Studies in Systems, Decision and Control 185

Andrea Bacciotti

Stability and Control of Linear Systems

Springer

Studies in Systems, Decision and Control

Volume 185

The series "Studies in Systems, Decision and Control" (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control–quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/13304

Andrea Bacciotti

Stability and Control of Linear Systems

Springer

Andrea Bacciotti
Dipartimento di Scienze Matematiche
"G.L. Lagrange" (DISMA: Dipartimento
di eccellenza 2018–22)
Politecnico di Torino
Turin, Italy

ISSN 2198-4182 ISSN 2198-4190 (electronic)
Studies in Systems, Decision and Control
ISBN 978-3-030-02404-8 ISBN 978-3-030-02405-5 (eBook)
https://doi.org/10.1007/978-3-030-02405-5

Library of Congress Control Number: 2018957665

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Giannina

Preface

This book is the natural outcome of a course I taught for many years at the Technical University of Torino, first for students enrolled in the aerospace engineering curriculum, and later for students enrolled in the applied mathematics curriculum. The aim of the course was to provide an introduction to the main notions of system theory and automatic control, with a rigorous theoretical framework and a solid mathematical background.

Throughout the book, the reference model is a finite-dimensional, time-invariant, multivariable linear system. The exposition is basically concerned with the time-domain approach, but also the frequency-domain approach is taken into consideration. In fact, the relationship between the two approaches is discussed, especially for the case of single-input–single-output systems. Of course, there are many other excellent handbooks on the same subject (just to quote a few of them, [3, 6, 8, 11, 14, 23, 25, 27, 28, 32]). The distinguishing feature of the present book is the treatment of some specific topics which are rare to find elsewhere at a graduate level. For instance, bounded-input–bounded-output stability (including a characterization in terms of canonical decompositions), static output feedback stabilization (for which a simple criterion in terms of generalized inverse matrices is proposed), controllability under constrained controls.

The mathematical theories of stability and controllability of linear systems are essentially based on linear algebra, and it has reached today a high level of advancement. During the last three decades of the past century, a great effort was done, in order to develop an analogous theory for nonlinear systems, based on differential geometry (see [7] for a historical overview). For this development, usually referred to as *geometric control theory*, we have today a rich literature ([2, 5, 13, 18–20, 26, 30]). However, I believe that the starting point for a successful approach to nonlinear systems is a wide and deep knowledge of the linear case. For this reason, while this book is limited to the linear context, in the presentation and organization of the material, as well as in the selection of topics, the final goal I had in mind is to prepare the reader for such a nonlinear extension.

Concerning the prerequisites, I assume that the reader is familiar with basic differential and integral calculus (for real functions of several real variables) and linear algebra. Some notions of complex analysis are required in the frequency-domain approach. The book can be used as a reference book for basic courses at a doctoral (or also upper undergraduate) level in mathematical control theory and in automatic control. More generally, parts of this book can be used in applied mathematics courses, where an introduction to the point of view of system theory and control philosophy is advisable. The perspective of control systems and the stability problem are indeed ubiquitous in applied sciences and witness a rapidly increasing importance in modern engineering. At a postdoctoral level, this book can be recommended for reading courses both for mathematician oriented to engineering applications and engineers with theoretical interests. To better focus on the main concepts and results, some more technical proofs are avoided or limited to special situations. However, in these cases, appropriate bibliographic references are supplied for the curious reader.

It follows a short description of the contents. The first chapter aims to introduce the reader to the “point of view” of system theory: In particular, the notions of input–output operator and external stability are given. The second chapter deals with systems without external forces which reduce, according to a more classical terminology, to homogeneous systems of linear differential equations. In view of the application, we are interested in, the representation of the general integral in terms of exponential matrix and Jordan form is crucial, and it is treated in detail. Chapter 3 is devoted to Lyapunov stability theory of the equilibrium position of a linear unforced system. The results reported in this chapter are classical but very important for the following chapters. In Chap. 4, we present some alternative approaches to the representation of solutions of a nonhomogeneous (i.e., with forcing term) system of linear differential equations: variation of constants, undetermined coefficients, Laplace transform. In Chap. 5 we finally begin the study of linear systems in a control perspective. We discuss the notions of controllability and observability, their analogies and characterizations, and the corresponding canonical forms. The final section treats shortly the controllability problem under constrained control, in view of possible applications to optimization theory. In Chap. 6, we address the bounded-input–bounded-output stability problem, and we propose a characterization using the canonical decompositions introduced in Chap. 5. Chapter 7 is devoted to various aspects of the stabilization problem: asymptotic controllability, static state feedback stabilization, static output feedback stabilization, dynamic output feedback stabilization. In particular, we re-propose in a new setting some old results about static output feedback stabilization. In author’s opinion, these results are very interesting, but neglected in the current literature. Finally, in Chap. 8, we introduce the frequency-domain approach and study the relationship with the time-domain approach. Two appendices follow. In the first one, the notions of internal stability are introduced. These notions are formulated with respect to a system of nonlinear ordinary differential equations. In fact, only in this part of the book nonlinear systems came into play. The reason of this choice is that all the aspects of the stability notions became more evident in the nonlinear context. The second appendix is a short list of useful facts about Laplace transform.

Finally, I wish to thank students, colleagues, and coworkers who contributed in many ways to improve the content of this book. A special thanks to Luisa Mazzi and Francesca Ceragioli.

Turin, Italy Andrea Bacciotti

Contents

Notations and Terminology

- We denote by **N**, **Z**, **R**, and **C**, respectively, the set of natural, integer, real, and complex numbers. The symbol $|a|$ denotes the modulus of a if $a \in \mathbf{C}$, and the absolute value of a if $a \in \mathbf{R}$. The symbol sgn a denotes the sign function (i.e., a if $a \geq 0$, $-a$ if $a < 0$). If $z \in C$, $\bar{z}$ Re z, im z denote respectively the conjugate, the real part and the imaginary part of z.
- If V is a (real or complex) vector space of finite dimension with dim $V = n$, the components of $v \in V$ with respect to a given basis are usually denoted by $v_1, \ldots, v_n$ and we also write $v = (v_1, \ldots, v_n)$. This notation is simple, but it may give rise to some ambiguity when we deal with several vectors distinguished by indices; in these cases, we write $(v_i)_j$ to denote the j-th component of the i-th vector. The subspace of V generated by a subset $U \subset V$ is denoted by span U.
- If v is an element of a finite-dimensional normed vector space V, the norm of v is generically denoted by $||v||_V$ or, when the space is clear from the context, simply by $||v||$. If v is a vector of $\mathbf{R}^n$, and if not differently stated, $||v||$ denotes the Euclidean norm, i.e., $||v|| = (\sum_{i=1}^n v_i^2)^{1/2}$.
- Let m and n be fixed. We denote $\mathcal{M}(\mathbf{R})$ the vector space of all the matrices with m rows and n columns, with real entries. A similar notation with **R** replaced by **C** is adopted for matrices with complex entries. If $M \in \mathcal{M}(\mathbf{R})$, we may also say that M is a $m \times n$ matrix. Of course, $\mathcal{M}(\mathbf{R})$ can be identified with $\mathbf{R}^{m \times n}$: However, note that to this end, $\mathbf{R}^{m \times n}$ should be considered different from $\mathbf{R}^{m \times n}$. A matrix M is said to be *square* if $n = m$. In this case, we may also say that M has dimension n.

  To specify the entries of a matrix, we write

$$M = (m_{ij})_{i=1,\ldots,n,\, j=1,\ldots,m}$$

  (the first index specifies the row, the second one the column). As for vectors, we may assign a norm to a matrix. In this book, we use the so-called Frobenius norm

$$||M|| = \sqrt{\sum_{i,j} |m_{ij}|^2} \, .$$

- The identity matrix of dimension n is denoted by I_n, or simply by I when the dimension is clear from the context. If M is a square matrix, we denote by $M^{\rm t}$ the transpose of M. We also denote respectively by tr M, det M, and rank M the trace, the determinant, and the rank of M. The symbol ker M represents the kernel of M, the symbol im M the image (range) of M. The characteristic polynomial of a $n \times n$ matrix M is written $p_M(\lambda)$. It is defined by

$$p_M(\lambda) = \det(M - \lambda I) = (-1)^n \det(\lambda I - M).$$

  Note in particular that $\deg p_M(\lambda) = n$ (where $\deg P(\lambda)$ denotes the degree of a polynomial $P(\lambda)$) and that for each n, $(-1)^n p_M(\lambda)$ is a *monic* polynomial (which means that the coefficient of λ^n is 1).
- Recall that the eigenvalues of a square matrix M are the roots of the characteristic polynomial of M, that is, the solutions of the algebraic equation $p_M(\lambda) = 0$. The set of distinct eigenvalues of M constitutes the *spectrum* of M. It is denoted by $\sigma(M)$, and it is, in general, a subset of the complex plane $\boldsymbol{C}$. Recall also that the matrices A and B are *similar* if there exists a nonsingular matrix P such that $B = P^{-1}AP$.
  An eigenvector of M corresponding to an eigenvalue λ is a nontrivial solution of the linear algebraic system $(M - \lambda I)v_0 = 0$. The dimension of the subspace generated by all the eigenvalues of an eigenvalue λ of A is called the *geometric multiplicity* of λ. The geometric multiplicity is less than or equal to the algebraic multiplicity of λ.
  Let v_0 be an eigenvector of M; the finite sequence of vectors $v_1, \ldots, v_k$ forms a chain of generalized eigenvectors generated by v_0 if $(M - \lambda I)v_1 = v_0$, $(M - \lambda I)v_2 = v_1, \ldots,$ $(M - \lambda I)v_k = v_{k-1}$.
- If A is a subset of $\mathbf{R}^n$, we denote respectively by $\overset{\circ}{A}$, $\bar{A}$, ∂A the set of the interior points of A, the closure of A, the boundary of A (in the topology of $\mathbf{R}^n$).
- If A and B are two arbitrary sets, $\mathcal{F}(A, B)$ denotes the set of all the functions from A to B. In particular:
  - $\mathcal{C}(I, U)$ denotes the set of all the continuous functions defined in I with values in U, where I is an interval (open or closed, bounded or unbounded) of real numbers and $U \subseteq \mathbf{R}^n$;
  - $\mathcal{PC}([a, b], U)$ denotes the set of all the piecewise continuous,[1] right-continuous functions defined on $[a, b]$ with values in U, where a and b are real numbers ($a < b$) and $U \subseteq \mathbf{R}^n$;

[1]Recall that a function is piecewise continuous on a compact interval $[a, b]$ if in this interval it has at most finitely many discontinuity points, and each possible discontinuity point is a jump.

- $\mathcal{PC}([a, +\infty), U)$, where $a \in \mathbf{R}$ and $U \subseteq \mathbf{R}^n$, denotes the set of all the piecewise continuous[2] right-continuous functions defined on the interval $[a, +\infty)$ with values in U (the sets $\mathcal{PC}((-\infty, +\infty), U)$ and $\mathcal{PC}((-\infty, b], U)$ are defined in analogous way);
- $\mathcal{B}(I, \mathbf{R}^n)$ denotes the set of all the bounded functions defined in the interval I with values in $\mathbf{R}^n$;
- we may use also the notations

$$\mathcal{CB}([a, +\infty), \mathbf{R}^n) \quad \text{and} \quad \mathcal{PCB}([a, +\infty), \mathbf{R}^n)$$

  to denote the sets of all the bounded functions which belong respectively to the sets

$$\mathcal{C}([a, +\infty), \mathbf{R}^n) \quad \text{and} \quad \mathcal{PC}([a, +\infty), \mathbf{R}^n).$$

- If the function $f(\cdot)$ is an element of a functional normed vector space $\mathcal{V}$, its norm is denoted by $||f(\cdot)||_{\mathcal{V}}$. In particular, if $f(\cdot) \in \mathcal{B}(I, \mathbf{R}^n)$, we will write $||f(\cdot)||_\infty = sup_{t\in I} ||f(t)||$ (norm of the uniform convergence).
- Depending on the circumstances, for the derivative of a function $f(t) : \mathbf{R} \to \mathbf{R}^n$, the following symbols can be used: $\frac{df}{dt}$, $f'(t)$, $\dot{f}(t)$, $(Df)(t)$. For higher-order derivatives, we write $f^{(k)}(t)$.
- A rational function has the form $R(s) = N(s)/D(s)$ where $N(s)$ and $D(s)$ are polynomials. It is usually thought of as a function from $\mathbf{C}$ to $\mathbf{C}$. A rational function is said to be *proper* if $\deg N(s) < \deg D(s)$. Other agreements about rational functions will be specified later in Chap. 8 (see in particular Remark 8.2)

[2] Recall that a function is piecewise continuous on a unbounded interval I if it is piecewise continuous on every compact interval $[c, b] \subset I$.

Chapter 1
Introduction

Many phenomena observed in the real world, regardless to their different nature, involve several physical quantities and result from the interaction of various components: for these reasons, in these situations the term "system" is generally used.

The experimental information obtained by studying a physical system gives often rise to the construction of a mathematical model. In this way, it can be easily communicated and elaborated qualitatively or numerically, and possibly employed to control the evolution of the system. In this book, the term *system* will be often referred to the mathematical model, rather than the represented real phenomenon.

Without any pretence of giving an axiomatic definition, the present introductory chapter aims to describe informally the main features of the notion of system, and the way we can take advantages of them.

1.1 The Abstract Notion of System

Direct experience shows that a system is often subject to time evolution. This means that the numerical values of the physical quantities characterizing the state of the system change while time passes. For this reason, they will be treated as variables. The changes are due, in general, to the action of internal forces and constraints, as well as of possible external forces or signals.

1.1.1 The Input-Output Operator

In order to provide an abstract description of the evolution of a system, we need to assign the following objects:

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_1

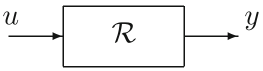
(1) a set $\mathcal{T}$ representing the time;
(2) a set $\mathcal{U}$, whose members represent the numerical values of the possible external variables (forces or signals);
(3) a set $\mathcal{X}$, whose members represent the numerical values of the internal state variables;
(4) a set $\mathcal{Y}$, whose members represent the numerical values of the variables carrying the information provided by the system about its internal state.

The action exerted on the system by the external world during the evolution is therefore represented by a function $u(\cdot) \in \mathcal{F}(\mathcal{T}, \mathcal{U})$: it is called the *input map*. The response of the system, that is the available information about the state of the system during the evolution, is represented by a function $y(\cdot) \in \mathcal{F}(\mathcal{T}, \mathcal{Y})$: it is called the *output map*. Finally, the internal state of the system, during the evolution, is represented by a function $x(\cdot) \in \mathcal{F}(\mathcal{T}, \mathcal{X})$, called the *state evolution map*. The sets $\mathcal{U}$, $\mathcal{Y}$ and $\mathcal{X}$ are respectively called the *input set*, the *output set* and the *state set*.

The system acts as on operator $\mathcal{R}$ transforming elements $u(\cdot) \in \mathcal{F}(\mathcal{T}, \mathcal{U})$ to elements $y(\cdot) \in \mathcal{F}(\mathcal{T}, \mathcal{Y})$. We will write

$$y(\cdot) = \mathcal{R}(u(\cdot)) \ , \quad \text{with} \quad \mathcal{R} : \mathcal{F}(\mathcal{T}, \mathcal{U}) \to \mathcal{F}(\mathcal{T}, \mathcal{Y}) \ .$$

The operator $\mathcal{R}$ is called the *input-output operator*. In system theory, the action of an input-output operator $\mathcal{R}$ is often represented graphically by means of a flow chart, as the following figure shows.



An input-output operator is not necessarily defined for each element $u(\cdot) \in \mathcal{F}(\mathcal{T}, \mathcal{U})$. For instance, in some applications the input maps need to satisfy some constraints, imposed by the nature of the problem either on their values or on their functional character. The subset of $\mathcal{F}(\mathcal{T}, \mathcal{U})$ formed by all the maps satisfying the required conditions constitutes the domain of $\mathcal{R}$. It is called the *set of admissible inputs*.

The following subsections aim to specify the nature of the sets $\mathcal{T}, \mathcal{U}, \mathcal{X}, \mathcal{Y}$.

Remark 1.1 When we assume that the output map can be uniquely and exactly determined by applying the operator $\mathcal{R}$, we are implicitly assuming that we have a full knowledge about the structure of the system and the physical laws governing its evolution. But in real situations this is not always true. In practice, it might happen that repeated experiments (with the same input map) give rise to different outputs, or that the output is affected by imprecisions, due to one or more of the following reasons: neglecting or simplifying some details during the modeling process; measurement errors; uncertainty in parameters identification; random phenomena. To face these or similar situations, suitable extensions of the theory need to be developed. But these will not be considered in this book.

Remark 1.2 A comment about terminology is in order. The word "system" is often used with lightly different meanings in the common language, and sometimes also in the technical literature. For instance, in Mathematics, "system" classically means "set of coupled equations". A system corresponding to an input-output operator as above, should be more properly called an *input-output system*. However, throughout this book, we prefer to use for simplicity the term "system" also in this case. The ambiguity is not serious. The right sense can be easily understood every time from the context.

1.1.2 Discrete Time and Continuous Time

The time can be represented by any totally ordered set $\mathcal{T}$, endowed with a group structure. In practice, we have two possible choices: either $\mathcal{T} = \mathbf{Z}$ or $\mathcal{T} = \mathbf{R}$. In the former case we speak about *discrete time systems*: the functions representing the input, the state and the output are actually sequences. In the latter case we speak about *continuous time systems*. It may happens that a physical system can be modeled both as a discrete time system and as a continuous time system. This may depends on the purposes of the search, on the measure scales and on the measure devices. Sometimes, different representations of the same physical system provide complementary information.

1.1.3 Input Space and Output Space

We can distinguish several types of inputs variables. A *disturbance* is a undesired signal, which cannot be governed, and sometimes not even measured in real time. A *reference signal* is an input representing the ideal evolution, to be tracked by the real evolution of the system. A *control* is an input completely determined by the decisions of a supervisor, which can be used to modify the behavior of the system during the evolution.

In general, we assume that the value of each single physical quantity involved in the model can be expressed by a real number. Moreover, it is convenient to order the input variables (and, separately, the output variables), and to rewrite them as the components of a vector. It is therefore natural to assume that the input set $\mathcal{U}$ and the output set $\mathcal{Y}$ are endowed with the structure of a real vector spaces.

1.1.4 State Space

In common applications, the number of the state variables is usually greater than the number of the input and output variables. Moreover, the state variables are difficult to identify, since in general they are not directly available to the observation. Sometimes, one should think of the state variables as mathematical idealizations, inherent to the model. We will assume that also the state set $\mathcal{X}$ has the structure of a real vector space.

1.1.5 Finite Dimensional Systems

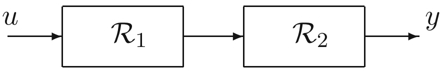
We say that a system is *finite dimensional* when the input variables, the output variables and the state variables can be represented as vectors with finitely many real components. Thus, for a finite dimensional system, it is natural to assume $\mathcal{X} = \mathbf{R}^n$, $\mathcal{U} = \mathbf{R}^m$, $\mathcal{Y} = \mathbf{R}^p$, where n, m, p are given integers, greater than or equal to 1. The sets of functions representing the input, the output and the state maps will be therefore respectively denoted by $\mathcal{F}(\mathbf{R}, \mathbf{R}^m)$, $\mathcal{F}(\mathbf{R}, \mathbf{R}^p)$, $\mathcal{F}(\mathbf{R}, \mathbf{R}^n)$. In particular, the system is said to be SISO (single-input-single-output) when $m = p = 1$; otherwise, the system is said to be MIMO (multi-input-multi-output).

Remark 1.3 From now on, by the term *system* we mean a finite dimensional, time continuous system.

1.1.6 Connection of Systems

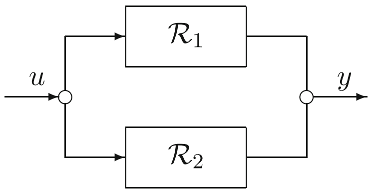
In some applications, it is necessary to manage simultaneously several systems, and to enable connections among them. The result of these manipulations may be often reviewed as a new system. On the contrary, it may be sometimes convenient to decompose a given system as the connection of certain subsystems. Let two systems, whose representing operators are denoted respectively by $\mathcal{R}_1$ and $\mathcal{R}_2$, be given. We describe below three basic types of connections.

(1) *Cascade connection*. The input of the second system coincides with the output of the first system.



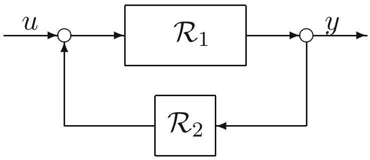
If we denote by $\mathcal{R}$ the operator representing respectively the new resulting system, we have $\mathcal{R} = \mathcal{R}_2 \circ \mathcal{R}_1$, where $\circ$ denotes the composition of maps.

(2) *Parallel connection*. The first and the second system have the same input and both contribute to determine the output.



If we denote as before by $\mathcal{R}$ the operators representing the resulting system, a typical parallel connection is obtained by taking $\mathcal{R} = \mathcal{R}_1 + \mathcal{R}_2$.

(3) *Feedback connection*. The output of the first system is conveyed into the second system; it is elaborated, possibly combined with other external inputs and finally re-injected into the input channel of the first system.



In this case, the operator $\mathcal{R}$ representing the result of the connection of $\mathcal{R}_1$ and $\mathcal{R}_2$, is implicitly defined by the relation $y(\cdot) = \mathcal{R}_1(\mathcal{R}_2(y(\cdot)) + u(\cdot))$.

These types of connection can be combined, to obtain very general patterns. As a rule, the complexity of a system becomes greater and greater, as the number of connections increases.

1.1.7 System Analysis

The purpose of the *analysis of a system* is the study of the properties of the input-output operator. For instance, it is interesting to estimate how the energy carried by the output signal depends on the energy carried by the input signal. To this end, it is necessary to assume that the spaces of the input maps and of the output maps are endowed with a structure of normed vector space. For the moment we do not need to chose a specific norm, which may depend on the particular application. For simplicity, we continue to use the notation $\mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ and $\mathcal{F}(\mathbf{R}, \mathbf{R}^p)$ for the space of the input maps and the space of the output maps, but remember that from now on they are normed space. The norms on these spaces are respectively denoted by $|| \cdot ||_{\mathcal{F}(\mathbf{R},\mathbf{R}^m)}$ and $|| \cdot ||_{\mathcal{F}(\mathbf{R},\mathbf{R}^p)}$.

Informally, it is used to say that a system is *externally stable* when each bounded input map generates a bounded output map. More precisely, we give the following definition.

Definition 1.1 A system, or its representing operator $\mathcal{R}$, is said to be *BIBO-stable* (i.e., *bounded-input-bounded-output-stable*) with respect to the norms $|| \cdot ||_{\mathcal{F}(\mathbf{R},\mathbf{R}^m)}$ and $|| \cdot ||_{\mathcal{F}(\mathbf{R},\mathbf{R}^p)}$ if for each real number $R > 0$ there exists a real number $S > 0$ such that for each input map $u(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ one has

$$||u(\cdot)||_{\mathcal{F}(\mathbf{R},\mathbf{R}^m)} \leq R \implies ||y(\cdot)||_{\mathcal{F}(\mathbf{R},\mathbf{R}^m)} \leq S$$

where $y(\cdot)$ is the output map of the system corresponding to the input map $u(\cdot)$.

Notice that according to Definition 1.1, the output is allowed to be different from zero, even if the input vanishes.

Proposition 1.1 *A system, or its representing operator* $\mathcal{R}$, *is* BIBO-stable *if and only if there exists a continuous and non-decreasing function*

$$\alpha(r) : [0, +\infty) \to [0, +\infty)$$

such that for each input map $u(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ *one has:*

$$||y(\cdot)||_{\mathcal{F}(\mathbf{R},\mathbf{R}^p)} \leq \alpha(||u(\cdot)||_{\mathcal{F}(\mathbf{R},\mathbf{R}^m)}). \tag{1.1}$$

The meaning of (1.1) can be explained as follows: if the energy carried by the input signal is bounded, then the energy of the output signal can be estimated in terms of the energy of the input signal. The value of $\alpha(0)$ is sometimes called the *bias* term, while the function $\alpha(r) - \alpha(0)$ is called the *gain function*.

In system analysis, it is very important to know the conditions under which a system is BIBO-stable and, in the positive case, to give information about the shape of the function α.

1.1.8 Control System Design

The *control system design* consists in the development of a control strategy, to be exerted throughout the input channel, in such a way that the output matches a reference signal as well as possible. Roughly speaking, we can distinguish two kinds of control strategies.

(1) *Open loop control*. The control is realized as a function of time $u(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$, and directly injected into the system.
(2) *Closed loop control*. The control is implemented by constructing a second system and establishing a feedback connection.

The closed loop control strategy is also called *automatic control*. It provides some advantages. Indeed, it enables the system to self-regulate, also in presence of unpredictable perturbations, without the need of intervention of a human supervisor. Let us use the term *plant* to denote the system to be controlled, and let us denote by $\mathcal{R}_P : \mathcal{F}(\mathbf{R}, \mathbf{R}^m) \to \mathcal{F}(\mathbf{R}, \mathbf{R}^p)$ the corresponding operator. Let us call *compensator* or *controller* the system to be designed, and let us denote by $\mathcal{R}_C : \mathcal{F}(\mathbf{R}, \mathbf{R}^p) \to \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ the representing operator. The closed loop control strategy consists basically in the following procedure. The output of the plant is monitored and compared with the reference signal; when a unacceptable difference between the two signals is detected, the compensator is activated and the necessary corrections are sent to the plant.

When it is possible to observe directly the state of the system, the compensator can be realized as an operator $\mathcal{R}_C : \mathcal{F}(\mathbf{R}, \mathbf{R}^n) \to \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$. We will use the terms *output feedback* or *state feedback* when we need to distinguish the two situations.

1.1.9 Properties of Systems

In this section we aim to discuss the properties that a generic operator $\mathcal{R} : \mathcal{F}(\mathbf{R}, \mathbf{R}^m) \to \mathcal{F}(\mathbf{R}, \mathbf{R}^p)$ is expected to satisfy in order to represent a real physical system.

1.1.9.1 Causal Systems

Usually, systems encountered in applications are *causal* (or *non anticipative*). This means that for each $t \in \mathbf{R}$ and for each pair of input maps $u_1(\cdot), u_2(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$, if

$$u_1(\tau) = u_2(\tau) \text{ for each } \tau \le t \ ,$$

then

$$y_1(t) = y_2(t)$$

where $y_1(\cdot) = \mathcal{R}(u_1(\cdot))$ and $y_2(\cdot) = \mathcal{R}(u_2(\cdot))$. In other words, the value of the output at any instant t is determined only by the values that the input map takes at the interval $(-\infty, t]$.

1.1.9.2 Time Invariant Systems

We say that a system, or its representing operator, is *time invariant* if for each $t, T \in \mathbf{R}$ and for each input map $u(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$, one has

$$z(t) = y(t - T)$$

where

$$v(t) = u(t - T), \quad y(\cdot) = \mathcal{R}(u(\cdot)), \quad z(\cdot) = \mathcal{R}(v(\cdot)).$$

In other words, if the input signal is delayed (or anticipated) of a fixed duration, also the output signal is delayed (or anticipated) of the same duration, but its shape is unchanged. Time invariant systems are also called *stationary*, or *autonomous*.

1.1.9.3 Linear Systems

A system is said to be *linear* when its input-output operator is linear, that is

$$a_1\mathcal{R}(u_1(\cdot)) + a_2\mathcal{R}(u_2(\cdot)) = \mathcal{R}(a_1u_1(\cdot) + a_2u_2(\cdot))$$

for each pair of input maps $u_1(\cdot), u_2(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ and each pair of scalars a_1, a_2.

Notice that this definition makes sense, since the input and the output sets are vector spaces.

1.2 Impulse Response Systems

In this section we try to make more concrete the description of a continuous time, finite dimensional system. More precisely, here we assume the existence of a matrix $h(t)$ with p rows and m columns, whose elements are continuous functions defined for each $t \in \mathbf{R}$, such that the response $y(\cdot) = \mathcal{R}(u(\cdot))$ corresponding to an input map $u(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ admits the representation

$$y(t) = \int_{-\infty}^{+\infty} h(t-\tau)u(\tau)\, d\tau. \tag{1.2}$$

Of course, here we are implicitly assuming that the integral is absolutely convergent.[1] A system for which such a matrix exists is called an *impulse response system*, and the matrix $h(t)$ is called an *impulse response matrix*. This terminology can be explained in the following way.

Let $e_1, \ldots, e_m$ be the canonical basis of $\mathbf{R}^m$, and let $u(t) = \delta(t)e_i$ (for some $i \in \{1, \ldots, m\}$), where $\delta(t)$ represents the *Dirac delta* function (see Appendix B). We have:

$$y(t) = \int_{-\infty}^{+\infty} h(t-\tau)u(\tau)\, d\tau = \int_{-\infty}^{+\infty} h(t-\tau)\delta(\tau)e_i\, d\tau = h(t)e_i\ .$$

This shows that the response of the system to the unit impulse in the direction of the vector e_i coincides with the i-th column of the matrix $h(t)$. Notice that for SISO systems (i.e., with $p = m = 1$), $h(t)$ is simply a real function of one real variable. The proof of the following proposition is straightforward.

Proposition 1.2 *For any impulse response system, the associated input-output operator $\mathcal{R}$ is linear.*

In particular, it follows from Proposition 1.2 that for an impulse response system with a vanishing input map, the output is zero for each t.

---

[1]This may require some restrictions on the nature of the system and the set of admissible inputs.

Proposition 1.3 *Each impulse response system is time invariant.*

Proof Let $u(t)$ be an input map, and let $y(t)$ be the corresponding output map. Let moreover $T \in \mathbf{R}$, $v(t) = u(t - T)$, and let $z(t)$ be the output corresponding to the input $v(t)$. We have:

$$z(t) = \int_{-\infty}^{+\infty} h(t-\tau)v(\tau)\,d\tau = \int_{-\infty}^{+\infty} h(t-\tau)u(\tau - T)\,d\tau.$$

Setting $\tau - T = \theta$, we finally obtain:

$$z(t) = \int_{-\infty}^{+\infty} h(t - T - \theta)u(\theta)\,d\theta = y(t - T).$$ ■

Remark 1.4 The two previous propositions provide a complete characterization of the class of impulse response systems. Indeed, it is possible to prove that each time invariant, linear system is an impulse response system (see for instance [32], pp. 152–154).

■

However, in general, an impulse response system is not causal.

Proposition 1.4 *Let an impulse response system be given, and let $h(t)$ be its impulse response matrix. The following properties are equivalent.*

(1) The system is causal.
(2) $h(t) = 0$ for each $t < 0$.
(3) For each input map $u(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ and each $t \in \mathbf{R}$

$$y(t) = \int_{-\infty}^{t} h(t-\tau)u(\tau)\,d\tau.$$

Proof We start to prove that (1) $\Longrightarrow$ (2). For convenience of exposition, we first discuss the case $m = p = 1$. Let $t > 0$ be fixed. Let

$$u_1(\tau) = 0 \quad \text{and} \quad u_2(\tau) = \begin{cases} 0 & \text{if } \tau < t \\ \operatorname{sgn} h(t-\tau) & \text{if } \tau \geq t \end{cases}$$

for $\tau \in \mathbf{R}$, be two input maps (note that u_2 depends on t).

Since $u_1(\tau) = u_2(\tau)$ for $\tau < t$ and the system is causal, we have $y_1(t) = y_2(t)$. On the other hand, it is evident that $y_1(t) = 0$, while

$$y_2(t) = \int_{-\infty}^{+\infty} h(t-\tau)u_2(\tau)\,d\tau = \int_{t}^{+\infty} |h(t-\tau)|\,d\tau.$$

We are so led to conclude that $h(t-\tau)=0$ for each $\tau > t$, that is $h(r)=0$ for $r<0$.

If m or p (or both) are not equal to 1, the proof is technically more complicated, but the basic idea is the same. One starts by fixing a pair of indices i, j, ($i=1,\dots,p$, $j=1,\dots,m$). As before, we chose $u_1(\tau)=0$ for each $\tau\in\mathbf{R}$, which implies that the corresponding output vanishes identically. Next we define $u_2(\tau)$ component-wise, according to the following rule: if $l\neq j$, then $(u_2)_l(\tau)=0$ for each $\tau\in\mathbf{R}$, while

$$(u_2)_j(\tau)=\begin{cases}0 & \text{if } \tau<t\\ \operatorname{sgn} h_{ij}(t-\tau) & \text{if } \tau\geq t.\end{cases}$$

The causality assumption implies that the output corresponding to $u_2(\tau)$ must be identically zero, as well. On the other hand, the i-th component of the output corresponding to $u_2(\tau)$ is

$$\begin{aligned}(y_2)_i(t)&=\int_{-\infty}^{+\infty} h_{ij}(t-\tau)\cdot(u_2)_j(\tau)\,d\tau\\ &=\int_{t}^{+\infty} h_{ij}(t-\tau)\cdot\operatorname{sgn} h_{ij}(t-\tau)\,d\tau=\int_{t}^{+\infty}|h_{ij}(t-\tau)|\,d\tau\end{aligned}$$

which is zero only if $h_{ij}(r)$ vanishes for each $r\in(-\infty,t)$. The conclusion is achieved, by repeating the argument for each choice of i, j.

The proof that (2) $\Longrightarrow$ (3) is straightforward. Thus, it remains to prove that (3) $\Longrightarrow$ (1). Let $t\in\mathbf{R}$ be fixed. If u_1, u_2 are input maps such that $u_1(\tau)=u_2(\tau)$ for each $\tau\leq t$, then the corresponding output maps y_1, y_2 satisfy

$$y_1(t)=\int_{-\infty}^{t} h(t-\tau)u_1(\tau)\,d\tau=\int_{-\infty}^{t} h(t-\tau)u_2(\tau)\,d\tau=y_2(t).$$

Hence, the system is causal. ■

A further simplification in the representation of impulse response systems is possible, if we limit ourselves to input maps $u(\cdot)\in\mathcal{F}(\mathbf{R},\mathbf{R}^m)$ satisfying the following condition: there exists $t_0\in\mathbf{R}$ such that $u(\tau)=0$ for each $\tau<t_0$. Indeed, in such a case, we will have $y(t_0)=0$ and, for each $t>t_0$,

$$y(t)=\int_{t_0}^{t} h(t-\tau)u(\tau)\,d\tau.$$

For impulse response systems, there is also a simple characterization of the external stability property. Recall that the Frobenius norm of a real matrix $M=(m_{ij})_{i=1,\dots,p,j=1,\dots,m}$ satisfies the following inequality:

$$||M||\leq\sum_{i,j}|m_{ij}|. \tag{1.3}$$

Proposition 1.5 *Let an impulse response system be given, and let $h(t)$ be its impulse response matrix. Let us assume in addition that the system is causal. Let $\mathcal{B}(\mathbf{R}, \mathbf{R}^m)$ and $\mathcal{B}(\mathbf{R}, \mathbf{R}^p)$ be respectively, the input maps and the output maps space, both endowed with the uniform convergence norm. The system is BIBO-stable if and only if the integral*

$$\int_0^{+\infty} ||h(r)||\, dr$$

is convergent or, equivalently, if and only if the function $\int_0^t ||h(r)||\, dr$ is bounded for $t \in [0, +\infty)$.

Proof Since the system is causal, for each $t \in \mathbf{R}$ we have:

$$\begin{aligned} ||y(t)|| = ||\int_{-\infty}^t h(t-\tau)u(\tau)\, d\tau|| &\le \int_{-\infty}^t ||h(t-\tau)u(\tau)||\, d\tau \\ &\le \int_{-\infty}^t ||h(t-\tau)|| \cdot ||u(\tau)||\, d\tau \\ &\le \int_{-\infty}^t ||h(t-\tau)||\, d\tau \cdot ||u(\cdot)||_\infty . \end{aligned}$$

By the substitution $t - \tau = r$, we get

$$\int_{-\infty}^t ||h(t-\tau)||\, d\tau = \int_0^{+\infty} ||h(r)||\, dr.$$

Hence, if $\int_0^{+\infty} ||h(r)||\, dr = \ell < \infty$, from the previous computation we obtain

$$||y(t)|| \le \ell ||u(\cdot)||_\infty$$

for each $t \in \mathbf{R}$ and, finally, $||y(\cdot)||_\infty \le \ell ||u(\cdot)||_\infty$. The BIBO-stability condition will be therefore satisfied taking, for each $R > 0$, $S = \ell R$.

As far as the reverse implication is concerned, let us consider first the case $m = p = 1$. Assuming that the system is BIBO-stable, let us fix $t > 0$ and define the input map

$$\tilde{u}(\tau) = \begin{cases} 0 & \text{if } \tau < 0 \\ \operatorname{sgn} h(t-\tau) & \text{if } \tau \in [0, t] \end{cases}$$

(notice that $\tilde{u}(\tau)$ depends on t). Let $\tilde{y}(t)$ be the corresponding output. Invoking again the causality assumption, we have:

$$\tilde{y}(t) = \int_{-\infty}^t h(t-\tau)\tilde{u}(\tau)\, d\tau = \int_0^t |h(t-\tau)|\, d\tau = \int_0^t |h(r)|\, dr. \tag{1.4}$$

Since $|\tilde{u}(\tau)| \leq 1$ for each t and τ, applying the BIBO-stability condition with $R = 1$ we find a constant $S > 0$ such that

$$|\tilde{y}(t)| \leq ||\tilde{y}(\cdot)||_\infty \leq S \tag{1.5}$$

for each $t > 0$. From (1.4) and (1.5) we infer that the integral $\int_0^{+\infty} |h(r)|\, dr$ is convergent.

With some notational complication, the proof extends to the general case where m or p (or both) are greater than 1. Since there are similarities with the proof of Proposition 1.4, we limit ourselves to sketch the reasoning. Let $h_{ij}(t)$ $(i = 1, \ldots, p,\ j = 1, \ldots, m)$ be the elements of the matrix $h(t)$, and let $t > 0$. Let a pair of indices i, j be fixed, and define the input map $\tilde{u}(\tau) = (\tilde{u}_1(\tau), \ldots, \tilde{u}_m(\tau))$ by taking $\tilde{u}_l(\tau) \equiv 0$ if $l \neq j$ and

$$\tilde{u}_j(\tau) = \begin{cases} 0 & \text{if } \tau < 0 \\ \operatorname{sgn} h_{ij}(t-\tau) & \text{if } \tau \in [0, t]. \end{cases}$$

Let finally $\tilde{y}(t) = (\tilde{y}_1(t), \ldots, \tilde{y}_p(t))$ the corresponding output map. Using the causality hypothesis we have

$$\tilde{y}_i(t) = \int_0^t h_{ij}(t-\tau)\tilde{u}_j(\tau)\, d\tau = \int_0^t |h_{ij}(r)|\, dr. \tag{1.6}$$

Since $\|\tilde{u}(\cdot)\|_\infty \leq 1$, the BIBO-stability condition allows us to determine a real number S such that

$$||\tilde{y}(t)|| \leq S \tag{1.7}$$

for each $t > 0$. Clearly, $\tilde{y}_i(t) = |\tilde{y}_i(t)| \leq \|\tilde{y}(t)\|$. As a consequence of (1.6) and (1.7) we conclude that

$$\int_0^t |h_{ij}(r)|\, dr \leq S$$

for each $t > 0$. Finally, by virtue of (1.3), we have

$$\int_0^t \|h(r)\|\, dr \leq \sum_{i,j} \int_0^t |h_{ij}(r)|\, dr \leq pmS$$

for each $t > 0$. The conclusion easily follows. ■

1.3 Initial Conditions

Representing a system as an operator $\mathcal{R} : \mathcal{F}(\mathbf{R}, \mathbf{R}^m) \to \mathcal{F}(\mathbf{R}, \mathbf{R}^p)$ is a very simple and attractive idea, but it is not realistic. In common applications indeed, the inputs are not known on the whole time axis, but only starting from some instant $t_0 \in \mathbf{R}$, assumed as the *initial instant*. Moreover, we are interested to study the behavior of the system in the future, that is for $t \geq t_0$. In these cases, in order to compensate the loss of information about the inputs for $t < t_0$, we need to assume the assignment of the *initial state*, that is the value $x_0 \in \mathbf{R}^n$ assumed by the state variable at the initial instant t_0.

1.3.1 Deterministic Systems

We may image that the *initial condition* i.e., the pair $(t_0, x_0) \in \mathbf{R} \times \mathbf{R}^n$, summarizes the past history of the system. It is also reasonable to presume that the assignment of the initial data, together with the assignment of the input map for $t \geq t_0$, is sufficient to determine uniquely the future evolution of the system. This is actually an assumption, similar to the causality assumption, but more appropriate to the new point of view.

Definition 1.2 We say that a system, or its representing operator, is *deterministic* if for each $t_0 \in \mathbf{R}$,

$$u_1(t) = u_2(t) \ \ \forall t \geq t_0 \ \textit{and} \ x_1(t_0) = x_2(t_0) \implies y_1(t) = y_2(t) \ \forall t \geq t_0$$

where $x_i(t)$, $y_i(t)$ are respectively the state evolution map and the output map corresponding to the input map $u_i(t)$, $i = 1, 2$.

Note that the deterministic hypothesis basically differs from the causality assumption, since it explicitly involves the state of the system. When a system is deterministic, it is convenient to interpret the input-output operator as an "initialized" operator $\mathcal{R}(t_0, x_0)(u(\cdot))$, mapping functions[2]

$$u(\cdot) \in \mathcal{F}([t_0, +\infty), \mathbf{R}^m)$$

to functions

$$y(\cdot) \in \mathcal{F}([t_0, +\infty), \mathbf{R}^p).$$

We write also $y(\cdot) = \mathcal{R}(t_0, x_0)(u(\cdot))$.

Remark 1.5 According to Definition 1.2, the so-called *delayed systems* (systems whose behavior for $t \geq t_0$ depends not only on the state of the system at the initial instant t_0, but also on the values assumed by the state variable on some interval $[t_0 - \theta, t_0]$, $(\theta > 0)$ and, more generally, systems with memory, cannot be considered deterministic. ■

When we want to make use of the notion of initialized operator, the definitions of time invariant system and of linear system need to be appropriately modified.

[2]Alternatively, we may agree that the admissible inputs are restricted to functions $u(\cdot) \in \mathcal{F}(\mathbf{R}, \mathbf{R}^m)$ vanishing for $t < t_0$.

1.3.2 Time Invariant Systems

A system represented by a deterministic initialized operator is *time invariant* if for each $t_0, T \in \mathbf{R}$, each $x_0 \in \mathbf{R}^n$, and each input map $u(\cdot)$, denoting

$$v(t) = u(t - T), \quad y(\cdot) = \mathcal{R}(t_0, x_0)(u(\cdot)), \quad z(\cdot) = \mathcal{R}(t_0 + T, x_0)(v(\cdot))$$

one has

$$z(t) = y(t - T).$$

Proposition 1.6 *Let $\mathcal{R}$ be a time invariant, deterministic initialized operator, and let $y(\cdot) = \mathcal{R}(t_0, x_0)(u(\cdot))$. Then,*

$$y(t) = z(t - t_0)$$

where $z(\cdot) = \mathcal{R}(0, x_0)(v(\cdot))$, and $v(t) = u(t + t_0)$.

In other words, dealing with a time invariant operator, we may assume, without loss of generality, that the initial instant coincides with the origin of the time axis.

1.3.3 Linear Systems

A system represented by means of a deterministic initialized operator $\mathcal{R}$, is *linear* if for each $t_0 \in \mathbf{R}$, $\mathcal{R}$ is linear as a map from $\mathbf{R}^n \times \mathcal{F}([t_0, +\infty), \mathbf{R}^m)$ to $\mathcal{F}([t_0, +\infty), \mathbf{R}^p)$, that is if for each $t_0 \in \mathbf{R}$, and for each choice of the pairs $x_1, x_2 \in \mathbf{R}^n$, $u_1(\cdot), u_2(\cdot) \in \mathcal{F}([t_0, +\infty), \mathbf{R}^m)$, and $a_1, a_2 \in \mathbf{R}$ one has

$$\begin{aligned} &\mathcal{R}\,(t_0, a_1x_1 + a_2x_2)(a_1u_1(\cdot) + a_2u_2(\cdot)) \\ &= a_1\mathcal{R}(t_0, x_1)(u_1(\cdot)) + a_2\mathcal{R}(t_0, x_2)(u_2(\cdot)). \end{aligned}$$

Proposition 1.7 *Let $\mathcal{R}$ be a linear, time invariant, deterministic initialized operator. For each $t_0 \in \mathbf{R}$, each $x_0 \in \mathbf{R}^n$ and each $u(\cdot) \in \mathcal{F}([t_0, +\infty), \mathbf{R}^m)$ one has:*

$$y(\cdot) = \mathcal{R}(t_0, x_0)(0) + \mathcal{R}(t_0, 0)(u(\cdot)).$$

Proof By applying the definition of linear initialized system with $x_1 = x_0$, $x_2 = 0$, $u_1(\cdot) = 0$, $u_2(\cdot) = u(\cdot)$, $a_1 = a_2 = 1$, we immediately have

$$y(\cdot) = \mathcal{R}(t_0, x_0)(u(\cdot)) = \mathcal{R}(t_0, x_0)(0) + \mathcal{R}(t_0, 0)(u(\cdot))$$

as required. ■

1.3.4 External Stability

We now update Definition 1.1, for the case of systems represented by initialized operators.

Definition 1.3 A system represented by a deterministic initialized operator $\mathcal{R}$, is *BIBO-stable* (uniformly with respect to the initial instant) if for each real number $R > 0$ there exists a real number $S > 0$ such that for each $t_0 \in \mathbf{R}$ and each input map $u(\cdot) \in \mathcal{B}([t_0, +\infty), \mathbf{R}^m)$ we have

$$||x_0|| \leq R, \ \ ||u(\cdot)||_\infty \leq R \implies ||y(\cdot)||_\infty \leq S$$

where $y(\cdot) = \mathcal{R}(t_0, x_0)(u(\cdot))$.

1.3.5 Zero-Initialized Systems and Unforced Systems

We may interpret Proposition 1.7 by saying that the response of a linear system corresponding to some initial state x_0 and some input map $u(\cdot)$ can be decomposed as the sum of

- the response corresponding to the initial state x_0 when the the input is set to be zero;
- the response corresponding to the input $u(\cdot)$ when the initial state is set to be zero.

In other words, when analyzing the behavior of a linear system, and more precisely when we are interested in the study of the external stability, the way the response is affected by the initial data and the way it is affected by the inputs can be analyzed separately. We will frequently refer to this principle in this book.

Therefore, in the study of linear systems we may conveniently distinguish two different steps. In the first step we may assume that the input vanishes, while in the second step we may assume that the initial state vanishes. In this way, we will be also able to recover some analogies with the theory of the impulse response systems.

We say that a deterministic system represented by a time invariant initialized operator is *zero-initialized* (or *initialized at zero*) if the initial state x_0 at the instant $t_0 = 0$ is set to be zero. We say that a deterministic system represented by a time invariant initialized operator is *unforced* if the input map is set to be equal to zero for each $t \geq 0$.

Unforced systems may present a non-zero evolution in time: indeed, because of the energy stored in the system at the initial instant, the initial state does not coincide, in general, with a rest point. In these circumstances, we expect that the unforced system evolves in such a way that the initial energy is dissipated, by approaching a rest point asymptotically. If this really happens, we will say informally that the system is *internally stable*. A more precise and formal definition of internal stability will be given later.

In the analysis of the qualitative properties of a system, the study of the behavior when the forcing terms are provisionally suppressed, is an essential preliminary step. As we shall see, the properties of internal stability and external stability are intimately related.

1.4 Differential Systems

In this section we focus on systems which are modeled by means of ordinary differential equations; they will be called *differential systems*. This class of systems is very important, because of the variety and the large amount of applications. Moreover, a well developed and complete theory is available, for these systems. However, its introduction requires some restrictions.

1.4.1 Admissible Inputs

Dealing with differential systems, by *admissible input map* we mean a function $u(\cdot) \in \mathcal{PC}([t_0, +\infty), \mathbf{R}^m)$, for some $t_0 \in \mathbf{R}$. For some applications, it is necessary to limit further the admissible inputs: a typical choice is $u(\cdot) \in \mathcal{PC}([t_0, +\infty), U)$, where U is a given nonempty, bounded subset of $\mathbf{R}^m$. The role of U is to represent possible limitations on the energy available in order to exert the control. Notice that if U is bounded, $\mathcal{PC}([t_0, +\infty), U)$ is not a vector space.

1.4.2 State Equations

Let $f(t, x, u) : \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n$ and $h(t, x) : \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^p$ be given functions. A *differential system* is defined by the equations

$$\dot{x} = \frac{dx}{dt} = f(t, x, u) \tag{1.8}$$

$$y = h(t, x). \tag{1.9}$$

Equation (1.8) is also called the *state equation*, while $h(t, x)$ is called the *observation map*. For each admissible input map $u(t)$, (1.8) becomes a system of ordinary differential equations of the first order in normal form

$$\dot{x} = f(t, x, u(t)) = g(t, x). \tag{1.10}$$

Concerning the functions f and h, it is customary to make the following assumptions:

(**A1**) f is continuous with respect to the pair of variables (x, u); the first partial derivatives of f with respect to all the components x_i of the state vector x exist and are continuous with respect to the pair of variables (x, u); moreover, f is piecewise continuous with respect to t;
(**A2**) h is continuous with respect to x, and piecewise continuous with respect to t;
(**A3**) for each admissible input $u(t)$, there exist continuous, positive real functions $a(t), b(t)$ such that

$$||f(t, x, u(t))|| \leq a(t)||x|| + b(t)$$

for each $(t, x) \in \mathbf{R} \times \mathbf{R}^n$.

Under these assumptions, for each pair of initial values $(t_0, x_0) \in \mathbf{R} \times \mathbf{R}^n$ and for each admissible input map $u(t)$ there exists a unique solution of the *Cauchy problem*

$$\begin{cases} \dot{x} = g(t, x) \\ x(t_0) = x_0 \end{cases} \tag{1.11}$$

defined on the whole interval[3] $[t_0, +\infty)$. When we want to emphasize the dependence of the solution of the problem (1.11) on the initial conditions and on the input map, we will use the notation

$$x = x(t; t_0, x_0, u(\cdot)). \tag{1.12}$$

When the dependence on the initial conditions and on the input map is clear from the context, we may also use the simplified notation $x = x(t)$. The initialized input-output operator associated to the differential system (1.8), (1.9)

$$y(\cdot) = \mathcal{R}(t_0, x_0)(u(\cdot)) \tag{1.13}$$

is given by $y(t) = h(t, x(t; t_0, x_0, u(\cdot)))$ for $t \geq t_0$. By analogy with (1.12), sometimes we may use the notation

$$y = y(t; t_0, x_0, u(\cdot)). \tag{1.14}$$

The following proposition summarizes the previous remarks.

Proposition 1.8 *Under the hypotheses* **(A1)***,* **(A2)***,* **(A3)***, the differential system (1.8), (1.9) defines a deterministic input-output operator on the set of admissible input maps. Moreover, the output map is continuous.*

---

[3]Provided that the input is defined for each $t \in \mathbf{R}$, existence and uniqueness of solutions is actually guaranteed on $(-\infty, +\infty)$.

Next proposition characterizes the differential systems which possess the time invariance property.

Proposition 1.9 *Assume that* **(A1)***,* **(A2)***,* **(A3)** *hold. The input-output operator (1.13) defined by the differential system (1.8), (1.9) is time invariant if the functions* f *e* h *do not depend explicitly on* t*, that is* $f(t, x, u) = f(x, u)$ *and* $h(t, x) = h(x)$.

Proof Let $t_0 \in \mathbf{R}$ and let $u(t) \in \mathcal{PC}([t_0, +\infty), \mathbf{R}^m)$. Assume that an initial state x_0 is given; let $x(t)$ be the corresponding solution of (1.8) and let $y(t) = h(x(t))$. Let finally T be a fixed real number. Setting $v(t) = u(t - T)$ and $\xi(t) = x(t - T)$, we have

$$\frac{d}{dt}\xi(t) = \frac{d}{dt}x(t - T) = f(x(t - T), u(t - T)) = f(\xi(t), v(t)).$$

In other words, $\xi(t)$ coincides with the solution corresponding to the translated input map $v(t)$ and to the initial condition $(t_0 + T, x_0)$. Setting finally $z(t) = h(\xi(t))$, it is clear that $z(t) = y(t - T)$. ■

By virtue of Propositions 1.6 and 1.9, if the functions f and h do not depend explicitly on t we may assume $t_0 = 0$ without loss of generality. In this case, the notation (1.12) and (1.14) can be simplified, by avoiding the explicit indication of the initial instant.

1.4.3 Linear Differential Systems

In the mathematical theory of differential systems, a prominent role is played by systems whose state equations are linear. The importance of linear systems is also supported by their interest in applications.

Definition 1.4 A time invariant differential system is said to be *linear* if there exist real matrices A, B, C of respective dimensions $n \times n$, $n \times m$, $p \times n$, such that $f(x, u) = Ax + Bu$ and $h(x) = Cx$.

In other words, a system is linear in the sense of Definition 1.4 when it can be written in the form

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx. \end{cases} \tag{1.15}$$

Proposition 1.10 *If a system is linear in the sense of Definition 1.4, then the associated input-output initialized operator (1.13) is linear.*

The proof of Proposition 1.10 will be given later. Beginning with Chap. 2, we focus our attention on the study of linear, time invariant differential systems.

Chapter Summary

The first part of this chapter constitutes a short introduction to systems theory. The basic notions of input, output and state variables are presented in abstract terms, as well as the notion of input-output operator. We discuss the main properties involved in the investigation of a system, and illustrate how distinct systems can be combined to give rise to a new system. In this framework, we also introduce the main concern of this book: how to exploit the input channel in order to control the evolution of a system.

The exposition becomes more concrete in the remaining part of the chapter, where we explain how a system can be represented by certain mathematical models: impulse response, state space equations. The role of initial conditions is emphasized, in a deterministic philosophy, in connection with the notion of state variable.

Chapter 2
Unforced Linear Systems

In this chapter we undertake a systematic study of finite dimensional, unforced, linear, time invariant differential systems. They are defined by a system of ordinary differential equations of the form

$$\dot{x} = Ax \, , \qquad x \in \mathbf{R}^n \, . \tag{2.1}$$

According to the mathematical tradition, (2.1) is called a linear homogeneous system of differential equations (with constant coefficients). In extended notation, (2.1) reads

$$\begin{cases} \dot{x}_1 = a_{11}x_1 + \cdots + a_{1n}x_n \\ \cdots \quad \cdots \quad \cdots \\ \dot{x}_n = a_{n1}x_1 + \cdots + a_{nn}x_n \, . \end{cases}$$

For a general system of ordinary differential equations, the notion of solution is recalled in Appendix A. In force of the special form of (2.1) the solutions enjoy some special properties.

2.1 Prerequisites

In this section we recall some important facts, concerning a system of equations of type (2.1) and its solutions.

Fact 1. *For each initial state x_0 there exists a unique solution $x = \varphi(t)$ of system* (2.1) *such that $\varphi(0) = x_0$; moreover, $\varphi(t)$ is defined for each $t \in \mathbf{R}$.*

Fact 2. *If $v \in \mathbf{R}^n$ ($v \neq 0$) is an eigenvector of A corresponding to the eigenvalue $\lambda \in \mathbf{R}$, then $\varphi(t) = e^{\lambda t}v$ represents the solution of* (2.1) *corresponding to the initial state v.*

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_2

Fact 3. *If $\varphi_1(\cdot)$, $\varphi_2(\cdot)$ are solutions of* (2.1) *and $\alpha_1, \alpha_2 \in \mathbf{R}$, then also $\alpha_1\varphi_1(\cdot) + \alpha_2\varphi_2(\cdot)$ is a solution of* (2.1).

Fact 4. *Let $\varphi_1(\cdot), \ldots, \varphi_k(\cdot)$ be k solutions of* (2.1). *The following statements are equivalent*:

- *there exists $\bar{t} \in \mathbf{R}$ such that the vectors $\varphi_1(\bar{t}), \ldots, \varphi_k(\bar{t})$ are linearly independent in $\mathbf{R}^n$*;
- *the functions $\varphi_1(\cdot), \ldots, \varphi_k(\cdot)$ are linearly independent, as elements of the space $\mathcal{C}(-\infty, +\infty, \mathbf{R}^n)$*;
- *for each $t \in \mathbf{R}$, the vectors $\varphi_1(t), \ldots, \varphi_k(t)$ are linearly independent, as elements of the space $\mathbf{R}^n$.*

When one of the above equivalent conditions holds, we simply say that $\varphi_1(\cdot), \ldots, \varphi_k(\cdot)$ are linearly independent.

Fact 5. *The set of all the solutions of the system* (2.1) *forms a subspace $\mathcal{S}$ of $\mathcal{C}(-\infty, +\infty, \mathbf{R}^n)$. The dimension of $\mathcal{S}$ is finite and, more precisely, it is equal to n. The subspace $\mathcal{S}$ is also called the general integral of system* (2.1).

Notice that system (2.1) makes sense even if we allow that x takes value into the n-dimensional complex space $\mathbf{C}^n$, and that the entries of A are complex numbers: apart from some obvious modifications, all the previous facts remain valid.[1] Actually, to this respect we may list some further properties.

Fact 6. *If the elements of A are real, and if $\varphi(\cdot)$ is a solution of* (2.1) *with nonzero imaginary part, then the conjugate function $\overline{\varphi}(\cdot)$ is a solution of* (2.1), *as well.*

Fact 7. *If the elements of A are real, and if $\varphi(\cdot)$ is a solution of* (2.1) *with nonzero imaginary part, then $\varphi(\cdot)$ and $\overline{\varphi}(\cdot)$ are linearly independent; in addition,*

$$\varphi_1(\cdot) = \frac{\varphi(\cdot) + \overline{\varphi}(\cdot)}{2} \quad \text{and} \quad \varphi_2(\cdot) = \frac{\varphi(\cdot) - \overline{\varphi}(\cdot)}{2\mathrm{i}} \tag{2.2}$$

are two real and linearly independent solutions of (2.1).

If A is a matrix with real elements and with a complex eigenvalue $\lambda = \alpha + \mathrm{i}\,\beta$ ($\beta \neq 0$) associated to an eigenvector $v = u + \mathrm{i}\,w$, we dispose of the complex solution $\varphi(t) = e^{\lambda t}v$. Then, using (2.2), we obtain the representation

$$\varphi_1(t) = e^{\alpha t}[(\cos\beta t)u - (\sin\beta t)w]\ , \quad \varphi_2(t) = e^{\alpha t}[(\cos\beta t)w + (\sin\beta t)u]\ .$$

Remark 2.1 The existence of non-real eigenvalues implies therefore the existence of real oscillatory solutions. In particular, if $\alpha = 0$ and $\beta \neq 0$, the eigenvalues are purely imaginary, and we have periodic solutions with minimal period equal to $2\pi/\beta$. ■

[1] The convenience of extending the search for the solutions to the complex field even if the elements of A are real numbers, is suggested by Fact 2: possible eigenvalues of A represented by conjugate complex numbers (with nonzero imaginary part) generates solutions which, otherwise, would be difficult to identify.

We are now able to conclude that the general integral of system (2.1) can be written as a linear combination

$$\varphi(t) = c_1\varphi_1(t) + \cdots + c_n\varphi_n(t) \tag{2.3}$$

where $c_1, \ldots, c_n$ represent arbitrary constants, and $\varphi_1, \ldots, \varphi_n$ represent n arbitrary solutions, provided that they are linearly independent. If A is real, it is not restrictive to assume that $\varphi_1, \ldots, \varphi_n$ are real valued: hence, Eq. (2.3) describes either the space of all the real solutions when the constants $c_1, \ldots, c_n$ are taken in the real field, or the space of all the complex solutions when the constants $c_1, \ldots, c_n$ are taken in the complex field.

A set formed by n linearly independent solutions of the system (2.1) is called a *fundamental set of solutions*. To each fundamental set of solutions φ_1, ..., φ_n, we associate a *fundamental matrix*

$$\Phi(t) = (\varphi_1(t), \ldots, \varphi_n(t))$$

whose columns are formed by the components of the vectors $\varphi_1(t)$, ..., $\varphi_n(t)$. Notice that if $\Phi(t)$ is a fundamental matrix and Q is a constant, nonsingular matrix, then also $\Phi(t)Q$ is a fundamental matrix. From this remark, it follows easily that, for each $t_0 \in \mathbf{R}$, there exists a unique fundamental matrix such that $\Phi(t_0) = I$. This is also called the *principal fundamental matrix relative to* t_0. The principal fundamental matrix relative to $t_0 = 0$ will be simply called *principal fundamental matrix*.

Let us introduce the constant vector $c = (c_1, \ldots, c_n)^{\mathbf{t}}$. If $\Phi(t)$ is any fundamental matrix, we can rewrite (2.3) as

$$\varphi(t) = \Phi(t)c \tag{2.4}$$

The particular solution satisfying the initial conditions $\varphi(t_0) = x_0$ can be recovered by solving the algebraic system

$$\Phi(t_0)c = x_0$$

with respect to the unknown vector c. If $\Phi(t)$ is the principal fundamental matrix relative to t_0, we simply have $c = x_0$.

2.2 The Exponential Matrix

Let $\mathcal{M}(\mathbf{C})$ be the finite dimensional vector space formed by the square matrices $M = (m_{ij})_{i,j=1,\ldots,n}$ of dimensions $n \times n$ with complex entries, endowed with the Frobenius norm. It is possible to prove that the series

$$\sum_{k=0}^{\infty} \frac{M^k}{k!} ,$$

converges for each $M \in \mathcal{M}(\mathbf{C})$ (see for instance [17], p. 83). Its sum is denoted e^M and it is called the *exponential matrix* of M. We list below the main properties of the exponential matrix.

- If the entries of M are real, then the entries of e^M are real.
- $e^0 = I$, where 0 denotes here a matrix whose entries are all equal to zero and I is the identity matrix.
- $e^{N+M} = e^M e^N$, provided that $MN = NM$.
- The eigenvalues of e^M are the complex numbers of the form e^λ, where λ is an eigenvalue of M.
- $e^M M = Me^M$.
- $\det e^M = e^{\operatorname{tr} M}$. As a consequence, $\det e^M \neq 0$ for each M.
- If P is a nonsingular matrix, $e^{P^{-1}MP} = P^{-1}e^M P$.

Let us come back to system (2.1). For each $t \in \mathbf{R}$, all the entries of the matrix e^{tA} are of class C^1. Moreover, the following proposition holds.

Proposition 2.1 *For each $A \in \mathcal{M}(\mathbf{C})$ and each $t \in \mathbf{R}$, we have*

$$\frac{d}{dt} e^{tA} = Ae^{tA} .$$

Thus, the exponential matrix provides a useful formalism, which allows us to represent the solutions of the system (2.1). Indeed, if $x = \varphi(t)$ is the solution of (2.1) such that $\varphi(t_0) = x_0$, then by using the uniqueness of solutions and the properties of the exponential matrix, we get

$$\varphi(t) = e^{(t-t_0)A} x_0 .$$

If $t_0 = 0$, we simply have

$$\varphi(t) = e^{tA} x_0 \tag{2.5}$$

for each $t \in \mathbf{R}$. In other words, computing the exponential matrix is equivalent to compute a fundamental matrix of the system (actually, the principal fundamental matrix).

In the following sections, we will see how to realize an explicit construction of the exponential matrix. The final result will be achieved through several steps. We start by examining some special situations.

2.3 The Diagonal Case

Let

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} = \mathrm{diag}\,(\lambda_1, \dots, \lambda_n)$$

where $\lambda_1, \dots, \lambda_n$ are not necessarily distinct numbers (real or complex).

Remark 2.2 For such a matrix, λ is an eigenvalue if and only if $\lambda = \lambda_i$ for some $i = 1, \dots, n$, and the algebraic multiplicity of λ indicates how many λ_i's are equal to λ. The eigenvectors corresponding to $\lambda_1, \dots, \lambda_n$ can be taken respectively coincident with the vectors of the canonical basis

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, v_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{2.6}$$

■

A fundamental set of solutions of (2.1) can be therefore written in the form

$$\varphi_1(t) = e^{\lambda_1 t} v_1, \ \dots, \varphi_n(t) = e^{\lambda_n t} v_n \, .$$

A system (2.1) defined by a diagonal matrix A is called *decoupled*, since the evolution of each component x_i of x depends on x_i, but not on x_j with $j \neq i$. A system of this type can be trivially solved by integrating separately the single equations. The fundamental set of solutions obtained by this method obviously coincides by the previous one. The same fundamental set of solutions can be obtained also by computing the exponential matrix. Indeed, it is easy to check that for each positive integer k,

$$A^k = \mathrm{diag}\,(\lambda_1^k, \dots, \lambda_n^k) \, ,$$

hence

$$e^{tA} = \mathrm{diag}\,(e^{\lambda_1 t}, \dots, e^{\lambda_n t}) \, .$$

2.4 The Nilpotent Case

If A is nilpotent, there exists a positive integer q such that $A^k = 0$ for each $k \geq q$. Thus, the power series which defines the exponential matrix reduces to a polynomial and can be computed in elementary way. A typical nilpotent matrix (for which $q = n$) is

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \\ . \end{pmatrix} \tag{2.7}$$

The direct computation of the exponential matrix shows that if A has the form (2.7), then

$$e^{tA} = \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \dots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \dots & \frac{t^{n-2}}{(n-2)!} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & t \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} .$$

Alternatively, we can write the corresponding system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dots \\ \dot{x}_n = 0 \end{cases}$$

and solve it by cascaded integration (from down to top). The two approaches obviously lead to the same result. A fundamental set of solutions can be written in the form

$$\varphi_1(t) = v_1 \ , \quad \varphi_2(t) = tv_1 + v_2 \ , \ \dots \ , \quad \varphi_n(t) = \frac{t^{n-1}}{(n-1)!} v_1 + \dots + tv_{n-1} + v_n \tag{2.8}$$

where the vectors $v_1, \dots, v_n$ are as in (2.6) the vector of the canonical basis.

Remark 2.3 Notice that zero is the unique eigenvalue of the matrix (2.7); the corresponding proper subspace is one dimensional. Moreover, $Av_1 = 0$ (which means that v_1 is an eigenvector of A), $Av_2 = v_1$, $Av_3 = v_2$ and so on. ■

The general integral of the system defined by the matrix (2.7) can be written as

$$\varphi(t) = c_1\varphi_1(t) + \dots + c_n\varphi_n(t) = d_1 + td_2 + \dots + \frac{t^{n-1}}{(n-1)!} d_n$$

where

$$d_1 = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{pmatrix}, \quad d_2 = \begin{pmatrix} c_2 \\ c_3 \\ \vdots \\ c_n \\ 0 \end{pmatrix}, \dots, \quad d_n = \begin{pmatrix} c_n \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

Notice that $Ad_1 = d_2, Ad_2 = d_3, \dots, Ad_n = 0$. Notice also that d_1 can be arbitrarily chosen, and that d_n is an eigenvector of A, regardless to the choice of d_1.

Remark 2.4 Combining the methods used for the cases of diagonal and nilpotent matrices, we are able to compute the exponential matrix for each matrix A of the form $\lambda I + T$ where λ is any real number, I is the identity matrix of dimensions $n \times n$, and T is nilpotent. In particular, if T has the form (2.7), then

$$e^{t(\lambda I+T)} = e^{\lambda t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \dots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \dots & \frac{t^{n-2}}{(n-2)!} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & t \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \tag{2.9}$$

■

2.5 The Block Diagonal Case

If M is block diagonal, that is

$$M = \begin{pmatrix} M_1 & 0 & \dots & 0 \\ 0 & M_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & M_k \end{pmatrix} = \operatorname{diag}(M_1, \dots, M_k),$$

then also its exponential matrix is block diagonal

$$e^M = \operatorname{diag}(e^{M_1}, \dots, e^{M_k}).$$

The exponential matrix of M is easily obtained, provided that we are able to construct the exponential matrix of every block M_i.

2.6 Linear Equivalence

To address the problem of computing the exponential matrix in the general case, we need to introduce the concept of linear equivalence.

Let us image system (2.1) as the mathematical model of a process evolving in a real vector space V of dimension n, where a basis has been fixed. The state of the system is represented, in this basis, by the n-tuple $x = (x_1, \dots, x_n)^{\mathbf{t}}$.

Assume that a new basis of V is given, and let $y = (y_1, \dots, y_n)^{\mathbf{t}}$ be the components of the state in this new basis. As well known, there exists a nonsingular matrix P such that for each element of V,

$$x = Py.$$

We want to see how (2.1) changes, when the state is represented in the new basis. We have

$$\dot{y} = P^{-1}\dot{x} = P^{-1}APy = By\ . \tag{2.10}$$

We therefore obtain again a linear system, defined by a matrix B which is similar to the given matrix A. Vice versa, two systems of the type (2.1) defined by similar matrices can be always thought of as two representations of the same system in two different systems of coordinates.

Definition 2.1 Two systems

$$\dot{x} = Ax \qquad \text{and} \qquad \dot{y} = By\ , \qquad x \in \mathbf{R}^n,\ y \in \mathbf{R}^n$$

are said to be *linearly equivalent* if A and B are similar, that is if $B = P^{-1}AP$ for some nonsingular matrix P.

The previous definition is actually an equivalence relation. It is clear that each solution $x = \varphi(t)$ of the first system is of the form $\varphi(t) = P\psi(t)$ where $y = \psi(t)$ is a solution of the second one and vice-versa. On the other hand, it is easy to see that

$$e^{tB} = P^{-1}e^{tA}P \qquad (\text{or, equivalently, } e^{tA} = Pe^{tB}P^{-1})\ . \tag{2.11}$$

Hence, as far as we are interested in solution representation, we can work with any system linearly equivalent to the given one, and finally we can use (2.11) in order to come back to the original coordinates.

The notion of linear equivalence, as well as the notion of similar matrices, can be immediately generalized to the case where $x \in \mathbf{C}^n$. Of course, if A and B are similar matrices, A is real and B contains complex elements, then the matrix P determining the change of basis must contain complex elements, as well.

2.7 The Diagonalizable Case

It is well known that a matrix A of dimensions $n \times n$ is *diagonalizable* (that is, similar to a diagonal matrix) if and only if there exist n linearly independent vectors $v_1, \dots, v_n$ such that each v_i, $i = 1, \dots, n$, is an eigenvector of A. In such a case, we say that the vectors $v_1, \dots, v_n$ constitute a *proper basis* of A. In particular, A is diagonalizable if it admits n distinct eigenvalues.

Denoting by P the matrix whose columns are $v_1, \dots, v_n$ (in this order), we have

$$P^{-1}AP = \text{diag}\,(\lambda_1, \dots, \lambda_n) = D$$

where λ_1 is the eigenvalue of A corresponding to v_1, λ_2 is the eigenvalue of A corresponding to v_2 and so on (it is not required that the numbers $\lambda_1, \dots, \lambda_n$ are distinct).

To compute e^{tA} we can proceed in the following way: first we diagonalize A by means of the change of coordinates determined by P, then we compute e^{tD}, and finally we come back to the original coordinates, making use of (2.11).

Remark 2.5 If A is real but it admits complex eigenvalues, then P and D will have complex elements, as well. However, by construction, the elements of e^{tA} must be real.

Notice that $\Phi(t) = Pe^{tD}$ is a fundamental matrix; its computation do not require to know the inverse of P. However, in general the elements of Pe^{tD} are not real, not even if A is real.

In conclusion, to determine explicitly the elements of the matrix e^{tA} and hence the general integral of (2.1) in the diagonalizable case, it is sufficient to know the eigenvalues of A and the corresponding eigenvectors.

Example 2.1 Let us consider the system

$$\begin{cases} \dot{x}_1 = -x_2 \\ \dot{x}_2 = x_1 \end{cases}$$

defined by the matrix

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

The eigenvalues of A are $+\mathrm{i}$, with eigenvector $\begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix}$, and $-\mathrm{i}$, with eigenvector $\begin{pmatrix} -\mathrm{i} \\ 1 \end{pmatrix}$. It is easy to identify two (complex conjugate) linearly independent solutions

$$\varphi_1(t) = e^{\mathrm{i}\,t} \begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} + \mathrm{i} \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

and

$$\varphi_2(t) = e^{-\mathrm{i}\,t}\begin{pmatrix}\mathrm{i}\\1\end{pmatrix} = \begin{pmatrix}-\sin t\\\cos t\end{pmatrix} - \mathrm{i}\begin{pmatrix}\cos t\\\sin t\end{pmatrix} .$$

Taking their real and imaginary parts we obtain two linearly independent real solutions

$$\psi_1(t) = \begin{pmatrix}-\sin t\\\cos t\end{pmatrix} \quad \text{and} \quad \psi_2(t) = \begin{pmatrix}\cos t\\\sin t\end{pmatrix} .$$

Alternatively, we can apply the diagonalization procedure. To this end, we need to compute the inverse matrix of

$$P = \begin{pmatrix}-\mathrm{i} & \mathrm{i}\\1 & 1\end{pmatrix}$$

given by

$$P^{-1} = -\frac{1}{2\mathrm{i}}\begin{pmatrix}1 & -\mathrm{i}\\-1 & -\mathrm{i}\end{pmatrix} .$$

We easily get

$$D = P^{-1}AP = \begin{pmatrix}\mathrm{i} & 0\\0 & -\mathrm{i}\end{pmatrix} ,$$

and

$$e^{tD} = \begin{pmatrix}e^{\mathrm{i}\,t} & 0\\0 & e^{-\mathrm{i}\,t}\end{pmatrix} .$$

Finally,

$$e^{tA} = Pe^{tD}P^{-1} = \begin{pmatrix}\cos t & -\sin t\\\sin t & \cos t\end{pmatrix} .$$

In this case, the exponential matrix could be also obtained directly, by applying the definition; indeed, it is not difficult to see that

$$A^4 = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} .$$

2.8 Jordan Form

In this section, for any $n \times n$ matrix A, we denote by $\lambda_1, \dots, \lambda_k$ $(1 \le k \le n)$ its distinct eigenvalues. For each eigenvalue λ_i of A, by μ_i and ν_i we denote respectively the algebraic and geometric multiplicity of λ_i $(1 \le \nu_i \le \mu_i)$. Moreover, we will write $\lambda_i = \alpha_i + \mathrm{i}\,\beta_i$.

If A possesses eigenvalues with algebraic multiplicity greater than one and with geometric multiplicity less than the algebraic multiplicity, then A is not diagonalizable. In other words, the number of linearly independent eigenvectors is not sufficient to form a basis of the space. To overcome the difficulty, we resort to generalized eigenvectors. The following theorem holds (see for instance [4]).

Theorem 2.1 *Each matrix A of dimension $n \times n$ is similar to a block-diagonal matrix of the form*

$$J = \begin{pmatrix} C_{1,1} & 0 & \dots & 0 \\ 0 & C_{1,2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & C_{k,\nu_k} \end{pmatrix}$$

where the blocks $C_{i,j}$ are square matrices of the form

$$C_{i,j} = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda_i \end{pmatrix} .$$

Only one eigenvalue appears in each block, but a single eigenvalue can appear in more than one block. More precisely, for each eigenvalue λ_i there are exactly ν_i blocks, and each block is associated to one proper eigenvector. The dimension of a block $C_{i,j}$ equals the length of the chain of generalized eigenvectors originating from the j-th eigenvector associated to λ_i. The eigenvalue λ_i appears exactly μ_i times on the principal diagonal of J.

The matrix J is called a *Jordan form* of A. From our point of view, it is important to remark that each block J has the form $\lambda_i I + T$, where I is the identity matrix (of appropriate dimension), and T is the nilpotent matrix of type (2.7). Taking into account the conclusions of Sect. 2.5, the strategy illustrated for the case of a diagonalizable matrix can be therefore extended to the present situation: we transform the given system (2.1) to the system

$$\dot{y} = Jy \tag{2.12}$$

by means of a suitable change of coordinates, then we find the solutions of (2.12) directly (or, alternatively, we compute e^{tJ}, and we come back to the original coordinates). It remains only the problem of identifying the matrix P which determines the similarity between A and J. To this purpose, as already sketched, we need to determine for each eigenvalue λ_i, a number of linearly independent eigenvectors and generalized eigenvectors equal to μ_i. These vectors must be indexed in accordance to the order of the indices of the eigenvalues and, for each eigenvector, in accordance with the order of generation of the generalized eigenvectors of a same chain.

$$\overbrace{\underbrace{\underbrace{v_{1,1,0}}_{eigenvector}\quad v_{1,1,1}\ \cdots}_{first\ chain}\ \underbrace{\underbrace{v_{1,2,0}}_{eigenvector}\quad v_{1,2,1}\cdots}_{second\ chain}\ \underbrace{\underbrace{v_{1,3,0}}_{eigenvector}\quad v_{1,3,1}\cdots}_{third\ chain}\ \cdots}^{\lambda_1}$$

$$\overbrace{\underbrace{\underbrace{v_{2,1,0}}_{eigenvector}\quad v_{2,1,1}\ \cdots}_{first\ chain}\ \cdots}^{\lambda_2}\quad\cdots$$

The set of all these vectors constitutes a basis of the space, called again a *proper basis*. The columns of the matrix P are formed by the vectors of a proper basis in the aforementioned order, that is

$$\begin{aligned} P &= (v_{1,1,0} \mid v_{1,1,1} \mid \ldots \mid v_{1,2,0} \mid v_{1,2,1} \mid \ldots \mid v_{1,3,0} \mid v_{1,3,1} \mid \ldots \\ &\quad \ldots \mid v_{2,1,0} \mid v_{2,1,1} \mid \ldots) \,. \end{aligned}$$

Another proper basis and another corresponding Jordan form can be obtained by permutations of the order of the eigenvalues or, for each eigenvalue, permutations of the order of the corresponding eigenvectors (but leaving unchanged the order of generation of the generalized eigenvectors). In this sense, the Jordan form is not unique.

After that a proper basis has been constructed and provided that the order of the various indices is correctly settled out, we have all the information we need in order to explicitly write the Jordan form. In fact, we do not need to perform the change of coordinates. However, the computation of P and P^{-1} is inevitable in order to recover e^{tA} in the original coordinates. The computation of P^{-1} can be avoided, if we may limit ourselves to write the fundamental (in general, complex) matrix Pe^{tJ}.

Keeping in mind (2.9), and the procedure illustrated in Sect. 2.1 (Fact 7), we can resume the conclusions achieved so far in the following proposition.

Proposition 2.2 *The generic element $\varphi_{r,s}(t)$ of the matrix e^{tA} $(r, s = 1, \ldots, n)$ reads as*

$$\varphi_{r,s}(t) = \sum_{i=1}^{k} (Z_{r,s})_i(t) e^{\lambda_i t}$$

where each term $(Z_{r,s})_i(t)$ is a polynomial (in general, with complex coefficients) whose degree is (strictly) less than the algebraic multiplicity of λ_i, and λ_i is an eigenvalue of A $(i = 1, \ldots, k)$.

If A is real, the generic element $\varphi_{r,s}(t)$ of the matrix e^{tA} can be put in the form

$$\varphi_{r,s}(t) = \sum_{i=1}^{k} e^{\alpha_i t} \left[(p_{r,s})_i(t) \cos \beta_i t + (q_{r,s})_i(t) \sin \beta_i t \right] \tag{2.13}$$

where $(p_{r,s})_i$ *and* $(q_{r,s})_i$ *are polynomials with real coefficients whose degree is (strictly) less than the algebraic multiplicity of* λ_i *(of course, the previous formula includes also the contributions of the real eigenvalues, for which* $\beta_i = 0$*).*

2.9 Asymptotic Estimation of the Solutions

To our purposes, one of the main applications of the conclusions of the previous section is the estimation of the asymptotic behavior of the solutions of (2.1) for $t \to +\infty$.

Lemma 2.1 *For each* $\varepsilon > 0$ *and each integer* $m \in \mathbf{N}$ *there exists a constant* $k > 0$ *such that* $t^m < ke^{\varepsilon t}$*, for each* $t \geq 0$.

Proof The proof can be carried on by mathematical induction. If $m = 1$ we can take $k = \frac{1}{\varepsilon}$. Indeed, setting

$$f(t) = \frac{e^{\varepsilon t}}{\varepsilon} - t$$

we have $f(0) = \frac{1}{\varepsilon}$ and $f'(t) = e^{\varepsilon t} - 1 > 0$ for $t > 0$. Let us assume that the result holds for $m - 1$, with $k = \bar{k}$. The function

$$f(t) = ke^{\varepsilon t} - t^m$$

is such that

$$f(0) = k \quad \text{and} \quad f'(t) = k\varepsilon e^{\varepsilon t} - mt^{m-1} = m\left(\frac{k\varepsilon}{m}e^{\varepsilon t} - t^{m-1}\right) > 0$$

for $t > 0$, provided that we choose $k = \frac{m\bar{k}}{\varepsilon}$. ■

Let α_0 be the maximum of the real parts α_i of the eigenvalues λ_i of the matrix A $(i = 1, \ldots, k)$ and let α be any real number greater than α_0:

$$\alpha > \alpha_0 \geq \alpha_i \quad \text{for each} \quad (i = 1, \ldots, k) .$$

Since $|\sin \beta_i t| \leq 1$ and $|\cos \beta_i t| \leq 1$ for each $i = 1, \ldots, k$, starting from (2.13) and using repeatedly the triangular inequality we get, for $t \geq 0$,

$$|\varphi_{r,s}(t)| \leq \sum_{i=1}^{k} e^{\alpha_i t}\left(|(p_{r,s})_i(t)| + |(q_{r,s})_i(t)|\right) \leq \sum_{i=1}^{k}(Q_{r,s})_i(t)e^{\alpha_i t}$$

where $(Q_{r,s})_i$ is a polynomial whose coefficients are nonnegative real numbers, which majorize the absolute values of the corresponding coefficients of the polynomials

$(p_{r,s})_i(t)$ and $(q_{r,s})_i(t)$. Even if not essential for the subsequent developments, we note that the degree of $(Q_{r,s})_i$ is less than the algebraic multiplicity of λ_i.

Let $0 < \varepsilon < \alpha - \alpha_0$. By Lemma 2.1, there are constants $k_{r,s}$ such that $|\varphi_{r,s}(t)| \le k_{r,s}e^{(\alpha_0+\varepsilon)t} \le k_{r,s}e^{\alpha t}$ for each $t \ge 0$. Hence $\|e^{tA}\| = \sqrt{\sum_{r,s}\varphi^2_{r,s}(t)} \le \sqrt{\sum_{r,s}k^2_{r,s}}e^{\alpha t}$ and, finally,

$$\|e^{tA}\| \le k_0 e^{\alpha t} \qquad \forall t \ge 0$$

where k_0 is a new constant.

Note that if all the eigenvalues λ_i whose real parts are exactly equal to α_0 (i.e., $\alpha_i = \operatorname{Re}\lambda_i = \alpha_0$) have algebraic multiplicity coincident with the geometric multiplicity, then the previous inequality holds even when α is replaced by α_0. Indeed, in this case the corresponding polynomials $(p_{r,s})_i(t)$ and $(q_{r,s})_i(t)$ reduce to constants. Hence, the term $(Q_{r,s})_i(t)e^{\alpha_i t}$ can be directly majorized by $e^{\alpha_0 t}$, apart from a multiplicative constant, without need of using Lemma 2.1. Concerning the eigenvalues λ_i for which $\alpha_i = \operatorname{Re}\lambda_i < \alpha_0$, we may apply Lemma 2.1 with $\varepsilon = \alpha_0 - \alpha_i$. The corresponding terms $(Q_{r,s})_i(t)e^{\alpha_i t}$ can therefore be majorized, apart form some multiplicative constants, by $e^{\alpha_i t}e^{\varepsilon t} = e^{\alpha_0 t}$. Summing up, we can state the following proposition.

Proposition 2.3 *Let A be a real matrix. For each $\alpha > \alpha_0$, there exists $k_0 > 0$ such that*

$$\|e^{tA}\| \le k_0 e^{\alpha t} \qquad \forall t \ge 0\,. \tag{2.14}$$

If all the eigenvalues of A with real part equal to α_0 have the algebraic multiplicity coincident with the geometric multiplicity, then in (2.14) we can take $\alpha = \alpha_0$.

From (2.14) it follows

$$\|e^{tA}c\| \le k_0\|c\|e^{\alpha t}\,, \quad t \ge 0 \tag{2.15}$$

for each real constant vector c.

2.10 The Scalar Equation of Order *n*

The scalar differential equation (with constant coefficients, $n > 1$)

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}y' + a_n y = 0 \tag{2.16}$$

can be thought of as a particular case of (2.1). Indeed, setting

$$y = x_1,\ y' = x_2,\ \ldots,\ y^{(n-1)} = x_n$$

and using (2.16) we have

$$\begin{aligned} x_1' &= y' = x_2 \\ x_2' &= y'' = x_3 \\ &\ldots\ldots\ldots\ldots \\ x_n' &= y^{(n)} = -a_1 x_n - \ldots - a_n x_1 \end{aligned}$$

that is, with vector notation,

$$\dot{x} = Cx \tag{2.17}$$

where we set $x = (x_1, \ldots, x_n)^{\mathbf{t}}$, and

$$C = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_n & -a_{n-1} & \ldots & \ldots & -a_1 \end{pmatrix} . \tag{2.18}$$

By *solution* of (2.16) we obviously mean a n-times continuously differentiable function $y(t) : \mathbf{R} \to \mathbf{R}$ for which (2.16) is identically satisfied for each $t \in \mathbf{R}$. The problem of determining the general integral (that is, the set of all the solutions) of (2.16) is clearly equivalent to the problem of finding n linearly independent solutions of (2.17).

A matrix exhibiting the structure (2.18) is called a *companion matrix*. More precisely, we say that (2.18) is the companion matrix associated to the Eq. (2.16).

The special structure of C displayed by (2.18) allows us to identify immediately an important algebraic object, invariant under similarity. Indeed, using mathematical induction, it is easy to check that the characteristic polynomial of C is $p_C(\lambda) = (-1)^n \left[\lambda^n + a_1\lambda^{n-1} + \cdots + a_n\right]$. We are especially interested in the eigenvalues of C, which are the roots of $p_C(\lambda)$; hence, the coefficient $(-1)^n$ can be neglected. In fact, it is customary (even is slightly confusing) to call $(-1)^n p_C(\lambda)$ the *characteristic polynomial of the differential equation* (2.16). From now on, we adopt the notation

$$p_{ch}(\lambda) = (-1)^n p_C(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n \,. \tag{2.19}$$

Note that $p_{ch}(\lambda)$ is monic for each n, and that it can be immediately written, without need of transforming (2.16) in the equivalent system (2.17), by replacing formally y by λ and reinterpreting the orders of the derivatives as powers. It is also customary to say that

$$p_{ch}(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n = 0 \tag{2.20}$$

is the *characteristic equation* of the differential equation (2.16), and that its solutions are the *characteristic roots* of (2.16).

Now we change the point of view. Let A be an arbitrary $n \times n$ real matrix, and let

$$p_A(\lambda) = (-1)^n \left[\lambda^n + a_1\lambda^{n-1} + \cdots + a_n\right] . \tag{2.21}$$

its characteristic polynomial. Write a matrix C_A of the form (2.18), reporting in the last row the coefficients $a_1, \ldots, a_n$ taken from (2.21). In this way, A and C_A will have the same characteristic polynomial and hence the same eigenvalues (with the same algebraic multiplicity). The matrix C_A is called the *the companion matrix associated to A*. Unfortunately, in general, A and C_A need not to be similar. For instance, the characteristic polynomial of the identity matrix is $p_I(\lambda) = \sum_{i=0}^{n}(-1)^i \binom{n}{i}\lambda^i$. We may write the associated companion matrix C_I; however the identity matrix I is not similar to C_I, the class of equivalence of I under the similarity relation being just the singleton $\{I\}$. It follows that not all the systems of linear differential equations in $\mathbf{R}^n$ can be reduced by linear transformations to a scalar equation of order n.

The following theorem provides conditions ensuring that a given matrix A is similar to its associated matrix C_A in companion form. This theorem has its own interest from an algebraic point of view, but it is also very important for our future developments.

Theorem 2.2 *Let A be a square matrix of dimensions $n \times n$. The following properties are equivalent.*

(i) *A is similar to its associated matrix in companion form.*
(ii) rank $(A - \lambda I) = n - 1$ *for each eigenvalue λ of A.*
(iii) *The geometric multiplicity of each eigenvalue of A is equal to* 1.
(iv) *The characteristic polynomial of A coincides with its minimal polynomial.*
(v) *There exists a vector $v \neq 0$ such that the n vectors*

$$v, Av, A^2v, \ldots, A^{n-1}v$$

are linearly independent.

The complete proof of Theorem 2.2 can be found for instance in [22]. To our future purposes, the equivalence between *(i)* and *(v)* is especially important.[2] A vector v enjoying the property stated in *(v)* is said to be *cyclic* for A.

Thus, if the $n \times n$ matrix A satisfies one of the assumptions of Theorem 2.2, solving the linear system defined by A is actually equivalent to solve a differential equation of order n of the form (2.16). Property *(iii)* of Theorem 2.2 implies in particular that for each eigenvalue λ of A there is a unique eigenvector v and hence a unique chain of possible generalized eigenvectors engendered by v.

[2]For reader's convenience, a proof of this equivalence will be given in the next section.

Of course, statements *(i),...,(v)* are fulfilled by any matrix C assigned in companion form. It follows that the general integral of (2.16) can be obtained as a linear combination of the n functions

$$
\begin{array}{l}
e^{\lambda_1 t},\ te^{\lambda_1 t}, \dots, t^{\mu_i - 1} e^{\lambda_1 t} \\
\dots\dots\dots\dots\dots\dots \\
e^{\lambda_k t},\ te^{\lambda_k t}, \dots, t^{\mu_k - 1} e^{\lambda_k t}
\end{array}
$$

where $\lambda_1, \dots, \lambda_k$ are the distinct roots of the characteristic Eq. (2.20) and $\mu_1, \dots, \mu_k$ the respective algebraic multiplicities.

Example 2.2 Let us consider in detail the case of a linear equation of order 2

$$y'' + ay' + by = 0 \,. \tag{2.22}$$

To write the general integral $y(t)$, first we need to discuss the characteristic equation

$$\lambda^2 + a\lambda + b = 0 \,. \tag{2.23}$$

If (2.23) has two distinct real solutions λ_1, λ_2, then we have

$$y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} \,. \tag{2.24}$$

If (2.23) has a unique real solution $\lambda_1 = \lambda_2 = \lambda$ of multiplicity 2, then we have

$$y(t) = (c_1 + tc_2) e^{\lambda t} \,. \tag{2.25}$$

Finally, if (2.23) has complex (not real) conjugate solutions[3] $\alpha \pm \mathrm{i}\,\beta$, then we have

$$y(t) = (c_1 \cos \beta t + c_2 \sin \beta t) e^{\alpha t} \,. \tag{2.26}$$

The behavior of the solutions for $t \geq 0$ depends on the signs of λ_1 and λ_2 in the case (2.24) and, respectively, an the signs of λ and α in the cases (2.25) (2.26).

For instance, if $\lambda_1, \lambda_2 < 0$ [respectively, $\lambda < 0, \alpha < 0$] the energy initially stored in the system (measured by the initial conditions) is dissipated:

---

[3]The expression (2.26) results from the application of formulæ (2.2). Alternatively, (2.26) can be obtained starting from the complex version of (2.24)

$$k_1 e^{\lambda t} + k_2 e^{\overline{\lambda} t}$$

using the fact that $e^{\alpha \pm \mathrm{i}\,\beta} = e^{\alpha t}(\cos \beta t \pm \mathrm{i}\, \sin \beta t)$ and setting

$$c_1 = k_1 + k_2 \qquad c_2 = -\mathrm{i}\,(k_1 - k_2) \,.$$

In particular, c_1 and c_2 turn out to be real if we take $k_2 = \overline{k}_1$.

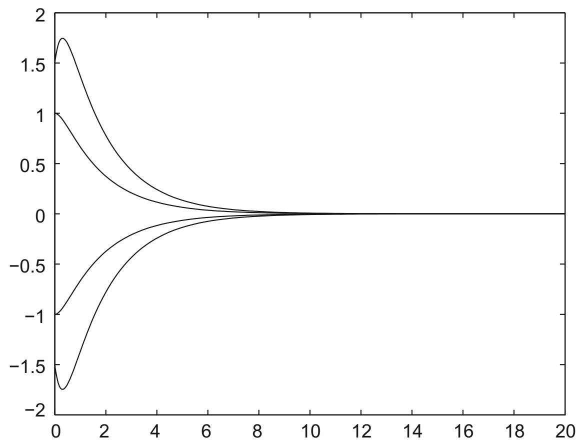


Fig. 2.1 Dissipation in the case of real characteristic roots

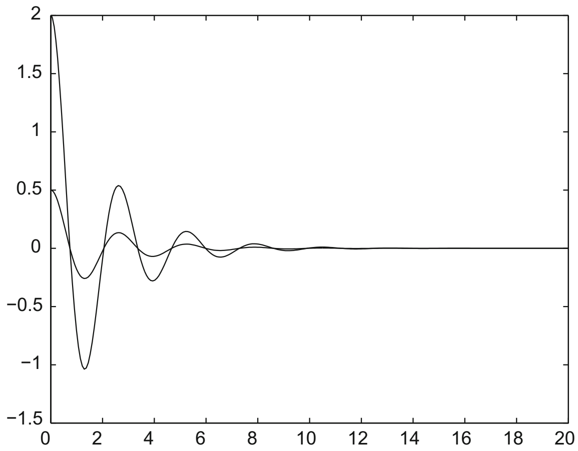


Fig. 2.2 Dissipation in the case of complex characteristic roots

- monotonically in the cases (2.24) and (2.25), after possible initial picks, whose occurrence depends on the choice of c_1 and c_2 (Fig. 2.1);
- with oscillatory decay in the case (2.26) (Fig. 2.2).

The case $a = 0$ and $b > 0$ is a particular instance of (2.26) with $\alpha = 0$ and $\beta = \sqrt{b}$: the solutions are periodic with minimal period $2\pi/\beta$. The general integral takes the form

$$y(t) = c_1 \cos \beta t + c_2 \sin \beta t = \rho \cos(\beta t + \theta) \tag{2.27}$$

where ρ and $\theta \in [0, 2\pi)$ are identified by the relations $c_1 = \rho \cos \theta, c_2 = \rho \sin \theta$. The numbers $\rho = \sqrt{c_1^2 + c_2^2}$ and θ are called *amplitude* and *phase* of the periodic function (2.27). The inverse of the minimal period is called the *frequency*. Note that in (2.27), the frequency depends on b while the amplitude depends on the initial conditions and remains constant. In other words, in this case we have conservation of the energy. The reader can easily check that these conclusions agree with those of Example 2.1.

Remark 2.6 Let us denote by D the derivative operator. Formally, (2.16) can be rewritten as

$$L(D)y = (D^n + a_1 D^{n-1} + \cdots + a_n)y = 0$$

where $(-1)^n L(D) = p_{ch}(D)$. Notice that $L(D)$ acts as a linear operator.

2.11 The Companion Matrix

In this section we show that a matrix A is similar to the associated matrix in companion form if and only if there exists a cyclic vector for A, that is a vector $v \neq 0$ such that $v, Av, \ldots, A^{n-1}v$ form a basis of $\mathbf{R}^n$ (this proves the equivalence of statements *(i)* and *(v)* of Theorem 2.2).

Lemma 2.2 *Let C be a matrix in companion form, and let $\lambda_1, \ldots, \lambda_n$ its eigenvalues (not necessarily distinct). Then, C is similar to a matrix of the form*

$$M = \begin{pmatrix} \lambda_1 & 1 & 0 \ldots & 0 & 0 \\ 0 & \lambda_2 & 1 \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \ldots & \lambda_{n-1} & 1 \\ 0 & 0 & 0 \ldots & \ldots & \lambda_n \end{pmatrix}. \tag{2.28}$$

Proof Let us start with Eq. (2.16). Let us show that by means of suitable linear substitutions, (2.16) can be transformed in a system of first order linear equations defined by the matrix (2.28). Let us set

$$\xi_1 = y, \quad \xi_2 = p_{2,1}y + y', \ldots, \quad \xi_n = p_{n,1}y + \cdots + p_{n,n-1}y^{(n-2)} + y^{(n-1)}$$

where the coefficients $p_{i,j}$ are recovered by the relations

$$p_{2,1}y + y' = (-\lambda_1 + D)y$$
$$p_{3,1}y + p_{3,2}y' + y'' = (-\lambda_1 + D)(-\lambda_2 + D)y$$
$$\ldots\ldots\ldots\ldots$$
$$p_{n,1}y + \cdots + p_{n,n-1}y^{(n-2)} + y^{(n-1)} = (-\lambda_1 + D) \cdot \cdots \cdot (-\lambda_{n-1} + D)y$$

and D is the derivation operator. We have

$$
\begin{aligned}
\xi_1' &= y' = y' + p_{2,1}y - p_{2,1}y = (-\lambda_1 + D)y + \lambda_1 y = \lambda_1\xi_1 + \xi_2 \\
\xi_2' &= (p_{2,1}y + y')' = \\
&= D(-\lambda_1 + D)y + \lambda_2(-\lambda_1 + D)y - \lambda_2(-\lambda_1 + D)y = \\
&= (-\lambda_2 + D)(-\lambda_1 + D)y + \lambda_2\xi_2 = \lambda_2\xi_2 + \xi_3 \\
&\qquad \ldots\ldots\ldots\ldots\ldots \\
\xi_n' &= D(p_{n,1}y + \cdots + p_{n,n-1}y^{(n-2)} + y^{(n-1)}) = \\
&= D(-\lambda_{n-1} + D)\cdot\cdots\cdot(-\lambda_1 + D)y = \\
&= (-\lambda_n + D)\cdot\cdots\cdot(-\lambda_1 + D)y + \lambda_n(-\lambda_{n-1} + D)\cdot\cdots\cdot(-\lambda_1 + D)y\,.
\end{aligned}
$$

The term $(-\lambda_n + D)\cdot\cdots\cdot(-\lambda_1 + D)y$ vanishes, since it coincides with (2.16). Hence we get

$$\xi_n' = \lambda_n\xi_n\,.$$

The statement easily follows.

We emphasize that (2.28) is not a Jordan form of C (it coincides with the Jordan form of C, only in the case where $\lambda_1 = \lambda_2 = \cdots = \lambda_n$). Let P be the matrix such that $C = P^{-1}MP$. Then P has the form

$$
\begin{pmatrix}
1 & 0 & 0 & \ldots & 0 & 0 \\
p_{2,1} & 1 & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
p_{n,1} & p_{n,2} & p_{n,3} & \ldots & p_{n,n-1} & 1
\end{pmatrix}
$$

where the numbers p_{ij} are the same as in the proof of Lemma 2.2. Let us remark that the companion form is not the unique way to rewrite (2.16) as a system of first order equations. We can take for instance

$$
\begin{cases}
z_1 = a_{n-1}y + a_{n-2}y' + \cdots + a_1 y^{(n-2)} + y^{(n-1)} \\
z_2 = a_{n-2}y + a_{n-3}y' + \cdots + a_1 y^{(n-3)} + y^{(n-2)} \\
\ldots\ldots\ldots\ldots \\
z_{n-1} = a_1 y + y' \\
z_n = y\,.
\end{cases}
$$

Then we have

$$\begin{cases} z_1' & = a_{n-1}y' + a_{n-2}y'' + \cdots + a_1 y^{(n-1)} + y^{(n)} = \\ & = -a_n y = -a_n z_n \\ z_2' & = a_{n-2}y' + \cdots + a_1 y^{(n-2)} + y^{(n-1)} = \\ & = a_{n-2}y' + \cdots + a_1 y^{(n-2)} + z_1 - \\ & -(a_{n-1}y + a_{n-2}y' + \cdots + a_1 y^{(n-2)}) = \\ & = z_1 - a_{n-1}y = z_1 - a_{n-1}z_n \\ & \ldots\ldots\ldots\ldots \\ z_n' & = y' = z_{n-1} - a_1 y = z_{n-1} - a_1 z_n \,. \end{cases}$$

This system corresponds to the matrix

$$C^{\mathbf{t}} = \begin{pmatrix} 0 & 0 & \ldots & 0 & -a_n \\ 1 & 0 & \ldots & 0 & -a_{n-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -a_1 \end{pmatrix} .$$

Since all these substitutions are linear and invertible, we have actually proved that C and $C^{\mathbf{t}}$ are similar (as a matter of fact, this is true for every square matrix).

We are now able to conclude the proof. Let us assume that A is similar to its companion form C_A. We know by Lemma 2.2 that A is similar to the matrix M given by (2.28), as well.

Let $w = (0, \ldots, 0, 1)^{\mathbf{t}}$. The result of the multiplication Mw is a vector coinciding with the last column of M. Let us perform the iterated multiplications $M^2 w = M(Mw)$, $M^3 w = M(M^2 w), \ldots$ and let us form a new matrix whose columns are given by the vectors $w, Mw, \ldots, M^{n-1}w$, in this order:

$$\begin{pmatrix} 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 1 & \ldots & * \\ 0 & 1 & \lambda_n + \lambda_{n-1} & \ldots & * \\ 1 & \lambda_n & \lambda_n^2 & \ldots & * \end{pmatrix}$$

where we denoted by $*$ some unessential functions of $\lambda_1, \ldots, \lambda_n$. This matrix is not singular, which means that $w, Mw, \ldots, M^{n-1}w$ are linearly independent. Now let P be the matrix transforming A in M, and let $v = Pw$. We have

$$\begin{aligned} (w|Mw|\ldots|M^{n-1}w) &= (P^{-1}Pw|P^{-1}APw|\ldots|P^{-1}A^{n-1}Pw) = \\ &= P^{-1}(v|Av|\ldots|A^{n-1}v) \end{aligned}$$

which yields the desired conclusion. Vice versa, we finally prove that A is similar to $C_A^{\mathbf{t}}$ (the transpose of the companion form of A) provided that the condition *(v)* of Theorem 2.2 holds. Setting $R = (v, Av, \dots, A^{n-1}v)$, we have to prove that $R^{-1}AR = C_A^{\mathbf{t}}$ or, equivalently,

$$AR = (Av|A^2v|\dots|A^nv) = RC_A^{\mathbf{t}} =$$
$$= (v|Av|\dots|A^{n-1}v) \cdot \begin{pmatrix} 0 & 0 & \dots & 0 & -a_n \\ 1 & 0 & \dots & 0 & -a_{n-1} \\ 0 & 1 & \dots & 0 & -a_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_1 \end{pmatrix} .$$

The computations are not difficult (for the last column we need to apply the Cayley-Hamilton Theorem). We already know that a matrix in companion form and its transpose are similar, but we can also proceed in a direct way. Indeed, it is sufficient to remark that a matrix in companion form satisfies *(v)* with $v = (0, \dots, 0, 1)^{\mathbf{t}}$. By repeating the same computations as before, we recover the required similarity.

Chapter Summary

This chapter is devoted to the mathematical problem of representing the solutions of a homogeneous system of linear differential equations by means of suitable explicit formulæ. This corresponds to the study of the qualitative behavior of a system when the evolution depends only on the internal forces and the external inputs are switched off. It is actually the first step in the investigation of the properties of a system.

Chapter 3
Stability of Unforced Linear Systems

In this chapter we continue our investigation of the properties of linear unforced differential systems

$$\dot{x} = Ax \tag{3.1}$$

where A is a square $n \times n$ matrix with real entries, and $x \in \mathbf{R}^n$. We focus in particular on the stability problem.

3.1 Equilibrium Positions

The equilibrium positions of system (3.1) coincide with the solutions of the algebraic equation $Ax = 0$. Hence, a system of the form (3.1) always have an equilibrium position for $x = 0$. Such an equilibrium position is unique (and hence isolated) if and only if $\det A \neq 0$. Otherwise, there are infinitely many equilibrium positions (none of which isolated): more precisely, the set of all the equilibrium positions of (3.1) constitutes a subspace of $\mathbf{R}^n$.

Remark 3.1 Assume that there is a point $\bar{x} \neq 0$ such that $A\bar{x} = 0$. Then, $\bar{x}$ is a stable equilibrium position for system (3.1) if and only if the origin is a stable equilibrium position for system (3.1). Indeed, setting $y = x - \bar{x}$, we have

$$\dot{y} = \dot{x} = Ax = Ax - A\bar{x} = Ay.$$

■

The displacements of x with respect to $\bar{x}$, that is the displacement of y with respect to $y = 0$, are determined by the same system which determines the displacements of x with respect to $x = 0$.

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_3

Remark 3.2 If an equilibrium point is attractive for system (3.1), then it must be isolated. Hence, if A is singular, there exist no attractive equilibrium positions. In other words, if system (3.1) possesses an attractive equilibrium position $\bar{x}$, then $\bar{x} = 0$, and there are no other equilibria $\bar{x} \neq 0$. ■

According to the previous remarks, when studying stability and asymptotic stability of linear systems, it is not restrictive to limit ourselves to the origin.

Proposition 3.1 *If the origin is stable [respectively, asymptotically stable] for system (3.1), then the origin is stable [respectively, asymptotically stable] for all the systems linearly equivalent to (3.1).*

Proof Let $B = P^{-1}AP$ and let $\psi(t)$ be any solution of the system $\dot{y} = By$. Let us fix $\varepsilon > 0$, and let $\varepsilon' = \varepsilon/\|P^{-1}\|$. Since (3.1) is stable at the origin, there exists $\delta' > 0$ such that $\|\varphi(0)\| < \delta' \Longrightarrow \|\varphi(t)\| < \varepsilon'$ for each $t \geq 0$ and each solution $\varphi(t)$ of (3.1). Let $\delta = \delta'/\|P\|$ and $\varphi(t) = P\psi(t)$. Then,

$$\|\psi(0)\| < \delta \Longrightarrow \|\varphi(0)\| = \|P\psi(0)\| \leq \|P\| \cdot \|\psi(0)\| < \delta'$$

so that

$$\|\psi(t)\| = ||P^{-1}\varphi(t)|| \leq \|P^{-1}\| \cdot \|\varphi(t)\| < \varepsilon \ .$$

Finally, assume that $\lim_{t\to+\infty} \varphi(t) = 0$ for a given solution of (3.1). Then, for each $\sigma > 0$ there exists $T > 0$ such that

$$t > T \Longrightarrow \|\varphi(t)\| < \sigma'$$

where $\sigma' = \sigma/\|P^{-1}\|$, and this implies that $\|\psi(t)\| \leq \|P^{-1}\| \cdot \|\varphi(t)\| < \sigma$. The reasoning is easily completed. ■

3.2 Conditions for Stability

For linear time invariant systems, the analysis of the stability properties can be carried on by means of purely algebraic tools.

Definition 3.1 We say that a real square matrix A possesses the *Hurwitz property* if all the eigenvalues of A have (strictly) negative real part.

In short, when A possesses the Hurwitz property we shall also say that A is a *Hurwitz matrix*. Note that every Hurwitz matrix is nonsingular.

Theorem 3.1 *If A is a Hurwitz matrix then the origin is a globally and exponentially stable equilibrium point for system (3.1). If the origin is a locally attractive equilibrium point for system (3.1), then A is a* Hurwitz matrix.

Proof Assume that all the eigenvalues of A have negative real part. Then we can choose α and k_0 in (2.15) in such a way that $\alpha_0 < \alpha < 0$. The global and exponential attraction of the origin trivially follows. As far as the stability property is concerned, we may use again (2.15). As already noticed, it is not restrictive to take $\alpha < 0$; for $t \geq 0$, we have therefore $e^{\alpha t} \leq 1$. Hence, for each $\varepsilon > 0$, it is sufficient to take $\delta = \varepsilon / k_0$.

We now pass to the second statement. Being the origin locally attractive, there exists a neighborhood Ω of the origin such that all the solutions issuing from a point of Ω asymptotically approach zero when $t \to +\infty$. We proceed by distinguishing several cases.

Assume first that there is an eigenvalue λ with strictly positive real part. If λ is real and if v is a corresponding (real) eigenvector, then we can construct a solution of the form $e^{\lambda t} v$. Note that the norm of v can be taken arbitrarily small. Instead, if $\lambda = \alpha + \mathrm{i}\,\beta$ is not real, then we can construct a solution of the form

$$e^{\alpha t}[(\cos \beta t)u + (\sin \beta t)w] ,$$

where u, w are certain real vectors, whose norm can be taken arbitrarily small, and $\alpha > 0$. In both cases, these solutions are unbounded for $t \geq 0$. This contradicts the assumptions.

In similar way we exclude the existence of eigenvalues λ with zero real part. Indeed, in this case either $\lambda = 0$, so that there is a nonzero constant solution, or λ is purely imaginary, so that we can construct a periodic solution $(\cos \beta t)u + (\sin \beta t)w$, which is bounded but does not approach zero. ■

From Theorem 3.1 and its proof we can infer other information, which can be resumed in the following way.

- For the linear system (3.1), the condition that A possesses the Hurwitz property is necessary and sufficient for the asymptotic stability of the origin.
- If the origin is locally attractive for the linear system (3.1), then it is globally and exponentially attractive, as well.
- For a linear system, if the origin is locally attractive then it is also stable.

Instead, even in the case of a linear system it may happen that the origin is stable but not attractive; very simple examples are given by the system $\dot{x} = 0$ with $x \in \mathbf{R}$, whose solutions are constant, and by the system in Example 2.1, whose solutions are periodic.

From the proof of Theorem 3.1, we also immediately see that if there exists an eigenvalue of A with strictly positive real part then the origin is unstable. Thus, it remains to discuss the case where all the eigenvalues of A have non-positive real part, and at least one among them has a real part exactly equal to zero.

Theorem 3.2 *The following statements are equivalent.*

(i) All the eigenvalues of A have non-positive real part, and for each possible eigenvalue with zero real part, the algebraic multiplicity and the geometric multiplicity coincide.

(ii) *The norm of the exponential matrix* e^{tA} *is bounded for* $t \geq 0$.
(iii) *The origin is stable for system (3.1).*

Proof $(i) \Longrightarrow (ii)$. If all the eigenvalues of A have nonpositive real part and the possible eigenvalues with zero real part have the same algebraic and geometric multiplicity, then we can apply (2.14) with $\alpha = \alpha_0 = 0$. The conclusion is straightforward.

$(ii) \Longrightarrow (iii)$. If there is a constant $k_0 > 0$ such that $||e^{tA}|| \leq k_0$ for $t \geq 0$, then for every $x_0 \in \mathbf{R}^n$ we have

$$\|e^{tA}x_0\| \leq k_0\|x_0\| \ . \tag{3.2}$$

The definition of stability is recovered taking $\delta = \varepsilon/k_0$.

Finally we prove by contradiction that $(iii) \Longrightarrow (i)$. We already know that if the origin is stable, there exist no eigenvalues with strictly positive real part. Assume that there is an eigenvalue λ with zero real part and whose geometric multiplicity is less than its algebraic multiplicity.

If $\lambda = \mathrm{i}\,\beta$ con $\beta \neq 0$, we could construct a complex solution of the form $(\cos\beta t + \mathrm{i}\,\sin\beta t)(tu + v)$, where v is an eigenvector corresponding to λ, and u is a generalized eigenvector; both v and u can be chosen of arbitrarily small norm. But then, we could also find a real solution

$$(\cos\beta t)[tu_1 + v_1] - (\sin\beta t)[tu_2 + v_2]$$

where v_1, v_2, u_1, u_2 are some real vectors. This solution corresponds to the initial state $x_0 = v_1$. Since u_1 and u_2 cannot be both zero, the solution exhibits an oscillatory behavior and the amplitude of the oscillations increases as t increases. This solution is not bounded for $t \geq 0$, so that the stability assumption is contradicted.

The case $\lambda = 0$ can be ruled out in similar way. ■

We may also prove the following proposition by analogous arguments.

Proposition 3.2 *The following statements are equivalent.*

(i) *All the eigenvalues of A have nonpositive real part, and for each possible eigenvalue with zero real part, the algebraic multiplicity and the geometric multiplicity coincide.*
(ii) *All the solutions of the system are bounded on* $t \geq 0$.

Remark 3.3 If the system at hand is defined by a scalar differential equation of order n like (2.16), the stability conditions of the equilibrium position $y = y' = \ldots = y^{(n-1)} = 0$ can be stated in terms of its characteristic roots. ■

3.3 Lyapunov Matrix Equation

In this section we present a different characterization of stable linear systems. Recall that a real symmetric matrix P is said to be:

- *positive definite* if for each $0 \neq x \in \mathbf{R}^n$ we have $x^{\mathbf{t}} P x > 0$.
- *positive semidefinite* if $x^{\mathbf{t}} P x \geq 0$ for each $x \in \mathbf{R}^n$.

If the conditions above are fulfilled by the matrix $-P$, then we say that, respectively, P is *negative definite, negative semidefinite*. Finally, we say that P is *indefinite* if $x^{\mathbf{t}} P x$ takes both strictly positive and strictly negative values.

Theorem 3.3 *The following statements are equivalent.*

(i) *A possesses the Hurwitz property.*

(ii) *There exists a positive definite, real symmetric matrix P such that for each nontrivial solution $\varphi(t)$ of (3.1), we have*

$$\left.\frac{d}{dt} g(t)\right|_{t=0} = -||\varphi(0)||^2 \tag{3.3}$$

where $V(x) = x^{\mathbf{t}} P x$ and $g(t) = V(\varphi(t))$.

(iii) *There exists a positive definite, real symmetric matrix P such that*

$$A^{\mathbf{t}} P + P A = -I\ . \tag{3.4}$$

Proof $(i) \Longrightarrow (ii)$. Assume that (i) holds and denote by p_{ij} the unknown elements of the matrix P. Let $\psi_1(t), \ldots, \psi_n(t)$ be the columns of the exponential matrix e^{tA}, and let us define

$$p_{ij} = \int_0^{+\infty} \psi_i^{\mathbf{t}}(s)\psi_j(s)ds\ .$$

The numbers p_{ij} are well defined: indeed, if the eigenvalues of A have negative real part, all the entries of e^{tA} go to zero exponentially and so, the integral converges. Let us check that the matrix P fulfils the required properties. Clearly, P is symmetric. The solution $\varphi(t)$ corresponding to the initial state $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$, can be written as

$$\varphi(t) = e^{tA}x = \sum_{i=1}^{n} \psi_i(t)x_i\ .$$

Thus,

$$\begin{aligned}
V(x) = x^{\mathbf{t}} P x = \sum_{i,j=1}^{n} p_{ij}x_i x_j &= \sum_{i,j=1}^{n} \left(\int_0^{+\infty} \psi_i^{\mathbf{t}}(s)\psi_j(s)ds \right) x_i x_j \\
&= \int_0^{+\infty} \sum_{i,j=1}^{n} \left(\psi_i^{\mathbf{t}}(s)\psi_j(s)\right) x_i x_j\, ds \\
&= \int_0^{+\infty} \left(\sum_{i=1}^{n} \psi_i(s)x_i \right)^{\mathbf{t}} \left(\sum_{j=1}^{n} \psi_j(s)x_j \right) ds
\end{aligned}$$

$$= \int_0^{+\infty} \|\varphi(s)\|^2\, ds\ .$$

The properties $V(0) = 0$ and $V(x) > 0$ for each $x \neq 0$ are easily checked. It remains to prove (3.3). Since

$$V(\varphi(t)) = \int_0^{+\infty} \|\varphi(t+s)\|^2\, ds = \int_t^{+\infty} \|\varphi(\sigma)\|^2\, d\sigma$$

we have, for each $x \in \mathbf{R}^n$,

$$\begin{aligned}\frac{d}{dt} V(\varphi(t))\Big|_{t=0} &= \frac{d}{dt}\int_t^{+\infty} \|\varphi(\sigma)\|^2 d\sigma\Big|_{t=0} \\ &= -\|\varphi(t)\|^2\Big|_{t=0} = -\|x\|^2\ .\end{aligned}$$

Now we prove that $(ii) \Longrightarrow (iii)$. We reconsider the already defined function $V(\varphi(t))$ and we compute its derivative in a different way. We have

$$\begin{aligned}\frac{d}{dt} V(\varphi(t)) &= \frac{d}{dt}\big((\varphi(t))^{\mathbf{t}} P\varphi(t)\big) = (\dot\varphi(t))^{\mathbf{t}} P\varphi(t) + (\varphi(t))^{\mathbf{t}} P\dot\varphi(t) \\ &= (\varphi(t))^{\mathbf{t}} A^{\mathbf{t}} P\varphi(t) + (\varphi(t))^{\mathbf{t}} PA\varphi(t)\ .\end{aligned}$$

Setting $t = 0$ and $\varphi(0) = x$, and taking the assumption into account, the identity above yields

$$x^{\mathbf{t}}[A^{\mathbf{t}}P + PA]x = -x^{\mathbf{t}}x\ .$$

Since the solution $\varphi(t)$ is arbitrary, we obtain $A^{\mathbf{t}}P + PA = -I$, as required.

Finally, we prove that $(iii) \Longrightarrow (i)$. Let λ be a (real or complex) eigenvalue of A, and let $Av = \lambda v$ ($v \neq 0$). we have

$$\begin{aligned}-\bar v^{\mathbf{t}} v &= \bar v^{\mathbf{t}}(A^{\mathbf{t}}P + PA)v = (A\bar v)^{\mathbf{t}} Pv + \bar v^{\mathbf{t}} PAv \\ &= \bar\lambda \bar v^{\mathbf{t}} Pv + \lambda \bar v^{\mathbf{t}} Pv = (\bar\lambda + \lambda)\bar v^{\mathbf{t}} Pv = 2\alpha\, \bar v^{\mathbf{t}} Pv\end{aligned}$$

where α denotes the real part of λ. Now, it is not difficult to check that if P is any positive definite, real symmetric matrix and if v is any (real or complex) nonzero vector, then $\bar v^{\mathbf{t}} Pv > 0$. Hence we must have $\alpha < 0$. ■

In what follows, we refer to (3.4) as the *Lyapunov matrix equation*. Indeed, it can be conveniently interpreted as an equation in the unknown P. It is equivalent to a system with $n(n+1)/2$ algebraic linear equations, whose unknowns are the elements of P; Theorem 3.3 states in particular that if A is a Hurwitz matrix, then such a system admits a solution. More precisely, under this condition there exists a unique positive definite solution of (3.4).

A function of the form $V(x) = x^{\mathbf{t}}Px$ where P is a positive definite, real symmetric matrix enjoying one of the properties (ii) or (iii) listed in Theorem 3.3, is called a *quadratic Lyapunov function*. It is indeed a homogeneous real polynomial of degree 2 with n variables.

The following corollary provides a generalized version of (3.4). It enlightens in particular the flexibility of the matrix Lyapunov equation.

Corollary 3.1 *If there exists a positive definite, real symmetric matrix Q such that the matrix equation*

$$A^{\mathbf{t}}P + PA = -Q \tag{3.5}$$

admits a (positive definite, real symmetric) solution P, then A is a Hurwitz matrix.

If A is a Hurwitz matrix, then for each positive definite, real symmetric matrix Q, there exists a unique (positive definite, real symmetric) solution P of the matrix equation (3.5).

Proof The proof of the first statement is a slight modification of the proof that $(iii) \Longrightarrow (i)$ of Theorem 3.3. As far as the second statement is concerned, we start by writing $Q = R^{\mathbf{t}}R$, where R is some nonsingular symmetric matrix (see [21] Chap. 11, or [6]). Since the eigenvalues of A have strictly negative real part, the same is true for the matrix $\tilde{A} = RAR^{-1}$. According to Theorem 3.3 (iii), there exists a matrix $\tilde{P}$ such that

$$\tilde{A}^{\mathbf{t}}\tilde{P} + \tilde{P}\tilde{A} = -I\ .$$

This implies

$$(R^{-1})^{\mathbf{t}}A^{\mathbf{t}}R^{\mathbf{t}}\tilde{P} + \tilde{P}RAR^{-1} = -I\ .$$

The conclusion follows, multiplying both sides of this equation by $R^{\mathbf{t}}$ and R (respectively, on the left and on the right), and setting $P = R^{\mathbf{t}}\tilde{P}R$. ■

An elegant representation of the solution of (3.5) is (see [31]) $P = \int_0^\infty e^{tA^{\mathbf{t}}} Qe^{tA}\,dt$. In order to characterize the stability property, we may use a weakened version of Theorem 3.3.

Theorem 3.4 *The following properties are equivalent.*

(i) All the eigenvalues of A have nonpositive real part, and for each possible eigenvalue with zero real part the algebraic multiplicity and the geometric multiplicity coincide.
(ii) There exists a positive definite, real symmetric matrix P such that the matrix $A^{\mathbf{t}}P + PA$ is negative semidefinite.

A function of the form $V(x) = x^{\mathbf{t}}Px$ with P positive definite, real symmetric, enjoying one of the properties listed in Theorem 3.4 is called a *quadratic weak Lyapunov function*.

3.4 Routh-Hurwitz Criterion

The results presented in this chapter emphasize the interest of criteria which enable us to predict the sign of the roots of a polynomial, without need of computing them explicitly. Recall that the eigenvalues of a matrix A coincide with the roots of the characteristic polynomial of A. Let

$$P(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$$

be a monic polynomial of degree n with real coefficients.

Proposition 3.3 *If all the roots of $P(\lambda)$ have strictly negative real part, then all the coefficients a_i must be strictly grater than zero.*

Proof Let $\lambda_1, \dots, \lambda_k$ be the (negative) real roots of $P(\lambda)$ and let $\alpha_1 \pm \mathrm{i}\,\beta_1, \dots, \alpha_h \pm \mathrm{i}\,\beta_h$ be the complex conjugate roots, with $\alpha_1 < 0, \dots, \alpha_h < 0$. Then we have

$$\begin{aligned} P(\lambda) = {} & (\lambda - \lambda_1)^{\mu_1} \cdot \ldots \cdot (\lambda - \lambda_k)^{\mu_k} \\ & \cdot(\lambda - (\alpha_1 + \mathrm{i}\,\beta_1))^{\nu_1} \cdot (\lambda - (\alpha_1 - \mathrm{i}\,\beta_1))^{\nu_1} \cdot \ldots \\ & \cdot(\lambda - (\alpha_h + \mathrm{i}\,\beta_h))^{\nu_h} \cdot (\lambda - (\alpha_h - \mathrm{i}\,\beta_h))^{\nu_h} \,. \end{aligned}$$

Every pair of linear factors where the complex roots appear can be replaced by a unique factor of degree 2

$$\lambda^2 + p_1\lambda + q_1 \,, \ \ldots \ , \ \lambda^2 + p_h\lambda + q_h$$

where, being $\alpha_1 < 0, \dots, \alpha_h < 0$, all the coefficients $p_1, \dots, p_h, \ q_1, \dots, q_h$ are positive.

Recovering the expanded form of the polynomial, we find of course that all the coefficients of $P(\lambda)$ are positive. ■

The necessary condition provided by Proposition 3.3 is also sufficient if the degree of the polynomial is $n = 1$ or $n = 2$, but not in general. For instance,

$$\left[\lambda - \frac{1 + \mathrm{i}\sqrt{11}}{2}\right] \cdot \left[\lambda - \frac{1 - \mathrm{i}\sqrt{11}}{2}\right] \cdot [\lambda + 2] = \lambda^3 + \lambda^2 + \lambda + 6 \,.$$

There are several necessary and sufficient conditions which allow us to establish when the roots of a polynomial belong to the half plane $\{z \in \mathbf{C} : \quad \mathrm{Re}\, z < 0\}$. They are generally referred to as *Routh* and *Hurwitz* criteria. We state one of these criteria without proof. It is based on the examination of the sign of the determinants of n matrices $\Delta_1, \Delta_2, \dots, \Delta_n$ of order $1, 2, \dots, n$, respectively. These matrices are computed starting from the coefficients of $P(\lambda)$, according to the following procedure.

First of all, for sake of convenience, we agree to write $a_j = 0$ for each value of $j > n$, and $a_0 = 1$. We define

$$\Delta_1 = a_1 \ ,$$

$$\Delta_2 = \begin{pmatrix} a_1 & a_3 \\ a_0 & a_2 \end{pmatrix},$$

$$\Delta_3 = \begin{pmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{pmatrix},$$

$$\Delta_4 = \begin{pmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & 1 & a_2 & a_4 \end{pmatrix},$$

$$\Delta_5 = \begin{pmatrix} a_1 & a_3 & a_5 & a_7 & a_9 \\ a_0 & a_2 & a_4 & a_6 & a_8 \\ 0 & a_1 & a_3 & a_5 & a_7 \\ 0 & 1 & a_2 & a_4 & a_6 \\ 0 & 0 & a_1 & a_3 & a_5 \end{pmatrix},$$

and so on, finishing with Δ_n. Let us remark that on the "odd" rows of these matrices (the first row, the third row, etc.) we find the coefficients with odd index, displayed in increasing order, while on the "even" rows we find the coefficients of even index. The elements which appear in the first two rows are repeated in the following rows, shifted of one position each time. The free positions at the beginning of any new row are filled with zeros, while the last element on the right of any row is eliminated at a new repetition.

Theorem 3.5 *All the roots of the polynomial $P(\lambda)$ belong to the half plane $\{z \in \mathbf{C} : \mathrm{Re} z < 0\}$ if and only if all the determinants of the matrices $\Delta_1, \ldots, \Delta_n$ are positive.*

For instance, in the case $n = 3$ the condition of Theorem 3.5 reduces to

$$a_1 > 0, \ a_3 > 0, \ a_1 a_2 - a_3 > 0 \ .$$

This form of the Routh-Hurwitz criterion can be found in [24] or in [10], where the reader can also find a proof of Theorem 3.5.

Chapter Summary

In this chapter the study of unforced linear systems is continued. We focus in particular on the stability properties of the equilibrium position (the origin). This corresponds to the study of the internal stability properties of a system with input and output. We state and prove the classical Lyapunov Theorem which allows us to reduce the stability analysis to an algebraic problem (computation of the eigenvalues of a matrix). We also introduce the quadratic Lyapunov functions and the Lyapunov matrix equation. The Routh-Hurwitz criterion is given without proof.

Chapter 4
Linear Systems with Forcing Term

The simplest way to model an external input is to introduce an additive term in the system equations. In this chapter we shall see how the solutions of a system of differential equations, whose right-hand side is the sum of a linear part and a time-varying term, can be explicitly found.

4.1 Nonhomogeneous Systems

A linear nonhomogeneous system[1] of differential equations has the general form

$$\dot{x} = Ax + b(t) \tag{4.1}$$

where $b(t)$, frequently referred to as the *forcing term*, belongs to the space $\mathcal{PC}(I, \mathbf{R})$. Here, I denotes in general any interval of $\mathbf{R}$ with nonempty interior, although for our purposes, the relevant cases are $I = \mathbf{R}$ and $I = [0, +\infty)$. We report below some basic facts.

Fact 1. *For each initial instant $t_0 \in I$ and each initial state $x_0 \in \mathbf{R}^n$ there exists a unique solution $x = \psi(t)$ of (4.1) such that $\psi(t_0) = x_0$. Moreover, $\psi(t)$ is defined for each $t \in I$.*

Fact 2. *If $\psi_1(t)$, $\psi_2(t)$ are solutions of (4.1) defined on I, then $\psi_1(t) - \psi_2(t)$ is a solution of the so-called* associated homogeneous system

$$\dot{x} = Ax. \tag{4.2}$$

[1] According to a more correct terminology, a system of the form (4.1) should be called an "affine" system; however, the term "linear nonhomogeneous" is very frequent in the literature.

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_4

Fact 3. *If $\varphi(t)$ is any solution of the associated homogeneous system* (4.2) *and $\psi^*(t)$ is any solution of the nonhomogeneous system* (4.1), *then $\varphi(t) + \psi^*(t)$ is a solution of the nonhomogeneous system* (4.1).

Fact 4. (Superposition principle) If $\psi_1(t)$ *is a solution of system* (4.1) *with $b(t) = b_1(t)$ and $\psi_2(t)$ is a solution of system* (4.1) *with $b(t) = b_2(t)$, then $\psi_1(t) + \psi_2(t)$ is a solution of system* (4.1) *with $b(t) = b_1(t) + b_2(t)$.*

From Facts 2 and 3 it follows that in order to determine the set of all the solutions of system (4.1), we need to find:

(a) a fundamental matrix $\Phi(t)$ of (4.2);
(b) a particular solution $\psi^*(t)$ of (4.1).

The set of all the solutions of system (4.1) can be therefore represented by the formula

$$x = \psi(t) = \Phi(t)c + \psi^*(t) \tag{4.3}$$

where c is a vector of arbitrary constants. It is called the *general integral* of system (4.1). The particular solution corresponding to a given initial condition (t_0, x_0) can be obtained solving the algebraic system

$$x_0 - \psi^*(t_0) = \Phi(t_0)c.$$

If $\Phi(t) = e^{(t-t_0)A}$, then $c = x_0 - \psi^*(t_0)$.

4.1.1 The Variation of Constants Method

The problem of determining a fundamental matrix of system (4.2) has been solved in Chap. 2. As far as point (b) is concerned, we have the following general result.

Proposition 4.1 *The function*

$$\psi_0^*(t) = \int_{t_0}^{t} e^{(t-\tau)A} b(\tau)\, d\tau \tag{4.4}$$

provides the solution of (4.1) such that $\psi_0^*(t_0) = 0$.

Taking into account this result, we can write the solution corresponding to the initial state (t_0, x_0) as

$$\psi(t) = e^{(t-t_0)A}x_0 + \int_{t_0}^{t} e^{(t-\tau)A} b(\tau) d\tau = e^{(t-t_0)A}(x_0 + \int_{t_0}^{t} e^{(t_0-\tau)A} b(\tau) d\tau). \tag{4.5}$$

This is called *Lagrange formula* or *variation of constants formula*. This formula is very well suited for theoretical purposes but sometimes not so convenient in practice, because of the presence of the integral that, for certain functions $b(t)$, might be hard or even impossible to compute explicitly.

Remark 4.1 Formula (4.5) is often used with $t_0 = 0$ (provided of course that $0 \in I$), that is in the form

$$\psi(t) = e^{tA}x_0 + \int_0^t e^{(t-\tau)A}b(\tau)\,d\tau. \tag{4.6}$$

There is an interesting interpretation of (4.6): it shows that each solution is the sum of two contributions. The first one depends on the initial state but not on the forcing term. On the contrary, the second one depends on the forcing term but not on the initial state. This suggests that the analysis of the dynamical behavior of a linear system can be carried out by investigating separately the effect of the initial conditions (with zeroed forcing term) and the effect of the forcing term (with zeroed initial state).[2] ■

Nonhomogeneous equations arise frequently in applications, both in classical physics and in system theory. In any case, it is natural to presume that $b(t)$ represents a signal generated by an *exosystem*, that is an external system connected to the main plant by a cascade connection. Exosystems are often simple linear devices without forcing terms. Hence it is reasonable to focus on forcing terms of the form

$$b(t) = \sum_{h=1}^{H} P_h(t)e^{\gamma_h t}$$

where each $P_h(t)$ is a polynomial with vector coefficients and $\gamma_h \in \mathbf{C}$. As illustrated in the next section, in such cases the computation of the integral in (4.5) can be avoided by virtue of the superposition principle and the use of some practical rules which allows us to find a particular solution in a more direct way. These rules are presented in the next section.

4.1.2 The Method of Undetermined Coefficients

We can limit ourselves to assume $b(t) = P(t)e^{\gamma t}$, where $P(t)$ is a polynomial with vector coefficients. We distinguish two cases.

Case 1: γ is not an eigenvalue of A. Then, there exists a particular solution of (4.1) with the following structure: $\psi^*(t) = Q(t)e^{\gamma t}$ where Q is a polynomial with vector coefficients and the same degree of P.

Case 2: γ is an eigenvalue of A, with algebraic multiplicity $\mu \geq 1$. Then, there exists a particular solution of (4.1) with the following structure: $\psi^*(t) = Q(t)e^{\gamma t}$ where Q

---

[2] This agrees with the conclusions of Chap. 1 (Sect. 1.3.5) provided that the forcing term is interpreted as an input and taking into account Proposition 1.10.

is a polynomial with vector coefficients, and $\deg Q = \deg P + \mu$: in this case, we say that the system exhibits *resonance*.

In both cases, the coefficients of Q depend on A, γ and the coefficients of $P(t)$; they can be determined by exploiting the identity $\dot{\psi}^*(t) = A\psi^*(t) + P(t)e^{\gamma t}$, which leads to a system of merely algebraic equations. This is the reason why this procedure is called the *method of undetermined coefficients*. We emphasize that these rules hold even if γ is a complex number. By virtue of the formulæ

$$\cos \omega t = \frac{e^{\mathrm{i}\,\omega t} + e^{-\mathrm{i}\,\omega t}}{2}, \qquad \sin \omega t = \frac{e^{\mathrm{i}\,\omega t} - e^{-\mathrm{i}\,\omega t}}{2\mathrm{i}}$$

the method of undetermined coefficients can be therefore extended to forcing terms of the form $b(t) = P_1(t)\cos \omega t + P_2(t)\sin \omega t$ ($P_1(t)$ and $P_2(t)$ being polynomials with real vector coefficients).

Example 4.1 We are especially interested in the case where the forcing term is a periodic function of the form

$$b(t) = (\cos \omega t)u + (\sin \omega t)v \tag{4.7}$$

where $u, v \in \mathbf{R}^n$ are constant vectors. This case occurs frequently in applications, and it will be further developed later in this chapter, in different situations. Let us apply the method of undetermined coefficients separately to the systems

$$\dot{x} = Ax + \frac{e^{\mathrm{i}\,\omega t}}{2}(u - \mathrm{i}\,v) \quad \text{and} \quad \dot{x} = Ax + \frac{e^{-\mathrm{i}\,\omega t}}{2}(u + \mathrm{i}\,v).$$

Assuming for simplicity that resonance does not occur and taking into account that the elements of A and u are real, we find respectively particular solutions of the type

$$\psi_1(t) = e^{\mathrm{i}\,\omega t}c \quad \text{and} \quad \psi_2(t) = e^{-\mathrm{i}\,\omega t}\bar{c}$$

for some constant vector $c \in \mathbf{C}^n$. According to the superposition principle, a particular solution of the system will be found of the form

$$\psi^*(t) = e^{\mathrm{i}\,\omega t}c + e^{-\mathrm{i}\,\omega t}\bar{c}.$$

This solution is actually real, since it is the sum of two conjugate terms. It can be rewritten as

$$\psi^*(t) = (\cos \omega t)a + (\sin \omega t)b \tag{4.8}$$

for some vector constants $a, b \in \mathbf{R}^n$. ■

Remark 4.2 It is important to notice that (4.8) is a periodic solution, with the <u>same</u> frequency as the forcing term (4.7). It should be also noticed that in (4.8) a and b may be both nonzero, even if in (4.7) one between u and v is zero. ■

Remark 4.3 Notice that in general it is not possible to preassign the initial state of the particular solution obtained by the method of undetermined coefficients. In general, we will have $\psi^*(t_0) \neq 0$, so that it does not coincide with the solution introduced in Proposition 4.1. More precisely, let $\psi^*(t)$ be a particular solution obtained by the method of undetermined coefficients, and let for simplicity $t_0 = 0$. We may rewrite (4.3), as

$$x = e^{tA}(x_0 - \psi^*(0)) + \psi^*(t) \tag{4.9}$$

where x_0 stands for the desired initial state. The particular solution provided by the method of variation of constants can be recovered as

$$\psi_0^*(t) = \psi^*(t) - e^{tA}\psi^*(0). \tag{4.10}$$

■

4.2 Transient and Steady State

Throughout this section, we assume that $I = [0, +\infty)$ and $t_0 = 0$. In addition, we assume that the matrix A in (4.1) possesses the Hurwitz property (Definition 3.1).

Using (2.14), we may give an asymptotic estimation of the solutions also for nonhomogeneous systems of type (4.1).

Proposition 4.2 *If A is a Hurwitz matrix and $b(t)$ is bounded on the interval $[0, +\infty)$, then for each solution $\psi(t)$ of (4.1) we have*

$$\|\psi(t)\| \leq k_0\|x_0\|e^{\alpha t} + k_1 \cdot b_0, \quad t \geq 0,$$

where k_0 and k_1 are positive constants, $\alpha < 0$, $b_0 = \sup_{\tau \geq 0} \|b(\tau)\|$, and $x_0 = \psi(0)$.

Proof The assumptions imply the existence of constants $\alpha < 0$ and $k_0 > 0$ such that for each t and each $\tau \in [0, t]$

$$\|e^{(t-\tau)A}b(\tau)\| \leq k_0\|b(\tau)\|e^{(t-\tau)\alpha}.$$

Since the initial state x_0 is assigned for $t_0 = 0$, we may use the version (4.6) of the variation of constants formula. We have:

$$\begin{aligned}\|\psi(t)\| \leq k_0\|x_0\|e^{\alpha t} + b_0k_0\int_0^t e^{(t-\tau)\alpha}\,d\tau &= k_0\|x_0\|e^{\alpha t} - \frac{b_0k_0}{\alpha}\left[e^{(t-\tau)\alpha}\right]_0^t \\ &= k_0\|x_0\|e^{\alpha t} + \frac{b_0k_0}{\alpha}\left[e^{\alpha t} - 1\right].\end{aligned}$$

Being $\alpha < 0$, we have $e^{\alpha t} \leq 1$ for $t > 0$. Setting $k_1 = \frac{k_0}{|\alpha|}$, the previous inequality reduces to the desired one. ■

Proposition 4.2 implies in particular that if the matrix A is Hurwitz and the forcing term is bounded, then every solution is bounded on $[0, +\infty)$. We want to focus on the following two particular cases:

(1) the forcing term $b(t)$ is constant;
(2) the forcing term $b(t)$ is a periodic function of the form (4.7).

Note that in force of the Hurwitz property, resonance does not occur neither in case (1) nor in case (2). As a consequence, the system admits a unique constant solution in case (1) and, respectively, a unique periodic solution[3] in case (2). One such solution $\psi^*(t)$ can be used in (4.9), in order to represent a generic solution. Recalling again that A is Hurwitz and using (2.14) with $\alpha < 0$, we have that

$$\lim_{t\to+\infty} e^{tA}c = 0$$

for each $c \in \mathbf{R}^n$. This means that in (4.9), for sufficiently large t, the term $e^{tA}(x_0 - \psi^*(0))$ can be neglected and the evolution of the system "becomes independent" of the initial state x_0. It is approximately constant or periodic, and it is essentially determined by the forcing term. It is customary to distinguish two stages in the time evolution of the system. The first stage, where the evolution is appreciably affected by the initial state x_0, is called the *transient*. The subsequent stage, where the effect of the initial state is no more perceptible, is called the *steady state*. Of course, the distinction between the transient and the steady state is not rigorous, since the term $e^{tA}(x_0 - \psi^*(0))$ in (4.9) will never be exactly equal to zero. Distinguishing the two stages depends on the admitted error margins and on the precision of the measurements, but it is very impressive and convenient, at least from the heuristic point of view.

Remark 4.4 The steady state solution is be more correctly thought of as a "limit" solution, asymptotically approached by all the solutions of system (4.1). As already noticed, such a limit solution does not necessarily vanish for $t = 0$, and so it does not coincide, in general, with the particular solution appearing in the variation of constants formula (4.6). Indeed, as we can understand from (4.10), further terms vanishing when $t \to +\infty$, could be hidden in the particular solution appearing in (4.6). These terms compensate for the gap between the assigned initial state and the initial state of the steady state solution. The steady state solution is found in a natural way when the method of undetermined coefficients is adopted. ■

Example 4.2 Consider the system represented by the scalar differential equation

$$\dot{x} = -x + 2 \tag{4.11}$$

with the initial condition $x(0) = 1$. The general integral of the associated homogeneous system is $x = e^{-t}c$, with c an arbitrary constant. A solution of the

[3] On the other hand, it is easy to check that there exists a constant or periodic solution only if the forcing term is, respectively, constant or periodic.

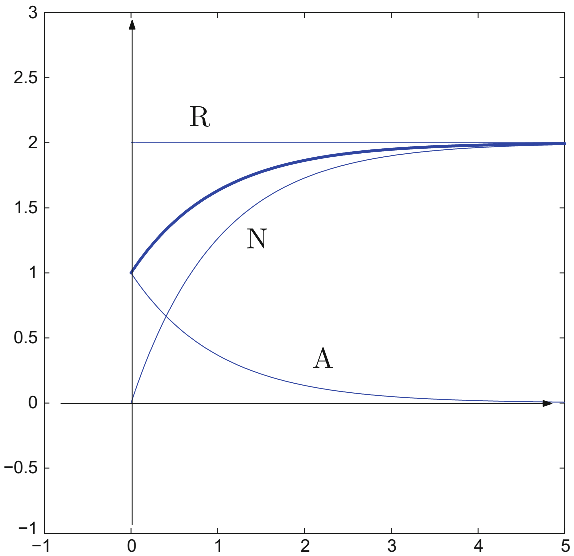


Fig. 4.1 The curve in bold represents the graph of the solution of (4.11) such that $x(0) = 1$; the curve marked by A represents the graph of the solution of the associated homogeneous equation with the same initial condition $x(0) = 1$; the curve marked by R represents the graph of the steady state solution; the curve marked by N represents the graph of the solution of (4.11) such that $x(0) = 0$

nonhomogeneous equation (4.11) can be found by applying Proposition 4.1: we obtain $x = 2\int_0^t e^{-(t-\tau)}\,d\tau = 2 - 2e^{-t}$. Since it vanishes for $t = 0$, we set $c = x(0) = 1$. According to (4.6), the required solution writes

$$x = e^{-t} - 2e^{-t} + 2. \tag{4.12}$$

Alternatively, we can use the method of undetermined coefficients. In this way we find directly the steady state solution $x = 2$. The general integral of (4.11) takes therefore the form

$$x = e^{-t}c + 2$$

and imposing the condition $x(0) = 1$, now we find $c = -1$. Of course, the two approaches lead to the same result. The graphs of the various components of the sum (4.12) are shown in Fig. 4.1. ■

4.3 The Nonhomogeneous Scalar Equation of Order *n*

By the same procedure illustrated in Sect. 2.10, the nonhomogeneous scalar equation of order n (with constant coefficients)

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}y' + a_n y = g(t) \tag{4.13}$$

can be rewritten as a system of the form (4.1) with a matrix A in companion form, and

$$b(t) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ g(t) \end{pmatrix}.$$

Thus, (4.13) can be considered a particular case of (4.1), the function $g(t)$ playing the role of the *forcing term*. It follows that for each function $g(\cdot) \in \mathcal{PC}(I, \mathbf{R})$ and for each set of initial conditions

$$y(t_0) = y_0,\ y'(t_0) = y_1, \ldots, y^{(n-1)}(t_0) = y_{n-1} \tag{4.14}$$

($t_0 \in I$) there is a unique solution defined for $t \in I$. We emphasize that by this procedure, we are led to identify the state of the system with the vector whose components are $(y, y', \ldots, y^{(n-1)})$.

In order to determine the solutions of (4.13), the methods described in the previous sections can be applied. However, if $g(t) = p(t)e^{\gamma t}$ where $\gamma \in \mathbf{C}$ and $p(t)$ is a polynomial with real or complex coefficients, it is more convenient to work directly with (4.13). Indeed, we can write the general integral as

$$y(t) = c_1 y_1(t) + \cdots + c_n y_n(t) + \chi^*(t) \tag{4.15}$$

where $y_1(t), \ldots, y_n(t)$ are linearly independent solutions of the *associated homogeneous (or unforced) equation*

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = 0, \tag{4.16}$$

$c_1, \ldots, c_n$ are arbitrary constants, and $\chi^*(t)$ is a particular solution of (4.13). Now, the method of undetermined coefficients gives rise to the following simplified rules: a particular solution $\chi^*(t)$ can be sought of the form

1. $\chi^*(t) = q(t)e^{\gamma t}$ provided that γ is not a characteristic root of (4.16);
2. $\chi^*(t) = t^\mu q(t)e^{\gamma t}$ provided that γ is a characteristic root of (4.16) with algebraic multiplicity μ (case of resonance).

In both cases, $q(t)$ represents a polynomial of the same degree as $p(t)$. Recall that the solution $\chi^*(t)$ obtained by the method of undetermined coefficients does not coincide, in general, with the solution of (4.13) vanishing at $t = t_0$.

Remark 4.5 If all the characteristic roots have strictly negative real part, then all the solutions of the associated homogeneous system (4.16) (and all their derivatives) go to zero when $t \to +\infty$. Hence, if the forcing term $g(t)$ is constant or periodic, the particular solution $\chi^*(t)$ can be interpreted, also in this case, as the steady state solution. ■

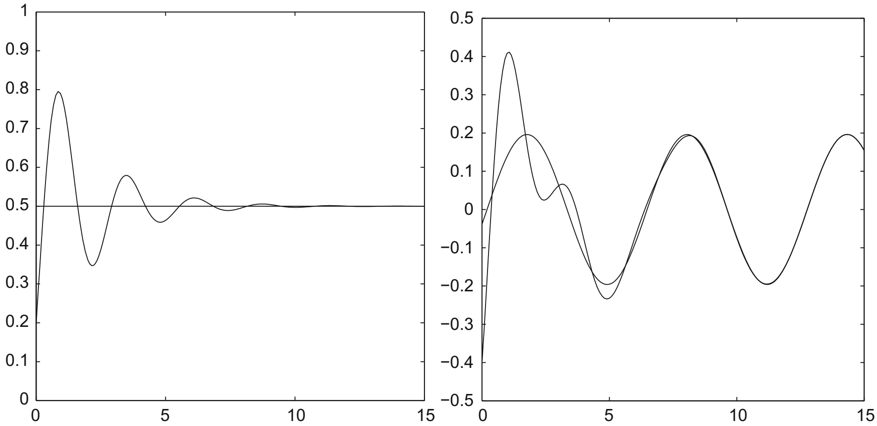


Fig. 4.2 Examples 4.3 and 4.4: steady state solution and transient

Example 4.3 The linear equation of order 2 in general form

$$y'' + ay' + by = g(t) \tag{4.17}$$

constitutes a model for a large variety of physical problems, and it is appropriate to illustrate the transient and the steady state phenomena. To this end, we assume that the characteristic polynomial of the associated homogeneous equation has a pair of complex conjugate roots $\alpha \pm \mathrm{i}\,\beta$ with $\alpha = -a/2 < 0$ and $\beta \neq 0$ (which necessarily yields $b \neq 0$).

If the forcing term is constant, say $g(t) = g_0$, the unique constant solution is $\chi^*(t) = g_0/b$. Its graph is drawn in Fig. 4.2 (left), for the case $a = 1$, $b = 6$, $g_0 = 3$. The figure shows also the graph of a solution corresponding to different initial conditions. The transient stage can be recognized in the interval where the two graphs can be clearly distinguished. ■

Example 4.4 Considered again the general second order equation (4.17) under the same assumptions about the coefficients a, b, but now with a periodic forcing term

$$g(t) = p_1 \cos \omega t + p_2 \sin \omega t, \quad p_1, p_2 \in \mathbf{R}.$$

By the same procedure of Example 4.1, we can find a particular solution of the form

$$\chi^*(t) = q_1 \cos \omega t + q_2 \sin \omega t, \quad q_1, q_2 \in \mathbf{R} \tag{4.18}$$

which can be recognized as the steady state solution. The general integral can be written as

$$y(t) = (c_1 \cos \beta t + c_2 \sin \beta t)e^{\alpha t} + q_1 \cos \omega t + q_2 \sin \omega t. \tag{4.19}$$

The transient will be shorter and shorter, as the absolute value of α becomes larger and larger.

The coefficients q_1 and q_2 in (4.18) depend on p_1, p_2 and ω (as well as on a and b) and can be computed by direct substitution.[4] Sometimes, it may be convenient to rewrite (4.18) as

$$\chi^*(t) = \rho \cos(\omega t + \theta) \tag{4.20}$$

where $q_1 = \rho \cos\theta$, $q_2 = \rho \sin\theta$. The quantities ρ and θ represent the amplitude and, respectively, the phase of the periodic function at hand (compare with Example 2.2). Also the forcing term can be rewritten in a similar way. Note that the initial conditions contribute to determine the values of c_1 and c_2 in (4.19), but not the values of q_1, q_2 (equivalently, ρ, θ) characterizing the shape of (4.18).

Consistently with Remark 4.2, we see that the frequency of the steady state solution is unchanged, when compared with the frequency of the forcing term. On the contrary, while the signal goes through the system, the phase and the amplitude may undergo a variation.

A simulation is presented in Fig. 4.2 (right), for the case $a = 1$, $b = 6$, $g_0 = \sin t$. The periodic steady state solution is $\chi^*(t) = (-\cos t + 5\sin t)/26$. ■

Example 4.5 Let us consider again the Eq. (4.17), with the same forcing term but now with $a = 0$. If $\omega^2 = b$ then $\mathrm{i}\,\omega$ is a solution of the characteristic equation. The system resonates. The form of the general integral is

$$\chi^*(t) = (c_1 + tq_1)\cos\omega t + (c_2 + tq_2)\sin\omega t. \tag{4.21}$$

The constants q_1, q_2 characterizing the particular solution can be easily determined by direct substitution. The solutions exhibit an oscillatory behavior, and the amplitude of the oscillations goes to $+\infty$ when $t \to \infty$. ■

4.4 The Laplace Transform Method

In this section we discuss a different approach to the problem of determining the solutions of (4.13), based on the Laplace transform (see Appendix B for notation and properties of the Laplace transform).

4.4.1 Transfer Function

Let us assume that the forcing term $g(\cdot)$ is defined for $t \geq 0$, and that it belongs to the set of subexponential functions of class $\mathcal{PC}([0, +\infty), \mathbf{R}^m)$. According to what

[4] Recall that $p_i = 0$ $(i = 1, 2)$ does not imply in general $q_i = 0$.

exposed in the previous sections, we know that the solution $y(t)$ is defined for $t \geq 0$, and that it is a subexponential function, as well. This justifies the use of the Laplace transform.

Let us apply the Laplace transform to both side of (4.13). Recalling (4.14) and (B.10), from

$$\mathcal{L}[y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y] = \mathcal{L}[g(t)]$$

we have

$$\begin{aligned} &(s^n + a_1 s^{n-1} + \cdots + a_n) Y(s) \\ -\big\{ &s^{n-1} y_0 + s^{n-2} y_1 + s^{n-3} y_2 + \cdots + y_{n-1} \\ &+ \big[s^{n-2} y_0 + s^{n-3} y_1 + \cdots + y_{n-2} \big] a_1 \\ &+ \big[s^{n-3} y_0 + \cdots + y_{n-3} \big] a_2 \\ &+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &+ y_0 a_{n-1} \big\} = G(s). \end{aligned}$$

We recognize that $s^n + a_1 s^{n-1} + \cdots + a_n$ is nothing else but the characteristic polynomial $p_{ch}(s)$ of the homogeneous equation (4.16) associated to (4.13). Thus we can write

$$p_{ch}(s) Y(s) - P_0(s) = G(s) \tag{4.22}$$

where

$$P_0(s) = A_0 s^{n-1} + A_1 s^{n-2} + \cdots + A_{n-1} \tag{4.23}$$

with

$$A_0 = y_0, \;\; A_1 = y_1 + a_1 y_0, \ldots, \;\; A_{n-1} = y_{n-1} + a_1 y_{n-2} + \cdots + a_{n-1} y_0.$$

We remark that:

(i) $p_{ch}(s)$ is independent of both the forcing term and the initial conditions;
(ii) $\deg P_0(s) < \deg p_{ch}(s)$;
(iii) $P_0(s)$ vanishes if and only if $y_0 = \cdots = y_{n-1} = 0$.

From (4.22) we obtain formally

$$Y(s) = \frac{P_0(s)}{p_{ch}(s)} + \frac{G(s)}{p_{ch}(s)}. \tag{4.24}$$

Formula (4.24) is well defined provided that s is not a solution of the characteristic equation $p_{ch}(s) = 0$. Since the characteristic equation has finitely many solutions, there exists a real number σ_0 such that (4.24) holds in the half plane $\{s \in \mathbf{C} : \operatorname{Re} s > \sigma_0\}$.

Formula (4.24) provides in a purely algebraic way the Laplace transform of the solution $y(t)$ corresponding to the given initial conditions. Therefore, the solution $y(t)$ can be now determined for $t \geq 0$ by applying the inverse of the Laplace transform $\mathcal{L}^{-1}$. It is convenient to set

$$\varphi(t) = \mathcal{L}^{-1}\left[\frac{P_0(s)}{p_{ch}(s)}\right] \qquad \text{and} \qquad \chi(t) = \mathcal{L}^{-1}\left[\frac{G(s)}{p_{ch}(s)}\right]$$

so that $y(t) = \varphi(t) + \chi(t)$. The following remarks point out the analogy between the structures of (4.24) and of (4.15).

Remark 4.6 The first summand of (4.24) contains the information about the initial conditions: it coincides with the solution of the homogeneous equation (4.16) associated to (4.13), with the same initial conditions. This term is a proper rational function: once it has been decomposed as a sum of partial fractions, we may easily go back to $\varphi(t)$ by means of the table of inverse Laplace transforms.

Of course, in this way we recover the well known conclusions about the structure of the form of the general integral of a linear homogeneous differential equation. Indeed, the inverse transform of the rational function $P_0(s)/p_{ch}(s)$ is given by the sum of functions of the form $Q_1(t)e^{\alpha t}\cos\beta t$ and $Q_2(t)e^{\alpha t}\sin\beta t$ where $Q_1(t)$, $Q_2(t)$ are polynomials of degree less than n, whose coefficients depend on the initial conditions. ■

Remark 4.7 The second summand of (4.24) depends on the forcing term. It coincides with the solution obtained solving (4.13) with zero initial state (instead of the conditions (4.14)). It is written as a product $H(s)G(s)$, where the function $H(s) = 1/p_{ch}(s)$ (defined on the half plane $\{s \in \mathbf{C} : \operatorname{Re} s > \sigma_0\}$) is called the *transfer function*. Let $h(t)$ be the function which coincides with the inverse Laplace transform of $H(s)$ for $t \geq 0$, and vanishes for $t < 0$. Then the solution of (4.13) corresponding to the initial conditions $y_0 = \cdots = y_{n-1} = 0$ can be represented by the formula

$$\chi(t) = \int_0^t h(t-\tau)g(\tau)\,d\tau \qquad \text{per } \ t \geq 0 \tag{4.25}$$

(recall (B.12)). In particular, if we interpret $g(t)$ as an input and we agree that it vanishes for $t < 0$, then $h(t)$ can be reviewed as the impulse response function of the system defined by (4.13).

Formula (4.25) can be considered as an extension of the variation of constants formula to the differential equation (4.13). ■

Example 4.6 We want to find the solution of the system defined by the linear differential equation of second order

$$y'' + 3y' + 2y = 1 \tag{4.26}$$

with initial conditions $y(0) = 1$, $y'(0) = 0$. We apply to both sides of (4.26) the operator $\mathcal{L}$. We have

$$\begin{aligned}\mathcal{L}\left[y''\right] + 3\mathcal{L}\left[y'\right] + 2\mathcal{L}\left[y\right] &= -y'(0) + s\mathcal{L}\left[y'\right] + 3\mathcal{L}\left[y'\right] + 2\mathcal{L}\left[y\right] \\ &= -y'(0) + (s+3)(sY(s) - y(0)) + 2Y(s) \\ &= (s^2 + 3s + 2)Y(s) - y(0)(s+3) - y'(0) \\ &= \frac{1}{s}.\end{aligned}$$

The Laplace transform of the forcing term requires the restriction $\operatorname{Re} s > 0$. In this region of the complex plane there is no solutions of the characteristic equation

$$s^2 + 3s + 2 = 0$$

which are both real and negative. So we obtain

$$\begin{aligned}Y(s) &= \frac{s+3}{s^2+3s+2} + \frac{1}{s(s^2+3s+2)} = \frac{s^2+3s+1}{s(s+2)(s+1)} \\ &= \frac{1}{2}\left(\frac{1}{s} - \frac{1}{s+2} + \frac{2}{s+1}\right).\end{aligned}$$

By applying the inverse transform $\mathcal{L}^{-1}$, we easily get

$$y(t) = \frac{1}{2}\left(1 - e^{-2t} + 2e^{-t}\right)$$

for $t \geq 0$. We recognize in this last expression the sum of a particular solution of (4.26) and a particular solution of the associated homogeneous equation. The computations above deserve some comments. In particular, we remark that $Y(s)$ was obtained as the sum of two terms: then we passed to a single rational expression and finally we performed the partial fraction decomposition. This approach is the most natural and convenient for practical purposes. However, we may also rearrange the computation in a different way. Consistently with the previous analysis (Remarks 4.6 and 4.7), we now maintain separate the term carrying the information about the initial conditions and the term carrying the information about the forcing term. We have

$$Y(s) = \left(-\frac{1}{s+2} + \frac{2}{s+1}\right) + \frac{1}{2}\left(\frac{1}{s} + \frac{1}{s+2} - \frac{2}{s+1}\right)$$

which yields

$$y(t) = \left(-e^{-2t} + 2e^{-t}\right) + \left[\frac{1}{2}\left(e^{-2t} - 2e^{-t}\right) + \frac{1}{2}\right].$$

Now it is easier to interpret the structure of $y(t)$. The first summand represents the solution corresponding to the zero input (that is, the solution of the associated homogeneous equation) and the same initial conditions. Since the roots of the characteristic polynomial are negative, this part affects only the transient.

The second summand represents the solution corresponding to zero initial conditions. In turn, it is formed by a constant term (the steady state solution) plus other terms whose effect can be appreciated only in the transient. As already mentioned, the presence of these terms is due to the need of compensating the difference between the initial data of the actual solution and the steady state solution.

When the forcing term is not constant, the problem of the factorization of a polynomial of higher degree arises, For instance, if we take an input signal $g(t) = \sin t$, we have:

$$Y(s) = \frac{s+3}{(s+2)(s+1)} + \frac{1}{(s+2)(s+1)(s^2+1)}$$
$$= \frac{s^2+3s^2+s+4}{(s+2)(s+1)(s^2+1)} = \frac{1}{10}\left(\frac{-12}{s+2} + \frac{25}{s+1} + \frac{1-3s}{s^2+1}\right)$$

and so

$$y(t) = \frac{1}{10}\left(-12e^{-2t} + 25e^{-t} + \sin t - 3\cos t\right).$$

■

4.4.2 Frequency Response Analysis

In this section we present some further developments about the study of a linear differential equation (4.13), with a periodic forcing term of the form

$$g(t) = p_1 \cos \omega t + p_2 \sin \omega t \tag{4.27}$$

under the assumption that all the solutions of the characteristic equation $p_{ch}(s) = 0$ have strictly negative real part. As well known, under these conditions the steady state solution is periodic, with the same frequency as the forcing term (4.27). One of the classical problems at the origin of system theory is the analysis of the solution (response) corresponding to a periodic forcing term (input) of this form.

The problem has been already studied in the case where the order of the equation is $n = 2$ (Example 4.4), as an application of the method of undetermined coefficients. For the general case, the method illustrated in this section, based on the Laplace transform, provides a very efficient tool which allows us to obtain further information, and in particular to determine the parameters of the system and of the forcing term in such a way that the solutions have preassigned amplitude and phase. This approach is the so-called *frequency response analysis*. Taking into account (4.27), we may rewrite (4.24) as

$$Y(s) = \frac{P_0(s)}{p_{ch}(s)} + \frac{1}{p_{ch}(s)} \frac{p_1 s + p_2 \omega}{s^2 + \omega^2}.$$

Since the solutions of the characteristic equation $p_{ch}(s) = 0$ lies in the negative complex half-plane, we have $p_{ch}(\mathrm{i}\,\omega) \neq 0$; hence $s^2 + \omega^2$ is not a divisor of $p_{ch}(s)$. We can rewrite the right-hand side as

$$Y(s) = \frac{P_0(s)}{p_{ch}(s)} + \frac{P(s)}{p_{ch}(s)} + \frac{q_1 s + q_2}{s^2 + \omega^2}$$

where $P(s)$ is a polynomial, and q_1, q_2 are constants such that

$$P(s)s^2 + q_1 p_{ch}(s)s + P(s)\omega^2 + q_2 p_{ch}(s) = p_1 s + p_2 \omega\ . \tag{4.28}$$

We already know (see Remark 4.6) that the inverse transform of the rational function $P_0(s)/p_{ch}(s)$ is the sum of functions of the form $Q_1(t)e^{\alpha t} \cos \beta t$ and $Q_2(t)e^{\alpha t} \sin \beta t$ where $Q_1(t)$, $Q_2(t)$ are polynomials of degree less than n. Our hypothesis that all the characteristic roots have negative real part implies that these terms go to zero when $t \to +\infty$. Formula (4.28) shows in particular that $\deg P < \deg p_{ch}$. Thus, the same reasoning can be repeated about the term $P(s)/p_{ch}(s)$, as well. We finally conclude that the contributions of the terms $P_0(s)/p_{ch}(s)$ and $P(s)/p_{ch}(s)$ can be ultimately neglected, and the steady state response depends essentially on the third summand $(q_1 s + q_2)/(s^2 + \omega^2)$, whose inverse transform is

$$\mathcal{L}^{-1}\left[\frac{q_1 s + q_2}{s^2 + \omega^2}\right] = q_1 \cos \omega t + \frac{q_2}{\omega} \sin \omega t = k \sin(\omega t + \theta)$$

being $q_1 = k \sin \theta$ and $q_2 = k\omega \cos \theta$. Recall that the term $P_0(s)/p_{ch}(s)$ represents the solution of the unforced system with the same initial conditions of the given system. The term $P(s)/p_{ch}(s)$ compensates the difference between the assigned initial conditions and the (in general, different) initial conditions of the steady state solution (compare these comments with those in Remark 4.4).

Finally, we show how to compute q_1 and q_2, and hence k and θ. Replacing $s = \mathrm{i}\,\omega$, from (4.28) we find

$$q_2 + \mathrm{i}\, q_1 \omega = \frac{\omega}{p_{ch}(\mathrm{i}\,\omega)}(p_2 + \mathrm{i}\, p_1),$$

which yields $q_2 = \mathrm{Re}\,[\frac{\omega}{p_{ch}(\mathrm{i}\,\omega)}(p_2 + \mathrm{i}\, p_1)]$ and $q_1 = \mathrm{Im}\,[\frac{1}{p_{ch}(\mathrm{i}\,\omega)}(p_2 + \mathrm{i}\, p_1)]$. Alternatively, we can compute p_1 and p_2 as functions of some desired values of q_1 and q_2.

Chapter Summary

This chapter constitutes a different development of Chap. 2. We consider the problem of representing the solutions of nonhomogeneous (i.e., with forcing term) systems of linear differential equations. We present the variation of constants formula and the method of undetermined coefficients. Moreover, we illustrate the qualitative notions of transient and steady state solution. Finally we present the Laplace transform method and, as an application, we discuss the frequency response analysis of a system under periodic input.

Chapter 5
Controllability and Observability of Linear Systems

In this chapter we begin to study differential systems with inputs and outputs. We focus in particular on the so-called structural properties of finite-dimensional, time invariant linear systems, that is systems of the form

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \tag{5.1}$$

where $x \in \mathbf{R}^n$ represents the state of the system, $u \in \mathbf{R}^m$ represents the input and $y \in \mathbf{R}^p$ represents the output (n, m and p arbitrary integers greater than or equal to 1). Throughout this chapter, the admissible inputs are functions $u(\cdot) \in \mathcal{PC}([0, +\infty), \mathbf{R}^m)$. Indeed, the qualifier "structural" identifies properties which depend only on the matrices A, B, C, and so are not affected by possible restrictions on the inputs variables.

5.1 The Reachable Sets

For each admissible input $u(\cdot) \in \mathcal{PC}([0, +\infty), \mathbf{R}^m)$ and for each initial state $x(0) = x_0$, there is a unique solution of the system

$$\dot{x} = Ax + Bu(t) \tag{5.2}$$

denoted by $x(t, x_0, u(\cdot))$, and defined for $t \geq 0$. System (5.2) can be thought of as a linear nonhomogeneous system with forcing term $b(t) = Bu(t)$. Hence, the variation of constants formula applies and we can represent the solution as

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_5

$$x(t, x_0, u(\cdot)) = e^{tA}\left(x_0 + \int_0^t e^{-\tau A} Bu(\tau)\, d\tau\right) . \tag{5.3}$$

As already mentioned (Chap. 4), it is natural to think of (5.3) as the sum of

$$x(t, x_0, 0) = e^{tA} x_0 \tag{5.4}$$

also called the *free* (or *unforced*) *solution*, and

$$x(t, 0, u(\cdot)) = \int_0^t e^{(t-\tau)A} Bu(\tau) d\tau . \tag{5.5}$$

We emphasize that (5.4) represents the solution corresponding to the input $u = 0$, while (5.5) represents the solution corresponding to the actual input but with zeroed initial state. To this respect, there is an analogue of Proposition A.1.

Proposition 5.1 *For each pair of real numbers* $t, \tau \in [0, +\infty)$*, for each admissible input* $u(\cdot) : [0, +\infty) \to \mathbf{R}^m$*, and for each initial state* x_0*, we have*

$$x(0, x_0, u(\cdot)) = x_0$$

and

$$x(t + \tau, x_0, u(\cdot)) = x(t, x(\tau, x_0, u(\cdot)), w(\cdot))$$

where we set $w(t) = u(t + \tau)$ *for* $t \in [0, +\infty)$. ■

We now introduce the first important notion of this chapter.

Definition 5.1 Let $x_0, \eta_0 \in \mathbf{R}^n$. We say that η_0 is *reachable* from x_0 at time $T > 0$ (or also that x_0 is *controllable* to η_0 at time T) if there exists an admissible input $u(\cdot) : [0, T] \to \mathbf{R}^m$ such that

$$\eta_0 = x(T, x_0, u(\cdot)) . \tag{5.6}$$

For fixed x_0 and T, the set of points reachable from x_0 at time T is denoted by $\mathrm{R}(T, x_0)$ and it is called the *reachable set*.

Intuitively, the "size" of the set $\mathrm{R}(T, x_0)$ provides a measure of our ability to control the performances of the system. We are in particular interested in the following definitions.

Definition 5.2 A system of the form (5.1) is said to be:

- *globally reachable* from x_0 at time T if $\mathrm{R}(T, x_0) = \mathbf{R}^n$;
- *globally reachable* at time T if $\mathrm{R}(T, x_0) = \mathbf{R}^n$ for each x_0.

In fact, the two notions introduced in the previous definition are equivalent.

Proposition 5.2 *If there exists a state x_0 such that the system is globally reachable from x_0 at time T, then the system is globally reachable at time T from the origin, as well. If the system is globally reachable from the origin at time T, then it is globally reachable at time T.*

Proof Assume that there exists a point x_0 such that the system is globally reachable at time T from x_0, and let η_0 be an arbitrary point of $\mathbf{R}^n$. Let $\bar{\eta} = \eta_0 + e^{TA}x_0$. By assumption, there exists an input function $u(\cdot)$ such that

$$\bar{\eta} = \eta_0 + e^{TA}x_0 = e^{TA}x_0 + \int_0^T e^{(T-\tau)A}Bu(\tau)\,d\tau\,.$$

This yields

$$\eta_0 = \int_0^T e^{(T-\tau)A}Bu(\tau)\,d\tau\,,$$

meaning that η_0 is reachable from the origin at time T.

Vice versa, assume that the system is globally reachable at time T from the origin. Let x_0 and η_0 be two arbitrary points of $\mathbf{R}^n$. Setting $\bar{\eta} = \eta_0 - e^{TA}x_0$, we can find an input function $u(\cdot)$ such that

$$\bar{\eta} = \eta_0 - e^{TA}x_0 = \int_0^T e^{(T-\tau)A}Bu(\tau)\,d\tau$$

that is

$$\eta_0 = e^{TA}x_0 + \int_0^T e^{(T-\tau)A}Bu(\tau)\,d\tau$$

and this means that η_0 is reachable from x_0 at time T. ■

Remark 5.1 Analogously, we may fix η_0 and T and then we may consider the set of points x_0 for which there exists an admissible input $u(\cdot) : [0, T] \to \mathbf{R}^m$ such that (5.6) holds. This is called the *controllable set* and it is denoted by $\mathrm{C}(T, \eta_0)$. Clearly, $\mathrm{C}(T, \eta_0)$ is nothing else that $\mathrm{R}(T, \eta_0)$ for the reversed time system, obtained replacing A, B by $-A$, $-B$ in (5.1). Indeed, multiplying by e^{-TA} both sides of the equality

$$\eta_0 = e^{TA}x_0 + \int_0^T e^{(T-\tau)A}Bu(\tau)\,d\tau$$

and transforming the integral by the substitution $\tau = T - \theta$, we get

$$x_0 = e^{T(-A)}\eta_0 + \int_0^T e^{(T-\theta)(-A)}(-B)u(T-\theta)\,d\theta\,.$$

■

5.1.1 Structure of the Reachable Sets

The reachability property introduced in the previous section involves uniquely the input variables and the state variables. It does not depend on the matrix C, and therefore it is natural to argue that it can be characterized only in terms of the matrices A and B. Moreover, Proposition 5.2 suggests that our attention can be focused on the set of points reachable from the origin.

Theorem 5.1 *Let the linear system (5.1) be given. The map which associates to each* $u(\cdot) \in \mathcal{PC}([0, +\infty), \mathbf{R}^m)$ *the function*

$$t \mapsto x(t, 0, u(\cdot)) = \int_0^t e^{(t-\tau)A} Bu(\tau)\, d\tau \;\in\; \mathcal{C}([0, +\infty), \mathbf{R}^n) \tag{5.7}$$

is linear.

Proof If $u_1(\cdot), u_2(\cdot) \in \mathcal{PC}([0, +\infty), \mathbf{R}^m)$ and $\alpha, \beta \in \mathbf{R}$, then also $\alpha u_1(\cdot) + \beta u_2(\cdot) \in \mathcal{PC}([0, +\infty), \mathbf{R}^m)$ and, according to the basic properties of the integral,

$$\int_0^t e^{(t-\tau)A} B(\alpha u_1(\tau) + \beta u_2(\tau))\, d\tau$$
$$= \alpha \int_0^t e^{(t-\tau)A} Bu_1(\tau)\, d\tau + \beta \int_0^t e^{(t-\tau)A} Bu_2(\tau)\, d\tau\,.$$

■

Fix now $T > 0$. We can reinterpret (5.7) as a map Λ which associates to each input function $u(\cdot) \in \mathcal{PC}([0, T], \mathbf{R}^m)$ the element of $\mathbf{R}^n$

$$x = \Lambda(u(\cdot)) = x(T, 0, u(\cdot)) = \int_0^T e^{(T-\tau)A} Bu(\tau)\, d\tau \in \mathbf{R}^n\,. \tag{5.8}$$

Corollary 5.1 *The map* $\Lambda : \mathcal{PC}([0, T], \mathbf{R}^m) \to \mathbf{R}^n$ *is linear.*

Corollary 5.2 *For each fixed* $T > 0$*, the set* $\mathrm{R}(T, 0)$ *is a linear subspace of* $\mathbf{R}^n$*. For each* $T > 0$ *and each* $x_0 \neq 0$*, the set* $\mathrm{R}(T, x_0)$ *is a linear manifold of* $\mathbf{R}^n$*.*

Proof For each fixed $T > 0$, the set $\mathrm{R}(T, 0)$ coincides with the image of the operator Λ and hence it is a linear subspace of $\mathbf{R}^n$. As far as the second statement is concerned, it is sufficient to remark that $\mathrm{R}(T, x_0)$ is the translation of $\mathrm{R}(T, 0)$ by means of the vector $v = e^{TA}x_0$. ■

According to these conclusions, it is natural to assume as a measure of the "size" of the set $\mathrm{R}(T, x_0)$ the dimension of $\mathrm{R}(T, x_0)$ as a linear manifold of $\mathbf{R}^n$. Moreover, $\mathrm{R}(T, 0)$ will be often referred to as the *reachable space*.

Corollary 5.3 *System (5.1) is globally reachable at time T if and only if* $\mathrm{R}(T, 0) = \mathbf{R}^n$*, that is if and only if the dimension of* $\mathrm{R}(T, 0)$ *is maximal.*

5.1.2 The Input-Output Map

The previous results enable us to prove Proposition 1.10. From (5.3), given any admissible input $u(\cdot) : [0, +\infty) \to \mathbf{R}^m$ and any initial state x_0, the following representation for the output function of system (5.1)

$$y(t, x_0, u(\cdot)) = Cx(t, x_0, u(\cdot)) = Ce^{tA}\big(x_0 + \int_0^t e^{-\tau A} Bu(\tau)\, d\tau\big) \tag{5.9}$$

can be readily deduced. Of course, $y(0, x_0, u(\cdot)) = Cx_0$ and

$$y(t, x_0, u(\cdot)) = y(t, x_0, 0) + y(t, 0, u(\cdot))\,. \tag{5.10}$$

Proof of Proposition 1.10 The map which associates to each $u(\cdot) \in \mathcal{PC}([0, +\infty), \mathbf{R}^m)$ the function $x(t, 0, u(\cdot))$ is linear, by virtue of Theorem 5.1. Hence, the map which associates to $u(\cdot)$ the function $y(t, 0, u(\cdot)) = Cx(t, 0, u(\cdot))$ is linear, as well. On the other hand, also the map which associates to x_0 the function $y(t, x_0, 0) = Ce^{tA}x_0$ is linear. To finish, it is sufficient to take into account (5.10) and the fact that if $f_1 : V_1 \to W$, $f_2 : V_2 \to W$ are linear maps, then $f_1 + f_2 : V_1 \times V_2 \to W$ is a linear map. ■

5.1.3 Solution of the Reachability Problem

Next theorem provides a first necessary and sufficient condition for the global reachability of a linear system.

Theorem 5.2 *System (5.1) is globally reachable at time $T > 0$ if and only if the matrix*

$$\Gamma(T) = \int_0^T e^{-\tau A} BB^{\mathbf{t}} e^{-\tau A^{\mathbf{t}}}\, d\tau$$

is nonsingular.

Proof First we show that if $\Gamma(T)$ is nonsingular, then for each pair of states $x_0, \eta_0 \in \mathbf{R}^n$ there exists an input function $u(\cdot)$ for which (5.6) holds. Let, for $\tau \in [0, T]$,

$$u(\tau) = -B^{\mathbf{t}} e^{-\tau A^{\mathbf{t}}} \Gamma^{-1}(T)[x_0 - e^{-TA}\eta_0] \tag{5.11}$$

and compute

$$
\begin{aligned}
e^{TA}x_0 + \int_0^T e^{(T-\tau)A} Bu(\tau)\, d\tau \qquad (5.12)\\
= e^{TA}x_0 - e^{TA}\left[\int_0^T e^{-\tau A} BB^{\mathbf{t}} e^{-\tau A^{\mathbf{t}}}\, d\tau\right]\Gamma^{-1}(T)[x_0 - e^{-TA}\eta_0]\\
= e^{TA}x_0 - e^{TA}\Gamma(T)\Gamma^{-1}(T)[x_0 - e^{-TA}\eta_0] = e^{TA}x_0 - e^{TA}x_0 + \eta_0\,.
\end{aligned}
$$

In conclusion,

$$e^{TA}x_0 + \int_0^T e^{(T-\tau)A} Bu(\tau)\, d\tau = \eta_0\,.$$

In order to prove the converse, we need some preliminary remarks. Clearly, $\Gamma(T)$ is symmetric, and the quadratic form

$$\xi^{\mathbf{t}}\Gamma(T)\xi = \int_0^T ||B^{\mathbf{t}}e^{-\tau A^{\mathbf{t}}}\xi||^2\, d\tau \qquad (5.13)$$

is, in general, positive semidefinite. If $\Gamma(T)$ is singular, then there exists a point $x_0 \in \mathbf{R}^n$ ($x_0 \neq 0$) such that ${x_0}^{\mathbf{t}}\Gamma(T)x_0 = 0$. Therefore, taking into account (5.13), we have $B^{\mathbf{t}}e^{-\tau A^{\mathbf{t}}}x_0 = 0$, identically for $\tau \in [0, T]$. The global reachability assumption implies that starting from x_0 it is possible to reach the origin at time T. This yields

$$e^{TA}x_0 = -\int_0^T e^{(T-\tau)A} Bu(\tau)\, d\tau \qquad (5.14)$$

for some admissible input $u(\cdot)$. From (5.14) it follows

$$x_0 = -\int_0^T e^{-\tau A} Bu(\tau)\, d\tau$$

and so

$$
\begin{aligned}
||x_0||^2 = {x_0}^{\mathbf{t}}x_0 = -\left(\int_0^T e^{-\tau A} Bu(\tau)\, d\tau\right)^{\mathbf{t}} x_0\\
= -\int_0^T u^{\mathbf{t}}(\tau)B^{\mathbf{t}}e^{-\tau A^{\mathbf{t}}}x_0\, d\tau = 0\,.
\end{aligned}
$$

This contradicts the assumption $x_0 \neq 0$. ■

Remark 5.2 Formula (5.11) provides an answer to the problem of determining a control function which allows us to steer the system from the state x_0 to the state η_0. ■

5.1.4 The Controllability Matrix

The condition stated in Theorem 5.2 is useful for theoretical developments, but not easy to apply in practice. From this point of view, the criterion we are going to present in this section is more convenient, since it amounts to purely algebraic computations involving only the matrices A and B which define the system.

Theorem 5.3 *For a system of the form (5.1), the set* $\mathrm{R}(T, 0)$ *is independent of* T. *Moreover, for each* $T > 0$ *we have*

$$\mathrm{R}(T, 0) = V \tag{5.15}$$

where

$$V = \mathrm{span}\left\{b_1, \dots, b_m, Ab_1, \dots, Ab_m, \dots, A^{n-1}b_1, \dots, A^{n-1}b_m\right\} \tag{5.16}$$

and $b_1, \dots, b_m$ *denote the columns of* B.

Proof Since both sides of (5.15) are subspaces of $\mathbf{R}^n$, it is sufficient to prove that the respective orthogonal spaces coincide. First we prove that $V \subset \mathrm{R}(T, 0)$. Let $\mu \neq 0$ be a vector orthogonal to $\mathrm{R}(T, 0)$. Then we have, for each admissible input,

$$\begin{aligned} 0 &= \mu^{\mathbf{t}} \int_0^T e^{(T-\tau)A} Bu(\tau)\, d\tau \\ &= \int_0^T \mu^{\mathbf{t}} e^{(T-\tau)A} Bu(\tau)\, d\tau = \int_0^T \mu^{\mathbf{t}} e^{\theta A} Bu(T-\theta)\, d\theta\,. \end{aligned} \tag{5.17}$$

Taking into account (5.17), now we show that

$$\mu^{\mathbf{t}} e^{\theta A} Bu = 0 \tag{5.18}$$

for each $\theta \in (0, T)$ and each $u \in \mathbf{R}^m$. Assume that this is false. Then we can find $\bar{\theta} \in (0, T)$ and $\bar{u} \in \mathbf{R}^m$ such that $\mu^{\mathbf{t}} e^{\bar{\theta} A} B\bar{u} \neq 0$ (say for instance, $\mu^{\mathbf{t}} e^{\bar{\theta} A} B\bar{u} > 0$). Then, by continuity, there exists $\delta > 0$ such that $(\bar{\theta} - \delta, \bar{\theta} + \delta) \subset (0, T)$ and the function

$$\theta \mapsto \mu^{\mathbf{t}} e^{\theta A} B\bar{u}$$

takes positive values for $\bar{\theta} - \delta < \theta < \bar{\theta} + \delta$. Setting $\bar{\tau} = T - \bar{\theta}$, we can therefore define

$$u(\tau) = \begin{cases} \bar{u} & \text{for } \bar{\tau} - \delta < \tau < \bar{\tau} + \delta \\ 0 & \text{otherwise}\,. \end{cases} \tag{5.19}$$

Then,

$$u(T-\theta) = \begin{cases} \bar{u} & \text{for } \bar{\theta} - \delta < \theta < \bar{\theta} + \delta \\ 0 & \text{otherwise}\,. \end{cases} \tag{5.20}$$

This yields

$$\int_0^T \mu^{\mathbf{t}} e^{\theta A} B u(T-\theta)\, d\theta = \int_{\bar{\theta}-\delta}^{\bar{\theta}+\delta} \mu^{\mathbf{t}} e^{\theta A} B \bar{u}\, d\theta > 0$$

and we have a contradiction to (5.17). Hence, (5.18) is true. Taking the limit for $\theta \to 0^+$, we get

$$\mu^{\mathbf{t}} B u = 0 \quad \forall u \in \mathbf{R}^m .$$

Choosing respectively $u = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, u = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$, this last expression indicates that μ is orthogonal to $b_1, \ldots, b_m$. Moreover, taking the derivative of (5.18) with respect to θ, we have

$$\mu^{\mathbf{t}} e^{\theta A} A B u = 0 \quad \forall \theta \in (0, T) \quad \text{and} \quad \forall u \in \mathbf{R}^m$$

which in turn implies, for $\theta \to 0^+$,

$$\mu^{\mathbf{t}} A B u = 0 .$$

Repeating the previous reasoning, we see that μ is orthogonal to the vectors $Ab_1, \ldots, Ab_m$, as well. The procedure can be iterated, until the conclusion is achieved.

Now we prove the opposite inclusion. Let μ be orthogonal to

$$b_1, \ldots, b_m, Ab_1, \ldots, Ab_m, \ldots, A^{n-1}b_1, \ldots, A^{n-1}b_m .$$

For each $u \in \mathbf{R}^m$, we have

$$\mu^{\mathbf{t}} B u = \cdots = \mu^{\mathbf{t}} A^{n-1} B u = 0 .$$

Moreover

$$\begin{aligned} \mu^{\mathbf{t}} e^{\theta A} B u &= \mu^{\mathbf{t}} \sum_{i=0}^{n-1} \frac{\theta^i A^i}{i!} B u + \mu^{\mathbf{t}} \sum_{i=n}^{\infty} \frac{\theta^i A^i}{i!} B u \\ &= \sum_{i=0}^{n-1} \frac{\theta^i}{i!} \mu^{\mathbf{t}} A^i B u + \sum_{i=n}^{\infty} \frac{\theta^i}{i!} \mu^{\mathbf{t}} A^i B u . \end{aligned}$$

Clearly, the terms of the first sum vanish. But also the terms of the second sum vanish since, by Cayley-Hamilton Theorem, for each $i \geq n$, the vector $A^i B u$ is a linear combination of the vectors $A^i Bu$ with $i < n$. In conclusion, $\mu^{\mathbf{t}} e^{\theta A} Bu = 0$, for all $\theta \in [0, T]$ and all $u \in \mathbf{R}^m$. But then also

$$\mu^{\mathbf{t}} \int_0^T e^{(T-\tau)A} Bu(\tau) d\tau = 0$$

for every $u(\cdot) \in \mathcal{PC}([0, T], \mathbf{R}^m)$. The theorem is finally proved. ■

Definition 5.3 System (5.1) is said to be *completely controllable* when

$$\text{rank } (B|AB|\dots|A^{n-1}B) = n \tag{5.21}$$

where $(B|AB|\dots|A^{n-1}B)$ is the matrix with n rows and nm columns formed by the columns of the matrices $B, AB, \dots, A^{n-1}B$.

The matrix $(B|AB|\dots|A^{n-1}B)$ is called the *controllability matrix* of system (5.1). The following corollary is a straightforward consequence of Theorem 5.3.

Corollary 5.4 *System (5.1) is completely controllable if and only if it is globally controllable for some (and hence for each)* $T > 0$.

Remark 5.3 The vectors $v_1, \dots, v_n$ of $\mathbf{R}^n$ form a linearly independent set if and only if

$$\det (v_1|\dots|v_n) \neq 0 .$$

Since the determinant depends continuously on the entries of the matrix, replacing the vectors $v_1, \dots, v_n$ by some other vectors $\tilde{v}_1, \dots, \tilde{v}_n$ such that $\tilde{v}_k$ is sufficiently close to v_k (for every $k = 1, \dots, n$), then also the vectors $\tilde{v}_1, \dots, \tilde{v}_n$ form a linearly independent set.

From this remark it follows that if system (5.1) is completely controllable and if the matrices $\tilde{A}$, $\tilde{B}$ are sufficiently close to, respectively, A and B, then the system defined by the matrices $\tilde{A}$, $\tilde{B}$ is completely controllable, as well. It is also clear that if system (5.1) is not completely controllable, then there exist pairs of matrices $\tilde{A}$, $\tilde{B}$ arbitrarily close to A, B, such that the system defined by $\tilde{A}$, $\tilde{B}$ is completely controllable. In other words, we can say that "generically", any linear system is completely controllable, in the sense that:

- the complete controllability property is preserved under arbitrary small perturbations of the coefficients;
- the complete controllability property can be achieved by means of suitable small perturbations of the coefficients.

These considerations can be also resumed by saying that complete controllability is an open-dense property. ■

Remark 5.4 If the input of system (5.1) is scalar i.e., $m = 1$, matrix B reduces to a single column b and the controllability matrix is square. Checking the complete controllability condition reduces to compute the determinant of the controllability matrix. It is easily seen that such a system is completely controllable if and only if b is cyclic for A (see Sect. 2.10).

Consider in particular a SISO system defined by a linear nonhomogeneous equation of order n

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = u(t) \tag{5.22}$$

where the forcing term plays the role of a (scalar) input, and y is reviewed as a (scalar) output. According to the procedure illustrated in Sects. 2.10 and 4.3, (5.22) can be rewritten in the form (5.1) with A a companion matrix,

$$b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

and C reduced to the row $(1\ 0 \ldots 0)$. The state variable coincides with the vector $x = (y, y', \ldots, y^{(n-1)})$.

We can easily check that every system of the form (5.22) is completely controllable regardless the choice of the coefficients $a_1, \ldots, a_n$. ■

In force of the conclusions of Theorem 5.3, we can slightly simplify our notation: from now on, we write R instead of $\mathrm{R}(T, 0)$.

5.1.5 Hautus' Criterion

Conditions equivalent to complete controllability of the system (5.1) can be given in several different forms. In this section we present a criterion which will be sometimes recalled in our future developments.

Theorem 5.4 (Hautus' criterion) *System (5.1) is completely controllable if and only if*

$$\forall \lambda \in \mathbf{C}, \qquad \text{rank}\,(A - \lambda I | B) = n\,. \tag{5.23}$$

We remark that (5.23) is trivially fulfilled if λ is not an eigenvalue of A. Note also that in general, $(A - \lambda I | B)$ is a matrix with complex entries. In order to prove the theorem, it is therefore advisable to interpret also A and B as operators acting on complex spaces.

Definition 5.4 A subspace V of $\mathbf{C}^n$ is called a *(complex) algebraic invariant* for A if $AV \subseteq V$.

Lemma 5.1 *If a subspace V is an algebraic invariant for A, then there exists an eigenvector $v \neq 0$ of A such that $v \in V$.*

Proof Let $q = \dim V$ and let $v_1, \ldots, v_n$ be a basis of $\mathbf{C}^n$, such that its first q elements $v_1, \ldots, v_q$ constitute a basis of V. With respect to this basis, A takes the form

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

by virtue of the invariance assumption. The operator acting from V to V and defined by the matrix A_{11} will necessarily have at least one eigenvector $v \in V$. It is not difficult to check that the same vector v, reinterpreted as an element of $\mathbf{C}^n$, is an eigenvector of A corresponding to the same eigenvalue λ. ■

Proof of Theorem 5.4 We show that (5.21) implies (5.23). Assume by contradiction that for some $\lambda \in \mathbf{C}$, the n rows of matrix $(A - \lambda I|B)$ are linearly independent. Then, there exists a vector $\eta \in \mathbf{C}^n$ ($\eta \neq 0$) such that

$$\eta^{\mathbf{t}} A = \lambda \eta^{\mathbf{t}} \quad \text{and} \quad \eta^{\mathbf{t}} B = 0\,.$$

In particular, the function

$$\varphi(t) = e^{\lambda t} \eta^{\mathbf{t}} B = (e^{\lambda t} \eta)^{\mathbf{t}} B = \eta^{\mathbf{t}} B + \lambda t \eta^{\mathbf{t}} B + \frac{\lambda^2 t^2}{2} \eta^{\mathbf{t}} B + \cdots$$

must vanish. Thus we will have

$$\varphi(0) = \varphi'(0) = \varphi''(0) = \cdots = 0\,.$$

By applying the theorem about the derivative of a power series, and taking into account

$$\eta^{\mathbf{t}} A = \lambda \eta^{\mathbf{t}} \Longrightarrow \eta^{\mathbf{t}} A^2 = \lambda \eta^{\mathbf{t}} A = \lambda^2 \eta^{\mathbf{t}} \quad \text{etc.}$$

we finally obtain

$$\eta^{\mathbf{t}} B = \eta^{\mathbf{t}} A B = \eta^{\mathbf{t}} A^2 B = \cdots = 0\,.$$

This implies that the n rows of matrix $(B|AB|\ldots|A^{n-1}B)$ are linearly dependent, so that its rank is not equal to n.

Finally, we show that (5.23) implies (5.21). According to Cayley-Hamilton Theorem, if (5.21) is false then the rows of all the matrices of the form $A^j B$ ($j = 0, 1, \ldots$) will belong to a same proper subspace of $\mathbf{C}^n$. In other words, we could find a vector $v \neq 0$ such that

$$v^{\mathbf{t}} B = v^{\mathbf{t}} A B = v^{\mathbf{t}} A^2 B = \cdots = 0\,. \tag{5.24}$$

Setting $w = A^{\mathbf{t}} v$, we have

$$w^{\mathbf{t}} B = (A^{\mathbf{t}} v)^{\mathbf{t}} B = v^{\mathbf{t}} A B = 0, \quad w^{\mathbf{t}} A B = (A^{\mathbf{t}} v)^{\mathbf{t}} A B = v^{\mathbf{t}} A^2 B = 0, \quad \text{etc.}$$

Denoted by V the subspace of $\mathbf{C}^n$ constituted by all vectors v for which (5.24) holds, we have so proved that if $v \in V$, also $w = A^{\mathbf{t}} v \in V$, and this in turn implies that V is an algebraic invariant with respect to the linear operator associated to matrix $A^{\mathbf{t}}$. Then by Lemma 5.1, it must exists a nonzero vector $\eta \in V$ and a complex number $\lambda \in \mathbf{C}$ such that

$$A^{\mathbf{t}}\eta = \lambda\eta \quad \text{that is} \quad \eta^{\mathbf{t}} A = \lambda\eta^{\mathbf{t}} .$$

As a consequence of the definition of V, we have in particular that $\eta^{\mathbf{t}} B = 0$. In conclusion, the rows of $(A - \lambda I | B)$ are linearly independent and (5.23) does not hold. ■

5.2 Observability

In common applications, the output variable does not coincide with the state variable. In these cases, the observability function plays an essential role.

Definition 5.5 We say that two points $x_0, \eta_0 \in \mathbf{R}^n$ are *indistinguishable* at time T if for each admissible input $u(\cdot) : [0, T] \to \mathbf{R}^n$ one has

$$y(t, x_0, u(\cdot)) = y(t, \eta_0, u(\cdot)) \qquad \forall t \in [0, T] .$$

The previous definition is inspired by the following idea: for each fixed input function $u(\cdot)$, if the initial state x_0 is replaced by η_0, then the system response remains unchanged. In other words, it is not possible in general to reconstruct exactly the initial state on the base of information obtained uniquely by monitoring the output corresponding to a known input.

Example 5.1 Consider the system

$$\begin{cases} \dot{x}_1 = x_1 + u \\ \dot{x}_2 = x_2 \end{cases}$$

with $y = x_1 - x_2$. The solution corresponding to an initial state of the form (a, a) is easily found:

$$x_1 = e^t \left(a + \int_0^t e^{-\tau} u(\tau) d\tau \right), \quad x_2 = a e^t .$$

Hence, we see that $y(t) = e^t \int_0^t e^{-\tau} u(\tau) d\tau$ is independent of a. In other words, two distinct arbitrary points on the line $x_1 = x_2$ are indistinguishable. ■

Let us emphasize that in practical applications, the knowledge of the initial state is an important issue. Assume that we have a physical system, and that a mathematical model has been constructed. In principle, the mathematical model should be used to simulate the evolution of the physical system and to predict the future behavior. To this end, we need to integrate analytically or numerically the system equations. But this is impossible, if we do not known how to set the initial state of the model, which should be the same as the initial state of the physical system.

5.2.1 The Unobservability Space

Our aim now is to characterize those systems for which there exist no pairs of indistinguishable points. First of all, we remark that in Definition 5.5, the role of the input function is unessential, in the sense explained by the following proposition.

Proposition 5.3 *The points x_0 and η_0 are indistinguishable at time T for the system (5.1) if and only if they are indistinguishable at time T for the unforced system*

$$\begin{cases} \dot{x} = Ax \\ y = Cx\,. \end{cases} \tag{5.25}$$

Proof If x_0 and η_0 are indistinguishable at time T for the system (5.1), then for each $u(\cdot) \in \mathcal{PC}([0,T], \mathbf{R}^m)$ and each $t \in [0,T]$ we have

$$y(t, x_0, u(\cdot)) = y(t, \eta_0, u(\cdot))$$

that is

$$C[e^{tA}(x_0 + \int_0^t e^{-\tau A} Bu(\tau)\,d\tau)] = C[e^{tA}(\eta_0 + \int_0^t e^{-\tau A} Bu(\tau)\,d\tau)]\,.$$

Getting rid of the common term, we obtain the identity

$$Ce^{tA}x_0 = Ce^{tA}\eta_0$$

for each $t \in [0,T]$. This actually means that x_0 and η_0 are indistinguishable with respect to the system (5.25). The reverse argument proves the vice versa. ■

Next proposition points out that in order to characterize the set of points which are indistinguishable from a fixed $x \in \mathbf{R}^n$, it is sufficient to characterize the set of points which are indistinguishable from the origin.

Proposition 5.4 *If x_0, η_0 are indistinguishable at the time $T > 0$, then $\xi = x_0 - \eta_0$ is indistinguishable from the origin at time T. Vice versa, if ξ is indistinguishable from the origin at the time T and if x_0 is any vector of $\mathbf{R}^n$, then x_0 and $\eta_0 = x_0 + \xi$ are indistinguishable each other at time T.*

Proof By virtue of Proposition 5.3, we can refer to system (5.25). From the assumption that x_0 and η_0 are indistinguishable at time T, we deduce that

$$Ce^{tA}x_0 = Ce^{tA}\eta_0$$

for each $t \in [0, T]$. This last equality rewrites

$$Ce^{tA}(\eta_0 - x_0) = 0 = Ce^{tA}0$$

for each $t \in [0, T]$. The statement follows, setting $\xi = x_0 - \eta_0$. Vice versa, if ξ is indistinguishable from the origin (which means that $Ce^{tA}\xi = 0$ for each $t \in [0, T]$), for each $x_0 \in \mathbf{R}^n$ the equality

$$Ce^{tA}(\xi + x_0) = Ce^{tA}\xi + Ce^{tA}x_0 = Ce^{tA}x_0$$

holds for each $t \in [0, T]$. This means that $\xi + x_0$ and x_0 are indistinguishable at time T for system (5.25), and so also for system (5.1). ■

We denote by $\mathrm{N}(T, 0)$ the set of the states indistinguishable from the origin at time T. A first characterization of $\mathrm{N}(T, 0)$ is provided by the following Theorem.

Theorem 5.5 *The following statements are equivalent.*

(i) ξ *is indistinguishable from the origin at time* T *for system (5.1);*
(ii) $\xi \in \ker\ Ce^{tA}$, $\forall t \in [0, T]$*;*
(iii) the output function of system (5.1) corresponding to the input $u(t) = 0$ *for each* $t \in [0, T)$ *and to the initial state* ξ*, vanishes on* $[0, T]$.

Proof The equivalence between *(i)* and *(ii)* follows from Proposition 5.3. The equivalence between *(ii)* and *(iii)* is straightforward. ■

Theorem 5.5 *(ii)* implies in particular that the set $\mathrm{N}(T, 0)$ coincides with

$$\bigcap_{t\in[0,T]} \ker Ce^{tA} . \tag{5.26}$$

But (5.26) is a subspace of $\mathbf{R}^n$. Hence $\mathrm{N}(T, 0)$ is a subspace of $\mathbf{R}^n$. It is called the *unobservability space*.

5.2.2 The Observability Matrix

Now, consider the matrices

$$C^{\mathbf{t}}, A^{\mathbf{t}}C^{\mathbf{t}}, \ldots, (A^{\mathbf{t}})^{n-1}C^{\mathbf{t}} .$$

Their columns can be interpreted as vectors of $\mathbf{R}^n$. Let $V \subset \mathbf{R}^n$ be the space engendered by these vectors.

Theorem 5.6 *Given the system (5.1), the unobservability space is independent of T. In fact, for each $T > 0$ one has* $\mathrm{N}(T, 0) = V^{\perp}$.

Proof We limit ourselves to sketch the main steps, since the proof is similar to that of Theorem 5.3. Let $v \in \mathrm{N}(T, 0)$. Then for all $\theta \in [0, T]$ and all $\mu \in \mathbf{R}^p$

$$(Ce^{\theta A}v)^{\mathbf{t}}\mu = 0$$

or

$$v^{\mathbf{t}}e^{\theta A^{\mathbf{t}}}C^{\mathbf{t}}\mu = 0\,.$$

For $\theta = 0$, we find $v^{\mathbf{t}}C^{\mathbf{t}}\mu = 0$ and, being μ arbitrary, v is orthogonal to all the columns of $C^{\mathbf{t}}$. Next step is to compute iteratively the derivatives of any order with respect to θ. Each derivative is evaluated at $\theta = 0$. Vice versa, if $v \in V^{\perp}$, then for each $\mu \in \mathbf{R}^p$

$$v^{\mathbf{t}}C^{\mathbf{t}}\mu = \cdots = v^{\mathbf{t}}(A^{\mathbf{t}})^{n-1}C^{\mathbf{t}}\mu = 0\,.$$

Making use, as in Theorem 5.3, of the series expansion of the exponential and of Cayley-Hamilton Theorem, this implies in turn that $v^{\mathbf{t}}e^{\theta A^{\mathbf{t}}}C^{\mathbf{t}}\mu = 0$ for each θ. Finally,

$$(Ce^{\theta A}v)^{\mathbf{t}}\mu = 0 \qquad \forall\theta \in \mathbf{R}$$

which implies $Ce^{\theta A}v = 0$, for each $\theta \in \mathbf{R}$. ■

Matrix $(C^{\mathbf{t}}|A^{\mathbf{t}}C^{\mathbf{t}}|\ldots|(A^{\mathbf{t}})^{n-1}C^{\mathbf{t}})$ is called the *observability matrix* of system (5.1). From now on, since $\mathrm{N}(T, 0)$ is independent of T, we write simply N.

Definition 5.6 The system is said to be *completely observable* when

$$\operatorname{rank}\,(C^{\mathbf{t}}|A^{\mathbf{t}}C^{\mathbf{t}}|\ldots|(A^{\mathbf{t}})^{n-1}C^{\mathbf{t}}) = n\,. \tag{5.27}$$

Corollary 5.5 *System (5.1) is completely observable if and only if for each pair of indistinguishable states x_0, η_0 we have $x_0 = \eta_0$, or, equivalently, when* $\mathrm{N} = \{0\}$.

Remark 5.5 Neither *(ii)* of Theorem 5.5 nor (5.27) depend on matrix B. This is not surprising, if we have in mind Proposition 5.3. ■

Remark 5.6 Since the unobservability space N does not depend on T, we may say that a point $x_0 \in \mathrm{N}$ if and only if $Ce^{\theta A}x_0 = 0$ for each $\theta \geq 0$. On the other hand, from Theorem 5.6 it follows that if we replace A by $-A$, the space N does not change. Hence, the previous statement can be strengthened, writing that $x_0 \in \mathrm{N}$ if and only if $Ce^{\theta A}x_0 = 0$ for each $\theta \in \mathbf{R}$. This implies in turn that N is dynamically invariant (compare with Definition A.4) with respect to the unforced system. Indeed, if $x_0 \in \mathrm{N}$ and $\eta = e^{tA}x_0$ for $t \in \mathbf{R}$, we have

$$Ce^{\theta A}\eta = Ce^{\theta A}e^{tA}x_0 = Ce^{(\theta+t)A}x_0 = 0$$

for each $\theta \in \mathbf{R}$, and so $\eta \in \mathrm{N}$. ■

Remark 5.7 It is easily checked that a system (5.1) defined by means of a scalar linear differential equation of order n,

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}y' + a_n y = u(t)$$

where u is interpreted as the input and y as the output, is completely observable, regardless the choice of the coefficients $a_1, \ldots, a_n$. ■

5.2.3 Reconstruction of the Initial State

For a fixed input function $u(\cdot) : [0, +\infty) \to \mathbf{R}^m$, let us look at the map which associates the output function $y(t)$ to a given initial state $x(0) = x_0$. Complete observability of system (5.1) implies that such a map is injective. Hence, we expect that monitoring $y(t)$ on the interval $[0, T]$ (for some $T > 0$) provides sufficient information in order to recover the exact value of x_0. Next we show how this can be actually done. Assume that $u(t)$ and $y(t)$ are known for $t \in [0, T]$. Recall that

$$y(t) = C\left[e^{tA}x_0 + \int_0^t e^{(t-\tau)A}Bu(\tau)\,d\tau\right].$$

Multiplying both sides by $e^{tA^{\mathbf{t}}}C^{\mathbf{t}}$ we get

$$e^{tA^{\mathbf{t}}}C^{\mathbf{t}}y(t) = e^{tA^{\mathbf{t}}}C^{\mathbf{t}}Ce^{tA}x_0 + e^{tA^{\mathbf{t}}}C^{\mathbf{t}}C\int_0^t e^{(t-\tau)A}Bu(\tau)\,d\tau$$

and integrating from 0 to T:

$$E(T)x_0 = \int_0^T e^{tA^{\mathbf{t}}}C^{\mathbf{t}}y(t)\,dt - \int_0^T e^{tA^{\mathbf{t}}}C^{\mathbf{t}}C\left(\int_0^t e^{(t-\tau)A}Bu(\tau)\,d\tau\right)dt\,, \quad (5.28)$$

where we set $E(T) = \int_0^T e^{tA^{\mathbf{t}}}C^{\mathbf{t}}Ce^{tA}\,dt$.

Theorem 5.7 *The following properties are equivalent.*

(i) System (5.1) is completely observable.
(ii) Matrix $E(T)$ is positive definite for each $T > 0$, and so invertible.

Proof A simple computation shows that

$$\xi^{\mathbf{t}}E(T)\xi = \int_0^T ||Ce^{tA}\xi||^2\,dt \geq 0\,.$$

The integral on the right-hand side is zero if and only if the integrand vanishes, that is if and only if $\xi \in \ker Ce^{tA}$ for each $t \in [0, T]$. If the system is completely observable, this may happen only if $\xi = 0$. The opposite statement can be easily proven by contradiction. ■

Therefore, if (5.1) is completely observable, from (5.28) it is possible to exactly compute x_0 taking the inverse of the matrix $E(T)$. We emphasize that in this procedure the input is absolutely arbitrary: the more natural choice is, of course, $u(t) = 0$ for each $t \in [0, T]$.

The approach to the observability problem described in this section has a drawback: indeed, (5.28) may be sometimes hard to apply, because of the need of computing the exponential matrix, some integrals and an inverse matrix.

5.2.4 Duality

The analogies between the notions of controllability and observability are evident. We point out that the observability matrix of system (5.1) coincides with the controllability matrix of system

$$\begin{cases} \dot{x} = A^{\mathbf{t}}x + C^{\mathbf{t}}u \\ y = B^{\mathbf{t}}x \end{cases} \tag{5.29}$$

where $u \in \mathbf{R}^p$ and $y \in \mathbf{R}^m$. Note that with respect to (5.1), the roles of B and C are exchanged. Thus, (5.1) is completely controllable if and only if (5.29) is completely observable and vice versa. System (5.29) is called the *dual* of (5.1). The properties of complete controllability and complete observability are also said to be dual properties.

We emphasize also the analogies (and the differences) between Theorems 5.2 and 5.7.

5.3 Canonical Decompositions

In the analysis of a system, it is important to find out certain canonical forms; they are particular representations which make possible to understand at a first glance the main structural properties of the system. This requires the search for suitable changes of coordinates.

5.3.1 Linear Equivalence

The systems

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \qquad \text{and} \qquad \begin{cases} \dot{z} = \tilde{A}z + \tilde{B}u \\ y = \tilde{C}z \end{cases}$$

are called *linearly equivalent* if there exists a linear change of coordinates $x = Pz$ ($\det P \neq 0$) such that $\tilde{A} = P^{-1}AP$, $\tilde{B} = P^{-1}B$, and $\tilde{C} = CP$. We recognize in this definition a generalization of a notion already introduced for linear unforced system, and fruitfully applied in Chap. 2.

We remark that such a transformation does not affect the controllability properties of the system. Indeed, one has

$$\begin{aligned} (\tilde{B}|\, \tilde{A}\tilde{B}| \dots |\tilde{A}^{n-1}\tilde{B}) &= (P^{-1}B|P^{-1}AB| \dots |P^{-1}A^{n-1}B) \\ &= P^{-1}(B|AB| \dots |A^{n-1}B) \end{aligned}$$

so that the controllability matrices of any pair of linearly equivalent systems have the same rank. Moreover, the subspaces engendered by the columns of these matrices (that is, the controllability spaces of these systems) are consistently transformed each other by the change of coordinates.

Similar conclusions can be achieved, of course, about the observability property and the unobservability space.

5.3.2 Controlled Invariance

Before to introduce the first important canonical form, we still need a definition. A subspace $W \subset \mathbf{R}^n$ is said to be a *controlled invariant* for system (5.1) if for each $x_0 \in W$ and for each admissible input we have:

$$x(t, x_0, u(\cdot)) \in W \qquad \forall t > 0\,.$$

We recognize in this definition an extension of Definition A.4. The space R is an example of controlled invariant. Indeed, assume by contradiction that there exist a point $x_0 \in$ R and an input function $u(\cdot) : [0, T] \to \mathbf{R}^m$ such that $x(T, x_0, u(\cdot)) = \eta \notin$ R. If $u_0(\cdot) : [0, t_0] \to \mathbf{R}^m$ is an input function for which $x(t_0, 0, u_0(\cdot)) = x_0$ (at least one such input function exists by hypothesis) we can take the new input

$$\tilde{u}(\tau) = \begin{cases} u_0(\tau) & \text{for} \quad 0 \leq \tau < t_0 \\ u(\tau - t_0) & \text{for} \quad t_0 \leq \tau \leq t_0 + T\,. \end{cases}$$

Clearly $x(t_0 + T, 0, \tilde{u}(\cdot)) = \eta$, and this is a contradiction.

Lemma 5.2 *Let the subspace $W \subset \mathbf{R}^n$ be a controlled invariant for system (5.1) and let*

$$x(t, x_0, u(\cdot))$$

a trajectory with $x_0 \in W$. Then the tangent vector of the curve $t \mapsto x(t, x_0, u(\cdot))$ for $t = 0$ belongs to W, as well.

Proof Let us consider the difference quotient

$$q(t) = \frac{x(t, x_0, u(\cdot)) - x_0}{t}$$

for $t \neq 0$. The tangent vector is defined as $\lim_{t \to 0} q(t)$. Since W is a subspace, $q(t) \in W$ for each $t \neq 0$. Hence, the limit must belong to W, as well, being a subspace a closed set. ■

5.3.3 Controllability Form

Theorem 5.8 *There exist a change of coordinates $x = Pz$ and an integer q $(0 \leq q \leq n)$ such that in the new coordinates z the system (5.1) takes the form*

$$\begin{cases} \dot{z}_1 = A_{11} z_1 + A_{12} z_2 + B_1 u \\ \dot{z}_2 = \phantom{A_{11} z_1 +} A_{22} z_2 \end{cases} \tag{5.30}$$

where $z = (z_1, z_2)$ with $z_1 \in \mathbf{R}^q$, $z_2 \in \mathbf{R}^{n-q}$, and where A_{11}, A_{12}, A_{22}, and B_1 are matrices of suitable dimensions. Moreover, the system

$$\dot{z}_1 = A_{11} z_1 + B_1 u \tag{5.31}$$

with state variable $z_1 \in \mathbf{R}^q$, is completely controllable.

Proof Let $q = \dim \mathrm{R}$. The limit cases $q = 0$ and $q = n$ correspond respectively to the cases where (5.1) is completely uncontrollable (that is $B = 0$) and the case where (5.1) is completely controllable. So, we can limit ourselves to assume $(0 < q < n)$. Consider a basis of the state space, such that the first q vectors form a basis of R. Let z be the coordinates in this new basis, partitioned in such a way that $\mathrm{R} = \{z : z_2 = 0\}$. In general, the representation of the system in these new coordinates can be written

$$\begin{cases} \dot{z}_1 = A_{11} z_1 + A_{12} z_2 + B_1 u \\ \dot{z}_2 = A_{21} z_1 + A_{22} z_2 + B_2 u \, . \end{cases}$$

We show that, according to the particular choice of the basis, $A_{21} = B_2 = 0$. Recall that R is a controlled invariant, and notice that this property does not depend on the choice of the coordinates. Assume that $A_{21} \neq 0$. Let us take an initial state $(z_1, 0) \in$ R with $z_1 \neq 0$, $A_{21}z_1 \neq 0$. Let moreover $u(t) = 0$ for each $t \geq 0$. The tangent vector to the corresponding solution, evaluated at $t = 0$, is

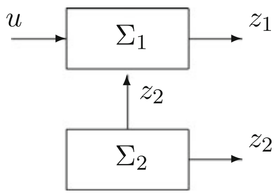
$$\begin{pmatrix} A_{11}z_1 \\ A_{21}z_1 \end{pmatrix} \notin \mathrm{R}\,.$$

This is a contradiction to Lemma 5.2. Hence $A_{21} = 0$. In a similar way, it is possible to show also that $B_2 = 0$. In practice, the matrix P which determines the change of coordinates can be written as

$$P = (v_1|\dots|v_q|v_{q+1}|\dots|v_n)$$

where $v_1, \dots, v_q$ are chosen in such a way that they form a basis of R, and $v_{q+1}, \dots, v_n$ in such a way that they form, together with $v_1, \dots, v_q$, a basis of $\mathbf{R}^n$. Notice that the basis $v_1, \dots, v_n$ is not uniquely determined by this construction. For instance, it is not restrictive (in fact, for future developments, it is strongly recommended) to take the vectors $v_1, \dots, v_n$ pairwise orthogonal.

It remains to prove that (5.31) is completely controllable. Let $z_1 \in \mathbf{R}^q$ be given. By construction, there exists an input function $u(\cdot)$ such that the corresponding solution of system (5.30) steers the origin of $\mathbf{R}^n$ to the state $(z_1, 0) \in$ R. Obviously, the same input applied to system (5.31) steers the origin of $\mathbf{R}^q$ in z_1. Hence, (5.31) is completely controllable. ■



The Eq. (5.31) can be obtained from (5.30) setting $z_2 = 0$. It can be therefore interpreted as a subsystem: it is called the *controllable part* of the overall system (in the figure above, it is denoted by Σ_1). Notice that the evolution of the component z_2 of the state in (5.30) does not depend at all on the action of the input function. It represents the *uncontrollable part* of the system (in the figure above, it is denoted by Σ_2).

The form (5.30) reveals the structure of the system, and in particular it allows us to separate and recognize the controllable and uncontrollable parts.

5.3.4 *Observability Form*

An analogous construction, based on linear equivalence, allows us to obtain a form of system (5.1) which reveals the observability properties. More precisely, it is possible to prove the existence of an integer $r (0 \leq r \leq n)$ and a nonsingular matrix P such that the change of coordinates $x = Pz$ gives rise to the form

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + B_1u \\ \dot{z}_2 = \qquad\qquad A_{22}z_2 + B_2u \\ y \;\; = C_2z_2 \end{cases} \tag{5.32}$$

where $z = (z_1, z_2)$, $z_1 \in \mathbf{R}^r$, $z_2 \in \mathbf{R}^{n-r}$, and the reduced order system

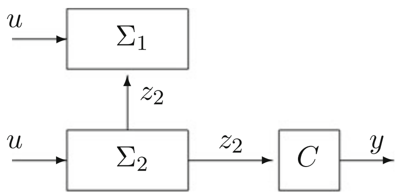
$$\begin{cases} \dot{z}_2 = A_{22}z_2 + B_2u \\ y \;\; = C_2z_2 \end{cases} \tag{5.33}$$

with state variable $z_2 \in \mathbf{R}^{n-r}$, is completely observable.

Note that if we put $z_2 = 0$ in the differential part of (5.32), the resulting reduced order system

$$\dot{z}_1 = A_{11}z_1 + B_1u \tag{5.34}$$

with state variable $z_1 \in \mathbf{R}^r$, does not produce any output. The reduced order systems (5.33) and (5.34) are called, respectively, the *observable part* (denoted by Σ_2 in the figure below) and the *unobservable part* (denoted by Σ_1 in the figure below) of the system.



However, this time the construction of the matrix P which determines the change of coordinates is more delicate. We start by computing the observability matrix. Let $n - r$ be its rank. Choose $n - r$ linearly independent columns of the observability matrix, and let us denote them by $v_{r+1}, \ldots, v_n$. Choose finally r linearly independent vectors $v_1, \ldots, v_r$ such that the subspace generated by $v_1, \ldots, v_r$ is orthogonal[1] to the subspace generated by $v_{r+1}, \ldots, v_n$. According to Theorem 5.6, the vector $v_1, \ldots, v_r$ constitute a basis of the non-observability subspace N. A possible choice of P is the matrix whose columns are

$$P = (v_1 | \ldots | v_r | v_{r+1} | \ldots | v_n) .$$

[1]We stress that in general, the construction does not work if the orthogonality requirement is neglected: this is an important difference with respect to the construction of the controllability form. The reason of this fact is implicit in the statement of Theorem 5.6.

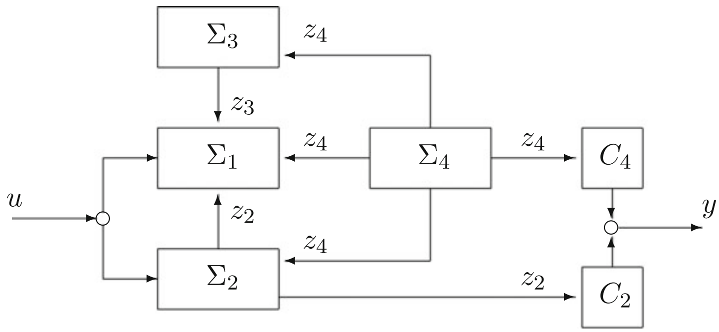
The construction can be easily modified, in order to replace P by a new matrix whose columns are all pairwise orthogonal. To prove that, after the change of coordinates, the block A_{21} becomes zero, we may argue as in the proof of Theorem 5.8, setting $u = 0$ and taking into account Remark 5.6.

5.3.5 Kalman Decomposition

The controllability form and the observability form discussed in the previous sections can be combined, giving rise to the form

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + A_{13}z_3 + A_{14}z_4 + B_1u \\ \dot{z}_2 = A_{22}z_2 + A_{24}z_4 + B_2u \\ \dot{z}_3 = A_{33}z_3 + A_{34}z_4 \\ \dot{z}_4 = A_{44}z_4 \\ y = C_2z_2 + C_4z_4 \,. \end{cases} \tag{5.35}$$

The special structure exhibited by (5.35) corresponds to the connections displayed in the figure below, where by $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ we have denoted the reduced order systems which determine the evolutions of the blocks of coordinates z_1, z_2, z_3, z_4, respectively.



We remark in particular that:

- the block of coordinates (z_1, z_2) identifies the completely controllable part: the form (5.30) is recovered as a particular case, by collecting the blocks of coordinates (z_1, z_2), (z_3, z_4);
- the block of coordinates (z_2, z_4) identifies the completely observable part: the form (5.32) is recovered by collecting the blocks of coordinates (z_1, z_3), (z_2, z_4) and rewriting the equations, after reordering the indices in the following way: 1, 3, 2, 4;
- the block of coordinates z_2 identified the completely controllable and completely observable part;
- the block of coordinates z_3 identifies the uncontrollable and unobservable part.

These remarks are trivial, with the exception of the third one, for which we provide a sketch of proof. Because of the block-triangular form of the matrix, we realize that the controllability matrix for the block of coordinates z_2 can be obtained taking suitable submatrices of the controllability matrix for the block (z_1, z_2). Such a matrix has a maximal rank, since the block (z_1, z_2) corresponds to the completely controllable part of the overall system. But this is possible only if the rank of the controllability matrix for the block z_2 is maximal. The complete observability is proved in similar way.

5.3.6 Some Examples

In order to illustrate the construction of the canonical forms presented in the previous sections we present some examples.

Example 5.2 Consider the system with scalar input defined by the matrices

$$A = \begin{pmatrix} 1 & 4 \\ 2 & -6 \end{pmatrix} \qquad b = \begin{pmatrix} 4 \\ 1 \end{pmatrix} .$$

We do not need to specify C, since in this example we will be interested only in the controllability form. The controllability matrix is $(b|Ab) = \begin{pmatrix} 4 & 8 \\ 1 & 2 \end{pmatrix}$ and its rank is equal to 1. The system is not completely controllable and we can proceed to the determination of a controllability form. We perform a change of coordinates by means of the matrices

$$P = \begin{pmatrix} 4 & -1 \\ 1 & 4 \end{pmatrix} \qquad P^{-1} = \frac{1}{17} \begin{pmatrix} 4 & 1 \\ -1 & 4 \end{pmatrix}$$

(notice that the columns of P are orthogonal). We obtain, as desired,

$$P^{-1}AP = \begin{pmatrix} 2 & -2 \\ 0 & -7 \end{pmatrix} \qquad P^{-1}b = \begin{pmatrix} 1 \\ 0 \end{pmatrix} .$$

If we chose a different matrix, with the first column parallel to b and the second column linearly independent (but not necessarily orthogonal) to the first one, we obtain again a controllability form, which in general may differ from the previous one for some unessential details. For instance, with

$$Q = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$$

we have

$$Q^{-1}AQ = \begin{pmatrix} 2 & 2 \\ 0 & -7 \end{pmatrix} \qquad Q^{-1}b = \begin{pmatrix} 1 \\ 0 \end{pmatrix} .$$

■

Example 5.3 Now consider the system with scalar input and scalar output defined by the matrices

$$A = \begin{pmatrix} 3 & 2 \\ 3 & 4 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad c = (1 \ \ 1)\,.$$

This system is completely controllable but not completely observable. Indeed, since $A^{\mathbf{t}}c^{\mathbf{t}} = \begin{pmatrix} 6 \\ 6 \end{pmatrix}$, the rank of the observability matrix is 1. We can proceed to determine an observability canonical form. Define a change of coordinates by the matrices

$$P = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \qquad P^{-1} = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}.$$

We get

$$P^{-1}AP = \begin{pmatrix} 1 & 1 \\ 0 & 6 \end{pmatrix} \qquad P^{-1}b = \frac{1}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad cP = (0 \ \ 2)\,.$$

Note that the columns of P are orthogonal. Note also that with this procedure, the unobservability space becomes coincident with the first component. ■

Example 5.4 Consider finally the system with scalar input and scalar output defined by the matrices

$$A = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad c = (-1 \ \ 2)\,.$$

In this case, both the controllability and the observability matrices have rank 1. The system is neither completely controllable nor completely observable. It is convenient to start by computing an observability form. However, now it is preferable to proceed in a lightly different way. Making use of the change of coordinates defined by the matrices

$$P = \begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix} \qquad P^{-1} = \frac{1}{5}\begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix}, \tag{5.36}$$

we may align, in the new coordinates, the unobservability space with the second component instead of the first one, as we did in Example 5.3. We have:

$$P^{-1}AP = \begin{pmatrix} 1 & 0 \\ -3 & 2 \end{pmatrix} = \hat{A} \quad P^{-1}b = \frac{1}{5}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \hat{b} \quad cP = (5 \ \ 0) = \hat{c}\,.$$

Notice that the columns of P are orthogonal. Notice also that this form does not allow to identify immediately the controllability space. Thus, we need to apply a further change of coordinates, to the purpose of achieving a complete Kalman decomposition. More precisely, we need to find a change of coordinates which, while keeping unchanged the position of the unobservability space, superposes the controllability space to the subspace orthogonal to the unobservability one (that is, with the first component). We can take:

$$Q = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \qquad Q^{-1} = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}$$

(the first column of Q is parallel to $\hat{b}$, the second generates the unobservability space in the new coordinates). We finally obtain the required form:

$$Q^{-1}\hat{A}Q = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \qquad Q^{-1}\hat{b} = \frac{1}{5}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{c}P = (5\ 0)\,.$$

It is now evident, in particular, that the system possesses a completely controllable and completely observable part, and a uncontrollable and unobservable part.

Of course, the transformation can be accomplished by a unique change of coordinates defined by the matrix

$$PQ = \begin{pmatrix} 5 & 2 \\ 5 & 1 \end{pmatrix}\,.$$

■

Example 5.5 As a last example we take the same matrices A and c as in Example 5.4, but

$$b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}\,.$$

By the first change of coordinates given by (5.36), we get

$$P^{-1}b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\,.$$

The controllability space is already coincident with the unobservability space. Of course, now it is not possible, by a further change of coordinates, to move the controllability space in such a way that it becomes orthogonal to the unobservability space. The complete Kalman form has been obtained with the first change of coordinates. The system possesses a controllable part which is not observable, and an observable part which is not controllable. ■

5.4 Constrained Controllability

One of the most important developments of control theory is optimization. Typical examples are the minimal time problem [16] and the quadratic regulator problem [6]. In particular, minimal time problems are strictly related to the controllability properties of the system. However, a minimal time problem makes sense only if the admissible control functions are constrained to take values in a bounded set. Although optimization is beyond the purposes of this book, in this section we discuss shortly how the geometric properties of the reachable sets change when the admissible controls are subject to constraints. Thus, in this section we consider linear systems

$$\dot{x} = Ax + Bu \tag{5.37}$$

with $x \in \mathbf{R}^n$ and $u \in U$, where U represents a nonempty, proper (in general, bounded) subset of $\mathbf{R}^m$. To begin with, we need to update the notion of reachability introduced in the previous Sect. 5.1 and the related notation.

Let $x_0, \eta_0 \in \mathbf{R}^n$. We say that η_0 is reachable from x_0 at time $T > 0$ with constrained controls if there exists $u(\cdot) \in \mathcal{PC}([0, +\infty), U)$ such that $x(T, x_0, u(\cdot)) = \eta_0$, and we denote $\mathrm{R}(T, x_0, U)$ the set of such points. This is called the reachable set of (5.37) with constrained controls.

The geometric properties of $\mathrm{R}(T, x_0, U)$ depend not only on the pair of matrices A and B, but also on the set U. For instance, it is no more true in general that $\mathrm{R}(T, 0, U)$ is a subspace of $\mathbf{R}^n$. As a consequence, we should enrich the notion of reachability introducing some new definitions.

Definition 5.7 A system (5.37), with admissible inputs constrained to a set U, is said to be:

- *accessible* from x_0 at time T when $\overset{\circ}{\mathrm{R}}(T, x_0, U) \neq \emptyset$;
- *locally reachable* from x_0 at time T when $x_0 \in \overset{\circ}{\mathrm{R}}(T, x_0, U)$;
- *locally reachable along the free solution* from x_0 at time T if $x(T, x_0, 0) \in \overset{\circ}{\mathrm{R}}(T, x_0, U)$.

Notice that when $x_0 = 0$, the definitions of local reachability and local reachability along the free solution coincide.

Example 5.6 Consider the simple scalar system

$$\dot{x} = x + u, \qquad x, u \in \mathbf{R}.$$

Let $u_0 > 0$ be fixed and assume that the control functions are subject to the constraint $|u(t)| \leq u_0$. Solving the equation, we have for $t \geq 0$

$$x(t) = e^t x_0 + \int_0^t e^{(t-\tau)} u(\tau)\, d\tau$$

that is

$$e^t x_0 - \int_0^t e^{(t-\tau)} u_0\, d\tau \leq x(t) \leq e^t x_0 + \int_0^t e^{(t-\tau)} u_0\, d\tau$$

namely

$$e^t(x_0 - u_0) + u_0 \le x(t) \le e^t(x_0 + u_0) - u_0$$

and finally

$$e^t x_0 - (e^t - 1)u_0 \le x(t) \le e^t x_0 + (e^t - 1)u_0 \,.$$

This shows that the system is accessible and locally reachable along the free solution, for each choice of x_0 and u_0 (recall that if $t \ge 0$ then $e^t - 1 \ge 0$).

If $-u_0 < x_0 < u_0$, then x_0 belongs to the interval $(e^t(x_0 - u_0) + u_0, e^t(x_0 + u_0) - u_0) = \overset{\circ}{\mathrm{R}}(t, x_0, U)$ for each t. Hence, the system is also locally reachable at x_0. Moreover, $x_0 - u_0 < 0 < x_0 + u_0$ so that $\mathrm{R}(x_0, U) = \mathbf{R}$. On the contrary, if $x_0 \ge u_0$ then $\mathrm{R}(x_0, U)$ coincides with the half line $[x_0, +\infty)$. In this case the system is not locally reachable. The conclusion is the same for $x_0 \le -u_0$. ■

The study of the reachability properties of a linear system with constrained input can be actually reduced to the case $x_0 = 0$.

Proposition 5.5 *Let a system of the form (5.37) be given, with admissible control functions constrained to a subset* $U \subset \mathbf{R}^m$.

(i) The system is accessible from x_0 *at time* T *if and only if it is accessible from the origin at the time* T.

(ii) The system is locally reachable from x_0 *at time* T *along the free solution if and only if it is locally reachable from the origin at time* T *or, equivalently,* $0 \in \overset{\circ}{\mathrm{R}}(T, 0, U)$.

The proof is a straightforward application of the variation of constants formula. Notice that if the constraints are relaxed, system (5.37) is accessible from any initial state if and only if $\mathrm{R}(T, 0, \mathbf{R}^m) = \mathbf{R}^n$, that is if and only if the system is globally reachable at time T.

We already noticed that if U is a proper subset of $\mathbf{R}^m$, then $\mathrm{R}(T, 0, U)$ is no more, in general, a subspace of $\mathbf{R}^n$. However, it preserves an important property.

Proposition 5.6 *Consider the system (5.37). For each* $T \ge 0$ *and for each nonempty constraint set* U*, the set* $\mathrm{R}(T, 0, U)$ *is convex.*

The proof is trivial if U is convex. Otherwise, some advanced results of measure theory are needed (see [23] p. 163, [12] p. 11).

Theorem 5.9 *Assume that system (5.37) is completely controllable. Assume in addition that* $0 \in \overset{\circ}{U}$. *Then,* $0 \in \overset{\circ}{\mathrm{R}}(T, 0, U)$.

Proof Let $\mathcal{B}$ be a ball centered at the origin. Let r be radius of $\mathcal{B}$, chosen in such a way $\mathcal{B} \subset U$. Since the system is completely controllable, for each unit vector $\mathbf{e}_i$ of the canonical basis of $\mathbf{R}^n$ there exists a control function $u_i(t) : [0, T] \to \mathbf{R}^m$ which steers the system from the origin to $\mathbf{e}_i$ at time T. The control functions $u_i(t)$ need not to met the prescribed constraints but, since they are piecewise continuous and hence bounded on $[0, T]$, there exists $M > 0$ such that

$$|u_i(t)| < M \qquad \forall t \in [0, T], \forall i = 1, \ldots, n\,.$$

Since the system is linear, the controls

$$\tilde{u}_i(t) = \frac{r}{M} u_i(t)$$

steer the system from the origin to some vectors $\tilde{\mathbf{e}}_i$ which still constitute a basis of $\mathbf{R}^n$. These new control functions satisfy the required constraints. Invoking again the linearity of the system, we finally see that the points $-\tilde{\mathbf{e}}_i$ can be reached by means of the control functions $-\tilde{u}_i(t)$. The conclusion follows, since $\mathrm{R}(T, 0, U)$ is convex. ■

Corollary 5.6 *Under the assumptions of Theorem 5.9, the system possesses the property of local reachability along the free solution.*

Corollary 5.7 *Assume that system (5.37) is completely controllable. Moreover, assume that $\overset{\circ}{U} \neq \emptyset$. Then, the system possesses the property of accessibility from the origin for each $T > 0$.*

Proof Let $u_0 \in \overset{\circ}{U}$. By assumption, there exists a ball of positive radius centered at u_0, which is contained in U. Replacing U by

$$U - \{u_0\} = \{v : v = u - u_0 \text{ with } u \in U\}$$

we obtain a system which satisfies the assumptions of Theorem 5.9. Thus, it is sufficient to remark that

$$\mathrm{R}(T, 0, U) = \mathrm{R}(T, 0, U - \{u_0\}) + \int_0^T e^{(T-s)A} B u_0 \, ds\,.$$

■

Chapter Summary

In this chapter we deal with the so-called structural properties of a linear system with input and output. These properties depend only on the coefficients of the mathematical model. We study in particular controllability (which provides a measure of our ability to control the system) and observability (which provides a measure of our ability of extracting information about the state of the system). We obtain algebraic characterizations of these properties. We also study canonical forms i.e., linear transformations of the state space which allow us to rewrite the model. This makes more evident, in this way, recognizing the controllability and observability properties.

Chapter 6
External Stability

The classical notions of stability and asymptotic stability are no more sufficient to describe the behavior of a system, in the presence of external input. Consider for instance a system for which the origin is stable when the external inputs are switched off. Likely, restoring the external forces, a deviation from the equilibrium will be observed. It seems natural to expect that the amplitude of the deviation is related to the amplitude of the input signal, and that it will be "small" in some sense, if the input signal is "small". As already mentioned in Chap. 1, this kind of behavior is informally qualified as *external stability*. However, simple stability is not sufficient to guarantee such a natural behavior, as shown for instance by the simple example $\dot{x} = u_0$, where $x \in \mathbf{R}$ and u_0 is a nonzero constant.

In Chap. 4 we proved that a linear differential system with a constant (respectively, periodic) forcing term has a constant (respectively, periodic) solution, provided that resonance does not occur: such solutions are, of course, bounded. For constant (or periodic) forcing terms, resonance is surely avoided if all the eigenvalues of the unforced system have negative real part. Moreover, in this case all the solutions approach the constant (or periodic) one and are so bounded. This remark points out that the right property to be considered in order to characterize external stability is internal stability.

In this chapter we are actually interested in studying the relationship between the external stability and the internal stability of a linear system of the form

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \tag{6.1}$$

where, with the usual notation, $x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$, $y \in \mathbf{R}^p$. Recall that informally, it is used to say that the system (6.1) is *internally stable* when the associate unforced system

$$\dot{x} = Ax \tag{6.2}$$

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_6

has an asymptotically stable equilibrium point at the origin. We will see in particular that internal stability actually implies external stability, but the converse does not hold in general.

To the purposes of this chapter, the space of the input maps is restricted to $\mathcal{PCB}([0, +\infty), \mathbf{R}^m)$ endowed with the norm of the uniform convergence. The norm of the uniform convergence will be adopted also for the output maps, provided that they belong to $\mathcal{CB}([0, +\infty), \mathbf{R}^p)$. Moreover, we denote respectively by $x(t, x_0, u(\cdot))$ and $y(t, x_0, u(\cdot))$ the solution of the differential system and the output map, corresponding to the initial state x_0 and the input $u(t)$.

6.1 Definitions

For reader's convenience, we recall the definition of BIBO-stability already given in Chap. 1 (see Definition 1.3).

Definition 6.1 We say that system (6.1) is *BIBO-stable* if for each $R > 0$ there exists $S > 0$ such that

$$\|x_0\| \le R, \quad \|u(\cdot)\|_\infty \le R \Longrightarrow \|y(t, x_0, u(\cdot))\| \le S \tag{6.3}$$

for each $t \ge 0$, where $x_0 \in \mathbf{R}^n$ and $u(\cdot) \in \mathcal{PCB}([0, +\infty), \mathbf{R}^m)$.

From (6.3) we infer in particular that for any BIBO-stable system the output map $y(\cdot, x_0, u(\cdot))$ necessarily belongs to $\mathcal{CB}([0, +\infty), \mathbf{R}^p)$, and $||y(\cdot, x_0, u(\cdot))||_\infty \le S$.

If $C = I$ (the unit matrix), inequality (6.3) holds with $y(t, x_0, u(\cdot))$ replaced by $x(t, x_0, u(\cdot))$; in such a case we say, more appropriately, that the system is *BIBS-stable* (*bounded-input-bounded-state-stable*). There is another possible definition of external stability for system (6.1).

Definition 6.2 We say that system (6.1) has the *finite gain stability* property if there exist two positive constant γ_1 and γ_2 such that

$$\|y(t, x_0, u(\cdot))\| \le \gamma_1 \|x_0\| + \gamma_2 \|u(\cdot)\|_\infty \tag{6.4}$$

for each $t \ge 0$, each $x_0 \in \mathbf{R}^n$ and each input $u(\cdot) \in \mathcal{PCB}([0, +\infty), \mathbf{R}^m)$ (without loss of generality, we may assume that $\gamma_1 = \gamma_2$).

At a first glance, inequality (6.4) appears more restrictive than (6.3): indeed, it requires that the norm of the output is proportional to the norms of the input and of the initial state. Our first task is to prove that, for linear systems, Definitions 6.1–6.2 are actually equivalent.

Theorem 6.1 *Given a system of the form (6.1), the finite gain stability property is necessary and sufficient for the BIBO-stability property.*

The proof of Theorem 6.1 requires several steps. The sufficient part is almost trivial.

Proof of the sufficient part Assume that (6.4) holds. Then, (6.3) is fulfilled, if we take for each $R > 0$, $S = R(\gamma_1 + \gamma_2)$. ■

To prove the necessary part, we need to establish some preliminary lemmas.

Lemma 6.1 *Let a system of the form (6.1) be given. If it is BIBO-stable, then there exists* $M > 0$ *such that* $||Ce^{tA}|| < M$ *for* $t \geq 0$.

Proof By assumption, there exists a constant S_1 such that if $||x_0|| \leq 1$ then $||y(t, x_0, 0)|| \leq S_1$ for each $t \geq 0$. Let us denote by $\eta_{ij}(t)$ a generic entry of the matrix Ce^{tA} and let us assume by contradiction that there exists a pair of indices i, j for which $\eta_{ij}(t)$ is not bounded for $t \geq 0$. Take as x_0 the j-th vector of the canonic basis, that is the vector e_j whose components are zero, except the j-th which is 1. The i-th component of the function $\psi(t) = Ce^{tA}e_j$ is equal to $\eta_{ij}(t)$ and we have

$$|\eta_{ij}(t)| \leq ||\psi(t)|| \,.$$

The function $\psi(t) = Ce^{tA}e_j$ is therefore unbounded for $t \geq 0$. But $\psi(t)$ coincides with $y(t, e_j, 0)$, and $||e_j|| = 1$. We get in this way a contradiction to the BIBO-stability assumption. Thus, we are led to conclude that all the entries of the matrix Ce^{tA} are bounded, and the conclusion easily follows. ■

Lemma 6.2 *If the system (6.1) is BIBO-stable, then there exists* $L > 0$ *such that*

$$\int_0^t \|W(\tau)\| \, d\tau < L \tag{6.5}$$

for each $t \geq 0$*, where* $W(\tau) = Ce^{\tau A}B$.

Remark 6.1 Note that $W(\tau)$ is a matrix with p rows and m columns. Note also that Eq. (6.5) holds if and only if the integral

$$\int_0^\infty \|W(\tau)\| \, d\tau$$

is convergent. ■

Proof of Lemma 6.2 A system of the form (6.1) can be interpreted as an impulse response system provided that:

(a) the initial state is equal to zero;
(b) the admissible input maps are assumed to vanish for $t < 0$.

Under these conditions, the impulse response matrix can be defined as

$$h(t) = \begin{cases} W(t) & \text{if } t \geq 0 \\ 0 & \text{if } t < 0. \end{cases} \tag{6.6}$$

Indeed, by virtue of the variation of constants formula, when the system is initialized to zero, the input-output operator writes

$$y(t) = \int_0^t Ce^{(t-\tau)A} Bu(\tau)\, d\tau = \int_0^t W(t-\tau)u(\tau)\, d\tau$$

and hence, taking into account (b) and (6.6),

$$y(t) = \int_{-\infty}^{+\infty} h(t-\tau)u(\tau)\, d\tau\,.$$

The proof can be therefore carried out by repeating the same arguments as in Chap. 1 (necessary part of the proof of Proposition 1.5). ■

In what follows, we sometimes refer the term "impulse response matrix" to the matrix $W(t)$. This is a little abuse, justified by the reasons explained in the previous proof.

Remark 6.2 Lemma 6.2 can be restated by saying that in the case of linear systems, BIBO-stability implies the boundedness of the impulse response. However in general, boundedness of the impulse response and BIBO-stability are not equivalent. As an example, we can consider the system

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

with $C = I$. This system has a bounded impulse response matrix, but it is not BIBO-stable.

Thus we see that the sufficient part of Proposition 1.5 of Chap. 1 does not hold for systems of the form (6.1): the reason of course is that the behavior of the system depends here also on the initial conditions, and not only on the input (as for the systems considered in Proposition 1.5). ■

We are now in a position to complete the proof of Theorem 6.1.

Proof of the necessary part Assuming that the system (6.1) is BIBO-stable, we show that it has the finite gain stability property, as well.

We first consider the output of the system (6.1) corresponding to the initial state $x_0 = 0$ and any admissible (bounded) input $u(\cdot)$. According to Lemma 6.2, with $L = \gamma_2$, we get

$$\|y(t, 0, u(\cdot))\| \le \int_0^t \|W(t-\tau)u(\tau)\|\, d\tau \le \int_0^t \|W(t-\tau)\|\ \|u(\cdot)\|_\infty\, d\tau$$
$$\le \|u(\cdot)\|_\infty \int_0^t \|W(\sigma)\|\, d\sigma \le \gamma_2 \|u(\cdot)\|_\infty$$

for each $t \geq 0$. Then we consider the output of system (6.1) corresponding to any initial state x_0 and the constant input $u(t) = 0$. According to Lemma 6.1, with $M = \gamma_1$, we have

$$||Ce^{tA}|| \leq \gamma_1$$

for each $t \geq 0$. This yields

$$||y(t, x_0, 0)|| \leq \gamma_1 ||x_0||$$

for each $x_0 \in \mathbf{R}^n$ and each $t \geq 0$. Since for a linear system

$$y(t, x_0, u(\cdot)) = y(t, x_0, 0) + y(t, 0, u(\cdot))$$

the conclusion is readily achieved. ■

From now on, when we say that a linear system is *externally stable* we mean that one of the two Definitions 6.1 and 6.2 is (and hence both are) fulfilled. To conclude this section, we show that external stability is invariant under linear changes of coordinates in the state space $\mathbf{R}^n$.

Proposition 6.1 *Let the system*

$$\begin{cases} \dot{z} = \tilde{A}z + \tilde{B}u \\ y = \tilde{C}z \end{cases} \tag{6.7}$$

be linearly equivalent to (6.1). System (6.7) is externally stable if and only if system (6.1) is externally stable.

Proof Since the systems are linearly equivalent, there exists a nonsingular matrix P such that $\tilde{A} = P^{-1}AP$, $\tilde{B} = P^{-1}B$, $\tilde{C} = CP$. Assume that (6.1) has the finite gain stability property. By the variation of constants formula, the output of the system (6.7) writes

$$y(t) = \tilde{C}e^{t\tilde{A}}z_0 + \int_0^t \tilde{C}e^{(t-\tau)\tilde{A}}\tilde{B}u(\tau)\,d\tau = Ce^{tA}Pz_0 + \int_0^t Ce^{(t-\tau)A}Bu(\tau)\,d\tau\,.$$

It coincides with the response of system (6.1) corresponding to the initial state Pz_0 and the same input map. We have

$$||y(t)|| \leq \gamma_1||Pz_0|| + \gamma_2||u(\cdot)||_\infty \leq \tilde{\gamma}_1||z_0|| + \gamma_2||u(\cdot)||_\infty$$

where $\tilde{\gamma}_1 = ||P||\gamma_1$. The proof of the converse statement can be achieved by exchanging the roles of the systems. ■

6.2 Internal Stability

Recall that the origin is asymptotically stable for the unforced system (6.2) if and only if all the eigenvalues of A have negative real part. In this case, we use to say that A has the Hurwitz property or, in short, that A is a Hurwitz matrix (Definition 3.1). We show that in general the internal stability property implies the external stability one.

Theorem 6.2 *Assume that the matrix A is Hurwitz. Then, the system (6.1) has the finite gain stability property.*

Proof By virtue of Proposition 4.2, for some constants $k_0, \alpha, \bar{b}$ we have

$$||x(t, x_0, u(\cdot))|| \leq k_0||x_0||e^{\alpha t} + \bar{b}||u(\cdot)||_\infty$$

for $t \geq 0$. If A is Hurwitz, α can be taken negative, and so $e^{\alpha t} < 1$. The conclusion follows, since $||y|| \leq ||C|| \cdot ||x||$. ■

Unfortunately, the converse of Theorem 6.2 is false in general.

Example 6.1 Let us consider the system

$$\begin{cases} \dot{x}_1 = 0 \\ \dot{x}_2 = -x_2 + u \end{cases}$$

with the identity as observation function. It is easy to check that this system has the finite gain stability property, but it is not internally stable. ■

Next we address the problem of identifying additional conditions which allow us to prove a partial converse of Theorem 6.2 or, alternatively, a property weaker than asymptotic stability of the unforced system, which could be proved to be equivalent to the external stability of (6.1). In the next section, we solve this problem for the particular case where the matrix C is the identity. The general case will be studied later in Sect. 6.4.

6.3 The Case $C = I$

If $C = I$, the output coincides with the state, and the system takes the form

$$\begin{cases} \dot{x} = Ax + Bu \\ y = x\,. \end{cases} \tag{6.8}$$

Recall that under this condition the BIBO-stability property reduces to the BIBS-stability one and, moreover, $W(t) = e^{tA}B$.

Proposition 6.2 *If system (6.8) is BIBO-stable, then the origin is a (in general, not asymptotically) stable equilibrium point for the unforced system (6.2).*

Proof Of course, if the system is BIBO-stable then the solutions of the unforced system (6.2) must be bounded for $t \geq 0$. Thus, the origin is stable (in Lyapunov sense) for, the unforced system (6.2). ∎

On the other hand, simple examples (take for instance the scalar system $\dot{x} = 1$) point out that the simple (not asymptotic) stability of the unforced system is not sufficient for the BIBO-stability of system (6.8). The property of the unforced system we are looking for, should be therefore intermediate between simple stability and asymptotic stability.

Lemma 6.3 *Assume that the system (6.8) is BIBO-stable, and let* $u(\cdot) : [0, +\infty] \to \mathbf{R}^m$ *be a piecewise continuous function. If there exists a number* $\bar{t} > 0$ *such that* $u(t) = 0$ *for every* $t \geq \bar{t}$, *then*

$$\lim_{t \to +\infty} x(t, 0, u(\cdot)) = 0\,. \tag{6.9}$$

Proof Let $\bar{u}$ be a constant such that $||u(t)|| \leq \bar{u}$ for $t \in [0, \bar{t}]$. If $t > \bar{t}$, we have

$$x(t, 0, u(\cdot)) = \int_0^{\bar{t}} W(t - \tau)u(\tau)\, d\tau = \int_{t-\bar{t}}^{t} W(\sigma)u(t - \sigma)\, d\sigma$$

which implies

$$||x(t, 0, u(\cdot))|| \leq \bar{u} \int_{t-\bar{t}}^{t} ||W(\sigma)||\, d\sigma\,. \tag{6.10}$$

Since the system is BIBO-stable, the integral $\int_0^\infty ||W(\sigma)||\, d\sigma$ converges (Lemma 6.2). Hence, the integral in (6.10) can be rendered arbitrarily small for sufficiently large t. ∎

We are now ready to prove a partial converse of Theorem 6.2 for systems of the form (6.8).

Proposition 6.3 *Assume that system (6.8) is completely controllable. If in addition the system is BIBO-stable, then the matrix A is Hurwitz.*

Proof We will show that the origin is globally attractive for the unforced system (6.2). In other words, we will show that for each $x_0 \in \mathbf{R}^n$,

$$\lim_{t \to +\infty} x(t, x_0, 0) = 0\,.$$

Since (6.8) is completely controllable, for any fixed instant $T_0 > 0$ there exists an admissible input map $u_0 : [0, T_0) \to \mathbf{R}^m$ such that $x(T_0, 0, u_0(\cdot)) = x_0$. Let us define

$$u(t) = \begin{cases} u_0(t) & \text{if } t \in [0, T_0] \\ 0 & \text{if } t > T_0. \end{cases}$$

According to Lemma 6.3, we must have

$$\lim_{t \to +\infty} x(t, 0, u(\cdot)) = 0\,.$$

Since $x(t, 0, u(\cdot)) \equiv x(t, 0, u_0(\cdot))$ for $t \in [0, T_0]$, we have $x(T_0, 0, u(\cdot)) = x_0$. Moreover, using time invariance,

$$x(t, 0, u(\cdot)) = x(t - T_0, x_0, 0) \quad \text{for} \quad t > T_0\,.$$

The conclusion is straightforward. ■

So far, we have established that under the complete controllability assumption, and continuing to assume $C = I$, internal stability is necessary and sufficient for external stability. However, a system may be internally stable even if not completely controllable (this happens for instance with the scalar system $\dot{x} = -x + bu$, if $b = 0$). A more precise necessary and sufficient condition of external stability for systems of the form (6.8) can be achieved by weakening the complete controllability assumption.

Recall that, apart from a linear change of coordinates, we can rewrite system (6.8) under the controllability canonical form

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + B_1 u \\ \dot{z}_2 = \qquad\quad A_{22}z_2 \end{cases} \tag{6.11}$$

where $z_1 \in \mathbf{R}^q$ e $z_2 \in \mathbf{R}^{n-q}$ (the notation is that of Chap. 5) and the pair of matrices A_{11}, B_1 defines a completely controllable system in $\mathbf{R}^q$.

Theorem 6.3 *System (6.8) is BIBO-stable if and only if both the following conditions hold:*

(i) all the eigenvalues of the matrix A_{11} of its controllable part have negative real part;

(ii) all the eigenvalues of the matrix A_{22} of its uncontrollable part have nonpositive real part and, for each possible eigenvalue of A_{22} with zero real part, the algebraic and geometric multiplicity coincide.

Proof If system (6.11) is BIBO-stable, the same is true for the subsystem

$$\dot{z}_1 = A_{11}z_1 + B_1 u\,. \tag{6.12}$$

To prove statement (i), it is sufficient to remark that since (6.12) is completely controllable, Proposition 6.3 applies. On the other hand, the eigenvalues of A_{22} form a subset of the set of the eigenvalues of A. Therefore, statement (ii) follows from Proposition 6.2.

Next we prove the converse. Let us fix a positive number R. Consider a point $\bar{z} \in \mathbf{R}^n$ and an input map $u(t)$ such that

$$\|\bar{z}\| < R, \quad \|u(\cdot)\|_\infty < R\,.$$

Denote by $z(t) = (z_1(t), z_2(t))$ the solution of the system, written in the form (6.11), corresponding to the initial condition $\bar{z} = (\bar{z}_1, \bar{z}_2)$ and the input $u(t)$. Notice that

$$\|z(t)\| \le \|z_1(t)\| + \|z_2(t)\|$$

and

$$\|\bar{z}_1\| \le \|\bar{z}\|, \quad \|\bar{z}_2\| \le \|\bar{z}\|$$

where, depending on the vectors, the norms are taken in $\mathbf{R}^n$, in $\mathbf{R}^q$ or in $\mathbf{R}^{n-q}$. The assumption about A_{22} implies the existence of a constant $\gamma > 0$ such that

$$\|z_2(t)\| \le \gamma\|\bar{z}_2\| < \gamma R\,.$$

Let $v(t) = A_{12}z_2(t) + B_1u(t)$. We have

$$\|v(t)\| \le \|A_{12}\|\gamma R + \|B_1\|R\,.$$

Let now $\tilde{R} = \max\{R, \|A_{12}\|\gamma R + \|B_1\|R\}$. Since A_{11} is Hurwitz, the subsystem

$$\dot{z}_1 = A_{11}z_1 + v$$

is BIBO-stable. Hence, there exists a constant $\tilde{S}$ such that if

$$\|v(t)\| < \tilde{R}, \quad \|\bar{z}_1\| < \tilde{R} \tag{6.13}$$

then $\|z_1(t)\| < \tilde{S}$. Note that in our case, conditions (6.13) are valid by construction. Hence,

$$\|z(t)\| \le \|z_1(t)\| + \|z_2(t)\| \le \tilde{S} + \gamma R$$

and the inequality (6.3) is recovered taking $S = \tilde{S} + \gamma R$. ■

6.4 The General Case

Now we come back to the general case (6.1). First, we establish a generalized version of Proposition 6.2.

Proposition 6.4 *Assume that system (6.1) is completely observable and, in addition, BIBO-stable. Then, the origin is a stable (in general, not asymptotically) equilibrium point for the unforced system (6.2).*

Proof By the external stability assumption, for each $R > 0$ there exists $S > 0$ such that if $||x_0|| < R$ and $u(t) = 0$ for each $t \geq 0$, then we have

$$||Ce^{tA}x_0|| < S$$

for $t \geq 0$. We distinguish several cases. First, we assume that A possesses a real eigenvalue $\lambda > 0$, and let $v \neq 0$ be an eigenvector of λ such that $||v|| < R$. We have

$$||Ce^{tA}v|| = ||C(e^{\lambda t}v)|| = e^{\lambda t}||Cv|| < S$$

for $t \geq 0$, which is possible only if $Cv = 0$. The properties of the exponential matrix can be invoked to infer that

$$\frac{d}{dt}Ce^{tA}v = CAe^{tA}v\,.$$

On the other hand, we also have

$$\frac{d}{dt}Ce^{tA}v = \frac{d}{dt}e^{\lambda t}Cv = \lambda e^{\lambda t}Cv\,.$$

From $Cv = 0$ we therefore obtain $CAe^{tA}v = 0$ for $t \geq 0$, and in particular, for $t = 0$, $CAv = 0$. This procedure can be iterated, until we arrive to conclude that:

$$\begin{pmatrix} C \\ CA \\ \dots \\ CA^{n-1} \end{pmatrix} v = 0\,.$$

This implies in turn that the rank of the matrix $(C^{\mathbf{t}}|A^{\mathbf{t}}C^{\mathbf{t}}|\dots|(A^{\mathbf{t}})^{n-1}C^{\mathbf{t}})$ is strictly less than n, which is impossible since the system is completely observable.

Second, assume that A has an eigenvalue $\lambda = 0$, whose geometric multiplicity is strictly less than the algebraic multiplicity. Then there exist an eigenvector $v_0 \neq 0$ and a generalized eigenvector $v_1 \neq 0$ such that

$$e^{tA}v_1 = tv_0 + v_1\,.$$

It is not restrictive to assume $||v_1|| < R$, so that $||Ce^{tA}v_1|| < S$ per $t \geq 0$. But

$$||Ce^{tA}v_1|| = ||tCv_0 + Cv_1|| \geq t||Cv_0|| - ||Cv_1||$$

which is bounded for $t \geq 0$ only if $Cv_0 = 0$. Of course,

$$\frac{d}{dt}e^{tA}v_1 = \frac{d}{dt}(tv_0 + v_1) = v_0$$

that is

$$\frac{d}{dt}Ce^{tA}v_1 = Cv_0 = 0\,.$$

On the other hand, we also have

$$\frac{d}{dt}Ce^{tA}v_1 = CAe^{tA}v_1 = CA(tv_0 + v_1)\,.$$

This last expression vanishes only if $CAv_0 = 0$. The remaining part of the proof can be carried out as in the previous case.

In analogous way, we exclude the cases where A possesses complex eigenvalues with positive real part, or imaginary eigenvalues with algebraic multiplicity strictly greater than their geometric multiplicity. ■

Lemma 6.4 *Assume that system (6.1) is completely observable and, in addition, BIBO-stable. Then system (6.8), where the matrices A and B are the same as those of system (6.1), is BIBO-stable.*

Proof Let us start by a preliminary remark, which will be used later. Let $E(t)$ be the matrix defined in Sect. 5.2.3. By assumption, $E(t)$ is positive definite for each $t \geq 0$. Let $M = \min_{||x||=1} x^{\mathbf{t}}E(1)x$. This minimum exists and it is strictly positive. Moreover, for each $x \in \mathbf{R}^n$,

$$x^{\mathbf{t}}E(1)x = \left(\frac{x^{\mathbf{t}}}{||x||^2}E(1)\frac{x}{||x||^2}\right)||x||^2 \geq M||x||^2\,.$$

The proof of the Lemma is by contradiction. Assume that there exists $R > 0$ such that for each $S > 0$ we can find an initial state x_0, an admissible input $u(\cdot) : [0, +\infty] \to \mathbf{R}^m$, and an instant $\tau > 0$ such that

$$||x_0|| < R, \quad ||u(\cdot)||_\infty < R, \quad \text{but} \quad ||x(\tau)|| > \sqrt{\frac{S}{M}}\,.$$

Consider the output $y(t)$ of system (6.1), corresponding to the same initial state x_0 and the admissible input

$$\tilde{u}(t) = \begin{cases} u(t) & \text{for } t \leq \tau \\ 0 & \text{for } t > \tau. \end{cases}$$

Of course, $||\tilde{u}(\cdot)||_\infty < R$. For $t \geq \tau$ we have

$$y(t) = Cx(t) = Ce^{(t-\tau)A}x(\tau)\,.$$

Taking the norms, and integrating over the interval $[\tau, \tau+1]$, we have

$$\begin{aligned}\int_\tau^{\tau+1} ||y(t)||\,dt &= \int_\tau^{\tau+1} ||Ce^{(t-\tau)A}x(\tau)||\,dt \\ &= \int_\tau^{\tau+1} x^{\mathbf{t}}(\tau)e^{(t-\tau)A^{\mathbf{t}}}C^{\mathbf{t}}Ce^{(t-\tau)A}x(\tau)\,dt \\ &= x^{\mathbf{t}}(\tau)\int_0^1 e^{\sigma A^{\mathbf{t}}}C^{\mathbf{t}}Ce^{\sigma A}\,d\sigma x(\tau) = x^{\mathbf{t}}(\tau)E(1)x(\tau)\; > 0\,.\end{aligned}$$

On the other hand, $\int_\tau^{\tau+1} ||y(t)||\,dt \le \max_{[\tau,\tau+1]} ||y(t)|| = ||y(\tilde{\tau})||$ for some $\tilde{\tau} \ge \tau$. Comparing these two conclusions, we get

$$||y(\tilde{\tau})|| \ge \int_\tau^{\tau+1} ||y(t)||\,dt \ge x^{\mathbf{t}}E(1)x \ge M||x||^2 > S\,.$$

This means that (6.1) is not externally stable, which contradicts one of the hypotheses. ■

Thanks to the previous Lemma, we can state a partial converse of Theorem 6.2 for the systems of the form (6.1); this is actually a generalization of Proposition 6.3.

Proposition 6.5 *Assume that system (6.1) is completely controllable and completely observable. Assume also that the system is BIBO-stable. Then, the matrix A is Hurwitz.*

Proof According to Lemma 6.4, system (6.8) is externally stable. If system (6.1) is also completely controllable, then system (6.8) is completely controllable, as well. The conclusion is easily established as a consequence of Proposition 6.3. ■

We have therefore realized that under the hypotheses of complete controllability and complete observability, a system of the form (6.1) is externally stable if and only if it is internally stable.

By a last effort, we can finally relax any restrictive assumption and formulate a general necessary and sufficient condition. Referring again to the Kalman decomposition and the related notation studied in Sect. 5.3.5, we consider the system

$$\begin{cases}\dot{z}_2 &= A_{22}z_2 + A_{24}z_4 + B_2u \\ \dot{z}_4 &= \qquad A_{44}z_4\end{cases} \tag{6.14}$$

with the observation function $y = C_2z_2 + C_4z_4$. By construction, (6.14) is completely observable.

Lemma 6.5 *System (6.1) is externally stable if and only if system (6.14) is externally stable.*

Proof First, we assume that system (6.14) possesses the BIBO-stability property. By Proposition 6.1, we may assume without loss of generality that (6.1) is in the Kalman canonical form

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + A_{13}z_3 + A_{14}z_4 + B_1 u \\ \dot{z}_2 = \qquad A_{22}z_2 \qquad + A_{24}z_4 + B_2 u \\ \dot{z}_3 = \qquad\qquad A_{33}z_3 + A_{34}z_4 \\ \dot{z}_4 = \qquad\qquad\qquad A_{44}z_4 \\ y \ = C_2 z_2 + C_4 z_4 \, . \end{cases}$$

Let $R > 0$. Focusing on system (6.1), we chose an initial state $\bar{z} = (\bar{z}_1, \bar{z}_2, \bar{z}_3, \bar{z}_4)$ and an admissible input $u(\cdot)$ such that

$$||\bar{z}|| < R \quad \text{and} \quad ||u(\cdot)||_\infty < R \, .$$

The components of the vector $(\bar{z}_2, \bar{z}_4)$ are a subset of the components of $\bar{z}$; hence we have

$$||(\bar{z}_2, \bar{z}_4)|| < R \, .$$

Then by hypothesis, there exists S such that $||y(t)|| < S$ for each $t \geq 0$. To conclude, it is sufficient to remark that (6.14) and (6.1) have the same output.

Vice versa, assume that (6.1) possesses the BIBO-stability property. Let us introduce R and consider the system (6.14). Chose an initial state $(\bar{z}_2, \bar{z}_4)$ and an admissible input $u(\cdot)$ such that

$$||(\bar{z}_2, \bar{z}_4)|| < R \quad \text{and} \quad ||u(\cdot)||_\infty < R \, .$$

Apply the same input to the system (6.1), choosing the initial condition $\bar{z} = (0, \bar{z}_2, 0, \bar{z}_4)$. Clearly $||\bar{z}|| = ||(\bar{z}_2, \bar{z}_4)||$. Hence, there exists S such that $||y(t)|| < S$ for each $t \geq 0$. Again, we get the conclusion noticing that the systems have the same output. ■

Lemma 6.6 *System (6.14) is BIBO-stable if and only if*

- *(i) all the eigenvalues of A_{22} have strictly negative real part;*
- *(ii) all the eigenvalues of A_{44} have nonpositive real part, and for each possible eigenvalue with zero real part, the algebraic and geometric multiplicities coincide.*

Proof Assume that the conditions (i) and (ii) hold. Matrix A_{44} defines a unforced system for which the origin is stable (in general, not asymptotically). In particular, the z_4 component of (6.14) remains bounded for $t \geq 0$. Let us consider $v = A_{24}z_4 + B_2 u$ as a new input. If $u(\cdot)$ is bounded, $v(\cdot)$ is bounded, as well. Since the subsystem

$$\dot{z}_2 = A_{22}z_2 + v \qquad (6.15)$$

is completely controllable, also z_2 remains bounded for $t \geq 0$. The conclusion is readily obtained.

Vice versa, assume that the system is BIBO-stable. Since it is completely observable, by Proposition 6.4, the matrix

$$\begin{pmatrix} A_{22} & A_{24} \\ 0 & A_{44} \end{pmatrix}$$

defines an unforced system for which the origin is stable (in general, not asymptotically). This implies in particular the validity of (ii). Condition (i) follows applying Proposition 6.5 to the completely controllable and completely observable subsystem

$$\begin{cases} \dot{z}_2 = A_{22}z_2 + B_1 u \\ y = C_2 z_2 \end{cases} \tag{6.16}$$

initialized to $(\bar{z}_2, 0)$. ■

The desired necessary and sufficient condition is finally obtained by combining Lemmas 6.5 and 6.6.

Theorem 6.4 *A system of the form (6.1) is BIBO-stable if and only if the following two conditions hold:*

(i) all the eigenvalues of the matrix A_{22} of the controllable and observable part of the system have strictly negative real part;
(ii) all the eigenvalues of the matrix A_{44} of the observable but not controllable part of the system have nonpositive real part, and for the possible eigenvalues of A_{44} with zero real part the algebraic and geometric multiplicities coincide.

Chapter Summary

In this chapter we address the study of the external stability. External stability is a natural consequence of the internal stability properties studied in Chap. 3. However, there are systems which are externally stable but not internally stable. Necessary and sufficient conditions for external stability are proven, in terms of the canonical decomposition introduced in Chap. 5.

Chapter 7
Stabilization

As already pointed out in Chap. 1, the behavior of a system can be regulated, without need of radical changes in its internal plant, by the construction of a suitable device which interacts with the system by means of a feedback connection. The action of such a device may have a *static* nature (and, in this case, it can be mathematically represented as a function) or a *dynamic* one (and so being interpreted as an auxiliary system). The feedback connection allows us to exert the control action in an automatic way (i.e., without need of the presence of a human operator), and requires the installation of sensors and actuators.

When all the state variables can be monitored and measured at each time, and all the information about their evolution can be used by the control device, we speak about *state* feedback. On the contrary, when the information about the state is only partially available (since they are, for instance, obtained by means of an observation function) we speak about *output* feedback.

7.1 Static State Feedback

In the static state feedback stabilization problem, the observation function is not involved. Hence, in this section we can limit ourselves to systems of the form

$$\dot{x} = Ax + Bu \ , \quad x \in \mathbf{R}^n \ , \ u \in \mathbf{R}^m \ . \tag{7.1}$$

First of all, we try to understand what happens when the feedback connection is implemented. Let $v(t)$ be an external signal, injected into the system through the input channel u, and let $x(t)$ be the solution representing the resulting state evolution. The feedback map $k(x) : \mathbf{R}^n \to \mathbf{R}^m$ generates another signal $w(t) = k(x(t))$. The signal actually received by the system is the sum of $v(t)$ and $w(t)$, that is $u = w(t) + v(t)$.

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_7

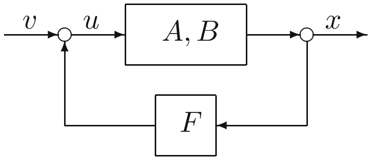
A static state feedback is said to be *linear* if $k(x) = Fx$, F being a $m \times n$ matrix. The implementation of a linear feedback can be mathematically interpreted as a substitution

$$u = Fx + v \tag{7.2}$$

which gives rise to a transformation of the system. Indeed, replacing (7.2) in (7.1), we obtain

$$\dot{x} = (A + BF)x + Bv \ , \tag{7.3}$$

the so-called *closed loop* system.



We stress that by virtue of the particular structure of the control (7.2), the transformed system is still of the form (7.1), with the matrix B unchanged, and the matrix A replaced by the new matrix $\tilde{A} = A + BF$. We also notice that the transformation induced by (7.2) is invertible; indeed, if we apply the feedback law $v = -Fx + u$ to (7.3) we recover the form (7.1) of the system. Thus, the transformation (7.2) defines an equivalence relation on the set of all the systems of the form (7.1); this fact can be formalized by the following definition.

Definition 7.1 We say that systems (7.1) and

$$\dot{x} = \tilde{A}x + Bu$$

are *feedback equivalent* if there exists a matrix F such that $\tilde{A} = A + BF$.

In this perspective, we can formulate the following problem pattern: assume that we are interested in a certain property, and that this property is not satisfied by the given system (7.1). We wonder whether the property is satisfied by system (7.3), for a suitable choice of the matrix F. More precisely, we want to find conditions under which the qualitative behavior of the given system can be modified in the desired way by means of a convenient feedback connection.

7.1.1 Controllability

As a first example, we ask whether a system can achieve the complete controllability property by means of a feedback transformation (diversely stated, whether in the same feedback equivalence class there may exist systems whose reachable spaces have different dimensions). The answer is negative; indeed, the following theorem holds.

Theorem 7.1 *For each matrix F, the reachable spaces of the systems (7.1) and (7.3), denoted here respectively by* $\mathrm{R}_{(7.1)}$ *and* $\mathrm{R}_{(7.3)}$*, coincide. As a consequence, we have*

$$\text{rank}\,(B|AB|\dots|A^{n-1}B) = \text{rank}\,(B|(A+BF)B|\dots|(A+BF)^{n-1}B)\,.$$

Proof According to Theorem 5.3

$$\mathrm{R}_{(7.1)} = \text{span}\,\{b^1,\dots,b^m;\ Ab^1,\dots,Ab^m;\ \dots;\ A^{n-1}b^1,\dots,A^{n-1}b^m\}\,.$$

Thus, $v \in \mathrm{R}_{(7.1)}$ if and only if v is a linear combination of the vectors

$$b^1,\dots,b^m;\ Ab^1,\dots,Ab^m;\ \dots;\ A^{n-1}b^1,\dots,A^{n-1}b^m\ .$$

Analogously,

$$\begin{aligned}\mathrm{R}_{(7.3)} = \text{span}\,\{&b^1,\dots,b^m;\ (A+BF)b^1,\dots,(A+BF)b^m;\\ &\dots;(A+BF)^{n-1}b^1,\dots,(A+BF)^{n-1}b^m\}\,.\end{aligned}$$

Notice that $(A+BF)b^j = Ab^j + BFb^j$. The vector Ab^j belongs to $\mathrm{R}_{(7.1)}$ and the vector

$$BFb^j = (b^1|\dots|b^m)Fb^j$$

belongs to $\mathrm{R}_{(7.1)}$ as well, since it is a linear combination of $b^1,\dots,b^m$. Continuing in this way, we notice that

$$(A+BF)^2b^j = (A+BF)(A+BF)b^j = A^2b^j + ABFb^j + BFAb^j + BFBFb^j\ .$$

The first term is in $\mathrm{R}_{(7.1)}$ by construction; the second because it is a linear combination of $Ab^1,\dots,Ab^m$; the third and the fourth term because they are linear combination of $b^1,\dots,b^m$. The same reasoning applies to each term of the form $(A+BF)^kb^j$.

In conclusion, $\mathrm{R}_{(7.3)} \subseteq \mathrm{R}_{(7.1)}$, since all the vectors of $\mathrm{R}_{(7.3)}$ are linear combinations of vectors of $\mathrm{R}_{(7.1)}$.

The opposite inclusion can be achieved by exchanging the roles of the systems (recall that (7.1) can be recovered from (7.3) by the inverse feedback transformation $v = -Fx + u$). ■

In other words, Theorem 7.1 states that the complete controllability property is invariant under feedback equivalence.

7.1.2 Stability

In the previous chapter we tried to characterize those systems of the form (7.1) which enjoy the external stability property. We noticed that this property is intimately linked to the internal stability properties of the system (Hurwitz property). This motivates the effort to elaborate models for which the eigenvalues of the system matrix A lie in the open left half of the complex plane and, in case this condition is not fulfilled, the interest in devising appropriate corrections.

The main purpose of this chapter is to show that feedback connections represent a convenient tool in order to improve the internal stability properties of a system.

7.1.3 Systems with Scalar Input

Consider first the case of a system with scalar input (i.e., with $m = 1$ and B reduced to a column vector b). Our approach is based on the following theorem.

Theorem 7.2 *Assume that $m = 1$, and that system (7.1) is completely controllable. Then, there exists a change of coordinates $x = P\zeta$ for which the system takes the form*

$$\dot{\zeta} = A_0\zeta + ub_0 \tag{7.4}$$

where A_0 is the companion matrix

$$\begin{pmatrix} 0 & 1 & 0 \dots & 0 \\ 0 & 0 & 1 \dots & 0 \\ \dots & \dots & \dots\dots & \dots \\ 0 & 0 & 0 \dots & 1 \\ -a_n & -a_{n-1} & \dots\dots & -a_1 \end{pmatrix} \quad \text{and} \quad b_0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Proof In Sects. 2.10 and 2.11 we saw that if there exists a cyclic vector for A (that is a vector v such that

$$v, Av, \dots, A^{n-1}v$$

are linearly independent), then A is similar to the matrix

$$\begin{pmatrix} 0 & 0 & \dots & 0 & -a_n \\ 1 & 0 & \dots & 0 & -a_{n-1} \\ 0 & 1 & \dots & 0 & -a_{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1 & -a_1 \end{pmatrix}$$

where the numbers $a_1, \dots, a_n$ are the coefficients of the characteristic polynomial of A, apart for a possible change of sign. Such a matrix is the transpose of the companion form.

The matrix associated to the change of coordinates is formed by the columns $v, Av, \dots, A^{n-1}v$. The complete controllability hypothesis states that the rank of the matrix

$$R = (b|Ab|\dots|A^{n-1}b)$$

is n. Hence, b is cyclic for A and $R^{-1}AR = A_0{}^{\mathbf{t}}$. Moreover,

$$R\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} = b \quad \text{that is} \quad R^{-1}b = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}.$$

We have so proved that our system is linearly equivalent to

$$\dot{w} = A_0^{\mathbf{t}}w + u\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}. \tag{7.5}$$

On the other hand, the matrix $Q = \big(b_0|A_0b_0|\dots|A_0^{n-1}b_0\big)$ has the form

$$\begin{pmatrix}0 & \dots\dots & 0 & 1\\ 0 & \dots\ 0 & 1 & *\\ \dots & \dots & \dots & \dots\\ 0 & 1 & * \ \dots & *\\ 1 & * & \dots\dots & *\end{pmatrix}$$

and so it is nonsingular (the stars stand for some numbers whose explicit expression is unessential). By the same arguments as before, we must have $Q^{-1}A_0Q = A_0^{\mathbf{t}}$ as well. Moreover,

$$Q\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} = b_0 \quad \text{that is} \quad Q^{-1}b_0 = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}.$$

Thus, the system (7.4) is linearly equivalent to (7.5), as well. Finally, (7.4) and the given system, being both linearly equivalent to (7.5), are equivalent each other. ■

Recall that the companion form characterizes the system representation of scalar linear differential equations. Theorem 7.2 states therefore that any completely controllable linear system with single input and state space dimension n is linearly equivalent to a system represented by a single linear differential equation of order n. We emphasize that the proof of Theorem 7.2 supplies an explicit expression for the matrix P which determines the similarity between A and its companion form A_0. Indeed, it is immediately seen that $P = RQ^{-1}$.

We rewrite for convenience system (7.4) as

$$\begin{cases} \dot{\zeta}_1 = \zeta_2 \\ \quad \vdots \\ \dot{\zeta}_{n-1} = \zeta_n \\ \\ \dot{\zeta}_n = -a_n\zeta_1 - \cdots - a_1\zeta_n + u \; . \end{cases} \tag{7.6}$$

Definition 7.2 A $2k$-tuple $\{\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_k\}$ is said to be *consistent* if:

(1) $1 \le k \le n$;
(2) $\lambda_1, \dots, \lambda_k$ are distinct complex numbers;
(3) $\mu_1, \dots, \mu_k$ are (not necessarily distinct) positive integers such that $\mu_1 + \cdots + \mu_k = n$;
(4) for each $i \in \{1, \dots, k\}$ there exists $j \in \{1, \dots, k\}$ such that $\lambda_j = \overline{\lambda_i}$ (the conjugate of λ_i) and $\mu_i = \mu_j$.

Given any consistent $2k$-tuple, it is easy to construct a monic polynomial with real coefficients

$$\lambda^n + b_1\lambda^{n-1} + \cdots + b_n$$

whose roots are exactly $\lambda_1, \dots, \lambda_k$, with respective multiplicities $\mu_1, \dots, \mu_k$. Now, let us apply to system (7.6) the feedback

$$u = (-b_n + a_n)\zeta_1 + \cdots + (-b_1 + a_1)\zeta_n + v \; . \tag{7.7}$$

The resulting closed-loop system is

$$\begin{cases} \dot{\zeta}_1 = \zeta_2 \\ \quad \vdots \\ \dot{\zeta}_n = -b_n\zeta_1 - \cdots - b_1\zeta_n + v \; . \end{cases}$$

Setting finally $v = 0$, we obtain an unforced linear system whose characteristic equation has exactly the roots $\lambda_1, \dots, \lambda_k$. Of course, the roots of the characteristic equation coincide with the eigenvalues of the matrix. If the numbers λ_i have been chosen in such a way that $\mathrm{Re}\lambda_i < 0$ for each i, we have obtained, by means of the feedback (7.7), a system for which the origin is asymptotically stable.

If fact, we have proven something more. For any preassigned real $n \times n$ matrix M, a completely controllable system with scalar input can be always transformed in a new system, such that the eigenvalues of the matrix of the new system coincide exactly with those of M.

7.1.3.1 System with Multiple Inputs

The discussion of the previous section motivates the following general definitions.

Definition 7.3 We say that (7.1) is *stabilizable* if there exists a static state feedback $u = Fx$ such that all the eigenvalues of the matrix $(A + BF)$ have negative real part.

We say that (7.1) is *superstabilizable* if for each $\alpha > 0$ there exists a static state feedback $u = Fx$ (with F dependent on α) such that the real part of each eigenvalue of the matrix $(A + BF)$ is less than $-\alpha$.

We say that (7.1) has the *pole assignment* property if for each given consistent $2k$-tuple there exists a static state feedback $u = Fx$ such that the eigenvalues of $A + BF$ are exactly the numbers $\lambda_1, \dots, \lambda_k$, with respective multiplicities $\mu_1, \dots, \mu_k$.

Systems which are superstabilizable are particularly interesting for applications. Indeed for these systems, it is not only possible to construct stabilizing feedback laws, but also to assign an arbitrary decay rate.

We already know that any completely controllable system with a scalar input possesses the pole assignment property, and hence it is stabilizable and superstabilizable. This result can be extended, with some technical complications in the proof, to systems with multiple input.

Theorem 7.3 *For any system of the form (7.1), the following properties are equivalent:*

(i) complete controllability
(ii) pole assignment
(iii) superstabilizability.

The reader interested in the full proof of Theorem 7.3 is referred, for instance, to [11], p. 145 or [28], p. 58. It follows in particular from Theorem 7.3 that for any system in the general form (7.1), complete controllability implies stabilizability by static state feedback. We give below an independent and direct proof of this fact.

Proposition 7.1 *If (7.1) is completely controllable, then it is stabilizable.*

Proof The completely controllability assumption amounts to say that for each $T > 0$ the matrix

$$\Gamma(T) = \int_0^T e^{-\tau A} B B^{\mathbf{t}} e^{-\tau A^{\mathbf{t}}} \, d\tau$$

is positive definite (Theorem 5.2). Write for simplicity $\Gamma = \Gamma(1)$. Let us show that the feedback law $u = Fx = -B^{\mathbf{t}}\Gamma^{-1}x$ actually stabilizes the system. Compute the derivative

$$\frac{d}{dt}\left(e^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}}\right) = -Ae^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}} - e^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}}A^{\mathbf{t}}$$

which yields

$$\int_0^1 \frac{d}{dt}\left(e^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}}\right)dt = \int_0^1 \Big(-Ae^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}} - e^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}}A^{\mathbf{t}}\Big)\,dt\,. \tag{7.8}$$

Clearly, the expression on the left hand side can be rewritten as

$$e^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}}\Big|_0^1 = e^{-A}BB^{\mathbf{t}}e^{-A^{\mathbf{t}}} - BB^{\mathbf{t}}\,.$$

By the definition of Γ, the right hand side of (7.8) is equal to

$$-A\int_0^1 e^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}}\,dt - \int_0^1 e^{-tA}BB^{\mathbf{t}}e^{-tA^{\mathbf{t}}}\,dt A^{\mathbf{t}} = -A\Gamma - \Gamma A^{\mathbf{t}}\,.$$

Hence,

$$e^{-A}BB^{\mathbf{t}}e^{-A^{\mathbf{t}}} - BB^{\mathbf{t}} = -A\Gamma - \Gamma A^{\mathbf{t}}\,. \tag{7.9}$$

On the other hand, Γ being a symmetric matrix,

$$(A - BB^{\mathbf{t}}\Gamma^{-1})\Gamma + \Gamma(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}} = A\Gamma + \Gamma A^{\mathbf{t}} - 2BB^{\mathbf{t}}\,. \tag{7.10}$$

From (7.9) and (7.10) we infer

$$(A - BB^{\mathbf{t}}\Gamma^{-1})\Gamma + \Gamma(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}} = -e^{-A}BB^{\mathbf{t}}e^{-A^{\mathbf{t}}} - BB^{\mathbf{t}}\,. \tag{7.11}$$

The matrix at the right hand side is (at least) negative semidefinite. According to Theorem 3.4, we can conclude that the origin is stable for the system

$$\dot{x} = (A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}}x \tag{7.12}$$

and so also for the system

$$\dot{x} = (A - BB^{\mathbf{t}}\Gamma^{-1})x \tag{7.13}$$

since any square matrix has the same eigenvalues as its transpose. However, on the base of (7.11), we are not able to conclude that (7.13) is asymptotically stable: there are indeed simple examples of completely controllable linear systems for which the matrix at the right hand side of (7.11) is actually not positive definite.[1] In other words, we cannot be sure that $V(x) = x^{\mathbf{t}}\Gamma x$ is a strict Lyapunov function for (7.12).

To finish the proof, we need therefore to try another way. We will resort directly to Theorem 3.1. More precisely, we will show that all the eigenvalues of $(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}}$ have strictly negative real part. To this end, we take advantage of the previous computations.

Let λ be an eigenvalue (real or complex) of $(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}}$, and let $v \neq 0$ be a corresponding eigenvector. We have

$$(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}} v = \lambda v\ . \tag{7.14}$$

From (7.11) we obtain

$$\bar{v}^{\mathbf{t}}[(A - BB^{\mathbf{t}}\Gamma^{-1})\Gamma + \Gamma(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}}]v = -\bar{v}^{\mathbf{t}}[e^{-A}BB^{\mathbf{t}}e^{-A^{\mathbf{t}}} + BB^{\mathbf{t}}]v\ . \tag{7.15}$$

On the other hand,

$$\begin{aligned} &\bar{v}^{\mathbf{t}}[(A - BB^{\mathbf{t}}\Gamma^{-1})\Gamma + \Gamma(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}}]v \\ &= \bar{\lambda}\bar{v}^{\mathbf{t}}\Gamma v + \lambda\bar{v}^{\mathbf{t}}\Gamma v = (\bar{\lambda} + \lambda)\bar{v}^{\mathbf{t}}\Gamma v = 2\,\mathrm{Re}\,\lambda(\bar{v}^{\mathbf{t}}\Gamma v)\ . \end{aligned} \tag{7.16}$$

Hence,

$$\bar{v}^{\mathbf{t}}[e^{-A}BB^{\mathbf{t}}e^{-A^{\mathbf{t}}} + BB^{\mathbf{t}}]v = -2\,\mathrm{Re}\,\lambda(\bar{v}^{\mathbf{t}}\Gamma v)\ . \tag{7.17}$$

Since Γ is real and positive definite, we can easily check that $\bar{v}^{\mathbf{t}}\Gamma v > 0$ (notice the analogies between this argument and the computation in the proof of Theorem 3.3). We recover in this way the previous conclusion that $\mathrm{Re}\,\lambda \leq 0$. Now, if it happens that $\mathrm{Re}\,\lambda = 0$ for some λ, then we should also have that

$$\bar{v}^{\mathbf{t}}[e^{-A}BB^{\mathbf{t}}e^{-A^{\mathbf{t}}} + BB^{\mathbf{t}}]v = 0$$

and so in particular

$$\bar{v}^{\mathbf{t}}BB^{\mathbf{t}}v = \bar{v}^{\mathbf{t}}\bar{B}B^{\mathbf{t}}v = 0\ .$$

This implies that $B^{\mathbf{t}}v = 0$. It follows

$$(A - BB^{\mathbf{t}}\Gamma^{-1})^{\mathbf{t}}v = A^{\mathbf{t}}v - (\Gamma^{-1})^{\mathbf{t}}BB^{\mathbf{t}}v = A^{\mathbf{t}}v\ . \tag{7.18}$$

Comparing (7.14) and (7.18), we conclude that v is also an eigenvector of $A^{\mathbf{t}}$ corresponding to the same eigenvalue λ. But then $e^{-tA^{\mathbf{t}}}v = e^{-\lambda t}v$. Finally we get

---

[1] One such example can be obtained taking $A = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$, $b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

$$\bar{v}^{\mathbf{t}}\Gamma v = \int_0^1 \bar{v}^{\mathbf{t}} e^{-tA} B B^{\mathbf{t}} e^{-tA^{\mathbf{t}}} v \, dt = \int_0^1 e^{-\lambda t} \bar{v}^{\mathbf{t}} e^{-tA} B B^{\mathbf{t}} v \, dt = 0$$

which is impossible since Γ is positive definite and $v \neq 0$. ■

7.1.4 Stabilizability

The stabilizability property is actually weaker than complete controllability; this can be easily realized looking at a system for which A is Hurwitz and $B = 0$. In this section we aim to characterize the stabilizability property by means of suitable and easy-to-check conditions.

To this end, it is convenient to apply a preliminary change of coordinates in order to put the system in the controllability canonical form. In other words, without loss of generality, we can assume for our system the form

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + B_1 u \\ \dot{z}_2 = \qquad\qquad A_{22}z_2 \end{cases} \tag{7.19}$$

where the subsystem corresponding to the block of coordinates z_1 is completely controllable.

Theorem 7.4 *System (7.19) is stabilizable if and only if the matrix A_{22} is Hurwitz.*

Proof The set of the eigenvalues of a matrix $A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$ is the union of the sets of the eigenvalues of the matrices A_{11} and A_{22}. By means of a feedback, there is no way to modify the eigenvalues of the matrix A_{22}. Hence, if the system is stabilizable, A_{22} must be Hurwitz.

Vice versa, since the subsystem corresponding to the components z_1 is completely controllable, we can construct a feedback $u = F_1 z_1$ (for instance, by the method illustrated in the proof of Proposition 7.1) in such a way that matrix $A_{11} + B_1 F_1$ is Hurwitz. The matrix A_{22} is Hurwitz by hypothesis. Hence, the matrix A is Hurwitz, as well. ■

Next we present (without proof) other necessary and sufficient conditions for stabilization.

Theorem 7.5 *Let V be the subspace of $\mathbf{C}^n$ generated by all the eigenvectors (including the generalized ones) associated to all the eigenvalues λ of A, having nonnegative real part. Moreover, let U be the subspace of $\mathbf{R}^n$ generated by all the vectors of the form* Re v *and* im v*, with $v \in V$. System (7.1) is stabilizable if and only if U is contained in its reachable space* R.

Theorem 7.6 *System (7.1) is stabilizable if and only if*

$$\text{rank}\,(A - \lambda I \mid B) = n$$

for every complex number λ with nonnegative real part.

It is interesting to compare Theorem 7.6 and Hautus' controllability criterion (Theorem 5.4).

Theorem 7.7 *System (7.1) is stabilizable if and only if there exists a symmetric, positive definite matrix P such that*

$$A^{\mathbf{t}}P + PA - PBB^{\mathbf{t}}P = -I\ . \tag{7.20}$$

In this case, a stabilizing feedback can be found of the form $u = Fx$, with $F = -\alpha B^{\mathbf{t}}P$ and $\alpha \geq \frac{1}{2}$.

It is not difficult to see that (7.20) is sufficient for stabilizability. Indeed, replacing the feedback $F = -\alpha B^{\mathbf{t}}P$ into the equation we obtain the closed-loop system in the form

$$\dot{x} = Ax - \alpha BB^{\mathbf{t}}Px\ . \tag{7.21}$$

Taking into account (7.20), we have

$$\begin{aligned}(A - \alpha BB^{\mathbf{t}}P)^{\mathbf{t}}P + P(A - \alpha BB^{\mathbf{t}}P) &= A^{\mathbf{t}}P + PA - 2\alpha PBB^{\mathbf{t}}P \\ &= -I + (1 - 2\alpha)PBB^{\mathbf{t}}P\ .\end{aligned}$$

The matrix $PBB^{\mathbf{t}}P = (B^{\mathbf{t}}P)^{\mathbf{t}}B^{\mathbf{t}}P$ corresponds to a positive semidefinite quadratic form. Hence, if $\alpha \geq \frac{1}{2}$, P solves a Lyapunov matrix equation for system (7.21). The conclusion follows by Corollary 3.1.

On the contrary, proving that the same condition is also necessary for stabilizability is more difficult (a proof can be found in [11], p. 133).

Equation (7.20) is called the *algebraic Riccati matrix equation* associated to system (7.1). We emphasize that (7.20) is nonlinear with respect to the entries of the unknown matrix P. We emphasize also that, once a solution of (7.20) has been found, the feedback law provided by Theorem 7.7 is explicit and simpler than the feedback provided in the proof of Proposition 7.1.

Corollary 7.1 *If there exists a symmetric, positive definite matrix Q such that the matrix equation*

$$A^{\mathbf{t}}P + PA - PBB^{\mathbf{t}}P = -Q \tag{7.22}$$

admits a symmetric, positive definite solution P, then system (7.1) is stabilizable.

On the other hand, if system (7.1) is stabilizable, then for each symmetric, positive definite matrix Q there exists a symmetric, positive definite solution P of the matrix equation (7.22).

Proof The proof of the first statement is similar to the proof of the sufficient part of Theorem 7.7. In order to prove the second statement, we write $Q = R^{\mathbf{t}}R$, with R nonsingular and symmetric. If the feedback $u = Fx$ stabilizes the given system, then the system

$$\dot{x} = \tilde{A}x + \tilde{B}u$$

where $\tilde{A} = RAR^{-1}$ and $\tilde{B} = RB$, is stabilized by the feedback $u = FR^{-1}$. Hence, according to the necessary part of Theorem 7.7, there must exists a symmetric, positive definite matrix $\tilde{P}$ such that

$$\tilde{A}^{\mathbf{t}}\tilde{P} + \tilde{P}\tilde{A} - \tilde{P}\tilde{B}\tilde{B}^{\mathbf{t}}\tilde{P} = -I \ .$$

The remaining part of the proof is similar to that of Corollary 3.1. ■

7.1.5 Asymptotic Controllability

If system (7.1) is stabilizable by means of a feedback $u = Fx$, then for each initial state $x_0 \in \mathbf{R}^n$ we can consider the solution $x(t, x_0)$ of the problem

$$\begin{cases} \dot{x} = (A + BF)x \\ x(0) = x_0 \ . \end{cases}$$

This solution $x(t, x_0)$ obviously coincides with the solution $x(t)$ of the problem

$$\begin{cases} \dot{x} = Ax + Bu_{x_0}(t) \\ x(0) = x_0 \end{cases} \tag{7.23}$$

where $u_{x_0}(t) = Fx(t, x_0)$.

Definition 7.4 We say that the system (7.1) is *asymptotically controllable* if for each $x_0 \in \mathbf{R}^n$ there exists an input map $u_{x_0}(t)$ such that the corresponding solution $x(t)$ of the problem (7.23) approaches the origin for $t \to +\infty$.

The previous reasoning shows that a stabilizable system is asymptotically controllable. But also the converse is true. Indeed, by an argument similar to that used in the proof of Theorem 7.4, it is not difficult to see that if system (7.1) is asymptotically controllable, then the uncontrollable part of the associated canonical controllability form must be asymptotically stable. Then, the conclusion follows in force of Theorem 7.4. We can therefore state a further necessary and sufficient condition for stabilizability.

Theorem 7.8 *System (7.1) is stabilizable if and only if it is asymptotically controllable.*

7.2 Static Output Feedback

In the previous section we studied the problem of stabilizability by means of a static state feedback. Obviously, this way is not feasible when the whole state of the system is not available. This usually happens when we deal with a system with an observation function

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \tag{7.24}$$

($x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$, $y \in \mathbf{R}^p$), and the requested feedback law has of the form $u = Ky$. The problem of stabilization by means of output feedback is more natural in view of the applications, but also much more difficult to study.

To become familiar with these new difficulties, we examine some simple examples.

Example 7.1 The two dimensional system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u \end{cases} \tag{7.25}$$

is completely controllable, and hence stabilizable by means of a feedback of the form $u = k_1x_1 + k_2x_2$. In fact, provided that the parameters k_1 and k_2 can chosen freely and independently each other, the system is superstabilizable.

However, it is not possible to stabilize the system if the choice is limited to feedback laws of the form $u = kx_1$. Indeed, by applying such a feedback, the system becomes

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = kx_1 \end{cases}$$

whose eigenvalues are both either on the imaginary axis, or on the real axis and, in this second case, they have opposite sign. In other words, since now only one parameter can be arbitrarily chosen, we do not have the degrees of freedom necessary to solve the problem.

As already suggested, the impossibility of implementing a feedback which uses all the state variables typically arises when we have an observation function. In this example, the feedback $u = kx_1$ can be interpreted as an output feedback, if we assume an observation function $y = c^{\mathbf{t}}x$ with $c = (1\ 0)$. We emphasize that the system, with respect to this observation function, is completely observable, as well; nevertheless, the system is not stabilizable by an output feedback. ■

Example 7.2 Consider again the system (7.25), but this time with the observation function $y = x_1 + x_2$. By applying the feedback $u = -ky = -k(x_1 + x_2)$, we obtain the system

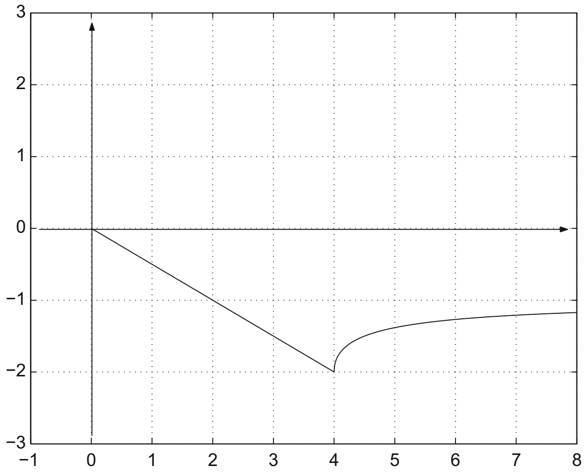


Fig. 7.1 Graph of Re $(\lambda_2(k))$

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -kx_1 - kx_2 \end{cases}$$

whose characteristic equation is $\lambda^2 + k\lambda + k = 0$. Thus, the system is stabilized if $k > 0$. However, the system is not superstabilizable. Indeed, if $0 < k < 4$ the characteristic roots are not real. Their real part is $\operatorname{Re}\lambda = -k/2$; it decreases when k increases, and

$$\lim_{k \to 4^-} \operatorname{Re}\lambda = -2 .$$

If $k = 4$ we have $\lambda_1 = \lambda_2 = -2$. If $k > 4$, the characteristic roots are real and can be represented as

$$\lambda_1 = \frac{-k - \sqrt{k^2 - 4k}}{2} \quad \text{and} \quad \lambda_2 = \frac{-k + \sqrt{k^2 - 4k}}{2} .$$

Clearly $\lambda_2 > \lambda_1$: moreover, λ_2 increases when k increases, and

$$\lim_{k \to 4^+} \lambda_2 = -2 , \qquad \lim_{k \to +\infty} \lambda_2 = -1 .$$

In conclusion, $\operatorname{Re}\lambda_2(k) \geq -2$ for each $k \geq 0$, and it attains the minimum for $k = -2$ (the graph of the real part of λ_2 as a function of k is shown in Fig. 7.1).

Notice that again in this example, the system is completely controllable and completely observable. ■

Of course, if a system is stabilizable by a static output feedback, it is stabilizable also by a static state feedback. Hence, all the static state feedback stabilizability conditions listed in the previous section can be reviewed as necessary (but no more sufficient) conditions for static output feedback stabilizability.

7.2.1 Reduction of Dimension

In this section we present a theorem which allows us to simplify, under particular conditions, the study of the static output feedback stabilization problem.

Given a system (7.24), we begin by applying a linear change of coordinates, with the purpose of rewriting the matrices A, B, C in Kalman's canonical form (5.35), that is

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + A_{13}z_3 + A_{14}z_4 + B_1u \\ \dot{z}_2 = \qquad\quad A_{22}z_2 \qquad\quad + A_{24}z_4 + B_2u \\ \dot{z}_3 = \qquad\qquad\qquad\quad A_{33}z_3 + A_{34}z_4 \\ \dot{z}_4 = \qquad\qquad\qquad\qquad\qquad A_{44}z_4 \\ y = C_2z_2 + C_4z_4\ . \end{cases} \tag{7.26}$$

Recall that according to the usual notation, the controllable part of the system is identified by the indices 1 and 2, while the completely observable part is identified by the indices 2 and 4.

Theorem 7.9 *The overall system (7.24) is stabilizable by static output feedback if and only if the following conditions are both satisfied:*

(1) the matrices A_{11}, A_{33} and A_{44} have all the eigenvalues with negative real part;
(2) the completely controllable and completely observable part of the system, that is the part corresponding to the subsystem

$$\begin{cases} \dot{z}_2 = A_{22}z_2 + B_2u \\ y = C_2z_2 \end{cases} \tag{7.27}$$

is stabilizable by static output feedback.

Proof We start by a preliminary remark. Let K be a matrix with p columns and m rows. If we apply the feedback $u = Ky = KC_2z_2 + KC_4z_4$ to system (7.26), the system matrix becomes

$$\begin{pmatrix} A_{11} & \tilde{A}_{12} & A_{13} & \tilde{A}_{14} \\ 0 & \tilde{A}_{22} & 0 & \tilde{A}_{24} \\ 0 & 0 & A_{33} & A_{34} \\ 0 & 0 & 0 & A_{44} \end{pmatrix} \tag{7.28}$$

where $\tilde{A}_{12} = A_{12} + B_1KC_2$, $\tilde{A}_{14} = A_{14} + B_1KC_4$, $\tilde{A}_{22} = A_{22} + B_2KC_2$, $\tilde{A}_{24} = A_{24} + B_2KC_4$.

By virtue of the triangular block form of (7.28), it is clear that the feedback $u = Ky$ stabilizes the system if and only if the matrices A_{11}, $\tilde{A}_{22}$, A_{33} and A_{44} have all their eigenvalues with negative real part. Taking into account of condition (1), this actually happens if and only if the feedback $u = Ky = KC_2z_2$ stabilizes the reduced order system (7.27). ■

In view of Theorem 7.9, as far as we are interested in the static output feedback stabilization problem, it is not restrictive to assume that the system at hand is completely controllable as well as completely observable. Then, the following sufficient condition may be of some help.

Proposition 7.2 *Let the system (7.24) be given. Assume that it is completely controllable, and that the matrix C is invertible. Then, the system is stabilizable by a static output feedback.*

Proof By the complete controllability hypothesis, there exists a matrix K such that the system is stabilizable by a static state feedback $u = Kx$. We can write $u = KC^{-1}Cx = KC^{-1}y$. We obtain in this way a static output feedback $u = Fy$ with $F = KC^{-1}$ whose effect on the system is the desired one. ■

In the previous statement, the assumption that C is invertible implies of course that $p = n$ and that the system is completely observable, as well.

Example 7.3 Consider the system

$$\begin{cases} \dot{x}_1 = x_1 + 4x_2 \\ \dot{x}_2 = 2x_1 - 6x_2 + u \\ y = 2x_1 + x_2 \ . \end{cases}$$

It is clear that it is completely controllable, but not completely observable. Using the change of coordinates $x = Pz$ where P is defined as

$$P = \begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix}$$

(according to the method explained in Sect. 5.3) we recover the observability canonical form

$$\begin{cases} \dot{z}_1 = -7z_1 - 2z_2 + \frac{2}{5}u \\ \dot{z}_2 = 2z_2 + \frac{1}{5}u \\ y = 5z_2 \ . \end{cases}$$

The unobservable part (first equation), when we set $u = 0$, is asymptotically stable. By Theorem 7.9, now it is clear that the system, in the new coordinates, is stabilizable by means of a static output feedback $u = ky$. Convenient values for the parameter k can be determined by direct computation: we find $k < -2$. Coming back to the original coordinates, the feedback to be applied is $u = ky = k(2x_1 + x_2)$ (again, $k < -2$). ■

As suggested by the previous example, once the reduction of dimension has been performed, if the dimension of the completely controllable and observable part turns out to be small, the existence of static output stabilizers can be checked by direct computation. An other example is given below.

Example 7.4 Consider a two-dimensional, completely controllable and completely observable system (in controllability canonical form)

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -a_2 x_1 - a_1 x_2 + u \\ y = c_0 x_1 + c_1 x_2 \ . \end{cases} \tag{7.29}$$

The complete observability assumption amount to say that $c_0^2 - c_0 c_1 a_1 + c_1^2 a_2 \neq 0$, which in turn implies that c_0 and c_1 cannot be both zero. By the substitution $u = ky$, the system matrix takes the form

$$\begin{pmatrix} 0 & 1 \\ -a_2 + kc_0 & -a_1 + kc_1 \end{pmatrix}$$

whose characteristic equation is

$$\lambda^2 - \lambda(-a_1 + kc_1) + a_2 - kc_0 = 0 \ .$$

By the Routh-Hurwitz criterion, it is easy to see that the system is stabilizable with static output feedback if and only if there exists $k \in \mathbf{R}$ such that $kc_0 < a_2$, and $kc_1 < a_1$. ■

The following statement is a dual version of Proposition 7.2.

Proposition 7.3 *Let the system (7.24) with $m = n$ be given. Assume that it is completely observable, and that the matrix B is invertible. Then, the system is stabilizable by a static output feedback.*

Proof The observability assumption about (7.24) implies that the dual system

$$\begin{cases} \dot{x} = A^{\mathbf{t}} x + C^{\mathbf{t}} v \\ y = B^{\mathbf{t}} x \end{cases}$$

is completely controllable. Hence, there is a matrix K such that $A^{\mathbf{t}} + C^{\mathbf{t}} K$ is Hurwitz. The matrix $A + K^{\mathbf{t}} C$ is Hurwitz, as well. Thus, if we apply to (7.24) the feedback law $u = B^{-1} K^{\mathbf{t}} y$, then the matrix of the closed loop system is $A + BB^{-1} K^{\mathbf{t}} C = A + K^{\mathbf{t}} C$. The statement is proven. ■

7.2.2 Systems with Stable Zero Dynamics

Concerning systems represented by Eq. (7.26), there is a further interesting remark. For a moment, let us limit ourselves to look at the differential part of the system, operated in open loop. The solution corresponding to an initial state of the form $\bar{z} = (\bar{z}_1, 0, \bar{z}_3, 0)$ and a vanishing input $u = 0$ evolves inside the subspace of equations $z_2 = z_4 = 0$, which is therefore invariant (in the sense of Definition A.4) for the unforced system. We point out that each solution lying in this subspace gives rise to a vanishing output. For this reason, the subspace of equations $z_2 = z_4 = 0$ is called the *space of the zero dynamics*, and the system

$$\begin{cases} \dot{z}_1 = A_{11} z_1 + A_{13} z_3 \\ \dot{z}_3 = \qquad\quad A_{33} z_3 \end{cases}$$

is called the *system of the zero dynamics*. The following statement is a straightforward consequence of Theorem 7.9 and Proposition 7.2.

Corollary 7.2 *Let the dimension of the observable but not controllable part of the system (7.24) be zero. Assume in addition that the matrix C_2 is invertible. Then, the system is stabilizable by static output feedback if and only if the origin is asymptotically stable for the system of the zero dynamics.*

7.2.3 A Generalized Matrix Equation

Other sufficient conditions for static output stabilization can be obtained by suitable generalizations of the Riccati matrix equation (7.20). Next we present one such generalization.

Let C be a matrix with p rows and n columns. A *generalized inverse* (or *pseudoinverse*) of C is any matrix $C^\dagger$ with n rows and p columns such that

$$CC^\dagger C = C \quad \text{and} \quad C^\dagger C C^\dagger = C^\dagger \ .$$

If in addition we require that $CC^\dagger$ is symmetric, then $C^\dagger$ is uniquely determined and it is called the *Moore-Penrose generalized inverse* of C (see [4] for properties of generalized inverse matrices). The space $\mathbf{R}^n$ can be decomposed as $\text{im}\,(C^{\mathbf{t}}) \oplus \ker\,(C)$. Moreover, the subspaces $\text{im}\,(C^{\mathbf{t}})$ and $\ker\,(C)$ are orthogonal each other. The square $n \times n$ matrix $E_{\text{im}} = C^\dagger C$ represents the orthogonal projection on $\text{im}\,(C^{\mathbf{t}})$, while the orthogonal projection on $\ker\,(C)$ is given by $E_{\text{ker}} = I - E_{\text{im}}$. The matrices E_{im} and E_{ker} are uniquely determined and symmetric.

Theorem 7.10 *Consider the system (7.24), and assume that there exist symmetric and positive definite matrices P and Q such that*

$$A^{\mathbf{t}} P + PA - E_{im} PBB^{\mathbf{t}} PE_{im} + E_{\text{ker}} PBB^{\mathbf{t}} PE_{\text{ker}} = -Q \ . \tag{7.30}$$

Then, the system is stabilizable by the static output feedback $u = Ky$*, with* $K = -B^{\mathbf{t}} P C^{\dagger}$.

Before proving the theorem, we point out the following matrix identity:

$$E_{\mathrm{im}} P B B^{\mathbf{t}} P E_{\mathrm{im}} - E_{\mathrm{ker}} P B B^{\mathbf{t}} P E_{\mathrm{ker}} + P B B^{\mathbf{t}} P \\ = E_{\mathrm{im}} P B B^{\mathbf{t}} P + P B B^{\mathbf{t}} P E_{\mathrm{im}} \tag{7.31}$$

which can be easily checked taking into account that $E_{\mathrm{ker}} = I - E_{\mathrm{im}}$.

Proof of Theorem 7.10 Applying the static output feedback $u = -B^{\mathbf{t}} P C^{\dagger} y$, the closed-loop system takes the form

$$\dot{x} = (A - B B^{\mathbf{t}} P C^{\dagger} C) x \ .$$

We show that, by virtue of (7.30) and (7.31), P solves the Lyapunov matrix equation for this system. Indeed, we have:

$$\begin{aligned} & (A - B B^{\mathbf{t}} P C^{\dagger} C)^{\mathbf{t}} P + P (A - B B^{\mathbf{t}} P C^{\dagger} C) \\ &= A^{\mathbf{t}} P + P A - [E_{\mathrm{im}} P B B^{\mathbf{t}} P + P B B^{\mathbf{t}} P E_{\mathrm{im}}] \\ &= A^{\mathbf{t}} P + P A - E_{\mathrm{im}} P B B^{\mathbf{t}} P E_{\mathrm{im}} + E_{\mathrm{ker}} P B B^{\mathbf{t}} P E_{\mathrm{ker}} - P B B^{\mathbf{t}} P \\ &= -Q - P B B^{\mathbf{t}} P \ . \end{aligned}$$

The conclusion follows by Theorem 3.3, since the matrix $P B B^{\mathbf{t}} P$ is clearly positive semidefinite. ∎

Notice that (7.30) reduces to (7.20) when $C = I$.

Example 7.5 The condition of Theorem 7.10 works, for the case considered in the previous Example 7.2. Since $C = (1 \ 1)$, as a generalized inverse we can take for instance

$$C^{\dagger} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

so that

$$E_{\mathrm{im}} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad \text{and} \quad E_{\mathrm{ker}} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} .$$

Writing

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix} \tag{7.32}$$

the left-hand side of (7.30) takes the form

$$\begin{pmatrix} -p_{12} p_{22} & p_{11} - \dfrac{p_{12}^2 + p_{22}^2}{2} \\ p_{11} - \dfrac{p_{12}^2 + p_{22}^2}{2} & 2p_{12} - p_{12} p_{22} \end{pmatrix} .$$

A solution of (7.30) with the required properties is obtained taking $p_{11} = 6$, $p_{22} = 3$, $p_{12} = 2$. The corresponding output feedback is $u = -\frac{5}{2}y$. ■

Note that assuming $Q = I$ in (7.30) would be restrictive, contrary to what happens in the case of the Lyapunov matrix equation (see Theorem 3.3 and Corollary 3.1) and in the case of the Riccati matrix equation. For instance, it is not difficult to check that in the previous example, there is no solutions for (7.30) if we set $Q = I$.

Example 7.6 The condition of Theorem 7.10 is not necessary for the existence of an output feedback stabilizer. Consider the two-dimensional system with

$$A = \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad C = (1\ 0)\ .$$

Clearly, this time we have

$$E_{\text{im}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad E_{\text{ker}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\ .$$

Assume that (7.30) has a positive definite solution P, that we write again in the form (7.32). Then it would be possible to construct an output stabilizer of the form $u = -B^{\mathbf{t}}PC^{\dagger}y$, which reduces in this case to $u = -p_{12}y$. By direct substitution, it is easily seen that the system can be actually stabilized by an output feedback of this form, provided that $p_{12} > 2$.

On the other hand, the left-hand side of (7.30) takes now the form

$$\begin{pmatrix} 2p_{11} - p_{12}^2 & p_{11} - p_{12} \\ p_{11} - p_{12} & 2p_{12} - 4p_{22} + p_{22}^2 \end{pmatrix}\ .$$

This matrix is definite negative only if the term $2p_{12} - 4p_{22} + p_{22}^2$ is negative: as easily seen, this requires that $p_{12} < 2$.

In conclusion, the system is stabilizable by an output feedback, but the coefficient of the feedback cannot be determined on the base of Theorem 7.10. ■

7.2.4 A Necessary and Sufficient Condition

The static output feedback stabilization problem is sometimes referred to in the literature as an unsolved problem. From a practical point of view, a numerical solution of a nonlinear matrix equation like (7.30) can be indeed very hard to find. On the contrary, theoretical characterizations of systems admitting static output stabilizing feedbacks expressed in the form of nonlinear generalized Riccati equations can be actually found in the existing literature. For instance, the following theorem appears in [29].

Theorem 7.11 *System (7.24) is stabilizable by a static output linear feedback if and only if there exist symmetric, positive definite $n \times n$ matrices P, Q and a matrix M (with m rows and n columns) such that*

$$A^{\mathbf{t}}P + PA - E_{im}M^{\mathbf{t}}B^{\mathbf{t}}P - PBME_{im} = -Q \, . \tag{7.33}$$

Moreover, when (7.33) holds, a stabilizing feedback can be taken of the form $u = Ky$, with $K = -MC^{\dagger}$.

Proof To prove the necessity of the condition, let us assume that the system is stabilized by a feedback of the form $u = Ky$ for some matrix K. Then we have also the static state stabilizer

$$u = KCx = KCC^{\dagger}Cx = -MC^{\dagger}Cx = -ME_{\mathrm{im}}x$$

where we used the definition of generalized inverse and we set $M = -KC$. Hence, the closed loop system must satisfy the Lyapunov matrix equation (Theorem 3.3) for some symmetric, positive definite matrix P; namely

$$\begin{aligned}&(A - BME_{\mathrm{im}})^{\mathbf{t}}P + P(A - BME_{\mathrm{im}})\\ &= A^{\mathbf{t}}P + PA - E_{\mathrm{im}}M^{\mathbf{t}}B^{\mathbf{t}}P - PBME_{\mathrm{im}} = -I\end{aligned}$$

which is (7.33) with $Q = I$. As far as the sufficiency is concerned, assuming that (7.33) holds for some matrices P, Q, M, we apply the static output feedback $u = -MC^{\dagger}y = -ME_{\mathrm{im}}x$. The conclusion can be easily achieved by repeating the same computation as before and using the Lyapunov matrix equation as a sufficient condition. ■

Remark 7.1 In [29], condition (7.33) is written in a different, but equivalent, way: indeed, the authors do not use the formalism of generalized inverse. ■

Remark 7.2 It is not difficult to see that (7.30) implies (7.33), setting $M = B^{\mathbf{t}}P$, and using the matrix identity (7.31). However, we notice that with respect to (7.30), the matrix equation (7.33) contains the additional unknown M. ■

Remark 7.3 Another necessary and sufficient condition for the existence of static output stabilizing feedbacks was given in [9]. Reformulated in terms of generalized inverse, this condition reads: there exist matrices P, Q, M (of the same dimensions as before) such that

$$A^{\mathbf{t}}P + PA - PBB^{\mathbf{t}}P + (B^{\mathbf{t}}P - ME_{\mathrm{im}})^{\mathbf{t}}(B^{\mathbf{t}}P - ME_{\mathrm{im}}) = -Q \, . \tag{7.34}$$

Of course, (7.34) is equivalent to (7.33), but not with the same P and Q, in general. ■

7.3 Dynamic Output Feedback

The practical difficulties encountered in the static output stabilization problem can be overcome resorting to a different approach, provided that the system is, in principle, stabilizable by means of a static state feedback law and a suitable (but natural) technical condition is met. The new approach we are going to describe in this section is *dynamic output feedback*.

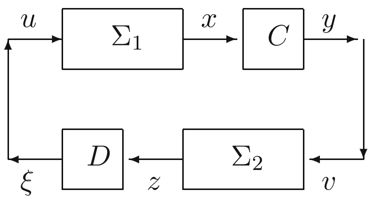
Definition 7.5 We say that system (7.24) is *stabilizable by dynamic output feedback* if there exists a system

$$\begin{cases} \dot{z} = Fz + Gv \\ \xi = Dz \end{cases} \tag{7.35}$$

with $z \in \mathbf{R}^{\nu}$, $v \in \mathbf{R}^{p}$, $\xi \in \mathbf{R}^{m}$ such that the composed system obtained by means of the substitutions $u = \xi$ and $v = y$

$$\begin{cases} \dot{x} = Ax + BDz \\ \dot{z} = Fz + GCx \end{cases}$$

has an asymptotically stable equilibrium position at the origin $(x, z) = (0, 0) \in \mathbf{R}^{n+\nu}$. The system (7.35) represents the *compensator*, or *controller*.



In the figure above, Σ_1 and Σ_2 denote respectively the differential parts of (7.24) and (7.35).

Example 7.7 The system (7.25) (Example 7.1) with the observation function $y = x_1$, can be dynamically stabilized by means of the compensator

$$\begin{cases} \dot{z}_1 = -z_1 + z_2 + v \\ \dot{z}_2 = -2z_1 - z_2 + v \\ \xi = -z_1 - z_2 \, . \end{cases}$$

Indeed, establishing the connection as explained above, we obtain the closed-loop system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -z_1 - z_2 \\ \dot{z}_1 = x_1 - z_1 + z_2 \\ \dot{z}_2 = x_1 - 2z_1 - z_2 \end{cases}$$

whose characteristic equation is $\lambda^4 + 2\lambda^3 + 3\lambda^2 + 2\lambda + 1 = 0$. It is not difficult to check, by the aid of the Routh-Hurwitz criterion, that all the roots have negative real part. ■

The remaining part of this section is devoted to illustrate how the stabilizing compensator can be constructed in practice, for a general system of the form (7.24).

7.3.1 Construction of an Asymptotic Observer

Definition 7.6 We say that system (7.24) has the *detectability* property (or that it is *detectable*) if there exists a matrix K of appropriate dimensions such that the matrix $L^{\mathbf{t}} = A^{\mathbf{t}} - C^{\mathbf{t}} K^{\mathbf{t}}$ is Hurwitz.

A system possesses the detectability property if and only if its dual system is stabilizable by static state feedback. In particular, each completely observable system is detectable.

Proposition 7.4 *Assume that the given system (7.24) is detectable. For each admissible open loop input $u(t) : [0, +\infty) \to \mathbf{R}^m$ and for each pair of vectors $x_0, z_0 \in \mathbf{R}^n$, we denote by $x(t)$ the solution of the system*

$$\dot{x} = Ax + Bu$$

corresponding to the input $u(t)$ and the initial state x_0, and by $z(t)$ the solution of the system

$$\dot{z} = Lz + Ky + Bu \tag{7.36}$$

corresponding to the input $u(t)$ and the initial state z_0. Then we have

$$\lim_{t\to+\infty} (x(t) - z(t)) = 0 \ .$$

Proof Denote by $e(t) = x(t) - z(t)$ the difference between $z(t)$ and the state $x(t)$. We have:

$$\dot{e} = \dot{x} - \dot{z} = Ax + Bu - Lz - KCx - Bu = (L + KC)x - Lz - KCx = Le \ . \tag{7.37}$$

Recall that the eigenvalues of a matrix are the same as the eigenvalues of its transpose. Since by assumption all the eigenvalues of L have negative real part, we conclude that $\lim_{t\to+\infty} e(t) = 0$ as desired. ■

Remark 7.4 Notice that the input of (7.36) is the sum of the same external input received by the given system and the output of the given system. Proposition 7.4 states that, regardless the initialization of the two systems and assuming that the external input is the same, the solutions of (7.36) asymptotically approximate the solutions of the given system. For this reason, system (7.36) is called an *asymptotic observer* and the quantity $e(t)$ introduced in the previous proof is called the *error* between the true state $x(t)$ and the observed state $z(t)$. ■

7.3.2 Construction of the Dynamic Stabilizer

Now assume that system (7.24) is stabilizable by static state feedback, as well as detectable. Under this additional hypothesis, we may find a matrix H such that the matrix $(A + BH)$ is Hurwitz.

If the full state vector is measurable and available for control purposes, we could directly apply the feedback $u = Hx$ and solve in this way the stabilization problem. Otherwise, it is natural to try the control law $u = Hz$, where z is the approximation of x provided by the asymptotic observer (7.36).

Replacing $u = Hz$ in (7.24) and in (7.36), and recalling that $y = Cx$, we obtain the two systems of differential equations

$$\dot{x} = Ax + BHz \, , \tag{7.38}$$

$$\dot{z} = Lz + KCx + BHz = KCx + (A - KC + BH)z \, . \tag{7.39}$$

Lemma 7.1 *The system composed by (7.38) and (7.39) is asymptotically stable at the origin.*

Proof Let us introduce, as above, the variable $e = x - z$. System (7.38) is equivalent to

$$\dot{x} = (A + BH)x - BHe \tag{7.40}$$

while system (7.39) is equivalent to

$$\begin{aligned} \dot{e} &= Ax + BHz - Az + KCz - BHz - KCx \\ &= Ax - Ax + Ae + KCx - KCe - KCx = Le \, . \end{aligned} \tag{7.41}$$

Systems (7.40) and (7.41) can be reviewed as a unique unforced system, whose matrix is

$$\begin{pmatrix} A + BH & -BH \\ 0 & A - KC \end{pmatrix} \, . \tag{7.42}$$

The set of the eigenvalues of the matrix (7.42) is the union of the sets of the eigenvalues of the matrices $A + BH$ and $A - KC$ which, by construction, are Hurwitz. The statement is so proven. ■

Theorem 7.12 *If system (7.24) is stabilizable by static state feedback, and if it is detectable, then it is stabilizable by a dynamic output feedback, as well.*

Proof The system composed by (7.38) and (7.39) can be interpreted as the result of the connection of the given system (7.24) and the dynamic compensator

$$\begin{cases} \dot{z} = (A - KC + BH)z + Kv \\ \xi = Hz \, . \end{cases}$$

We have so really constructed a dynamic output feedback: the consistency with the notation of Definition 7.5, is easily recovered setting $F = A - KC + BH$, $G = K$, $D = H$. The argument is finished, thanks to Lemma 7.1. ■

We therefore see that the construction of a stabilizing dynamic output feedback reduces to the construction of a stabilizing static state feedback $u = Hx$ for the given system, and the construction of a stabilizing static state feedback $w = -K^{\mathbf{t}}z$ for the dual system

$$\dot{z} = A^{\mathbf{t}}z + C^{\mathbf{t}}w \, .$$

This conclusion is known as the *separation principle*. We emphasize once more that in order to construct H and K we need to know A, B and C, but in order to practically implement the connection, it is sufficient to dispose of the output y.

At a first glance, Theorem 7.12 seems to suggest that the method of dynamic feedback is more general than the method of static state feedback. As a matter of fact, these two methods are (theoretically but, recall, not practically) equivalent.

Theorem 7.13 *Let the system (7.24) be given, and assume that it is stabilizable by means of a dynamic output feedback (7.35). Then, the system is stabilizable by means of a static state feedback, as well.*

Proof Assume that a stabilizing dynamic output feedback exists. The closed-loop system writes as

$$\begin{pmatrix} \dot{x} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & F \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} + \begin{pmatrix} B & 0 \\ 0 & G \end{pmatrix} \begin{pmatrix} 0 & D \\ C & 0 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} \, .$$

Hence, the system in $\mathbf{R}^{n+\nu}$ defined by the matrices

$$\tilde{A} = \begin{pmatrix} A & 0 \\ 0 & F \end{pmatrix} \quad \text{and} \quad \tilde{B} = \begin{pmatrix} B & 0 \\ 0 & G \end{pmatrix}$$

is stabilizable by a static state feedback. Then, according to Theorem 7.6, we must have

$$\text{rank} \begin{pmatrix} A + \lambda I & 0 & | & B & 0 \\ 0 & F + \lambda I & | & 0 & G \end{pmatrix} = n + \nu$$

for each complex number λ with nonnegative real part. This yields

$$\text{rank}\,(A + \lambda I \mid B) = n$$

for each complex number λ with nonnegative real part. ■

7.4 PID Control

As already mentioned, many physical systems of interest in practical applications can be represented by a single linear differential equation

$$\ddot{\xi} + a\dot{\xi} + b\xi = u\ . \tag{7.43}$$

In (7.43) ξ represents the main variable of interest, while u is a scalar input. In the early literature, the control of such systems is based on the following ideas.

1. A feedback proportional to the main variable, that is

$$u = k_0\xi\ . \tag{7.44}$$

2. A feedback proportional to the derivative of the main variable, that is

$$u = k_1\dot{\xi}\ . \tag{7.45}$$

3. An input proportional to the integral of $\xi(t)$, that is

$$u = k_2 \int_0^t \xi(\tau)\, d\tau\ . \tag{7.46}$$

Here, k_0, k_1, k_2 are suitable real constants, often referred to as the *gains*. The feedback (7.44) is called a P control. It can be reviewed as a static output feedback, assuming that (7.43) is associated to the observation function $y = \xi$.

The feedback (7.45) is called a D control. The sum of (7.44) and (7.45), that is the feedback

$$u = k_0\xi + k_1\dot{\xi} \tag{7.47}$$

is called a PD control. Since the full state of (7.43) is the pair $(\xi, \dot{\xi})$, (7.47) can be reviewed as a static state feedback for (7.43).

The function defined in (7.46) is called a I control. Notice that (7.46) can be thought of as a signal generated by the dynamic compensator

$$\dot{z} = k_2\xi\ . \tag{7.48}$$

The combination of (7.44), (7.45), (7.46), that is

$$u = k_0\xi + k_1\dot{\xi} + k_2 \int_0^t \xi(\tau)\, d\tau \tag{7.49}$$

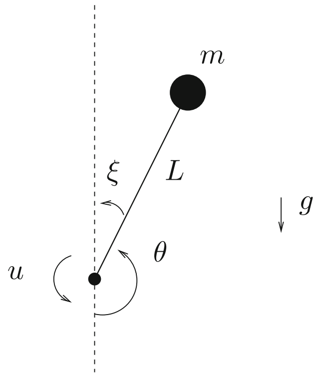
is called a PID control. A PID control can be reviewed as a static state feedback for the system formed by the composition of (7.43) and (7.48). The following example illustrates the use of the PID control.

Consider the equation

$$\ddot{\xi} + \varepsilon\dot{\xi} - \frac{L}{g}\xi = u \quad (0 < \varepsilon << 1)\ . \tag{7.50}$$

It represents the linearized equation of an inverted pendulum with (small) friction. Here, L denotes the length of the bar of the pendulum (of mass m), and g is the gravity constant.

The main variable $\xi = \pi - \theta$ represents the angle with respect to the upward oriented vertical line (see the figure). The control u is exerted by a torque applied to the pivot.



Assume for simplicity that $\frac{L}{g} = 1$. The free system (i.e., with $u = 0$) is clearly unstable. The system can be stabilized by means of a P control, with gain $k_0 < -1$. However, by means of such a control the decay rate cannot be improved, since it depends on the coefficient of the derivative $\dot{\xi}$, which it is not affected by a P control.

Now we try a PI control (that is, a linear combination of P and I controls). To this end, we add the Eq. (7.48) to the system, and write a new system with variables $x_1 = z$, $x_2 = \xi$, $x_3 = \dot{\xi}$. The matrices involved in this three-dimensional representation are

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -\varepsilon \end{pmatrix}, \qquad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad c = (1\ 1\ 0)\ .$$

The choice of c is related to the form $u = k_2x_1 + k_0x_2$ of the desired feedback. The system is completely controllable and observable, and can be actually stabilized by static output feedback. However, as before, the decay rate cannot be improved.

Finally, we can easily see that the system is superstabilizable if a PID control is used. Unfortunately, feedbacks involving a D control are not easy to implement, because measuring the derivative of a variable is usually in practice a hard task. Nevertheless, even today PID control is very popular in industrial applications.

Chapter Summary

In this chapter, the two main topics studied in this book (stability and control) encounter each other. We address the stabilization problem, that is the problem of improving the stability performances of a system by applying a suitable feedback law. We consider several approaches: static state feedback, static output feedback and dynamic feedback. Finally we revisit in this framework the classical PID control method.

Chapter 8
Frequency Domain Approach

As illustrated in the last section of Chap. 4, one possible approach to the analysis of linear differential equations of any order with forcing term makes use of the Laplace transform. This approach reveals a great significance when the forcing term is interpreted as a control. In this chapter, we first generalize the Laplace transform approach to MIMO systems with external inputs. For the particular case of SISO systems which are completely controllable and completely observable, we show how, by this approach, we can obtain a simple, alternative solution of the *synthesis problem*, which means the explicit construction of static output stabilizing feedback laws.

8.1 The Transfer Matrix

Consider a finite dimensional, time invariant linear system represented by the equations

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \ , \end{cases} \tag{8.1}$$

where, as usual, $x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$ and $y \in \mathbf{R}^p$, with n, m and p positive integers. Making use of the variation of constants formula, the output corresponding to an initial state x_0 and an input $u(\cdot)$ can be written

$$y(t) = Cx(t) \qquad \text{where} \qquad x(t) = \int_0^t e^{(t-\tau)A} Bu(\tau)\, d\tau + e^{tA} x_0 \ . \tag{8.2}$$

An alternative representation can be obtained by applying the vector Laplace transform (Sect. B.4) to both side of the first equation in (8.1). To this end, we

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5_8

assume, as we did in Sect. 4.4, that $u(\cdot)$, and hence also $x(\cdot)$ and $y(\cdot)$, are defined for $t \geq 0$. Moreover, we restrict the set of admissible inputs to subexponential functions of the class $\mathcal{PC}([0, +\infty), \mathbf{R}^m)$. By virtue of (8.2), this implies in turn that also $x(t)$ and $y(t)$ are subexponential.

Let $X(s) = \mathcal{L}[x(t)]$, $U(s) = \mathcal{L}[u(t)]$ and $Y(s) = \mathcal{L}[y(t)]$. We have

$$\mathcal{L}[\dot{x}(t)] = sX(s) - x_0 = \mathcal{L}[Ax(t) + Bu(t)] = AX(s) + BU(s)$$

which yields

$$-(A - sI)X(s) = x_0 + BU(s)\ .$$

The matrix $(A - sI)$ is invertible, except for those values of $s \in \mathbf{C}$ which coincides with some eigenvalue of A. Let us denote by σ_0 the maximal real part of the eigenvalues of A. Therefore, if $\mathrm{Re}\, s > \sigma_0$, we may write

$$X(s) = -(A - sI)^{-1}x_0 - (A - sI)^{-1}BU(s)$$

and

$$Y(s) = CX(s) = T(s)U(s) + C(sI - A)^{-1}x_0 \tag{8.3}$$

where we set

$$T(s) = C(sI - A)^{-1}B\ . \tag{8.4}$$

The analogy between (8.2) and (8.3) is not surprising: both contain the sum of two terms; one of them depends on the initial state, the other depends on the input function. As a matter of fact, (8.3) can be alternatively obtained by applying the Laplace transform to (8.2) for $s > \sigma_0$, and making use of Proposition B.4.

Comparing (8.3) with (4.24) of Chap. 4, we see that the role of the polynomial $p_{ch}(s)$ is now played by the matrix $(A - sI)$.

Remark 8.1 Summing up, we have at our disposal two ways in order to represent a physical system with input and output: the matrix (8.4) and the Eq. (8.1) identified in short, in what follows, by the triplet of matrices (A, B, C). When (8.4) is used, we say that the system is represented according to the *frequency domain* approach. This terminology comes from the classical problem of frequency response analysis illustrated in Sect. 4.4.2. When (8.1) is used, we say that the system is represented according to the *time domain* approach. In principle, we should expect that both representations supply the same information about the behavior of the system but, as we shall see later, this is only partially true.

Notice also that in the frequency domain approach, the notion of "state" of the system is not explicitly involved. On the other hand, (8.3) requires purely algebraic computations, while in order to solve (8.1) integral calculus is needed. ■

The matrix $T(s)$ given by (8.4) is called the *transfer matrix*. Notice that $T(s)$ is independent of the initial conditions, so that in order to compute it, we may assume $x_0 = 0$. In the case of a SISO system i.e., when $p = m = 1$, the transfer matrix reduces to a unique element. When in addition the system is defined by a single linear differential equation of order n, it coincides with the transfer function already introduced in Remark 4.7.

In principle, it is possible to compute explicitly the transfer matrix making use of the formula (8.4): the main difficulty rests on the computation of the inverse of $(sI - A)$. Such inverse matrix is sometimes called the *resolvent* of A. For our purposes, the following proposition is sufficient.

Proposition 8.1 *If $s \in \mathbf{C}$ is not an eigenvalue of A, then*

$$(sI - A)^{-1} = \frac{1}{s^n + a_1 s^{n-1} + \cdots + a_n} M(s) \tag{8.5}$$

where $s^n + a_1 s^{n-1} + \cdots + a_n = (-1)^n p_A(s)$, and $M(s)$ is a matrix, whose entries are polynomials of degree less than or equal to $n-1$ (recall that $p_A(s)$ denotes the characteristic polynomial of A).

The proof of (8.5) is easily obtained, having in mind the construction of the inverse of a matrix based on cofactors. A more precise formula for $(sI - A)^{-1}$ can be found for instance in [25], p. 12.

Example 8.1 Consider the system

$$\begin{cases} \dot{x}_1 = x_1 - x_2 + u_1 \\ \dot{x}_2 = x_1 + u_2 \end{cases}$$

with observation map $y = (x_1 + x_2)/2$. The input, state and output spaces have respectively dimension equal to 2, 2, 1. The matrices which define the system are

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad C = \left(\tfrac{1}{2}\ ,\ \tfrac{1}{2}\right) .$$

The system is completely controllable and completely observable. The matrix

$$sI - A = \begin{pmatrix} s-1 & 1 \\ -1 & s \end{pmatrix}$$

is invertible for $s \neq (1 \pm \mathrm{i}\sqrt{3})/2$, and

$$(sI - A)^{-1} = \frac{1}{s^2 - s + 1} \begin{pmatrix} s & -1 \\ 1 & s-1 \end{pmatrix} .$$

The matrix $T(s)$ coincides with the row-vector

$$T(s) = \frac{1}{2(s^2 - s + 1)} \big(s + 1 \, , \, s - 2\big)$$

■

8.2 Properties of the Transfer Matrix

Given a system of the form (8.1), we now list some important properties of its transfer matrix $T(s)$ defined in (8.4).

Property 1 *Each element of $T(s)$ is a proper rational function of $s \in \mathbf{C}$. Moreover, the denominator of each element of $T(s)$ is a polynomial of degree less than or equal to n.*

Property 1 is a straightforward consequence of Proposition 8.1.

Remark 8.2 Property 1 points out that rational functions play a relevant role in the frequency domain approach. Thus, it is convenient to fix some terminology. Let $R_1(s)$, $R_2(s)$ be two rational functions of the complex variable s. We agree that the expression "$R_1(s)$ and $R_2(s)$ are *equal* (or *coincide*, or are the *same function*)" means that there exists a real number r_0 such that

$$R_1(s) = R_2(s) \quad \forall s \in \mathbf{C} \quad \text{with } \operatorname{Re} s > r_0 \, . \tag{8.6}$$

To understand the meaning of this definition, we may look at the following example. Let

$$R_1(s) = \frac{s}{s(s-1)} \, , \qquad R_2(s) = \frac{1}{s-1} \, .$$

These functions cannot be considered "the same function" in the usual sense, since they do not have the same domain. However, they are "equal" in the aforementioned sense. Notice that if the numerator and the denominator of a rational function are polynomial of high degree, it may be very hard to recognize the existence of possible common factors and to get rid of them.

Let $N(s)$, $D(s)$ be two polynomials. We say that $N(s)$ and $D(s)$ are *coprime polynomials* if they do not have common factors. In this case, we also say that the rational function $N(s)/D(s)$ is written in *lowest terms*.

This terminology extends in the obvious way to matrices whose elements are rational functions. For instance, we say that a matrix $T(s)$ is written in lowest terms if the common factors have been canceled in all its entries. ■

Property 2 *Let σ_0 be the maximum of the real parts of the eigenvalues of A, and let $W(\tau) = Ce^{\tau A}B$. Then*

$$T(s) = \mathcal{L}[W(t)] \tag{8.7}$$

for each $s \in \mathbf{C}$ such that $s > \sigma_0$.

Formula (8.7) is an easy consequence of (B.26). The matrix $W(t)$ has been already encountered in Chap. 6 (Lemma 6.2 and subsequent comments). It can be interpreted as the matrix of the *impulse response* (Sect. 1.2): indeed, the columns of $W(t)$ coincide with the output functions of the system corresponding to the initial state $x_0 = 0$ and the impulsive inputs $\delta(t)e_1, \dots, \delta(t)e_m$.

Property 1 could be also recovered from Property 2 taking into account the rules of the Laplace transform. Indeed, the elements of e^{tA} are obtained as sum of terms of the type $q_1(t)e^{\mu t}\cos\omega t$ and $q_2(t)e^{\mu t}\sin\omega t$, where $q_1(t), q_2(t)$ are polynomials and $\lambda = \mu + \mathrm{i}\,\omega$ is an eigenvalue of A.

Notice that $T(s)$ coincides with the Laplace transform of $W(t)$ only if $s > \sigma_0$ but its natural domain, as a rational function, contains all the points of the complex plane, with finitely many exceptions.

Property 3 *Let the system*

$$\begin{cases} \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}u \\ y = \tilde{C}\tilde{x} \end{cases} \tag{8.8}$$

be linearly equivalent to (8.1), according to the definition of Sect. 5.3. Then, (8.1) and (8.8) have the same transfer matrix.

Indeed, if (8.1) and (8.8) are linearly equivalent, then there exists a nonsingular matrix P such that $\tilde{A} = P^{-1}AP$, $\tilde{B} = P^{-1}B$, and $\tilde{C} = CP$. To compute the transfer matrix of system (8.8) we may apply the usual procedure, starting with the differential part of the system. Without loss of generality, we assume a vanishing initial state. We have

$$s\tilde{X}(s) = \tilde{A}\tilde{X}(s) + \tilde{B}U(s) = P^{-1}AP\tilde{X}(s) + P^{-1}BU(s)$$

which yields

$$(sI - P^{-1}AP)\tilde{X}(s) = P^{-1}BU(s)$$

or

$$P^{-1}(sI - A)P\tilde{X}(s) = P^{-1}BU(s)\ .$$

Then we proceed in the following way: we multiply both sides on the left first by P, then by $(sI - A)^{-1}$, and finally by P^{-1}. We obtain

$$\tilde{X}(s) = P^{-1}(sI - A)^{-1}BU(s)\ .$$

In conclusion,

$$Y(s) = \tilde{C}\tilde{X}(s) = CPP^{-1}(sI - A)^{-1}BU(s) = C(sI - A)^{-1}BU(s) .$$

Property 4 *The transfer matrix $T(s)$ depends only on the completely controllable and completely observable part of system (8.1).*

The proof of the statement above makes use of the decomposition of Sect. 5.3, which can be recovered by linear equivalence. Recall that if the matrix A has a triangular block structure, the exponential matrix e^{tA} has an analogous triangular block structure, too. Then, it is not difficult to see that $W(t) = Ce^{tA}B = C_2e^{tA_{22}}B_2$ (see Sect. 5.3 for the notation).

Definition 8.1 Assume that every element of $T(s)$ has been reduced to lowest terms. We say that the complex number s_0 is a *pole* of the system with multiplicity $\mu \geq 1$ if:

1. the denominator of at least one element of $T(s)$ can be exactly divided by $(s - s_0)^\mu$;
2. there exists no element of $T(s)$ whose denominator can be exactly divided by $(s - s_0)^{\mu+1}$.

In other words, the poles of a system are the points of the complex plane where at least one of the elements of the matrix $T(s)$ is not defined.

Property 5 *If s_0 is a pole of $T(s)$ with multiplicity μ, then s_0 is an eigenvalue of A with algebraic multiplicity greater than or equal to μ.*

Property 5 follows directly from Proposition 8.1. On the contrary, it may happens that A possesses some eigenvalue s_0 which is not a pole $T(s)$. We shall come on this point very soon.

Remark 8.3 Properties 4 and 5 imply that if s_0 is a pole of $T(s)$ with multiplicity μ, then s_0 is an eigenvalue of the matrix A_{22} (i.e., the matrix of the completely controllable and completely observable part of system) with algebraic multiplicity greater than or equal to μ. ■

8.3 The Realization Problem

In the previous section we saw how to determine the transfer matrix of a system given under the form (8.1). Now we address the inverse problem. Namely, we want to know if (and how) it is possible to recover the representation (8.1), when the system is assigned by means of its transfer matrix.

Definition 8.2 Let $T(s)$ be a matrix with p rows and m columns, whose elements are proper rational functions of the variable $s \in \mathbf{C}$. The triplet of matrices (A, B, C) whose dimensions are respectively $n \times n$, $n \times m$, $p \times n$, is said to be a *realization* of $T(s)$ if $T(s)$ coincides with the transfer function of the system (8.1) defined by means of the matrices A, B, C. The number n is said to be the *dimension* of the realization.

In order to illustrate some difficulties of the problem, we propose two examples.

Example 8.2 Let us consider the SISO system defined by the equations

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -2x_1 + 3x_2 + u \\ y = x_1 - x_2 \ . \end{cases}$$

The differential part of the system is equivalent to the scalar equation $\xi'' - 3\xi' + 2\xi = u$, where we set $\xi = x_1$. The transfer function is

$$T(s) = \frac{1-s}{s^2 - 3s + 2} = \frac{1-s}{(s-1)(s-2)} = \frac{1}{2-s} \ .$$

The eigenvalues of the system matrix

$$A = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}$$

are 1 and 2, while the unique pole of the transfer function is 2. This example shows that the number of the poles can be actually less than the number of the eigenvalues: this is related to the cancelation of common factors appearing at the numerator and at the denominator of the transfer function. Notice that the system at hand is completely controllable but not completely observable. We also notice that the given system and the system represented by the single equation

$$\begin{cases} \dot{\xi} = 2\xi - u \\ y = \xi \end{cases}$$

have the same transfer function. ■

Example 8.3 Consider the SISO system

$$\begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_2 \\ y = x_1 + x_2 \ . \end{cases}$$

Denoting by A the system matrix, we readily obtain

$$sI - A = \begin{pmatrix} s & 0 \\ 0 & s-1 \end{pmatrix} .$$

This matrix is invertible for $s \notin \{0, 1\}$, and

$$(sI - A)^{-1} = \frac{1}{s(s-1)} \begin{pmatrix} s-1 & 0 \\ 0 & s \end{pmatrix} .$$

It follows that $T(s) = 1/s$. This time, we notice that the system is completely observable, but not completely controllable. The transfer matrix is the same as the transfer matrix of the system

$$\begin{cases} \dot{\xi} = u \\ y = \xi . \end{cases}$$

■

Examples 8.2 and 8.3 point out that systems with different time domain representation may have the same transfer function. In other words, the realization problem does not have, in general, a unique solution. Of course, taking into account that the transfer function depends only on the completely controllable and completely observable part of the system (Property 4 above), this is not surprising. Examples 8.2 and 8.3 suggest also that the presence of uncontrollable or unobservable parts may lead to the cancelation of common factors at the numerator and the denominator of some element of the transfer matrix: this implies in turn a loss of information about the evolution of the internal state of the system. To distinguish among different realizations of the matrix $T(s)$, the following definitions are useful.

Definition 8.3 Let $T(s)$ be a matrix with p rows and m columns, whose elements are proper rational functions of the variable $s \in \mathbf{C}$. A realization (A, B, C) of $T(s)$ is said to be *minimal* if, denoting by n its state space dimension, the state space dimension of any other realization of $T(s)$ is greater than (or equal to) n.

A realization (A, B, C) is said to be *canonical* if the system (8.1) defined by the triplet A, B, C is completely controllable and completely observable.

Theorem 8.1 *Let $T(s)$ be a matrix with p rows and m columns, whose elements are proper rational functions of the variable $s \in \mathbf{C}$. A realization of $T(s)$ is minimal if and only if it is canonical.*

Proof If the realization (A, B, C) is minimal then it must be completely controllable and completely observable. Indeed, assume the contrary. Then according to Property 4, the completely controllable and completely observable part of the system defined by the triplet (A, B, C) provides another realization of $T(s)$, whose state space dimension is strictly less than the previous one.

To prove the converse, let us first recall (see Sects. 5.1.3 and 5.2.3) that a system of the form (8.1) is:

(1) completely controllable if and only if the matrix

$$\Gamma(T) = \int_0^T e^{-\sigma A} B B^{\mathbf{t}} e^{-\sigma A^{\mathbf{t}}} \, d\sigma$$

is nonsingular for some (and hence for each) $T > 0$;

(2) completely observable if and only if the matrix

$$E(T) = \int_0^T e^{\sigma A^{\mathbf{t}}} C^{\mathbf{t}} C e^{\sigma A} \, d\sigma$$

is nonsingular for some (and hence for each) $T > 0$.

Even in this case we may argue by contradiction. Let (A, B, C) be a canonical realization of dimension n and let us assume that for some system

$$\begin{cases} \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}u \\ y = \tilde{C}\tilde{x} \end{cases}$$

where $\tilde{A}$ is a square matrix of dimensions $\nu \times \nu$ with $\nu < n$, we have

$$T(s) = \mathcal{L}[Ce^{tA}B] = \mathcal{L}[\tilde{C}e^{t\tilde{A}}\tilde{B}] \, .$$

By applying $\mathcal{L}^{-1}$ to both sides, we get $Ce^{tA}B = \tilde{C}e^{t\tilde{A}}\tilde{B}$ for each $t \geq 0$, and so also for each $t \in \mathbf{R}$. Now let τ, σ be two arbitrary real numbers. We have:

$$Ce^{\tau A}e^{-\sigma A}B = Ce^{(\tau-\sigma)A}B = \tilde{C}e^{(\tau-\sigma)\tilde{A}}\tilde{B} = \tilde{C}e^{\tau\tilde{A}}e^{-\sigma\tilde{A}}\tilde{B} \, .$$

Multiplying by $e^{\tau A^{\mathbf{t}}}C^{\mathbf{t}}$ to the left and by $B^{\mathbf{t}}e^{-\sigma A^{\mathbf{t}}}$ to the right we obtain

$$e^{\tau A^{\mathbf{t}}}C^{\mathbf{t}}Ce^{\tau A}e^{-\sigma A}BB^{\mathbf{t}}e^{-\sigma A^{\mathbf{t}}} = e^{\tau A^{\mathbf{t}}}C^{\mathbf{t}}\tilde{C}e^{\tau\tilde{A}}e^{-\sigma\tilde{A}}\tilde{B}B^{\mathbf{t}}e^{-\sigma A^{\mathbf{t}}} \, .$$

Next we integrate both sides on the square $0 \leq \tau \leq T$, $0 \leq \sigma \leq T$. The result can be written as

$$E(T) \cdot \Gamma(T) = \tilde{E}(T) \cdot \tilde{\Gamma}(T) \tag{8.9}$$

where $\tilde{E}(T) = \int_0^T e^{\tau A^{\mathbf{t}}}C^{\mathbf{t}}\tilde{C}e^{\tau\tilde{A}}\, d\tau$ and $\tilde{\Gamma}(T) = \int_0^T e^{-\sigma\tilde{A}}\tilde{B}B^{\mathbf{t}}e^{-\sigma A^{\mathbf{t}}}\, d\sigma$.

The matrix $E(T) \cdot \Gamma(T)$ is, by hypothesis, the product of two nonsingular matrices. Hence, it has maximal rank equal to n. On the other hand, $\tilde{E}(T)$ has only ν columns, and $\tilde{\Gamma}(T)$ only ν rows. Hence, the rank of their product cannot be greater than $\nu < n$. Therefore, we conclude that the identity (8.9) is false and the statement is proved. ■

The realization problem has not a unique solution, not even if we limit ourselves to minimal realizations. The reason is clear, having in mind Property 3. Indeed, the first essential step in the construction of a realization is the choice of a real vector space to serve as state space. But, in order to write the Eq. (8.1), one needs also to fix a system of coordinates in this space. This choice is, of course, absolutely arbitrary: for different choices of the system of coordinates the system equations will look formally different, although of the same dimension. The following results clear up the situation.

Theorem 8.2 *Let $T(s)$ be a matrix with p rows and m columns, whose elements are proper rational functions of the variable $s \in \mathbf{C}$. Then there exists at least one realization of $T(s)$.*

Of course, if there exists one realization of $T(s)$, then there exists also a minimal (and canonical) realization.

Proposition 8.2 *Under the same assumptions of Theorem 8.2, if (A, B, C) (with state space dimension equal to n) and $(\tilde{A}, \tilde{B}, \tilde{C})$ (with state space dimension equal to $\tilde{n}$) are two canonical realizations of $T(s)$, then $n = \tilde{n}$ and, moreover, the systems respectively defined by the triplets (A, B, C) and $(\tilde{A}, \tilde{B}, \tilde{C})$ are linearly equivalent.*

For the proofs of these results we refer to [3, 6, 23].

8.4 SISO Systems

In this section we focus our attention on SISO systems. Thus in what follows, we always have $m = p = 1$. Moreover, the transfer matrix reduces to a unique element represented by a proper rational function, referred to as the *transfer function*.

8.4.1 The Realization Problem for SISO Systems

Consider the rational function of the variable $s \in \mathbf{C}$

$$T(s) = \frac{N(s)}{D(s)} = \frac{c_0 + c_1 s + \cdots + c_k s^k}{s^n + a_1 s^{n-1} + \cdots + a_n} \tag{8.10}$$

where $N(s)$ and $D(s)$ may possibly have common factors. If $k < n$, $T(s)$ is proper, and so according to Theorem 8.2 there must exist a realization. Next proposition provides a simple, explicit construction for such a realization, and so it provides also a proof of Theorem 8.2 for the particular case of SISO systems.

Proposition 8.3 *If $k < n$, the system*

$$\begin{cases} \dot{x}_1 = x_2 \\ \dots \\ \dot{x}_{n-1} = x_n \\ \dot{x}_n = -a_n x_1 - \dots - a_1 x_n + u(t) \end{cases} \tag{8.11}$$

with the observation function

$$y = c_0 x_1 + c_1 x_2 + \dots + c_k x_{k+1} \tag{8.12}$$

represents a realization of (8.10).

Remark 8.4 We emphasize that the dimension of realization (8.11) is equal to the degree of the polynomial $D(s)$. ■

Proof of Proposition 8.3 Let C be the matrix of system (8.11). Since C is in companion form, system (8.11) is equivalent to the differential equation of order n

$$\xi^{(n)} + a_1 \xi^{(n-1)} + \dots + a_{n-1}\xi' + a_n \xi = u(t) \tag{8.13}$$

where we set $\xi = x_1$. Taking the Laplace transform of both sides, with the usual notation, we get

$$\Xi(s) = \frac{1}{p_{ch}(s)} U(s)$$

where $p_{ch}(s) = s^n + a_1 s^{n-1} + \dots + a_n = D(s)$ coincides with the characteristic polynomial of (8.13) and, by a possible change of sign, with the characteristic polynomial $p_C(s)$ of the matrix C. Taking into account (8.12), we easily get

$$Y(s) = \frac{c_0 + c_1 s + \dots + c_k s^k}{p_{ch}(s)} U(s) = \frac{N(s)}{D(s)} U(s) = T(s)U(s) \ .$$

Thus we see that $T(s)$ coincides with the transfer function of system (8.11), (8.12). ■

Clearly, the realization (8.11), (8.12) provided by Proposition 8.3 is completely controllable, but not necessarily completely observable. The following proposition concludes the reasoning.

Proposition 8.4 *Let the rational function (8.10) be given, and let $k < n$. The polynomials $N(s)$ and $D(s)$ are coprime if and only if system (8.11) with the observation function (8.12) represents a minimal realization of (8.10).*

Proof Let us assume that numerator and denominator of $T(s)$ do not have common factors. We already know that the system (8.11) with observation function (8.12) is a completely controllable realization of (8.10). If it is not completely observable, taking the observable part we can obtain another realization for which the dimension of the state space is $\tilde{n} < n$. Because of Property 1, we could rewrite $T(s)$ as $T(s) = \tilde{N}(s)/\tilde{D}(s)$, where $\tilde{D}(s)$ is a polynomial of degree not greater than $\tilde{n}$. This is impossible, since by assumption there is no common factors to cancel in $N(s)$ and $D(s)$.

Vice versa, let us assume that the system (8.11), (8.12) satisfies the complete observability condition, in addition to the complete controllability one. By Proposition 8.3, the transfer function of system (8.11), (8.12) coincides with $T(s) = N(s)/D(s)$. Recall that the dimension of the state space of (8.11) is, by construction, equal to the degree of $D(s)$.

If there are common factors to cancel in $N(s)$ and $D(s)$, we could rewrite $T(s)$ as $\tilde{N}(s)/\tilde{D}(s)$, where $\tilde{D}(s)$ is a polynomial of degree $\tilde{n} < n$. But then, using again Proposition 8.3, we can construct another realization of dimension $\tilde{n} < n$. This would imply that the realization (8.11), (8.12) is not minimal, and hence not canonical. This contradicts the assumption. ■

Remark 8.5 In particular, if in (8.10) we have $k = 0$ and $c_0 = 1$, (8.12) reduces to $y = x_1$ and the realization provided by Proposition 8.3 is canonical. ■

Finally, we show that in the case of SISO systems, Property 5 of the previous section admits a partial converse.

Proposition 8.5 *Let a system (8.1) with $m = p = 1$ be given, and let $T(s)$ be its transfer function. Assume that the system is completely controllable and completely observable. If λ is an eigenvalue of A of algebraic multiplicity μ, then λ is a pole of $T(s)$ of multiplicity μ.*

Proof According to Property 1, the transfer function can be written in the form

$$T(s) = \frac{N(s)}{D(s)}$$

where the degree of the polynomial $D(s)$ is not greater than n. If the degree of $D(s)$ is strictly less than n, then $T(s)$ would admit realizations of dimension strictly less than n, a contradiction to the complete controllability and complete observability assumptions. The unique possibility is therefore that the degree of $D(s)$ is exactly equal to n.

It follows that $T(s)$ has exactly n poles (counting possible multiplicities). Let us denote by $s_1, \ldots, s_k$ the distinct poles of $T(s)$, and by $\mu_1, \ldots, \mu_k$ their multiplicities, so that $\mu_1 + \cdots + \mu_k = n$. By Property 5, every s_i is an eigenvalue of A and its algebraic multiplicity is greater than or equal to μ_i. But the eigenvalues of A (counting multiplicities) cannot be more than n. Then, if there is some eigenvalue of A different from $s_1, \ldots, s_k$, or if for some index i, s_i regarded as an eigenvalue of A would have multiplicity strictly greater than μ_i, we get a contradiction. ■

Example 8.4 Consider a system for which the set of the admissible inputs is restricted to the set of functions $u(t)$ of class C^k on the interval $[0, +\infty)$. Assume that the evolution of the system is determined by a linear differential equation of order n

$$\xi^{(n)} + a_1\xi^{(n-1)} + \cdots + a_{n-1}\xi' + a_n\xi = c_k u^{(k)} + \cdots + c_0 u \tag{8.14}$$

where $a_1, \ldots, a_n, c_0, \ldots, c_k$ are real numbers, with $n > k$. Moreover, we assume that the output y coincides with ξ. Such a model, because of the presence of the derivative of the input, seems not to be covered by the form (8.1).

Let us apply the Laplace transform to both sides of (8.14). Assuming $\xi(0) = \xi'(0) = \cdots = \xi^{(n-1)}(0) = 0$ and $u(0) = u'(0) = \cdots = u^{(k)}(0) = 0$, we obtain

$$(s^n + a_1 s^{n-1} + \cdots + a_n)\Xi(s) = (c_k s^k + \cdots + c_0)U(s)$$

that is

$$Y(s) = \Xi(s) = \frac{c_k s^k + \cdots + c_0}{s^n + a_1 s^{n-1} + \cdots + a_n} U(s)$$

where, according to the assumption $m = p = 1$, the transfer matrix reduces to the scalar function

$$T(s) = \frac{c_k s^k + \cdots + c_0}{s^n + a_1 s^{n-1} + \cdots + a_n} . \tag{8.15}$$

This is a proper rational function, possibly with some common factors in the numerator and the denominator. Note that the denominator coincides with the characteristic polynomial $p_{ch}(s)$ of the homogeneous equation associated to (8.14). Therefore, system (8.14) can be realized by means of (8.11) with observation function (8.12). Recall that such a realization is completely controllable, but could be not completely observable.

In the modeling of physical systems, it is not rare the case where the derivative of the input appears explicitly in the equations; this happens for instance when a nonholonomic constraint is modeled as an input. ■

8.4.2 External Stability

As a consequence of Proposition 8.5, and recalling the conclusions of Chaps. 3 and 6 (in particular, Theorem 3.1, Theorem 6.2 and Proposition 6.5), we can immediately state the following theorem.

Theorem 8.3 *Let a system (8.1), with $m = p = 1$, be given. Let $T(s)$ be its transfer function. Assume that (8.1) is completely controllable and completely observable. Then the following statements are equivalent:*

(1) all the eigenvalues of A have negative real part;
(2) all the poles of the transfer function have negative real part;

(3) the system is internally stable;
(4) the system is BIBO-stable.

Remark 8.6 Recall that for systems which are not completely controllable or not completely observable, the previous statements are no more equivalent, in general. For instance, the system

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

is not BIBO-stable, but its transfer function $T(s) = 1/(s+1)$ has the unique pole $s = -1$. ■

Remark 8.7 Theorem 8.3 applies in particular to BIBO systems of the special form

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = u(t) \tag{8.16}$$

where y is taken as the output variable. Indeed, systems of this form are recognized to be completely controllable and completely observable (Remarks 5.4 and 5.7). Hence, for such systems, external stability and internal stability are equivalent properties. ■

The equivalence (2) $\Longleftrightarrow$ (4) of Theorem 8.3 can be proved by using only frequency domain methods. Consider, for simplicity, the case of a system defined by the second order equation

$$y'' + a_1 y' + a_2 y = u \tag{8.17}$$

with output variable y, under the assumption that the characteristic roots s_1, s_2 of the associated unforced equation are real and distinct. Recall that the solutions of (8.17) can be put in the form

$$y(t) = \varphi(t) + \chi(t) \tag{8.18}$$

where $\varphi(t)$ and $\chi(t)$ represent respectively the general solution of the associated homogeneous equation i.e., the unforced solution, and $\chi(t)$ represents the inverse Laplace transform of the function $U(s)/p_{ch}(s)$ i.e., the solution corresponding to the zeroed initial state (compare with (4.24)). Here, with the usual notation, $U(s)$ denotes the Laplace transform of the input $u(t)$. The following lemma exploits the linearity of the system; the argument is similar to that used in the proof of Theorem 6.1.

Lemma 8.1 *The solution $y(t)$ is bounded on the interval $[0, +\infty)$ for each initial condition and each bounded input $u(t)$ if and only if both the following conditions are met:*

(i) the function $\varphi(t)$ is bounded on $[0, +\infty)$ for each initial condition;
(ii) the function $\chi(t)$ is bounded on $[0, +\infty)$ for each bounded input $u(t)$.

Proof The sufficient part is straightforward, by virtue of the triangular inequality

$$|y(t)| \leq |\varphi(t)| + |\chi(t)| .$$

The necessary part can be proved by contradiction. Indeed, assume that there is a choice of the initial conditions y_0, y_1, for which $\varphi(t)$ is not bounded. Let us apply the zero input and let the system evolve with these initial conditions. We have $y(t) = \varphi(t)$ for each $t \geq 0$, so that $y(t)$ is unbounded, as well. To prove that also $\chi(t)$ must be bounded, we can argue in a similar way. ■

The poles of the transfer function coincide with the characteristic roots of the unforced equation associated to (8.17), and the condition that all the characteristic roots have negative real part is necessary and sufficient for internal stability. Therefore, under this condition, the function $\varphi(t)$ is bounded in $[0, +\infty)$.

Let us show that the same condition implies the boundedness of $\chi(t)$ as well, provided that the inputs are bounded. Recalling that $\chi(t)$ can be written in the form (4.25) we recall Proposition 1.5, which states that proving the boundedness of $\chi(t)$ is equivalent to proving that the integral of the function $h(\rho)$ is absolutely convergent. To this end, we need an explicit expression of $h(\rho)$. Since in this case $p_{ch}(s) = s^2 + a_1 s + a_2 = (s - s_2)(s - s_1)$, the Laplace transform of $\chi(t)$ is

$$\frac{1}{s - s_1} \cdot \frac{1}{s - s_2} \cdot U(s) . \tag{8.19}$$

This last expression tells us that the system acts as a cascade connection of two systems of the first order

$$\dot{y} = s_2 y + v \qquad \text{and} \qquad \dot{v} = s_1 v + u .$$

Solving independently these two systems, we obtain

$$y(t) = \int_0^t e^{(t-\tau_2)s_2} v(\tau_2)\, d\tau_2 \qquad \text{and} \qquad v(t) = \int_0^t e^{(t-\tau_1)s_1} u(\tau_1)\, d\tau_1 .$$

Combining these two expressions leads to

$$y(t) = \int_0^t e^{(t-\tau_2)s_2} \left(\int_0^{\tau_2} e^{(\tau_2-\tau_1)s_1} u(\tau_1)\, d\tau_1 \right) d\tau_2 .$$

We can eliminate the variable τ_2 by changing the order of integration and applying the substitution $\tau_2 - \tau_1 = r$. We find

$$\begin{aligned} y(t) &= \int_0^t \left(\int_{\tau_1}^t e^{(t-\tau_2)s_2} e^{(\tau_2-\tau_1)s_1}\, d\tau_2 \right) u(\tau_1)\, d\tau_1 \\ &= \int_0^t \left(\int_0^{t-\tau_1} e^{(t-\tau_1-r)s_2} e^{rs_1}\, dr \right) u(\tau_1)\, d\tau_1 . \end{aligned}$$

Comparing this last expression with (4.25) and applying further the substitution $\rho = t - \tau_1$, we finally get[1]

$$h(\rho) = \int_0^\rho e^{(\rho - r)s_2} e^{r s_1}\, dr \,. \tag{8.20}$$

Now it is straightforward to see that the integral $\int_0^\infty h(\rho)\, d\rho$ is absolutely convergent, by virtue of the assumption $s_1 < 0$, $s_2 < 0$. To finish the proof, we show that the negativity of the real part of the characteristic roots of the unforced system associated to (8.17) is also necessary for the boundedness of $\chi(t)$. Computing the integral in (8.20), we immediately find

$$h(\rho) = \frac{e^{s_1 \rho} - e^{s_2 \rho}}{s_1 - s_2} \,. \tag{8.21}$$

If both s_1 and s_2 (that are distinct by assumption) are not zero, we have with a further integration

$$\int_0^\infty h(\rho)\, d\rho = \lim_{T \to \infty} \frac{1}{s_1 - s_2} \left(\frac{e^{s_1 T}}{s_1} - \frac{e^{s_2 T}}{s_2} - \frac{1}{s_1} + \frac{1}{s_2} \right) .$$

Hence, if at least one is positive, the integral is not convergent. If one is equal to zero (say, s_1), we have:

$$\int_0^\infty h(\rho)\, d\rho = \lim_{T \to \infty} \frac{1}{s_1 - s_2} \left(T - \frac{e^{s_2 T}}{s_2} + \frac{1}{s_2} \right) .$$

and also in this case the integral $\int_0^\infty h(\rho)\, d\rho$ does not converge, regardless the sign of s_2.

[1] Alternatively, performing a decomposition to partial fractions, we can rewrite (8.19) as

$$\frac{1}{s_2 - s_1} \left(\frac{U(s)}{s - s_2} - \frac{U(s)}{s - s_1} \right) .$$

Taking the inverse transform we obtain

$$\frac{1}{s_2 - s_1} \left(\int_0^t e^{s_2 (t-\tau)} u(\tau)\, d\tau - \int_0^t e^{s_1 (t-\tau)} u(\tau)\, d\tau \right)$$
$$= \frac{1}{s_2 - s_1} \left(\int_0^t [e^{s_2 (t-\tau)} - e^{s_1 (t-\tau)}] u(\tau)\, d\tau \right)$$

and we can recover (8.20), comparing again with (4.25).

8.4.3 *Nyquist Diagram*

The so-called Nyquist diagram is a graphic criterion which, applied to the transfer function, allows us to recognize whether a given system possesses the BIBO stability property.

Assume that the system is given by means of its state equations (8.1), where $m = p = 1$, and that it is completely controllable and completely observable, so that the conclusions of Theorem 8.3 hold. Moreover, by virtue of Proposition 8.4, under the same conditions as before, there is no loss of generality assuming that the numerator and the denominator of the transfer function are coprime.

Now let $T(s)$ be a proper rational function of the variable $s \in \mathbf{C}$, with no common factors. A number $s_0 \in \mathbf{C}$ is said to be a *zero*[2] of $T(s)$ if $T(s_0) = 0$.

Let us denote by $w = T(s) \in \mathbf{C}$ the dependent variable. Any complex number s can be thought of as a point of a plane, where a system of coordinates has been fixed $(\mathrm{Re}\, s, \mathrm{Im} s)$. Analogously, any complex number w will be thought of as a point in a plane referred to the coordinates $(\mathrm{Re}\, w, \mathrm{Im} w)$. A continuous map $s = \gamma(t)$ from $\mathbf{R}$ to $\mathbf{C}$ can be interpreted as a planar curve. Analogously, the image of $\gamma(t)$ throughout T can be interpreted as a planar curve $w = \delta(t) = T(\gamma(t))$.

If $s = \gamma(t)$ is simple and closed, it surrounds an open and bounded region $\Gamma \subset \mathbf{C}$. Of course, if $s = \gamma(t)$ is simple and closed, $\delta(t)$ is closed, but it is not necessarily simple.

Example 8.5 Let $T(s) = 1/(s-1)(s-2)$. Figure 8.1 shows the curve δ obtained by applying T to the circumference

$$\begin{cases} \mathrm{Re}\, s = 2 + 2\cos t \\ \mathrm{Im} s = 2\sin t \ . \end{cases}$$

■

Example 8.6 Let $T(s) = 1/(s-1)^2$, and let δ be now obtained by applying T to the circumference

$$\begin{cases} \mathrm{Re}\, s = 1 + \cos t \\ \mathrm{Im} s = \sin t \ . \end{cases}$$

Figure 8.2 may give the wrong impression of a simple and closed curve. Actually the curve is run twice. ■

Let $s = \gamma(t)$ be a simple and closed curve. Let us denote by Z the number of zeros of $T(s)$ lying in Γ and by P the number of poles of $T(s)$ lying in Γ. For simplicity, we assume that there is neither zeros nor poles on the contour of Γ. We need the

[2] If $T(s)$ is the transfer function of a system, its zeros give useful information about the behavior of the system: the interested reader is referred to [8].

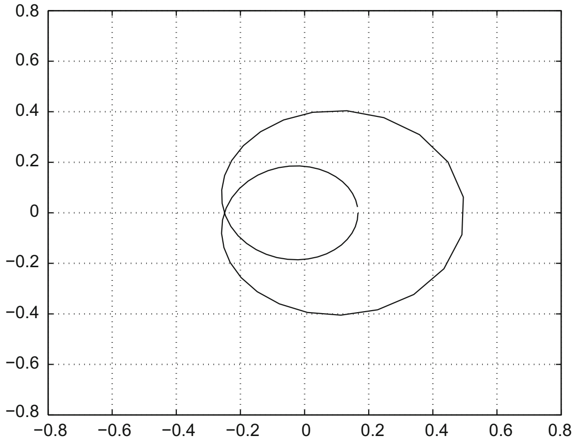


Fig. 8.1 The curve δ of Example 8.5

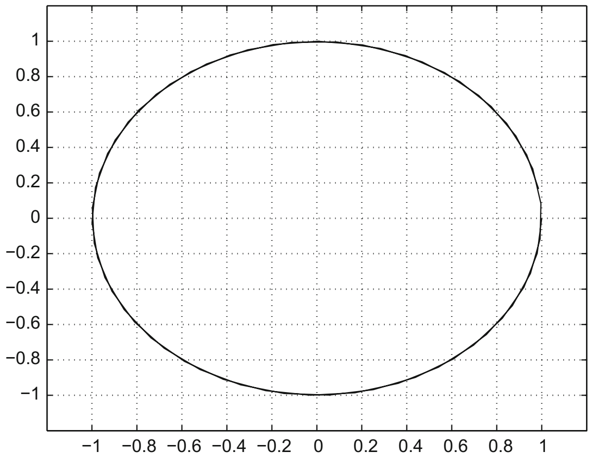


Fig. 8.2 The curve δ of Example 8.6

following, classical result from the theory of functions of a complex variable (see for instance [1]).

Argument principle *Let Q be the integer number denoting how many times the curve $\delta(t)$ encircles the origin in counterclockwise sense, while the contour of Γ is run once in the counterclockwise sense. Then, $Q = Z - P$.*

Definition 8.4 The *Nyquist diagram* of a proper rational function $T(s)$ is the image of the curve $w = T(\gamma(t)) = \delta(t)$, when $\gamma(t) = -\mathrm{i}\, t$ $(t \in \mathbf{R})$.

The curve $\gamma(t) = -\mathrm{i}\, t$ generating the Nyquist diagram is not closed. Nevertheless, the image δ of γ obtained by composition with T, surrounds a bounded region of the complex plane. Indeed, since T is proper, we have

$$\lim_{t \to \pm\infty} \delta(t) = 0 \,.$$

In fact, we may also think of $\gamma(t)$ as a closed curve, by adding to its domain the infinity point: completed in this way, we may imagine that γ surrounds the right half plane of **C** (the contour being run in the counterclockwise sense). Notice that by construction, $\delta(t) = (\mathrm{Re}\, T(-\mathrm{i}\, t), \mathrm{Im} T(-\mathrm{i}\, t))$.

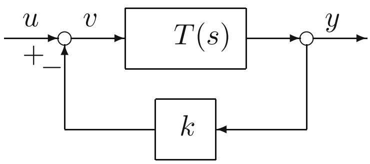
Let $T(s)$ be a proper rational function without zeros or poles on the imaginary axis. Drawing the Nyquist diagram and assuming that Z is known, we can now easily check whether the right half plane of **C** contains some poles of $T(s)$.

By some suitable modifications, these conclusions can be extended to the case where $T(s)$ possesses purely imaginary poles or zeros.

8.4.4 Stabilization by Static Output Feedback

Continuing to deal with a SISO system of the form (8.1) satisfying the complete controllability and the complete observability assumption, in this section we show how to take advantages of the Nyquist criterion in order to determine a static output feedback which stabilizes the given system in the BIBO (and hence also in the internal) sense.

As usual, we denote by $u \in \mathbf{R}$ the input variable and by $y \in \mathbf{R}$ the output variable. First, we examine how the transfer function changes, when a feedback of the form $-ky$ is added to the external input u: here, k is a positive constant, sometimes called the *gain*; the choice of the minus sign is conventional.



Let $T(s)$ be the transfer function of the given system. Let $v = u - ky$. By the aid of the figure above, we easily see that

$$Y(s) = T(s)V(s) = T(s)(U(s) - kY(s))$$

so that

$$Y(s) + kT(s)Y(s) = T(s)U(S) \,.$$

As a consequence, for each $s \in C$ such that $1 + kT(s) \neq 0$,

$$Y(s) = G(s)U(s) = \frac{T(s)}{1 + kT(s)} U(s) = \frac{1}{k} \cdot \frac{T(s)}{\frac{1}{k} + T(s)} U(s)$$

where $G(s)$ denotes the transfer function of the closed loop system. While the value of parameter k varies, the positions of the poles of the resulting transfer function $G(s)$ vary in a continuous way. Thus, to accomplish the desired goal, we need to find a value of k, if any, in such a way that all the poles of the $G(s)$ are moved to the left half of the complex plane.

For simplicity, we assume that $G(s)$ does not have poles on the imaginary axis, and we write

$$H(s) = \frac{1}{k} + T(s) .$$

Lemma 8.2 *The poles of $G(s)$ coincide with the zeros of $H(s)$.*

Proof Write $T(s) = N(s)/D(s)$, where $N(s)$ and $D(s)$ are polynomial. We have

$$H(s) = \frac{1}{k} + \frac{N(s)}{D(s)} = \frac{D(s) + kN(s)}{kD(s)} .$$

Hence, s_0 is a zero of $H(s)$ if and only if $D(s_0) + kN(s_0) = 0$. On the other hand

$$G(s) = \frac{N(s)}{D(s)} \cdot \frac{1}{1 + k\dfrac{N(s)}{D(s)}} = \frac{N(s)}{D(s) + kN(s)} .$$

Hence, s_0 is a pole of $G(s)$ if and only if $D(s_0) + kN(s_0) = 0$. ■

In addition, it is straightforward to realize that the poles of $H(s)$ coincide with the poles of $T(s)$. Applying the Argument Principle to the rational function $H(s)$ leads to the following conclusions:

$$\begin{aligned} &number\ of\ the\ poles\ of\ G\ with\ positive\ real\ part \\ &= number\ of\ poles\ of\ T\ with\ positive\ real\ part\ +\ Q \end{aligned}$$

where Q denotes the number of times the curve $H(-\mathrm{i}\,t)$ encircles the origin in the counterclockwise sense, while the parameter t moves from $-\infty$ to $+\infty$.

On the other hand, it is evident that $-Q$ represents the number of times the curve $T(-\mathrm{i}\,t)$ encircles the point of coordinates $(-\frac{1}{k}, 0)$ of the complex plane in clockwise sense, while the parameter t moves from $-\infty$ to $+\infty$. The following statement resumes the conclusions.

Proposition 8.6 *The static output feedback $-ky$ stabilizes in BIBO (and so also in internal) sense the SISO system (8.1) if the number of times the Nyquist diagram of its transfer function $T(s)$ encircles the point $(-\frac{1}{k}, 0)$ in clockwise sense while the parameter t moves from $-\infty$ to $+\infty$, is equal to the number of poles of the given system lying in the open right half of the complex plane.*

In practical applications, one draws the Nyquist diagram of the given system, and then checks whether there exists a region $\mathcal{D}$ encircled by the diagram the required number of times. If this region exists and intersects the negative real axis, the system is stabilizable. A stabilizing feedback is provided by any value of k such that $(-\frac{1}{k}, 0) \in \mathcal{D}$.

8.5 Disturbance Decoupling

In this last section we discuss an important application which involves both frequency domain and time domain techniques. Consider the system

$$\begin{cases} \dot{x} = Ax + Gd \\ y = Cx \end{cases} \tag{8.22}$$

where $x \in \mathbf{R}^n$, $y \in \mathbf{R}^p$, $d \in \mathbf{R}^q$. The input $d(t) : [0, +\infty) \to \mathbf{R}^q$ is now interpreted as a disturbance. In other words, $d(t)$ is a unknown and undesired input; we just assume that it is piecewise continuous and right continuous, in order to guarantee existence of solutions. For each initial state x_0, the variation of constants formula yields

$$y(t, x_0, d(\cdot)) = Ce^{tA}x_0 + \int_0^t Ce^{(t-\tau)A}Gd(\tau)\, d\tau$$

which reduces to

$$y_0(t) = Ce^{tA}x_0$$

when $d(t) = 0$ for each $t \geq 0$. The function $y_0(t)$ is called the *uncorrupted output signal*. It may happen that $y(t) = y_0(t)$ even for not vanishing disturbances $d(t)$.

Example 8.7 Clearly, the output of the (not completely observable) system

$$\begin{cases} \dot{x}_1 = x_1 - x_2 + d \\ \dot{x}_2 = x_2 \\ \\ y = x_2 \end{cases}$$

is not affected by the disturbance. ■

Definition 8.5 Let us denote, as before, by $y_0(t)$ the uncorrupted output, that is the output corresponding to some initial state x_0 and the vanishing input $d(t) = 0$. We say that the system is *disturbance decoupled* if we have $y(t, x_0, d(\cdot)) = y_0(t)$ for each $t \geq 0$, each initial state x_0 and each input $d(t)$.

Proposition 8.7 *The following statements are equivalent:*

(i) *the system (8.22) is disturbance decoupled;*
(ii) *the impulse response matrix $W(t) = Ce^{tA}G$ vanishes for $t \geq 0$ (and hence, being a real analytic function, for each $t \in \mathbf{R}$);*
(iii) *the transfer matrix $T(s) = C(sI - A)^{-1}G$ vanishes for $s \in \mathbf{C}$;*
(iv) *for each integer $k \geq 0$, one has $CA^kG = 0$.*

Proof The equivalences *(i)* $\Longleftrightarrow$ *(ii)* $\Longleftrightarrow$ *(iii)* are straightforward. Thus, we focus on the statement *(iv)*, and we will prove that it is equivalent to *(ii)*. Assume first that the identity

$$W(t) = Ce^{tA}G = 0 \tag{8.23}$$

holds for each $t \in \mathbf{R}$. To begin with, the substitution $t = 0$ yields $CG = 0$. Coming back to (8.23) and taking the derivative, we obtain

$$CAe^{tA}G = 0 \tag{8.24}$$

for $t \in \mathbf{R}$, which implies $CAG = 0$ by the substitution $t = 0$. We repeat the procedure, taking now the derivative of (8.24) and letting again $t = 0$. This time we obtain $CA^2G = 0$. Continuing in this way, we conclude finally that $CA^kG = 0$ for each integer $k \geq 0$. The converse implication is immediate, since

$$W(t) = Ce^{tA}G = \sum_{k=0}^{\infty} \frac{t^k}{k!} CA^kG$$

for each $t \in R$. ■

Remark 8.8 According to the Cayley-Hamilton Theorem, it is sufficient to check condition *(iv)* of Proposition 8.7 for $k = 0, \dots, n - 1$. ■

Next we establish a necessary and sufficient condition.

Definition 8.6 Let A be a real matrix of dimensions $n \times n$. A subspace V of $\mathbf{R}^n$ is said to be an *algebraic (o geometric) invariant* for A if $AV \subseteq V$.

The subspace V is said to be a *dynamic invariant* for A if from $x_0 \in V$ it follows $e^{tA}x_0 \in V$ for each $t \geq 0$ (and hence for each $t \in \mathbf{R}$).

Proposition 8.8 *The subspace V is an* algebraic invariant *for A if and only if it is a* dynamic invariant *for A.*

Proof Let V be an algebraic invariant. For each $x_0 \in V$, we clearly have $Ax_0 \in V$, $A^2x_0 \in V$, and so on. Hence, $e^{tA}x_0 = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^kx_0$ belongs to V. On the other hand, let V be a dynamic invariant, and let $x_0 \in V$. Then, for each $t \neq 0$, we also have

$$\frac{e^{tA}x_0 - x_0}{t} \in V .$$

Taking the limit for $t \to 0$, we get $Ax_0 \in V$. ■

Theorem 8.4 *The given system is disturbance decoupled if and only if there exists a subspace V of $\mathbf{R}^n$ which is an algebraic invariant for A, and such that* $\operatorname{im} G \subseteq V \subseteq \ker C$.

Proof Assume that the system is disturbance decoupled. Let us introduce a matrix H, whose columns coincide with the columns of the matrices G, AG, $A^2G, \dots, A^{n-1}G$, in this order. The matrix H can be interpreted as a linear map from $\mathbf{R}^{n\times q}$ in $\mathbf{R}^n$. Let $V = \operatorname{im} H$. By the Cayley-Hamilton Theorem, V is an algebraic invariant. The inclusion $\operatorname{im} G \subseteq V$ is obvious, while the other one $V \subseteq \ker C$ follows from Proposition 8.7, *(iv)*.

To prove the converse, we first remark that if a subspace V is an algebraic invariant and $\operatorname{im} G \subseteq V$, then clearly $\operatorname{im}(A^kG) \subseteq V$ for each positive integer k. As a consequence, since $V \subseteq \ker C$, we also have $CA^kGx = 0$ for each integer $k \geq 0$ and each $x \in \mathbf{R}^n$. The conclusion follows, using again Proposition 8.7, *(iv)*. ■

There is an other characterization of disturbance decoupled systems. By means of a linear change of coordinates, we can put the system in the observability canonical form

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + G_1 d \\ \dot{z}_2 = A_{22}z_2 + G_2 d \\ \\ y = C_2 z_2 \end{cases} \tag{8.25}$$

where $z_1 \in \mathbf{R}^{n-r}$, $z_2 \in \mathbf{R}^r$ for some nonnegative integer $r \leq n$, and the subsystem

$$\begin{cases} \dot{z}_2 = A_{22}z_2 + G_2 d \\ \\ y = C_2 z_2 \end{cases} \tag{8.26}$$

is completely observable. The case $r = 0$ is trivial, so we can assume $r > 0$.

Theorem 8.5 *The system (8.22) is disturbance decoupled if and only if $G_2 = 0$, where G_2 is the matrix appearing in (8.25).*

Proof The sufficient part is evident (to be formal, it can be easily obtained as an application of Theorem 8.4). Let us prove the necessary part.

Assume that the system is disturbance decoupled. Taking into account the form (8.25), for each integer $k \geq 0$, we see that CA^k can be written as a row block matrix $(0 \mid C_2A_{22}^k)$, where 0 denotes here a block of $n-r$ zero columns. From this, it easily follows that $CA^kG = C_2A_{22}^kG_2$ for each integer $k \geq 0$. Since the subsystem (8.26) is completely observable, the matrix

$$M = \begin{pmatrix} C_2 \\ C_2A_{22} \\ \dots \\ C_2A_{22}^{r-1} \end{pmatrix}$$

has a maximal rank i.e., rank $M = r$. Now, assume by contradiction that $v = G_2d \neq 0$ for some $d \in \mathbf{R}^q$ (note that $v \in \mathbf{R}^r$ and that, necessarily, $d \neq 0$). The vector $Mv \in \mathbf{R}^{p\times r}$ is a linear combination of the r linearly independent columns of M, so that being $v \neq 0$, we also have $Mv \neq 0$. But

$$Mv = \begin{pmatrix} C_2G_2d \\ C_2A_{22}G_2d \\ \dots \\ C_2A_{22}^{r-1}G_2d \end{pmatrix} = \begin{pmatrix} C_2G_2 \\ C_2A_{22}G_2 \\ \dots \\ C_2A_{22}^{r-1}G_2 \end{pmatrix} d \tag{8.27}$$

with $d \neq 0$. On the other hand, the disturbance decoupling assumption implies

$$C_2G_2 = C_2A_{22}G_2 = \cdots = C_2A_{22}^{r-1}G_2 = 0\ . \tag{8.28}$$

Clearly, (8.27) and (8.28) are in contradiction. Therefore, we must have $G_2d = 0$ for each $d \in \mathbf{R}^q$, and this means that $G_2 = 0$. ■

If the given system is not disturbance decoupled, we can try to achieve this property by the use of a suitable feedback law. In other words, we add a control term in the system equation

$$\begin{cases} \dot{x} = Ax + Bu + Gd \\ y = Cx \end{cases} \tag{8.29}$$

where with the usual notation $u \in \mathbf{R}^m$, and we ask whether it is possible to find a static state feedback of the form $u = Fx$ such that the closed-loop system

$$\begin{cases} \dot{x} = (A + BF)x + Gd \\ y = Cx \end{cases}$$

is disturbance decoupled. The conditions for answering this question rest on the introduction of a new notion of invariance, concerning the state equation

$$\dot{x} = Ax + Bu\ . \tag{8.30}$$

Definition 8.7 A subspace $V \subseteq \mathbf{R}^n$ is said to be a *strong controlled invariant* for the system (8.30) if for each $x_0 \in V$ and each admissible input $u(t) : [0, +\infty) \to \mathbf{R}^m$ we have $x(t, x_0, u(\cdot)) \in V$ for each $t \geq 0$.

Apart from the modified terminology, the definition above coincides with the notion already introduced in Sect. 5.3.2.

Definition 8.8 A subspace $V \subseteq \mathbf{R}^n$ is said to be a *weak controlled invariant* for the system (8.30) if for each $x_0 \in V$ there exists an admissible input $u(t) : [0, +\infty) \to \mathbf{R}^m$ such that $x(t, x_0, u(\cdot)) \in V$ for each $t \geq 0$.

Example 8.8 The subspace $V = \{(x_1, x_2) : x_2 = 0\} \subseteq \mathbf{R}^2$ is a weak controlled invariant, but not a strong controlled invariant, for the system

$$\begin{cases} \dot{x}_1 = x_1 + x_2 \\ \dot{x}_2 = u \ . \end{cases}$$

Note that this system is completely controllable. ■

The weak controlled invariant subspaces can be characterized in the following way.

Proposition 8.9 *The following statements are equivalent.*

(i) *V is a weak controlled invariant;*
(ii) *$AV \subseteq V + \operatorname{im} B$;*
(iii) *there exists a matrix F with n columns and m rows such that $(A + BF)V \subseteq V$.*

Proof First we prove that *(i)* $\Longrightarrow$ *(ii)*. Let $x_0 \in V$ and let $u(t) : [0, +\infty) \to \mathbf{R}^m$ be an input such that $x(t, x_0, u(\cdot)) \in V$ for each $t \geq 0$. Without loss of generality, we can extend continuously $u(t)$ on a small interval $(-\varepsilon, 0)$, so that $x(t, x_0, u(\cdot))$ can be considered of class C^1 at $t = 0$. Then

$$\lim_{t \to 0^+} \frac{x(t, x_0, u(\cdot)) - x_0}{t} = \dot{x}(0) \in V$$

that is $Ax_0 + Bu(0) \in V$, or $Ax_0 \in V - Bu(0)$.

Next we prove that *(ii)* $\Longrightarrow$ *(iii)*. Let $\dim V = k \leq n$. Let $e_1, \ldots, e_n$ be a basis of $\mathbf{R}^n$, such that its first k elements $e_1, \ldots, e_k$ constitute a basis of V. Then for each $i = 1, \ldots, k$ one has $Ae_i = g_i + Bu_i$ for some $g_i \in V$ and some $u_i \in \mathbf{R}^m$. Let us chose other vectors $u_{k+1}, \ldots, u_n \in \mathbf{R}^m$ in arbitrary way, and define the matrix F by the relations $Fe_j = -u_j$, for $j = 1, \ldots, n$. Then we have, for $i = 1, \ldots, k$,

$$(A + BF)e_i = Ae_i + BFe_i = g_i + Bu_i - Bu_i = g_i \in V \ .$$

Finally we prove that *(iii)* $\Longrightarrow$ *(i)*. Let $x_0 \in V$ and let $x(t)$ be the solution of the closed loop system

$$\begin{cases} \dot{x} = (A + BF)x \\ x(0) = x_0 \ . \end{cases} \tag{8.31}$$

Of course, $x(t)$ is also a solution of the problem

$$\begin{cases} \dot{x} = Ax + Bu(t) \\ x(0) = x_0 \end{cases}$$

where $u(t) = Fx(t)$. The proof is completed, by noticing that V is a dynamic invariant with respect to system (8.31). ■

We are finally able to state the main result of this section.

Theorem 8.6 *System (8.29) can be rendered disturbance decoupled by means of a linear feedback if and only if there exists a subspace $V \subseteq \mathbf{R}^n$ which is a weak controlled invariant for the system (8.30) and such that* $\operatorname{im} G \subseteq V \subseteq \ker C$.

Proof Let us prove first the necessary part. So let F be a matrix such that the system

$$\begin{cases} \dot{x} = (A + BF)x + Gd \\ y = Cx \end{cases} \tag{8.32}$$

is disturbance decoupled. According to Theorem 8.4, there exists an algebraic invariant subspace V, such that $\operatorname{im} G \subseteq V \subseteq \ker C$. This implies that $(A + BF)V \subseteq V$, and this in turn means that V is a weak controlled invariant, by Proposition 8.9.

Then we prove the sufficient part. If V is weak controlled invariant, then by Proposition 8.9 there exists F such that $(A + BF)V \subseteq V$. Together with the inclusions $\operatorname{im} G \subseteq V \subseteq \ker C$, this implies finally that the system (8.32) is disturbance decoupled by Theorem 8.4. ■

Chapter Summary

The subject of the last chapter is the relationship between two possible approaches to the analysis of a system: the time domain approach (developed in the previous chapters) and the more traditional frequency domain approach based on the Laplace transform. We study in particular the realization problem. For the case of SISO systems, we also give a different solution to the stabilization problem by output feedback. Finally, we illustrate the decoupling problem, whose solution takes advantages of both approaches.

Appendix A
Internal Stability Notions

The purpose of this appendix is to introduce the basic properties arising in the characterization of the long-term qualitative behavior of solutions of unforced, time invariant differential systems. Notation and terminology are those of the Introduction (Chap. 1). However, since the interest is focused on the state variable, in this appendix the observation map is ignored. From the mathematical point of view, the systems considered in this chapter reduce therefore to systems of ordinary differential equations (in general, nonlinear)[1]

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

where $x \in \mathbf{R}^n$. Recall that a *solution* of (A.1) is any differentiable function $x = \varphi(t)$ defined on some interval $I \subseteq \mathbf{R}$ such that $\dot{\varphi}(t) = f(\varphi(t))$ for each $t \in I$. We will assume that the function f in (A.1) is defined and continuous together with its first partial derivatives, for each $x \in \mathbf{R}^n$; moreover, we assume that it satisfies the inequality

$$||f(x)|| \leq a||x|| + b$$

for some positive constants a, b. Under these assumptions, for each initial pair (t_0, x_0) existence and uniqueness of solutions are guaranteed, and we may further take $I = \mathbf{R}$ without loss of generality [24]. Moreover, since the function f does not depend explicitly on t, according to Proposition 1.9, the system (A.1) is time invariant; therefore it is not restrictive to assume $t_0 = 0$.

The notions introduced in this appendix are often referred to as *internal stability notions*, in order to emphasize the difference with the notion of *external stability* introduced in Chap. 1 and studied in detail in Chap. 6.

[1]The notions we are going to introduce are applied in this book essentially for the case of linear systems; however, they can be better understood when referred to a general system of the type (A.1).

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5

A.1 The Flow Map

A solution of (A.1) can be regarded as a parameterized curve $x = \varphi(t)$ of $\mathbf{R}^n$. For each $t \in \mathbf{R}$, the tangent vector to such a curve at the point x coincides with $f(x)$. For this reason, the function $f : \mathbf{R}^n \to \mathbf{R}^n$ which defines (A.1) is also called a *vector field*.

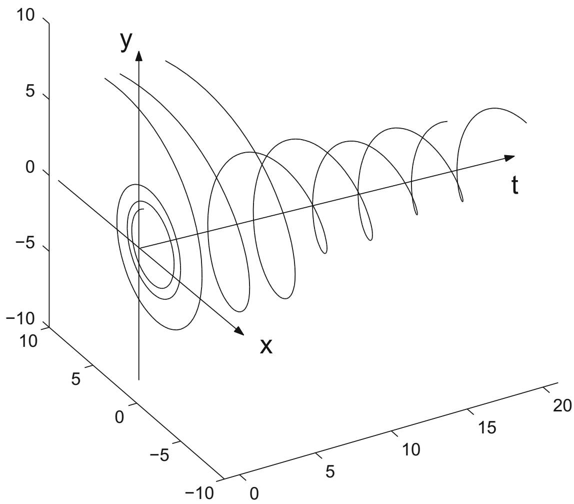
The image of a solution $x = \varphi(t)$ of (A.1) is called an *orbit* or a *trajectory*. It is important do not confuse the graph of a solution $\varphi(t)$, which is a subset of $\mathbf{R}\times\mathbf{R}^n$, with the orbit of $\varphi(t)$, which coincides with the set $\varphi(\mathbf{R})$ and it is a subset of $\mathbf{R}^n$. We may also view the orbit of φ as the orthogonal projection of the graph of φ on $\mathbf{R}^n$, along the time axis (see Fig. A.1).

We already mentioned that under the stated assumptions, for each initial condition (A.1) has a unique solution. In order to emphasize its global validity, we my reformulate this property writing that if $x = \varphi(t)$ and $x = \psi(t)$ are two arbitrary solutions of (A.1), then

$$\exists \bar{t} : \quad \varphi(\bar{t}) = \psi(\bar{t}) \quad \Longrightarrow \quad \varphi(t) = \psi(t) \quad \forall t \in \mathbf{R}\,. \tag{A.2}$$

The geometric interpretation of (A.2) is that if the graphs of the two solutions have a common point, then they must coincide. We may also interpret the time invariance property from a geometrical point of view: the time translation of the graph of a solution is again the graph of a (in general, different) solution. All the solutions obtained as time translation of a fixed solution obviously are equivalent parametrization of the same curve, and so they define the same orbit (see again Fig. A.1). This fact admits a converse.

Fig. A.1 Two solutions and their projections



Lemma A.1 *Let $\varphi(t)$ and $\psi(t)$ be two arbitrary solutions of (A.1) defined for each $t \in \mathbf{R}$. Then,*

$$\exists t_1, t_2 : \quad \varphi(t_1) = \psi(t_2) \implies \psi(t) = \varphi(t+T) \ \ \forall t \in \mathbf{R} \,, \tag{A.3}$$

where we set $T = t_1 - t_2$.

Proof Let $T = t_1 - t_2$ and $\chi(t) = \varphi(t+T)$. Clearly, $\chi(t)$ is a solution. It satisfies the initial condition

$$\chi(t_2) = \varphi(t_2 + t_1 - t_2) = \varphi(t_1) \ .$$

But also $\psi(t)$ is a solution which, by hypothesis, satisfies the same condition. Because of the uniqueness property, we have

$$\chi(t) = \varphi(t+T) = \psi(t) \qquad \forall t \in \mathbf{R} \ .$$

■

The meaning of Lemma A.1 is that if two orbits have a common point, then they must coincide (the reader is warned to notice the difference between (A.2) and (A.3)). In other words, there is a unique orbit passing through every point of $\mathbf{R}^n$. The orbits of the system (A.1) fill the space and are displayed in such a way to form a partition of the space. We might define an equivalence relation, saying that two points are equivalent when they lie on the same orbit. In the particular case $n = 2$, we can image that the orbits form a picture in the plane. This picture is also called *state configuration* or *phase portrait*. To denote the solution of the Cauchy problem

$$\begin{cases} \dot{x} = f(x) \\ x(0) = x_0 \end{cases} \tag{A.4}$$

we use the notation[2]

$$x = x(t, x_0) \tag{A.5}$$

which has the advantage of emphasizing, beside the time variable t, also the initial state x_0. Equation (A.5) define a function from $\mathbf{R} \times \mathbf{R}^n$ to $\mathbf{R}^n$: this is called the *flow map* generated by the vector field f. It can be interpreted as a function of t for each fixed x_0, or as a function from $\mathbf{R}^n$ to $\mathbf{R}^n$, parameterized by t.

Remark A.1 In (A.5), the variable t should be thought of not as the indication of a precise instant of time, but rather as the indication of the duration of a process, that is the length of the time interval needed to transfer the state of the system from x_0 to $x(t, x_0)$. ■

[2]Note that (A.5) is nothing else than (1.12) adapted to the case of (A.1).

Proposition A.1 *The flow map of the vector field f satisfies the following properties:*

$$x(0, x_0) = x_0 \tag{A.6}$$

$$x(t, x(\tau, x_0)) = x(t + \tau, x_0) \tag{A.7}$$

for each $t, \tau \in \mathbf{R}$ and $x_0 \in \mathbf{R}^n$.

A.2 Equilibrium Points and Stability in Lyapunov Sense

Roughly speaking, *internal stability* means that in the absence of external energy supply, the state of a system evolves remaining in a neighborhood of a rest point, and eventually approaches a rest point.

Let the unforced, time invariant differential system (A.1) be given. We say that $\bar{x} \in \mathbf{R}^n$ is an *equilibrium point* if the constant function $\varphi(t) \equiv \bar{x}$ is a solution. Sometimes, equilibrium points are also called *rest* or *singular*, or even *critical points*. If $\bar{x}$ is an equilibrium point, then the orbit issuing from $\bar{x}$ reduces to the singleton $\{\bar{x}\}$.

Proposition A.2 *The point $\bar{x}$ is an equilibrium point if and only if $f(\bar{x}) = 0$.*

We say that the equilibrium point $\bar{x}$ is *isolated* if there exists a neighborhood $\mathcal{O}$ of $\bar{x}$ such that $f(x) \neq 0$ for each $x \in \mathcal{O}$, $x \neq \bar{x}$.

Definition A.1 Let $\bar{x}$ be an equilibrium point. We say that $\bar{x}$ is *stable (in Lyapunov sense)* for the system (A.1) if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|x_0 - \bar{x}\| < \delta \quad \Longrightarrow \quad \|x(t; x_0) - \bar{x}\| < \varepsilon, \quad \forall t \geq 0.$$

Definition A.2 Let $\bar{x}$ be an equilibrium point. We say that $\bar{x}$ is *attractive* if there exists $\delta_0 > 0$ such that, for each initial state x_0 for which $\|x_0 - \bar{x}\| < \delta_0$, one has

$$\lim_{t \to +\infty} x(t; x_0) = \bar{x}\ . \tag{A.8}$$

Definition A.3 Let $\bar{x}$ be an equilibrium point. If $\bar{x}$ is both stable and attractive, we say that it is *asymptotically stable*. Moreover, if (A.8) holds for each $x_0 \in \mathbf{R}^n$, $\bar{x}$ is called *globally asymptotically stable*. Finally, if the solutions converge to $\bar{x}$ with an exponential decay i.e., there exist $M > 0$, $\alpha > 0$ such that

$$\forall x_0 \ \text{ with } \ \|x_0 - \bar{x}\| < \delta_0 \ \text{ we have } \ \|x(t; x_0) - \bar{x}\| \leq M e^{-\alpha t}\ , \tag{A.9}$$

we speak about *exponential stability*. The supremum of the numbers α such that (A.9) holds for some suitable M, is called the *decay rate*.

We end this chapter by the following notion, very useful in the analysis of the qualitative behavior of the unforced system (A.1).

Definition A.4 Let K be a closed subset of $\mathbf{R}^n$. We say that K is *dynamically invariant* for the system (A.1) if for each $x_0 \in K$ one has $x(t; x_0) \in K$ for every $t \in \mathbf{R}$.

Appendix Summary

Appendix A recalls some mathematical definitions concerning stability. The informal term "stability" actually involves the notion of stability in the sense of Lyapunov and the notion of attraction. In general, these notions are mutually independent, but in the case of linear systems the latter implies the former. Moreover, in the case of linear systems there is no way to distinguish the local and global aspects. For these reasons, in this Appendix, and only in this Appendix, we refer to general (nonlinear) systems of ordinary differential equations.

Appendix B
Laplace Transform

In this appendix we recall some basic facts about Laplace transform, that are needed for the applications considered in this book. In view of our limited goals and for sake of simplicity, the subject will not be treated with the maximal generality and mathematical rigor. In particular, the Dirac-delta function and its Laplace transform will be introduced only at heuristic level. For a more formal presentation, the reader can be addressed to one of many existing books on this topic, for instance [15].

B.1 Definition and Main Properties

Let $f : [0, +\infty) \to \mathbf{R}$ be a piecewise continuous function.

Definition B.1 We say that f is a *subexponential function* if there exist real constants $M > 0$ and α such that

$$|f(t)| \leq Me^{\alpha t} \qquad \forall t \in [0, +\infty) \ . \tag{B.1}$$

To each piecewise continuous, subexponential function we can associate a real number σ_0, defined as the infimum of the numbers α for which there exists M such that (B.1) holds. This number σ_0 is called the *order* of f.

Lemma B.1 *Let f be a subexponential, piecewise continuous function, and let s be any complex number such that* $\operatorname{Re} s > \sigma_0$. *Then,*

$$\lim_{\xi \to +\infty} f(\xi)e^{-s\xi} = 0 \ .$$

Proof If $H(\xi)$ is a complex function of one real variable, $\lim_{\xi \to +\infty} H(\xi) = 0$ is equivalent to

$$\lim_{\xi \to +\infty} \operatorname{Re} H(\xi) = \lim_{\xi \to +\infty} \operatorname{Im} H(\xi) = 0 \ .$$

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5

But for each $z \in \mathbf{C}$, we have

$$|\mathrm{Re}\, z| \leq |z| \qquad \text{and} \qquad |\mathrm{Im}\, z| \leq |z|\ .$$

Thus, it is sufficient to prove that

$$\lim_{\xi \to +\infty} |f(\xi)e^{-s\xi}| = \lim_{\xi \to +\infty} |f(\xi)|e^{-\xi \mathrm{Re}\, s} = 0\ .$$

Let α be a real number such that $\sigma_0 < \alpha < \mathrm{Re}\, s$. We have

$$|f(\xi)|e^{-\xi \mathrm{Re}\, s} \leq M e^{\xi(\alpha - \mathrm{Re}\, s)}$$

and the statement follows since $\alpha - \mathrm{Re}\, s < 0$.

■

The Laplace transform allows us to associate a function $F : \mathbf{C} \to \mathbf{C}$ to each piecewise continuous, subexponential function $f : [0, +\infty) \to \mathbf{R}$. Before giving the formal definition, we still need some preliminary results.

Lemma B.2 *Let f be a piecewise continuous, subexponential function, whose order is σ_0. For each complex number s such that* $\mathrm{Re}\, s > \sigma_0$*, the improper integral*

$$\int_0^{+\infty} f(t)e^{-st}dt$$

is absolutely convergent.

Proof The absolute convergence of the improper integral $\int_0^{+\infty} H(t)dt$ of a function $H : \mathbf{R} \to \mathbf{C}$ is equivalent to the convergence of both the integrals

$$\int_0^{+\infty} |\mathrm{Re}\, H(t)|\, dt \qquad \text{and} \qquad \int_0^{+\infty} |\mathrm{Im}\, H(t)|\, dt\ . \tag{B.2}$$

As already noticed in the proof of Lemma B.1, it is therefore sufficient to show that the integral

$$\int_0^{+\infty} |H(t)|dt$$

is convergent. In our case,

$$|f(t)e^{-st}| = |f(t)||e^{-st}| = |f(t)|e^{-t\mathrm{Re}\, s} \leq M e^{t(\alpha - \mathrm{Re}\, s)}\ .$$

According to the definition of σ_0, we can chose α in such a way that $\sigma_0 < \alpha < \mathrm{Re}\, s$, so that $\alpha - \mathrm{Re}\, s < 0$. The convergence of the two integrals (B.2) is guaranteed by comparison. Notice that if we set $s = \sigma + \mathrm{i}\,\omega$, we have

$$\int_0^{+\infty} |\text{Re}\,(f(t)e^{-st})|dt = \int_0^{+\infty} |f(t)e^{-t\sigma}\cos(-\omega t)|dt$$

and

$$\int_0^{+\infty} |\text{Im}\,(f(t)e^{-st})|dt = \int_0^{+\infty} |f(t)e^{-t\sigma}\sin(-\omega t)|dt\ .$$

■

We are finally ready to introduce the main definition of this Appendix.

Definition B.2 Let f be a piecewise continuous, subexponential function, defined in $[0, +\infty)$, whose order is σ_0. The *Laplace transform* of f is the complex function

$$s \mapsto F(s) = \int_0^{+\infty} f(t)e^{-st}dt \tag{B.3}$$

defined on the domain $\{s \in \mathbf{C} : \text{Re}\, s > \sigma_0\}$.

It is convenient to remark that F could be coincident with the restriction to the half-plane $\{s \in \mathbf{C}\ :\ \text{Re}\, s > \sigma_0\}$ of a function $\tilde{F} : \mathbf{C} \to \mathbf{C}$ defined in a broader domain (to this respect, see Remark B.1). We also remark that (B.3) is meaningful even if f is a complex function of one real variable.

The operator defined by (B.3) will be denoted by the symbol $\mathcal{L}$. We will also agree to denote by the same letter (respectively, small and capital) the function to be transformed and its Laplace transform. Hence, we write

$$F(s) = \mathcal{L}[f(t)] = \int_0^{+\infty} f(t)e^{-st}dt.$$

Next we review some important properties of the Laplace transform.

Property 1 (Linearity) *Let f and g be two piecewise continuous, subexponential functions, defined on the interval* $[0, +\infty)$*, of order respectively σ_1 and σ_2. Then, for each $a, b \in \mathbf{R}$ the function $af + bg$ is subexponential of order* $\sigma_0 = \max\{\sigma_1, \sigma_2\}$. *Moreover,*

$$aF(s) + bG(s) = \mathcal{L}[af(t) + bg(t)] \tag{B.4}$$

for $\text{Re}\, s > \sigma_0$*, where $F(s) = \mathcal{L}[f(t)]$ and $G(s) = \mathcal{L}[g(t)]$.*

Proof It is immediate to check that $af + bg$ is subexponential (take a value of α greater than both σ_1 and σ_2). Formula (B.4) is a trivial consequence of the properties of the integrals.

■

Property 2 (Rescaling) *Let f be a piecewise continuous, subexponential function, defined in* $[0, +\infty)$*, of order σ_0, and let $F(s) = \mathcal{L}[f(t)]$ for* $\text{Re}\, s > \sigma_0$*. Then, for each* $a > 0$*, $g(t) = f(at)$ is a subexponential function of order $a\sigma_0$, and*

$$\mathcal{L}[f(at)] = \frac{1}{a} F\left(\frac{s}{a}\right) \qquad for \;\; \text{Re}\, s > a\sigma_0 \, . \tag{B.5}$$

Proof From (B.1) we have easily $|f(at)| \le Me^{\alpha a t} = Me^{\beta t}$ for $\beta = \alpha a > a\sigma_0$. Setting $\tau = at$, we therefore have

$$\int_0^{+\infty} f(at)e^{-st}\, dt = \frac{1}{a}\int_0^{+\infty} f(\tau)e^{-\frac{s}{a}\tau}\, d\tau = \frac{1}{a}\int_0^{+\infty} f(\tau)e^{-r\tau}\, d\tau = \frac{1}{a}F(r)$$

provided that $\text{Re}\, r > \sigma_0$, where $r = s/a$. But $\text{Re}\, r = \text{Re}\,(s/a) = (\text{Re}\, s)/a$, and hence requiring $\text{Re}\, r > \sigma_0$ it is equivalent to require $\text{Re}\, s > a\sigma_0$.

■

Property 3 (Right translation) *Let f be a piecewise continuous, subexponential function, defined in* $[0, +\infty)$*, of order σ_0, and let $F(s) = \mathcal{L}[f(t)]$ for* $\text{Re}\, s > \sigma_0$. *In addition, let*

$$g(t) = \begin{cases} 0 & for\; 0 \le t \le c \\ f(t-c) & for\; t > c \, , \end{cases}$$

where $c > 0$. Then g is a subexponential function of order σ_0, and $\mathcal{L}[g(t)] = e^{-cs}F(s)$ for $\text{Re}\, s > \sigma_0$.

Proof The reader can easily check that g is a subexponential function. Moreover, by the definition of g, we have

$$\int_0^{+\infty} g(t)e^{-st}\, dt = \int_c^{+\infty} f(t-c)e^{-st}\, dt \, .$$

Finally, the substitution $\tau = t - c$ yields

$$\int_0^{+\infty} g(t)e^{-st}\, dt = \int_0^{+\infty} f(\tau)e^{-s(\tau+c)}\, d\tau = e^{-cs}F(s).$$

■

Property 4 (Multiplication by t^n) *Let f be a piecewise continuous, subexponential function, defined in* $[0, +\infty)$*, of order σ_0, and let $F(s) = \mathcal{L}[f(t)]$ for* $\text{Re}\, s > \sigma_0$. *For each $n \in \mathbf{N}$, the function $t^n f(t)$ is a subexponential function of order σ_0, and*

$$\mathcal{L}[t^n f(t)] = (-1)^n \frac{d^n F(s)}{ds^n} \, , \qquad \text{Re}\, s > \sigma_0 \, . \tag{B.6}$$

Proof The subexponential property of $t^n f(t)$ is a consequence of Lemma 2.1. Formula (B.6) can be proved by induction. The case $n = 0$ follows immediately by definition. Assuming that (B.6) holds for $n = k$, the case $n = k + 1$ can be obtained by computing the derivative[3] of both sides with respect to s.

■

[3]Here and in other following proofs, the crucial point consists in exchanging the order of certain operations like limits, derivatives, integrals. The correctness of such exchanges requires some uni-

Property 5 (Multiplication by e^{at}) *Let f be a piecewise continuous, subexponential function, defined in* $[0, +\infty)$*, of order σ_0, and let $F(s) = \mathcal{L}[f(t)]$ for* $\mathrm{Re}\, s > \sigma_0$. *Let moreover $a \in \mathbf{R}$. Then, $e^{at} f(t)$ is a subexponential function of order $\sigma_0 + a$, and*

$$\mathcal{L}[e^{at} f(t)] = F(s-a) \qquad \mathrm{Re}\, s > \sigma_0 + a\ . \tag{B.7}$$

Proof It is easy to check that $e^{at} f(t)$ is subexponential. Moreover,

$$\int_0^{+\infty} e^{at} f(t) e^{-st}\, dt = \int_0^{+\infty} f(t) e^{-t(s-a)}\, dt = \int_0^{+\infty} f(t) e^{-tr}\, dt = F(r)$$

with $\mathrm{Re}\, r > \sigma_0$, where $r = s - a$. But $\mathrm{Re}\, r = \mathrm{Re}\, s - a$ and so $\mathrm{Re}\, r > \sigma_0$ is equivalent to $\mathrm{Re}\, s > \sigma_0 + a$.

■

Property 5 extends in the case $a \in \mathbf{C}$, with $\mathrm{Re}\, s > \sigma_0 + \mathrm{Re}\, a$.

Next properties are the most important from our point of view, since they refer to the behavior of the operator $\mathcal{L}$ with respect to the operations of the differential and integral calculus.

Property 6 *Assume that f is a piecewise continuous, subexponential function defined in* $[0, +\infty)$*, of order σ_0. Assume further that its derivative f' exists and it is piecewise continuous in* $[0, +\infty)$*. If $F(s) = \mathcal{L}[f(t)]$ for* $\mathrm{Re}\, s > \sigma_0$*, then the Laplace transform of f' exists, and it is given by*

$$\mathcal{L}[f'(t)] = -f(0) + sF(s) \qquad for\ \ \mathrm{Re}\, s > \sigma_0\ . \tag{B.8}$$

Proof We want to compute

$$\lim_{\xi \to +\infty} \int_0^{\xi} f'(t) e^{-st}\, dt\ .$$

Integrating by parts we have

$$\begin{aligned}\int_0^{\xi} f'(t) e^{-st}\, dt &= f(t) e^{-st}\Big|_0^{\xi} + s \int_0^{\xi} f(t) e^{-st}\, dt \\ &= f(\xi) e^{-s\xi} - f(0) + s \int_0^{\xi} f(t) e^{-st}\, dt\ .\end{aligned}$$

The statement is proved, taking the limit for $\xi \to +\infty$, and taking into account Lemma B.1.

■

formity assumptions, which are not difficult to ensure when we work with continuous functions defined on compact intervals. In our framework (complex variables, unbounded intervals) there are some additional technical difficulties. We do not enter in these details..

Property 7 *Let f be a piecewise continuous, subexponential function, defined in* $[0, +\infty)$*, of order* σ_0*, and let* $F(s) = \mathcal{L}[f(t)]$ *for* $\operatorname{Re} s > \sigma_0$*. Then, each antiderivative of f is subexponential of order* $\max\{\sigma_0, 0\}$ *and we have*

$$\mathcal{L}\left[\int_0^t f(\rho)\,d\rho\right] = \frac{F(s)}{s} \qquad for\ \operatorname{Re} s > \max\{\sigma_0, 0\}\,. \tag{B.9}$$

Proof The first statement is left as an exercise. As far as (B.9) is concerned, we can apply again the integration by part rule:

$$\begin{aligned}\int_0^\xi \left(\int_0^t f(\rho)\,d\rho\right) e^{-st}\,dt &= \int_0^\xi h(t)e^{-st}\,dt \\ &= \frac{h(t)e^{-st}}{-s}\Big|_0^\xi + \frac{1}{s}\int_0^\xi h'(t)e^{-st}\,dt \\ &= \frac{h(\xi)e^{-s\xi}}{-s} - \frac{h(0)}{-s} + \frac{1}{s}\int_0^\xi f(t)e^{-st}\,dt\,.\end{aligned}$$

Noticing that $h(0) = 0$, the conclusion follows by taking the limit for $\xi \to +\infty$.

■

Obviously, (B.8) and (B.9) can be iterated, which gives:

$$\mathcal{L}[f^{(k)}(t)] = -f^{(k-1)}(0) - sf^{(k-2)}(0) - \cdots - s^{k-1}f(0) + s^k F(s)\,, \tag{B.10}$$

$$\mathcal{L}\left[\int_0^t \int_0^{t_1} \cdots \int_0^{t_{k-1}} f(t_k)\,dt_k\,dt_{k-1}\ldots dt_1\right] = \frac{F(s)}{s^k}\,. \tag{B.11}$$

Property 8 *Let f and g be two piecewise continuous, subexponential functions, defined in* $[0, +\infty)$*, of order respectively* σ_1 *and* σ_2*. Let* $F(s) = \mathcal{L}[f(t)]$ *for* $\operatorname{Re} s > \sigma_1$ *and* $G(s) = \mathcal{L}[g(t)]$ *for* $\operatorname{Re} s > \sigma_2$*. Let moreover*

$$h(t) = \int_0^t f(t-\rho)g(\rho)\,d\rho\,. \tag{B.12}$$

Then, h is a subexponential function, and $\mathcal{L}[h(t)] = F(s)G(s)$ *for* $\operatorname{Re} s > \max\{\sigma_1, \sigma_2\}$.

Property 8 answers the question of finding a function $h(t)$ such that $\mathcal{L}[h(t)] = F(s)G(s)$, assuming that $F(s) = \mathcal{L}[f(t)]$ and $G(s) = \mathcal{L}[g(t)]$ are known. We remark that (B.12) is well defined, since for $\rho \in [0, t]$ we have $t - \rho \geq 0$. Introducing the following extensions of the functions f and g:

$$\tilde{f}(t) = \begin{cases} 0 & \text{if } t < 0 \\ f(t) & \text{if } t \geq 0 \end{cases} \qquad \text{and} \qquad \tilde{g}(t) = \begin{cases} 0 & \text{if } t < 0 \\ g(t) & \text{if } t \geq 0\,, \end{cases}$$

we can write

$$h(t) = \int_0^t f(t-\rho)g(\rho)\,d\rho = \int_{-\infty}^{+\infty} \tilde{f}(t-\rho)\tilde{g}(\rho)\,d\rho\,.$$

Now let $p(\cdot)$ and $q(\cdot)$ be two piecewise continuous arbitrary functions defined on the whole of **R**. The *convolution* between p and q is defined by

$$(p * q)(t) = \int_{-\infty}^{+\infty} p(t-\rho)q(\rho)\,d\rho\,,$$

provided that the integral is convergent. Thus, we may reformulate (B.12) by writing

$$\mathcal{L}[(\tilde{f} * \tilde{g})(t)] = F(s)G(s)\,.$$

Proof of Property 8 To prove that $h(t)$ is a subexponential function of order $\max\{\sigma_1, \sigma_2\}$ is a simple exercise. With the notation above, we may also write $h(t) = \int_0^{+\infty} \tilde{f}(t-\rho)g(\rho)\,d\rho$. Thus

$$\begin{aligned}\mathcal{L}[h(t)] &= \int_0^{+\infty} \left(\int_0^{+\infty} \tilde{f}(t-\rho)g(\rho)\,d\rho\right) e^{-st}\,dt \\ &= \int_0^{+\infty} g(\rho) \left(\int_0^{+\infty} \tilde{f}(t-\rho)e^{-st}\,dt\right) d\rho\,.\end{aligned}$$

By virtue of Property 3, we finally get

$$\mathcal{L}[h(t)] = \int_0^{+\infty} g(\rho)e^{-\rho s}F(s)\,d\rho = F(s)\int_0^{+\infty} g(\rho)e^{-\rho s}\,d\rho = F(s)G(s)\,.$$

■

B.2 A List of Laplace Transforms

B.2.1 Elementary Functions

We now compute the Laplace transform of some elementary functions.

Proposition B.1 *Let* $f(t) \equiv 1$ *for* $t \geq 0$. *Then*

$$\mathcal{L}[f(t)] = \frac{1}{s} \qquad for\;\; \mathrm{Re}\, s > 0\,.$$

Proof We have

$$\mathcal{L}[f(t)] = \int_0^{+\infty} e^{-st}\,dt = \lim_{\xi\to+\infty}\int_0^{\xi} e^{-st}\,dt = \lim_{\xi\to+\infty}\frac{e^{st}}{-s}\Big|_0^{\xi} = \lim_{\xi\to+\infty}\frac{e^{-s\xi}}{-s} + \frac{1}{s}\,.$$

The conclusion follows from the remark that if $\operatorname{Re} s > 0$, then we have $\lim_{\xi\to+\infty} e^{-s\xi} = 0$.

■

Proposition B.2 *The Laplace transform of the restrictions to the interval* $[0, +\infty)$ *of the power functions, the exponential function, and the trigonometric functions are given by:*

$$\mathcal{L}[at] = \frac{a}{s^2} \qquad for\ \ \operatorname{Re} s > 0\,; \tag{B.13}$$

$$\mathcal{L}[t^n] = \frac{n!}{s^{n+1}} \qquad for\ \ \operatorname{Re} s > 0 \qquad (n \in \mathbf{N})\,; \tag{B.14}$$

$$\mathcal{L}[e^{at}] = \frac{1}{s-a} \qquad for\ \ \operatorname{Re} s > a\,; \tag{B.15}$$

$$\mathcal{L}[\cos\omega t] = \frac{s}{s^2+\omega^2} \qquad for\ \ \operatorname{Re} s > 0\,; \tag{B.16}$$

$$\mathcal{L}[\sin\omega t] = \frac{\omega}{s^2+\omega^2} \qquad for\ \ \operatorname{Re} s > 0\,. \tag{B.17}$$

Proof We will prove formulæ (B.13), (B.14) and (B.15) as applications of Proposition B.1 and Properties 1, 2, 4 and 5. We begin with (B.13). We have

$$\mathcal{L}[at] = a\mathcal{L}[t] = a\mathcal{L}[t\cdot 1]\,.$$

By applying Property 4 and recalling that $\mathcal{L}[1] = 1/s$, we conclude

$$\mathcal{L}[at] = \frac{a}{s^2}\,.$$

Now consider formula (B.14). We have

$$\mathcal{L}[t^n] = \mathcal{L}[t^n\cdot 1] = (-1)^n\frac{d^nF(s)}{ds^n} \qquad \text{where} \qquad F(s) = \frac{1}{s}\,.$$

From this, by mathematical induction, we get $F^{(n)}(s) = (-1)^n n! s^{-(n+1)}$. As far as (B.15) is concerned, we just need to remark that

$$\mathcal{L}[e^{at}] = \mathcal{L}[e^{at}\cdot 1] = \frac{1}{s-a}\,.$$

Now we consider the Laplace transform of the trigonometric functions. As already noticed, the operator $\mathcal{L}$ applies to functions $f : \mathbf{R} \to \mathbf{C}$, as well. Hence we can compute $\mathcal{L}[\cos \omega t]$ making use of the Euler formula

$$\cos \omega t = \frac{e^{\mathrm{i}\,\omega t} + e^{-\mathrm{i}\,\omega t}}{2} \, .$$

We have

$$\mathcal{L}[\cos \omega t] = \frac{1}{2}\left(\mathcal{L}[e^{\mathrm{i}\,\omega t}] + \mathcal{L}[e^{-\mathrm{i}\,\omega t}]\right) = \frac{1}{2}\left(\frac{1}{s + \mathrm{i}\,\omega} + \frac{1}{s - \mathrm{i}\,\omega}\right)$$

for $\mathrm{Re}\, s > 0$. This yields

$$\mathcal{L}[\cos \omega t] = \frac{s}{s^2 + \omega^2} \, .$$

The proof of (B.17) is similar. ■

B.2.2 *Discontinuous functions*

From Propositions B.1 and B.2 it is possible to deduce the Laplace transform of some functions which are commonly used in signal theory. For instance,

$$U(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$

is called the *unit step* or also the *Heaviside function*. It represents a signal which instantaneously jumps from zero to 1 (switch-on). The function $f(t) \equiv 1$ considered in Proposition B.1 coincides with the restriction of $U(t)$ to $[0, +\infty)$. Taking into account the definition of Laplace transform, with a little abuse of notation we will write

$$\mathcal{L}(U(t)) = \frac{1}{s} \qquad (\mathrm{Re} s > 0) \, .$$

Remark B.1 Let us remark that the complex function of a complex variable which associates s to its inverse $1/s$ is defined for each $s \neq 0$. Nevertheless, it is not correct to say that such a function is the Laplace transform of $U(t)$. Indeed the identity $\mathcal{L}[U(t)] = 1/s$ holds only for $\mathrm{Re}\, s > 0$. In other words, $\mathcal{L}[U(t)]$ coincides with the restriction to the positive complex half plane of the function $1/s$.

■

The function $U(t)$ allows us to represent other types of discontinuous signals, whose Laplace transform can be easily computed by applying Properties 1 and 3. For instance,

(i) *switch-on* for $t = c$

$$f(t) = \begin{cases} 0 & \text{for } t < c \\ 1 & \text{for } t \geq c \end{cases}$$

is equivalent to $f(t) = U(t - c)$;

(ii) *switch-off* for $t = c$

$$f(t) = \begin{cases} 1 & \text{for } t \leq c \\ 0 & \text{for } t > 0 \end{cases}$$

is equivalent to $f(t) = 1 - U(t - c)$ (or $= U(c - t)$);

(iii) *rectangular impulse*

$$f(t) = \begin{cases} 0 & \text{for } t < a \\ 1 & \text{for } a \leq t \leq b \\ 0 & \text{for } t > b \end{cases}$$

is equivalent to $f(t) = U(t - a) - U(t - b)$.

The function $U(t)$ is also useful to represent piecewise elementary functions. For instance the function

$$f(t) = \begin{cases} t & \text{for } t < 1 \\ t^2 & \text{for } t \geq 1 \end{cases}$$

can be written as

$$f(t) = t[1 - U(t - 1)] + t^2 U(t - 1) \ .$$

With the same abuse of notation as above, we can think of (B.13) as the Laplace transform of a signal of the form

$$f(t) = \begin{cases} 0 & \text{for } t < 0 \\ at & \text{for } t \geq 0 \end{cases}$$

while (B.16), (B.17) provide the Laplace transform of signal of sinusoidal shape (but vanishing for $t < 0$).

B.2.3 Dirac Delta Function

One of the most important signals typically employed in system theory is the *unit impulse* function, denoted by the symbol $\delta(t)$ and also called Dirac δ function.

$$\delta(t) = \begin{cases} +\infty & \text{if } t = 0 \\ 0 & \text{if } t \neq 0\,. \end{cases} \tag{B.18}$$

This definition is an ideal representation of a signal of very large energy, concentrated at a single point. Of course, (B.18) is nonsense from a rigorous point of view, since it really does not define a function $\mathbf{R} \to \mathbf{R}$. There is a theory, called *distribution theory*, based on a generalization of the notion of function, which allows us to formally introduce and study objects like (B.18). To our purposes, it is sufficient to think of the function $\delta(t)$ as the limit of suitable sequences; for instance

$$\delta(t) = \lim_{\varepsilon \to 0^+} \frac{1}{2\varepsilon} \left(U(t+\varepsilon) - U(t-\varepsilon) \right) \,. \tag{B.19}$$

From (B.19) we infer in particular that

$$\int_{-\infty}^{+\infty} \delta(t)\, dt = \int_{-k}^{k} \delta(t)\, dt = 1 \qquad \forall k > 0\,. \tag{B.20}$$

It is possible to define the sum and the multiplication between generalized functions like $\delta(t)$. It is also possible to give a sense to certain operators of the differential calculus for generalized functions, but this is not required in this book. We limit ourselves to recall some facts and properties related to the Dirac delta function.

An impulse of intensity k concentrated at a point $a \in \mathbf{R}$ is represented by $k \cdot \delta(t-a)$. We have

$$\int_{-\infty}^{+\infty} k \cdot \delta(t-a) dt = k \tag{B.21}$$

and

$$\int_{-\infty}^{+\infty} f(t)\delta(t-a) dt = f(a) \tag{B.22}$$

provided that the function f is continuous at the point $t = a$. Finally,

$$\int_{-\infty}^{t} \delta(\tau) d\tau = U(t) \tag{B.23}$$

and

$$\mathcal{L}[\delta(t)] = 1\,. \tag{B.24}$$

Formulæ (B.20)–(B.24) can be formally proved in the context of distribution theory. They can be also justified heuristically on the base of (B.19). For instance, concerning (B.24) we suggest the following argument.

$$\begin{aligned}\mathcal{L}[\delta(t)] &= \mathcal{L}\left[\lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\big(U(t)-U(t-\varepsilon)\big)\right] \\ &= \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\{\mathcal{L}[U(t)]-\mathcal{L}[U(t-\varepsilon)]\} = \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\left\{\frac{1}{s}-\frac{e^{-\varepsilon s}}{s}\right\} \\ &= \lim_{\varepsilon\to 0^+}\frac{1}{s}\cdot\frac{1-e^{-\varepsilon s}}{\varepsilon} = -\lim_{\varepsilon\to 0^+}\frac{e^{-\varepsilon s}-1}{\varepsilon s} = 1\,.\end{aligned}$$

B.3 Inverse Transform

The following proposition guarantees the existence of the inverse transformation $\mathcal{L}^{-1}$.

Proposition B.3 *Let f and g be two continuous functions defined on $[0, +\infty)$, and let $F(s) = \mathcal{L}[f(t)]$, $G(s) = \mathcal{L}[g(t)]$, both defined for $\operatorname{Re} s > \sigma_0$. If $F(s) = G(s)$ for all s such that $\operatorname{Re} s > \sigma_0$, then $f(t) = g(t)$ for all $t \geq 0$.*

This proposition states that $f(t)$ can be uniquely reconstructed from $F(s)$, and so it allows us to define $\mathcal{L}^{-1}$. However, it should be noted that if $f(t)$ is the restriction to $[0, +\infty)$ of a function $\varphi(t)$ defined on a larger interval $[a, +\infty)$ with $-\infty \leq a < 0$, in general it is not possible to reconstruct $\varphi(t)$ on the interval $(a, 0)$ by applying $\mathcal{L}^{-1}$ to $F(s)$. Nevertheless, in many applications this problem can be overcome if it is known a priori that $f(t)$ is real analytic.

The inverse transformation $\mathcal{L}^{-1}$ can be explicitly represented by a suitable formula, but the use of this formula is not needed in this book. We limit ourselves to remark that $\mathcal{L}^{-1}$ is, like $\mathcal{L}$, a linear operator.

B.4 The Laplace Transform of a Vector Function

The extension of the Laplace transform to vector functions

$$f : [0, +\infty) \to \mathbf{R}^n \quad \text{or} \quad f : [0, +\infty) \to \mathbf{C}^n$$

where $f = (f_1, \ldots, f_n)$, is straightforward: under the assumption that each component is a subexponential function, we set $\mathcal{L}[f] = (\mathcal{L}[f_1], \ldots, \mathcal{L}[f_n])$.

The aforementioned properties of the Laplace transform can be easily extended, as well. In addition, we have

$$\mathcal{L}[Mf(t)] = M\mathcal{L}[f(t)] \tag{B.25}$$

for each matrix M with real or complex constant entries. We are especially interested in the transform of the exponential matrix and in the convolution product formula (Property 8).

Proposition B.4 *Let A be a square matrix (real or complex), c be a constant vector and $b(t)$ be a vector function whose components are piecewise continuous, subexponential functions. Let σ_0 be the maximal real part of the eigenvalues of A. Then, for each $s \in \mathbf{C}$ such that* $\mathrm{Re}\, s > \sigma_0$, *we have:*

$$\mathcal{L}[e^{tA}c] = -(A - sI)^{-1}c\ , \tag{B.26}$$

$$\mathcal{L}\left[\int_0^t e^{(t-\tau)A}b(\tau)d\tau\right] = -(A - sI)^{-1}B(s)\ . \tag{B.27}$$

Proof By definition,

$$\mathcal{L}[e^{tA}c] = \int_0^{+\infty} e^{-st}e^{tA}cdt = \int_0^{+\infty} e^{t(A-sI)}c\,dt = \lim_{\xi\to\infty}\int_0^{\xi} e^{t(A-sI)}c\,dt\ .$$

If $\mathrm{Re}\, s > \sigma_0$, then det $(A - sI) \neq 0$ and the inverse $(A - sI)^{-1}$ exists. On the other hand, it is well known that for each square matrix M, the exponential matrix e^{tM} admits a derivative and $(e^{tM})' = Me^{tM} = e^{tM}M$. This implies that if M is invertible, $\int e^{tM} = M^{-1}e^{tM} = e^{tM}M^{-1}$. We can therefore proceed in the following way:

$$\begin{aligned}\mathcal{L}[e^{tA}c] &= \lim_{\xi\to\infty}(A - sI)^{-1}e^{t(A-sI)}c\Big|_0^{\xi}\\ &= \lim_{\xi\to\infty}\left[(A - sI)^{-1}e^{\xi(A-sI)}c - (A - sI)^{-1}c\right]\ .\end{aligned}$$

The assumption that $\mathrm{Re}\, s > \sigma_0$ also implies that all the eigenvalues of $A - sI$ have negative real part. Indeed, it is clear that the eigenvalues μ of $A - sI$ have the form $\mu = \lambda - s$ where λ is an eigenvalue of A . But then $\mathrm{Re}\,\mu = \mathrm{Re}\,\lambda - \mathrm{Re}\, s < 0$.

We know that if all the eigenvalues of a matrix M have negative real part, then for each c we have $\lim\limits_{\xi\to+\infty} e^{\xi M}c = 0$. In conclusion,

$$\mathcal{L}[e^{tA}c] = -(A - sI)^{-1}c$$

as required. As far as (B.27) is concerned, we remark that

$$\begin{aligned}\mathcal{L}\left[\int_0^t e^{(t-\tau)A}b(\tau)d\tau\right] &= \int_0^{+\infty} e^{-st}\left(\int_0^t e^{(t-\tau)A}b(\tau)d\tau\right)dt\\ &= \int_0^{+\infty} e^{-\tau A}\left(\int_{\tau}^{+\infty} e^{t(A-sI)}\,dt\right)b(\tau)\,d\tau\ .\end{aligned}$$

Note the change of the integration interval due to the change of integration order. Making use of the assumption that $\mathrm{Re}\, s > \sigma_0$, we finally conclude

$$\begin{aligned}\mathcal{L}\left[\int_0^t e^{(t-\tau)A} b(\tau)d\tau\right] &= -\int_0^{+\infty} e^{-\tau A}(A-sI)^{-1}e^{\tau(A-sI)}b(\tau)d\tau \\ &= -\int_0^{+\infty}(A-sI)^{-1}e^{-s\tau}b(\tau)d\tau \\ &= -(A-sI)^{-1}\int_0^{+\infty} e^{-s\tau}b(\tau)d\tau \\ &= -(A-sI)^{-1}B(s)\ .\end{aligned}$$

■

Appendix Summary

Appendix B recalls the definition of Laplace transform and its main properties. Moreover, we give a list of the Laplace transforms of some elementary functions.

References

1. Ahlfors LV (1979) Complex analysis: an introduction to the theory of analytic functions of one complex variable. McGraw-Hill, New York
2. Bacciotti A, Rosier L (2005) Liapunov functions and stability in control theory. Springer, Berlin
3. Barnett S (1975) Introduction to mathematical control theory. Clarendon Press, Oxford
4. Bernstein DS (2005) Matrix mathematics: theory facts and formulas with application to linear systems theory. Princeton University Press, Princeton
5. Bressan A, Piccoli B (2007) An introduction to the mathematical theory of control. American Institute of Mathematical Sciences, Springfields
6. Brockett RW (1970) Finite dimensional linear systems. Wiley, New York
7. Brockett R (2014) The early days of geometric nonlinear control. Automatica 50(9):2203–2224
8. Callier FM, Desoer CA (1982) Multivariable feedback systems. Springer, New York
9. Cao YY, Lam J, Sun YX (1998) Static output feedback stabilization: an ILMI approach. Automatica 34(12):1641–1645
10. Coppel WA (1965) Stability and asymptotic behaviour of differential equations. Heath Mathematical Monographs, Boston
11. Conti R (1976) Linear differential equations and control. Academic Press, London
12. Conti R (1985) Processi di controllo lineari in R^n, quaderni UMI 30. Pitagora Editrice, Bologna (in Italian)
13. Coron JM (2007) Control and nonlinearity. American Mathematical Society, Providence
14. Dragan V, Halanay A (1999) Stabilization of linear systems. Birkhäuser, Boston
15. Guest PB (1991) Laplace transforms and an introduction to distributions. Ellis Horwood, New York
16. Hermes H, LaSalle JP (1969) Functional analysis and time optimal control. Academic Press, New York
17. Hirsch MW, Smale S (1974) Differential equations, dynamical systems and linear algebra. Academic Press, New York
18. Isidori A (1989) Nonlinear control systems. Springer, Berlin
19. Isidori A (2017) Lectures in feedback design for multivariable systems. Springer

20. Jurdjevic V (1997) Geometric control theory. Cambridge University Press, Cambridge
21. Lang S (1966) Linear algebra. Addison Wesley, Reading Mass
22. LaSalle J (1986) The stability and control of discrete process. Springer, New York
23. Lee EB, Markus L (1967) Foundations of optimal control theory. Wiley, New York
24. Pontrjagin L (1975) Équations différentielles ordinaires. MIR, Moscou
25. Rosenbrock HH (1970) State-space and multivariable theory. Nelson, London
26. Sontag ED (1990) Mathematical control theory. Springer, New York
27. Terrel WJ (2009) Stability and stabilization: an introduction. Princeton press, Princeton
28. Trentelman HJ, Storvogel AA, Hautus M (2001) Control theory for linear systems. Springer, London
29. Tsinias J, Kalouptsidis N (1990) Output feedback stabilization. IEEE Trans Auto Control 35(8):951–954
30. Vidyasagar M (1993) Nonlinear systems analysis. Prentice Hall, Englewood Cliffs
31. Wonham WM (1979) Linear multivariable control: a geometric approach. Springer, New York
32. Zadeh LA, Desoer CA (1963) Linear system theory. McGraw-Hill, New York

Index

A. Bacciotti, *Stability and Control of Linear Systems*, Studies in Systems, Decision and Control 185, https://doi.org/10.1007/978-3-030-02405-5