \geq J_{opt}(x,T) \geq J_{opt}(x,1) = \min_{u(t)\in\mathcal{U}} g(x(t),u(t))$. To show ii), consider the case of $J_{opt}(x,T)$ and take two initial states $x_1(k)$ and $x_2(k)$ and the optimal control (constraint-admissible) sequences from these initial states: $u_1(k)$, $u_1(k+1)$, $\ldots u_1(k+T-1)$ and $u_2(k)$, $u_2(k+1)$, $\ldots$, $u_2(k+T-1)$. Then, for any initial state $x(k) = \alpha x_1(k) + (1-\alpha)x_2(k)$, $0 \leq \alpha \leq 1$, consider the control sequence $u(i) = \alpha u_1(i) + (1-\alpha)u_2(i)$. By linearity the state sequence will be $x(i) = \alpha x_1(i) + (1-\alpha)x_2(i)$, where $x_1(i)$ and $x_2(i)$ are the optimal (constraint-admissible) trajectories associated with $x_1(k)$ and $x_2(k)$. Since the constraints are convex, $\alpha x_1(i) + (1-\alpha)x_2(i)$ and $\alpha u_1(i) + (1-\alpha)u_2(i)$ are admissible. Since the cost is the sum of convex terms, then

$$\begin{aligned}
&\sum_{i=k}^{k+T-1} g(x(i),u(i)) \leq \\
&\leq \alpha \left[\sum_{i=k}^{k+T-1} g(x_1(i),u_1(i))\right] + (1-\alpha)\left[\sum_{i=k}^{k+T-1} g(x_2(i),u_2(i))\right] \leq \\
&\leq \alpha J_{opt}(x_1,T) + (1-\alpha)J_{opt}(x_2,T),
\end{aligned}$$

thus $J_{opt}(x,T)$ is convex. The same argument is valid for $J_{opt}(x)$

Property iii) is a standard property of the cost-to-go function and the proof can be carried out by standard arguments [May01, Ber00].

We consider now the problem of receding horizon control for linear systems in the presence of a convex cost and convex constraints:

$$\min J = \sum_{k=t}^{t+T-1} g(x(k),u(k)) \tag{10.15}$$

s.t.

$$x(k+1) = Ax(k) + Bu(k) \tag{10.16}$$

$$u(k) \in \mathcal{U},\ \ x(k) \in \mathcal{X},\ t \geq 0, \tag{10.17}$$

Recursive feasibility and stability can be assured by means of an additional constraint associated with a contractive set. Consider a λ-contractive set $\mathcal{S} \subseteq \mathcal{X}$ (i.e. which is admissible with respect to the constraints, in the sense that it can be associated with a control $\Phi(x) \in \mathcal{U}$, for all $x \in \mathcal{S}$). Denoting by $\Psi(x)$ the Minkowski

[3] Actually, in Section 3, we did not consider functions which may assume values $+\infty$ as J_{opt} could do. The reader is referred to [Roc70] for an extended definition.

function associated with $\mathcal{S}$, we require the contraction of the first state of the planning horizon with respect to this set. More precisely, the following additional initial constraint is added to the scheme

$$\Psi(x(t+1)) \leq \lambda \Psi(x(t)) \tag{10.18}$$

The following proposition holds.

Proposition 10.21. *The receding horizon scheme (10.15)–(10.17) with the additional constraint (10.18) is recursively feasible and locally stable with domain of attraction $\mathcal{S}$.*

Proof. Assume that $x(t) \in \mathcal{S}$. Then there exists a sequence of control inputs $u(t+i) \in \mathcal{U}, i = 0, 1, 2, \ldots$ for which $x(t+h) \in \lambda^h \mathcal{S} \subseteq \mathcal{S}$, thus there exists a feasible solution to the optimization problem. Stability of the scheme is assured since the first control of the computed sequence is actually applied and therefore condition (10.18) holds at each step.

A different approach to solve the problem is to include the final state of the planning horizon in a contractive set. Given a λ-contractive $\mathcal{S} \subseteq \mathcal{X}$ which is admissible with respect to the constraints, in the sense that it can be associated with a control $\Phi(x) \in \mathcal{U}$, for all $x \in \mathcal{S}$, the following additional final constraint is added to the scheme

$$x(t+T) \in \mathcal{S}. \tag{10.19}$$

Proposition 10.22. *The receding horizon scheme (10.15)–(10.17) with the additional constraint (10.19), with $\mathcal{S}$ controlled-invariant, is recursively feasible.*

Proof. Assume that at time t the condition $x(t+T) \in \mathcal{S}$ can be assured, precisely that there exists a sequence of T inputs which drives the state inside $\mathcal{S}$ in T steps. Then, by construction, at time $t+1$ there exists a sequence of of $T-1$ inputs which drives the state inside $\mathcal{S}$ is in $T-1$ steps. But since $\mathcal{S}$ is invariant there exists a further control which keeps the state inside $\mathcal{S}$, so the sequence can be extended to the original length T. Then the problem is recursively feasible.

Stability can be assured by reducing the time horizon of one unit at each step, so that the set $\mathcal{S}$ is eventually reached. Once the state is in $\mathcal{S}$, the typical procedure is to switch to a local controller. Another technique with a similar property is that of scaling, at each step, the set $\mathcal{S}$ by reducing its size. This can be done by storing the value $\mu_{prec} = \Psi(x(t+T))$ computed at step t and, at the next step $t+1$, based on the actual state $x(t+1)$, imposing the constraint

$$\Psi(x(t+T+1)) \leq \lambda \mu_{prec}.$$

This recursive scheme assures both recursive feasibility and asymptotic stability. A further possibility is that of modifying the cost by adding on top of the final constraint the additional final state cost

$$\alpha h(x(T)) = \alpha \Psi(x(T))^p,$$

where $\Psi(x)$ is the Minkowski function of the computed contractive set. When the final state reaches a neighborhood of the origin $\mathcal{N}[\Psi, \mu]$ where the constraints are not active, then we know that at the next step we can guarantee $x(T+1) \in \lambda \mathcal{N}[\Psi, \mu]$, equivalently $\Psi(x(T+1)) \leq \lambda \Psi(x(T))$. Assuming g positively homogeneous, for α large enough and p equal to the degree of g, the optimizer will be forced to reduce the cost by reducing $\Psi(x(T))$.

There is an interesting method to choose the final invariant set. Let us consider the unconstrained cost-to-go function, precisely

$$\psi(x) = \inf \sum_{t=0}^{\infty} g(x(t), u(t)) \tag{10.20}$$

s.t.

$$x(t+1) = f(x(t), u(t)) \tag{10.21}$$

$$x(0) = x \tag{10.22}$$

Function $\psi(x)$ is the optimal cost with initial condition x. Let us assume that $\psi(x)$ is continuous and positive definite. It is clear that, if $\mathcal{X}$ and $\mathcal{U}$ are C-sets, in a region close to the origin the constraints have no role and thus this function turns out to be the effective optimal cost for our original problem. In particular, consider the set of points of the form $\mathcal{N}[\psi, \mu]$, where μ is small enough to assure both conditions

$$\mathcal{N}[\psi, \mu] \subseteq \mathcal{X}$$

and

$$\Phi(x) \in \mathcal{U},$$

where $\Phi(x)$ is the unconstrained optimal control. Inside $\mathcal{N}[\psi, \mu]$, the optimal unconstrained trajectory and the optimal constrained one are the same. In fact the following proposition (reported without the obvious proof) holds.

Proposition 10.23. *The set $\mathcal{N}[\psi, \mu]$ is controlled-invariant.*

Therefore the infinite-horizon optimal control can be solved as follows [SD87].

$$J_{opt}(x) \doteq \min \sum_{k=0}^{T-1} g(x(k), u(k)) + \psi(x(T)) \tag{10.23}$$

$$s.t.$$

$$x(k+1) = f(x(k), u(k)) \tag{10.24}$$

$$x(0) = x \tag{10.25}$$

$$u(k) \in \mathcal{U}, \;\; x(k) \in \mathcal{X} \tag{10.26}$$

$$x(T) \in \mathcal{N}[\psi, \mu] \tag{10.27}$$

The following theorem holds.

Theorem 10.24. *For a given initial condition we have the following*

i) The optimal value J_{opt} is non-increasing with T.
ii) There exists $\hat{T}$ such that, for $T \geq \hat{T}$ the optimal cost remains unchanged.
iii) For any $T \geq \hat{T}$, the control achieved by solving (10.23)–(10.27) with the receding-horizon strategy is globally optimal.
iv) For any $T \geq \hat{T}$, the optimal cost of the optimization problem achieved by solving on-line (10.23)–(10.27) is equal to the infinite-time optimal cost with the current state as initial condition.

Proof. See [SD87, May01]

The importance of the previous theorem lies in the fact that we can compute the optimal constrained strategy in a receding horizon fashion if we are able to solve, locally, an unconstrained problem to determine $\psi(x)$ and if the horizon is large enough.

A typical case in which this scheme can be easily implemented is the constrained linear-quadratic optimal control with

$$J_{opt}(x) \doteq \min \sum_{k=0}^{\infty} x(k)^T Q x(k) + u(k)^T R u(k) \tag{10.28}$$

$$s.t.$$

$$x(k+1) = Ax(k) + Bu(k) \tag{10.29}$$

$$x(0) = x \tag{10.30}$$

$$u(k) \in \mathcal{U}, \;\; x(k) \in \mathcal{X}, \tag{10.31}$$

$$x(t+T)^T \; P \; x(t+T) \leq \mu \tag{10.32}$$

where $P \succ 0$ is any positive definite matrix associated with a controlled-invariant ellipsoid. An interesting method to choose the matrix P was proposed in [SD87].

For this problem, the unconstrained optimal cost function is given by

$$J_{opt}(x) = x^T P x$$

where P is the solution of the Riccati equation

$$P = Q + A^T PA - A^T PB \left[R + B^T PB\right]^{-1} B^T PA,$$

and the corresponding optimal control is given by

$$u = Kx = \left[R + B^T PB\right]^{-1} B^T PAx$$

Therefore one has to identify an ellipsoid $\mathcal{E}[P, \mu]$ such that

$$\mathcal{E}[P, \mu] \subseteq \mathcal{X}, \text{ and } K\mathcal{E}[P, \mu] \subseteq \mathcal{U}$$

by choosing a suitably small μ.

The method of imposing the arrival to a controlled-invariant set does not provide a domain of attraction to the origin as in the technique previously shown of forcing the first state inside a contractive set. A domain of attraction can be computed indirectly since any state for which the finite-horizon optimization problem has a feasible solution (including the arrival to the controlled invariant set) is included in the domain of attraction. However, in the linear problem with convex constraints and cost, we can under-estimate the domain of attraction by means of the following property.

Proposition 10.25. *Consider the receding horizon problem (10.23)–(10.27) and further assume that (10.24) is linear with $\mathcal{X}$ and $\mathcal{U}$ convex. Consider a polytope $\mathcal{V}[X]$. If for any vertex (i.e. any column of matrix X) the finite-horizon control problem with arrival to a controlled-invariant C-set $\mathcal{P}$ has a feasible solution, then each state in the polytope is driven to $\mathcal{P}$ (and then to 0) by the receding horizon strategy and with a cost which is upper bounded by the largest of the costs associated with the vertices.*

Proof. The proof follows by simple convexity arguments and is left as an exercise.

When $\mathcal{X}$ and $\mathcal{U}$ are polytopes, the receding horizon problem can be solved by linear-quadratic optimization in an efficient way. Indeed, if the following constraints are assumed

$$\mathcal{X} = \{x : Fx \leq g\} \text{ and } \mathcal{U} = \{x : Hx \leq k\},$$

the associated receding horizon problem has the following aspect

$$\begin{aligned}
x_1 &= Ax_0 \quad + Bu_0 \\
x_2 &= A^2x_0 + ABu_0 \quad + Bu_1 \\
x_3 &= A^3x_0 + A^2Bu_0 \quad + ABu_1 \quad + Bu_2 \\
&\vdots \\
x_T &= A^Tx_0 + A^{T-1}Bu_0 + A^{T-2}Bu_1 + \ldots + Bu_{T-1} \\
Fx_i &\leq g \\
Hu_i &\leq k
\end{aligned} \tag{10.33}$$

In the quadratic cost case, minimizing (10.28) subject to these constraints is a classical problem for which specific software exists.

It is worth mentioning that the constrained linear-quadratic problem can be solved in a different way in view of the result in [BMDP02]. Indeed, in the case of linear constraints (i.e., polytopic $\mathcal{X}$ and $\mathcal{U}$), the cost-to-go function is piecewise quadratic and the optimal compensator is piecewise-linear, therefore their evaluation is computationally feasible. A further interesting approach to face constrained optimal control problems is that based on the so-called multiparametric programming (see, for instance, [TJB03] and [BBBM05]).

10.3.3 Receding horizon control in the presence of disturbances

We consider now the case in which the system is affected by disturbances, and precisely the case of systems of the form

$$x(t+1) = Ax(t) + Bu(t) + Ed(t),$$

where $d \in \mathcal{D}$ a C-set. We still assume constraint of the form $x \in \mathcal{X}$ and $u \in \mathcal{U}$. For brevity we assume that both $\mathcal{X}$ and $\mathcal{D}$ are polytopes. In this case it is possible to take into account the effect of the disturbance in the trajectory computation without basically increasing the complexity of the problem. Consider the linear constraint

$$F_i x(t+k) \leq g_i, \quad \text{i.e. } x(t+k) \in \mathcal{P}[F, g]$$

$k = 1, \ldots, T$. Assume $x(t)$ is known. From the system equation we get for each k

$$F\left[A^k x(t) + \sum_{h=0}^{k-1} A^{k-h-1}Bu(h+t) + \sum_{h=0}^{k-1} A^{k-h-1}Ed(h+t)\right] \leq g \tag{10.34}$$

Constraints (10.34) have to be satisfied for all possible $d \in \mathcal{D}$. The last term in the square brackets belongs to the k-step reachability set with the disturbance input, and then (10.34) is equivalent to

$$F\left[A^k x(t) + \sum_{h=0}^{k-1} A^{k-h-1} B u(h+t) + z(t+k)\right] \leq g,$$

where

$$z(t+k) \in \mathcal{R}_k,$$

and $\mathcal{R}_k$ is the k-step reachability set of the system (A, E) with constrained input $d \in \mathcal{D}$ (see Subsection 6.2.1). Consider a vector $r^{(k)}$, whose components are defined as

$$r_i^{(k)} \doteq \max_{z \in \mathcal{R}_k} F_i z = \phi_{\mathcal{R}_k}(F_i).$$

Then the set of linear constraints (10.34) becomes equivalent to

$$F\left[A^k x(t) + \sum_{h=0}^{k-1} A^{k-h-1} B u(h+t)\right] \leq g - r^{(k)}, \quad k = 1, 2, \ldots, T. \tag{10.35}$$

This trick was proposed in [Bla90a]. Note that the restricted constraints have the same form of the original ones and that only the known term g is replaced by the more stringent $g - r^{(k)}$. This allows to discharge the disturbance in the trajectory planning. The idea is pictorially represented in Fig. 10.3. Since disturbances are present, asymptotic stability cannot be achieved. To assure ultimate boundedness to the whole scheme one can consider, along with the regular constraints (10.35), the

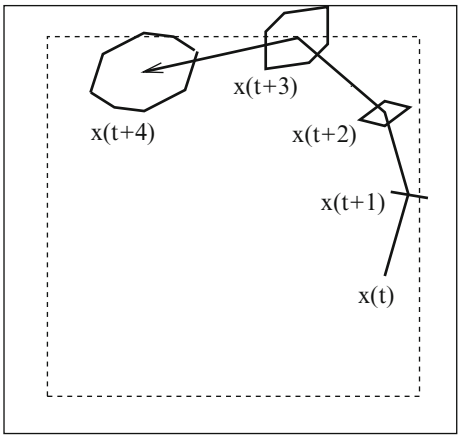


Fig. 10.3 The reduced constraints

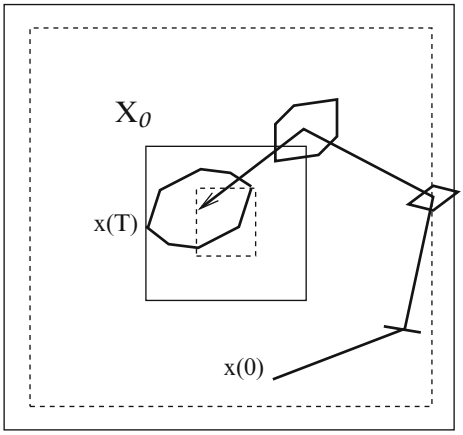


Fig. 10.4 The ultimate constraint

additional final constraint

$$x(t+T) \in \mathcal{X}_0 = \{x : \ F_0 x \leq g_0\}$$

where $\mathcal{X}_0$ is a compact set compatible with the constraints. Such a final constraint is equivalent to (see Fig. 10.4)

$$F_0 \left[A^T x(0) + \sum_{h=0}^{T-1} A^{T-h-1} B u(h) \right] \leq g_0 - r_0^{(T)}$$

where $r_0^{(T)}$ is the analogous of $r^{(T)}$ computed for $\mathcal{X}_0$. The following proposition holds [Bla90a].

Proposition 10.26. *If the set $\mathcal{X}_0$ is such that, for each vertex x_i of $\mathcal{X}_0$ assumed as initial condition, the problem of finding a sequence which drives the state again to $\mathcal{X}_0$ in T steps has a feasible solution, then the condition*

$$x(kT) \in \mathcal{X}_0, \quad k = 1, 2, \ldots$$

can be recursively guaranteed for all *the initial conditions inside $\mathcal{X}_0$. Therefore it is possible to assure ultimate boundedness, for all the initial states for which the problem with the terminal constraint $x(T) \in \mathcal{X}_0$ is feasible.*

The property which assures that the state can be periodically driven inside $\mathcal{X}_0$ is named U–D-invariance in T steps in [Bla90a]. Ultimate boundedness inside $\mathcal{X}_0$ can be assured if $\mathcal{X}_0$ is controlled invariant.

The main troubles arising with this approach are due to the fact that if the system is not asymptotically stable then the reachable sets $\mathcal{R}_k$ become larger and larger as the planning horizon increases. Therefore an improvement of the idea can be achieved by pre-stabilizing the system as suggested in [MSR05]. The system is pre-

stabilized by a certain feedback control (typically an optimal gain) so that the actual control action is $u(t) = Kx(t) + v(t)$. Clearly this means that the residual control authority is reduced and precisely

$$Kx(t) + v(t) \in \mathcal{U}$$

Setting $\hat{A} = A + BK$, the optimization problem takes the form

$$\begin{aligned}
x_1 &= \hat{A}x_0 \quad + Bv_0 \\
x_2 &= \hat{A}^2 x_0 + \hat{A}Bv_0 \quad + Bv_1 \\
x_3 &= \hat{A}^3 x_0 + \hat{A}^2 Bv_0 \quad + \hat{A}Bv_1 \quad + Bv_2 \\
&\vdots \\
x_T &= \hat{A}^T x_0 + \hat{A}^{T-1} Bv_0 + \hat{A}^{T-2} Bv_1 + \ldots + Bv_{T-1}
\end{aligned}$$

$$\begin{aligned}
Fx_i &\leq g - r_i \quad i = 1, 2, \ldots, T-1 \\
F_0 x_T &\leq g_0 - r_T \\
HKx_i + Kv_i &\leq k
\end{aligned}$$

Note that now joint state-control constraints are present. These constraints have been already considered in Section 5.3. In view of those results we can compute the reduced constraints as shown above and then compute a contractive set compatible with the above constraints.

The advantage of this approach is that since the system is stable, the reachable sets are bounded. In practice the new shifting values r_i are much smaller than those computed for the uncontrolled system (see Subsection 6.2.1). The new picture is that of Fig. 10.5. Since the propagation of the uncertainty effect, due to the prestabilization, is strongly reduced, a more accurate prediction is possible. Therefore a much smaller set can be assumed as final target. This fact can be seen

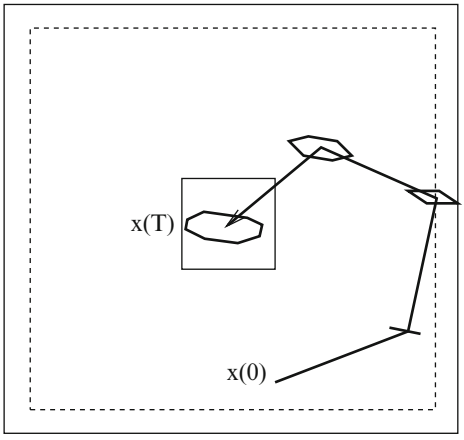


Fig. 10.5 The ultimate constraint with prestabilization

by means of a very simple example. Consider again system (5.21) in Section 5.3 now with an additional disturbance term

$$x(t+1) = 2x(t) + u(t) + d(t), \tag{10.36}$$

where $|u| \leq 4$ and $|d| \leq 1$. If we consider the prestabilization (already introduced in Section 5.3)

$$u = -2x(t) + v(t)$$

the resulting system is

$$x(t+1) = v(t) + d(t).$$

For this system the problem of finding an open-loop v-sequence $v(0)$, $v(1)$, $v(2)$ which keeps the state inside the interval $[-2, 2]$ can be solved on any arbitrarily large horizon. Indeed for $-2 \leq x(0) \leq 2$, one can just take $v(t) = 0$ for $t = 0, 1, 2, \ldots, T$. Simple computations prove that no constraint violation occurs and that, in fact, the condition $-1 \leq x(t) \leq 1$ is assured for $t \geq 1$. It is immediate to see that no decision can be made over a T-horizon for the original control input u. Indeed, no open-loop sequence $u(0), u(1), \ldots, u(T)$ for $T > 2$ will produce feasibility as long as the constraint $-2 \leq x(t) \leq 2$ must be satisfied, since there is no way to fit the disturbance-reachability set in this interval.

Note that this example does not mean that the technique of considering the prestabilization produces any miracle, since the new system is equivalent to the original one as long as the constraints remain. It just means that, without pre-stabilization, the prediction fades as the horizon increases and the trick of taking into account the disturbance effect during prediction by means of the reachability sets cannot be applied. For more detailed discussions about robust model-predictive control, in particular on the so-called tube-based model-predictive control technique, the reader is referred to [MSR05, MRFA06, CBKR11, RKFC12, RKC+12, RKCP12].

Using set-theoretic methods in receding-horizon control is a standard approach and only a brief overview is given here. After the first edition of the present book, a lot of contributions have been published in the literature. Examples of recent developments are [GFA+11, SOH11, CKRC11, LAAC08, JM10, RFFT14]. A detailed survey is out of the scope of this book and the reader is referred to specific literature [AZ00, KH05, GSDD06, Ala06, BM99a].

10.4 Relatively optimal control

In this section, a recent idea to solve optimal control problems in the presence of constraints is discussed. The motivating idea is that finding a control which is optimal for *any* initial condition might be computationally hard. However, achieving

Fig. 10.6 Systems with dominating tasks

optimality for *any* initial condition is not really important in many cases of systems (Fig. 10.6) whose main goal is to perform specific tasks and which thus work with special initial conditions (such as lifts, bascule bridges, automatic gates, floodgates, cranes, ...). Therefore in such cases optimality is important only when starting from specific nominal initial states. The Relatively Optimal control (ROC) problem is that of finding a feedback control (i.e., without feedforward actions) such that

- it is optimal for the nominal initial condition;
- it is stabilizing for all the other initial states.

Consider the discrete-time reachable system

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{10.37}$$

and the following convex functions of the output

$$g(y), \quad l_i(y), \ \ i = 1, 2, \ldots, s,$$

together with the constraint

$$y(t) \in \mathcal{Y}, \tag{10.38}$$

where $\mathcal{Y}$ is a convex and closed set. Consider then the next optimization problem with assigned initial condition $x(1) = \bar{x} \neq 0$ (the initial time now is set to $k = 1$ for notation coherence)

$$J_{opt}(\bar{x}) = \min \sum_{k=1}^{N} g(y(k)) \quad \text{s.t.} \tag{10.39}$$

$$x(k+1) = Ax(k) + Bu(k), \ \ k = 1, \ldots, N \tag{10.40}$$

$$y(k) = Cx(k) + Du(k), \ \ k = 1, \ldots, N \tag{10.41}$$

$$\sum_{k=1}^{N} l_i(y(k)) \leq \mu_i, \ \ i = 1, 2, \ldots, s \tag{10.42}$$

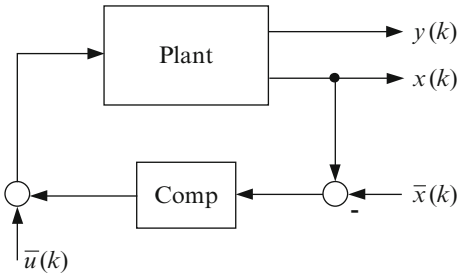


Fig. 10.7 Feedforward + static feedback (trajectory-based) scheme

$$y(k) \in \mathcal{Y}, \quad k = 1, 2, \ldots, N \tag{10.43}$$

$$x(1) = \bar{x} \tag{10.44}$$

$$x(N+1) = 0 \tag{10.45}$$

$$N \geq 0, \quad \text{assigned (or free).} \tag{10.46}$$

In principle this problem is very easy to solve by means of a feedback-and-feedforward control as in Fig. 10.7. Let $\bar{x}(k)$ and $\bar{u}(k)$, $k = 1, 2, \ldots, N$, be the state and control optimal trajectories, K be any feedback matrix such that $A+BK$ is stable and consider the static control

$$u(t) = \bar{u}(t) + K(x(t) - \bar{x}(t)). \tag{10.47}$$

It is obvious that, by definition of $\bar{u}(t)$, if the system is initialized with initial condition $x(1) = \bar{x}$, the corresponding trajectory is the optimal one. Unfortunately, this scheme presents well-known disadvantages. Firstly, the resulting controller is time-varying. Secondly, the initial condition $\bar{x}$ may be just the effect of a disturbing event, for instance an impulse, whose precise instant is unknown a priori. Therefore a solution based on feedforward is not viable. Thirdly, the controller basically tracks the optimal trajectory originating in $\bar{x}$ and therefore, if $x(1)$ is far from $\bar{x}$, the resulting transient will be completely unsatisfactory. Fourthly, when magnitude constraints on u and y are present, a natural request is that, whenever $x(1)$ is "reduced" (for instance $x(1) = \bar{x}\alpha$, $0 < \alpha < 1$), the constraints remain satisfied. The scheme (10.47) does not assure this property, as shown in the next example.

Example 10.27. Consider the problem above for the system with

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

Assume $|y_2| \leq 2$ and $|y_3| = |u| \leq 1$, and let the initial condition be $x(1) = \bar{x} = [\ -8\ \ 0\ ]^T$. Take $g(y) = |y_1|$ and $N = 6$. The optimal state and control trajectories,

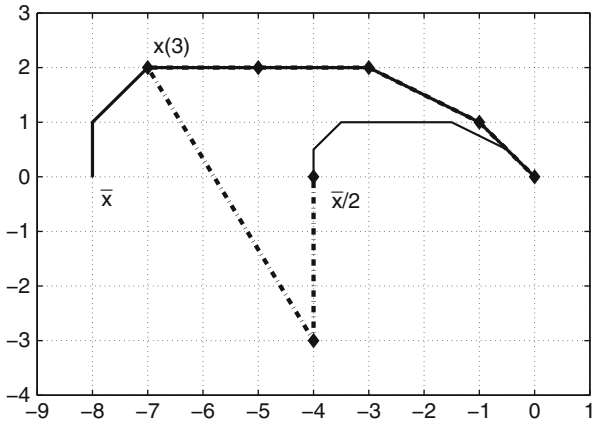


Fig. 10.8 The trajectories for the example

$x(1),\ x(2)\ldots$ and $u(1),\ u(2)\ldots$, respectively, are given by the columns of the following matrices

$$\begin{bmatrix} -8 & -8 & -7 & -5 & -3 & -1 & 0 \\ 0 & 1 & 2 & 2 & 2 & 1 & 0 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 1 & 0 & 0 & -1 & -1 & 0 \end{bmatrix}.$$

The corresponding optimal cost is $J_{opt} = 32$. Consider the control (10.47), with feedback matrix $K = [\ -1\ \ -2\ ]$ that renders $A + BK$ nilpotent. Clearly, for the assigned $x(1)$ this control yields the same optimal trajectory reported in Fig. 10.8 (plain line). Assume now to scale down the initial condition as $x(1) = \bar{x}/2$. The goal of the control scheme (10.47) is that of reaching the previous optimal trajectory. The resulting modified trajectory (dashed-dot curve in Figure 10.8) reaches the original one in two steps at the cost of $J = 24$. However it is apparent that this solution is absurd: the first component $x_1(k)$ is initially enlarged to reach the optimal trajectory. Furthermore, both state and input constraints are violated. Note that the scaled trajectory $\bar{x}(1)/2,\ \bar{x}(2)/2,\ \bar{x}(3)/2,\ \ldots$, corresponding to the plain line in Figure 10.8, not only moves the first component immediately in the right direction, but achieves a cost of $J = 16$ with no constraint violation. It is apparent that if the optimal trajectory for $x(1) = \bar{x}$ is achieved by a linear controller, with zero initial condition, the same controller produces the scaled trajectory when $x(1) = \bar{x}/2$. This second behavior is, of course, highly preferable. We do not even report the solution for $x(1) = -\bar{x}$, since it goes out of range with $x(2) = [\ 8\ \ -15\ ]^T$, with deep constraints violation. It can be shown that changing the stabilizing gain K (i.e., not assuming a dead-beat controller) does not improve the situation up to an acceptable level.

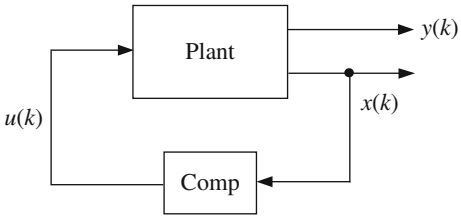


Fig. 10.9 Pure (dynamic) feedback scheme

Therefore, since the feedforward idea doesn't seem to produce reasonably good results, we resort to the feedback scheme in Fig. 10.9. The basic request for this scheme is that the system has to be stable and the transient has to be optimal if the initial condition is $\bar{x}$. To avoid some of the problems evidenced in the last example, we also require that our (possibly dynamic) compensator cannot be initialized based on the knowledge of $\bar{x}$, say the compensator initial condition is zero. We are now in the position of formalizing the problem. Denote by $z(t)$ the state of the compensator.

Problem 10.28 (Relatively Optimal Control Problem). Find a state feedback compensator having the structure in Fig. 10.9 such that

1. for $x(1) = \bar{x}$ and for $z(1) = 0$ the control and state trajectories are the optimal constrained ones
2. it is stabilizing[4] for all initial conditions.

We remark that the constraints are fulfilled only for the nominal condition. Thus they should be considered as *soft constraints* for performance specifications rather than hard ones.

10.4.1 The linear dynamic solution

We show how to obtain a stabilizing (actually dead-beat) state feedback compensator which does not require feedforward or state initialization and is optimal for the nominal initial condition. This technique has been proposed in [BP03]. For brevity, here we work under a technical assumption (in [BP03] it is shown how to remove it).

Assumption. The initial state $\bar{x} \neq 0$ does not belong to any proper (A, B)-invariant subspace[5] of $\mathbb{R}^n$.

[4] Although with possible constraint violations.

[5] A subspace $\mathcal{S}$ is said (A, B)-invariant if for all $x \in \mathcal{S}$ there exists $u(x)$ such that $Ax + Bu(x) \in \mathcal{S}$. It is said proper if $\mathcal{S} \neq \mathbb{R}^n$ [BM92].

Denote by

$$\bar{X} = \begin{bmatrix} \bar{x}(1) & \bar{x}(2) & \dots & \bar{x}(N) \end{bmatrix} \tag{10.48}$$

the $n \times N$ matrix containing the optimal state trajectory and by

$$\bar{U} = \begin{bmatrix} \bar{u}(1) & \bar{u}(2) & \dots & \bar{u}(N) \end{bmatrix} \tag{10.49}$$

the $m \times N$ matrix containing the optimal input sequence. By construction, $\bar{x}(k+1) = A\bar{x}(k) + B\bar{u}(k)$ and $\bar{x}(N+1) = A\bar{x}(N) + B\bar{u}(N) = 0$. This means that the matrices $\bar{X}$ and $\bar{U}$ satisfy the next equation

$$A\bar{X} + B\bar{U} = \bar{X}P, \tag{10.50}$$

where the square matrix P is an N dimensional Jordan block associated with the 0 eigenvalue,

$$P = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \tag{10.51}$$

and is a stable matrix since $P^k = 0$ for all $k \geq N$. In view of the introduced assumption, we can state the next Lemma, whose proof can be found in [BP03].

Lemma 10.29. *Matrix $\bar{X}$ has rank n, namely it has full row rank.*

Note that the technical assumption is sufficient, but not necessary, say there might be cases in which $x(1)$ belongs to an (A, B) invariant subspace but still $\bar{X}$ results in a full row rank matrix[6]. The previous lemma obviously implies that $N \geq n$. For the moment being, let us assume $N > n$ (the simple case $N = n$ will be briefly discussed later). Let us consider any $(N-n) \times N$ matrix $\bar{Z}$ of the form

$$\bar{Z} = \begin{bmatrix} 0 & \bar{z}(2) & \bar{z}(3) & \dots & \bar{z}(N) \end{bmatrix} \tag{10.52}$$

(note that the first column is zero) such that the next square matrix is invertible

$$T = \begin{bmatrix} \bar{X} \\ \bar{Z} \end{bmatrix} = \begin{bmatrix} \bar{x}(1) & \bar{x}(2) & \bar{x}(3) & \dots & \bar{x}(N) \\ 0 & \bar{z}(2) & \bar{z}(3) & \dots & \bar{z}(N) \end{bmatrix}. \tag{10.53}$$

[6] Basically, only via academic examples it is possible to provide non full-row-rank X. For instance by considering a pure LQ control and by taking as initial condition an eigenvector of the optimal closed-loop matrix.

Clearly finding a $\bar{Z}$ in such a way that T is invertible is always possible because, by Lemma 10.29, $\bar{X}$ has full row rank. The fact that we can always take $\bar{Z}$ having the zero vector as first column is also readily seen. Since $x(1) = \bar{x}$ is non-zero, there is at least one non-zero component in the first column of $\bar{X}$. By means of the row of $\bar{X}$ corresponding to this entry we can apply Gaussian elimination to $\bar{Z}$ in order to render null each entry of the first column of $\bar{Z}$. Matrix T remains invertible after this operation. Denote by

$$\bar{V} \doteq \bar{Z}P \tag{10.54}$$

and consider the linear compensator

$$z(k+1) = Fz(k) + Gx(k) \tag{10.55}$$

$$u(k) = Hz(k) + Kx(k) \tag{10.56}$$

where F, G, H, K are achieved as the unique solution of the linear equation

$$\begin{bmatrix} K & H \\ G & F \end{bmatrix} \begin{bmatrix} \bar{X} \\ \bar{Z} \end{bmatrix} = \begin{bmatrix} \bar{U} \\ \bar{V} \end{bmatrix}. \tag{10.57}$$

The next theorem states that such a compensator is stabilizing and that, denoting by $x(k)$, $z(k)$, $u(k)$ the generic solution of the closed-loop system, for $x(1) = \bar{x}$ and $z(1) = 0$ the resulting trajectory is the optimal one, namely $x(k) = \bar{x}(k)$, $u(k) = \bar{u}(k)$ and $z(k) = \bar{z}(k)$.

Theorem 10.30. *The compensator given by (10.55)–(10.57), with T invertible, is a solution of the ROC Problem.*

Proof. By combining (10.50) and (10.54) the following equation is obtained

$$\begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{X} \\ \bar{Z} \end{bmatrix} + \begin{bmatrix} B & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \bar{U} \\ \bar{V} \end{bmatrix} = \begin{bmatrix} \bar{X} \\ \bar{Z} \end{bmatrix} P. \tag{10.58}$$

The state matrix of the closed-loop system, when the compensator (10.55)–(10.56) is used, is then

$$A_{cl} = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} B & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} K & H \\ G & F \end{bmatrix}, \tag{10.59}$$

which, by (10.57) and (10.58), satisfies the condition

$$A_{cl}\begin{bmatrix}\bar{X}\\ \bar{Z}\end{bmatrix}=\begin{bmatrix}\bar{X}\\ \bar{Z}\end{bmatrix}\begin{bmatrix}0\,0\ldots0\,0\\ 1\,0\ldots0\,0\\ 0\,1\ldots0\,0\\ :\ :\ldots:\ :\\ 0\,0\ldots1\,0\end{bmatrix}=TP. \tag{10.60}$$

Since T is invertible, the matrix $A_{cl}=TPT^{-1}$ is similar to P, hence it is stable. If we consider the expression of the kth column in (10.60) and the expression of T in (10.53), denoting by P_k the kth column of P, it is readily seen that if the initial condition is

$$\begin{bmatrix}x(1)\\ z(1)\end{bmatrix}=\begin{bmatrix}\bar{x}(1)\\ \bar{z}(1)\end{bmatrix}=\begin{bmatrix}\bar{x}\\ 0\end{bmatrix}$$

then the solution

$$A_{cl}\begin{bmatrix}x(k)\\ z(k)\end{bmatrix}=\begin{bmatrix}x(k+1)\\ z(k+1)\end{bmatrix}, \tag{10.61}$$

for $k=1,2,\ldots,N-1$ is such that

$$x(k)=\bar{x}(k),\ \text{ and }\ z(k)=\bar{z}(k)$$

and

$$A_{cl}\begin{bmatrix}x(N)\\ z(N)\end{bmatrix}=0, \tag{10.62}$$

therefore the state sequence is the optimal one. The corresponding control sequence is achieved by considering (10.55)–(10.56) and in particular

$$u(k)=Hz(k)+Kx(k)=H\bar{z}(k)+K\bar{x}(k),$$

so that, from (10.57), the control sequence is $u(k)=\bar{u}(k),\ \ k=1\ldots N$, the optimal one.

The case $N=n$ is very simple. Indeed no augmentation is necessary. Since X is square invertible and $U=KX$, we immediately derive that

$$u=Kx=[\bar{U}\bar{X}^{-1}]x \tag{10.63}$$

is the desired control.

Once we have established that for the nominal initial condition no constraint violations occur, a further point worth investigating is the determination of the set

of states for which the same applies. The next proposition provides an answer in the case of linear constraints.

Proposition 10.31. *Assume that the constraint set on y, namely the set $\mathcal{Y}$, is a polyhedron including the origin in its interior*

$$\mathcal{Y} = \{y : \ P_j y \leq q_j, \ j = 1, 2, \ldots, s, \quad q_j > 0\} \tag{10.64}$$

Then the set of all the initial conditions which are driven to the origin without constraint violation is a polyhedron $\mathcal{X}_{max}$ expressed by a set of $s \times (n-1)$ inequalities as follows:

$$\mathcal{X}_{max} = \left\{ \begin{array}{c} P_j \ [C + DK \ \ DH \,] \ (A_{cl})^k \ \begin{bmatrix} I \\ 0 \end{bmatrix} x \leq q_j, \\ j = 1, 2, \ldots, s, \ k = 1, 2, \ldots, N-1 \end{array} \right\} \tag{10.65}$$

Proof. To prove the theorem, we note that

$$y(k) = Cx(k) + Du(k) = (C + DK)x(k) + DHz(k) = [C + DK \ \ DH \,]\zeta(k)$$

where $\zeta \doteq [x^T \ z^T]^T$. By exploiting the main result presented in Section 5.4 and due to [GT91] it can be seen that the set of initial states $\mathcal{X}_{max}$ for which the constraints are satisfied is defined by the set of inequalities

$$P_j \, [C + DK \ \ DH \,] \ (A_{cl})^k \, \zeta(k) \leq q_j, \quad j = 1, 2, \ldots, s, \ \ k = 1, 2, \ldots, \hat{N}$$

for some $\hat{N}$ (which is finite, under appropriate assumptions unnecessary here, but not known a priori [GT91]). Since A_{cl} is nilpotent, k ranges up to $N-1$ and for $k \geq N$ the inequalities become trivial (since $q_j > 0$). Furthermore we assume that the compensator is initialized with $z(1) = 0$ and therefore the admissible set of initial conditions is the intersection of this set and the x-space (namely the intersection with the subspace $z = 0$) which is given by the inequalities in (10.65).

In the problem formulation the final constraint $x(N+1) = 0$ was imposed and this produced a dead-beat compensator (the closed-loop matrix is similar to P). This constraint can be removed and, as a consequence, the closed-loop stability problem arises. In [BP06] it is shown how this problem can be solved via characteristic polynomial assignment. Briefly we report the main idea, which basically consists in replacing constraint $x(N+1) = 0$ by the new constraint

$$x(N+1) = \sum_{i=1}^{N} \ x(N-i+1)p_i, \tag{10.66}$$

where p_i are real numbers. Then it can be shown that the closed-loop matrix turns out to be similar to the following matrix P

$$P = \begin{bmatrix} 0 & 0 & \dots & 0 & p_N \\ 1 & 0 & \dots & 0 & p_{N-1} \\ 0 & 1 & \dots & 0 & p_{N-2} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & p_1 \end{bmatrix}. \tag{10.67}$$

Concerning the numbers p_i, there are two options.

- they can be assigned arbitrarily, say an ROC problem can be solved with characteristic polynomial assignment;
- they can be determined a posteriori. Under some conditions (see [BP06]) it can be shown that they can be taken as free in the optimization problem (thus not affecting it with the additional constraint (10.66)) and determined as the coefficients of a stable polynomial.

We finally stress that, since the zero terminal constraint is not imposed, the cost can be equipped by a weight for the final state and precisely

$$J_{opt}(\bar{x}) = \min \sum_{k=1}^{N} g(y(k)) + h(y(N+1)). \tag{10.68}$$

Clearly the control is optimal only on the considered horizon. Therefore the technique is suitable only in those cases in which "it does not matter what happens close to the target." The reader is referred to [BP06] for further motivations and details.

Example 10.32. As already mentioned, the proposed approach turns out to be useful when a system mainly operates between some prescribed positions for whom optimality is required, but it is also important to guarantee a "reasonable behavior" from any initial condition. As an example, consider the cart-pole system depicted in Fig. 10.10, and suppose that the system "normal" operation is moving from A

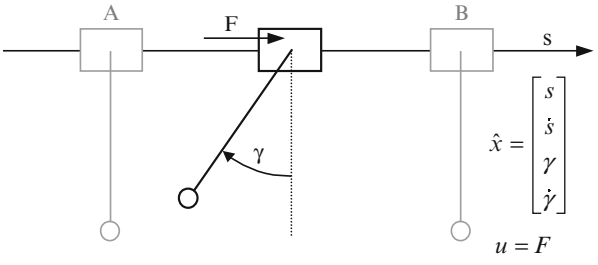


Fig. 10.10 The cart-pole system

to B (and vice versa). In the following, a stabilizing dynamic compensator which guarantees time-optimality for such an operation is computed and a comparison of its performance for a non-nominal condition versus the behavior of a trajectory-based scheme is made. The state vector $\hat{x}$ for the continuous-time system is reported in Fig. 10.10, while the zero-order-hold sampling of the linearized system leads to the following state-space representation (the parameters are: mass of the cart $0.3kg$, mass of the pole $0.1kg$, length of the pole $1m$, gravity acceleration $9.81m/s^2$, friction neglected, sampling time $0.2s$):

$$A = \begin{bmatrix} 1 & 0.2 & 0.06259 & 0.00425 \\ 0 & 1 & 0.59844 & 0.06259 \\ 0 & 0 & 0.74961 & 0.18301 \\ 0 & 0 & -2.3938 & 0.74961 \end{bmatrix}, \quad B = \begin{bmatrix} 0.065953 \\ 0.65251 \\ -0.06381 \\ -0.61004 \end{bmatrix}.$$

The target and nominal initial conditions considered are $x_B = [0\ 0\ 0\ 0]^T$ and $x_A = [-0.9\ 0\ 0\ 0]^T$. The following constraints on both the input (force applied to the cart) and the third component of the state (angle of the pole) are present:

$$|u(k)| \leq 3.5, \quad |x_3(k)| \leq 0.4. \tag{10.69}$$

By choosing $\bar{x} = x_A$, and

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \; D = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$g(y(k)) = 1, \; \mathcal{Y} = \{[y_1\ y_2]^T : |y_1| \leq 0.4, |y_2| \leq 3.5\},$$

with N free, the problem of finding the minimum-time optimal trajectory is cast into the form (10.39)–(10.46). The resulting optimal trajectory, steering the system from $\bar{x}$ to 0 in $N = 7$ steps, is reported in Fig. 10.11 (a) and (b), for the pairs (x_1, x_2) and (x_3, x_4) respectively. The optimal control sequence turns out to be:

$$\bar{U} = \begin{bmatrix} 2.9199 & -1.2314 & -1.7968 & 0.0000 & 1.7968 & 1.2314 & -2.9199 \end{bmatrix}$$

and the corresponding matrix $\bar{X}$ is

$$\bar{X} = \begin{bmatrix} -0.9 & -0.70742 & -0.42682 & -0.39439 & -0.50561 & -0.47318 & -0.19258 \\ 0 & 1.9053 & 0.87873 & -0.53399 & -0.53399 & 0.87873 & 1.9053 \\ 0 & -0.18632 & -0.38707 & -0.20076 & 0.20076 & 0.38707 & 0.18632 \\ 0 & -1.7812 & -0.138 & 1.9192 & 1.9192 & -0.138 & -1.7812 \end{bmatrix}.$$

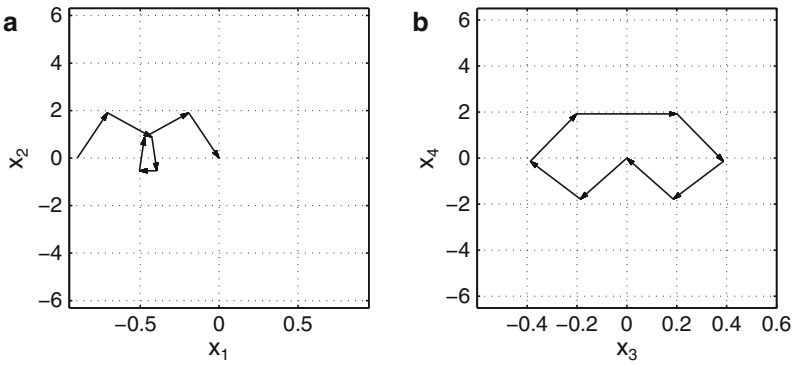


Fig. 10.11 The optimal trajectory from $\bar{x} = [-0.9\ 0\ 0\ 0]^T$

The order of the compensator is $N - n = 3$. By means of the augmentation matrix

$$\bar{Z} = \begin{bmatrix} 0 & -0.26741 & -0.14709 & 0.089061 & 0.2518 & 0.21588 & 0.068797 \\ 0 & 0.13377 & -0.047075 & -0.0082171 & 0.19501 & 0.32251 & 0.21912 \\ 0 & 0.23974 & 0.048655 & -0.028113 & 0.12015 & 0.28393 & 0.23175 \end{bmatrix},$$

we get an invertible T which yields the following matrix F:

$$F = \begin{bmatrix} 0.039928 & 0.058252 & 0.031516 \\ 0.22618 & 0.20217 & 0.0027141 \\ 0.062665 & 0.17021 & 0.20636 \end{bmatrix},$$

(whose eigenvalues are -0.00551, 0.30822, and 0.14574). Note that this matrix is stable. This fact did not occur by chance only. In [BP03], a technique to determine a dynamic compensator which is actually stable is reported. The overall compensator matrix:

$$\begin{bmatrix} K & H \\ G & F \end{bmatrix} =$$

$$= \left[\begin{array}{cccc|ccc} -3.2443 & 801.79 & 2678.68 & 583.548 & -2344.82 & -2566.59 & -1152.60 \\ \hline 0.2971 & 0.2659 & -0.06448 & 0.25835 & 0.03992 & 0.05825 & 0.03151 \\ -0.1486 & 0.01446 & 0.08137 & 0.07400 & 0.22618 & 0.20217 & 0.00271 \\ -0.2663 & -0.08883 & 0.14637 & -0.0007 & 0.06266 & 0.17021 & 0.2063 \end{array}\right]$$

By construction, the closed-loop system, initialized in $[x_A^T\ 0\ 0\ 0]^T$ (4 plant variables and 3 compensator zero variables) gives rise to the trajectory depicted in Fig. 10.11, the same trajectory obtained by means of the trajectory-tracking controller (10.47). However, starting from a non-nominal initial condition the situation is different:

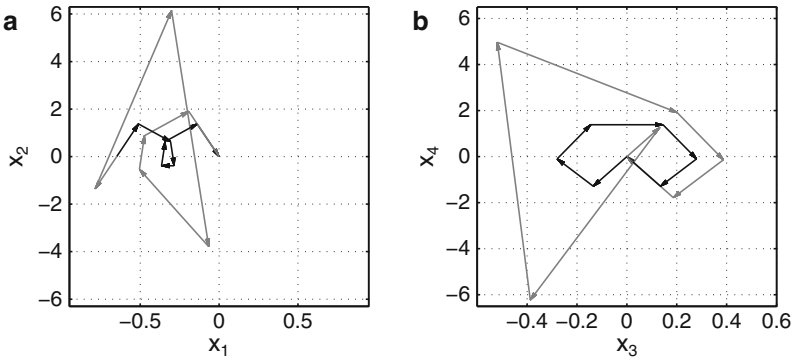


Fig. 10.12 The trajectory from $\tilde{x} = [-.65\ 0\ 0\ 0]^T$ (non-nominal initial condition) using the proposed compensator (black) and a trajectory tracking approach (gray)

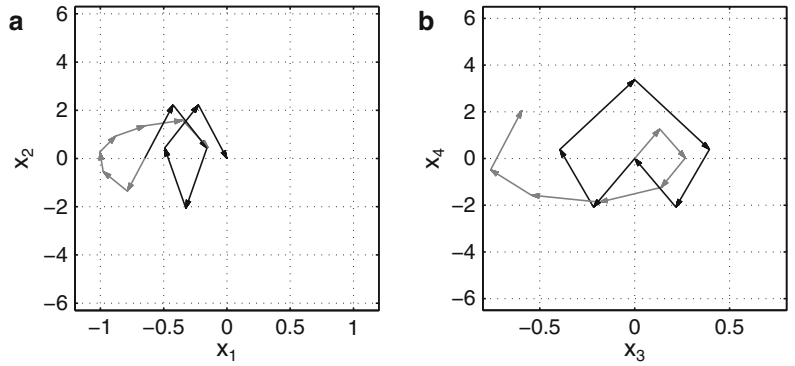


Fig. 10.13 The optimal minimum-time trajectory from $x = [-0.65\ 0\ 0\ 0]^T$ (black) and that obtained using the trajectory tracking approach saturating the input according to the first of the (10.69) (gray)

Fig. 10.12 shows the behavior of the system from $\tilde{x} = [-0.65\ 0\ 0\ 0]^T \neq x_A$ under (black line) the proposed feedback controller and (gray line) the controller (10.47) (for $K = [-19.9688\ -9.9844\ -1.056\ -6.9919]$, which assigns zero eigenvalues to $A + BK$). It is clear that the proposed compensator, producing a downscaled version of the optimal trajectory, is highly preferable to the other, which causes huge steps and violates the constraints, both on the state and the control variables. Note that the true minimum-time trajectory from $\tilde{x}$, reported in black in Fig. 10.13, is one step shorter than that obtained with the proposed compensator. Finally, we show that the trajectory-based controller produces a highly undesirable trajectory with constraint violation if we saturate the control at its assigned bound $|u| \leq 3.5$ (gray line), notwithstanding the fact that the new initial condition is closer to the target.

10.4.2 The nonlinear static solution

In the previous section, a dynamic ROC was derived. Here it is shown how to compute a static ROC, which must be necessarily nonlinear. To this aim, the following additional assumptions are made

a) the objective function is convex, positive definite, and 0-symmetric;
b) the constraints are all 0-symmetric;
c) the optimal trajectory is such that the residual cost is strictly decreasing, i.e.

$$\sum_{k=h}^{N} g(y(k)) < \sum_{k=h+1}^{N} g(y(k)).$$

The latter assumption is absolutely reasonable and avoids trivialities (it is obviously true, for instance, if $g(x,u)$ is positive definite with respect to x).

The problem can be solved as follows: starting from the points of the optimal trajectory and their opposites (connected by the plain line in Figure 10.14), the state-space is partitioned into disjoint regions. The convex hull of the points of the trajectory (the shaded hexagon in Figure 10.14) includes the nominal initial state $\bar{x}$ (possibly in its interior) and can be divided into simplices (see Chapter 3 for the definition of simplex and simplicial sector, which will be used shortly) in each of whom the control is affine. The simplicial partition induces the corresponding simplicial cones in the external part. In such convex cones, centered in the origin and "truncated to keep the outer part," the controller is linear. Besides being linear in each cone, the overall so derived control is Lipschitz-continuous. Before stating

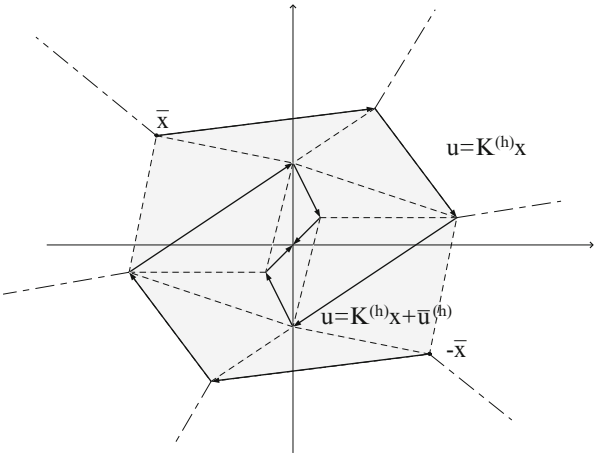


Fig. 10.14 The state-space partition.

the main result, we recall that the inequality $p \leq 0$, when p is a vector, has to be intended componentwise and that $\bar{1}$ denotes the vector (the dimension depending on the context) whose components are all equal to 1

$$\bar{1} = [\, 1 \;\; 1 \;\; \ldots \;\; 1 \,]^T \tag{10.70}$$

(so that the expression $\bar{1}^T p$ is the sum of the components of vector p).

Together with the standard simplex and simplicial sector notation, we need to consider the complement (the outer part) of the "unit simplex" in a simplicial cone, which is the closure of the complement in $\mathcal{C}$:

$$\tilde{\mathcal{C}} = \{x = Xp : \;\; p \geq 0, \;\; \bar{1}^T p \geq 1\}. \tag{10.71}$$

A pictorial explanation of the construction is provided in Fig. 10.14.

If $\Phi : \mathbb{R}^n \to \mathbb{R}$ is a function and X is an $n \times m$ matrix we now denote by $\Phi(X)$ the following matrix

$$\Phi(X) = [\, \Phi(x_1) \;\; \Phi(x_2) \;\; \ldots \;\; \Phi(x_m) \,]. \tag{10.72}$$

The next theorem states that there exists a control which is optimal for $x(1) = \bar{x}$ and locally stabilizing.

Theorem 10.33. *There exists a convex and compact polyhedron $\mathcal{P}$ including the origin in its interior such that $\bar{x} \in \mathcal{P}$ and which is partitioned into simplices $\mathcal{S}^{(h)}$, each generated by an $n \times (n+1)$ matrix $X^{(h)}$ whose columns are vectors properly chosen among the vectors of the optimal state trajectory and such that the intersection of two different simplices has empty interior, say*

$$\mathcal{P} = \bigcup \mathcal{S}^{(h)} = \bigcup \mathcal{S}(X^{(h)}), \;\; int\{\mathcal{S}^{(h)} \bigcap \mathcal{S}^{(k)}\} = \emptyset, \quad h \neq k \tag{10.73}$$

It is possible to associate with each simplex $\mathcal{S}^{(h)}$ an $m \times (n+1)$ matrix $U^{(h)}$ whose columns are vectors properly chosen among those of $\Phi(X)$, the inputs of the optimal trajectory, and to define the following piecewise affine static controller

$$u = \Phi_{\mathcal{P}}(x) = K^{(h)} x + \bar{u}^{(h)} = U^{(h)} \begin{bmatrix} X^{(h)} \\ \bar{1}^T \end{bmatrix}^{-1} \begin{bmatrix} x \\ 1 \end{bmatrix}, \quad for \;\; x \in \mathcal{S}^{(h)} \tag{10.74}$$

which is Lipschitz-continuous and relatively optimal inside $\mathcal{P}$. More precisely, it is stabilizing with domain of attraction $\mathcal{P}$ and, for $x(1) = \bar{x}$, produces the optimal trajectory. Moreover, for each $x(1) \in \mathcal{P}$, the constraints are satisfied and the transient cost is bounded as

$$J(x(1)) \leq \max_{i=1,\ldots,n+1} J_{opt}(x_{k_i}), \tag{10.75}$$

where $x_{k_1}, x_{k_2}, \ldots, x_{k_{n+1}}$ are the vertices of a simplex including $x(1)$ and $J_{opt}(x_{k_i})$ is the optimal cost associated with the initial condition x_{k_i}.

The next theorem states that the same control can be globally extended over $\mathbb{R}^n$ (the external part in Fig. 10.14).

Theorem 10.34. *The control (10.74) can be extended onto $\mathbb{R}^n$ as follows. The complement of the polytope $\mathcal{P}$ can be partitioned into complements of simplices inside a cone*

$$\tilde{\mathcal{C}}^{(h)} = \tilde{\mathcal{C}}(X^{(h)}) = \{x = X^{(h)}p : \ p \geq 0, \ \bar{1}^T p \geq 1\}, \tag{10.76}$$

each generated by a square invertible matrix $X^{(h)}$, having intersection with empty interior among one another

$$int\{\tilde{\mathcal{C}}^{(h)} \bigcap \tilde{\mathcal{C}}^{(k)}\} = \emptyset, \quad h \neq k \tag{10.77}$$

and having intersection with empty interior with $\mathcal{P}$

$$int\{\tilde{\mathcal{C}}^{(h)} \bigcap \mathcal{P}\} = \emptyset \tag{10.78}$$

and such that

$$\mathcal{P} \bigcup \left[\bigcup_h \tilde{\mathcal{C}}^{(h)}\right] = \mathbb{R}^n. \tag{10.79}$$

To each set $\tilde{\mathcal{C}}^{(h)}$ it is possible to associate an $m \times n$ matrix $U^{(h)}$ whose columns are vectors properly chosen among the inputs of the optimal trajectory, thus obtaining the control law

$$u = \Phi(x) = K^{(h)}x = U^{(h)} \left[X^{(h)}\right]^{-1} x \tag{10.80}$$

The extended control obtained in this way is globally Lipschitz-continuous and relatively optimal.

A sketch of the proof of the theorem and of the construction of the controller is now reported. Denote by $\bar{x}(1), \ldots, \bar{x}(N)$ the optimal state trajectory obtained by solving the open-loop optimal control problem when the system is initialized at $\bar{x} = \bar{x}(1)$. Revert and shift by one the time index and for $i = 1, \ldots, N$ define

$$x_i = \bar{x}(N - i + 1), \ \ u_i = \bar{u}(N - i + 1), \ \ x_{-i} = -x_i, \tag{10.81}$$

so that $x_0 = 0$ and

$$x_{i-1} = Ax_i + Bu_i, \ \ i = 1, \ldots, N.$$

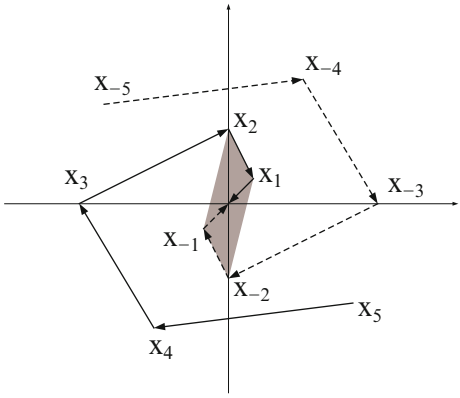


Fig. 10.15 Example of the set $\mathcal{P}_n$ (gray area) in a two-dimensional space.

To keep things simple, the following assumption (which can be easily removed as shown in [BP07]) is introduced: matrix $X_n = [x_1\ x_2\ \dots\ x_n]$, formed by the last n states of the optimal trajectory, is invertible. Assume also that the residual cost is strictly decreasing along the optimal trajectory, more precisely that the optimal cost with initial state x_i is strictly greater than the cost with initial state x_{i-1} and consider the polyhedral set

$$\mathcal{P}_n = \{x = X_n p : \|p\|_1 \leq 1\}, \tag{10.82}$$

which is the convex hull of the last n states of the optimal trajectory and their opposites. Such a set clearly contains the origin in its interior and it is 0-symmetric. An example for $n = 2$ is shown in Figure 10.15: $\mathcal{P}_n$ is the convex hull of the last two states of the optimal trajectory (connected by the plain line) and their opposite (connected by the dashed line). Thanks to the assumption of invertibility of matrix X_n, the following lemma holds.

Lemma 10.35. *The linear control (proposed in [GC86a])*

$$u(x) = [U_n X_n^{-1}]x, \tag{10.83}$$

where $U_n = [u_1\ u_2\ \dots\ u_n]$, renders positively invariant the set $\mathcal{P}_n$ and is such that no constraint violation occurs for all the initial conditions inside the set. In particular, it is dead-beat and steers the state to zero in at most n steps.

Proof. The control law $u(x) = U_n X_n^{-1} x$ is a control-at-the-vertices strategy. All $x \in \mathcal{P}_n$ can be written in a unique way as a linear combination of the columns of X_n, namely the last n states of the optimal trajectory:

$$x = X_n p. \tag{10.84}$$

Since X_n is invertible, it follows that

$$p(x) = X_n^{-1} x, \tag{10.85}$$

hence the control law $u(x) = U_n X_n^{-1} x$ basically computes a control which is a linear combination of the controls at the vertices of $\mathcal{P}_n$ according to the coefficients $p(x)$. Positive invariance is a consequence of the fact that, by construction, the control at each vertex keeps the state inside the set [GC86a, Bla99]. Constraints satisfaction is guaranteed for all initial conditions inside the set since input and state constraints are convex and 0-symmetric. To prove that the control is dead-beat, note that if at time k we have

$$x(k) = x_n p_n + \cdots + x_2 p_2 + x_1 p_1, \tag{10.86}$$

then the computed control will be

$$u(k) = u_n p_n + \cdots + u_2 p_2 + u_1 p_1. \tag{10.87}$$

Since $x_{i-1} = Ax_i + Bu_i$, by linearity,

$$x(k+1) = x_{n-1} p_n + \cdots + x_1 p_2 + 0 p_1, \tag{10.88}$$

and at the next step, by repeating the same reasoning,

$$x(k+2) = x_{n-2} p_n + \cdots + x_1 p_3 + 0 p_2 + 0 p_1, \tag{10.89}$$

and so on and therefore after at most n steps the system will reach the origin.

Remark 10.36. The control law defined above is such that inside $\mathcal{P}_n$, at each step, the state is a convex combination of points with decreasing index and 0.

Note that, if the system reaches the state $x_n = \bar{x}(N - n + 1) \in \mathcal{P}_n$, it starts following the last n steps of the optimal trajectory. Note also that $\mathcal{P}_n$ (which will be the first element of a sequence of sets) is affine to a diamond[7] and thus can be partitioned into simplices. The next sets of the sequence are computed as follows. Consider the state x_{n+1} (corresponding to the state x_3 in the example of Figure 10.15). Since x_{n+1} and its opposite $-x_{n+1}$ are outside $\mathcal{P}_n$ (as it will be shown later on), they can be connected with a certain number of vertices of $\mathcal{P}_n$ without crossing such a set, thus simplices are formed by some vertices of $\mathcal{P}_n$ and the two points x_{n+1} and $x_{-(n+1)}$ (in the example of Figure 10.16, such simplices are the triangles (x_3, x_2, x_{-1}) and (x_3, x_{-1}, x_{-2}) and their symmetric).

[7] $\mathcal{P}$ is affine to $\mathcal{Q}$ if $\mathcal{Q} = T\mathcal{P} + q_0$ with T invertible.

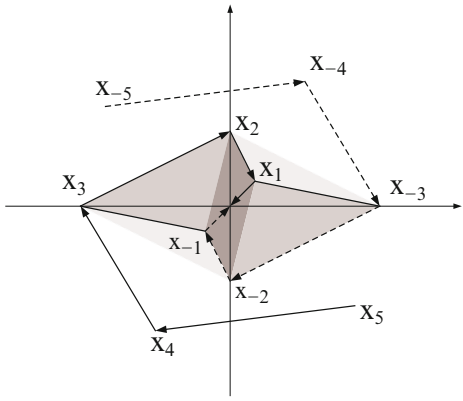


Fig. 10.16 Considering x_3 and its opposite x_{-3}, four simplices can be constructed.

Denoting by $\mathcal{S}^j_{n+1}, j = 1 \dots m_{n+1}$, the simplices having x_{n+1} as a vertex and with $\mathcal{S}^j_{n+1}, j = -m_{n+1} \cdots -1$ those having $x_{-(n+1)}$ as a vertex, we can define the set $\mathcal{P}_{n+1}$ as follows:

$$\mathcal{P}_{n+1} = \bigcup_{j=\pm 1 \cdots \pm m_{n+1}} \mathcal{S}^j_{n+1} \cup \mathcal{P}_n. \tag{10.90}$$

The procedure goes on exactly in the same manner to generate the sequence of sets $\mathcal{P}_k, k = n+1, n+2, \dots N$, ordered by inclusion and the corresponding simplicial partition: if we define the set

$$\mathcal{P}_k = \text{conv}\{x_1, x_2, \dots, x_k, x_{-1}, x_{-2}, \dots, x_{-k}\}, \quad k < N, \tag{10.91}$$

we can consider the vector x_{k+1}, and form a new simplicial partition for $\mathcal{P}_{k+1}$ by adding new simplices. It is fundamental to note that each new simplicial partition of $\mathcal{P}_{k+1}$ preserves all the simplices forming the simplicial partition for $\mathcal{P}_k$. To prove that the construction is well-defined we need the following lemma.

Lemma 10.37. *The vector x_{k+1} in the construction is outside $\mathcal{P}_k$.*

Proof. Assume by contradiction that $x_{k+1} \in \mathcal{P}_k$. Then x_{k+1} could be written as a convex combination of the vertices of $\mathcal{P}_k$. So if we take x_{k+1} as initial state, since we considered convex constraints, then x_{k+1} could be driven to zero in a time not exceeding k at a cost not exceeding the maximum cost of all the vertices of $\mathcal{P}_k$. This is in contradiction with the assumption that the cost is strictly decreasing along the optimal trajectory.

Therefore the procedure is such that $\{\mathcal{P}_k\}$ is a strictly increasing (in the sense of inclusion) sequence of sets, each of which preserves the simplicial partition of the former. This construction terminates once $\mathcal{P} \doteq \mathcal{P}_N$ is constructed.

The next step is to show how to associate a control with each simplex. For each of the simplices $\mathcal{S}_k^j$:

1. order (arbitrarily) the vertices;
2. associate a matrix X_k^j whose columns are the ordered vertices;
3. associate a matrix U_k^j whose columns are the controls corresponding to the ordered vertices (if the vertex belongs to the optimal trajectory, take the corresponding control, if it belongs to the opposite of the optimal trajectory, take the opposite of the corresponding control).

Consider the following control strategy. Given $x \in \mathcal{P}$,

- if $x \in \mathcal{P}_n$ then

$$\Phi_{\mathcal{P}}(x) = U_n X_n^{-1} x, \tag{10.92}$$

- otherwise, if $x \in \mathcal{S}_k^j$ then

$$\Phi_{\mathcal{P}}(x) = U_k^j p, \tag{10.93}$$

where $p \geq 0$ is the (unique) vector such that

$$x = X_k^j p, \;\; \bar{1}^T p = 1. \tag{10.94}$$

Note that p is such that

$$\begin{bmatrix} X_k^j \\ \bar{1}^T \end{bmatrix} p = \begin{bmatrix} x \\ 1 \end{bmatrix}, \tag{10.95}$$

so that the proposed control is of the form (10.74).

To show that the control is dead-beat, we need to introduce the index $\text{In}(\mathcal{S})$ of a sector $\mathcal{S}$ as the maximum of the absolute values of the indices of its generating vectors. Formally, if $\mathcal{S}$ is generated by corners $x_{k_1}, x_{k_2}, \ldots, x_{k_n}$, then

$$\text{In}(\mathcal{S}) = \max\{|k_1|, |k_2|, \ldots, |k_n|\} \tag{10.96}$$

For reasons that will be clear soon, $\text{In}(\mathcal{S})$ will be referred to as the *distance* of $\mathcal{S}$ from 0.

Remark 10.38. The notion of "sector" deserves some comments. Indeed we now consider possibly degenerate simplices that can have empty interior being formed by some points x_k and the origin repeatedly considered. For instance, $\mathcal{S}$ could be generated by $[0\ 0\ x_1\ x_2]$, representing a two-dimensional degenerate simplex in the three dimensional space. Note also that $\text{In}(\mathcal{S}) \leq k$ for all sectors inside $\mathcal{P}_k$.

The next lemma shows that, with the proposed control, if the system state is inside a sector, then it jumps to another one closer to zero.

Lemma 10.39. *The proposed strategy is such that if $x \in \mathcal{S}$, a sector of $\mathcal{P}_k$, then $Ax + Bu(x) \in \mathcal{S}'$ with*

$$In(\mathcal{S}') < In(\mathcal{S}), \tag{10.97}$$

as long as $In(\mathcal{S}) \neq 0$, and therefore if $x(1) \in \mathcal{P}_k$, the control steers the system to zero in at most k steps.

Proof. As a first step we remind that, according to Lemma 10.35 and Remark 10.36, the jump to a sector closer to 0 occurs $\forall x \in \mathcal{P}_n$. Now we proceed by induction. Assume that $x \in \mathcal{P}_{n+1}$. If $x \in \mathcal{P}_n$ there is nothing to prove, otherwise x is necessarily in a sector $\mathcal{S}$ generated by x_{n+1} or its opposite $x_{-(n+1)}$ and other vertices of smaller indices

$$x = \sum_{i=1}^{n+1} x_{k_i} p_i, \quad \sum_{i=1}^{n+1} p_i = 1, \quad p_i \geq 0, \tag{10.98}$$

with $|k_i| \leq n$, $i = 1, 2, \ldots, n$, and $|k_{n+1}| = n + 1$. Then we have, by construction,

$$Ax + B\Phi_{\mathcal{P}}(x) = A[\sum_{i=1}^{n+1} x_{k_i} p_i] + B[\sum_{i=1}^{n+1} u_{k_i} p_i] =$$

$$= \sum_{i=1}^{n+1} p_i \underbrace{[Ax_{k_i} + Bu_{k_i}]}_{\in \mathcal{P}_n} \in \mathcal{P}_n.$$

Therefore $Ax + B\Phi_{\mathcal{P}}(x)$ is necessarily in a sector with index $\text{In} \leq n$. The rest of the proof proceeds in the same way. Any point x in $\mathcal{P}_{k+1}$, if not in $\mathcal{P}_k$, is included in a sector $\mathcal{S}$ with index $\text{In}(\mathcal{S}) = k + 1$ and, by means of the same machinery, we can show that $Ax + B\Phi_{\mathcal{P}}(x) \in \mathcal{S}'$ with $\text{In}(\mathcal{S}') \leq k$. The fact that if $x(1) \in \mathcal{P}_k$, the state converges to 0 in at most k steps is an immediate consequence.

The procedure for the construction of the controller and the state partition can be summarized as follows.

Procedure. Construction of a static ROC.

Given the optimal open-loop trajectory, which satisfies the assumptions, perform the following operations.

1. Let the set $\mathcal{P}_n = \{x : x = X_n p, \|p\|_1 \leq 1\}$, where $X_n = [x_1\ x_2\ \ldots\ x_n]$, be the convex hull of the last n states of the optimal trajectory and their opposite.
2. Let $U_n = [u_1\ u_2\ \ldots\ u_n]$ be the matrix whose columns are the control vectors corresponding to the last n states of the optimal trajectory.
3. Take $i = n + 1$.

4. Construct the simplices $\mathcal{S}_i^j$, $j = \pm 1 \cdots \pm m_i$ by connecting x_i and x_{-i} to the vertices of $\mathcal{P}_{i-1}$ without crossing such set. This is always possible, since $x_i, x_{-i} \notin \mathcal{P}_{i-1}$.
5. Let X_i^j be the matrix whose columns are the vertices of $\mathcal{S}_i^j$ in an arbitrary order and U_i^j the controls corresponding to the vertices in the same order. For vertices belonging to the opposite of the optimal trajectory, take the opposite of the control.
6. Let $\mathcal{P}_i = \bigcup_j \mathcal{S}_i^j \cup \mathcal{P}_{i-1}$.
7. Increase i.
8. If $i \leq N$, go back to step 4.

Note that, by construction, the sets $\mathcal{P}_i$, $i = n, \ldots, N$ are convex, 0-symmetric and such that $\mathcal{P}_i \subset \mathcal{P}_{i+1}$. The set $\mathcal{P}_{i+1} \setminus \mathcal{P}_i$, the difference between $\mathcal{P}_{i+1}$ and $\mathcal{P}_i$, is composed of simplices S_i^j each of whom has all the vertices but one (precisely x_{i+1} or $x_{-(i+1)}$) belonging to $\mathcal{P}_i$.

For $x \in \mathcal{P} = \mathcal{P}_N$, the controller described above is relatively optimal. However, the control law is not defined for $x \notin \mathcal{P}$. This control can be actually extended outside since $\mathcal{P}$ is a polytope including the origin in its interior. Then we can still use the control induced by this polytope [GC86a]. It can be shown that the derived control is globally Lipschitz (over the whole $\mathbb{R}^n$) and globally stabilizing (although it might violate the constraints for initial states outside $\mathcal{P}$) [BP07].

Example 10.40 (Static ROC construction). Consider the double integrator:

$$x(k+1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k), \tag{10.99}$$

under the constraints $|x(k)| \leq 5, |u(k)| \leq 3$. Given the initial state $x(1) = [-2 \ 5]^T$, the horizon $N = 5$, the final state $x(N+1) = 0$, and the cost function $J = \sum_{i=1}^{N} u(k)^2$, the optimal (open-loop) control and trajectory, found by solving a quadratic-programming problem, are respectively:

$$\bar{U} = \begin{bmatrix} -3 & -2.9 & -1.3 & 0.3 & 1.9 \end{bmatrix} \tag{10.100}$$

and

$$\bar{X} = \begin{bmatrix} -2 & 3 & 5 & 4.1 & 1.9 \\ 5 & 2 & -0.9 & -2.2 & -1.9 \end{bmatrix}. \tag{10.101}$$

The optimal trajectory is reported in Figure 10.17. By means of the proposed procedure the triangulation reported in Figure 10.18 is obtained; the number of triangles is 12 (including the four triangles in which the darkest region, i.e. $\mathcal{P}_2$, can be split). The piecewise affine control law obtained by applying a control-at-the-vertices strategy inside each of the triangles, as stated above, is relatively optimal, hence is optimal for the nominal initial condition and guarantees convergence

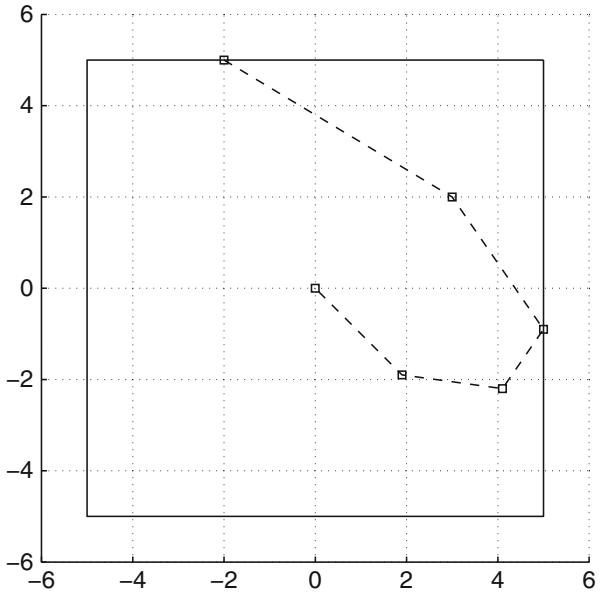


Fig. 10.17 The optimal trajectory.

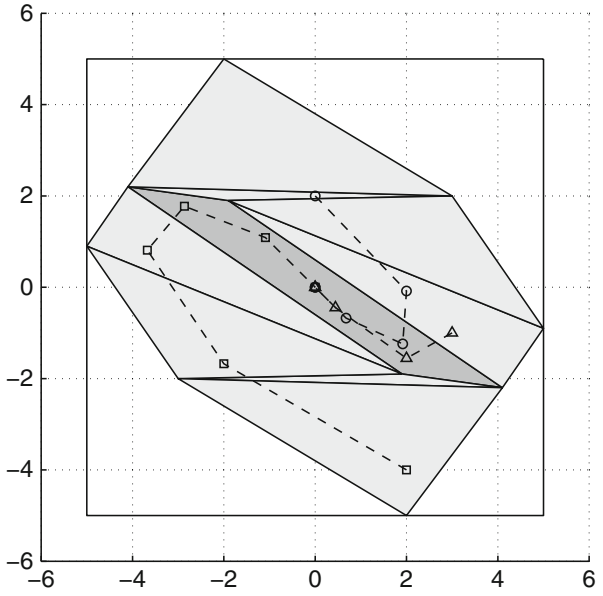


Fig. 10.18 The triangulation induced by the optimal trajectory and the trajectories from three non-nominal initial conditions inside $\mathcal{P}$.

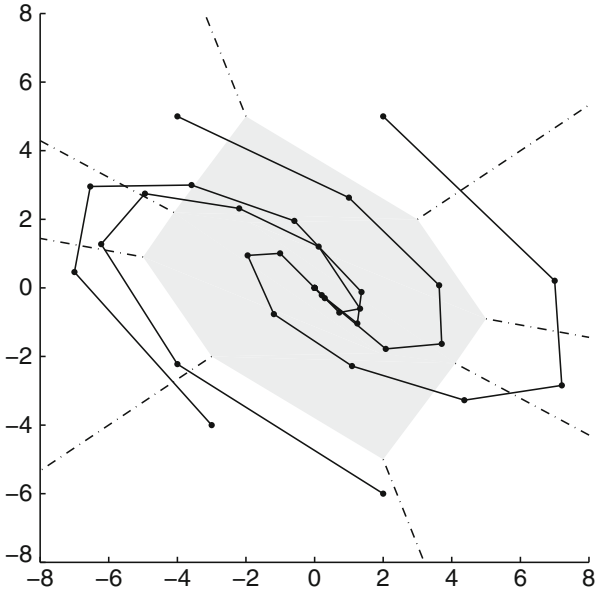


Fig. 10.19 Some trajectories starting from outside $\mathcal{P}$.

and constraint satisfaction for the other initial conditions. In Figure 10.18, the trajectories from three non-nominal initial conditions are reported. Note that the number of steps required to reach the origin depends on the triangle the initial state belongs to. Figure 10.19 shows the effectiveness of the *extended* control, reporting some trajectories starting from outside $\mathcal{P}$; the dash-dotted lines represent the boundaries between the simplicial cones $\mathcal{C}^{(h)}$ in the complement of $\mathcal{P}$.

10.5 Merging Lyapunov function

In constrained optimal control, constraints are in general active when the state is far from the origin. Typically a Lyapunov function which is suitable to face constraints maximizing the size of the domain of attraction is not suitable to assure performances when the state is close to zero, the attractor. In this section, we present an interesting idea proposed in [BCG12], named Lyapunov function merging. Merging two control Lyapunov functions Ψ_1 and Ψ_2 basically consists in producing a new control Lyapunov function which is close to Ψ_2 near the origin and close to Ψ_1 far away.

Given two smooth radially unbounded control Lyapunov functions Ψ_1 and Ψ_2, we say that the function Ψ is a gradient-type merging of Ψ_1 and Ψ_2 if it is positive definite, smooth, radially unbounded and if

$$\nabla\Psi(x) = \rho_1(x)\nabla\Psi_1(x) + \rho_2(x)\nabla\Psi_2(x), \quad \forall x$$

with $\rho_1(x) > 0$ and $\rho_2(x) > 0$ for $x \neq 0$. A gradient type merging is easy to find. For instance, take

$$\Psi(x) = \alpha_1(\Psi_1(x), \Psi_2(x))\Psi_1(x) + \alpha_2(\Psi_1(x), \Psi_2(x))\Psi_2(x)$$

with smooth positive-definite scalar functions α_1 and α_2. Then

$$\nabla\Psi = \left(\alpha_1 + \Psi_1\frac{\partial\alpha_1}{\partial\Psi_1} + \Psi_2\frac{\partial\alpha_2}{\partial\Psi_1}\right)\nabla\Psi_1(x) + \left(\alpha_2 + \Psi_1\frac{\partial\alpha_1}{\partial\Psi_2} + \Psi_2\frac{\partial\alpha_2}{\partial\Psi_2}\right)\nabla\Psi_2(x)$$

Unfortunately, the new function is not necessarily a control Lyapunov function, even if the two original functions are such, unless proper conditions are satisfied.

Consider the case of a system of the form

$$\dot{x} = f(x) + G(x)u \tag{10.102}$$

We say that Ψ_1 and Ψ_2 have the control-sharing property if the two conditions

$$\begin{aligned}\nabla\Psi_1(x)(f(x) + G(x)\Phi(x)) &\leq -\phi_1(\|x\|)\\ \nabla\Psi_2(x)(f(x) + G(x)\Phi(x)) &\leq -\phi_2(\|x\|)\end{aligned}$$

are satisfied by the same feedback control Φ for some κ functions ϕ_1 and ϕ_2. Then we have the following.

Proposition 10.41. *The following two conditions are equivalent.*

1. *Any gradient-type merging of Ψ_1 and Ψ_2 is a control Lyapunov function.*
2. *Ψ_1 and Ψ_2 have the control-sharing property.*

Sufficiency is immediate. Indeed

$$\begin{aligned}&\nabla\Psi(x)(f(x) + G(x)\Phi(x))\\ &= \rho_1(x)\nabla\Psi_1(x)(f(x) + G(x)\Phi(x)) + \rho_2(x)\nabla\Psi_2(x)(f(x) + G(x)\Phi(x)) \leq\\ &-(\rho_1(x)\phi_1(\|x\|) + \rho_2(x)\phi_2(\|x\|)) \doteq -\phi(\|x\|)\end{aligned}$$

with ϕ a κ function. For necessity see [GBC14].

An interesting example of merging is achieved by the "R-composition" [BCG12]. Let Ψ_1 be a convex positive definite, function positively homogeneous of order 2, which is associated with $\mathcal{N}[\Psi_1, 1]$, a "large" domain of attraction. Let Ψ_2 be a

positively homogeneous function of order 2 which is associated with an optimal control, for instance, in the linear case, $\Psi_2(x) = x^T P x$, with P the positive definite solution of the Riccati equation. Assume that

$$\mathcal{N}[\Psi_1, 1] \subset \mathcal{N}[\Psi_2, 1]$$

or equivalently $\Psi_2(x) < \Psi_1(x) \leq 1$ for $x \neq 0$. This is not a restriction since scaling the function is equivalent to scaling the cost function. Then

1) Define $R_1, R_2 : \mathbb{R}^n \to \mathbb{R}^n$ as $R_i(x) = 1 - \Psi_i(x), \;\; i = 1, 2$.
2) For fixed $\phi > 0$, define

$$R_\wedge(x) := \rho(\phi) \left(\phi R_1(x) + R_2(x) - \sqrt{\phi^2 R_1(x)^2 + R_2(x)^2} \right),$$

where $\rho(\phi) := \left(\phi + 1 - \sqrt{\phi^2 + 1} \right)^{-1}$ is a normalization factor.
3) Define the "R-composition" as

$$\Psi_\wedge(x) := 1 - R_\wedge(x).$$

Then we have that [BCG12]

$$\nabla \Psi_\wedge(x) = \rho(\phi) \left[\phi c_1(\phi, x) \nabla \Psi_1(x) + c_2(\phi, x) \nabla \Psi_2(x) \right],$$

where $c_1, c_2 > 0$ are defined as

$$c_1(\phi, x) := 1 + \frac{-\phi R_1(x)}{\sqrt{\phi^2 R_1(x)^2 + R_2(x)^2}}, \;\; c_2(\phi, x) := 1 + \frac{-R_2(x)}{\sqrt{\phi^2 R_1(x)^2 + R_2(x)^2}}.$$

The so defined R-function is positive definite and differentiable. It is not convex in general even if the "parents" Ψ_1 and Ψ_2 are convex. Such a function is a gradient-type merging candidate.

According to Proposition 10.41, if Ψ_1 and Ψ_2 are CLFs and share a (possibly constrained) control Φ, then Φ is admissible as well for $\Phi_\wedge$, which turns out to be a CLF for (10.102) (possibly under constraints). In this case, we will refer to the CLF $\Psi_\wedge$ as Control Lyapunov R-Function (CLRF).

It is not difficult to see that, independently of $\phi > 0$, the unit sublevel set of $\Psi_\wedge$ is the same as that of Ψ_1

$$\mathcal{N}[\Psi_\wedge, 1] = \mathcal{N}[\Psi_1, 1],$$

so the domain of attraction is preserved.

Conversely, close to the origin, the level set of $\Psi_\wedge$ are close to those of Ψ_2. Parameter ρ imposes a trade-off between the shape of the level sets of Ψ_1 and of Ψ_2. Namely, in light of [BCG12], we have

$$\Psi_\wedge(x) \overset{\phi\to\infty}{\longrightarrow} \Psi_2(x) \quad \text{and} \quad \Psi_\wedge(x) \overset{\phi\to 0^+}{\longrightarrow} \Psi_1(x)$$

point-wise in $int\{\mathcal{N}[\Psi_\wedge, 1]\}$. Moreover [BCG12], we have

$$\nabla\Psi_\wedge(x) \overset{\phi\to\infty}{\longrightarrow} \nabla\Psi_2(x), \quad \text{and} \quad \nabla\Psi_\wedge(x) \overset{\phi\to 0^+}{\longrightarrow} \nabla\Psi_1(x)$$

uniformly on $\mathcal{N}[\Psi_\wedge, 1-\epsilon]$ for any $\epsilon > 0$ small.

This particular property of fixing the "external" shape, while making the "inner" one "close" to any given choice can be exploited to fix a "large" DoA while achieving "locally-optimal" closed-loop performances.

10.5.1 Controller design under constraints

We now investigate the existence of a continuous locally optimal control under constraints $x \in \mathcal{N}[\Psi_1, 1]$ and $u \in \mathcal{U}$ which is closed (possibly compact) and convex. For simplicity, we consider (10.102) with $G(x) = B$. Since the CLF $\Psi_\wedge$ is differentiable, in principle, the existence of a stabilizing control law Φ, continuous with the exception of the origin, or including $x = 0$ if $\Psi_\wedge$ satisfies the small control property[8], could be proved by using the arguments in [FK96b, Chapters 2–4].

We scale Ψ_2 in such a way that $\mathcal{N}[\Psi_1, 1] \subset \mathcal{N}[\Psi_2, 1]$. Assume that functions Ψ_1 and Ψ_2 have the control-sharing property. Associated with Ψ_2 there is an "optimal" continuous control law Φ_2. The following convex-valued mapping of admissible (constrained) controls is non-empty for all $x \in \mathcal{N}[\Psi_1, 1]$

$$\Omega(x) := \{u \in \mathcal{U} : \ \nabla\Psi_\wedge(x)\,(f(x) + Bu) + \phi(\|x\|) \le 0\}\,. \tag{10.103}$$

We then propose the control law

$$\Phi(x) := \arg\min\{\|\upsilon - \Phi_2(x)\| : \ \upsilon \in \Omega(x)\}\,. \tag{10.104}$$

Then the control law (10.104) associated with $\Psi_\wedge$ is continuous, satisfies the constraints in $\mathcal{N}[\Psi_1, 1]$, and is locally optimal [GBC14]. The same results hold without essential restriction for differential inclusions of the form $\dot{x} = A(w)x + Bu$.

[8] A CLF Ψ satisfies the small control property if, for $u := \Phi(x)$, we have that for all $\upsilon > 0$ there exists $\epsilon > 0$ so that, whenever $\|x\| < \epsilon$ we have $\|u\| < \upsilon$ [Son98].

10.5.2 Illustrative example

We address the constrained stabilization of a simplified inverted pendulum proposed in [GBC14], whose dynamics is given by the nonlinear differential equation

$$I\ddot{\theta}(t) = mgl\sin(\theta(t)) + \tau(t)$$

The goal is the stabilization of $(\theta, \dot{\theta})$ to the origin, under the constraints $|\theta| \leq \frac{\pi}{4}$, $|\dot{\theta}| \leq \frac{\pi}{4}$ and $|\tau| \leq 2$. With the notation $x_1 = \theta$, $x_2 = \dot{\theta} = \dot{x}_1$, $u = \tau$ and

$$w(x) := \left\{ \frac{\sin(x_1)}{x_1} : \ |x_1| \leq \frac{\pi}{4} \right\}$$

we can consider the following constrained uncertain "absorbing" model:

$$\dot{x} \in \begin{bmatrix} 0 & 1 \\ aw(x) & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ b \end{bmatrix} u, \tag{10.105}$$

where $a = (mgl/I), b = (1/I); w(x) \simeq [0.89, 1], w(0) = 1; |x_1| \leq \pi/4, |x_2| \leq \pi/4$, $|u| \leq 2$. The numerical parameters used in the simulation are $I = 0.05, m = 0.5$, $g = 9.81, l = 0.3$.

We adopt the infinite-horizon quadratic performance cost

$$J(x, u) := \int_0^\infty (\|x(t)\|_Q^2 + \|u(t)\|_R^2)dt,$$

with weighting matrices $Q = I_2$, $R = 10$. Let us define the locally optimal (i.e., for $w \equiv 1$) cost function $\bar{\Psi}_2(x) = x^\top Px$, where P is the unique solution of the Algebraic Riccati Equation. To accommodate constraints, we consider the function $\bar{\Psi}_1(x) = \|Fx\|_\infty^2$, with

$$F = \begin{bmatrix} 0 & 1.53 & 4/\pi \\ 4/\pi & 0.51 & 0 \end{bmatrix}^\top,$$

which turns out to be a polyhedral control Lyapunov function for the constrained linear differential inclusions (10.105) and therefore also for the constrained nonlinear system. Then we smooth this function by taking the polyhedral control Lyapunov function $\Psi_1(x) = \|Fx\|_{40}^2$ [BM99c], achieving almost the same DoA as that assured by $\mathcal{N}[\bar{\Psi}_1, 1]$. Let us also define Ψ_2 scaling $\bar{\Psi}_2$, so that $\mathcal{N}[\Psi_1, 1] \subset \mathcal{N}[\Psi_2, 1]$. Functions Ψ_1 and Ψ_2 share a constrained control [GBC14], therefore any gradient-type merging is a CLF.

Now, Ψ_1 has a "large" DoA but it induces a "poor" performance when used with gradient-based controllers of the kind (10.104) (Figure 10.20 in fact shows that the constraint $u \in \Omega(x)$ in (10.103) with Ψ_1 in place of $\Psi_\wedge$ may be "too restrictive").

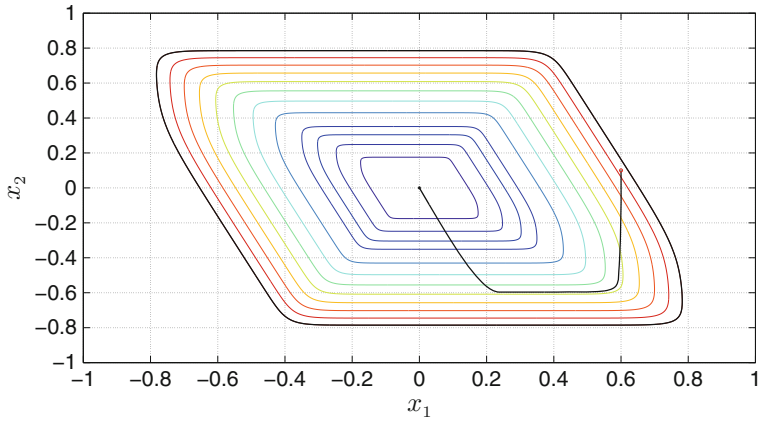


Fig. 10.20 A controlled state trajectory starting from $x_0 = (0.6,\ 0.1)^\top$ and converging to the origin. The state is actually "forced" to always "enter" the level sets of the smoothed polyhedral CLF Ψ_1.

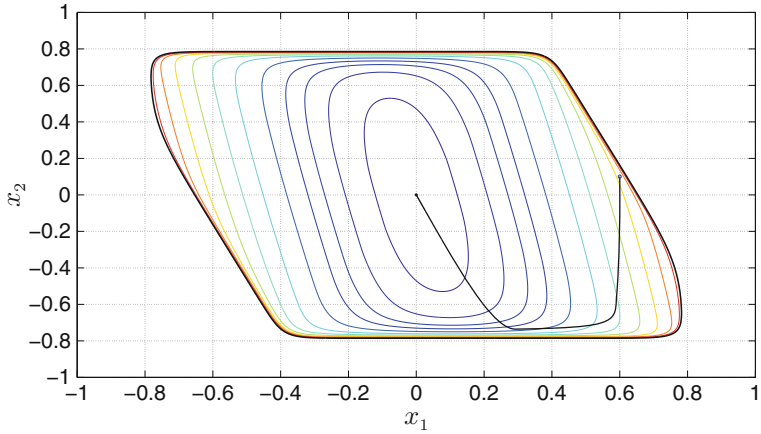


Fig. 10.21 A controlled state trajectory starting from $x_0 = (0.6,\ 0.1)^\top$ and converging to the origin "in accordance" with the level sets of the control Lyapunov R-function $\Psi_\wedge$.

On the other hand, Ψ_2 is locally optimal, but both gradient-based controllers, for instance (10.104) with Ψ_2 in place of $\Psi_\wedge$, and the standard LQ regulator, yield constraint violations.

If we take the merging $\Psi_\wedge$ with $\phi = 20$ (see Figure 10.21), we notice that $\Psi_\wedge$, with controller (10.104), inherits the benefits of Ψ_1 ("large" DoA under constraints) and Ψ_2 (local optimality). For the linearized system (i.e., for $w \equiv 1$), extensive Monte Carlo numerical experiments show that the closed-loop performance is "quite close" to the constrained "global optimal" (obtained via a receding "long"-horizon controller, under a "fine" system discretization).

10.6 Exercises

1. In principle, we could say that $\mathcal{S}$ is reachable from $\mathcal{P}$ in T steps if, for all $x \in \mathcal{S}$, there exist u and $x(0)$ in $\mathcal{P}$ such that $x(T) = \bar{x}$. If there are no uncertainties, we can consider an open-loop control u which is a function of the final state $x = x(T)$, namely $u(t) = \Phi(x(T), t)$ and this is equivalent to the controllability to $\mathcal{S}$ of $\mathcal{P}$ for the time-reversed system (if it is defined). Could you figure out which are the difficulties in properly defining worst-case reachability if there are uncertainties? (Note that open-loop and closed-loop definitions are not equivalent anymore ...).
2. Prove by means of a simple example that the 0-reaching minimum-time problem for an uncertain discrete-time system has no solution even if $E = 0$, since the origin cannot be reached in a finite number of steps.
3. Find an example of a control problem for which $T_R(x_0) > T_U(x_0)$ where $T_R(x_0)$ and $T_U(x_0)$ are defined in Subsection 10.1.3.
4. Find the values of $\beta > 0$ in Example 10.19 for which the described "escape" phenomenon occurs. Find the values of $\beta > 0$ for which it does not.
5. Find a sequence of values of $\beta > 0$ in the Example 10.19 for which the receding horizon scheme converges but the global (infinite-time) cost of the corresponding receding horizon schemes becomes arbitrary high, and therefore arbitrarily far from the infinite-time optimal one.
6. Show that $\Omega(x)$ in Eq. (10.2) is a polyhedral set if the constraints are polyhedral. What about the case of an uncertain A?
7. Prove that a convex positive-definite function is radially unbounded as claimed at the beginning of Section 10.3.
8. Show that control (10.63) is relatively optimal and dead-beat if $n = N$.
9. In the proof of Th. 10.34 it is assumed that the last non-zero vectors of the trajectory are linearly independent. Prove that this condition is assured if x_n does not belong to a proper (A, B)-invariant subspace.
10. Prove that a for a system of the form $\dot{x} = Bu$, $u \in \mathcal{U}$, $\mathcal{U}$ a C-set, the cost-to-go minimum-time function to the origin is $T_{min}(x) = \Phi_{\mathcal{V}}(x)$, the Minkowski function of $\mathcal{V} = -B\mathcal{U}$.
11. If we modify the system of the previous exercise as $\dot{x} = Bu - d$, $d \in \mathcal{D}$, $\mathcal{D}$ a C-set, the "worst case" cost-to-go function (under state feedback) is $\Phi_{\tilde{\mathcal{V}}(x)}$, where $\tilde{\mathcal{V}}(x)$ is the erosion of $\mathcal{V}$ with respect to $\mathcal{D}$. This is a little bit harder to show ... (see [MS82] and [BMR04]).

Chapter 11
Set-theoretic estimation

This chapter is mainly devoted to evidencing difficulties of the set-theoretic approach in dealing with output feedback problems. Therefore the chapter (apparently the shortest of the book) will be essentially conceptual rather than technical. We will be interested in set-theoretic state estimation, and we will not consider, if not in passing, the problem of set-theoretic system identification. Roughly, given an uncertain system with output measurements, a set-theoretic estimator provides a set which includes at each time all the state vectors which are compatible with available measurements and given disturbance specification.

It would be unfair to claim that set-theoretic techniques for state estimation are not well developed, it is rather the opposite. In particular as far as the worst case techniques for estimation are concerned, there are many valid contributions which started in the 70s' papers [Wit68b, BR71a, BR71b, GS71, Sch73]. From a practical standpoint, set-theoretic techniques for state estimation suffer from computational problems that are worse than those for set-theoretic control. We try to concisely express these difficulties in two points.

- In set-theoretic-control non-conservative solutions are known. The resulting state feedback compensators can have arbitrarily high complexity. However, the complexity of the compensator is fixed at the end of the (off-line) design stage.
- In set-theoretic-estimation non-conservative solutions[1] for the evaluation of the state not only may have arbitrarily high complexity, but the complexity increases on-line as more output data are collected.

We claim to be unable to overcome this difficulty at this moment, and apparently many excellent researchers failed as well. Our contribution here is to try our best to enlighten the problem, hoping for a solution in the near future.

[1] We will provide definitions to formalize the concept of non-conservativeness for a state-estimator.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_11

11.1 Worst case estimation

Let us formulate the problem in a general framework. Here we mainly consider discrete-time systems, although the concepts we present apply to continuous-time as well. Consider the system

$$x(t+1) = f(x(t), u(t), d(t)) \tag{11.1}$$

$$y(t) = g(x(t), w(t)) \tag{11.2}$$

where we assume $d(t) \in \mathcal{D}$ and $w(t) \in \mathcal{W}$, with $\mathcal{D}$ and $\mathcal{W}$ assigned C-sets. The control $u(t) \in \mathcal{U}$ is always assumed to be known in the present context (possibly determined by a suitable control algorithm).

Together with the information given by equations (11.1)–(11.2) we assume to have a further a priori information regarding the initial state, precisely

$$x(0) \in \hat{\mathcal{X}}_0. \tag{11.3}$$

Note that in the complete blindness we can just write $\hat{\mathcal{X}}_0 = \mathbb{R}^n$. Now we introduce two operators from set to set. Given a set $\mathcal{X} \subseteq \mathbb{R}^n$, define the set of all reachable states in one step for all possible $d(t) \in \mathcal{D}$ given the control action $u(t) \in \mathcal{U}$, according to equation (11.1)

$$Reach[\mathcal{X}, \mathcal{D}](u) \doteq \{z = f(x, u, d), \;\; x \in \mathcal{X}, \;\; d \in \mathcal{D}\}. \tag{11.4}$$

Given a guess about the state in t, the previous equation propagates this information at $t+1$. Clearly, in this way the information about the actual state spreads. However by means of (11.2) we can cut out a portion of the new guess region that is not compatible with measurements. Let us introduce the set of all the states compatible with measurements as

$$Comp[\mathcal{W}](y) \doteq \{x : g(x, w) = y, \;\; \text{for some} \;\; w \in \mathcal{W}\} \tag{11.5}$$

By means of (11.4) and (11.5) we can describe, theoretically, the estimation set. Let us formally define the concept.

Definition 11.1 (Estimation region). The set $\hat{\mathcal{X}}(t)$ is an estimation region (set) at time t given the information (11.1), (11.2), and (11.3) over a prescribed horizon $0, 1, \ldots, t$, if the condition

$$x(t) \in \hat{\mathcal{X}}(t)$$

is assured for all $w \in \mathcal{W}$ and $d \in \mathcal{D}$.

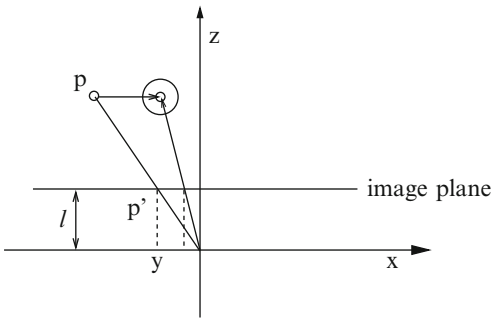


Fig. 11.1 The estimation problem

Definition 11.2 (Non-conservative estimation region). The estimation region $\hat{\mathcal{X}}(t)$ is non-conservative (with respect to the available information) if there does not exist a proper subset of it which is also an estimation region.

Before starting with an algorithm, some examples of the previous concept are reported.

Example 11.3. A position detection problem. Let us study the problem of detecting the position from an image (see Fig. 11.1). We consider for brevity a 2-dimensional problem in which a point P of coordinates (x, z) is moving in a plane. The only information available is its image, which is a point P' on the line which is at a distance l from the observation point $(0, 0)$. Thus, the output is given by

$$y(t) = l\frac{x(t)}{z(t)} + w(t), \tag{11.6}$$

where $|w| \leq \bar{w}$ is the error on the image plane (typically due to pixel quantization). We assume $z \geq l$ to avoid unnecessary complications.

The system position cannot be detected by this equation, not even for $\bar{w} = 0$, if we do not assume some kind of motion. Indeed, the best we could say (for $w = 0$) is that the point is on a line $yz = lx$, given the current measured y. Let us consider the following equations for the motion:

$$\begin{aligned} x(t+1) &= x(t) + u(t) + d_x(t) \\ z(t+1) &= z(t) + d_z(t) \end{aligned} \tag{11.7}$$

where the vector $d = (d_x, d_z)$ represents the proper (unknown) motion of the object and $u(t)$ represents a controlled horizontal motion. We assume the bound

$$\|d\| = \sqrt{d_x^2 + d_z^2} \leq \bar{d},$$

which means that the object has a limited speed: precisely, at each sampling time $\bar{d}$ is the maximum translation, that can occur in any direction.

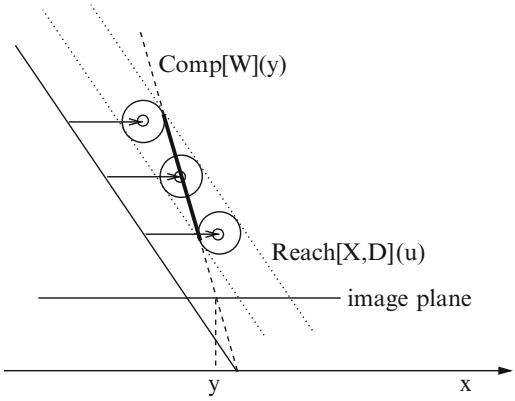


Fig. 11.2 The estimation set without output noise

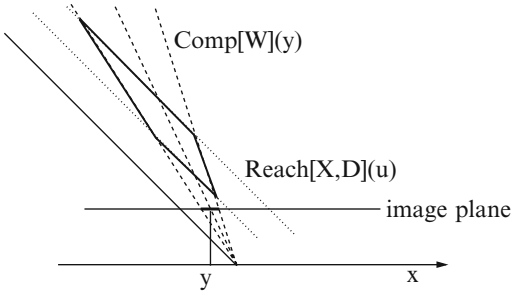


Fig. 11.3 The estimation set with output noise

To give an explanation to the input u (not to make this example seem too fuzzy), we are not thinking of a controlled motion of the object but rather of a controlled horizontal motion of the observer (typically a camera or videocamera). Indeed only a relative motion (without stereoscopy) can lead to the detection of the position with finite error. Assume that the initial position is unknown up to the fact that the object is on the ray starting from the viewpoint. Let us denote this ray by $\mathcal{X}$.

Given the horizontal translation u it is then known that, at the next step, if $d = 0$ the object will be on a line that is nothing but the set $\mathcal{X}$ translated by u, say $\mathcal{X} + u$. Since $\|d\| \leq \bar{d}$ also comes into play, the actual set is the translated line to which we have to sum the $\mathcal{D}$ circle (thus, we end up with $\mathcal{X} + u + \mathcal{D}$). We achieve in this way the set $Reach[\mathcal{X}, \mathcal{D}](u)$ which is the dotted strip represented in Fig. 11.2. If we assume, for the moment being, that $\bar{w} = 0$, i.e., there are no direct errors on the measurement, it is apparent that the new observation y identifies a new line, which is the set $Comp[\mathcal{W}](y)$. The intersection of such a line with the strip produces the set of possible states that are compatible with the new measurement (the thick segment in Fig. 11.2).

If we consider also the measurement error, we do not have a single segment, but a figure which is generated by the union of all the segments generated by lines that are compatible with the current observation (see Fig. 11.3). The set $Comp[\mathcal{W}](y)$ is then the sector represented by the dashed lines. The intersection with the strip gives a more complex figure (pictured in thick lines) that represents the set of all possible states at time $t = 1$ which are consistent with the given information.

The set we have computed in the example has a special meaning. It represents the set of all possible states at time $t = 1$ that are compatible with the measurements at steps $t = 0, 1$. This is a *worst case* type of approach, since it is basically based on the pessimistic idea that hiding the true object position is the main purpose of the uncertain inputs w and d (thought as "clever" agents). It is apparent that this kind of analysis naturally leads to the concept of *set-membership estimation*.

In general, we say that a set-membership estimator is a mechanism (algorithm) that provides the estimation region $\hat{\mathcal{X}}(t)$ as in Definition 11.1. The basic idea to compute one of such sets which is non-conservative is essentially that of iterating *forward in time* the procedure sketched in the example. We formally introduce this procedure. For the sake of generality[2], it is assumed that the system observation starts at time k, and that the following a priori information is available

$$x(k) \in \hat{\mathcal{X}}(k)$$

As a trivial (extreme) case one can just set $\hat{\mathcal{X}}(k) = \mathbb{R}^n$. Other possible extreme choices are those corresponding to huge portions of the state space as we did, in some sense, in the example (where $z \geq l$ as a primary information was assumed), or to a singleton $\hat{\mathcal{X}}(k) = x_0$. At this point, one computes the set of all the states compatible with the measurements. Once this set is found, the next step is to propagate it according to the system equation. Among all the states in the new set, one takes those compatible with the new measurements and so on. The procedure is formalized next. We denote by $\hat{\mathcal{X}}(t|k)$ the set of all states at time t that are consistent with the system equation from time k up to t.

Procedure: Non-conservative state-membership estimation.

For the dynamic system described by (11.1) (11.2), given the initial estimation $\hat{\mathcal{X}}(k)$, let $\hat{\mathcal{X}}(k|k)$ be the set of estimated states at step k which are output compatible, say

$$\hat{\mathcal{X}}(k|k) = \hat{\mathcal{X}}(k) \bigcap Comp[\mathcal{W}](y(k)).$$

Set $t = k$ and perform the following steps.

1. Given the current value of the control $u(t)$, propagate this set forward

$$\hat{\mathcal{Z}}(t+1|k) := Reach[\hat{\mathcal{X}}(t|k), \mathcal{D}](u(t)).$$

2. Compute the new estimation set as the intersection with the compatible set

$$\hat{\mathcal{X}}(t+1|k) := \hat{\mathcal{Z}}(t+1|k) \bigcap Comp[\mathcal{W}](y(t+1)).$$

3. Set $t := t + 1$ and go to Step 1.

[2] A waste of generality is typical when non-constructive results are provided.

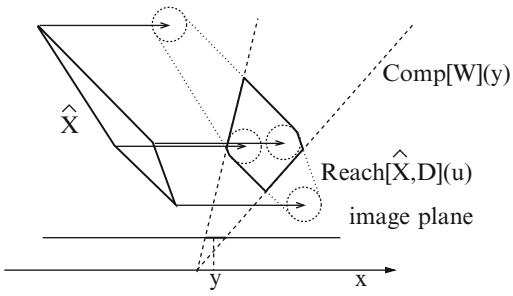


Fig. 11.4 A further iteration for the example

Easy to state, hard to compute[3]. The difficulty of iterating the method is apparent. To show it in a simple way, we reconsider the first example and we allow a further iteration. Consider the last computed set $\hat{\mathcal{X}}$ and assume that a new value of the control u is given (its action is represented by the arcs in Fig. 11.4). If this action is combined with d, we derive the figure which is delimited by both curve and straight lines that are represented via dotted lines. If we intersect with the cone of output compatible states we derive a figure which is delimited, again, by straight lines and circle arcs. This case is quite a lucky one, since it is apparent that all the future sets will be delimited by lines and arcs and will be therefore reasonably simple. However, in general not even convexity and connection of the set $\hat{\mathcal{X}}(t+1|k)$ are assured since, as it has been evidenced in Section 6.1.2 in Chapter 6, the propagation of uncertainty in general results in non-convex sets even in the LPV case.

We try now to compare the set-theoretic estimation procedure with that used to generate the largest (contractive) invariant set inside a domain for the derivation of state feedback control laws. As expected, there are well-established connections in the literature [Wit68b, BR71a, BR71b, GS71, Sch73] and actually many aspects are in some sense dual. Although we avoided acronyms in most of the book, we make an exception here and we refer to SEP to mean State Estimation Procedure and to CIP, Controlled-Invariant (generation) Procedure.

- The SEP proceeds forward in time while the CIP proceeds backward in time. This is a well-known circumstance in control theory. For instance, exactly the same dichotomy holds for the optimal control and the optimal state estimation. For instance, the Riccati equation for optimal control is integrated backward in time while the Riccati equation for the Kalman–Bucy filter is integrated forward in time [KS72].
- Both SEP and CIP provide objects that may have arbitrarily large complexity of description.
- The SEP has to be performed on-line while the CIP is performed off-line. This is in practice a substantial difference, since once the CIP has stopped,

[3]The reader should be familiar at this point with this feeling.

one can analyze the control and check if it is implementable with the available hardware (and decide if its implementation is worthwhile). Conversely, the SEP is performed on-line and leads to a system crash if the complexity of $\mathcal{X}(t+1|k)$ reaches the computer limit.

A further fact worth pointing out is that the set-theoretic procedure is strongly dependent on the choice of the control action u. This is apparent in the example we considered. If one leaves the controlled input $u(t) = 0$, then no state estimation is possible (if we do not have information on the size), since it is possible to detect the direction of the object but not the distance. It can be easily understood that the set-membership estimation, for $u(t) = 0$, would be formed by the cone of all output-compatible states. The motion plays a fundamental role in the identification. Now, if one considers a control problem, for instance the alignment with the object, in which one wants to track a relative object position $\bar{x}$, then the situation is messy. Assume that at a certain point the condition $x(t) = \bar{x}$ holds. Then the controller has to perform some actions anyway *only for the purpose of estimating* the position. We will come back to this control-estimation interaction soon.

11.1.1 Set membership estimation for linear systems with linear constraints

The previous procedure can be actually implemented via linear programming in the case of linear systems with linear constraints

$$x(t+1) = Ax(t) + Bu(t) + Ed(t) \tag{11.8}$$

$$y(t) = Cx(t) + w(t) \tag{11.9}$$

when $\mathcal{D}$ and $\mathcal{W}$ are polytopes. For convenience we adopt the vertex and plane representation, respectively.

$$\mathcal{D} = \mathcal{V}[D] \quad \mathcal{W} = \mathcal{P}[W, q]$$

It is a well-assessed fact that the solution of system (11.8) can be split in to two terms

$$x(t) = x_u(t) + x_d(t)$$

which are the solutions, respectively, of

$$x_u(t+1) = Ax_u(t) + Bu(t), \quad x_u(t_0) = 0,$$
$$x_d(t+1) = Ax_d(t) + Ed(t), \quad x_d(t_0) = x(t_0),$$

therefore, since u is known, only $x_d(t)$ needs to be estimated. Henceforth, without restriction, it is possible to consider only the problem of estimating the state of the next system

$$x(t+1) = Ax(t) + Ed(t) \tag{11.10}$$

$$y(t) = Cx(t) + w(t) \tag{11.11}$$

Given the polyhedron $\hat{\mathcal{X}}$ in its vertex representation $\hat{\mathcal{X}} = \mathcal{V}[X]$, the one-step forward propagation is a polytope

$$\mathcal{Z} = Reach[\hat{\mathcal{X}}, \mathcal{D}] = \mathcal{V}[Z] = A\hat{\mathcal{X}} + E\mathcal{D}$$

(note that the dependence on u was eliminated due to the fact that the control is being considered separately) where Z is the matrix whose columns are generated as follows

$$Z_k = AX_i + ED_j, \tag{11.12}$$

by combining the columns of X and of D in all possible ways. It is clear that, in performing this operation, eliminating the redundant vertices is fundamental. The output admissible state set, given the measure y, has the representation

$$Comp[\mathcal{W}](y) = \{x : W(y - Cx) \leq q\}.$$

Let us now consider a plane representation for $\mathcal{Z} = \mathcal{P}[S, r]$. Then the intersection of $\mathcal{Z}$ with the output-admissible sets provides the updated state estimator

$$\hat{\mathcal{X}}_{new} = \mathcal{P}[S, r] \bigcap Comp[\mathcal{W}](y) = \{x : \ \ Sx \leq r, \text{ and } -WCx \leq q - Wy\},$$

or, in a compact form,

$$\mathcal{X}_{new} = \mathcal{P}\left[\begin{bmatrix} S \\ -WC \end{bmatrix}, \begin{bmatrix} r \\ q - Wy \end{bmatrix}\right].$$

Therefore the state estimation region for linear systems can be described by polytopes. Still there is a major problem. The complexity of the set-membership estimation in general increases at each step and can consequently grow arbitrarily.

A different way of proceeding is the so-called batch approach in which all the measurements from time k to time t are simultaneously processed. This works as follows. For brevity consider the system

$$x(i+1) = Ax(i) + d(i), \ \ i = k, k+1, \ldots, t,$$

where

$$d \in \mathcal{P}[D, e], \tag{11.13}$$

and the output measurements

$$y(i) = Cx(i) + w(i), \quad i = k, k+1, \ldots, t,$$

with

$$w \in \mathcal{P}[W, q]. \tag{11.14}$$

Then write

$$\begin{bmatrix} I & -A & 0 & 0 & \ldots & 0 & 0 \\ 0 & I & -A & 0 & \ldots & 0 & 0 \\ 0 & 0 & I & -A & \ldots & 0 & 0 \\ : & : & : & : & \ldots & -A & 0 \\ 0 & 0 & 0 & 0 & \ldots & I & -A \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-1) \\ x(t-2) \\ : \\ x(k) \end{bmatrix} = \begin{bmatrix} d(t-1) \\ d(t-2) \\ d(t-3) \\ : \\ d(k) \end{bmatrix} \tag{11.15}$$

and

$$\begin{bmatrix} I & 0 \ldots & 0 & -C & 0 & 0 & 0 \\ 0 & I \ldots & 0 & 0 & -C & 0 & 0 \\ : & : \ldots & : & : & \ldots & : & : \\ 0 & 0 \ldots & I & 0 & \ldots & 0 & -C \end{bmatrix} \begin{bmatrix} y(t) \\ y(t-1) \\ : \\ y(k) \\ x(t) \\ x(t-1) \\ : \\ x(k) \end{bmatrix} = \begin{bmatrix} w(t) \\ w(t-1) \\ : \\ w(k) \end{bmatrix} \tag{11.16}$$

Since $d(i)$ and $w(i)$ are such that $Dd(i) \leq e$ and $Ww(i) \leq q$, equations (11.15) and (11.16) become

$$\begin{bmatrix} D & -DA & 0 & 0 & \ldots & 0 & 0 \\ 0 & D & DA & 0 & \ldots & 0 & 0 \\ 0 & 0 & D & -DA & \ldots & 0 & 0 \\ : & : & : & - : & \ldots & -DA & 0 \\ 0 & 0 & 0 & 0 & \ldots & D & -DA \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-1) \\ x(t-2) \\ : \\ x(k) \end{bmatrix} \leq \begin{bmatrix} e \\ e \\ e \\ : \\ e \end{bmatrix}$$

$$\begin{bmatrix} -WC & 0 & 0 & 0 \\ 0 & -WC & 0 & 0 \\ : & : & \ldots & : \\ 0 & 0 & \ldots & -WC \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-1) \\ : \\ x(k) \end{bmatrix} \leq \begin{bmatrix} q \\ q \\ : \\ q \end{bmatrix} - \begin{bmatrix} W & 0 & 0 & 0 \\ 0 & W & 0 & 0 \\ : & \ldots & : & : \\ 0 & \ldots & 0 & W \end{bmatrix} \begin{bmatrix} y(t) \\ y(t-1) \\ : \\ y(k) \end{bmatrix}$$

The above inequalities identify a polyhedron in the space of the sequences

$$\bar{x}(t,k) = [x(k)\ x(k+1)\ \dots\ x(t)],$$

which is the set of all admissible states given the measured output sequence

$$\bar{y}(t,k) = [y(k)\ y(k+1)\ \dots\ y(t)].$$

This polyhedron is uniquely identified once $\bar{y}(t,k)$ is given and so we denote it as

$$\Xi(\bar{y}(t,k)). \tag{11.17}$$

Now consider the "projection operator" defined as

$$Pr\,(\bar{x}(t,k)) = x(t)$$

which selects the last element of the sequence. Then the set of all the states compatible with the measurements is

$$\hat{\mathcal{X}}(t|k) = Pr\,(\Xi(\bar{y}(t,k)))\,,$$

The difference with the recursive procedure previously described is, basically, that the former projects at each step to achieve the intermediate set $\hat{\mathcal{X}}(i|k)$, $i = k+1, k+2, \dots, t$, while the latter identifies the set $\hat{\mathcal{X}}(t|k)$, by means of a single (huge) projection operation.

Example 11.4. Consider the system

$$\begin{aligned} x(t+1) &= Ax(t) + Ed(t) \\ y(t) &= Cx(t) + w(t) \end{aligned}$$

where

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1\ 0 \end{bmatrix}.$$

Assume that $|d(t)| \le \bar{d} = 1$ and $|w(t)| \le \bar{w} = 1$ for all t.

For simulation purposes, we initialize the system at $x(0) = [1\ 0]^T$ and we choose the sequences d and w as $d(0) = -1$, $d(1) = 0$, $d(2) = 0$ and $w(0) = 0$, $w(1) = 0$, $w(2) = 0$, respectively. This information is clearly unknown to the estimation algorithm. We indeed assume that, as a priori information, one knows that $\|x(0)\|_\infty \le 4$, precisely that the state is in the square depicted in Fig. 11.5, which is the given $\hat{\mathcal{X}}(0)$. Let us see how this information is propagated. The true state is represented by the thick dot. For pictorial purposes we have represented: the set $A\hat{\mathcal{X}}$ by dotted lines, the set $A\hat{\mathcal{X}} + E\mathcal{D}$ by dashed lines, the output admissible

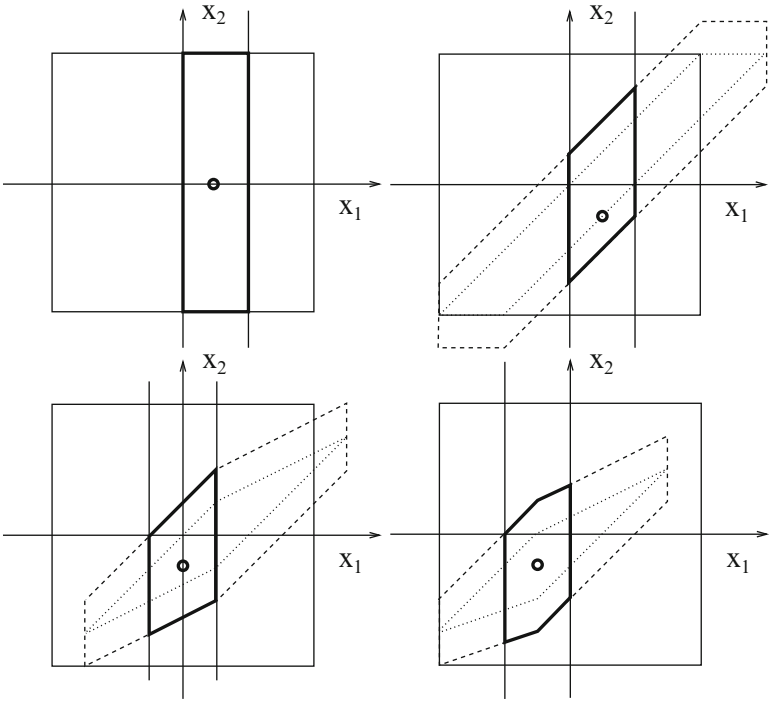


Fig. 11.5 The sequence of non-conservative estimation regions

set $Comp[w](y) = \{x : |Cx - y| \leq \bar{w}\}$ by plain lines and the actual estimation set by a thick plain line. This set will be denoted by $\mathcal{V}[X^{(k)}]$, where $X^{(k)}$ is the vertex matrix.

At time $t = 0$, $y(0) = Cx + w(0) = 1$, that is to say $|Cx - 1| = |w(0)|$. Since $w(0)$ is unknown we can only deduce that the system is in the intersection of the square and the strip $|Cx - 1| \leq 1$, see Fig. 11.5, left-top. This set $\hat{\mathcal{X}}(0|0)$ is a rectangle $\mathcal{V}[X^{(0)}]$ and it has vertex matrix

$$X^{(0)} = \begin{bmatrix} 0 & 2 & 2 & 0 \\ -4 & -4 & 4 & 4 \end{bmatrix}$$

Now we compute the set $A\hat{\mathcal{X}}(0|0)$ depicted by a dotted line in Fig. 11.5, right-top. In the same figure we have represented by dashed line the set $A\hat{\mathcal{X}}(0|0) + E\mathcal{D}$. We computed the intersection with the admissibility set given by the output $y(1) = 1$ corresponding to the new state $x(0) = [1 \ -1]^T$ and the value $w(1) = 0$ (unknown to the estimation algorithm). Such and output-admissible strip is $|Cx - 1| \leq 1$. Finally, the thick line represents the new estimation set $\hat{\mathcal{X}}(1|0)$ having vertex matrix

$$X^{(1)} = \begin{bmatrix} 0 & 2 & 2 & 0 \\ -3 & -1 & 3 & 1 \end{bmatrix}$$

A further iteration provides the new state $x(2) = [0\ -1]^T$ and output $y(2) = Cx(2) + w(2) = 0$. The new set $\hat{\mathcal{X}}(2|0)$, computed as previously shown, is characterized by the vertex matrix

$$X^{(2)} = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 0 & -3 & -2 & 2 \end{bmatrix}.$$

At time $t = 3$ the true state is $x(2) = [-1\ \ -1]^T$, the output $y(3) = -1$. The new set $\hat{\mathcal{X}}(3|0)$ has the following vertex matrix.

$$X^{(3)} = \begin{bmatrix} -2 & -1 & -1 & 0 & 0 & -1 \\ 0 & -3.33 & -3 & -2 & 1.5 & 1 \end{bmatrix}$$

Note that the latest set has more vertices than the previous.

To explain the troubles associated with this kind of set-membership estimators, we computed the estimation set sequence for a different initial condition and disturbance sequence. Precisely, we assumed that $x(0) = [3\ 2]^T$ and that the first elements of sequences d and w are $d(0) = -1$, $d(1) = -1$, $d(2) = 0$ and $w(0) = -1$, $w(2) = 0$, $w(3) = 0.5$, respectively. The corresponding true state sequence is $x(1) = [5\ 1]^T$, $x(2) = [6\ 0]^T$, $x(3) = [6\ 0]^T$, the output sequence is $y(0) = Cx(0) + w(0) = 3$, $y(1) = 4$, $y(2) = 6$ and $y(3) = 6.5$ The corresponding set sequence is depicted in Fig. 11.6 and is represented by the following matrices. The first set is $\hat{\mathcal{X}}(0|0) = \mathcal{V}[X^{(0)}]$, where

$$X^{(0)} = \begin{bmatrix} 2 & 4 & 4 & 2 \\ -4 & -4 & 4 & 4 \end{bmatrix}.$$

At time $t = 1$ the set is $\hat{\mathcal{X}}(1|0) = \mathcal{V}[X^{(1)}]$, where

$$X^{(1)} = \begin{bmatrix} 3 & 5 & 5 & 3 \\ -2 & 0 & 4 & 2 \end{bmatrix}.$$

At time $t = 2$ the set is $\hat{\mathcal{X}}(2|0) = \mathcal{V}[X^{(2)}]$ where

$$X^{(2)} = \begin{bmatrix} 5 & 7 & 7 & 5 \\ -1 & 1 & 4 & 3 \end{bmatrix}.$$

Finally, at $t = 3$ the set is $\hat{\mathcal{X}}(3|0) = \mathcal{V}[X^{(3)}]$ where

$$X^{(3)} = \begin{bmatrix} 5.50 & 7.50 & 7.5 & 7 & 5.5 \\ -1.25 & -0.25 & 4.125 & 4 & 2.0 \end{bmatrix}.$$

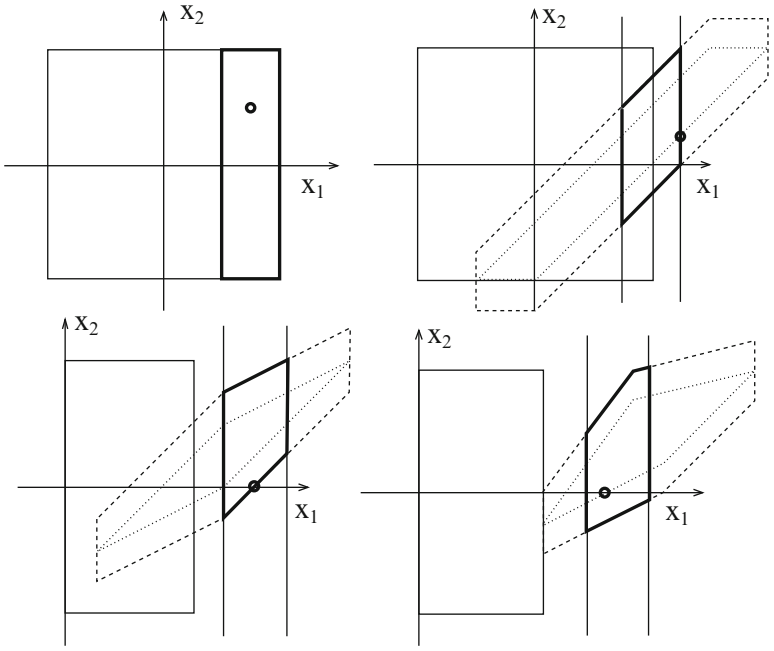


Fig. 11.6 The new sequence of non-conservative estimation regions

It is apparent that not only the positions but also the shapes of the new sets are different from those computed before.

We point out the following property, which is a quite obvious consequence of what has been said up to now.

Proposition 11.5. *Given the initial information set $\hat{\mathcal{X}}(k)$ on the state $x(k)$, and the corresponding output-admissible subset $\hat{\mathcal{X}}(k|k)$ assumed as the initial condition, $\hat{\mathcal{X}}(t|k)$ represents the set of all the possible states and it is the exact estimation set (consistent with the information from k to t). In other words, any element $x \in \hat{\mathcal{X}}(t|k)$ can be actually achieved, namely $x(i) = x$, by means of admissible sequences $w(i) \in \mathcal{W}$ and $d(i) \in \mathcal{D}$, starting from a proper $x(k) \in \hat{\mathcal{X}}(k|k)$, with a transient which produces the observed output. Conversely, any value $x \notin \hat{\mathcal{X}}(t|k)$ cannot be reached without a constraint violation.*

Note also the following point. The method is based on precise assumptions on

i) the model: the matrices A, E, C (and possibly B) are known exactly;
ii) the disturbances bound: the sets $\mathcal{W}$ and $\mathcal{D}$ are known exactly;
iii) the initial guess: the set $\hat{\mathcal{X}}(k)$ is known exactly.

From an ideal point of view, as long as the assumptions on the disturbances $w(t) \in \mathcal{W}$ and $d(t) \in \mathcal{D}$ and the guess on the initial state $\hat{\mathcal{X}}(k|k)$ are correct, the corresponding sequence $\hat{\mathcal{X}}(t|k)$ must be non-empty. In its practical application, this kind of set-theoretic estimation has the following limitations.

- it might be possible that $\hat{\mathcal{X}}(t|k) = \emptyset$. This means that either the assumptions on the disturbances are not verified or the model is non-correct.
- It might be possible that the true state $x(t) \notin \hat{\mathcal{X}}(t|k)$. This also means that either the assumptions on the disturbances are not verified or the model is non-correct.

If one of the previous shortcomings occurs, then the only way to overcome it is to enlarge the uncertainty bound. It is clear that for $\mathcal{W}$ and $\mathcal{D}$ large enough the problem becomes necessarily consistent. In particular, uncertainty in the model, i.e. on (A, C, E, B), can be compensated by augmenting the bounds on the uncertainties, either additive or in the measurements (or both).

Once the set of admissible states is evaluated, the consequent problem consists in selecting an estimated value. This is fundamental if this set-theoretic state estimation has to be used for control. This is nothing else than a selection problem. Precisely a proper estimation procedure requires computing $\hat{x}(t)$ as a selection

$$\hat{x}(t) = \Phi_{est}(y(t), y(t-1), \ldots, y(k), \hat{\mathcal{X}}(k)) \in \mathcal{X}(t|k) \tag{11.18}$$

where Φ_{est} is a proper map with values in $\mathcal{X}(t|k)$. Since any x is possible, any selection Φ_{est} is a possible solution. There are several ways to select a proper value, normally based on some optimization criteria, for instance:

- $\hat{x}$ is taken as the value of smallest norm;
- $\hat{x}$ is taken as the center of the ellipsoid of greatest volume inside $\hat{\mathcal{X}}(t|k)$ (known as Chebyshev center when a ball instead of an ellipsoid is considered);
- $\hat{x}$ is taken as the center of the ellipsoid of smallest volume containing $\hat{\mathcal{X}}(t|k)$ (such an ellipsoid is known as Löwner-John ellipsoid);
- $\hat{x}$ is taken as the center of the smallest box including $\hat{\mathcal{X}}(t|k)$;
- $\hat{x}$ is taken as the solution of a min–max problem

$$\max_{\hat{x} \in \hat{\mathcal{X}}(t|k)} \min_{z \in \partial \hat{\mathcal{X}}(t|k)} \|x - z\|$$

Discussions about the choice of the center are proposed in [MT85] and [BV04]. However, the main issue remains unchanged: the complexity of the set $\hat{\mathcal{X}}(t|k)$ increases with time and it is often impossible to use this algorithm in most practical real-time applications. Some approximations are necessary and will be discussed next.

11.1.2 Approximate solutions

In this section we discuss some kind of approximate solutions to the problem. Possible ways to overcome the problem of arbitrary increasing of complexity of the exact solution (with respect to the information from k to t) are the following.

- Discharge part of the information when $\hat{\mathcal{X}}(t|k)$ becomes too complex. This eliminates some of the constraints and provides a set which is an over-bound of the true one.
- Consider an observer given by some optimality criteria (i.e., a Kalman–Bucy filter) and then evaluate the error bound a posteriori.

A possible way to follow the first idea is to choose a fixed horizon $[t-h,t]$ and compute at each time the set $\hat{\mathcal{X}}(t|t-h)$ which has finite complexity. This can be achieved by considering, for all t, the set $\Xi(\bar{y}(t-T,t))$ defined in (11.17), with $T>0$ fixed. For large T the projection $Pr(\Xi(\bar{y}(t-T,t)))$ defines an accurate estimation region.

We illustrate this method by means of an example, assuming for brevity $d=0$ (measurement errors only) and A invertible. Then we have

$$x(i)=A^{i-t}x(t),\quad i=t-k,\dots,t-1,t.$$

The measurements are

$$y(i)=CA^{i-t}x(t)-w(i).$$

Write the above equalities in a compact form

$$\underbrace{\begin{bmatrix} C \\ CA^{-1} \\ CA^{-2} \\ \vdots \\ CA^{-k} \end{bmatrix}}_{\Omega_{t,t-k}} x(t) - \underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ y(t-2) \\ \vdots \\ y(t-k) \end{bmatrix}}_{\bar{y}_{t,t-k}} = \underbrace{\begin{bmatrix} w(t) \\ w(t-1) \\ w(t-2) \\ \vdots \\ w(t-k) \end{bmatrix}}_{\bar{w}_{t,t-k}}$$

and write the standard regression model

$$\Omega_{t,t-k}x-\bar{y}_{t,t-k}=\bar{w}_{t,t-k}.$$

Then if we assume $w\in\mathcal{W}$, a polytope, the admissible set for $x(t)$ is also a polytope. If, to keep the exposition simple, we assume $\mathcal{W}$ as the ω-ball of the ∞-norm, so that $|w_i|\le\omega$, then

$$\hat{\mathcal{X}}_\omega(t|t-k)=\{x:\|\Omega_{t,t-k}x-\bar{y}_{t,t-k}\|_\infty\le\omega\}.$$

Note that we can deal with the problem in an optimization setup by minimizing the uncertain size ω, precisely by computing

$$\min \ \{\omega \geq 0 \text{ s.t. } \hat{\mathcal{X}}_\omega(t|t-k) \text{ is non-empty}\}.$$

To use this scheme in practice, in particular to find an estimated value $\hat{x}(t)$ at each time, when the new output $y(t+1)$ is available, *all the available data must be re-processed*. There is no known method to process these data in a recursive way.

In the case of the 2-norm, when $\hat{x}(t)$ is taken as the minimizer

$$\hat{x}(t) = \arg\min_x \ \|\Omega_{t,k-t}x - \bar{y}_{t,k-t}\|_2,$$

the least-square minimizer is

$$\hat{x}(t) = \left[\Omega_{t,k-t}^T \Omega_{t,k-t}\right]^{-1} \Omega_{t,k-t}^T y_{t,k}.$$

It also well known that the above formula admits a recursive implementation and the reader is referred to specialized literature (e.g., [Lju99]).

Coming back to the case of ∞ norms (in general the polytopic case), we have to deal with the problem of selecting an element inside a polytope. Unfortunately, for fixed k, at each t the problem increases its complexity. However, if we fix the difference $h = t-k$, then the algorithm has a fixed complexity, but this is clearly an approximation since we basically *keep the information up to* $t-k$.

A more efficient method is to discharge, among the constraints defining the feasible set, those which appear the "most useless" according to some optimality criterion, which is not that of removing the oldest ones. For instance, one possible way is to discharge constraints in order to achieve the minimum volume among all the possible over-bounding sets [VZ96].

The second mentioned approach, precisely that of designing an observer to evaluate the error bound a posteriori is more practical. Consider the system

$$x(t+1) = Ax(t) + Bu(t) + Ed(t), \quad y(t) = Cx(t) + w(t)$$

and consider the observer

$$\hat{x}(t+1) = (A - LC)\hat{x}(t) + Bu(t) + Ly(t), .$$

Then the observation error is

$$e(t) \doteq \hat{x}(t) - x(t)$$

subject to the equation

$$e(t+1) = (A - LC)e(t) - Ed(t) + Lw(t) \tag{11.19}$$

Under standard detectability assumptions, $(A - LC)$ can be rendered stable. However, the error size bound can be quite different depending on the choice of matrix L.

One possible way to determine an error bound is to compute an invariant set $\mathcal{E}$ for system (11.19). In this case

$$e(0) \in \mathcal{E} \Rightarrow e(t) \in \mathcal{E}$$

If $A - LC$ is stable, $e(t) \to \mathcal{E}$ and $\mathcal{E}$ can be assumed as the ultimate observer error.

However the condition $e(0) \in \mathcal{E}$ is not necessarily assured. If we are granted the a priori information

$$e(0) \in \mathcal{E}_0,$$

where $\mathcal{E}_0$ is not necessarily invariant, a viable solution is that of computing the smallest invariant set including $\mathcal{E}_0$ as follows. In view of the results in Chapter 6, in particular in Subsections 6.1.2 and 6.2.1, we can propagate the set $\mathcal{E}_0$, computing the reachable sets $\mathcal{E}_k$ with bounded inputs. If in a finite time we have

$$\mathcal{E}_{\bar{k}} \subset \mathcal{E}_0,$$

then the convex hull of the union of these sets

$$conv \left\{ \bigcup_{k=0}^{\bar{k}} \mathcal{E}_k \right\}$$

is positively invariant (see [BMM95]). This means that we can provide an over-bound for the error.

An easier problem is to see if we can assure an ultimate observer bound. Assume that we wish to assure the bound

$$\|e(t)\|_\infty \leq \mu, \quad \text{for } t \geq T, \tag{11.20}$$

where T clearly depends on the initial condition. The following proposition holds.

Proposition 11.6. *Assume that $A - LC$ is asymptotically stable. Denote by μ_{inf} the infimum of the values for which (11.20) holds for some T. Then μ_{inf} is the smallest value for which there exists an invariant set for (11.19) inside the μ-ball of the $\|\cdot\|_\infty$ norm.*

The proof of this proposition is easy and is left to the reader.

In a set-theoretic framework it is very important to reduce the complexity of the estimation algorithm whenever possible. In the case of an observer it is possible, for instance, to resort to a reduced order observer. Consider the following standard formulation

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} A_{11}\ A_{12} \\ A_{21}\ A_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} d(t), \quad y(t) = x_1(t) + w(t).$$

Consider the variable $z(t) = x_2(t) - Lx_1(t)$. This variable satisfies

$$\begin{aligned} z(t+1) &= x_2(t+1) - Lx_1(t+1) = \\ &= A_{21}x_1(t) + A_{22}x_2(t) + E_2 d(t) - LA_{11}x_1(t) - LA_{12}x_2(t) - LE_1 d(t) \\ &= (A_{22} - LA_{12})z(t) + (A_{21} - LA_{11} + A_{22}L - LA_{12}L)x_1(t) \\ &+ (E_2 - LE_1)d(t) = (A_{22} - LA_{12})z(t) + P_L x_1(t) + Q_L d(t) \end{aligned}$$

where $Q_L = [E_2 - LE_1]$ and $P_L = A_{21} - LA_{11} + A_{22}L - LA_{12}L$. For this system we consider the observer

$$\hat{z}(t+1) = (A_{22} - LA_{12})\hat{z}(t) + P_L y(t), \tag{11.21}$$

where $\hat{z}(t)$ is an estimation value for $z(t)$ and $\hat{x}_2(t) \doteq \hat{z}(t) + Lx_1(t)$ is an estimation value for $x_2(t)$. This leads to the error

$$e_2(t) = \hat{z}(t) - z(t) = \hat{x}_2(t) - x_2(t),$$

which satisfies the equation

$$e_2(t+1) = (A_{22} - LA_{12})e_2(t) - Q_L d(t) + P_L w(t),$$

Finding an estimation set for this system $\mathcal{E}_2$ is, in general, easier in view of the reduction of the state space dimension. Clearly, we have also to take into account the error on x_1. One reasonable possibility is to trust the available measure $\hat{x}_1 = y = x_1 + w$. This means that

$$e_1(t) = \hat{x}_1(t) - x_1(t) \in \mathcal{W}$$

Example 11.7. Consider the system of the previous example

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} 1\ 1 \\ 0\ 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} d(t), \quad y(t) = x_1(t) + w(t).$$

Again we assume $|w(t)| \leq \bar{w} = 1$ and $|d(t)| \leq \bar{d} = 1$. The estimation of x_1 is $\hat{x}_1(t) = y(t)$, with error

$$|e_1(t)| = |\hat{x}_1(t) - x_1(t)| \leq 1.$$

Let us now estimate x_2. Consider the variable $z(t) = x_2(t) - Lx_1(t)$. Then

$$\begin{aligned} z(t+1) &= [1-L]z(t) + [-L^2]x_1(t) + [1]d(t) \\ \hat{z}(t+1) &= [1-L]\hat{z}(t) + [-L^2]x_1(t) + [-L^2]w(t) \\ e_2(t+1) &= [1-L]e_2(t) + [-L^2]w(t) - d(t) \end{aligned}$$

This is a single dimensional problem very easy to solve. For instance, it is quite easy to see that (assuming asymptotic stability, i.e. $|1-L| < 1$) the smallest invariant interval for this system has extrema $\pm\bar{e}_2$, where

$$\bar{e}_2 = [1-L]\bar{e}_2 - L^2\bar{w} - \bar{d}$$

namely $\bar{e}_2 = [-L\bar{w} - \bar{d}/L]$. In our case ($\bar{w} = 1$ and $\bar{d} = 1$) $\bar{e}_2 = -[L + 1/L]$. This means that we can achieve a limit interval of the form

$$|e_2| \leq L + 1/L$$

We have derived in this way the asymptotic estimation error set, which is

$$\{(e_1, e_2) : |e_1| \leq 1, \text{ and } |e_2| \leq L + 1/L\}.$$

Note that the feasible set previously computed for this example is necessarily greater than this one, since the bound given here holds only asymptotically.

11.1.3 Bounding ellipsoids

We can easily exploit the results of Section 6.1.3 to describe the estimation error bounds via ellipsoid. Consider the error equation

$$\dot{e}(t) = (A - LC)e(t) - Ed(t) + Lw(t) = (A - LC)e(t) + E_{aug}d_{aug}(t)$$

where $E_{aug} = [-E \;\; L]$ and d_{aug} is an uncertain input bounded as

$$d_{aug} \in \mathcal{E}(0, G^{-1}, 1) = \{d_{aug} : d_{aug}^T G^{-1} d_{aug} \leq 1\}$$

According to the results in Section 6.1.3 [Sch73]

$$x(t) \in \mathcal{E}(Q^{-1}(t), 1) \tag{11.22}$$

and

$$\dot{Q}(t) = (A - LC)Q(t) + Q(t)(A - LC)^T + \beta(t)Q(t) + \beta(t)^{-1}E_{aug}GE_{aug}^T,$$

where $\beta(t)$ is an arbitrary positive function. If we assume for G a block-diagonal structure, $G = diag\{G_E, G_L\}$, we get

$$\dot{Q}(t) = (A - LC)Q(t) + Q(t)(A - LC)^T + \beta(t)Q(t) + \beta(t)^{-1}(EG_EE^T + LG_LL^T).$$

Assuming $\beta(t) \to \bar{\beta}$, the limit equation is

$$(A - LC)Q + Q(A - LC)^T + \bar{\beta}Q + \bar{\beta}^{-1}(EG_EE^T + LG_LL^T) = 0,$$

which provides an invariant ellipsoid $\mathcal{D}(Q)$ for the error.

11.1.4 Energy bounded disturbances

For the sake of justice, we consider here the continuous-time case (the chapter is almost completely devoted to discrete-time). The extension to the discrete-time case is simple. Consider now the case of energy-bounded disturbances, precisely, an integral constraint of the form

$$\int_0^t [w(\sigma)^T w(\sigma) + d(\sigma)^T d(\sigma)]d\sigma \leq 1,$$

for all t. Reconsider the (continuous-time) observer error equation

$$\dot{e}(t) = (A - LC)e(t) - Ed(t) + Lw(t) \tag{11.23}$$

The set $\mathcal{E}_\sqcup$ of all the states $e(t)$ reachable from $e(0) = 0$ with bounded energy can be written as follows:

$$\mathcal{E}_\sqcup = \left\{ e(t) = \int_0^t e^{(A-LC)\sigma} [-Ed(t-\sigma) + Lw(t-\sigma)]d\sigma \right\}$$

with w, d as above. Consider the support function representation. For any unit vector z, the corresponding support hyperplane of $\mathcal{E}_t$ is given by

$$\begin{aligned} z^T e &\leq \max_{d,w} z^T \int_0^t e^{(A-LC)\sigma} [-Ed(t-\sigma) + Lw(t-\sigma)]d\sigma \\ &= \left[\int_0^t z^T e^{(A-LC)\sigma}(EE^T + LL^T)e^{(A-LC)^T\sigma} z \, d\sigma \right]^{\frac{1}{2}}, \end{aligned}$$

where the maximum is derived by the same argument of Theorem 6.18. We conclude that

$$z^T e \leq \mu(t) \doteq \sqrt{z^T Q(t) z}$$

where $Q(t)$ satisfies equation (6.13) which now becomes (we assume $A-LC$ stable)

$$\dot{Q}(t) = (A-LC)Q(t) + Q(t)(A-LC)^T + EE^T + LL^T. \tag{11.24}$$

The equation can be manipulated as follows:

$$\begin{aligned}\dot{Q}(t) &= AQ(t) + Q(t)A^T + EE^T + LL^T - LCQ(t) - Q(t)C^TL^T \\ &= AQ(t) + Q(t)A^T + EE^T + LL^T - LCQ(t) - Q(t)C^TL^T \pm Q(t)C^TCQ(t) \\ &= AQ(t) + Q(t)A^T + EE^T - Q(t)C^TCQ(t) + \\ &+ (L - Q(t)C^T)(L - Q(t)C^T)^T\end{aligned}$$

Note that, without difficulties, we could consider a time-varying gain $L = L(t)$ and the expression would remain unchanged. For each unit vector z, let us define the function

$$\phi(z,t) = z^TQ(t)z = \mu^2(t).$$

Intuitively, since

$$\phi(z,t) = \phi(z,0) + \int_0^t \dot{\phi}(z,\sigma)d\sigma,$$

by minimizing the derivative we achieve the "smallest" ellipsoid. Such a derivative is

$$\dot{\phi}(z,t) = z^T\dot{Q}(t)z$$

and we can minimize it over $L(t)$ independently of z by considering the expression of $\dot{Q}$. Indeed we immediately derive

$$\dot{\phi}(z,t) = z^T[AQ + QA^T + EE^T - QC^TCQ]z + z^T[(L - QC^T)(L - QC^T)^T]z$$

(where (t) has been dropped to avoid burdening of notations). The minimum value is achieved when $L(t)$ zeroes the last non-negative term, precisely

$$L(t) = Q(t)C^T.$$

The resulting equation

$$\dot{Q}(t) = AQ(t) + Q(t)A^T + EE^T - Q(t)C^TCQ(t) \tag{11.25}$$

is well known in filtering[4] and is the Riccati equation This means that the solution of the Riccati equation provides the smallest ellipsoid which is reachable with energy-bounded disturbances. The asymptotic equation

$$AQ + QA + EE^T - QC^T CQ = 0, \tag{11.26}$$

along with the corresponding gain

$$L = QC^T,$$

provides the steady state solution along with the infinite-time ellipsoid.

11.2 Including observer errors in the control design

The problem of determining an error bound for the observer is fundamental to extend state feedback techniques to the case in which the state variables are not measured. Assume that the state is not known exactly but

$$\hat{x}(t) = x(t) + e(t),$$

where the error due to the adopted estimation algorithm is $e(t) \in \mathcal{E}$. Assume we are given the state equation

$$x(t+1) = Ax(t) + Bu(t) + Ed(t),$$

with $d(t) \in \mathcal{D}$ and that both $\mathcal{E}$ and $\mathcal{D}$ are 0-symmetric. A standard approach is that of designing a state feedback and applying it to the estimated state

$$u(t) = \Phi(\hat{x}(t))$$

We show now how to treat the observer error as an additional disturbance [BR71a, GS71] in the control scheme. Assume that we have computed a contractive C-set $\mathcal{P}$ for the system

$$x(t+1) = Ax(t) + Bu(t) + Ed(t) - Ae_c(t) + e_n(t), \tag{11.27}$$

where e_c and e_n are assumed to be independent signals that represent the current and the next estimation assuming arbitrary values in $\mathcal{E}$. This means that there exists $\lambda < 1$ such that, for all $x \in \mathcal{P}$, there exists $u = \phi(x)$ for which

$$Ax + Bu + Ed - Ae_c + e_n \in \lambda\mathcal{P}, \tag{11.28}$$

[4] Being associated with the well-known Kalman–Bucy filter.

for all possible disturbances $d \in \mathcal{D}$, $e_c \in \mathcal{E}$ and $e_n \in \mathcal{E}$. Now assume that both $x(t)$ and $\hat{x}(t)$ are in $\mathcal{P}$,

$$(x(t), \hat{x}(t)) \in \mathcal{P} \times \mathcal{P},$$

along with $x(t) - \hat{x}(t) \in \mathcal{E}$. Consider the equation for the state

$$\begin{aligned} x(t+1) &= Ax(t) + B\Phi(\hat{x}(t)) + Ed(t) \\ &= A\hat{x}(t) + B\Phi(\hat{x}(t)) + Ed(t) - Ae(t) \in \lambda\mathcal{P} \end{aligned}$$

(this inclusion holds in view of (11.28) because $e_n = 0$ is an admissible value), which means that the true state is inside $\lambda\mathcal{P}$.

For the estimation $\hat{x}$, as long as $\hat{x}(t) \in \mathcal{P}$, we can write the condition

$$\begin{aligned} \hat{x}(t+1) &= Ax(t) + B\Phi(\hat{x}(t)) + Ed(t) + e(t+1) \\ &= A\hat{x}(t) + B\Phi(\hat{x}(t)) + Ed(t) - Ae(t) + e(t+1) \in \lambda\mathcal{P} \end{aligned} \quad (11.29)$$

by construction (again from (11.28)). Thus we have the condition

$$(x(t), \hat{x}(t)) \in \mathcal{P} \times \mathcal{P} \quad \Rightarrow \quad (x(t+1), \hat{x}(t+1)) \in \lambda(\mathcal{P} \times \mathcal{P}),$$

which implies that the product set $\mathcal{P} \times \mathcal{P}$ is contractive. Therefore, we can find a control law $u = \Phi(x)$, positively homogeneous of order one and such that the Minkowski function of $\mathcal{P} \times \mathcal{P}$ is a Lyapunov function outside this set. We can conclude this reasoning as follows.

Proposition 11.8. *Assume that a contractive C-set $\mathcal{P}$ is known for system (11.27) and that the estimation error is bounded as $e(t) \in \mathcal{E}$, for all t. Then the set $\mathcal{P} \times \mathcal{P}$ is contractive for the extended state. Therefore there exists a control law, positively homogeneous of order one, which can be computed by considering system (11.27), where $d(t)$, e_c and e_n are assumed bounded disturbances, which assures uniform ultimate boundedness of both $x(t)$ and $\hat{x}(t)$ in $\mathcal{P}$.*

We now introduce the following claim, about separation principles, which is not supported by any formal result but only by evidence about the state of art. Unless we consider conservative solution (error upper bounds), a correct state estimator has a complexity that varies over time on-line and this is a serious obstacle to any reasonable implementation.

Claim: *There are no neat separation principles in a set-theoretic framework. In other words, if one wishes to design a compensator based on state feedback, one can only rely on conservative estimated bounds in order to design the estimated state feedback.*

Actually, this is a claim that is universal in the context of uncertain systems. For (even linear) uncertain systems, the state estimation problem is usually harder than the state feedback as long as the uncertainties entering the systems are not measurable.

The situation is different if the uncertainties w and d are unknown a priori but measurable on line or, more in general, available to the compensator. A typical case is the tracking problem, in which the tracking signal is provided to the control algorithm. Consider the case

$$x(t+1) = Ax(t) + Bu(t) + Er(t), \quad y(t) = Cx(t) + Dr(t)$$

Then a proper observer turns out to be

$$\hat{x}(t+1) = (A - LC)\hat{x}(t) + Bu(t) + Er(t) + L(y(t) - Dr(t)),$$

that obeys the equation

$$e(t+1) = (A - LC)e(t).$$

This case is virtually identical to the state feedback, since under detectability assumption the error vanishes asymptotically.

We remind that in Section 7.4 we have seen that in the case in which there are parameter variations (LPV) but the parameter value is known to the controller, then some kind of duality and separation principle can be stated.

11.3 Literature review

The first contributions about set membership estimation in connection with control trace back to the early 70s with the seminal papers [Wit68b, BR71a, BR71b, GS71, GS71, Sch73] (undoubtedly the most cited papers in this book). For several years this area has remained almost inactive, with few exceptions. The problem of set-theoretic estimation via linear programming techniques has been considered later [BM82, MT85]. Ellipsoidal techniques have been studied as well and the reader is referred to [KV97] for a comprehensive presentation.

Very recently the worst-case state estimation problem for control has received a renewed attention. Set-based estimators have been included in receding-horizon schemes [BU93, BG00, KWL00, CZ02, MRFA06]. We point out that such a technique is conceptually different from the receding-horizon estimation (see, for instance, [ABB03] and the references therein). The problem of disturbance rejection has been recently faced by means of set-theoretic estimators [ST97, ST99, AR11].

11.4 Exercises

1. Find an example in which the estimation set $\hat{\mathcal{X}}(t+1|k)$ is non-connected (and therefore non-convex).
2. State estimation in the presence of model uncertainty is not only "a difficult problem," but in some cases it is unsolvable. Find an example of a system, involving a parameter w, whose state is detectable when w is known but cannot be detected when w is unknown (one of such examples can be found in [BM03]).
3. Could you figure out an estimation scheme for $x(t+1) = Ax(t)$ and $y(t) = Cx(t) + w(t)$, where $w(t)$ is an unknown noise bounded in the C-set $\mathcal{W}$, without assuming A invertible as we did? Assume (A, C) observable and discuss the consequences on the boundedness of the estimation region.
4. In Example 11.7 show that $\bar{e}_2$ is indeed the bound for the interval. Find the minimum value of $\bar{e}_2$ and the corresponding L.
5. Assume that $\Psi_{\mathcal{S}}$ is the Minkowski function of $\mathcal{P}$ in Proposition 11.8. Find the Minkowski function $\Psi_{\mathcal{SS}}$ of the product set $\mathcal{P} \times \mathcal{P}$ and write a bound for $\Psi_{\mathcal{SS}}(x(t), \hat{x}(t))$.

Chapter 12
Related topics

In this chapter, some problems and examples which are in some way related to the exposed theory are presented.

12.1 Adaptive control

The basic motivation for which we are considering this issue is to evidence the fact that Lyapunov functions can be used not only to prove stability, but also to prove boundedness of some variables, a step which is necessary in proving the convergence of some adaptive schemes. The theory sketched here is exposed in [BW84, Ilc93, IR94]. A very specific problem which is normally solved by means of Lyapunov techniques is reported next. Consider the dynamic system

$$\dot{z}(t) = F(w(t))z(t) + G(w(t))y(t)$$
$$\dot{y}(t) = H(w(t))z(t) + J(y(t), w(t)) + M(w(t))u(t)$$

where w is an uncertain time-varying parameter ranging in a compact set $\mathcal{W}$, $u(t)$ is the input and $y(t)$ is the output. We admit that matrices F, G, H, J, and M are completely unknown, but satisfy the following assumptions.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9_12

- F, G, H, J, and M are Lipschitz continuous (this assumption can be weakened).
- $F(w(t))$ is quadratically stable, i.e. there exists $P \succ 0$ such that

$$F(w)^T P + PF(w) = -Q(w), \quad \forall w \in \mathcal{W}$$

  with $Q(w) \succ Q_0 \succ 0$.[1]
- $M(w(t))$ is positive definite, precisely, for all u,

$$u^T M(w) u \geq \alpha u^T u$$

  for some $\alpha > 0$. This is a sort of relative-degree-one assumption.
- The function J grows at most linearly with respect to y

$$\|J(y, w)\| \leq \beta \|y\|, \quad \forall w \in \mathcal{W}$$

  (note that this implies that $J(0, w) = 0$.)

We stress that α, β, P, and $Q(w)$ are all unknown and that the only quantity available for feedback is the output $y(t)$. This system includes as special case linear square systems with minimum-phase zeros and relative degree one which can written in the following form

$$\begin{aligned} \dot{z}(t) &= Fz(t) + Gy(t) \\ \dot{y}(t) &= Hz(t) + Jy(t) + Mu(t) \end{aligned} \tag{12.1}$$

with F stable. The assumption on M is basically equivalent to M non-singular, since its positive definiteness can be achieved by means of an input transformation[2]. Let $\varepsilon > 0$ be a tolerance for the error on y and for (12.1) consider the following high-gain adaptive control

$$\begin{aligned} u(t) &= -\kappa(t) y(t) \\ \dot{\kappa}(t) &= \sigma_\varepsilon^2(y(t)), \\ \kappa(0) &= \kappa_0 \geq 0, \end{aligned} \tag{12.2}$$

where (here we consider the Euclidean norm $\|y\| = \sqrt{y^T y}$)

$$\sigma_\epsilon(y) = \begin{cases} 0 & \text{if } \|y\| \leq \varepsilon \\ \|y\| - \varepsilon & \text{if } \|y\| > \varepsilon \end{cases} \tag{12.3}$$

[1] This is a sort of minimum-phase zeros assumption which is equivalent to saying that the zero-dynamics of the system which produces $y \equiv 0$ is governed by the system $\dot{z} = F(w)z$, which has to be stable (see [Kha96], Section 12.2 and Exercise 10).

[2] Just replace u by the new input v, where $u = M^T v$.

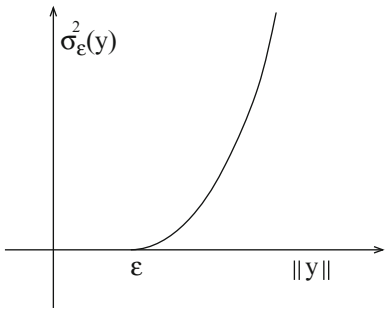


Fig. 12.1 The function $\sigma_\varepsilon^2(y)$

The function $\sigma_\varepsilon^2(y)$, depicted in Figure 12.1, represents the square of the distance from the ε-ball. It is 0 as long as $\|y\| \leq \varepsilon$ and grows quadratically when this condition is not true. It is then apparent that no gain adaptation occurs when y is in the tolerance ε-ball. We wish to prove the following.

Proposition 12.1. *The following conditions hold true.*

i) κ *converges monotonically from below to a finite value:* $\kappa(t) \to \kappa_\infty < +\infty$
ii) z *and* y *are bounded, precisely for all* $z(0)$ *and* $y(0)$ *there exists* μ *and* ν *such that* $\|z(t)\| \leq \mu$ *and* $\|y(t)\| \leq \nu$, *for all* $t > 0$.
iii) y *converges to the* ε*-ball or, equivalently,*

$$\sigma_\varepsilon(y(t)) \to 0$$

Proof. Without restriction, it is possible to assume that $P = I$, because if this were not the case, one might consider the (unknown) transformation $z = P^{-1/2}\hat{z}$ so that the F matrix becomes $\hat{F} = P^{1/2}FP^{-1/2}$ and then

$$P^{-1/2}(F^TP + PF)P^{-1/2} = P^{-1/2}F^TP^{1/2} + P^{1/2}FP^{-1/2} = \hat{F}^T + \hat{F} =$$
$$= -P^{-1/2}Q(w)P^{-1/2} \prec -P^{-1/2}Q_0P^{-1/2} \preceq -2\gamma I.$$

where the last inequality holds for $\gamma = \sigma_{min}/2$ where σ_{min} is the smallest eigenvalue of $P^{-1/2}Q_0P^{-1/2}$.

The first step of the proof is to show that κ is bounded. In view of the second equation in (12.2), $\kappa(t)$ is non-decreasing. Then it either converges to a finite limit or grows up to $+\infty$. To show that the latter option is not possible, consider the Lyapunov-like function

$$\Psi_1(z, y) = \frac{z^Tz + y^Ty}{2},$$

whose derivative is

$$\begin{aligned}\dot{\Psi}_1(z,y) = z^T\dot{z} + y^T\dot{y} = z^TFz + z^TGy + y^THz + y^TJy - \kappa y^TMy \leq \\ \leq -\gamma z^Tz + (\|G\| + \|H\|)\|z\|\|y\| + \|y^T\|\|Jy\| - \kappa\alpha y^Ty \leq \\ -\gamma\|z\|^2 + \rho\|z\|\|y\| + (\beta - \kappa\alpha)\|y\|^2 = \\ \underbrace{-\gamma\|z\|^2 + \rho\|z\|\|y\| + (\beta - \bar{\kappa}\alpha)\|y\|^2}_{\phi(z,y)} - (\kappa - \bar{\kappa})\alpha\|y\|^2.\end{aligned}$$

The new parameter $\bar{\kappa}$ is selected in such a way that the quadratic term denoted by $\phi(z,y)$ is negative definite. Assume by contradiction that $\kappa \to \infty$. Then, necessarily, $(\kappa - \bar{\kappa}) \to +\infty$ and hence the derivative becomes negative definite. In turn this means that $\Psi_1(z,y) \to 0$, say there exists $\bar{t}$ such that, for all $t \geq \bar{t}$, $\sigma_\epsilon(y(t)) = 0$, thus $\dot{\kappa} = 0$, in contradiction with the assumption $\kappa \to +\infty$. Therefore, the limit is $\kappa_\infty < +\infty$.

We now show that the state variables are bounded. Consider the new Lyapunov-like function

$$\Psi_2(z,y,\kappa,\hat{\kappa}) = \Psi_1(z,y) + (\kappa - \hat{\kappa})^2 = \frac{z^Tz + y^Ty}{2} + (\kappa - \hat{\kappa})^2$$

where $\hat{\kappa} > \kappa_\infty$ is a value that will be specified next. The variables y and z are clearly bounded if, for a proper value of $\hat{\kappa}$, function Ψ_2 remains bounded. This is what will be shown next.

As a first step, take t large enough in such a way that $\kappa > \kappa_\infty/2$ and define $\xi = (\beta - \alpha\kappa_\infty/2) > \beta - \alpha\kappa$. For $\|y\| \geq \varepsilon$, by exploiting the expression for $\dot{\Psi}_1(z,y)$ just computed, we get

$$\begin{aligned}\dot{\Psi}_2(z,y,\kappa,\hat{\kappa}) &= \dot{\Psi}_1(z,y) + 2(\kappa - \hat{\kappa})\sigma_\varepsilon^2(y) \\ &\leq -\gamma\|z\|^2 + \rho\|z\|\|y\| + (\beta - \kappa\alpha)\|y\|^2 - 2(\hat{\kappa} - \kappa)\sigma_\varepsilon^2(y) \\ &\leq -\gamma\|z\|^2 + \rho\|z\|\|y\| + \xi\|y\|^2 - 2(\hat{\kappa} - \kappa)\sigma_\varepsilon^2(y)\end{aligned}$$

The next step is to prove that this function becomes non-positive in the complement of the rectangle

$$\mathcal{R} = \{(z,y) : \|z\| \leq \zeta, \ \ \|y\| \leq 2\varepsilon\}$$

(see Fig. 12.2) if ζ is large enough. First note that, since $(\hat{\kappa} - \kappa)\sigma_\varepsilon^2(y) \geq 0$, the derivative is necessarily non-positive in the "butterfly region" (represented by the shaded region in Fig. 12.2)

$$\{(y,z) : \ \ \theta\|z\| \geq \|y\|\}$$

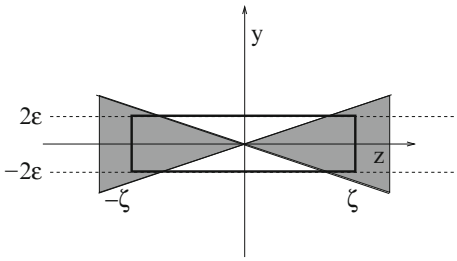


Fig. 12.2 The rectangle $\mathcal{R}$ and the "butterfly" region

for $\theta > 0$ small enough. Indeed

$$\dot{\Psi}_2(z, y, \kappa, \hat{\kappa}) \leq -\gamma\|z\|^2 + \rho\|z\|\|y\| + \xi\|y\|^2 \leq [-\gamma + \rho\theta + \xi\theta^2]\|z\|^2$$

and the right term is clearly negative for $0 < \theta < \bar{\theta}$, where $\bar{\theta}$ is the positive root of $-\gamma + \rho\theta + \xi\theta^2 = 0$.

Now we prove that the derivative cannot be positive when $\|y\| \geq 2\varepsilon$ (the border of such a region is represented by dashed lines in Fig. 12.2), for $\hat{\kappa}$ large enough. In this region $\sigma_\varepsilon^2(y) = (\|y\| - \varepsilon)^2 \geq \|y\|^2/4$ and then

$$\dot{\Psi}_2(z, y, \kappa, \hat{\kappa}) \leq -\gamma\|z\|^2 + \rho\|z\|\|y\| + \xi\|y\|^2 - 2(\hat{\kappa} - \kappa)(\|y\| - \varepsilon)^2 \leq$$
$$-\gamma\|z\|^2 + \rho\|z\|\|y\| + \xi\|y\|^2 - \frac{\hat{\kappa} - \kappa}{2}\|y\|^2 \leq 0,$$

where the last inequality holds for $\hat{\kappa}$ large enough to render negative the quadratic form in the variables $\|y\|$ and $\|z\|$.

Take now ζ such that $\zeta > 2\varepsilon/\theta$. In this way, for $\|z\| \geq \zeta$, (z, y) either belongs to the butterfly set $\theta\|z\| \geq \|y\|$ or lies outside the strip $\|y\| \leq 2\varepsilon$. Therefore, if we are outside rectangle $\mathcal{R}$ (which is disposed as in Fig. 12.2), one of these two conditions must occur so that

$$\dot{\Psi}_2(z, y, \kappa, \hat{\kappa}) \leq 0, \quad \text{for } (z, y) \notin \mathcal{R}$$

By taking $a > 0$ such that $\mathcal{N}[\Psi_2, a]$ includes the rectangle, $\Psi_2(z, y, \kappa, \hat{\kappa})$ results then non-increasing outside $\mathcal{N}[\Psi_2, a]$ and then both vectors $z(t)$ and $y(t)$ are bounded.

Finally, we have to prove that $y(t) \to [-\varepsilon, \varepsilon]$. Since z and y are bounded, then the time derivative of y,

$$\dot{y} = Hz + Jy - \kappa My$$

as well as the derivative

$$\frac{d}{dt}\sigma_\varepsilon^2(y(t)) = \frac{d}{dy}\sigma_\varepsilon^2(y(t))\dot{y}(t) = 2\sigma_\varepsilon(y(t))\dot{y}(t),$$

are bounded. Let us write

$$\kappa_\infty - \kappa(0) = \int_0^\infty \sigma_\varepsilon^2(y(t))dt < \infty$$

Thus we have that the integral of the non-negative function σ_ε^2 whose derivative is bounded on $[0, \infty)$, is bounded. In view of Barbalat's Lemma [Bar59], this means that

$$\lim_{t\to\infty} \sigma_\varepsilon(y(t)) = 0.$$

The previous theorem can be extended to the problem of tracking a proper signal r. We do not consider this case here but we refer the reader to specialized literature [Ilc93, IR94].

It has to be noticed that in the proof of the convergence of the scheme we assumed a quadratic growth for the adaptation law, so we considered the function $\sigma_\varepsilon^2(y)$. This is fundamental to prove boundedness. Quadratic growth can be dangerous in a real context, since in practice too high gain values can be unsuitable in practice. In the next section we consider a case in which a linear growth for the gain is sufficient.

12.1.1 A surge control problem

This subsection is completely dedicated to a real problem to which the high-gain adaptive control was successfully applied. Consider the compressor-plenum plant proposed in [Gre76], represented by the following second order model

$$\begin{aligned} \dot{x}_1(t) &= -B[x_2(t) - \Psi_s(x_1(t))] \\ \dot{x}_2(t) &= \tfrac{1}{B}[x_1(t) - u(t)\Gamma_s(x_2(t))], \end{aligned} \tag{12.4}$$

where x_1 is the dimensionless flow rate in a compressor duct, that pumps a compressible fluid in a plenum, whose dimensionless pressure is x_2. The coefficient $B > 0$ is known as the Greitzer parameter, $\Psi_s(x_1)$ and $\Gamma_s(x_2)$ are the static characteristics of the compressor and the throttle valve, respectively. The function $u(t)$, which is the system input, represents the throttle valve fraction opening and then it is subject to the constraints $0 \leq u \leq 1$. Model (12.4) can be associated with the output equation

$$y(t) = x_2(t) - \Psi_s(x_1(t)), \tag{12.5}$$

which expresses the dimensionless total pressure at the compressor inlet. It is well known that, due to compressor stall, the characteristic $\Psi_s(x_1)$ may present a drop. As a consequence, for small values of u the system may become unstable. In some conditions the system may exhibit limit cycles known as surge cycles.

This phenomenon has been investigated and some solutions based on feedback control have been proposed to suppress the surge.

We work under the conditions of qualitative knowledge of the system, assuming most of the parameters and curves unknown. The following assumptions are introduced.

Function $\Psi_s(x_1)$ is defined for $x_1 \leq x_1^+$, where $x_1^+ > 0$ is an unknown value for which $\Psi_s(x_1^+) = 0$, and it is piecewise-smooth on its domain of definition. Furthermore, $\Psi_s(x_1) > 0$ for $x_1 < x_1^+$ and

$$\lim_{x_1 \to -\infty} \Psi_s(x_1) = +\infty.$$

Function $\Gamma_s(x_2)$ is piecewise-smooth, strictly increasing and it is such that

$$x_2 \Gamma_s(x_2) \geq 0.$$

The static value u_0 can be chosen within the interval $[u^-, u^+]$, where $0 < u^-$ and $u^+ < 1$, with the exception of a single value u_{stall} (representing the value in which the compressor presents a stall). For all points

$$u_0 \in [u^-, u^+], \quad u_0 \neq u_{stall},$$

the system has an isolated equilibrium point $P = (x_1^*, x_2^*)$ characterized by

$$x_2 - \Psi_s(x_1) = 0, \quad x_1 - u_0 \Gamma_s(x_2) = 0.$$

We also assume that the two curves cannot have common tangent lines in P.

Denote by X_{amm} the next admissibility region in the x_1–x_2 plane (see Fig. 12.3)

$$X_{amm} = \{x : x_1 \leq x_1^+, \ x_2 \geq x_2^- > 0\},$$

where x_2^- is an assigned small positive value. In the following, it will be assumed that (these assumptions are commented in [BG02]): there exists a compact set $X_0 \subset X_{amm}$, depending on u_0, including the unique isolated equilibrium point $P \in X_0$, and for each initial condition $x(0) \in X_0$ the corresponding natural (i.e., uncontrolled) solution with $u(t) = u_0$ remains admissible, i.e., $x(t) \in X_{amm}$, and is bounded, namely, $\|x(t)\| \leq M$, for some positive M depending on u_0. The above assumptions are in practice equivalent to the fact that the plant characteristics are only qualitatively known, as it is often the case in real problems. Define the quantity

$$v(t) = u(t) - u_0.$$

It can be shown, via local stability analysis, that for $v = \kappa y$, with $\kappa \geq 0$ sufficiently large, k the equilibrium point is stable. However if one seeks for a more general solution (described later) local analysis is not sufficient. We consider the following

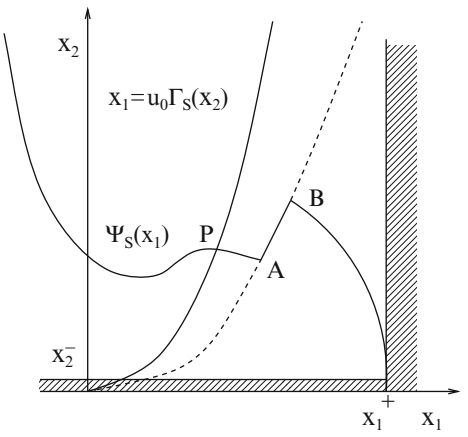


Fig. 12.3 The compressor and throttle characteristics

adaptive control law

$$v(t) = \kappa(t)y(t), \tag{12.6}$$

$$\dot{\kappa}(t) = \mu\sigma_\varepsilon(y(t)), \;\; \kappa(0) = 0, \tag{12.7}$$

where μ is an adaptation factor, and $\sigma_\varepsilon(y)$ is the function (previously defined)

$$\sigma_\varepsilon(y) = \begin{cases} 0 & \text{if } |y| \le \varepsilon \\ |y| - \varepsilon & \text{if } |y| > \varepsilon \end{cases} \tag{12.8}$$

Note that, by considering $\sigma_\varepsilon(y)$ instead of $\sigma_\varepsilon^2(y)$ in the adaptation rule, we are penalizing the distance from the interval of radius ε, not its square as in the previous case. We will now prove in a set-theoretic framework that this solution actually works.

Proposition 12.2. *The closed-loop system with the above adaptation scheme is such that*

i) from any initial condition $x(0) \in X_{amm}$ the closed-loop solution remains admissible and bounded.
ii) the parameter κ remains bounded $\kappa(t) \to \kappa_\infty < +\infty$.
iii) $y(t) \to [-\varepsilon, \varepsilon]$.

Proof. To prove this proposition, we need a preliminary (but fundamental) step. We show that for all initial conditions in the unknown set X_0 for which the natural solution $\bar{x}(t)$ (achieved by $\kappa(t) = 0$) remains admissible (i.e., in the admissible region) then also the solution $x(t)$ with the control $\kappa(t) > 0$ remains *bounded and admissible*, precisely, for each $x(0) = x_0 \in X_0$, $\|x(t)\| \le M$ and $x(t) \in X_{amm}$.

We prove this fact by determining some suitable compact invariant sets (for the controlled system) enclosing the initial state $x(0)$. These sets are determined (or partially determined) by the trajectories of the uncontrolled solution.

First recall that, by assumption, the norm of the uncontrolled solution $\bar{x}(t)$, achieved by $\kappa(t) = 0$ and $\bar{x}_0 \in X_0$, is bounded by M. Therefore, the uncontrolled solution $\bar{x}(t)$ either converges to a limit cycle or to the unique equilibrium point P (see [Kha96], Section 7.1). We distinguish two cases:

1. solution $\bar{x}(t)$ converges to a limit cycle, say $\mathcal{C}$, and $\bar{x}(0) \in \bar{\mathcal{C}}$, the closed area encircled by $\mathcal{C}$;
2. solution $\bar{x}(t)$ converges to a limit cycle, say $\mathcal{C}$, but $\bar{x}(0) \notin \bar{\mathcal{C}}$, or $\bar{x}(t)$ converges to P.

In the first case we can show that any solution corresponding to $\kappa(t) \geq 0$ does not exit from the area $\bar{\mathcal{C}}$ enclosed by $\mathcal{C}$. To this aim, we apply Nagumo's theorem to show that $\bar{\mathcal{C}}$ is positively invariant, not only, as it is obvious, for the uncontrolled system, but also for the time-varying system with any $\kappa(t) \geq 0$. Since $\mathcal{C}$ is a differentiable curve, the tangent cone for x on the boundary of $\bar{\mathcal{C}}$ is

$$\mathcal{T}_S(x) = \{z : n^T z \leq 0\},$$

where n is the external normal vector orthogonal to $\mathcal{C}$ (see the first case in Fig. 12.4) in point $x = (x_1, x_2)$, given by

$$n^T = \gamma(x_1, x_2) \left[\tfrac{1}{B}[x_1 - u_0 \Gamma_s(x_2)] \quad B[x_2 - \Psi_s(x_1)] \right] \tag{12.9}$$

(precisely the unit vector orthogonal to the "natural" derivative $\dot{\bar{x}}$), where $\gamma(x_1, x_2) = 1/\sqrt{\dot{\bar{x}}_1^2 + \dot{\bar{x}}_2^2} > 0$ is a normalizing factor. By applying Nagumo's condition we achieve for $x \in \mathcal{C}$,

$$n^T \dot{x} = -\kappa(t)\gamma(x_1, x_2)\Gamma_s(x_2)[x_2 - \Psi_s(x_1)]^2 \leq 0. \tag{12.10}$$

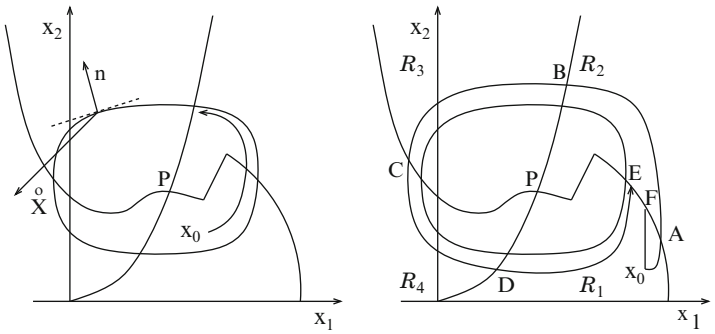


Fig. 12.4 The system trajectories

Thus, $\bar{\mathcal{C}}$ is positively invariant for the controlled system and the solution is bounded for all $x(0)$ inside this set.

The second case is more involved and we deal with it intuitively, since a formal proof would be too long (the reader is referred to [BG02] for more details). As a first step, note that we can distinguish four subregions $\mathcal{R}_i$ according to the signs of the derivative components $\dot{\bar{x}}_1$ and $\dot{\bar{x}}_2$ in the uncontrolled system (see Fig. 12.4) as follows:

$$\mathcal{R}_1 : (+,+), \quad \mathcal{R}_2 : (-,+), \quad \mathcal{R}_3 : (-,-), \quad \mathcal{R}_4 : (+,-).$$

Assume now that a limit cycle exists and consider any initial condition $\bar{x}_0 \in X_0$, outside the limit cycle. The corresponding uncontrolled trajectory $\bar{x}(t)$ intersects infinitely many times all regions $\mathcal{R}_i$ and the characteristic branches in strict periodic sequence.

Assume for brevity that $\bar{x}(0) \in \mathcal{R}_1$ (the other four cases can be dealt with in the same way). Denote by A, B, C, D, and E the intersections with the characteristic branches. Let E be the second intersection point with the compressor characteristic on the right of P. Necessarily E is on the left of A, because the trajectory cannot intersect itself, and must converge to the limit cycle. Now let F be the point on the compressor characteristic having the same abscissa of x_0. We claim that the region enclosed by the closed curve $\bar{x}(t)$ through point x_0, A, B, C, D, E, arc E–F, and vertical segment F–x_0 is positively invariant for the *controlled system*. To show this we have to analyze Nagumo's condition along the boundary. Along the boundary given by the curve associated with $\bar{x}(t)$, we derive exactly the same condition (12.9) obtained before. Let us consider the curve E–F. In any point $x = (\bar{x}_1, \bar{x}_2)$ in which it is differentiable, the tangent cone (actually a half-plane) is $\mathcal{T}_{\bar{\mathcal{C}}}(x) = \{z : \ z_2 - \Psi_s(\bar{x}_1)z_1 \geq 0\}$. Since $v = 0$ on the characteristic $x_2 - \Psi_s(x_1) = 0$ the derivative is $\dot{x} = [0 \ \ (\bar{x}_1 - u_0\Gamma_s(\bar{x}_2))/B]^T$ and its second component is positive, so $\dot{x} \in \mathcal{T}_{\bar{\mathcal{C}}}$. If the curve E–F is not differentiable in the point $(\bar{x}_1, \bar{x}_2)$ (for instance in the points associated with the stall) the tangent cone is characterized by the two inequalities $\mathcal{T}_{\bar{\mathcal{C}}}(x) = \{z : \ z_2 - \Psi_s'(\bar{x}_1^-)z_1 \geq 0, \quad z_2 - \Psi_s'(\bar{x}_1^+)z_1 \geq 0\}$ associated with the left and the right derivative and the same reasoning applies.

On the boundary formed by the segment F–x_0, the tangent cone is the set $\mathcal{T}_{\bar{\mathcal{C}}} = \{z : \ z_1 \geq 0\}$. Since the derivative $\dot{x}_1$ is non-negative in this sector, we have $\dot{x} \in \mathcal{T}_{\bar{\mathcal{C}}}$. We need only to consider the points E, F and x_0. It is very easy to see that in the point E the tangent cone is the union of the second and third quadrant and the half-plane $\mathcal{H}$ delimited by the line tangent to the curve in point E[3]. In the point F the tangent cone is the union of the first quadrant, the fourth quadrant and $\mathcal{H}$. Checking the condition $\dot{x} \in \mathcal{T}_{\bar{\mathcal{C}}}$ is very easy in both points E and F, since the derivative vector is vertical (we remind that $v = 0$).

[3] $\mathcal{H} = \{z : z_2 \geq \Psi_s'(\bar{x}_1)z_1\}$.

As far as the point x_0 is concerned, the tangent cone is delimited by the line passing through the origin and parallel to the tangent line of the natural solution passing through x_0 and by the vertical line

$$\mathcal{T}_{\bar{\mathcal{C}}}(x_0) = \{z : z_1 \geq 0, \quad n^T z \leq 0\},$$

where n in the unit vector already defined in (12.9). Checking the condition $\dot{x} \in \mathcal{T}_{\bar{\mathcal{C}}}(x_0)$ is immediate.

The case in which a limit cycle does not exist for the natural trajectory, so that $\bar{x}(t) \to P$ asymptotically, can be handled in the same way by considering P as a "degenerate limit cycle." Even in this case we can identify an area, partially delimited by $\bar{x}(t)$, which is invariant for the controlled solution. Therefore the situation is as in Fig. 12.4, Case 2, where point E (as well as A, B, C, D) may be equal to P.

Once we have established the boundedness of the controlled solution, we can use argument similar to those previously used. We first show that κ is bounded.

By contradiction, assume that $\kappa(t) \to +\infty$ monotonically. To this aim we need to rewrite the equations by changing variables

$$\begin{aligned} y(t) &= x_2(t) - \Psi_s(x_1(t)) \\ z(t) &= \qquad x_1(t), \end{aligned}$$

we get an equation of the form

$$\begin{aligned} \dot{z} &= -By \\ \dot{y} &= g(z,y) - \Gamma_s(y + \Psi_s(z))u = g(z,y) - \kappa\Gamma_s(y + \Psi_s(z))y. \end{aligned} \tag{12.11}$$

Since we have proved that x_1 and x_2 are bounded, so are z and y and the term $|g(z,y)| \leq \rho$ and then, since the variables remain admissible, $|\Gamma_s(y + \Psi_s(z))| \geq \xi > 0$. Consider the Lyapunov-like function $\Psi(y) = y^2/2$. Then

$$\dot{\Psi}(y) = yg(z,y) - \frac{\kappa}{B}\Gamma_s(y + \Psi_s(z))y^2 \leq \rho|y| - \xi\kappa y^2.$$

By taking $\bar{\kappa} > B\rho/\varepsilon\xi$ we have that, for $\kappa \geq \bar{\kappa}$, $\dot{\Psi}(y)$ is strictly negative outside the open interval $(-\varepsilon, \varepsilon)$. Therefore, it is a Lyapunov function outside this interval, which in turn implies that $y(t)$ is ultimately bounded inside the interval $[-\varepsilon, \varepsilon]$. This means that there exists $T > 0$ such that, for all $t \geq T$, $|y(t)| \leq \varepsilon$ and so, for all $t \geq T$, $\dot{\kappa}(t) = 0$, in contradiction with the fact that $\kappa(t) \to +\infty$

Since the variables are bounded, in view of the equation

$$\dot{y} = g(z,y) - \Gamma_s(y + \Psi_s(x_1))\kappa y/B,$$

$\dot{y}$ is bounded.

Consider equation (12.7) and write it in the equivalent form

$$\kappa_\infty - \kappa(0) = \mu \int_0^\infty \sigma_\varepsilon(y(t))dt$$

Then we can apply Barbalat's Lemma and conclude that $\sigma_\varepsilon(y(t)) \to 0$, which is what was to be proved.

It can be shown that if we take $\varepsilon \to 0$, the state variables approaches the desired equilibrium point (see [BG02] for details and experimental results). We also remark that the considered system, in view of Equation (12.11), does not satisfy the strict minimum-phase conditions imposed in the previous case (cf. footnote 1 in this section).

12.2 The domain of attraction

Throughout the book we have quite often mentioned the concept of domain of attraction, especially in Chapter 8, when constrained control problems have been considered. In view of the importance of the subject, we decided to dedicate a section to a concept which is clearly related to it: the region of asymptotic stability. The reader is referred to specialized literature, for instance the excellent survey [GTV85] and book [Che11a].

Given a system $\dot{x}(t) = f(x)$, with $f(0) = 0$, for which the origin is an asymptotically stable equilibrium point, we define as Region of Asymptotic Stability (RAS) the set of all the initial conditions x_0 such that if $x(0) = x_0$ the corresponding solution $x(t) \to 0$ as $t \to \infty$. The first investigation about RAS is due to the seminal work of Zubov [Zub64] and La Salle [LL61]. In particular it is known that if a region of asymptotic stability is sought as $\mathcal{N}[\Psi, 1]$, the 1-sublevel set of a Lyapunov smooth function $\Psi(x)$, then the following equation, known as Zubov equation, comes into play

$$\nabla\Psi(x)f(x) = -\phi(x)[1 - \Psi(x)],$$

where $\phi(x)$ is a positive definite function. It is also well known that solving this partial differential equation (without brute-force methods) is hopeless and therefore, approximate methods are necessary. These methods typically allow for the determination of what we call here a Domain of Attraction. A domain of attraction $\mathcal{S}$ is any set $\mathcal{S}$ for which

i) $0 \in \mathcal{S}$;
ii) $\mathcal{S}$ is positively invariant;
iii) $x(0) = x_0 \in \mathcal{S}$ implies $x(t) \to 0$.

Obviously, any domain of attraction is a subset of the region of asymptotic stability. Several techniques can be used to achieve a domain of attraction and the set-theoretic approach can be very helpful to solve the problem. We sketch some possible applications.

12.2.1 Systems with constraints

The case of systems with constraints has been already considered in this book and we just propose a new point of view. Here we assume that the system is locally stable and that constraints of the form

$$g(x) \in \mathcal{Y}$$

are assigned. Under the obvious assumption that $g(0) \in \mathcal{Y}$, an interesting problem is the determination of a set of initial states for which the constraints are satisfied during the transient. This requires the determination of the constrained domain of attraction, precisely a set which is admissible and positively invariant according to Theorem 8.2 at the beginning of Chapter 8. That chapter has been basically devoted to the determination of a proper control action for which the constraints are satisfied. In some cases the controller is given and thus we can only determine the corresponding domain of attraction.

In the case of asymptotically stable linear systems with linear constraints this problem can be solved exactly for discrete-time systems as shown in Section 5.4 and approximately for continuous-time systems. We have seen that solving this problem is possible even in the case of uncertain linear systems, so the case of nonlinear systems is considered next.

A possible solution is achieved by the model-absorbing technique presented in Section 2.1.2. This technique is based on the possibility of determining a compact $\mathcal{W}$ such that, in a proper domain $\mathcal{X}$, the following equality holds:

$$f(x) = A(w)x, \quad \text{for some } \; w \in \mathcal{W}$$

If the above holds, then any domain of attraction for the linear differential inclusion is a domain of attraction for the original dynamics. One positive aspect of the above way of proceeding is that uncertain systems can also be dealt with. The shortcoming is that the approximation can be very rough, as shown in the next example.

Example 12.3. Consider the following system

$$\begin{aligned}\dot{x}_1(t) &= -[x_1(t) - x_1^3(t)] - x_2(t)\\ \dot{x}_2(t) &= x_1(t) - x_2(t)\end{aligned}$$

and the local Lyapunov function

$$\Psi(x) = x_1^2 + x_2^2$$

(determined via linearization), whose derivative is

$$\dot{\Psi}(x) = -2x_1^2 - 2x_2^2 + 2x_1^4.$$

To determine a domain of attraction of the form $\mathcal{N}[\Psi, \kappa]$, we seek for the supremum value κ_{sup} of κ for which $\dot{\Psi}(x) < 0$ for all non-zero $x \in \mathcal{N}[\Psi, \kappa]$. Simple computations yield $\kappa_{sup} = 1$, so that the corresponding domain is the open unit circle (see Fig. 12.5). A different possibility is the following. Write the nonlinearity between square brackets as

$$-[x_1(t) - x_1^3(t)] = -[1 - w]x_1, \quad \text{with } w = x_1^2$$

By imposing bounds of the form

$$|x_1| \leq \xi \tag{12.12}$$

we have

$$|w| \leq \bar{w} = \xi^2 \tag{12.13}$$

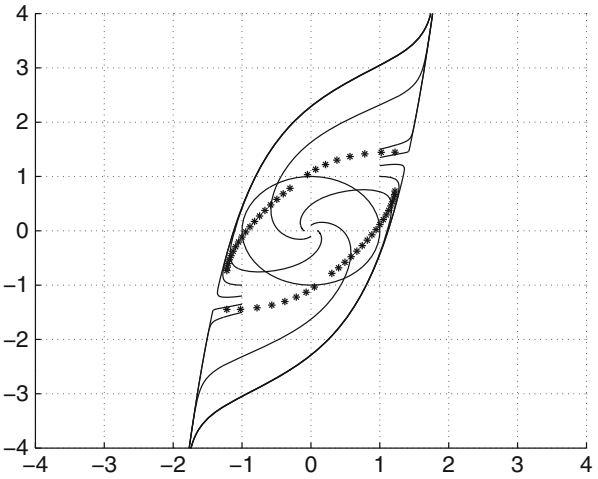


Fig. 12.5 The true and the estimated domains of attractions

The uncertain linear absorbing system, valid under condition (12.12), has the following state matrix

$$A(w) = \begin{bmatrix} -[1-w] & -1 \\ 1 & -1 \end{bmatrix}, \quad |w| \leq \bar{w},$$

and its largest domain of attraction subject to the constraints (12.12) and (12.13), was computed for $\xi = 3/2$ and $\bar{w} = (3/2)^2$ by means of a polyhedral Lyapunov function. The vertices of the unit ball are denoted by "*" in Figure 12.5. Note that, although it is "slightly better" (since it has a larger area), the domain achieved by absorbing does not include the domain of attraction achieved by the local Lyapunov function.

The "true" domain of attraction has been also represented. In such kind of simple examples, this domain can be "numerically determined" by considering several trajectories computed by backward simulation starting from several "final" conditions inside the previous domain of attraction. This example clearly shows the limits of the determination of the domain of attraction achieved by guessing the shape of the function or by the merging in an uncertain linear system.

In the next example we show a possible application to a real problem.

Example 12.4. Let us consider the magnetic levitator of Figure 2.1 in Subsection 2.1.2 and the corresponding absorbing equation (2.19) derived there. We rewrite here the state update and input matrices,

$$A = \begin{bmatrix} 0 & 1 \\ a & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ -b \end{bmatrix},$$

with

$$1507 \leq a \leq 2177, \quad 17.8 \leq b \leq 28.$$

The two variables of this system, x_1 and x_2, represent the position and the speed of an iron ball, while the input is the magnet current. We remind that position and speed are oriented downwards (i.e., increasing x_1 means lowering the sphere). For this system we impose the control constraints $|u| \leq \omega$, which are obviously due to the current limitation (actually only the upper bound is really important, since the "true current is $\bar{i} + u$, where $\bar{i}$ is the steady-state current). We also assume the following position constraints $|x_1| \leq \xi$. The lower limit is imposed by the fact that the sphere cannot get too close to the magnet since, due to residual magnetic field, the ball is attracted even for zero current. The upper level is due to the fact that to start the experiment the ball is carried by a platform which has been placed at a distance ξ from the nominal position. The fact that the limits are symmetrical has been imposed for convenience.

For $\omega = 0.7\text{A}$, $\xi = 0.005\text{m}$, and $\bar{i} = 1\text{A}$, we determined the polygon whose vertices are the columns of the matrix $[X \ -X]$, where

$$X = \begin{bmatrix} 0.005 & 0.005 \\ 0 & -0.398 \end{bmatrix}$$

which is contractive and can be associated with the linear control law

$$u = 120x_1 + 2.71x_2$$

This information is important because it means that if we place the ball on the platform at a distance $\xi = 5mm$ from the nominal one then, since the corresponding state $x(0) = [0.005\ 0]^T$ is in the computed set (actually a vertex), we can lift the sphere without constraint violation (as experimentally verified). In Figure 12.6, the estimated domain of attraction is depicted along with some kind of model-based "true" region computed as follows.

We have computed the full-force trajectory with maximum current $u = \omega$ from the lower limit of $x_1 = -\xi$ and we considered the maximum initial speed for which the breaking effect is such that the peak of x_1 (lowest position) is on the maximum value ξ. We have repeated the same experiment by integrating the trajectory with $i = 0$ (this means that $u = -1$ and then it is out of the considered limits) starting from $x_1 = \xi$ with negative speed and we computed the maximum negative speed such that the sphere reaches the lower limit $-\xi$ without crossing it. These extremal curves characterize a "safety region" which appears, as expected, much greater than the estimated one.

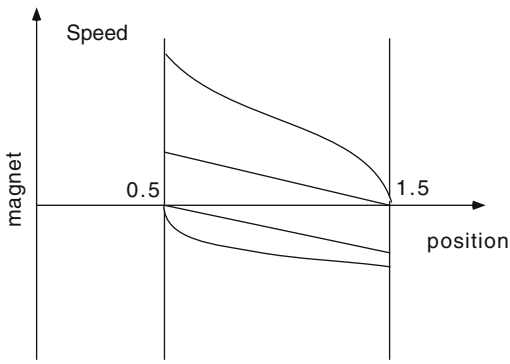


Fig. 12.6 The estimated (polygonal) and the true and safety regions

12.3 Obstacle avoidance

Obstacle avoidance is relevant in many applications in which regions in the state space must be avoided for safety reasons. Such a problem is encountered, for instance, in robotics, navigation, automated vehicles, and flight control.

Conceptually, the problem formulation is identical to the problem of keeping the state in a region, as we have seen in Section 5.1. Consider the dynamical system

$$\dot{x}(t) = F(x(t), u(t), w(t)),$$

where u is the control, $u(t) \in \mathcal{U}$, and w is the disturbance, $w(t) \in \mathcal{W}$. For such system, given a region $\mathcal{X}$ in the state space to be avoided, the problem faced here can then be translated into

$$x(t) \in \tilde{\mathcal{X}},$$

where $\tilde{\mathcal{X}}$ is the complement of $\mathcal{X}$, and therefore we are dealing with a constrained control problem.

The essential difference is that, while the convexity assumptions considered in the constrained control case are reasonable in most cases, the convexity assumption on $\tilde{\mathcal{X}}$ would be unacceptable in most situations. This fact renders the problem both difficult and challenging.

Often obstacles are just associated with regions of the physical space one should avoid. If, for instance, a convex obstacle is given, then avoiding it is a constrained problem with respect to the complement of a convex set.

Example 12.5 (Obstacle avoidance: pit-jumping dilemma). Consider the problem of jumping a pit as in Fig. 12.7. No biker-experience is needed to see that the bike or a runner may cross the pit provided that its (his/her) speed is large enough. In simple words, crossing is possible provided that the speed at the pit bound is greater than a certain quantity. Thus the expression for the obstacle is

$$\mathcal{O} = \{(y, v) : \ |y| \leq d, \ \ |v| \leq v_{min}\},$$

where y is the distance from the center of the ditch, $v = \dot{y}$ is the speed, v_{min} is the minimum speed at the boundary which is necessary to jump on the other side

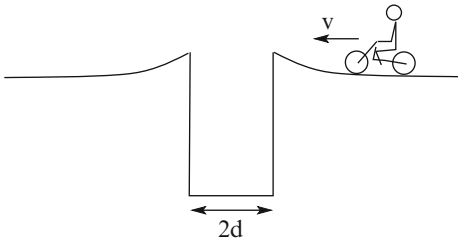


Fig. 12.7 Pit-jumping

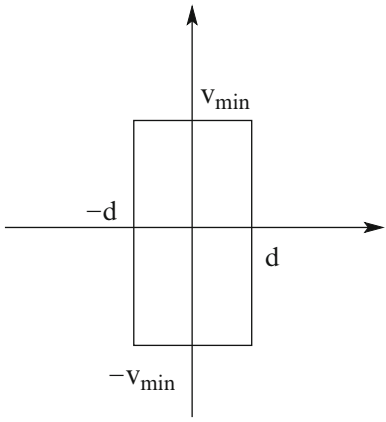


Fig. 12.8 The obstacle associated with the pit-jumping

of the ditch. Therefore, a (reasonable) problem for the biker is to be outside the convex set $\mathcal{O}$. The obstacle is represented in the state-space (position-speed space) in Fig. 12.8.

Let us now assume that the following equation is associated with the bike: $\ddot{y} = u$, with $-u_{min} \leq u \leq u_{max}$, where u_{min} and u_{max} are the minimum (i.e., braking) and maximum acceleration.

The dynamic problem is intuitive (as experience teaches us). There is a dangerous zone in terms of speed and position. If the bike approaching the ditch is not fast/slow enough within a certain distance from the ditch it might get into trouble. How is it possible to determine such a region and avoid it?

A possibility to face the previous question is to use the dynamic programming ideas developed in Chapter 5. These ideas have been around for many years [BR71a] with pursuit-evasion games. See also [CDS02, RBCM07] for more recent developments. Assume that the region to be avoided is $\mathcal{O}_0 = \mathcal{O}$. Consider the dynamics represented by the system

$$x(t+1) = f(x(t), u(t)), \qquad u \in \mathcal{U}.$$

Then, in discrete-time, let us consider the 1-step no-return region $\mathcal{O}_{-1}$, namely the region for which, no matter how u is taken, we will have that the next step x will be in the "pit" $\mathcal{O}_0$

$$\mathcal{O}_{-1} = \{x : f(x,u) \in \mathcal{O}_0, \ \ \forall u \in \mathcal{U}\}.$$

Needless to say this is a "suicide" region to be avoided. If the biker is inside, then there is no way to avoid falling into the ditch. The rest of the procedure is similar. Recursively let us compute

$$\mathcal{O}_{-(k+1)} = \{x : f(x,u) \in \mathcal{O}_k, \ \ \forall u \in \mathcal{U}\}$$

This is the set such that $x(t) \in \mathcal{O}_{-(k+1)}$ implies $x(t+1) \in \mathcal{O}_{-k}$, $x(t+2) \in \mathcal{O}_{-k+1}$ and so on. The set

$$\mathcal{D} \doteq \bigcup_{k=0}^{\infty} \mathcal{O}_{-k}$$

is the region to be avoided. If the biker is in this region, she/he cannot avoid the obstacle. It is immediately seen that if the dynamics is linear and if the initial set $\mathcal{O}_0$ is convex, the sets $\mathcal{O}_{-k}$ are convex, although $\mathcal{D}$ is not necessarily convex. If the sets $\mathcal{O}$ and $\mathcal{U}$ are polyhedral C-sets, then all the $\mathcal{O}_{-k}$ are polyhedral C-sets. More precisely, assuming the plane representation $\mathcal{O}_{-k} = \mathcal{P}[F, g]$, then $\mathcal{O}_{-k-1}$ can be computed by first determining its erosion (see Subsection 3.1.1)

$$\mathcal{E}_{-k} = \{y : \ y + Bu \in \mathcal{O}_{-k}, \quad \forall u \in \mathcal{U}\} = \mathcal{P}[F, \tilde{g}]$$

and then the pre-image

$$\mathcal{O}_{-(k+1)} = \{x : \ Ax \in \mathcal{E}_{-k}\} = \mathcal{P}[FA, \tilde{g}].$$

Interestingly enough, for systems without uncertainties, the complexity of the sets $\mathcal{O}_{-k}$ does not increase with time, say the complexity is linear in the time horizon.

Example 12.6 (Pit-jumping dilemma continued). Let us consider the discrete-time equivalent for the pit-jumping problem

$$x(t+1) = Ax(t) + Bu(t)$$

with

$$A = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} T^2/2 \\ T \end{bmatrix}.$$

Assuming $v_{min} = 5$, $d = 2$ and $|u| \leq u_{max} = 1$, the sequence of sets depicted in Fig. 12.9 is obtained. It turns out that the sets $\mathcal{O}_{-k}$ have empty interior for $k \geq 21$. The union of these sets is the region $\mathcal{D}$ which has to be avoided. The form of the dangerous set has an immediate physical interpretation. Take, for instance, the point $(-4, 4)$. Being in this state means that the bike is approaching the obstacle with positive speed and the engine has not enough power to reach the 'jumping" speed nor the brake has enough power to stop before the ditch. Note that for points outside the dangerous set the strategy might be different. In the state $(-5, 2)$ the bike is safe if it brakes, reaching the zero speed before the ditch, while in point $(-4, 6)$ a full force acceleration is "suggested" to jump over the ditch.

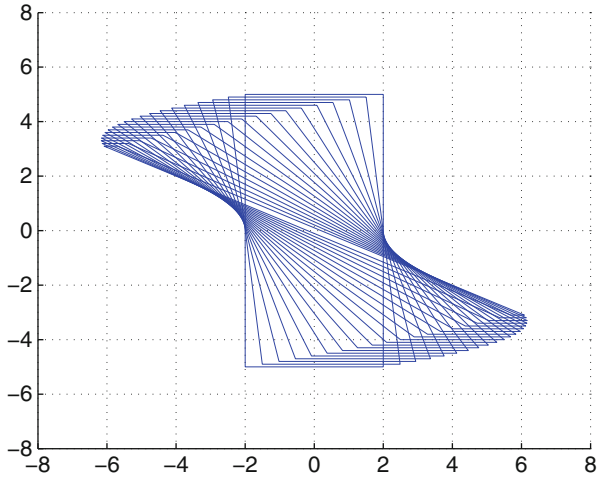


Fig. 12.9 The sequence of sets $\mathcal{O}_{-k}$ whose union is the region $\mathcal{D}$

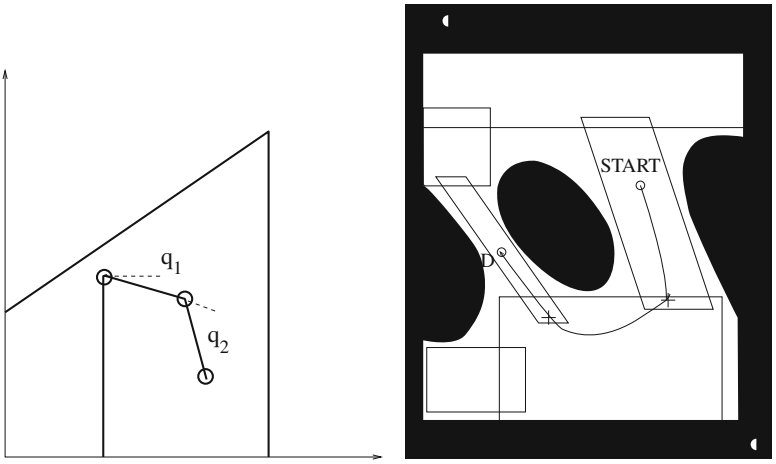


Fig. 12.10 Robot in a constrained environment

There are some cases in which the form of the obstacle is so complex that the previous procedure would become computationally unfeasible. This is the case of obstacle avoidance in robotics.

Assume that a manipulator has to operate in a constrained environment as in Fig. 12.10 left. Assume that the angles q_1 and q_2 are the free coordinates of the robot which has to move in a constrained environment. Generally speaking, even though the allowable physical space is simple, the corresponding allowable region in the coordinate space might be very hard to describe. Typically it is non-convex as the white set in Fig. 12.10 right.

One possible solution is to fill the admissible region with a family of simple overlapping sets, typically convex and compact (see, e.g., Fig. 12.10 right). This set-covering technique has been suggested in [MADF00] and the idea is described in [BPV04, BMPVA08]. It is based on constructing regions with crossing points between regions and on equipping the system with a hierarchical control with

- a high-level global controller, which decides a path of connected sets in which the first includes the starting point and the last the destination point (D in the figure). The high level control makes use of a connection graph.
- a low-level local controller, active in each convex set, which tracks the reference (if the reference is inside the current set) or tracks a "crossing" point to another set of the sequence which is closer to the final set.

The technical problem in the case of the robot is that the sets are naturally defined in the free coordinate q-space, not in the state-space (whose components are usually the coordinates and their derivatives $(q, \dot{q})$). Therefore they have to be expanded ("inflated") in order to become convex sets with non-empty interior in the state-space, roughly the $(q, \dot{q})$-space. We do not proceed further in describing the details of this problem. The interested reader is referred to [BPV04, BMPVA08]. We rather analyze the simplest version of the problem: a point moving in a constrained space q subject to the equation

$$\dot{q} = v, \tag{12.14}$$

where $q \in \mathbb{R}^n$ is the coordinate vector and v is the speed assumed as a control signal, so that the q-space is also the state space[4]. Assuming that a maximum speed $\|v\|_2 \leq \bar{v}$ for the robot is assigned, we can adopt as local controller

$$v = -\bar{v}\frac{\kappa(q - \bar{q})}{\max\{1, \kappa\|q - \bar{q}\|\}}, \tag{12.15}$$

where $\kappa > 0$ is a parameter and $\bar{q}$ is either the target point or a crossing point to the next set of the sequence.

This low-level control obviously satisfies the input constraints $\|v\|_2 \leq \bar{v}$. It is not difficult to see that it satisfies the space constraints too. If $\mathcal{S}_i$ is the set of the family in which the state $q(t_0)$ and the target or crossing point $\bar{q}$ are included at time t_0, then the resulting trajectory is

$$q(t) = \theta(t)(q(t_0) - \bar{q}) + \bar{q} \tag{12.16}$$

[4]This is typically reasonable when the system is equipped by a "fast" speed-loop control, so that the speed can be approximately assumed as input.

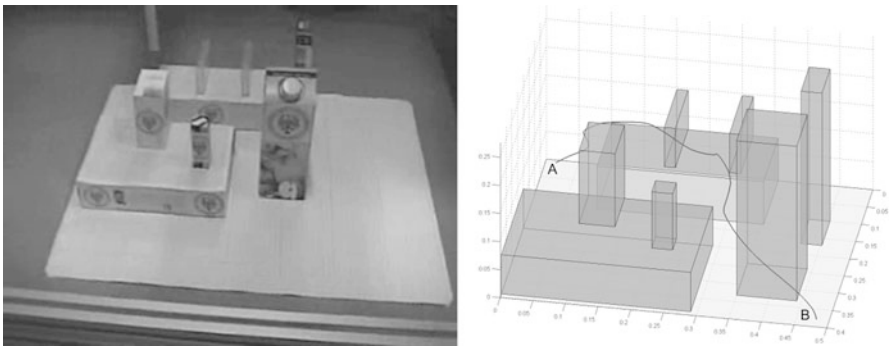


Fig. 12.11 The constrained environment and the (experimental) trajectory in a constrained environment

for some scalar function $\theta(t)$ such that $\theta(t_0) = 1$ and $\theta(\infty) = 0$. By replacing q as in (12.16) in (12.15) and then replacing v into (12.14), we see that function $\theta(t)$ satisfies the following equation

$$\dot{\theta}(t) = -\bar{v}\frac{\kappa\theta(t)}{\max\{1, \kappa\theta(t)\|q(t_0) - \bar{q}\|\}}. \tag{12.17}$$

The so derived closed-loop system satisfies the constraint as long as both $q(t_0)$ and $\bar{q}$ are in the convex set $\mathcal{S}_i$, because the trajectory lays on the segment having extrema $q(t_0)$ and $\bar{q}$. Note also that the speed of the process is saturated, $|v| = v_{max}$, when $q(t)$ is far from the target $\bar{q}$, and it is proportional to $\bar{q} - q$ when $q(t)$ is close to $\bar{q}$, so it converges exponentially.

This class of controllers has been successfully experimented on a Cartesian Robot in a constrained environment (Fig. 12.11).

12.4 Biological models

Biological systems, actually biological models, are a mine of interesting (and funny) problems for both engineers and mathematicians. There are many books which are devoted to the topic, for instance [EK05, CWLA05, CB11, DVM14]. Perhaps mathematics will not solve the major problems humans are encountering in biology and nature, but it might help understanding some principles underlying the behavior of many systems.

One fundamental field in which mathematics provides essential tools is that of chemical networks and biomolecular chemistry [DVM14]. Mathematical chemistry produced nice theorems, such as the 0 deficiency theorem [Fei87], which proves stability for a large class of networks governed by mass-action kinetics using the entropy as a Lyapunov function.

Here we wish to sketch some examples and ideas about problems which may benefit from the theory presented in the book. Some ideas in this direction have been presented in [ATS07]. We speak in general of "chemical reactions models" with the understanding that these principles are general in nature.

Consider the following reaction

$$A + B \to C + D$$

in which a molecule of the species A and a molecule of the species B react to produce a molecule C and a molecule D. If one denotes by a, b, c, and d the concentrations of these species, the reaction reduces the concentrations a and b of A and B and increases the concentrations c and d of C and D. Then we can write

$$\dot{a} = -g(a,b), \quad \dot{b} = -g(a,b), \quad \dot{c} = g(a,b), \quad \dot{d} = g(a,b),$$

where $g(a,b)$ is the reaction speed. An accepted assumption is that $g(a,b) = \kappa_{AB}\, ab$ (mass-action kinetics), with κ_{AB} constant. In general mass-action kinetics means that the reaction

$$pA + qB \to \quad \text{whatsoever},$$

where p molecules of A and q molecules of B are combined to produce some reaction products, has reaction speed $\kappa_{AB} a^p b^q$, with κ_{AB} constant.

Several reactions combined together form a chemical reaction network.

Example 12.7 (Chemical reaction chain with feedback). Consider the family of reactions

$$\emptyset \to A \to B \to C \to D \to \emptyset, \quad A + D \to \emptyset$$

We denoted by A a chemical species introduced from the external environment with rate a_0. In turn, A generates B, with rate $g_a(a)$; B generates C, with rate $g_b(b)$; C finally generates D with rate $g_c(c)$. D degrades with rate $g_d(d)$ and combines again with A thus consuming it and itself with rate $g_{ad}(a,d)$ so creating a negative feedback for its own production. The corresponding equations are the following

$$\begin{aligned}
\dot{a} &= a_0 - g_a(a) - g_{ad}(a,d) \\
\dot{b} &= g_a(a) - g_b(b) \\
\dot{c} &= g_b(b) - g_c(c) \\
\dot{d} &= g_c(c) - g_{ad}(a,d) - g_d(d)
\end{aligned}$$

It is assumed that all the functions $g_\cdot(\cdot)$ are defined for positive arguments, are smooth and strictly increasing with positive partial derivatives for positive arguments. We also assume that $g = 0$ if and only if any of its arguments is 0, otherwise it is positive.

A chemical reaction network, as the previous one, is generally modeled as

$$\dot{x} = Sg(x) + g_0,$$

where S is the stoichiometric matrix, $g(x)$ is the vector of the reaction rate functions, g_0 is the external input. In the case of the example

$$S = \begin{bmatrix} -1 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 \end{bmatrix}, \quad g = \begin{bmatrix} g_a \\ g_b \\ g_c \\ g_{ad} \\ g_d \end{bmatrix}, \quad g_0 = \begin{bmatrix} a_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

It is reasonable to assume that, for positive arguments $x > 0$ (componentwise), we have

$$\frac{\partial g_i}{\partial x_j} > 0$$

A typical question for a network is whether it is possible to infer some properties based on its structure, without the explicit knowledge of the function g.

In particular, the following questions are typical.

- Is the overall solution of the system non-negative for non-negative initial conditions?
- Is the overall solution of the system bounded?
- Assuming there exists an equilibrium point, is it (at least locally) stable?

A set-theoretic approach can be of help in answering these questions.

First of all we immediately stress that the first question should be better named "dogma," in the sense that positivity of the concentrations is necessary for the model to be meaningful, therefore a preliminary positivity test on the system is mandatory. Once positivity is checked, boundedness and stability can be considered.

Example 12.8 (Chemical reaction chain with feedback continued). Consider again the previous example and let us prove positivity and boundedness altogether. Let us sequentially define

- $\hat{a}$ the unique solution of the equation $a_0 - g_a(a) = 0$;
- $\hat{b}$ the unique solution of the equation $g_a(\hat{a}) - g_b(b) = 0$;
- $\hat{c}$ the unique solution of the equation $g_b(\hat{b}) - g_c(c) = 0$;
- $\hat{d}$ the unique solution of the equation $g_c(\hat{c}) - g_d(d) = 0$;

It will now be shown that the box defined by the inequalities

$$0 \le a \le \hat{a}, \quad 0 \le b \le \hat{b}, \quad 0 \le c \le \hat{c}, \quad 0 \le d \le \hat{d},$$

is positively invariant. On the lower bounds, "left inequality" Nagumo's conditions are immediate:

$$a = 0 \Rightarrow \dot{a} > 0,\ b = 0 \Rightarrow \dot{b} \ge 0,\ c = 0 \Rightarrow \dot{c} \ge 0 \text{ and } d = 0 \Rightarrow \dot{d} \ge 0$$

Note that this proves that the positive orthant is positively invariant.

As far as the upper bounds are concerned, the "right inequality" Nagumo's conditions still hold. Indeed

$$a = \hat{a} \Rightarrow \dot{a} = -g_{ad}(\hat{a}, d) \le 0$$

and

$$b = \hat{b} \Rightarrow \dot{b} = g_a(a) - g_b(\hat{b}) \le 0, \text{ for } 0 \le a \le \hat{a}.$$

Similarly

$$c = \hat{c} \Rightarrow \dot{c} \le 0, \text{ for } 0 \le b \le \hat{b},$$

and

$$d = \hat{d} \Rightarrow \dot{d} = g_c(c) - g_d(\hat{d}) - g_{ad}(a, \hat{d}) \le g_c(c) - g_d(\hat{d}) \le 0, \text{ for } 0 \le c \le \hat{c}.$$

It follows that an equilibrium point exists inside this positively invariant box. Denote by $(\bar{a}\ \bar{b}\ \bar{c}\ \bar{d})^T$ such an equilibrium.

This equilibrium must be component-wise positive for $a_0 = const > 0$. By contradiction, assume $\bar{a} = 0$; then $\dot{a} = a_0 > 0$, thus $\bar{a} > 0$. In the same way if $\bar{b} = 0$, then $\dot{b} = g_a(\bar{a}) > 0$, thus $\bar{b} > 0$. Similarly we can see that $\bar{c} > 0$ and $\bar{d} > 0$.

We are now in the position of checking whether this equilibrium is stable. To this aim (see [BG14]) the system is absorbed in a differential inclusion as follows (see Subsection 2.1.2). Write

$$g_a(a) - g_a(\bar{a}) = \alpha(a)(a - \bar{a}), \quad g_b(b) - g_b(\bar{b}) = \beta(b)(b - \bar{b}),$$

and

$$g_c(c) - g_c(\bar{c}) = \gamma(c)(c - \bar{c}), \quad g_d(d) - g_d(\bar{d}) = \delta(d)(d - \bar{d}).$$

Moreover

$$g_{ad}(a, d) - g_{ad}(\bar{a}, \bar{d}) = \mu(a, d)(a - \bar{a}) + \nu(a, d)(d - \bar{d}),$$

where the functions $\alpha(a) > 0$, $\beta(b) > 0$, $\gamma(c) > 0$, $\delta(d) > 0$, $\mu(a,d) > 0$ and $\nu(a,d) > 0$ are strictly positive and upper bounded inside the box. Then the system in the variables $x_a = a - \bar{a}$, $x_b = b - \bar{b}$, $x_c = c - \bar{c}$ and $x_d = d - \bar{d}$ can be absorbed in a differential inclusion with matrix

$$A = \begin{bmatrix} -(\alpha+\mu) & 0 & 0 & -\nu \\ \alpha & -\beta & 0 & 0 \\ 0 & \beta & -\gamma & 0 \\ -\mu & 0 & \gamma & -(\delta+\nu) \end{bmatrix}$$

This matrix is weakly column-diagonally dominant hence the 1-norm $\|\cdot\|_1$ is a polyhedral Lyapunov function in the weak sense, namely not increasing.

We assume that the following uniform bound is satisfied:

$$0 < \rho \leq \alpha, \beta, \gamma, \delta, \mu, \nu \leq \kappa$$

The upper bound is obvious. With a deeper discussion we could convince the reader that also the lower bound is true. Consider, for instance, the function g_{ad}

$$g_{ad}(a,d) - g_{ad}(\bar{a},\bar{d}) = \underbrace{\frac{g_{ad}(a,d) - g_{ad}(\bar{a},d)}{(a-\bar{a})}}_{=\mu}(a-\bar{a}) + \underbrace{\frac{g_{ad}(\bar{a},d) - g_{ad}(\bar{a},\bar{d})}{(d-\bar{d})}}_{=\nu}(d-\bar{d})$$

and notice that the terms μ and ν are uniformly lower bounded for a,b, c, $d > \rho$ small and positive.

Note that matrix A is irreducible and then, proceeding along the lines of Theorem 4.60, it is possible to apply the transformation $\hat{A} = T^{-1}AT$ with $T^{-1} = \mathrm{diag}\{1, 1, \lambda_3, \lambda_4\}$ so as to get the matrix

$$\hat{A} = \begin{bmatrix} -(\alpha+\mu) & 0 & 0 & -\nu/\lambda_4 \\ \alpha & -\beta & 0 & 0 \\ 0 & \beta\lambda_3 & -\gamma & 0 \\ -\mu\lambda_4 & 0 & \gamma\lambda_4/\lambda_3 & -(\delta+\nu) \end{bmatrix}$$

It is immediately seen that, if we take $0 < \lambda_4 < \lambda_3 < 1$, this matrix becomes strictly diagonally dominant, provided that λ_4 is close enough to 1, in such a way that $\delta + \nu > \nu/\lambda_4$ so that the element 4–4 "dominates" the fourth column for any choice of the parameters. Hence the 1-norm $\|\cdot\|_1$ is a strict Lyapunov function and the equilibrium point is thus stable.

For a more general discussion about the construction of Lyapunov functions for biochemical networks, the reader is referred to [BF11, BG14].

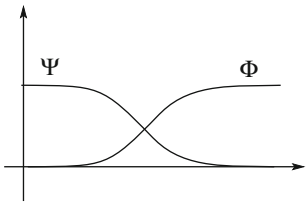


Fig. 12.12 A sigmoid Φ and a complementary sigmoid Ψ.

Example 12.9 (Biochemical oscillator). Consider the following model:

$$\begin{aligned}
\dot{x}_1 &= -\alpha_1 x_1 + \Phi(x_4) \\
\dot{x}_2 &= -\alpha_2 x_2 + \beta_2 x_1 \\
\dot{x}_3 &= -\alpha_3 x_3 + \Psi(x_2) \\
\dot{x}_4 &= -\alpha_4 x_4 + \beta_4 x_3
\end{aligned}$$

where the constants are all positive. The two functions Φ and Ψ are a smooth sigmoidal function and complementary sigmoidal function, respectively. A function Φ is sigmoidal if $\Phi(0) = 0$, $\Phi'(0) = 0\Phi(\infty) = \kappa > 0$, it is strictly increasing and its second derivative has a single positive root $\hat{x}$: $\Phi''(x)$ is positive for $0 < x < \hat{x}$ and negative for $x > \hat{x}$ (see Fig. 12.12). A function Ψ is a complementary sigmoid if $\Psi(0) - \Psi(x)$ is a sigmoid (Fig. 12.12). The model represents the negative feedback of two sub-systems: system x_1–x_2 activates the sub-system x_3–x_4, which in turn inhibits the former. It is well known that this type of systems may produce sustained oscillations.

The first step, to see if the model is a potential oscillator, is to investigate if it is a positive system and if its solutions are bounded. To prove positivity and boundedness we show that the simplex

$$\mathcal{S} = \left\{ x : \ \ x_i \geq 0, \quad c^T x \leq \rho \right\},$$

where

$$c^T = [\, 1 \ \ \mu \ \ 1 \ \ \nu \,],$$

is positively invariant for ρ large and for a suitable constants μ and ν. It is rather easy to see that for $x_i = 0$ we have $\dot{x}_i \geq 0$. Now we have to show that the Nagumo's conditions are met on the upper bound $c^T x = \rho$, for large ρ. We have

$$\begin{aligned}
&\frac{d}{dt} c^T x = \frac{d}{dt} \left(x_1 + \mu x_2 + x_3 + \nu x_4 \right) = \\
&= -(\alpha_1 - \mu\beta_2) x_1 - \mu\alpha_2 x_2 - (\alpha_3 - \nu\beta_4) x_3 - \nu\alpha_4 x_4 + \Phi(x_4) + \Psi(x_2)
\end{aligned}$$

Now, if we take $\mu, \nu > 0$ small enough so that $(\alpha_1 - \mu\beta_2) > 0$ and $(\alpha_3 - \nu\beta_4) > 0$ and define the positive vector $d^T = [(\alpha_1 - \mu\beta_2)\ \mu\alpha_2\ (\alpha_3 - \nu\beta_4)\ \nu\alpha_4 x_4]$, the derivative of the co-positive function $V(x) = c^T x$ is

$$\frac{d}{dt}c^T x = -d^T x + \Phi(x_4) + \Psi(x_2).$$

Since the term $\Phi(x_4) + \Psi(x_2) \leq \Phi(\infty) + \Psi(0)$ is bounded, and since c and d are positive vectors, for $c^T x = \rho$ large enough, the expression is negative (see Exercise 4).

Boundedness implies the existence of an equilibrium [Srz85, Hal88, RW02] and it can be seen that the system admits a single equilibrium point. Indeed, equating $\dot{x}_i = 0$ and eliminating x_1 and x_3, we get the following conditions

$$x_2 = \frac{\beta_2}{\alpha_1\alpha_2}\Phi(x_4), \quad x_4 = \frac{\beta_4}{\alpha_3\alpha_4}\Psi(x_2)$$

for the unique equilibrium (the two functions Φ and Ψ are, respectively, increasing and decreasing). The stability analysis for this system requires more information about the curves. This can in turn be done by considering the expression of the Jacobian and its characteristic polynomial

$$p(s) = \prod_{i=1}^{4}(s + \alpha_i) - \beta_2\beta_4\Psi'\Phi',$$

where Ψ' and Φ' are the derivatives evaluated at the equilibrium. Noticing that $-\beta_2\beta_4\Psi'\Phi' > 0$, we see that the characteristic polynomial has positive coefficients. Therefore unstable roots can be only complex conjugate (as expected in an oscillator).

Basically, this system is potentially an oscillator, but it does not have necessarily sustained oscillations. The oscillatory behavior is assured if $-\beta_2\beta_4\Psi'\Phi' > 0$ and this term is large enough. This in turn depends on the slopes Ψ' and Φ' of the curves Ψ and Φ at the intersection.

12.5 Monotone systems

Monotone systems form an important class. They arise in several contexts, including biology. Consider the system

$$\dot{x}(t) = f(x(t), u(t)), \quad y(t) = g(x(t))$$

and denote by $\varphi(u, x_0, t)$ its solution with input u and initial state x_0.

Definition 12.10. The system is said input–output monotone if

- given two initial conditions $x_a \leq x_b$ and two input functions $u_a \leq u_b$, both inequalities intended component-wise, then $\varphi(u_a, x_a, t) \leq \varphi(u_b, x_b, t)$;
- for $x_a \leq x_b$ we have $g(x_a) \leq g(x_b)$.

The above definition includes the special case of monotone systems in which there are no inputs or outputs.

The definition can be generalized if we define the order with respect to a convex cone. Given a convex cone $\mathcal{C}$, centered in 0, with a non-empty interior, we say that x_b is greater than x_a with respect to the partial order induced by $\mathcal{C}$, with abuse of notations $x_a \leq x_b$, if $x_b - x_a \in \mathcal{C}$.

The monotonicity defined in Definition 12.10 is just monotonicity with respect to the positive orthant (which is of course a convex cone).

A linear continuous-time system is input–output monotone iff A is a Metzler matrix and B, C are non-negative. In the linear discrete-time case, matrix A, B and C must be non-negative.

The following Theorem holds.

Theorem 12.11 ([Smi95]). *The system $\dot{x}(t) = f(x(t))$, with f defined in a convex domain, is monotone iff it satisfies the Kamke–Muller conditions: given two vectors $x \leq y$,*

$$f_i(x) \leq f_i(y), \quad \forall \ i \ \text{such that} \ x_i = y_i$$

Under smoothness assumption on f, a necessary and sufficient condition is that the system Jacobian $J(x)$ is a Metzler matrix at any point x, $\partial f_k(x)/\partial x_h \geq 0$ for $k \neq h$.

The first claim comes from Nagumo's theorem. Let $y(0) \geq x(0)$ and denote by $z(t) = y(t) - x(t)$. Monotonicity requires that the z dynamic system

$$\dot{z}(t) = f(z(t) + x(t)) - f(x(t))$$

is positive, namely $z(0) \geq 0$ implies $z(t) \geq 0$, for all $t \geq 0$. For the second claim the reader is referred to [Smi95].

Monotone systems are often referred to as cooperative, in the sense that there is a non-negative interaction between each pair of variables. This is basically the meaning of condition $\partial f_k(x)/\partial x_h \geq 0$ for $k \neq h$. They have a lot of remarkable properties.

Proposition 12.12. *Let $\dot{x}(t) = f(x(t))$ be a monotone system. Let $\bar{x}$ be an equilibrium point, i.e. $0 = f(\bar{x})$. The two sets*

$$\mathcal{S}_L = \{x \leq \bar{x}\}, \quad \mathcal{S}_U = \{x \geq \bar{x}\}$$

are both positively invariant.

No proof is required being the proposition an immediate consequence of the definition, because $x(t) \equiv \bar{x}$ is a trajectory of the system. Note that if there are two equilibria $x^- \leq x^+$, the corresponding box $\mathcal{S} = \{x : x^- \leq x \leq x^-\}$ is positively invariant.

A simple extension of the previous result is the following.

Proposition 12.13. *Assume that $\dot{x}(t) = f(x(t), u(t))$, with output $y = x$, is input–output monotone. Assume that $u^- \leq u \leq u^+$. If x^- and x^+ are equilibria corresponding to u^- and u^+, respectively (so $x^- \leq x^+$), then the set $\mathcal{S} = \{x : x^- \leq x \leq x^-\}$ is robustly positively invariant for all $u^- \leq u(t) \leq u^-$.*

Another remarkable fact is that in some cases uniqueness of the equilibrium implies its global stability. Roughly speaking, under mild assumptions, if a monotone system has globally bounded solutions and it has a single equilibrium, then this equilibrium is globally stable. We refer to specialized literature for further details [Smi95].

Example 12.14 (Biochemical switch). Consider the following model.

$$\begin{aligned}
\dot{x}_1 &= -\alpha_1 x_1 + \Phi(x_4) + u \\
\dot{x}_2 &= -\alpha_2 x_2 + \beta_2 x_1 \\
\dot{x}_3 &= -\alpha_3 x_3 + \Psi(x_2) \\
\dot{x}_4 &= -\alpha_4 x_4 + \beta_4 x_3
\end{aligned}$$

where the constant are all positive. For the moment, let us assume $u = 0$. The functions Φ and Ψ are now both sigmoidal functions, as defined in Example 12.9. Indeed in some sense this system is the dual of that in Example 12.9. Here the two sub-systems are in a positive feedback: sub-system x_1–x_2 activates sub-system x_3–x_4, which activates the former. This type of systems may produce bi-stability.

The solutions of this system are bounded. This can be proved exactly as it has been done in Example 12.9. Boundedness implies the existence of an equilibrium which is not necessarily unique since the two steady-state equations

$$x_2 = \frac{\beta_2}{\alpha_1 \alpha_2} \Phi(x_4) = \bar{\Phi}(x_4), \quad x_4 = \frac{\beta_4}{\alpha_3 \alpha_4} \Psi(x_2) = \bar{\Psi}(x_2)$$

may have multiple intersections. The stability analysis for this system requires again more information about the curves. To this aim, we write the expression of the Jacobian and we compute the characteristic polynomial

$$p(s) = \prod_{i=1}^{4} (s + \alpha_i) - \beta_2 \beta_4 \Psi' \Phi',$$

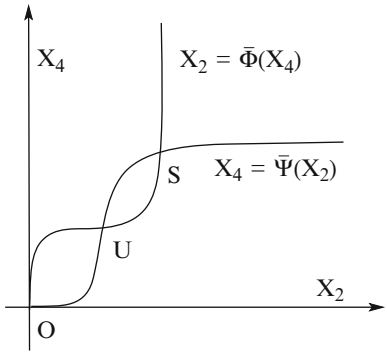


Fig. 12.13 The bistable case: equilibrium conditions

where Ψ' and Φ' are the derivatives evaluated at the equilibrium. Noticing that $-\beta_2\beta_4\Psi'\Phi' < 0$, we see that the characteristic polynomial has a negative constant term.

More interestingly, the Jacobian has non-negative non-diagonal entries and therefore the system is monotone. We consider the intersecting case in which the system has three equilibria, precisely the intersections of the curves are as in Fig. 12.13. The three equilibria denoted by O, the zero equilibrium, U (unstable as we will see soon) and S (stable) are ordered component wise.

$$x_O \leq x_U \leq x_S.$$

the inequality are trivially seen for components x_2 and x_4, since both curves Ψ and Φ are increasing. The same property holds for components x_1 and x_3 since they are increasing function of x_2 and x_4 at steady state.

It is an exercise to see that the equilibrium 0 is stable, since the Jacobian evaluated at 0 is a diagonal matrix with negative entries on the diagonal (note that $\Psi'(0) = 0$ and $\Phi'(0) = 0$).

We wish to show that U and S are unstable and stable respectively. We need the following proposition

Proposition 12.15. *A Metzler matrix is stable (i.e. it has negative real part eigenvalues) iff the coefficients of the characteristic polynomial are positive.*

The condition is well know to be necessary. To prove sufficiency we just need to remind that a Metzler (stable) matrix always has a real (negative) dominant eigenvalue (Theorem 4.65). If we assume that the coefficients of the characteristic polynomial $p(s)$ are all positive, it is readily seen that $p(s)$ cannot have non-negative (real) roots, hence it is a stable polynomial.

In view of the above proposition, it is thus possible to conclude that the stability of the characteristic polynomial (hence local stability) depends only on the sign of the constant term

$$p_0 = \prod_{i=1}^{4} \alpha_i - \beta_2\beta_4\Psi'\Phi'.$$

We show in turn that the sign of this term depends on the type of the intersection between the two curves in Fig. 12.13.

Consider point U. In such a point the slope of the Ψ curve is larger (roughly "more increasing") than the slope of the Φ curve (thought as a function of x_2). If we consider the inverse function

$$x_4 = \Phi^{-1}\left(\frac{\alpha_1\alpha_2}{\beta_2}x_2\right),$$

this means that

$$\frac{\beta_4}{\alpha_3\alpha_4}\Psi' > \frac{\alpha_1\alpha_2}{\beta_2}[\Phi']^{-1},$$

which in turn implies that $p_0 < 0$. So the equilibrium is unstable.

In the same way one can see that the equilibrium point S is stable, since we have $p_0 > 0$.

From Proposition 12.12 it is immediate to see that the equilibrium points characterize three positively invariant sets

$$\mathcal{P}_1 = \{x : \ 0 \le x \le x_U\}, \quad \mathcal{P}_2 = \{x : \ x_U \le x \le x_S\}, \quad \mathcal{P}_3 = \{x : \ x_S \le x\}.$$

To go on with our analysis, we need to explain why this is a toggle switch. Consider now the input u and take it such that

$$u(t) = \mu, \ \text{ for } 0 \le t \le T, \quad \text{and} \quad u(t) = 0, \ \text{ for } t > T.$$

We can show that

- for T and $\mu > 0$, both large enough, the state reaches the upper region $\mathcal{P}_3$;
- the state remains in $\mathcal{P}_3$ even after T.

Let $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4)$ be the steady-state values corresponding to the stable upper equilibrium S (region $\mathcal{P}_3$ is defined by $x_i \ge \bar{x}_i$).

To show the first property, note that, for μ and T large, x_1 grows large because

$$\dot{x}_1 \ge -\alpha_1 x_1 + u$$

and stays above $\bar{x}_1$ for an arbitrarily large period. Since

$$\dot{x}_2 = -\alpha_2 x_2 + \beta_2 x_1,$$

also x_2 grows arbitrarily large above $\bar{x}_2$ for an arbitrary large period, if T is large. Hence $\Psi(x_2)$ gets arbitrarily close to the saturation value $\Psi(\infty)$ and, since

$\Psi(\infty) > \Psi(\bar{x}_2)$, the condition $\Psi(x_2) - \Psi(\bar{x}_2) > 0$ is reached and preserved for a large period. From the equation (adding the zero term $\alpha_3\bar{x}_3 - \Psi(\bar{x}_2)$)

$$\frac{d}{dt}(x_3 - \bar{x}_3) = -\alpha_3 x_3 + \Psi(x_2) = -\alpha_3(x_3 - \bar{x}_3) + (\Psi(x_2) - \Psi(\bar{x}_2)),$$

we see that in due time $x_3 > \bar{x}_3$ and this condition will hold for large time values. Finally from the equation

$$\frac{d}{dt}(x_4 - \bar{x}_4) = -\alpha_4(x_4 - \bar{x}_4) + \beta_4(x_3 - \bar{x}_3)$$

we argue that x_4 will exceed $\bar{x}_4$. So the upper region $\mathcal{P}_3$ is reached.

Region $\mathcal{P}_3$ is positively invariant for $u = 0$, therefore the state will remain inside $\mathcal{P}_3$ even if u is switched off. Note that indeed the region is invariant for any $u \geq 0$, in view of the system monotonicity, hence $x(t)$ is trapped in $\mathcal{P}_3$ once this region is reached.

For pictorial reasons (and as an exercise for the reader), we suggest a simplified version of the problem in which the dynamics of x_2 and x_4 are fast enough to assume $-\alpha_2 x_2 + \beta_2 x_1 = 0$ and $-\alpha_4 x_4 + \beta_4 x_3 = 0$. This basically means accepting the following reduced-order model

$$\dot{x}_1 = -\frac{\alpha_1\alpha_2}{\beta_2}x_2 + \Phi(x_4) + u$$
$$\dot{x}_3 = -\frac{\alpha_3\alpha_4}{\beta_4}x_4 + \Psi(x_2)$$

The equilibrium value for x_2 and x_4 would be the same and, again, three positively invariant regions, represented in Fig. 12.14, could be found. In Fig. 12.14 it is also represented the fact that an active input can "switch on" the system, which remains in such an activated state even after u becomes inactive.

Further details about this type of models can be found in [FB12].

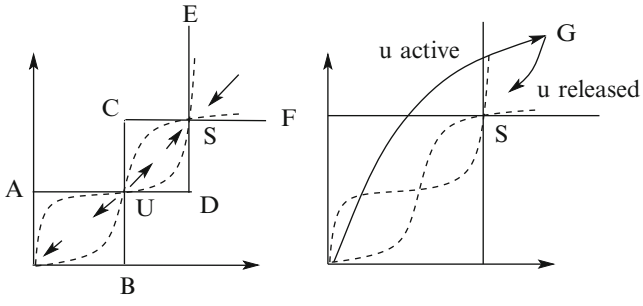


Fig. 12.14 The invariant boxes (left) and the toggle switch (right)

12.6 Communication and network problems

The control of networks is a really interesting area for control researchers and the set-theoretic framework can be quite helpful when dealing with network control problems. This seems surprising, since the dimension of realistic flow networks is often large and thus it is not so clear how the techniques presented in this book, which are often computationally demanding, might be of some benefit.

Indeed, flow networks or production–distribution systems have in general a large number of control variables, often exceeding the number of states, a fact which renders the problem peculiar and challenging at the same time.

12.6.1 Production–distribution systems

A typical model for production–distribution systems has the following form:

$$\dot{x}(t) = Bu(t) + Ed(t), \tag{12.18}$$

where $u \in \mathbb{R}^m$ is the controlled flow vector, B is the controlled flow matrix, $d(t) \in \mathbb{R}^q$ is the external input (typically demand or disturbances), and E is the corresponding input matrix. Note that the state dynamic matrix A is zero, namely the system is driftless. The state vector $x(t) \in \mathbb{R}^n$ represents the buffer levels or inventories.

We always assume that $m \geq n$, otherwise the system wouldn't be stabilizable. More precisely, a necessary and sufficient condition for stabilizability is that B has full row rank.

Clearly finding a stabilizing feedback control for this application is trivial. For instance, $u = -\gamma B^T x$, for any $\gamma > 0$, is a stabilizing control.

The problem becomes more interesting if we consider constraints on control and state variables and if we assume that d can be uncertain.

Example 12.16. Consider the network represented in Fig. 12.15, whose dynamic equation is (12.18), with

$$B = \begin{bmatrix} -1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Plain arcs represent controlled flows, dashed arcs uncontrolled flows. The boxes represent warehouses. This situation describes the flow of a resource, such as water, which is naturally supplied to the system through arc d_1 (uncontrolled) and artificially supplied, in case of shortage, through arc u_8 (controlled). The flow is to

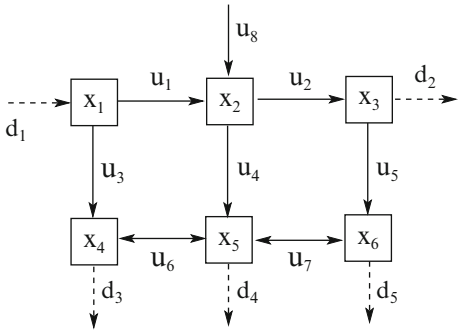


Fig. 12.15 A simple distribution system

be distributed by the controlled arcs $u_1 \dots u_7$ to satisfy the (uncontrolled) demands $d_2 \dots d_5$. It is legitimate to assume that state and control variables have known upper and lower bounds

$$x_i^- \leq x_i \leq x_i^+, \quad u_j^- \leq u_j \leq u_j^+.$$

As far as the uncontrolled flow in most cases this is not known but it can vary. A possible assumption is that

$$d_k^- \leq d_k \leq d_k^+,$$

with known bounds.

The following problem naturally arises. Assume that we are assigned the bounds

$$x \in \mathcal{X} = \{x : x^- \leq x \leq x^+\}, \qquad u \in \mathcal{U} = \{u : x^- \leq u \leq u^+\}, \ \ d \in \mathcal{D},$$

with $\mathcal{D}$ a convex polytope and with $\mathcal{X}$ and $\mathcal{U}$ having a non-empty interior. Does there exist a feedback control which keeps the state within the constraints and satisfies the flow constraints?

The following necessary and sufficient condition holds [BMU00].

Proposition 12.17. *There is a control which keeps $x(t)$ in the interior of $\mathcal{X}$, at least starting from a proper set of initial conditions $\mathcal{X}_0$, if and only if*

$$E\mathcal{D} \subset int\{B\mathcal{U}\}$$

We provide a simple argument to explain the result, by means of the Lyapunov function $\Psi(x) = \|x\|_2$, the Euclidean norm. Assume without restriction that $0 \in int\{\mathcal{X}\}$. Consider the following discontinuous (actually bang-bang) control law:

$$u_{BB}(x) = \arg\min_{u \in \mathcal{U}} x^T B u$$

The Lyapunov derivative is (we drop the "2")

$$\dot{\Psi}(x) = \frac{x^T}{\|x\|}\left[Bu_{BB}(x) + Ed\right] \leq \frac{x^T}{\|x\|}Bu_{BB}(x) + \max_{d\in\mathcal{D}}\frac{x^T}{\|x\|}Ed$$

On the other hand, the last term is

$$\min_{u\in\mathcal{U}}\frac{x^T}{\|x\|}Bu + \max_{d\in\mathcal{D}}\frac{x^T}{\|x\|}Ed \leq -\beta, \quad \text{for some } \beta > 0$$

where the last inequality comes from the assumption $E\mathcal{D} \subset int\{B\mathcal{U}\}$.

Indeed the proposition holds even if we assume that $\mathcal{D}$ and $\mathcal{U}$ are polytopes. In this case, the special form of $\mathcal{U}$ allows us to show a remarkable property. Indeed, given two vectors $a < b$ define the following function componentwise

$$\sigma_{a,b}(\xi) = \begin{cases} a_i & \text{if } \xi_i > 0 \\ b_i & \text{if } \xi_i < 0 \\ \text{any value in } [a_i, b_i] & \text{if } \xi_i = 0 \end{cases}$$

Then

$$u_{BB}(x) = \sigma_{u^-,u^+}(B^T x)$$

A relevant fact is that any control component u_i makes its decision based on the quantity $B_j^T x$, where B_j is the jth column of B. This control is decentralized in the sense of networks [ID90, Ift99, BRU97, BMU00, BDVP09], namely each control agent u_k makes its decision based only on the buffers which are directly affected by it. For instance, in the example, u_1 decides its action based on the knowledge of the two variables $x_1(t)$ and $x_2(t)$, u_2 decides its action based on the knowledge of the two variables $x_2(t)$ and $x_3(t)$ and so on.

Example 12.18. Consider again the network represented in Fig. 12.15, with bounds on u given by

$$u^+ = [\ 4\ \ 4\ \ 3\ \ 3\ \ 3\ \ 2\ \ 2\ \ 10\ ]^T \qquad u^- = [\ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ -2\ \ -2\ \ 0\ ]^T$$

and bounds on d given by

$$d^+ = [\ 4\ \ 2.5\ \ 3.5\ \ 1.5\ \ 2\ ]' \qquad d^- = [\ 2.5\ \ 1.5\ \ 1.5\ \ 0\ \ 1\ ]'.$$

Take as initial condition

$$x(0) = [\ 2\ \ -1\ \ 0\ \ 2\ \ -1\ \ 1\ ]^T.$$

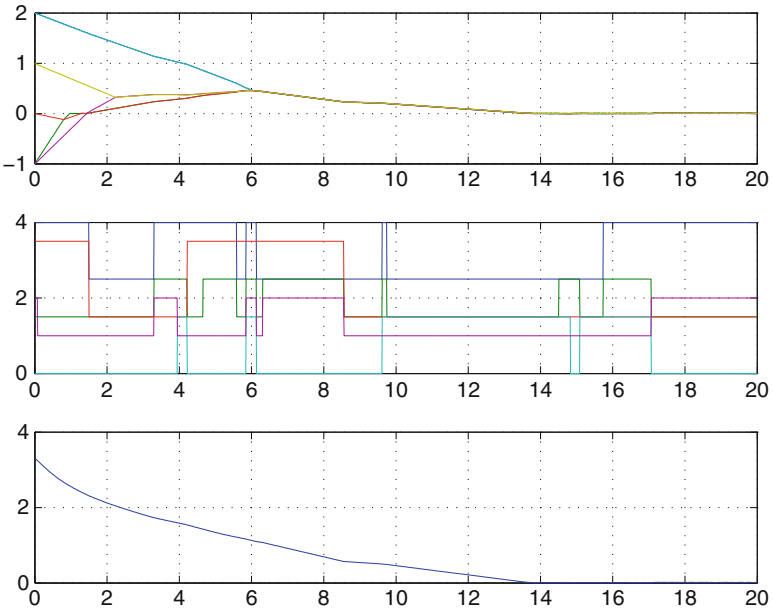


Fig. 12.16 The network transient: the buffer evolution (top), the disturbance d (middle), and the norm evolution (bottom).

It is clear that these initial levels are referred to a desired reference level. The transient, assuming random disturbances at the extrema, is represented in Fig. 12.16.

12.6.2 P-persistent communication protocol

Consider the problem of n transmitters (nodes) sharing the same channel which has a maximum capacity normalized at 1. The rule of the game is that each transmitting node has to regulate its own transmission rate x_i based on band occupancy. The goals are that the overall band has to be (almost) exploited; the regulation has to be fully decentralized; the maximum band has not to be exceeded. The following scheme can be considered [BDCMP12]:

$$\dot{x}_i(t) = -\alpha[(1+\epsilon)x_i(t) + z_i(t) - 1]$$

where

$$z_i(t) = \sum_{j \neq i} x_j(t)$$

is the complementary transmission rate. This control law is implemented at each node and is based on the knowledge of the local transmission rate x_i and of the aggregate transmission rate z_i, namely, the total transmission rate of all other nodes.

Since at steady state we have $x_i = x_j$ for all i and j, the steady state condition is

$$\bar{x}_i = \frac{1}{n+\epsilon}$$

where n is the number of nodes (unknown to each node). The details of the implementation are discussed in [BDCMP12]. We limit here ourselves to discuss the properties of the control.

For $\epsilon > 0$ small the goal is achieved at steady state. We wish to analyze what is happening during the transient. Consider the Lyapunov like function

$$y(t) = \sum_{j=1}^{n} x_j(t),$$

namely the overall transmission rate. Then

$$\dot{y} = -\alpha \sum_{j=1}^{n} [(1+\epsilon)x_j(t) + z_j(t) - 1] = -\alpha[(n+\epsilon)y - n]$$

and $y(t) \to n/(n+\epsilon) < 1$. Thus for ϵ small the band is almost fully exploited. By means of the same Lyapunov function it is possible to see that the bound $y \leq 1$ is never violated whenever $y(0) \leq 1$, because $\dot{y} < 0$ when $y = 1$. It can indeed be shown that the set

$$\mathcal{S} = \{x \geq 0 : y = \sum_{j=1}^{n} x_j \leq 1\}$$

is positively invariant. Since $y \leq 1$ cannot be violated, it is sufficient to show that $x_i \geq 0$ is not violated. This is in turn immediate, since

$$\dot{x}_i = -\alpha[z_i - 1] \geq 0$$

when $x_i = 0$.

Finally, to show that the system state actually converges to the equilibrium, consider the squared variance as a Lyapunov-like function

$$V = \sum_{i=1}^{n} (x_i - y/n)^2.$$

Then, exploiting the expression of $\dot{y}$,

$$\begin{aligned}\dot{V} &= 2\sum_{i=1}^{n}(x_i - \frac{y}{n})(\dot{x}_i - \frac{\dot{y}}{n}) \\ &= -2\alpha\sum_{i=1}^{n}(x_i - \frac{y}{n})\left[\epsilon x_i(t) + y(t) - 1 - \frac{(n+\epsilon)y - n}{n}\right] \\ &= -2\alpha\epsilon\sum_{i=1}^{n}(x_i - \frac{y}{n})\left[x_i(t) - \frac{y}{n}\right] = -2\alpha\epsilon V.\end{aligned}$$

Then $V \to 0$. It should be noticed that, while the band exploitation y converges quickly, if ϵ is small the variance converges slowly. Still the system tends to the fairness condition, i.e. $V = 0$ not only at steady state, but even under transient.

12.6.3 Clock-synchronization and consensus

We have already met the consensus problem in Section 2.5. We wish to consider the more involved problem of clock synchronization.

Suppose that we have n clocks which exchange information to synchronize their time (Fig. 12.17). This is the case of a computer network in which each machine wishes to have its internal clock synchronized with the others. Denoting by $\tau_i(t)$ the instantaneous time indication of the ith machine clock at the "true" time t and denoting by ω_i the time speed of the same machine, each isolated machine would obey the following dynamic equation

$$\dot{\tau}_i(t) = \omega_i$$

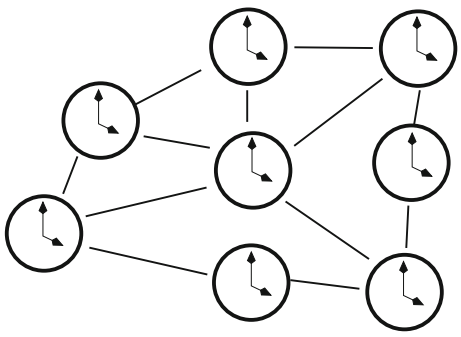


Fig. 12.17 The network synchronization problem

(mind that the derivative is taken with respect to the true time). If each clock exchanges information with some of the others, it can adapt its time and speed. A possible adaptation law is

$$\dot{\tau}_i(t) = \alpha \sum_{j \in \mathcal{C}_i} (\tau_j - \tau_i) + d_i \omega_i$$
$$\dot{\omega}_i(t) = \beta \sum_{j \in \mathcal{C}_i} (\tau_j - \tau_i)$$

where $\mathcal{C}_i$ denotes the set of indices of all the clocks which communicate with clock i and d_i represents the error in the speed. Namely if a clock considers its time variation as 1, this may result in a speed d_i measured with respect to the absolute time[5]. The resulting model is

$$\dot{\tau}(t) = -\alpha L \tau + D\omega$$
$$\dot{\omega}(t) = -\beta L \tau$$

where L is the so-called Laplacian matrix. The Laplacian matrix of a graph is a symmetric matrix of the same dimension of the system associated with the communication graph. L has positive diagonal elements, each equal to the number of edges leaving the corresponding node (the cardinality of $\mathcal{C}_i$). The non-diagonal entries are $L_{ih} = -1$ iff node i is connected to node i and $L_{ih} = 0$ otherwise. The Laplacian matrix has the property that $L\bar{1} = 0$ and that this is the only element of its kernel, up to a scaling factor, iff the graph is connected.

We assume that $D = I$ for brevity. For a more detailed study, the reader is referred to specialized literature (see [CCSZ11] and the references therein). Apply the transformation $\hat{\omega} = \sqrt{\beta}\omega$ and leave τ unchanged, so as to get the system matrix

$$\begin{bmatrix} I & 0 \\ 0 & \sqrt{\beta}^{-1} I \end{bmatrix} \begin{bmatrix} -\alpha L & I \\ -\beta L & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \sqrt{\beta} I \end{bmatrix} = \begin{bmatrix} -\alpha L & \sqrt{\beta} I \\ -\sqrt{\beta} L & 0 \end{bmatrix}.$$

Consider now the Lyapunov-like function

$$V = \tau^T L \tau + \hat{\omega}^T \hat{\omega}.$$

Then

$$\dot{V} = -2\alpha \tau^T L^2 \tau \leq 0$$

[5] We assume that clocks are resting or moving much slower than the speed of light.

and x must converge to the set in which $\tau^T L^2 \tau = \|L\tau\|^2 = 0$. If the clock communication graph is connected, this set is the subspace aligned with $\bar{1}$. This means that asymptotically

$$\tau(t) = \theta(t)\bar{1}.$$

Furthermore we also have

$$\dot{\omega}(t) = -\beta L\tau(t) \to 0$$

namely $\omega \to \bar{\omega} = \text{const}$. Since asymptotically $L\tau = 0$, then $\omega - \dot{\tau} \to 0$, hence

$$\bar{\omega} - \dot{\tau} = \bar{\omega} - \bar{1}\dot{\theta} \to 0.$$

From this condition we see that, asymptotically, $\dot{\theta}$ is constant[6] and $\bar{\omega}$ has all equal components

$$\bar{\omega} = \bar{1}\bar{a}$$

for some constant $\bar{a}$. Now consider the average value of the speeds $a(t) = \bar{1}^T\omega(t)/n$. We have for all t

$$\dot{a} = \bar{1}^T L\beta\tau/n = 0,$$

since $\bar{1}^T L = 0$. Then the average value of ω is constant and so $\bar{a} = a(0)$, namely the final speed vector has all components equal to the average of the components of the initial vector. In Fig. 12.18 we show the transient for a system with n clocks and a randomly generated graph. It is apparent that typically "time synchronization" is much faster than "speed synchronization."

12.6.4 Other applications and references

Any book must end at a certain page and, as it is expected, there is always something which could have been considered but it has not been included for space reasons.

We indeed believe that the list of possible applications and subjects which are in some way related could be at least twice as long. Consequently, we decided to briefly remind other subjects or applications which have been not considered or only marginally mentioned.

An intensive line of research concerns the so-called reachability on polytopes and [HvS04, BR06, Bro10]. The main problem considered in those references is that of

[6] Roughly $\theta(t)$ is asymptotically linear.

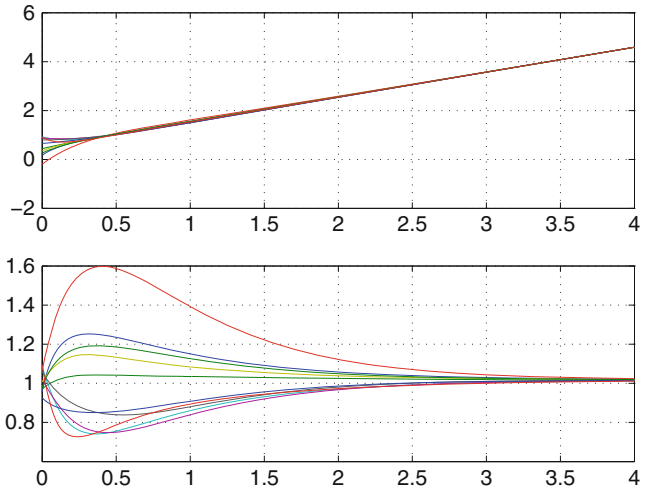


Fig. 12.18 The clock synchronization transient, with $n = 10$, $\alpha = 2$ and $\beta = 2$

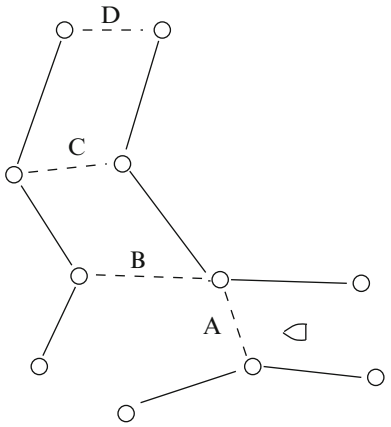


Fig. 12.19 A navigation problem

reaching a specified face of a polytope, starting from inside, without crossing any other face. An example of application is the motion among safe regions described in Section 12.3. Other possible applications can be found in the general problem of navigation with constraints, such as the problem of a natant which has to navigate in a channel as in Fig. 12.19.

Recent quite interesting applications of set-theoretic techniques are in fault detection and fault tolerant systems.

For instance, assume that we are given a system

$$\dot{x} = Ax + \sum_{j=1}^{m} B_j f_j u_j = Ax + BFu$$

where u_j is the control signal for the actuator j. The numbers f_j are fault coefficient which are grouped in a diagonal matrix F. In faulty conditions, $f_j = 0$ while the normal actuator condition is $f_j = 1$. Partial faults can be considered if we assume that f_j can be a continuous value, $0 \leq f_j \leq 1$.

Typically a fault assumption is introduced. For brevity assume that there are two actuators and only one actuator can be currently under fault. This means that the possible values of F are $F = diag\{1, 1\}$, $F = diag\{0, 1\}$ and $F = diag\{1, 0\}$. The problem of finding a stabilizing control law can be seen as a robust control problem in which the input matrix is affected by uncertainties.

Clearly the system is fault-tolerant if there is a control for which the system is stable with the three possible fault configurations. A possibility is that of determining a robust Lyapunov function, a quite stringent requirement since, in the case of linear feedback control $u = Kx$, $A + BFK$ might be stable for any admissible F even if there is no common Lyapunov function. However, a common Lyapunov function assures stability even under intermittent fault, i.e. when F changes in time. The reader is referred to specialized literature for further details [ODDSS10, SO13].

Also the fault detection and isolation problem can be handled via set-theoretic techniques. A typical approach for detecting faults is based on observers. Consider a linear system

$$\dot{x} = A_k x + B_k u + Ed, \quad y = C_k x + w,$$

where d and w are disturbances and where the usual matrices A_k, B_k, C_k are labelled with the index k. If $k = 0$, then the system is operating in the nominal condition, while $k = 1, 2, \ldots, N$ are the possible faulty conditions. If an observer is adopted

$$\dot{z} = (A_0 + LC_0)z - Ly + B_0 u,$$

it is possible to measure on-line the residual

$$r = C_0 z - y,$$

so that, defining the error $e = z - x$, its dynamics is described by

$$\begin{aligned} \dot{e} &= (A_0 + LC_0)e + (A_0 - A_k)x + (B_0 - B_k)u + L(C_0 - C_k) - Lw - Ed \\ r &= C_0 e - w + (C_0 - C_k)x \end{aligned}$$

In nominal conditions, say $k = 0$, with zero disturbances (and a properly designed observer), e and the residual converge to 0. They will get close to zero with small noises d and w.

This is not true if the system is under fault, so that the residual is far form 0. By means of set-theoretic methods and in particular those related with the set-theoretic estimation described in this book, it is in general possible to detect and even recognize (isolate) a fault. The reader is referred to specialized literature for further details [SO13].

Another topic which has not been considered, if not in passing, is that of control and analysis of time delay system. Systems with time delays are difficult to handle via set-theoretic methods because their state space is infinite dimensional, at least in the continuous-time case. In the discrete-time case it is well known that a system with time delays can be seen as an extended linear undelayed system. Some results on the definition and characterization of invariant sets for time delay systems are available since many years [HBB95, HT97, HT97] and it was still an active topic at the moment of writing the present edition of the book [LOBN12, GLO12, GOL+10].

There are several specific applications of set invariance, many of which have been described in the book. A converter control design technique has been proposed in [SALB12]. An application to engine speed control has been suggested in [DSDBP06].

From a theoretical standpoint, the investigation based on set invariance is deeply applied to nonlinear systems [FAC10, BT11] and hybrid systems [DSDBP06, CDS09, BDSCD07, DS08]. Recent results in the definition and construction of probabilistic invariant sets are found in [CKRC11, KDDS12].

12.7 Exercises

1. The considered adaptive control schemes are quite involved in order to prove boundedness of the signals to apply Barbalat's Lemma. Show that there exists non-negative continuously differentiable functions whose integral on the positive axis is bounded but which still do not converge to 0. This clearly cannot happen if the function has a bounded derivative.
2. Consider the Zubov equation for the scalar system $\dot{x} = -x + x^3$ with $\phi(x) = x^2$ and verify that the domain of attraction is $-1 < x < 1$.
3. Find the solution of (12.17).
4. Given two positive vectors $c, d > 0$ and the corresponding co-positive linear functions $c^T x$ and $d^T x$, show that for $x \geq 0$ we have, for some positive ξ and η, $\xi c^T x \leq d^T x \leq \eta d^T$. Finalize the proof of boundedness in Example 12.9.
5. Prove that (12.18) is stabilizable iff B has full row rank. If B has full row rank, show that $u = -\gamma B^T x$, for γ positive, is a stabilizing control.
6. The p-persistent protocol of Subsection 12.6.2 could have been analyzed in a simpler, although not so informative, way. The closed-loop system has equation $\dot{x} = -\alpha(\epsilon I + \bar{1}\bar{1}^T)x + \alpha\bar{1}$. Try to derive the same conclusions obtained in Section 12.6.2.
7. Try to devise an automated setup for the natant in Fig. 12.19 based on the "reachability of a face" technique.

Appendix

A.1 Remarkable properties of the Euler auxiliary system

In this section we summarize some properties of the Euler auxiliary system which have been already mentioned and used throughout the book. Given a continuous-time system

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), w(t)) \\ y(t) &= g(x(t), w(t)) \end{aligned}$$

we define the Euler Auxiliary System (EAS) as the discrete-time system

$$\begin{aligned} x(t+1) &= x(t) + \tau f(x(t), u(t), w(t)) \\ y(t) &= g(x(t), w(t)) \end{aligned}$$

where, as usual, $w(t) \in \mathcal{W}$. This system is introduced in the basic elementary analysis as the most natural approximation of a continuous-time system, based on the fact that $\dot{x}(t) \approx (x(t+\tau) - x(t))/\tau$. It is also well known that this is a very rough approximation and more appropriate numerical methods resort to much more sophisticated schemes. Still there are remarkable properties which are worth mentioning. In this section we remind a few of them.

The EAS has been often considered in the book because several synthesis techniques, essentially devoted to discrete-time systems, can be extended to the continuous-time case (often in a sub-optimal way). To avoid confusion we immediately stress that parameter τ **is not a sampling time**. In practice, assume that a (static) controller

$$u = \Phi(y(t), w(t))$$

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9

has been computed by means of a "discrete-time" technique. Then it has to be thought as a continuous-time one. If, as in practice happens, the controller is digitally applied, then it is neither requested nor recommended that the sampling time T is equal to τ. In practice we should have

$$T << \tau.$$

Note that, if we consider a dynamic compensator (which can be handled by means of an equivalent state augmentation) then if the "discrete-time" compensator is

$$\begin{aligned} z(t+1) &= h_{DT}(z(t), y(t), w(t)) \\ u(t) &= k_{DT}(z(t), y(t), w(t)) \end{aligned}$$

the actual "continuous-time" compensator must be

$$\begin{aligned} \dot{z}(t) &= \frac{h_{DT}(z(t), y(t), w(t)) - z(t)}{\tau} = h_{CT}(z(t), y(t), w(t)) \\ u(t) &= k_{DT}(z(t), y(t), w(t)) = k_{CT}(z(t), y(t), w(t)) \end{aligned}$$

In brief we have the following.

Lemma A.1. *The EAS of a continuous-time closed-loop system is the same system achieved by closing the loop of the corresponding EAS (plant and compensator) systems.*

The previous property is not true for other type of discretization techniques. We start now with some open-loop properties of the EAS. The first one establishes a connection between the stability of the continuous-time system and of the EAS.

Lemma A.2. *Assume that the EAS is stable for some $\tau > 0$ (with $u = 0$) and it admits a global convex Lyapunov function $\Psi(x)$. Then the continuous-time systems is asymptotically stable.*

Proof. It is an immediate consequence of the fact that the difference quotient is a non-decreasing function of τ [Roc70]

$$\frac{\Psi(x + hz) - \Psi(x)}{h} \leq \frac{\Psi(x + \tau z) - \Psi(x)}{\tau}$$

for $\tau \geq h$. Then if we consider the derivative

$$\limsup_{h \to 0^+} \frac{\Psi(x + hf(x(t), w(t)) - \Psi(x)}{h} \leq \frac{\Psi(x + \tau f(x(t), w(t)) - \Psi(x)}{\tau} \leq -\psi(\|x\|)$$

where, by definition of a Lyapunov function, ψ is a κ-function.

Obviously the previous result can be extended to ultimate boundedness and local stability. The converse is not true. For instance, the continuous-time system

$$\dot{x}(t) = -x(t) - x(t)^3$$

is globally exponentially stable. But if we consider its EAS

$$x(t+1) = (1-\tau)x(t) - \tau x(t)^3$$

this is not globally stable no matter how $\tau > 0$ is taken. Indeed for $x(0)$ large enough the discrete-time solution diverges.

The next result is concerned with the level of (sub-)optimality one can achieve via EAS.

Lemma A.3. *Consider the system*

$$\dot{x}(t) = f(x(t), u(t)), \quad y(t) = g(x(t))$$

associated with the cost

$$J_{CT} = \int_0^\infty h(x(t), u(t))dt,$$

where h is strictly convex and positive definite. Consider the EAS

$$x(t+1) = x(t) + \tau f(x(t), u(t)), \quad u(t) = g(x(t))$$

and the corresponding discrete-time cost

$$J_{DT} = \sum_{k=0}^{\infty} h(x(k), u(k))\ \tau.$$

Assume that the cost-to-go function[1] $\Psi_{DT}(x)$ *associated with the EAS is convex, a condition always assured if the system is linear, and that* $\Phi(x(t))$ *is the corresponding optimal control law, then this control applied to the continuous-time system assures a transient cost*

$$J_{CT} \leq \Psi_{DT}(x)$$

for every initial condition. This implies that the continuous-time cost-to-go function $\Psi_{CT}(x)$ *is upper bounded by that associated with the EAS:* $\Psi_{CT}(x) \leq \Psi_{DT}(x)$

[1]We remind that the cost-to-go function is the optimal value of the optimization with initial condition x.

To provide an outline of the proof, let us consider that, if $\Psi_{DT}(x)$ is the cost-to-go function and Φ is the optimal control, then

$$\Psi_{DT}(x(t+1)) - \Psi_{DT}(x(t)) = -\tau h(x(t), \Phi(x(t)))$$

namely

$$\frac{\Psi_{DT}(x(t) + \tau f(x(t), \Phi(x(t))) - \Psi_{DT}(x(t))}{\tau} = -h(x(t), \Phi(x(t))).$$

Now we remind that, if Ψ_{DT} is convex, the difference quotient is non-decreasing then

$$D^{+}\Psi_{DT} = \limsup_{h \to 0} \frac{\Psi_{DT}(x + hf(x, \Phi(x)) - \Psi_{DT}(x)}{h} \leq -h(x, \Phi(x)).$$

Since the function h is positive definite, this implies stability. Furthermore, we can integrate according to Theorem 2.11 and, assuming $x(0) = x_0$, we have

$$\Psi_{DT}(x(t)) - \Psi_{DT}(x_0) \leq -\int_0^t h(x(\sigma), \Phi(x(\sigma)))\, d\sigma$$

and, since $x(t) \to 0$,

$$\int_0^\infty h(x(\sigma), \Phi(x(\sigma)))\, d\sigma \leq \Psi_{DT}(x_0).$$

Thus the optimal discrete-time cost is an upper bound for the optimal continuous-time cost. The reader is referred to [BMP03] for more details including the case with convex constraints. In [BMP03] it is also shown that, under appropriate regularity assumptions, since the EAS solution converges to the continuous-time solution, then the discrete-time cost converges to the continuous-time from above.

Let us consider the exponential discretization of the continuous-time linear system

$$\dot{x}(t) = Ax(t) + Bd(t), \quad z(t) = Cx(t) + Dd(t),$$

precisely

$$x(t+1) = A_D x(t) + B_D d(t), \quad z(t) = Cx(t) + Dd(t),$$

with $A_D = e^{A\tau}$ and $B_D = \int_0^\tau e^{A\sigma} d\sigma B$. Then continuous-time asymptotic stability is equivalent to discrete-time asymptotic stability. Conversely, reachability or observability (hence stabilizability and detectability) is not assured for all $\tau > 0$ even if the continuous-time system is reachable or observable. For the EAS $A_D = I + \tau A$ and $B_D = \tau B$, the situation is the opposite. Stability of A does not imply

stability of $I + \tau A$ (unless we take $\tau > 0$ small enough). Conversely for all $\tau > 0$ reachability (observability) of the continuous-time system is equivalent to reachability (observability) of the EAS. This can be seen by means of the Popov criterion for reachability (observability).

Now we analyze some induced norms for linear systems. For brevity we consider the SISO case namely we assume that d and z are scalar. The results can be extended to the MIMO case with appropriate modifications.

Assume that A is asymptotically stable and define the following norms:

$$\|(A,B,C,D)\|_{\mathcal{L}_1} = \sup_{|d(t)|\le 1,\ x(0)=0} \sup_{t\ge 0} |y(t)|$$

$$\|(A,B,C,D)\|_{\mathcal{H}_\infty} = \sup_{\omega\ge 0} |C(j\omega I - A)^{-1}B + D|$$

$$\|(A,B,C)\|_{\mathcal{L}_\infty} = \sup_{t\ge 0} |Ce^{At}B|$$

$$\|(A,B,C)\|_{\mathcal{L}_2}^2 = \int_0^\infty \left(Ce^{At}B\right)^2 dt.$$

Define the corresponding discrete-time norms as

$$\|(A_D,B_D,C,D)\|_{l_1} = \sup_{|d(t)|\le 1,\ x(0)=0} \sup_{t\ge 0} |y(t)|$$

$$\|(A_D,B_D,C,D)\|_{\mathcal{H}_\infty} = \sup_{-\pi\le\theta\le\pi} |C(e^{j\theta}I - A_D)^{-1}B_D + D|$$

$$\|(A_D,B_D,C)\|_{l_\infty} = \sup_{t\ge 0} |CA_D^tB_D|$$

$$\|(A_D,B_D,C)\|_{l_2}^2 = \sum_{t=0}^{\infty} \left(CA_D^tB_D\right)^2 \tau$$

Then we have the following.

Lemma A.4. *If the discrete-time system is the EAS of the continuous-time one, namely $A_D = I + \tau A$ and $B_D = \tau B$ with $\tau > 0$, then all the mentioned discrete-time norms are upper bounds for the corresponding continuous-time norms for any $\tau > 0$. Furthermore for $\tau \to 0$ any of the discrete-time norms converges from above to the corresponding continuous-time value.*

Proof. The case of the $\mathcal{L}_1$-norm has been already presented in Subsection 6.4.1. Note that the result is valid also for systems with parametric uncertainties [BMS97]. The case of the $\mathcal{L}_\infty$ norm is proved in a simple way. Assume (without restriction) that (A,B) is reachable. Consider the initial conditions $x(0) = B$ and $x(0) = -B$ and let $x(k) = A^kB$ and $-x(k)$ be the corresponding free evolutions. Consider the set

$$\mathcal{R} = conv\{\pm x(k), k \ge 0\}.$$

Since (A, B) is reachable, $\mathcal{R}$ has a non-empty interior and, since A is stable, there is $\bar{k}$ such that $\pm x(k) \in \mathcal{R}$ for $k \geq \bar{k}$. Then $\mathcal{R}$ is a polytope. By construction it is positively invariant for the EAS, because $A(\pm x(k)) = \pm x(k+1) \in \mathcal{R}$, for any vertex $x(k)$. This implies that the polytope $\mathcal{R}$ is positively invariant for the continuous-time system. Then for $x(0) = B$ the continuous-time solution $x(t) = e^{At}B \in \mathcal{R}$. Now consider the strip

$$\mathcal{Y}(\mu) = \{x : \ |Cx| \leq \mu\}.$$

We have just to notice that

$$\|(A, B, C, 0)\|_{\mathcal{L}_\infty} = \sup |Ce^{At}B| = \inf\{\mu : \ e^{At}B \in \mathcal{Y}(\mu), \ t \geq 0\}$$

and that

$$\|(A_D, B_D, C)\|_{l_\infty} = \inf\{\mu : \ \mathcal{R} \subset \mathcal{Y}(\mu)\}$$

Then $\|(A, B, C)\|_{\mathcal{L}_\infty} \leq \|(A_D, B_D, C)\|_{l_\infty}$.

The $\mathcal{H}_\infty$ norm case can be handled as follows. Consider the discrete-time transfer function norm

$$\begin{aligned}\|(A_D, B_D, C, D)\|_{\mathcal{H}_\infty} &= \sup_{|z|=1} C(zI - [I + \tau A])^{-1}\tau B + D = \\ &= \sup_{-\pi \leq \theta \leq \pi} C\left(\frac{e^{j\theta} - 1}{\tau} - A\right)^{-1} B + D\end{aligned}$$

The set of all the points $(e^{j\theta} - 1)/\tau$ is a circle passing through the origin and centered in $-1/\tau$. For $\tau \to 0$ this circle "converges" to the imaginary axis. Therefore it includes all the eigenvalues of A for all $\tau < \bar{\tau}$ for $\bar{\tau} > 0$ large enough. Then the transfer function $C(sI - A)^{-1}B + D$ is analytic outside such a circle. By the maximum modulus theorem, $|C(sI - A)^{-1}B + D|$ is maximum on such a circle. Note that the imaginary axis is outside the circle (actually intersects it in $s = 0$); then

$$|C(j\omega I - A)^{-1}B + D| \leq \|(A_D, B_D, C, D)\|_{\mathcal{H}_\infty}$$

for all ω. For the proof of the bound on the $\mathcal{L}_2$ norm, the reader is referred to [SAI03] and the references therein.

The properties mentioned above have easy extensions in the case of $\mathcal{L}_1$ and $\mathcal{L}_\infty$ norms for uncertain polytopic systems.

A.2 MAXIS-G: a software for the computation of invariant sets for constrained LPV systems

Contributed by **Carlo Savorgnan**
Dipartimento di Ingegneria Elettrica Gestionale e Meccanica
Università di Udine,
33100 Udine Italy

MAXIS-G is a software that implements the algorithms presented in Chapter 5 to calculate the maximal invariant set for systems of the form

$$x(t+1) = A(w(t))x(t) + B(w(t))u(t) + Dd(t), \tag{A.1}$$

where $x(t)$, $u(t)$, and $d(t)$ are respectively, the state, the input and the disturbance vectors. The time-varying parameter $w(t)$ represents the uncertainty. The system matrices must have the polytopic structure

$$A(w(t)) = \sum_{i\in\mathcal{V}} A_i w_i(t), \quad B(w(t)) = \sum_{i\in\mathcal{V}} B_i w_i(t), \tag{A.2}$$

where $\mathcal{V} = \{1, 2, \cdots\}$.

The software can be applied to switching, gain-scheduling, and robust stabilization cases. Also sort of mixed robust/gain-scheduling problems can be considered. Indeed the system matrices $A(w(t))$ and $B(w(t))$ are allowed to attain their values in different matrix sets. More precisely, divide the matrices (A_i, B_i) in clusters $\mathcal{C}_k = \{(A_i, B_i) : i \in \mathcal{I}_k\}$ ($\mathcal{I}_k \cap \mathcal{I}_j = \emptyset$ and $\bigcup_k \mathcal{I}_k = \mathcal{V}$). The uncertainty is such that

$$w_i(t) = 0 \quad \text{if } i \notin \mathcal{I}_k, \qquad \sum_{i\in\mathcal{V}} w_i(t) = 1, \tag{A.3}$$

where k is a time-varying parameter indicating the current cluster. The only information available to the controller are the state of the system and the value of k. Special cases are then achieved by properly choosing the sets $\mathcal{I}_k$:

- when there is only one cluster, we have the robust stabilizability case;
- when every cluster is composed by only one couple of matrices (A_i, B_i), we have the switching gain-scheduling stabilizability case.

Due to the complexity of the calculation of an invariant set via polytopes, the main target in the implementation of MAXIS-G was computational efficiency. Different devices are used to increase the speed.

- The software is implemented in C++;
- Two polytopic representations are available: the double description and the inequality representations. The first one uses all the information about the inequalities and the vertices, while the latter uses only the inequalities.

The routines that use the double description method are implemented to adapt to the algorithm and perform well for low dimensional systems. The routines using the inequality representation use the GLPK library [a] and can be used for higher dimensional systems.

- All the redundant operations are eliminated. Starting from an initial polytope the algorithm at every step introduces new inequalities to possibly stop the procedure when an invariant set is found. An observation that can be made by analyzing the polytope at every step is that quite often the set of inequalities doesn't change considerably (see Figure 5.4). By implementing the algorithm in a straightforward manner, part of the operations are repeated for the inequalities that last in the polytope description for more than one step. The expansion to the state-input space of these inequalities doesn't change in following steps and combining the same expanded inequalities doesn't bring to generation of new inequalities. Experimentally, we have noticed that, by eliminating the redundant operations, the computational time can considerably decrease (for certain systems the reduction is more than 95%).

Beside systems of the form (A.1), the software can be used to calculate invariant sets for autonomous systems and continuous-time systems. While in the first case it is enough to set the input dimension to zero, in the latter case we have to set the time constant for the discretization, that is automatically done by the software with the Euler auxiliary system.

A.2.1 Software availability

MAXIS-G is available under GPL2 license and can be freely downloaded [b]. The software package is composed by two parts:

- a command line program that can be used for all the operations (like writing the input files and calculating the invariant sets);
- a simple interface designed for the MATLAB environment (beside all the operations available in the command line program, the graphic interface contains some facilities to plot the invariant sets).

A.2.2 Web addresses

a "GLPK: GNU Linear Programming Kit". www.gnu.org/software/glpk/.
b C. Savorgnan. MAXIS-G. www.diegm.uniud.it/savorgnan or www.diegm.uniud.it/smiani, 2005.

References

[ABB03] Alessandri, A., Baglietto, M., & Battistelli, G. (2003). Receding-horizon estimation for discrete-time linear system. *IEEE Transactions on Automatic Control, 46*(3), 473–478.

[AC84] Aubin, J. P., & Cellina, A. (1984). *Differential inclusions: Set-valued maps and viability theory*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* (Vol. 264). Berlin: Springer-Verlag.

[ACLC06] Alamo, T., Cepeda, A., Limon, D., & Camacho, E. F. (2006). A new concept of invariance for saturated systems. *Automatica Journal of IFAC, 42*(9), 1515–1521.

[ACMS97] Amato, F., Corless, M., Mattei, M., & Setola, R. A. (1997). A multivariable stability margin in the presence of time-varying, bounded rate gains. *International Journal of Robust and Nonlinear Control, 7*(2), 127–143.

[AG95] Apkarian, P., & Gahinet, P. (1995). A convex characterization of gain-scheduled H_∞ controllers. *IEEE Transactions on Automatic Control, 40*(5):853–864.

[Ala06] Alamir, M. (2006). *Stabilization of nonlinear systems using receding horizon technique*. *Lecture notes in control and information sciences* (Vol. 339). Berlin: Springer.

[Alt13] Altafini, C. (2013). Stability analysis of diagonally equipotent matrices. *Automatica Journal of IFAC, 49*(9), 2780–2785.

[AM81] Anderson, B. D. O., & Moore, J. B. (1981). Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM Journal on Control and Optimization, 19*(1), 20–32.

[ANP96] Abedor, J., Nagpal, K., & Poolla, K. (1996) A linear matrix inequality approach to peak-to-peak gain minimization. linear matrix inequalities in control theory and applications. *International Journal of Robust and Nonlinear Control, 6*(9–10), 899–927.

[AR08] Artstein, Z., & Raković, S. (2008). Feedback and invariance under uncertainty via set iterates. *Automatica, 44*(2), 520–525

[AR11] Artstein, Z., & Raković, S. V. (2011). Set invariance under output feedback: A set-dynamics approach. *International Journal of Systems Science, 42*(4), 539–555.

[Art83] Artstein, Z. (1983). Stabilization with relaxed controls. *Nonlinear Analysis, 7*(11), 1163–1173.

[ATRC07] Alamo, T., Tempo, R., Ramirez, D. R., & Camacho, E. F. (2007, July 2–5). A new vertex result for robustness problems with interval matrix uncertainty. In *Proceedings of the European Control Conference 2007*. Kos, Greece.

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9

[ATS07] Abate, A., Tiwari, A., & Sastry, S. (2007). Box invariance for biologically-inspired dynamical systems. In *Proceedings of the IEEE Conference on Decision and Control* (pp. 5162–5167).

[Aub91] Aubin, J. P. (1991). *Viability theory*. *Systems & control: Foundations & applications*. Boston, MA: Birkhäuser Boston Inc.

[AZ00] Allgöwer, F., & Zheng, A. (2000). *Nonlinear model predictive control* (Vol. 26). Basel: Birkhäuser.

[Bar59] Barbalat, I. (1959). Systems d'equations differentielles d'oscillations non lineaires. *Revue Roumaine de Mathématiques Pures et Appliquées Bucharest, IV*, 267–270.

[Bar85] Barmish, B. R. (1985) Necessary and sufficient conditions for quadratic stabilizability of an uncertain systems. *Journal of Optimization Theory and Applications, 46*(4), 399–408.

[Bar88a] Barabanov, N. E. (1988). Lyapunov indicator of discrete inclusion. I. *Automation and Remote Control, 49*(2), 152–157.

[Bar88b] Barabanov, N. E. (1988). Lyapunov indicator of discrete inclusion. II. *Automation and Remote Control, 49*(4), 283–287.

[Bar88c] Barabanov, N. E. (1988). Lyapunov indicator of discrete inclusion. III. *Automation and Remote Control, 49*(5), 558–566.

[BB89] Benzaouia, A., & Bourgat, C. (1989). Existence of non-symmetrical Lyapunov functions for linear system. *International Journal of Systems Science, 20*, 597–607.

[BB99] Benzaouia, A., & Baddou, A. (1999). Piecewise linear constrained control for continuous-time systems. *IEEE Transactions on Automatic Control, 44*(7), 1477–1481.

[BBBM05] Borrelli, F., Baotić, M., Bemporad, A., & Morari, M. (2005). Dynamic programming for constrained optimal control of discrete-time linear hybrid systems. *Automatica Journal of IFAC, 41*(10), 1709–1721.

[BC98] Brockman, M. L., & Corless, M. (1998). Quadratic boundedness of nominally linear systems. *International Journal of Control, 71*(6), 1105–1117.

[BC06] Bacciotti, A., & Ceragioli, F. (2006). Closed loop stabilization of planar bilinear switched systems. *International Journal of Control, 79*(1), 14–23.

[BCC+14] Blanchini, F., Casagrande, D., Colaneri, P., Gardonio, P., & Miani, S. (2014). Switching gains for semi-active damping via non-convex Lyapunov functions. *IEEE Transactions on Control Systems Technology, 59*(1), 107–119.

[BCG12] Balestrino, A., Caiti, A., & Grammatico, S. (2012). A new class of Lyapunov functions for the constrained stabilization of linear systems. *Automatica Journal of IFAC, 48*(11), 2951–2955.

[BCGM12] Blanchini, F., Casagrande, D., Gardonio, P., & Miani, S. (2012). Constant and switching gains in semi-active damping of vibrating structures. *International Journal of Control, 85*(12), 1886–1897.

[BCL83] Barmish, B. R., Corless, M., & Leitmann, G. (1983). A new class of stabilizing controllers for uncertain dynamical systems. *SIAM Journal on Control Optimization, 21*(2), 246–255.

[BCM97] Bemporad, A., Casavola, A., & Mosca, E. (1997). Nonlinear control of constrained linear systems with predictive reference management. *IEEE Transactions on Automatic Control, 42*(3), 340–349.

[BCMV10] Blanchini, F., Casagrande, D., Miani, S., & Viaro, U. (2010). Stable LPV realization of parametric transfer functions and its application to gain-scheduling control design. *IEEE Transactions on Automatic Control, 55*(10), 2271–2281.

[BCV12] Blanchini, F., Colaneri, P., & Valcher, M. E. (2012). Co-positive Lyapunov functions for the stabilization of positive switched systems. *IEEE Transactions on Automatic Control, 57*(12), 3038–3050.

[BCV13] Blanchini, F., Colaneri, P., & Valcher, M. E. (2013). Co-positive Lyapunov functions for the stabilization of positive switched systems. In *Proceedings of the 45th Conference on Decision and Control*. IEEE.

[BDCMP12] Blanchini, F., De Caneva, D., Montessoro, P. L., & Pierattoni, D. (2012). Control-based p-persistent adaptive communication protocol. *Transactions on Autonomous and Adaptive Systems, 22*(8).

[BDSCD07] Benzaouia, A., De Santis, E., Caravani, P., & Daraoui, N. (2007). Constrained control of switching systems: A positive invariant approach. *International Journal of Control, 80*(9), 1379–1387.

[BDVP09] Borrelli, F., Del Vecchio, C., & Parisio, A. (2009). Robust invariant sets for constrained storage systems. *Automatica Journal of IFAC, 45*(12), 2930–2936.

[BEGFB04] Boyd, S., El Ghaoui, L., Feron, E., & Balakrishnan, V. (2004). *Linear matrix inequalities in system and control theory*. Philadelphia: SIAM.

[Bel97] Beltrami, E. (1997). *Mathematics for dynamical modeling*. *Classics in applied mathematics*. San Diego: Academic Press.

[Bem98] Bemporad, A. (1998). Reference governor for constrained nonlinear systems. *IEEE Transactions on Automatic Control, 43*(3), 415–419.

[Ben94] Benzaouia, A. (1994). The resolution of equation $XA + XBX = HX$ and the pole assignment problem. *IEEE Transactions on Automatic Control, 39*(10), 2091–2095.

[Ber72] Bertsekas, D. P. (1972). Infinite-time reachability of state-space regions by using feedback control. *IEEE Transactions on Automatic Control, 17*, 604–613.

[Ber00] Bertsekas, D. P. (2000). *Dynamic programming and optimal control*. Belmont, MA: Athena Scientific.

[BF11] Blanchini, F., & Franco, E. (2011). Structurally robust biological networks. *Bio Med Central Systems Biology, 5*(1), 74.

[BG84] Barabanov, N. E., & Granichin, O. N. (1984). Optimal controller for a linear plant with bounded noise. *Automation and Remote Control, 45*(5), 578–584.

[BG86] Barmish, B. R., & Galimidi, A. R. (1986). Robustness of Luenberger observers: Linear systems stabilized via non-linear control. *Automatica Journal of IFAC, 22*(4), 413–423.

[BG95] Bitsoris, G., & Gravalou, E. (1995). Comparison principle, positive invariance and constrained regulation of nonlinear systems. *Automatica Journal of IFAC, 31*(2), 217–222.

[BG99] Bitsoris, G., & Gravalou, E. (1999). Design techniques for the control of discrete-time systems subject to state and control constraints. *IEEE Transactions on Automatic Control, 44*(5), 1057–1061.

[BG00] Bemporad, A., & Garulli, A. (2000). Output-feedback predictive control of constrained linear systems via set-membership state estimation. *International Journal of Control, 73*(8), 655–665.

[BG02] Blanchini, F., & Giannattasio, P. (2002). Adaptive control of compressor surge instability. *Automatica Journal of IFAC, 38*(8), 1373–1380.

[BG14] Blanchini, F., & Giordano, G. (2014). Piecewise-linear Lyapunov functions for structural stability of biochemical networks. *Automatica Journal of IFAC, 50*(10), 2482–2494.

[BGL69] Blaquière, A., Gérard, F., & Leitmann, G. (1969). *Quantitative and qualitative games*. New York: Academic Press. Includes bibliographical references (pp. 165–167) and indexes.

[BH93] Benzaouia, A., & Hmamed, A. (1993). Regulator problem for linear continuous-time systems with nonsymmetrical constrained control. *IEEE Transactions on Automatic Control, 38*, 1556–1560.

[BHSB96] Berg, J. M., Hammet, K. D., Schwartz, C. A., & Banda, S. (1996). An analysis of the destabilizing effect of daisy chained rate-limited actuators. *Transactions on Control Systems Technology, 4*(2), 171–176.

[Bit88] Bitsoris, G. (1988). On the positive invariance of polyhedral sets for discrete-time systems. *Systems Control Letters, 11*(3), 243–248.

[Bit91] Bitsoris, G. (1991). Existence of positively invariant polyhedral sets for continuous-time linear systems. *Control Theory Advanced Technology, 7*(3), 407–427.

[Bla90a] Blanchini, F. (1990). Constrained control for perturbed linear system. In *Proceedings of the 29th Conference on Decision and Control* (pp. 3464–3467). IEEE.

[Bla90b] Blanchini, F. (1990). Feedback control for linear time-invariant systems with state and control bounds in the presence of disturbances. *IEEE Transactions on Automatic Control, 45*(11), 2061–2070.

[Bla91] Blanchini, F. (1991). Constrained control for uncertain linear systems. *Journal of Optimization Theory and Applications, 71*(3), 465–483.

[Bla92] Blanchini, F. (1992). Minimum-time control for uncertain linear discrete-time linear systems. In *Proceedings of the 31st Conference on Decision and Control* (pp. 2629–2634). IEEE.

[Bla94] Blanchini, F. (1994). Ultimate boundedness control for discrete-time uncertain system via set-induced Lyapunov functions. *IEEE Transactions on Automatic Control, 39*(2), 428–433.

[Bla95] Blanchini, F. (1995). Non-quadratic Lyapunov function for robust control. *Automatica Journal of IFAC, 31*(3), 451–461.

[Bla99] Blanchini, F. (1999). Set invariance in control—a survey. *Automatica Journal of IFAC, 35*(11), 1747–1767.

[Bla00] Blanchini, F. (2000). The gain scheduling and the robust state feedback stabilization problems. *IEEE Transactions on Automatic Control, 45*(11), 2061–2070.

[BM82] Belforte, G., & Milanese, M. (1982). Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators. *IEEE Transactions on Automatic Control, 27*(2), 408–414.

[BM92] Basile, G., & Marro, G. (1992). *Controlled and conditioned invariants in linear system theory*. Englewood Cliffs, NJ: Prentice Hall

[BM94] Bhaya, A., & Mota, C. (1994). Equivalence of stability concepts for discrete time-varying system. *International Journal of Robust and Nonlinear Control, 4*(6), 725–740.

[BM96a] Blanchini, F., & Miani, S. (1996). Constrained stabilization of continuous-time linear systems. *Systems Control Letters, 28*(2), 95–102.

[BM96b] Blanchini, F.,& Miani, S. (1996). On the transient estimate for linear systems with time-varying uncertain parameters. *IEEE Transactions on Circuits and Systems, 43*(7), 592–596.

[BM98] Blanchini, F., & Miani, S. (1998). Constrained stabilization via smooth Lyapunov functions. *Systems Control Letters, 35*, 155–163.

[BM99a] Bemporad, A., & Morari, M. (1999). Robust model predictive control: A survey. In A. Garulli & A. Tesi (Eds.), *Robustness in identification and control*. *Lecture notes in control and information sciences* (Vol. 245, pp. 207–226). London: Springer.

[BM99b] Blanchini, F., & Megretski, A. (1999). Robust state feedback control of LTV systems: Nonlinear is better than linear. *IEEE Transactions on Automatic Control, 44*(4), 802–807.

[BM99c] Blanchini, F., & Miani, S. (1999). A universal class of smooth functions for robust control. *IEEE Transactions on Automatic Control, 44*(3), 641–647.

[BM00] Blanchini, F., & Miani, S. (2000). Any domain of attraction for a linear constrained system is a tracking domain of attraction. *SIAM Journal on Control Optimization, 38*(3), 971–994.

[BM03] Blanchini, F., & Miani, S. (2003). Stabilization of LPV systems: State feedback, state estimation and duality. *SIAM Journal on Control Optimization, 32*(1), 76–97.

[BMDP02] Bemporad, A., Morari, M., Dua, V., & Pistikopoulos, E. N. (2002). The explicit linear quadratic regulator for constrained systems. *Automatica Journal of IFAC, 38*(1), 3–20.

[BMM95] Blanchini, F., Mesquine, F., & Miani, S. (1995). Constrained stabilization with an assigned initial condition sets. *International Journal of Control, 62*(3), 601–617.

[BMM09] Blanchini, F., Miani, S., & Mesquine, F. (2009). A separation principle for linear switching systems and parametrization of all stabilizing controllers. *IEEE Transactions on Automatic Control, 54*(2), 279–292.

[BMP03] Blanchini, F., Miani, S., & Pellegrino, F. A. (2003). Suboptimal receding horizon control for continuous-time systems. *IEEE Transactions on Automatic Control, 48*(6), 1081–1086.

[BMPVA08] Blanchini, F., Miani, S., Pellegrino, F. A., & Van Arkel, B. (2008). Enhancing controller performance for robot positioning in a constrained environment. *IEEE Transaction on Control System Technology, 16*(5), 1066–1074

[BMR04] Blanchini, F., Miani, S., & Rinaldi, F. (2004). Guaranteed cost control for multi-inventory systems with uncertain demand. *Automatica Journal of IFAC, 40*(2), 213–224.

[BMS96] Blanchini, F., Miani, S., & Sznaier, M. (1996). Worst case l_∞ to l_∞ gain minimization: Dynamic versus static state feedback. In *Proceedings of the 35th Conference on Decision and Control*, Kobe, Japan (pp. 2395–2400).

[BMS97] Blanchini, F., Miani, S., & Sznaier, M. (1997). Robust performance with fixed and worst-case signals for uncertain time varying systems. *Automatica Journal of IFAC, 33*(12), 2183–2189.

[BMT96] Benzaouia, A., Mesquine, F., & Tadeo, F. (1996). The regulator problem for linear systems with saturations on the control and its increments or rate. *IEEE Circuit and systems I, 53*(12), 2681–2691.

[BMU00] Blanchini, F., Miani, S., & Ukovich, W. (2000). Control of production-distribution systems with unknown inputs and system failures. *IEEE Transactions on Automatic Control, 45*(6), 1072–1081.

[BN05] Blondel, V., & Nesterov, Y. (2005). Computationally efficient approximations of the joint spectral radius. *SIAM Journal of Matrix Analysis, 27*(1), 256–272.

[BNT05] Blondel, V., Nesterov, Y., & Theys, J. (2005). On the accuracy of the ellipsoid norm approximation of the joint spectral radius. *Linear Algebra and its Applications, 394*, 91–107.

[BO99] Başar, T., & Olsder, G. J. (1995). *Dynamic noncooperative game theory*. *Classics in applied mathematics* (Vol. 23). Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). Reprint of the second edition (1995).

[Bon69] Bony, J. M. (1969). Principe du maximum, inègalite de Harnack et unicité du probléme de Cauchy pour les opérateurs elliptiques dégénérés. *Annales de l'Institut Fourier, Grenoble, 19*, 277–304.

[Bos02] Boscain, U. (2002). Stability of planar switched systems: The linear single input case. *SIAM Journal on Control Optimization, 41*(1), 89–112.

[Bou32] Bouligand, G. (1932). *Introducion a la geometrie infinitesimale directe*. Paris: Gauthiers-Villars.

[BP94] Becker, G., & Packard, A. (1994). Robust performance of linear parametrically varying systems using parametrically-dependent linear feedback. *Systems & Control Letters, 23*, 05–215.

[BP98] Baras, J. S., & Patel, N. S. (1998). Robust control of set-valued discrete-time systems. *IEEE Transactions on Automatic Control, 43*(1), 61–75.

[BP03] Blanchini, F., & Pellegrino, F. A. (2003). Relatively optimal control and its linear implementation. *IEEE Transactions on Automatic Control, 48*(12), 2151–2162.

[BP06] Blanchini, F., & Pellegrino, F. A. (2006). Relatively optimal control with characteristic polynomial assignment and output feedback. *IEEE Transactions on Automatic Control, 51*(2), 183–191.

[BP07] Blanchini, F., & Pellegrino, F. A. (2007). Relatively optimal control: The piecewise affine solution. *SIAM Journal on Control Optimization, 42*(2), 585–603.

[BPF83] Barmish, B. R., Petersen, I., & Feuer, A. (1983). Linear ultimate boundedness control of uncertain systems. *Automatica Journal of IFAC, 19*(5), 523–532.

[BPV04] Blanchini, F., Pellegrino, F. A., & Visentini, L. (2004). Control of manipulators in a constrained workspace by means of linked invariant set. *International Journal on Robust and Nonlinear Control, 14*, 1185–1205.

[BR71a] Bertsekas, D. P., & Rhodes, I. B. (1971). On the minmax reachability of target set and target tubes. *Automatica Journal of IFAC, 7*, 233–247.

[BR71b] Bertsekas, D. P., & Rhodes, I. B. (1971). Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control, 16*, 117–128.

[BR06] Broucke, M., & Roszak, B. (2006). Necessary and sufficient conditions for reachability on a simplex. *Automatica Journal of IFAC, 42*(11), 1913–1918.

[Bre70] Brezis, H. (1970). On a characterization of flow-invariant sets. *Communications on Pure and Applied Mathematics, 23*(261–263), 261–263.

[BRK99] Burridge, R. R., Rizzi, A. A., & Koditschek, D. E. (1999). Sequential composition of dynamically dexterous robot behaviors. *International Journal on Robotics Research, 18(6)*, 534–555.

[Bro10] Broucke, M. (2010). Reach control on simplices by continuous state feedback. *SIAM Journal on Control Optimization, 48*(5), 3482–3500.

[BRU97] Blanchini, F., Rinaldi, F., & Ukovich, W. (1997). Least inventory control of multistorage systems with non-stochastic unknown inputs. *IEEE Transactions on Robotics and Automation, 13*(5), 633–645.

[BS79] Barmish, B. R., & Sankaran, J. (1979). The propagation of parametric uncertainty via polytopes. *IEEE Transactions on Automatic Control, 24*(2), 346–349.

[BS80] Barmish, B. R., & Schmitedorf, W. E. (1980). Null controllability of linear systems with constrained controls. *SIAM Journal on Control Optimization, 18*(2), 327–345.

[BS94] Blanchini, F., & Sznaier, M. (1994). Rational $\mathcal{L}_1$ suboptimal compensators for continuous-time systems. *IEEE Transactions on Automatic Control, 39*(7), 1487–1482.

[BS95] Blanchini, F., & Sznaier, M. (1995). Persistent disturbance rejection via static state feedback. *IEEE Transactions on Automatic Control, 40*(6), 1127–1131.

[BS04] Bolzern, P., & Spinelli, W. (2004). Quadratic stabilization of a switched affine system about a nonequilibrium point. In *Proceedings of the 2004 American Control Conference*, Boston, USA (pp. 3890–3895).

[BS08] Blanchini, F., & Savorgnan, C. (2008). Stabilizability of switched linear systems does not imply the existence of convex Lyapunov functions. *Automatica Journal of IFAC, 44*(4), 1166–1170.

[BT80] Brayton, R. K., & Tong, C. H. (1980). Constructive stability and asymptotic stability of dynamical systems. *IEEE Transactions on Circuits and Systems, 27*(11), 1121–1130.

[BT94] Bourgat, C., & Tarbouriech, S. (1994). Positively invariant sets for constrained continuous-time systems with cone properties. *IEEE Transactions on Automatic Control, 39*(2), 21–22.

[BT00] Blondel, V. D., & Tsitsiklis, J. N. (2000). A survey of computational complexity results in systems and control. *Automatica Journal of IFAC, 36*(9), 1249–1274.

[BT11] Bitsoris, G., & Truffet, L. (2011). Positive invariance, monotonicity and comparison of nonlinear systems. *Systems Control Letters, 60*(12), 960–966.

[BU93] Blanchini, F., & Ukovich, W. (1993). A linear programming approach to the control of discrete-time periodic system with state and control bounds in the presence of disturbance. *Journal of Optimization Theory and Applications, 73*(3), 523–539.

[BV90] Bitsoris, G., & Vassilaki, M. (1990). Constrained regulation of linear systems. In *Proceedings of the 11th IFAC World Congress*, Tallin, Estonia (pp. 287–292).

[BV95] Bitsoris, G., & Vassilaki, M. (1995). Constrained regulation of linear systems. *Automatica Journal of IFAC, 31*(3), 223–227.

[BV04] Boyd, S., & Vandenberghe, L. (2004) *Convex optimization*. Cambridge: Cambridge University Press.

[BW84] Byrnes, C. I., & Willems, J. C. (1984). Adaptive stabilization of multivariable linear systems. In *Proceedings of the 23rd Conference on Decision and Control* (pp. 1574–1577). Las Vegas: IEEE.

[CB11] Cosentino, C., & Bates, D. (2011). *Feedback control in systems biology*. Boca Raton: Taylor & Francis.

[CBKR11] Cannon, M., Buerger, J., Kouvaritakis, B., & Raković, S. (2011). Robust tubes in nonlinear model predictive control. *IEEE Transactions on Automatic Control, 56*(8), 1942–1947.

[CCG+12] Chesi, G., Colaneri, P., Geromel, J. C., Middleton, R. H., & Shorten, R. (2012). A nonconservative LMI condition for stability of switched systems with guaranteed dwell time. *IEEE Transactions on Automatic Control, 57*(5), 1297–1302.

[CCSZ11] Carli, R., Chiuso, A., Schenato, L., & Zampieri, S. (2011). Optimal synchronization for networks of noisy double integrators. *IEEE Transactions on Automatic Control, 56*(5), 1146–1152.

[CDS02] Caravani, P., & De Santis, E. (2002). Doubly invariant equilibria of linear discrete-time games. *Automatica Journal of IFAC, 38*(9), 1531–1538.

[CDS09] Caravani, P., & De Santis, E. (2009). Observer-based stabilization of linear switching systems. *International Journal of Robust and Nonlinear Control, 19*(14), 1541–1563.

[CGTV03] Chesi, G., Garulli, A., Tesi, A., & Vicino, A. (2003). Homogeneous Lyapunov functions for systems with structured uncertainties. *Automatica Journal of IFAC, 39*(6), 1027–1035.

[CGTV09] Chesi, G., Garulli, A., Tesi, A., & Vicino, A. (2009). *Homogeneous polynomial forms for robustness analysis of uncertain systems*. Berlin: Springer.

[CH92] Castelan, E. B., & Hennet, J. C. (1992). Eigenstructure assignment for state constrained linear continuous time systems. *Automatica Journal of IFAC, 28*(3), 605–611.

[CH93] Castelan, E. B., & Hennet, J. C. (1993). On invariant polyhedra of continuous-time linear systems. *IEEE Transactions on Automatic Control, 38*(11), 1680–1685.

[Che81] Chernousko, F. L. (1981). Ellipsoidal estimates of a controlled system's attainability domain. *Journal of Applied Mathematics and Mechanics, 45*(1), 11–19.

[Che94] Chernousko, F. L. (1994). *State estimation of dynamic systems*. Boca Raton, FL: CRC Press.

[Che10] Chesi, G. (2010). LMI techniques for optimization over polynomials in control: A survey. *IEEE Transactions on Automatic Control, 55*(11), 2500–2510.

[Che11a] Chesi, G. (2011). *Domain of attraction: Analysis and control via SOS programming*. *Lecture notes in control and information sciences*. Berlin: Springer.

[Che11b] Chesi, G. (2011). LMI conditions for time-varying uncertain systems can be non-conservative. *Automatica Journal of IFAC, 47*(3), 621–624.

[Che13] Chesi, G. (2013). Sufficient and necessary LMI conditions for robust stability of rationally time-varying uncertain systems. *IEEE Transactions on Automatic Control, 58*(6), 1546–1551.

[CKRC11] Cannon, M., Kouvaritakis, B., Raković, S. V., & Cheng, Q. (2011). Stochastic tubes in model predictive control with probabilistic constraints. *IEEE Transactions on Automatic Control, 56*(1), 194–200.

[CL03] Cao, Y. Y., & Lin, Z. (2003). Stability analysis of discrete-time systems with actuator saturation by a saturation-dependent Lyapunov function. *Automatica Journal of IFAC, 39*(7), 1235–1241.

[Cla83] Clarke, F. H. (1983). *Optimization and non smooth analysis*. New York: Wiley.

[CMA00] Casavola, A., Mosca, E., & Angeli, D. (2000). Robust command governors for constrained linear systems. *IEEE Transactions on Automatic Control, 45*(11), 2071–2077.

[CMP04] Casavola, A., Mosca, E., & Papini, M. (2004). Control under constraints: An application of the command governor approach to an inverted pendulum. *IEEE Transactions on Control Systems Technology, 12*, 193–204.

[CO04] Chernousko, F. L., & Ovseevich, A. I. (2004). Properties of the optimal ellipsoids approximating the reachable sets of uncertain systems. *Journal of Optimization Theory and Applications, 120*(2), 223–246.

[Col09] Colaneri, P. (2009). Dwell time analysis of deterministic and stochastic switched systems. *European Journal of Control, 15*, 228–248.

[CT95] Castelan, E. B., & Tarbouriech, S. (1995). An eigenstructure assignment approach for constrained linear continuous-time singular systems. *Systems Control Letters, 24*(5), 333–343.

[CWLA05] Chen, L., Wang, R., Li, C., & Aihara, K. (2005). *Modeling biomolecular networks in cells*. New York: Springer.

[CZ02] Chisci, L., & Zappa, G. (2002). Feasibility in predictive control of constrained linear systems: The output feedback case. *International Journal of Robust and Nonlinear Control, 12*(5), 465–487.

[DB01] Daafouz, J., & Bernussou, J. (2001) Parameter dependent Lyapunov functions for discrete time systems with time varying parametric uncertainties. *Systems Control Letters, 5*(43), 355–359.

[DBD92] Diaz-Bobillo, I. J., & Dahleh, M. A. (1992). State feedback l_1-optimal controllers can be dynamic. *Systems Control Letters, 19*,87–93.

[DBPL00] Decarlo, R. A., Branicky, M. S., Pettersson, S., & Lennartson, B. (2000). Perspectives and results on the stability and stabilizability of hybrid systems. *Proceedings of the IEEE, 88*(7), 1069–1082.

[DCAF94] Dasgupta, S., Chockalingam, G., Anderson, B., & Fu. M. (1994). Lyapunov functions for uncertain systems with applications to the stability of time varying systems. *IEEE Transactions on Circuits and Systems, 41*(2), 93–105.

[DDB95] Dahleh, M. A., & Diaz-Bobillo, I. D. (1995). *Control of uncertain systems: A linear programming approach*. Englewood Cliffs, NJ: Prentice-Hall.

[DDS96] D'Alessandro, P., & De Santis, E. (1996). General closed loop optimal solutions for linear dynamic systems with linear constraints and functional. *Journal of Mathematical Systems, Estimation, and Control, 6*(2), 1–14.

[DGD11] Daecto, G. S., Geromel, J. C., & Daafouz, J. (2011). Dynamic output feedback H_∞ control of switched linear system. *Automatica Journal of IFAC, 47*, 1713–1720.

[DH99] Dórea, C. E. T., & Hennet, J. C. (1999). (A, B)-invariant polyhedral sets of linear discrete-time systems. *Journal of Optimization Theory and Applications, 103*(3), 521–542.

[DH01] Dórea, C. E. T., & Hennet, J. C. (2001). (A, B)-invariance conditions of polyhedral domains for continuous-time systems. *European Journal of Control, 5*(1),70–81.

[DL92] Dontchev, A., & Lempio, F. (1992). Difference methods for differential inclusions: A survey. *SIAM Review, 34*(2), 263–294.

[dOBG99] de Oliveira, M. C., Bernussou, J., & Geromel, J. C. (1999). A new discrete-time robust stability condition. *Systems and Control Letters, 37*, 261–265.

[DP87] Dahleh, M. A., & Pearson, J. B. (1987). l^1-Optimal feedback controllers for MIMO discrete-time system. *IEEE Transactions on Automatic Control, 32*(4), 314–322.

[DS08] De Santis, E. (2008). Invariant dual cones for hybrid systems. *Systems and Control Letters, 57*(12), 971–977.

[DSDB09] De Santis, E., & Di Benedetto, M. D. (2009). Editorial: Observability and observer-based control of hybrid systems. *International Journal of Robust and Nonlinear Control, 19*(14), 1519–1520.

[DSDBB04] De Santis, E., Di Benedetto, M. D., & Berardi, L. (2004) Computation of maximal safe sets for switching systems. *IEEE Transactions on Automatic Control, 49*(2), 184–195.

[DSDBP06] De Santis, E., Di Benedetto, M. D., & Pola, G. (2006). Digital idle speed control of automotive engines: A safety problem for hybrid systems. *Nonlinear Analysis, 65*(9), 1705–1724.

[DVM14] Del Vecchio, D., & Murray, R. M. (2014). *Biomolecular feedback systems*. New Jersey: Princeton University Press.

[EK05] Edelstein-Keshet, L. (2005). *Mathematical models in biology*. Philadelphia, PA: SIAM.

[FAC10] Fiacchini, M., Alamo, T., & Camacho, E. F. (2010). On the computation of convex robust control invariant sets for nonlinear systems. *Automatica Journal of IFAC, 46*(8), 1334–1338.

[FAG96] Feron, E., Apkarian, P., & Gahinet, P. (1996). Analysis and synthesis of robust control systems via parameter-dependent Lyapunov functions. *IEEE Transactions on Automatic Control, 41*(7), 1041–1046.

[FB97] Farina, L., & Benvenuti, L. (1997). Polyhedral reachable set with positive controls. *Mathematics of Control, Signals, and Systems, 10*(4), 364–380.

[FB04] Farina, L., & Benvenuti, L. (2004). A tutorial on the positive realization problem. *IEEE Transactions on Automatic Control, 49*(5), 651–664.

[FB12] Franco, E., & Blanchini, F. (2012). Structural properties of the MAPK pathway topologies in PC12 cells. *Journal of Mathematical Biology,* 1–36.

[Fei87] Feinberg, M. (1987). Chemical reaction network structure and the stability of complex isothermal reactors: The deficiency zero and deficiency one theorems. *Chemical Engineering Science, 42*, 2229–2268.

[FG95] Fialho, I. J., & Georgiou, T. (1995). On the L_1 norm of uncertain linear systems. *IEEE Transactions on Automatic Control, 40*(6), 1142–1147.

[FG97] Fialho, I. J., & Georgiou, T. (1997). l_1 state-feedback control with a prescribed rate of exponential convergence. *IEEE Transactions on Automatic Control, 42*(10), 1476–1481

[FH76a] Feuer, A., & Heymann, M. (1976). Admissible sets in linear feedback systems with bounded control. *International Journal of Control, 23*(3), 381–393.

[FH76b] Feuer, A., & Heymann, M. (1976). ω-Invariance in control systems with bounded control. *Journal of Mathematical Analysis and Applications, 53*(3), 266–276.

[FJ14] Fiacchini, M., & Jungers, M. (2014). Necessary and sufficient condition for stabilizability of discrete-time linear switched systems: A set-theory approach. *Automatica Journal of IFAC, 50*(1), 75–83.

[FK96a] Freeman, R. A., & Kokotović, P. V. (1996). Inverse optimality in robust stabilization. *SIAM Journal on Control Optimization, 34*(4), 1365–1391.

[FK96b] Freeman, R. A., & Kokotović, P. V. (1996). *Robust nonlinear control design. State-space and Lyapunov techniques*. *Systems and control: Foundations and applications* (xii + 256 pp.). Boston, MA: Birkhäuser Inc. ISBN: 0-8176-3930-6.

[FMC09] Fainshil, L., Margaliot, M., & Chigansky, P. (2009). On the stability of positive linear switched systems under arbitrary switching laws. *IEEE Transactions on Automatic Control, 54*, 807–899.

[FR00] Farina, L., & Rinaldi, S. (2000). *Positive linear systems: Theory and applications*. New York: Wiley.

[FV12] Fornasini, E., & Valcher, M. E. (2012). Stability and stabilizability criteria for discrete-time positive switched systems. *IEEE Transactions on Automatic Control, 57*(5), 1208–1221.

[FZ87] Fernandes, M. L. C., & Zanolin, F. (1987). Remarks on strongly flow-invariant sets. *Journal of Mathematical Analysis and Applications, 128*(1), 176–188.

[Gar80] Gard, T. C., Strongly flow invariant sets. *Applicable Analysis, 10*, 285–293.

[Gay86] Gayek, J. E. (1986). Approximating reachable sets for a class of linear control systems. *International Journal of Control, 43*(2), 441–453.

[Gay91] Gayek, J. E. (1991). A survey of techniques for approximating reachable and controllable sets. In *Proceedings of the 30th Conference on Decision and Control*, Brighton, UK (pp. 1724–1730).

[GBC14] Grammatico, S., Blanchini, F., & Caiti, A. (2014). Control-sharing and merging control Lyapunov functions. *IEEE Transactions on Automatic Control, 59*(1), 1–13.

[GC86a] Gutman, P., & Cwikel, M. (1986). Admissible sets and feedback control for discrete-time linear dynamical systems with bounded controls and states. *IEEE Transactions on Automatic Control, 31*(4), 373–376.

[GC86b] Gutman, P., & Cwikel, M. (1986). Convergence of an algorithm to find maximal state constraint sets for discrete-time linear dynamical systems with bounded control and state. *IEEE Transactions on Automatic Control, 31*(5), 457–459.

[GC87] Gutman, P., & Cwikel, M. (1987). An algorithm to find maximal state constraint sets for discrete-time linear dynamical systems with bounded control and state. *IEEE Transactions on Automatic Control, 32*(3), 251–254.

[GC05] Geromel, J. C., & Colaneri, P. (2005). Stabilization of continuous-time switched systems. In *Proceedings of 2005 IFAC Conference*.

[GC06] Geromel, J. C., & Colaneri, P. (2006). Stability and stabilization of discrete time switched systems. *International Journal of Control, 79*(7), 719–728.

[GCB08] Geromel, J. C., Colaneri, P., & Bolzern, P. (2008). Dynamic output feedback control of switched linear systems. *IEEE Transactions on Automatic Control, 53*(3), 720–733.

[GdST99] Gomes da Silva, J. M., & Tarbouriech, S. (1999). Polyhedral regions of local stability for linear discrete-time systems with saturating controls. *IEEE Transactions on Automatic Control, 44*(11), 2081–2085.

[GdST01] Gomes da Silva, J. M., & Tarbouriech, S. (2001). Local stabilization of discrete-time linear systems with saturating controls: An LMI-based approach. *IEEE Transactions on Automatic Control, 46*(1), 119–125.

[GdSTG03] Gomes da Silva, J. M., Tarbouriech, S., & Garcia, G. (2003). Local stabilization of linear systems under amplitude and rate saturating actuators. *IEEE Transactions on Automatic Control, 48*(5), 842–847.

[GFA+11] Gonzalez, R., Fiacchini, M., Alamo, T., Guzman, J. L., & Rodriguez, F. (2011). Online robust tube-based MPC for time-varying systems: A practical approach. *International Journal of Control, 84*(6), 1157–1170.

[GH85] Gutman, P., & Hagander, P. (1985). A new design of constrained controllers for linear systems. *IEEE Transactions on Automatic Control, 30*(1), 22–33.

[GK91a] Georgiou, C., & Krikelis, N. J. (1991). A design approach for constrained regulation in discrete singular systems. *IEEE Transactions on Automatic Control, 17*(4), 297–304.

[GK91b] Graettinger, T. J., & Krogh, B. H. (1991). Hyperplane method for reachable state estimation for linear time-invariant systems. *Journal of Optimization Theory and Applications, 69*(3), 555–587.

[GK92] Graettinger, T. J., & Krogh, B. H. (1992). On the computation of reference signal constraints for guaranteed tracking performance. *Automatica Journal of IFAC, 28*(6), 1125–1141.

[GK02] Gilbert, E. G., & Kolmanovsky, I. V. (2002). Nonlinear tracking control in the presence of state and control constraints: A generalized reference governor. *Automatica Journal of IFAC, 38*(12), 2063–2073.

[GKT95] Gilbert, E. G., Kolmanovsky, I. V., & Tan, K. K.(1995). Discrete-time reference governors and the nonlinear control of systems with state and control constraints. *International Journal of Robust and Nonlinear Control, 5*(5), 487–504.

[GLO12] Gielen, R. H., Lazar, M., & Olaru, S. (2012). Set-induced stability results for delay difference equations. In *Time delay systems: Methods, applications and new trends* (pp. 73–84). Heidelberg: Springer.

[GM04] Gasparetto, A., & Miani, S. (2004). Dynamic model of a rotating channel used in the steel industry and controller implementation. *Journal of Vibration and Control, 10*(3), 423–445.

[GOL+10] Gielen, R. H., Olaru, S., Lazar, M., Heemels, W. P. M. H., Van de Wouw, N., & Niculescu, S. I. (2010). On polytopic inclusions as a modeling framework for systems with time-varying delays. *Automatica Journal of IFAC, 46*(3), 615–619.

[GPB91] Geromel, J. C., Peres, P. L. D., & Bernussou, J. (1991). On a convex parameter space method for linear control design of uncertain systems. *SIAM Journal on Control Optimization, 29*(2), 381–402.

[GPM89] Garcia, C. E., Prett, D. M., & Morari, M. (1989). Model predictive control theory and practice—a survey. *Automatica Journal of IFAC, 25*(3), 335–348.

[Gre76] Greitzer, E. M. (1976). Surge and rotating stall in axial flow compressors, part i: Theoretical compression system mode. *ASME Journal of Engineering for Power, 98*(2), 190–198.

[GS71] Glover, D., & Schweppe, F. (1971). Control of linear dynamic systems with set constrained disturbances. *IEEE Transactions on Automatic Control, 16*(5), 411–423.

[GS83] Glattfelder, A. H., & Schaufelberger, W. (1983). Stability analysis of single loop control systems with saturations and antireset-windup circuits. *IEEE Transactions on Automatic Control, 28*(12), 1074–1081.

[GS88] Glattfelder, A. H., & Schaufelberger, W. (1988). Stability of discrete override and cascade-limiter single-loop control systems. *IEEE Transactions on Automatic Control, 33*(6), 532–540.

[GSDD06] Goodwin, G. C., Seron, M. M., & De Doná, J. A. (2006). *Constrained control and estimation: An optimisation approach*. Berlin: Springer.

[GSM07] Gurvits, L., Shorten, R., & Mason, O. (2007). On the stability of switched positive linear systems. *IEEE Transactions on Automatic Control, 52*, 1099–1103.

[GT91] Gilbert, E. G., & Tan, K. K. (1991). Linear systems with state and control constraints: The theory and the applications of the maximal output admissible sets. *IEEE Transactions on Automatic Control, 36*(9), 1008–1020.

[GTC11] Garone, E, Tedesco, F., & Casavola, A. (2011). Sensorless supervision of linear dynamical systems: The feed-forward command governor approach. *Automatica Journal of IFAC, 47*(7), 1294–1303.

[GTV85] Genesio, R., Tartaglia, M., & Vicino, A. (1985). On the estimate of asymptotic stability regions: State of art and new proposal. *IEEE Transactions on Automatic Control, 30*(8), 437–443.

[Gut79] Gutman, S. (1979). Uncertain dynamical systems—a Lyapunov min-max approach. *IEEE Transactions on Automatic Control, 24*(3), 437–443.

[Gut86] Gutman, P. (1986). A linear programming regulator applied to hydroelectric reservoir level control. *Automatica Journal of IFAC, 22*(3), 373–376.

[GY93] Gerasimov, O. I., & Yunger, I. B. (1993). A matrix criterion for the absolute stability of impulse automatic control systems. *Automation and Remote Control, 54*(2), 290–294.

[Hah67] Hahn, W. (1967). *Stability of motion*. Berlin: Springer Verlag.

[Haj91] Hajek, O. (1991). *Control theory in the plane*. Berlin: Springer Verlag.

[Hal69] Hale, J. K. (1969). *Ordinary differential equation*. New York: Wiley Interscience.

[Hal88] Hale, J. K. (1988). *Asymptotic behavior of dissipative systems*. *Mathematical surveys and monographs* (Vol. 25). Providence, RI: American Mathematical Society.

[Har72] Hartman, P. (1972). On invariant sets and on a theorem of Wazeski. *Proceedings of American Mathematical Society, 32*(2).

[HB76] Horisberger, H. P., & Belanger, P. R. (1976). Regulators for linear, time invariant plants with uncertain parameters. *IEEE Transactions on Automatic Control, 21*, 705–708.

[HB91] Hennet, J. C., & Beziat, J. P. (1991). A class of invariant regulators for the discrete-time linear constrained regulation problem. *Automatica Journal of IFAC, 27*(3), 549–554.

[HBB95] Hmamed, A., Benzaouia, A., & Bensalah, H. (1995). Regulator problem for linear continuous-time delay systems with nonsymmetrical constrained control. *IEEE Transactions on Automatic Control, 40*(9), 1615–1619.

[HBR83] Hrovat, D., Barak, P., & Rabins, M. (1983). Semi-active versus passive or active tunes mass dampers for structural control. *Journal of Engineering Mechanics, 109*(3), 691–705.

[Hel98] Helmersson, A. (1998). μ synthesis and LFT gain scheduling with real uncertainties. *International Journal of Robust and Nonlinear Control, 8*(7), 631–642.

[HKH87] Hanus, R., Kinnaert, M., & Henrotte, J. L. (1987). Conditioning technique, a general anti-windup and bumpless transfer method. *Automatica, 23*(6), 729–739.

[HL01] Hu, T., & Lin, Z. (2001). *Control of systems with actuator saturation*. Boston, MA: Birkhauser.

[HL02] Hu, T., & Lin, Z. (2002). Exact characterization of invariant ellipsoids for single input linear systems subject to actuator saturation. *IEEE Transactions on Automatic Control, 47*(1), 164–169.

[HL03] Hu, T., & Lin, Z. (2003). Composite quadratic Lyapunov functions for constrained control systems. *IEEE Transactions on Automatic Control, 48*(3), 440–450.

[HL08] Hu, T., & Lin, Z. (2008). Stabilization of switched systems via composite quadratic functions. *IEEE Transactions on Automatic Control, 53*(11), 2571–2585.

[HM02] Hespanha, J. P., & Morse, A. S. (2002). Switching between stabilizing controllers. *Automatica Journal of IFAC, 38*(11), 1905–1917.

[HR71] Hamza, M. H., & Rasmy, M. E. (1971). A simple method for determining the reachable set for linear discrete systems. *IEEE Transactions on Automatic Control, 16*(3), 281–282.

[HT97] Hennet, J. C., & Tarbouriech, S. (1997). Stability and stabilization of delay differential systems. *Automatica, 33*(3), 347–354.

[HT98] Hennet, J. C., & Tarbouriech, S. (1998). Stability conditions of constrained delay systems via positive invariance. *International Journal of Robust and Nonlinear Control, 8*(3), 265–278.

[HTZ06] Hu, T., Teel, A., & Zaccarian, L. (2006). Stability and performances for saturated systems via quadratic and non-quadratic Lyapunov functions. *IEEE Transactions on Automatic Control, 51*(11), 1770–1786.

[Hu07] Hu, T. (2007). Nonlinear control design of linear differential inclusions via convex hull of quadratics. *Automatica Journal of IFAC, 43*(4), 685–692.

[HVCMB11] Hernandez-Vargas, E., Colaneri, P., Middleton, R., & Blanchini, F. (2011). Discrete-time control for switched positive systems with application to mitigating viral escape. *International Journal of Robust and Nonlinear Control, 21*(10, SI), 1093–1111.

[HvS04] Habets, L. C. G. J. M., & van Schuppen, J. H. (2004). A control problem for affine dynamical systems on a full-dimensional polytope. *Automatica Journal of IFAC, 40*(1), 21–35.

[ID90] Iftar, A., & Davison, E. J. (1990). Decentralized robust control for dynamic routing of large scale networks. In *Proceedings of the American Control Conference*. San Diego, CA, USA (pp. 441–446).

[Ift99] Iftar, A. (1999). A linear programming based decentralized routing controller for congested highways. *Automatica Journal of IFAC, 35*(2), 279–292.

[Ilc93] Ilchmann, A. (1993). *Non-identifier-based high-gain adaptive control*. *Lecture notes in control and information sciences* (Vol. 189). London: Springer-Verlag.

[IR94] Ilchmann, A., & Ryan, E. P. (1994). Universal λ-tracking for non-linearly-perturbed systems in the presence of noise. *Automatica Journal of IFAC, 30*(2), 337–346.

[JM10] Jones, C. N., & Morari, M. (2010). Polytopic approximation of explicit model predictive controllers. *IEEE Transactions on Automatic Control, 55*(11), 2542–2553.

[Joh00] Johansen, T. A. (2000). Computation of Lyapunov functions for smooth nonlinear systems using convex optimization. *Automatica Journal of IFAC, 36*(11), 1617–1626.

[Joh03] Johansson, M. (2003). *Piecewise linear control*. *Lecture notes in control and information sciences* (Vol. 284). Berlin: Springer.

[JW02] Jiang, Z. P., & Wang, Y. (2002) A converse Lyapunov theorem for discrete-time systems with disturbances. *Systems and Control Letters, 45*(1), 49–58.

[KAS88] Kapasouris, P., Athans, M., & Stein, G. (1988). Design of feedback control systems for stable plants with saturating actuators. In *Proceedings of 27th Conference on Decision and Control* (pp. 469–479). Austin: IEEE.

[KAS89] Kapasouris, P., Athans, M., & Stein, G. (1990). Design of feedback control systems for unstable plants with saturating actuators. In *Proceedings of IFAC Symposium on Nonlinear Control System Design*. Tarrytown, NY: Pergamon Press.

[KAS92] Kiendl, H., Adamy, J., & Stelzner, P. (1992). Vector norms as Lyapunov functions for linear systems. *IEEE Transactions on Automatic Control, 37*(6), 839–842.

[KDDS12] Kofman, E., De Doná, J. A., & Seron, M. M. (2012). Probabilistic set invariance and ultimate boundedness. *Automatica Journal of IFAC, 48*(10), 2670–2676.

[KG87] Keerthi, S. S., & Gilbert, E. G. (1987). Computation of minimum-time feedback control laws for discrete-time systems with state-control constraints. *IEEE Transactions on Automatic Control, 32*(5), 432–435.

[KG97] Kolmanovski, I. V., & Gilbert, E. G. (1997). Multimode regulators for systems with state and control constraints and disturbance input. In A. S. Morse (Ed.), *Control using logic-based switching*. *Lecture notes in control and information science* (Vol. 222, pp. 104–117). London: Springer-Verlag.

[KGC97] Kolmanovski, I. V., Gilbert, E. G., & Cook, J. A. (1997). Reference governor for supplemental torque source control in turbocharged diesel engines. In *American Control Conference* (pp. 652–656). Albuquerque, NM: IEEE.

[KH05] Kwon, W. H., & Han, S. (2005). *Receding Horizon Control*. London: Springer.

[Kha96] Khalil, H. (1996). *Nonlinear systems*. Upper Saddle River, NJ: Prentice Hall. Translated from the Russian by Scripta Technica.

[KPZ90] Khargonekar, P. P., Petersen, I. R., & Zhou, K. (1990). Robust stabilization of uncertain linear systems: Quadratic stabilizability and H_∞ control theory. *IEEE Transactions on Automatic Control, 35*(3), 356–361.

[Kra63] Krasowski, N. N. (1963). *Problems of the theory of stability of motion*. Stanford, CA: Stanford University Press. Translated from the Russian.

[Kra68] Krasnoselskii, M. A. (1968). *The operator of translation along the trajectories of differential equation*. Providence, RI: American Mathematical Society, Translations of Mathematical Monograph. Translated from the Russian by Scripta Technica.

[KS72] Kwakernaak, H., & Sivan, R. (1972). *Linear optimal control systems*. New York/London/Sydney: Wiley-Interscience.

[KV97] Kurzhanski, A., & Vályi, I. (1997). *Ellipsoidal calculus for estimation and control*. Boston, MA: Birkhäuser Boston.

[KWL00] Kouvaritakis, B., Wang, W., & Lee, Y. I. (2000). Observers in nonlinear model-based predictive control. *International Journal of Robust and Nonlinear Control, 10*(10), 749–761.

[LA09] Lin, H., & Antsaklis, P. (2009). Stability and stabilizability of switched linear systems: A survey of recent results. *IEEE Transactions on Automatic Control, 54*(2), 308–322.

[LAAC08] Limón, D., Alvarado, I., Alamo, T., & Camacho, E. F. (2008). MPC for tracking piecewise constant references for constrained linear systems. *Automatica Journal of IFAC, 44*(9), 2382–2387.

[Las87] Lasserre, J. B. (1987). A complete characterization of reachable sets for constrained linear time-varying systems. *IEEE Transactions on Automatic Control, 32*(9), 836–838.

[Las93] Lasserre, J. B. (1993). Reachable, controllable sets and stabilizing control of constrained systems. *Automatica Journal of IFAC, 29*(2), 531–536.

[Lay82] Lay, S. R. (1982). *Convex sets and their applications*. New York: Wiley.

[Lei79] Leitmann, G. (1979). Guaranteed asymptotic stability for some linear systems with bounded uncertainties. *Transactions of the ASME, 101*, 212–216.

[Lei81] Leitmann, G. (1981). On the efficacy of nonlinear control in uncertain systems. *ASME Journal of Dynamic Systems, Measurement and Control, 102*, 95–102.

[Lib03] Liberzon, D. (2003). *Switching in systems and control*. Boston: Birkhäuser.

[Liu68] Liu, R. W. (1968). Convergent systems. *IEEE Transactions on Automatic Control, 13*(4), 384–191.

[Lju99] Ljung, L. (1999). *System identification. Information and system science*. Upper Saddle River: Prentice Hall.

[LL61] LaSalle, J., & Lefschetz, S. (1961). *Stability by Lyapunov's direct method*. New York: Academic Press.

[LM99] Liberzon, D., & Morse, A. S. (1999). Basic problems in stability and design of switched systems. *IEEE Control Systems Magazine, 19*(5), 59–70.

[LO05] Limpiyamitr, A., & Ohta, Y. (2005, December 12–15). The duality relation between maximal output admissible set and reachable set. In *IEEE 44th Conference on CDC-ECC '05* (pp. 8282–8287).

[LOBN12] Lombardi, W., Olaru, S., Bitsoris, G., & Niculescu, S. I. (2012). Cyclic invariance for discrete time-delay systems. *Automatica Journal of IFAC. 48*(10), 2730–2733.

[LR95] Lawrence, D. A., & Rugh, W. J. (1995). Gain scheduling dynamic linear controllers for a nonlinear plant. *Automatica Journal of IFAC, 31*(3), 381–390.

[LS95] Lin, Y., & Sontag, E. D. (1995). Control-Lyapunov universal formulas for restricted inputs. *Control Theory Advanced Technology, 10*(4, Part 5), 1981–2004.

[LSS96] Lin, Z., Saberi, A., & Stoorvogel, A. (1996). Semiglobal stabilization of linear discrete-time systems subject to input saturation, via linear feedback—an ARE-based approach. *IEEE Transactions on Automatic Control, 41*(8), 1203–1207.

[LSW96] Lin, Y., Sontag, E. D., & Wang, Y. (1996). A smooth converse Lyapunov theorem for robust stability. *SIAM Journal on Control Optimization, 34*(1), 124–160.

[Lu98] Lu, W. M. (1998). Rejection of persistent $\mathcal{L}_\infty$-bounded disturbances for nonlinear systems. *IEEE Transactions on Automatic Control, 43*(12), 1692–1702.

[Lue69] Luenberger, D. G. (1969). *Optimization by vector space methods*. New York, USA: John Wiley & Sons Inc.

[Lya66] Lyapunov, A. M. (1966). *Stability of motions*. New York: Academic Press.

[MADF00] McConley, M. W., Appleby, B. D., Dahleh, M. A., & Feron, E. (2000). A computationally efficient Lyapunov-based scheduling procedure for control of nonlinear systems with stability guarantees. *IEEE Transactions on Automatic Control, 45*(1), 33–49.

[May01] Mayne, D. Q. (2001). Control of constrained dynamic systems. *European Journal of Control, 7*, 87–99.

[MB76] Morris, R., & Brown, R. F. (1976). Extension of validity of the GRG method in optimal control calculation. *IEEE Transactions on Automatic Control, 21*(4), 420–422.

[Meı74] Meılakhs, A. M. (1974). On the stabilization of linear controlled systems under condition of indeterminacy. *Automation and Remote Control, 35*(2), 182–184.

[Meı79] Meılakhs, A. M. (1979). Design of stable control systems subject to parametric perturbation. *Automation and Remote Control, 39*(10), 1409–1418.

[Mil95] Milnor, J. (1995). A Nobel prize for John Nash. *The Mathematical Intelligencer, 17*(3), 11–17.

[Mil02a] Milani, B. (2002). Computation of contractive polyhedra for discrete-time linear systems with saturating controls. *International Journal of Control, 75*(16–17), 1311–1320.

[Mil02b] Milani, B. (2002). Piecewise-affine Lyapunov functions for discrete-time linear systems with saturating controls. *Automatica Journal of IFAC, 38*(12), 2177–2184.

[MM96] Mesquine, F., & Mehdi, D. (1996). Constrained observer for linear continuous time systems. *International Journal of Systems Science, 27*(12), 1363–1369.

[MNV84] Michel, A. N., Nam, B. H., & Vittal, V. (1984). Computer generated Lyapunov functions for interconnected systems: improved results with applications to power systems. *IEEE Transactions on Circuits and Systems, 31*(2), 189–198.

[MP86a] Molchanov, A. P., & Pyatnitskii, E. S. (1986). Lyapunov functions that define necessary and sufficient conditions for absolute stability of nonlinear nonstationary control systems. I. *Automation and Remote Control, 47*(3), 344–354.

[MP86b] Molchanov, A. P., & Pyatnitskii, E. S. (1986). Lyapunov functions that define necessary and sufficient conditions for absolute stability of nonlinear nonstationary control systems. II. *Automation and Remote Control, 47*(4), 443–451.

[MP86c] Molchanov, A. P., & Pyatnitskii, E. S. (1986). Lyapunov functions that define necessary and sufficient conditions for absolute stability of nonlinear nonstationary control systems. III. *Automation and Remote Control, 47*(5), 620–630.

[MRFA06] Mayne, D. Q., Raković, S. V., Findeisen, R., & Allgöwer, F. (2006). Robust output feedback model predictive control of constrained linear systems. *Automatica, 42*(7), 1217–1222.

[MRRS00] Mayne, D. Q., Rawlings, J. B., Rao, C. V., & Scokaert, P. O. M. (2000). Constrained model predictive control: Stability and optimality. *Automatica Journal of IFAC, 36*, 789–814.

[MS82] Moss, F. H., & Segall, A. (1982). An optimal control approach to dynamic routing in networks. *IEEE Transactions on Automatic Control, 27*(2), 329–339.

[MS97] Mayne, D. Q., & Schroeder, W. R. (1997). Robust time-optimal control of constrained linear system. *Automatica Journal of IFAC, 33*(12), 2103–2118.

[MSP01] Martinelli, F., Shu, C., & Perkins, J. R. (2001). On the optimality of myopic production controls for single-server, continuous-flow manufacturing systems. *IEEE Transactions on Automatic Control, 46*(8), 1269–1273.

[MSR05] Mayne, D. Q., Seron, M. M., & Raković, S. (2005). Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica Journal of IFAC, 41*, 291–224.

[MT85] Milanese, M., & Tempo, R. (1985). Optimal algorithms theory for robust estimation and prediction. *IEEE Transactions on Automatic Control, 30*(5), 730–738.

[MTB04] Mesquine, F., Tadeo, F., & Benzaouia, A. (2004). Regular problem for linear systems with constraints on control and its increment or rate. *Automatica Journal of IFAC, 40*(8), 1387–1395.

[Nag42] Nagumo, M. (1942). Uber die lage der integralkurven gewöhnlicher differentialgleichungen. *Proceedings of the Physical-Mathematical Society of Japan, 24*(3), 272–559.

[NGOH13] Nguyen, H. N., Gutman, P. O., Olaru, S., & Hovd, M. (2013). Implicit improved vertex control for uncertain, time-varying linear discrete-time systems with state and control constraints. *Automatica Journal of IFAC, 49*(9), 2754–2759.

[NJ99] Nguyen, T., & Jabbari, F. (1999). Disturbance attenuation for systems with input saturation: An LMI approach. *IEEE Transactions on Automatic Control, 44*(4), 852–857.

[NJ00] Nguyen, T., & Jabbari, F. (2000). Output feedback controllers for disturbance attenuation with actuator amplitude and rate saturation. *Automatica Journal of IFAC, 36*(9), 1339–1346.

[ODDSS10] Olaru, S., De Doná, J. A., Seron, M. M., & Stoican, F. (2010). Positive invariant sets for fault tolerant multisensor control schemes. *International Journal of Control, 83*(12), 2622–2640.

[OIGH93] Ohta, Y., Imanishi, H., Gong, L., & Haneda, H. (1993). Computer generated Lyapunov functions for a class of nonlinear system. *IEEE Transactions on Circuits and Systems, 40*(5), 343–354.

[Ola92] Olas, A. (1992). On robustness of systems with structured uncertainties. In *Mechanics and control (Los Angeles, CA, 1991). Lecture notes in control and inform. sci.* (Vol. 170, pp. 156–169). Berlin: Springer.

[OM02] O'Dell, B. D., & Misawa, E. A. (2002). Semi-ellipsoidal controlled invariant sets for constrained linear systems. *ASME's Journal of Dynamic Systems, Measurement and Control, 124*(1), 98–103.

[O'R83] O'Reilly, J. (1983). *Observers for linear systems*. London/New York: Academic Press.

[Pad99] Padberg, M. (1999). *Linear optimization and extensions*. Berlin: Springer-Verlag.

[PB87] Petersen, I. R., & Barmish, B. R. (1987). Control effort considerations in the stabilization of uncertain dynamical systems. *Systems and Control Letters, 9*(5), 417–422.

[Pet85] Peteresen, I. R. (1985). Quadratic stabilizability of uncertain linear systems: Existence of a nonlinear stabilizing control does not imply existence of a linear stabilizing control. *IEEE Transactions on Automatic Control, 30*(3), 291–293.

[PH86] Petersen, I. R., & Hollot, C. (1986). A Riccati equation approach to the stabilization of uncertain system. *Automatica Journal of IFAC, 22*(3), 397–411.

[PHSG88] Parlos, A. G., Henry, A. F., Schweppe, F. C., & Gould, L. A. (1988). Nonlinear multivariable control of nuclear power on the unknown-but-bounded disturbances mode. *IEEE Transactions on Automatic Control, 33*(2), 130–137.

[Pic93] Piccardi, C. (1993). Infinite-horizon minimax control with pointwise cost functional. *Journal of Optimization Theory and Applications, 78*(2), 317–336.

[PN71] Pecsvaradi, T., & Narendra, K. S. (1971). Reachable sets for linear dynamical systems. *Information and Control, 19*, 319–345.

[PP14] Pin, G., & Parisini, T. (2014). On the robustness of nominal nonlinear minimum-time control and extension to non-robustly controllable target sets. *IEEE Transactions on Automatic Control, 59*(4), 863–875.

[PPA14] Poznyak, A., Polyakov, A., & Azhmyakov, V. (2014). *Attractive ellipsoids in robust control*. Charm: Springer International Publishing.

[Pro63] Propoi, A. I. (1963). Use of linear programming methods for synthesizing sampled data automatic systems. *Automation and Remote Control, 24*, 837–844.

[PV04] Pastravanu, O., & Voicu, M. (2004) Necessary and sufficient conditions for componentwise stability of interval matrix systems. *IEEE Transactions on Automatic Control, 49*(6), 1016–1021.

[PV06] Pastravanu, O., & Voicu, M. (2006). Generalized matrix diagonal stability and linear dynamical systems. *Linear Algebra and its Applications, 419*(2–3), 299–310.

[Qu98] Qu, Z. (1998). *Robust control of nonlinear uncertain systems*. New York: Wiley.

[Rac91] Rachid, A. (1991). Positively invariant polyhedral sets for uncertain discrete time systems. *Control Theory and Advanced Technology, 7*(1), 191–200.

[Ran95] Rantzer, A. (1995). Uncertain real parameters with bounded rate of variation. *The IMA Volumes in Mathematics and Its Applications, 74*, 345–349.

[RB10] Raković, S. V., & Barić, M. (2010). Parameterized robust control invariant sets for linear systems: theoretical advances and computational remarks. *IEEE Transactions on Automatic Control, 55*(7), 1599–1614.

[RBA07] Ren, W., Beard, R. W., & Atkins, E. M. (2007). Information consensus in multivehicle cooperative control. *IEEE Control Systems Magazine, 27*(2), 71–82.

[RBCM07] Raković, S. V., Blanchini, F., Cruck, E., & Morari, M. (2007). Robust obstacle avoidance for constrained linear discrete time systems: A set-theoretic approach. In *Proceedings of the 46th Conference on Decision and Control* (pp. 188–193). IEEE.

[RFFT14] Riverso, S., Farina, M., & Ferrari-Trecate, G. (2014). Plug-and-play model predictive control based on robust control invariant sets. *Automatica Journal of IFAC, 51*(8), 2179–2186.

[RHL77] Rouche, N., Habets, P., & Laloy, M. (1977). *Stability theory by Liapunov's direct method. Applied mathematical sciences* (Vol. 22, xii + 396 pp.). New York: Springer-Verlag. ISBN 0-387-90258-9.

[RJ98] Rantzer, A., & Johansson, M. (1998). Computation of piecewise quadratic functions for hybrid system. *IEEE Transactions on Automatic Control, 43*(4), 555–559.

[RK89] Rotea, M. A., & Khargonekar, P. P. (1989). Stabilization of uncertain systems with norm bounded uncertainty—a control Lyapunov function approach. *SIAM Journal on Control Optimization, 27*(6), 1462–1476.

[RK07] Raković, S., & Kouramas, K. (2007). The minimal robust positively invariant set for linear discrete-time systems: Approximation methods and control applications. In *Proceedings of the CDC06*, San Diego, California.

[RKC+12] Raković, S. V., Kouvaritakis, B., Cannon, M., Panos, C., & Findeisen, R. (2012). Parameterized tube model predictive control. *IEEE Transactions on Automatic Control, 57*(11), 2746–2761.

[RKCP12] Raković, S. V., Kouvaritakis, B., Cannon, M., & Panos, C. (2012). Fully parameterized tube model predictive control. *International Journal of Robust and Nonlinear Control, 22*(12), 1330–1361.

[RKFC12] Raković, S. V., Kouvaritakis, B., Findeisen, R., & Cannon, M. (2012). Homothetic tube model predictive control. *Automatica Journal of IFAC, 48*(8), 1631–1638.

[RKK+05] Raković, S. V., Kouramas, K. I., Kerrigan, E. C., Allwright, J. C., & Mayne, D. Q. (2005). On the minimal robust positively invariant set for linear difference inclusions. In *Proceedings of the CDCECC05*, Seville, Spain (pp. 2296–2301).

[RKKM05a] Raković, S., Kerrigan, E., Kouramas, K., & Mayne, D. (2005). Invariant approximations of the minimal robust positively invariant set. *IEEE Transactions on Automatic Control, 50*(3), 406–410.

[RKKM05b] Raković, S. V., Kerrigan, E. C., Kouramas, K. I., & Mayne, D. Q. (2005). Invariant approximation of the minimal robust positively invariant set. *IEEE Transactions on Automatic Control, 50*(3), 406–410.

[RKMK07] Raković, S., Kerrigan, E., Mayne, D., & Kouramas, K. (2007). Optimized robust invariance for linear discrete-time systems: theoretical foundations. *Automatica Journal of IFAC, 43*(5), 831–841.

[RKML06] Raković, S., Kerrigan, E., Mayne, D., & Lygeros, J. (2006). Reachability analysis of discrete-time systems with disturbances. *IEEE Transactions on Automatic Control, 51*(4), 546–561.

[RL14] Raković, S. V. & Lazar, M. (2014). The Minkowski-Lyapunov equation for linear dynamics: Theoretical foundations. *Automatica Journal of IFAC, 50*(8), 2015–2024.

[RM05] Raković, S., & Mayne, D. (2005). Robust time-optimal obstacle avoidance problem for constrained discrete-time systems. In *Proceedings of the Joint CDC–ECC 05*, Seville, Spain (pp. 831–841).

[Roc70] Rockafellar, R. T. (1970). *Convex analysis*. Princeton, NJ: Princeton University Press.

[RS60] Rota, G. C., & Strang, W. G. (1960). A note on the joint spectral radius. *Indagationes Mathematicae, 22*, 379–381.

[RW98] Rockafellar, R. T., & Wets, R. J. B. (1998). *Variational analysis*. New York: Springer.

[RW02] Richeson, D., & Wiseman, J. (2002). A fixed point theorem for bounded dynamical systems. *Illinois Journal of Mathematics, 46*(2), 491–495.

[SA90] Shamma, J. S., & Athans, M. (1990). Analysis of gain scheduled control for nonlinear plants. *IEEE Transactions on Automatic Control, 35*(8), 898–907.

[SA91] Shamma, J. S., & Athans, M. (1991). Guaranteed properties of gain scheduled control for linear parameter-varying plants. *Automatica Journal of IFAC, 27*(3), 559–564.

[SAI03] Sznaier, M., Amishima, T., & Inanc, T. (2003). $\mathcal{H}_2$ control with time-domain constraints: Theory and an application. *IEEE Transactions on Automatic Control, 48*(3), 355–368.

[SALB12] Spinu, V., Athanasopoulos, N., Lazar, M., & Bitsoris, G. (2012). Stabilization of bilinear power converters by affine state feedback under input and state constraints. *IEEE Transactions on Circuits and Systems, 59-II*(8), 520–524.

[Sav07] Savorgnan, C. (2007). *Synthesis of non-quadratic Lyapunov functions for robust constrained control*. PhD Thesis. Udine, Italy: University of Udine.

[SB92] Shahruz, S. M., & Behtash, S. (1992). Design of controllers for linear parameter-varying systems by the gain scheduling technique. *Journal of Mathematical Analysis and Applications, 168*(1), 195–217.

[Sch73] Schweppe, F. C. (1973). *Uncertain dynamic systems*. Englewood Cliff, NJ: Prentice Hall.

[SD87] Sznaier, M., & Damborg, M. J. (1987). Control of linear systems with state and control inequality constraint. In *Proceedings of the 26th Conference on Decision and Control*, Los Angeles, CA (pp. 761–762).

[SD90] Sznaier, M., & Damborg, M. J. (1990). Heuristically enhanced feedback control of constrained discrete-time systems. *Automatica Journal of IFAC, 26*(3), 521–532.

[SEG98] Scorletti, G., & El Ghaoui, L. (1998). Improved LMI conditions for gain scheduling and related control problems. *International Journal of Robust and Nonlinear Control, 8*(10), 845–877.

[SG05] Sun, Z., & Ge, S. S. (2005). *Switched linear systems control and design. Communications and control engineering*. London: Springer-Verlag.

[SG11] Sun, Z., & Ge, S. S. (2011). *Stability theory of switched dynamical systems*. London: Springer.

[Sha96a] Shamma, J. S. (1996). Linearization and gain scheduling. In W. Levine (Ed.), *The control handbook* (pp. 388–396). Boca Raton: CRC Press.

[Sha96b] Shamma, J. S. (1996). Optimization of the l^∞-induced norm under full state feedback. *IEEE Transactions on Automatic Control, 41*(4), 533–544.

[SHS02] Saberi, A., Han, J., & Stoorvogel, A. A. (2002). Constrained stabilization problems for linear plants. *Automatica Journal of IFAC, 38*(4), 639–654.

[Smi95] Smith, H. L. (1995). *Monotone dynamical systems: An introduction to the theory of competitive and cooperative systems. Mathematical surveys and monographs*. Providence: American Mathematical Society.

[SO13] Stoican, F., & Olaru, S. (2013). *Set-theoretic fault-tolerant control in multisensor systems*. Hoboken, NJ: John Wiley & Sons.

[SOH11] Scibilia, F., Olaru, S., & Hovd, M. (2011). On feasible sets for MPC and their approximations. *Automatica Journal of IFAC, 47*(1), 133–139.

[Son84] Sontag, E. D., An algebraic approach to bounded controllability of linear systems. *International Journal of Control, 39*(2), 181–188.

[Son98] Sontag, E. D. (1998). *Mathematical control theory: Deterministic finite-dimensional systems*. *Texts in applied mathematics* (Vol. 6, 2nd ed.). New York, USA: Springer-Verlag.

[SPS98] Pena, R. S., & Sznaier, M. (1998). *Robust systems, theory and applications*. New York: Wiley.

[SR00] Shamma, J. S., & Rugh, W. J. (2000). Research on gain scheduling. *Automatica Journal of IFAC, 36*(6), 1401–1425.

[Srz85] Srzednicki, R. (1985). On rest points of dynamical systems. *Fundamenta Mathematicae, 126*(1), 69–81.

[SS90a] Sabin, G. C. W., & Summers, N. (1990). Optimal technique for estimating the reachable set of a controlled n-dimensional linear system. *International Journal of Systems Science, 21*(4), 353–357.

[SS90b] Senin, E. I., & Soldunov, V. A. (1990). Attainable estimates of sets of feasible states of linear systems under limited disturbances. *Automation and Remote Control, 50*(11), 1513–1521.

[SSMAR99] Sznaier, M., Suárez, R., Miani, S., & Alvarez-Ramírez, J. (1999). Optimal l^∞ disturbance attenuation and global stabilization of linear systems with bounded control. *International Journal of Robust and Nonlinear Control, 9*(10), 659–675.

[SSS00] Saberi, A., Stoorvogel, A., & Sannuti, P. (2000). Communication and control engineering series. In *Control of linear systems with regulation and input constraints*. London: Springer-Verlag London, Ltd.

[ST97] Shamma, J. S., & Tu, K. Y. (1997). Approximate set-valued observers for nonlinear systems. *IEEE Transactions on Automatic Control, 42*(5), 648–658.

[ST99] Shamma, J. S., & Tu, K. Y. (1999). Set-valued observers and optimal disturbance rejection. *IEEE Transactions on Automatic Control, 44*(2), 253–264.

[Sta95a] Stalford, H. L. (1995). Robust asymptotically stabilizability of Petersen's counterexample via a linear static controller. *IEEE Transactions on Automatic Control, 40*(8), 1488–1491.

[Sta95b] Stalford, H. L. (1995). Scalar-quadratic stabilizability of the Petersen's counterexample via a linear static controller. *Journal of Optimization Theory and Applications, 86*(2), 327–346.

[Sto95] Stoorvogel, A. A. (1995). Nonlinear L_1 optimal controllers for linear systems. *IEEE Transactions on Automatic Control, 40*(4), 694–696.

[SWM+07] Shorten, R., Wirth, F., Mason, O., Wulff, K., & King, C. (2007). Stability criteria for switched and hybrid systems. *SIAM Review, 49*(4), 545–592.

[Szn93] Sznaier, M. (1993). A set induced norm approach to the robust control of constrained systems. *SIAM Journal on Control Optimization, 31*(3), 733–746.

[TB97] Tsitsiklis, J. N., & Blondel, V. D. (1997). The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate. *Math. Control Signals Systems, 10*(1), 31–40.

[TCD04] Tempo, R., Calafiore, G., & Dabbene, F. (2004). *Randomized algorithms for analysis and control of uncertain systems. Communications and control engineering series*. London: Springer-Verlag.

[TJB03] Tøndel, P., Johansen, T. A. & Bemporad, A. (2003). An algorithm for multi-parametric quadratic programming and explicit MPC solutions. *Automatica Journal of IFAC, 39*(3), 489–497.

[USGW82] Usoro, P. B., Schweppe, F. C., Gould, L. A., & Wormley, D. N. (1982). A Lagrange approach to set theoretic control synthesis. *IEEE Transactions on Automatic Control, 27*(2), 393–399.

[VB89] Vassilaki, M., & Bitsoris, G. (1989). Constrained regulation of linear continuous-time dynamical systems. *Systems and Control Letters, 13*, 247–252.

[VHB88] Vassilaki, M., Hennet, J. C., & Bitsoris, G. (1988). Feedback control of linear discrete-time systems under state and control constraints. *International Journal of Control, 47*(6), 1727–1735.

[VJ14] Vlassis, N., & Jungers, R. (2014). Polytopic uncertainty for linear systems: New and old complexity results. *Systems and Control Letters, 67*(0), 9–13.

[Voi84] Voicu, M. (1984). Componentwise asymptotic stability of linear constant dynamical systems. *IEEE Transactions on Automatic Control, 29*(10), 937–939.

[Vor98] Vorotnikov, V. I. (1998). *Partial stability and control*. Boston, MA: Birkhauser.

[VZ96] Vicino, A., & Zappa, G. (1996). Sequential approximation of feasible parameter sets for identification with set membership uncertainty. *IEEE Transactions on Automatic Control, 41*(6), 774–785.

[WB94] Wredenhagen, G. F., & Belanger, P. R. (1994). Piecewise-linear LQ control for systems with input constraint. *Automatica Journal of IFAC, 30*(3), 403–416.

[Wit68a] Witsenhausen, H. S. (1968). A minimax control problem for sampled linear systems. *IEEE Transactions on Automatic Control, 13*(1), 5–21.

[Wit68b] Witsenhausen, H. S. (1968). Set of possible states of linear systems given perturbed observation. *IEEE Transactions on Automatic Control, 13*, 556–558.

[XA00] Xu, X., & Antsaklis, P. J. (2000). Stabilization of second-order LTI switched systems. *International Journal of Control, 73*(14), 1261–1279.

[XHH00] Xiao, L., Hassibi, A., & How, J. P. (2000). Control with random communication delays via a discrete-time jump system approach. In *Proceedings of the American Control Conference, 2000* (Vol. 3, pp. 2199–2204).

[Yfo10] Yfoulis, C. A. (2010). Constrained switching stabilization of linear uncertain switched systems using piecewise linear Lyapunov functions. *Transactions of the Institute of Measurement and Control, 32*(5), 529–566.

[Yor67] Yorke, J. (1967). Invariance for ordinary differential inclusions. *Mathematical Systems Theory, 1*(4), 353–372.

[Yos75] Yoshizawa, T. (1975). *Stability theory and existence of periodic solutions and almost periodic solution*. New York: Springer Verlag.

[Zan87] Zanolin, F. (1987). Bound sets, periodic solutions and flow-invariance for ordinary differential equations in $\mathbb{R}^n$: Some remarks. *Rendiconti dell'Istituto di Matematica dell'Università di Trieste, XIX*.

[ZDG96] Zhou, K., Doyle, J. C., & Glower, K. (1996). *Robust and optimal control*. Englewood Cliff, NJ: Prentice-Hall.

[Zel94] Zelentsowsky, A. L. (1994). Nonquadratic Lyapunov functions for robust stability analysis of linear uncertain system. *IEEE Transactions on Automatic Control, 39*(1), 135–138.

[ZSCH05] Zhang, L., Shi, Y., Chen, T., & Huang, B. (2005). A new method for stabilization of networked control systems with random delays. *IEEE Transactions on Automatic Control, 50*(8), 1177–1181.

[Zub64] Zubov, V. I. (1964) *Methods of A.M. Lyapunov and their applications*. The Netherlands: Noordhoff.

Index

F. Blanchini, S. Miani, *Set-Theoretic Methods in Control*, Systems & Control: Foundations & Applications, DOI 10.1007/978-3-319-17933-9