Mathematical Systems Theory

Mathematical Systems Theory

fourth edition

G.J. Olsder
J.W. van der Woude
J.G. Maks
D. Jeltsema

Faculty of
Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

VSSD

Fourth edition 2011

Published by VSSD
Leeghwaterstraat 42, 2628 CA Delft, The Netherlands
tel. +31 15 27 82124, telefax +31 15 27 87585, e-mail: hlf@vssd.nl
internet: http://www.vssd.nl/hlf
URL about this book: http://www.vssd.nl/hlf/a003.htm

Printed version 4th edition
ISBN-13 978-90-6562-280-8

NUR 919

Key words: mathematical systems theory

Preface

Fourth edition

The major difference between this fourth edition and its predecessor is the presentation of the material in Chapter 3. The method of linearization of a system described in this chapter has now been restricted to solution-input pairs that are constant in time (equilibrium pairs) and the presentation of the analytical method of solving a linear system is restricted to the linear systems that are time-invariant. We firmly believe that both these restrictions are a great improvement from a didactic point of view. Another important change in this chapter concerns the introductory text, which we believe to be an improvement in the sense that the connection with the previous chapter on modeling is now described more explicitly. A change in Chapter 4 which is worth mentioning is given by the passage where we deal with the duality between the concepts of controllability and observability. We have given a qualitative interpretation of this phenomenon of duality which we believe to be a useful addition to the merely symbolic method using the transposition of matrices.

Delft, November 2011 J.W. van der Woude, J.G. Maks and D. Jeltsema

Third edition

Compared to the second edition, the presentation of material in this third edition has been changed significantly. For a start, based on feedback by students, certain topics, like linearization, Routh's criterion, interval stability, observer and compensator design, have been discussed in some more detail than in the second edition. Further, in each chapter theorems, lemmas, examples, and so on, are numbered consecutively now, and exercises have been moved towards the end of chapters. Also additional exercises have been included. Finally, errors and typos, found in the second edition, have been corrected. A.A. Stoorvogel and J.G. Maks are greatly acknowledged for their remarks on the second edition. We also thank VSSD for its willingness to publish these notes as a book. We hope that this third edition will be as successful as the previous ones.

Delft, November 2004 G.J. Olsder and J.W. van der Woude

Second edition

The main changes of this second edition over the first one are (i) the addition of a chapter with MATLAB®[1] exercises and possible solutions, and (ii) the chapter on 'Polynomial representations' in the first edition has been left out. A summary of that chapter now appears as a section in chapter 8. The material within the chapter on 'Input/output representations' has been shifted somewhat such that the parts dealing with frequency methods form one section now. Moreover, some exercises have been added and some mistakes have been corrected. I hope that this revised edition will find its way as its predecessor did.

Delft, December 1997 G.J. Olsder

First edition

These course notes are intended for use at undergraduate level. They are a substantial revision of the course notes used during the academic years 1983-'84 till 1993-'94. The most notable changes are an omission of some abstract system formulations and the addition of new chapters on modelling principles and on polynomial representation of systems. Also changes and additions in the already existing chapters have been made. The main purpose of the revision has been to make the student familiar with some recently developed concepts (such as 'disturbance rejection') and to give a more complete overview of the field.

A dilemma for any author of course notes, of which the total contents is limited by the number of teaching hours and the level of the students (and of the author!), is what to include and what not. One extreme choice is to treat a few subjects in depth and not to talk about the other subjects at all. The other extreme is to touch upon all subjects only very briefly. The choice made here is to teach the so-called state space approach in reasonable depth (with theorems and proofs) and to deal with the other approaches more briefly (in general no proofs) and to provide links of these other approaches with the state space approach.

The most essential prerequisites are a working knowledge of matrix manipulations and an elementary knowledge of differential equations. The mathematics student will probably experience these notes as a blend of techniques studied in other (first and second year) courses and as a solid introduction to a new field, viz. that of mathematical system theory, which opens vistas to various fields of application. The text is also of interest to the engineering student, who will, with his background in applications, probably experience these notes as more fundamental. Exercises are interspersed throughout the text; the student should not skip them. Unlike many mathematics texts, these notes contain more exercises (61) than definitions (31) and more examples (56) that theorems (36).

For the preparation of these notes various sources have been consulted. For the first edition such a source was, apart from some of the books mentioned in the bibliography, 'Inleiding wiskundige systeemtheorie' by A.J. van der Schaft, Twente University of

[1] MATLAB is a registered trademark of The MathWorks, Inc.

Technology. For the preparation of these revised notes, also use was made of 'Course d'Automatique, Commande Linéaire des Systèmes Dynamiques' by B. d'Andréa-Novel and M. Cohen de Lara, Ecole Nationale Supérieure des Mines de Paris. The contents of Chapter 2 have been prepared by J.W. van der Woude, which is gratefully acknowledged. The author is also grateful to many of his colleagues with whom he had discussions about the contents and who sometimes proposed changes. The figures have been prepared by Mrs T. Tijanova, who also helped with some aspects of the LaTeX document preparation system by means of which these notes have been prepared.

Parallel to this course there are computer lab sessions, based on MATLAB, by means of which the student himself can play with various examples such as to get a better feeling for concepts and for designing systems himself. This lab has been prepared by P. Twaalfhoven and J.G. Braker.

Delft, April 1994 G.J. Olsder

Contents

Chapter 1

Introduction

1.1 What is mathematical systems theory?

A system is part of reality which we think to be a separated unit within this reality. The reality outside the system is called the surroundings. The interaction between system and surroundings is realized via quantities, quite often functions of time, which are called input and output. The system is influenced via the input(-functions) and the system itself has an influence on the surroundings by means of the output(-functions).

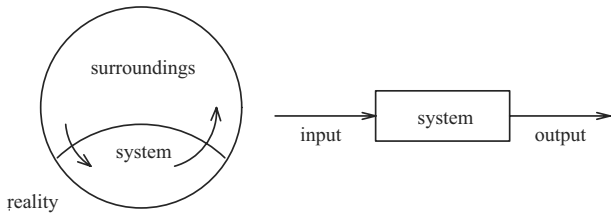

Figure 1.1 A system in interaction with its environment.

Three examples:

- How to fly an aeroplane: the position of the control wheel (the input) has an influence on the course (the output).
- In economics: the interest rate (the input) has an influence on investment behavior (the output).
- Rainfall (the input) has an influence on the level of the water in a river (the output).

In many fields of study, a phenomenon is not studied directly but indirectly through a model of the phenomenon. A model is a representation, often in mathematical terms, of what are felt to be the important features of the object or system under study. By the manipulation of the representation, it is hoped that new knowledge about the modelled phenomenon can be obtained without the danger, cost, or inconvenience of manipulating the real phenomenon itself. In mathematical system theory we only work with models and when talking about a system we mean a modelled version of the system as part of reality.

Most modelling uses mathematics. The important features of many physical phenomena can be described numerically and the relations between these features can be described by equations or inequalities. Particularly in natural sciences and engineering, properties such as mass, acceleration and forces can be described in mathematical terms.

To successfully utilize the modelling approach, however, knowledge is required of both the modelled phenomena and properties of the modelling technique. The development of high-speed computers has greatly increased the use and usefulness of modelling. By representing a system as a mathematical model, converting it into instructions for a computer and running the computer, it is possible to model larger and more complex systems than ever before.

Mathematical system(s) theory is concerned with the study and control of input/output phenomena. There is no difference between the terminologies 'system theory' and 'systems theory'; both are used in the (scientific) literature and will be used interchangeably. The emphasis in system(s) theory is on the dynamic behavior of these phenomena, i.e., how do characteristic features (such as input and output) change in time and what are the relationships between them, also as functions of time. One tries to design control systems such that a desired behavior is achieved. In this sense mathematical system(s) theory (and control theory) distinguishes itself from many other branches of mathematics by the fact that is prescriptive rather than descriptive.

Mathematical system theory forms the mathematical base for technical areas such as automatic control and networks. It is also the starting point for other mathematical subjects such as optimal control theory and filter theory. In optimal control theory one tries to find an input function which yields an output function that satisfies a certain requirement as well as possible. In filter theory the input to the system, then being a so-called filter, consists of observations with measurement errors, while the system itself tries to realize an output which equals the 'ideal' observations, i.e., as much as possible without measurement errors. Mathematical system theory also plays a role in economics (specially in macro-economic control theory and time series analysis), theoretical computer science (via automaton theory and Petri-nets) and management science (models of firms and other organizations). Lastly, mathematical system theory forms the hard, mathematical, core of more philosophically oriented areas such as general systems theory and cybernetics.

Example 1.1 [Autopilot of a boat] An autopilot is a device which receives as input the heading $\alpha(t)$ of a boat at time t (measured by an instrument such as a magnetic compass or a gyrocompass) and the (fixed) desired heading α_c (reference point) by the navigator. Using this information, the device automatically yields, as a function of time t, the positioning command $u(t)$ of the rudder in order to achieve the smallest possible heading error $e(t) = \alpha_c - \alpha(t)$. Given the dynamics of the boat and the external perturbations (wind, swell, etc.) the theory of automatic control helps to determine a control input command $u = f(e)$ that meets the imposed technical specifications (stability, accuracy, response time, etc.). For example, this control might be bang-bang:

$$u(t) = \begin{cases} +u_{\max} & \text{if} \quad e(t) > 0, \\ -u_{\max} & \text{if} \quad e(t) < 0. \end{cases}$$

(The arrows in the left-hand picture in Figure 1.2 point in the positive direction of the quantities concerned.) Alternatively, the control might be proportional:

$$u(t) = Ke(t),$$

where K is a constant. It has tacitly been assumed here that for all e-values of interest, $-u_{\max} \leq Ke(t) \leq u_{\max}$. If this is not the case, some kind of saturation must be intro-

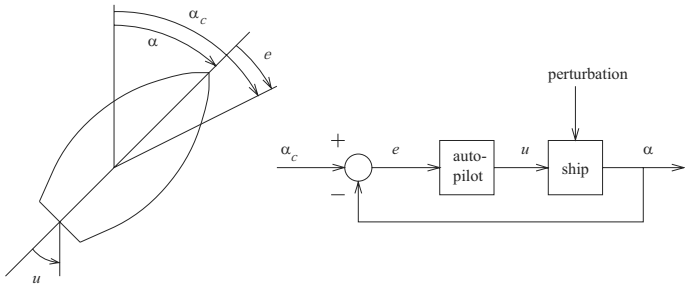

Figure 1.2 Autopilot of a boat.

duced. The control law might also consist of a proportional part, an integrating part and a differentiating part:

$$u(t) = Ke(t) + K' \int^t e(s)ds + K'' \frac{d}{dt} e(t), \tag{1.1}$$

where K, K' and K'' are constants. This control law is sometimes referred to as a PID controller, where P stands for the proportional part, I for the integrating part and D for the differentiating part. The lower bound of the integral in (1.1) has not been given explicitly; various choices are possible. In all these examples of a control law, a signal (the error in this case) is fed back to the input. One speaks of control by feedback.

Automatic control theory helps in the choice of the best control law. If the ship itself is considered as a system, then the input to the ship is the rudder setting u (and possibly perturbations) and the output is the course α. The autopilot is another system. Its input is the error signal e and output is the rudder setting u. Thus, we see that the output of one system can be the input of another system. The combination of ship, autopilot and the connection from α to α_c, all depicted in the right-hand side of Figure 1.2, can also be considered as a system. The inputs of the combined system are the desired course α_c and possible perturbations, and the output is the real course α. □

Example 1.2 [Optimal control problem] The motion of a ship is described by the differential equation

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

where the so-called state vector $x = (x_1,x_2)^\top \in \mathbb{R}^2$ equals the ship's position with respect to a fixed reference frame, the vector $u = (u_1,u_2)^\top \in \mathbb{R}^2$ denotes the control and t is the time. The superscript $^\top$ refers to 'transposed'. If not explicitly stated differently, vectors are always supposed to be column vectors. Although not specifically indicated, both x and u are supposed to be functions of time. The notation $\dot{x}$ refers to the time derivative of the (two) state components. In this example one control variable to be chosen is the ship's heading u_1, whereas the other one, u_2, is the ship's velocity. The problem now is to choose u_1 and u_2 in such a way that the ship uses as little fuel as possible such that, if it leaves Rotterdam at a certain time, it arrives in New York not more than 10 days later. The functions u_1 and u_2 may depend on available information such as time, weather forecast,

ocean streams, and so on. Formally, $u = (u_1, u_2)^\top$ must be chosen such that

$$\int_{t_0}^{t_f} g(x,u,t)dt$$

is minimized, where the (integral) criterion describes the amount of fuel used. The function g is the fuel consumption per time unit, t_0 is the departure time and t_f is the arrival time. □

Example 1.3 [Filtering] NAVSAT is the acronym for NAVigation by means of SATellites. It refers to a worldwide navigation system studied by the European Space Agency (ESA). During the 1980s the NAVSAT system was in the development phase with feasibility studies being performed by several European aerospace research institutes. At the National Aerospace Laboratory (NLR), Amsterdam, the Netherlands, for instance, a simulation tool was developed by which various alternative NAVSAT concepts and scenarios could be evaluated.

Recently, the United States and the European Union have reached an agreement on sharing their satellite navigation services, i.e., the current U.S. Global Positioning System and Europe's Galileo system, which is scheduled to be in operation by 2008. NAVSAT can be seen as a forerunner of Galileo.

The central idea of satellite based navigation system is the following. A user (such as an airplane, a ship or a car) receives messages from satellites, from which he can estimate his own position. Each satellite broadcasts its own coordinates (in some known reference frame) and the time instant at which this message is broadcast. The user measures the time instant at which he receives this message on his own clock. Thus, he knows the time difference between sending and receiving the message, which yields the distance between the position of the satellite and the user. If the user can calculate these distances with respect to at least three different satellites, he can in principle calculate his own position. Complicating factors in these calculations are: (i) different satellites send messages at different time instants, while the user moves in the meantime, (ii) several different sources of error present in the data, e.g. unknown ionospheric and tropospheric delays, the clocks of the satellites and of the user not running exactly synchronously, the satellite position being broadcast with only limited accuracy.

The problem to be solved by the user is how to calculate his position as accurately as possible, when he gets the information from the satellites and if he knows the stochastic characteristics of the errors or uncertainties mentioned above. As the satellites broadcast the information periodically, the user can update also periodically the estimate of his position, which is a function of time. □

1.2 A brief history

Feedback - the key concept of system theory - is found in many places such as in nature and in living organisms. An example is the control of the body temperature. Also, social and economic processes are controlled by feedback mechanisms. In most technical equipment use is made of control mechanisms.

In ancient times feedback was already applied in for instance the Babylonic water wheels and for the control of water levels in Roman aqueducts. According to historian

Otto Mayr, the first explicit use of a feedback mechanism has been designed by Cornelis Drebbel [1572–1633], both an engineer and an alchemist. He designed the 'Athanor', an oven in which he optimistically hoped to change lead into gold. Control of the temperature in this oven was rather complex and the method invented by Drebbel could be viewed as a feedback design.

Drebbel's invention was then used for commercial purposes by his son in law, Augustus Kuffler [1595–1677], being a contemporary of Christian Huygens [1629–1695]. It was Christian Huygens who designed a fly-wheel for the control of the rotational speed of windmills. This idea was refined by R. Hooke [1635–1703] and J. Watt [1736–1819], the latter being the inventor of the steam engine. In the middle of the 19th century more than 75,000 James Watt's fly-ball governors (see Exercise 1.4.2) were in use. Soon it was realized that these contraptions gave problems if control was too rigid. Nowadays one realizes that the undesired behavior was a form of instability due to a high gain in the feedback loop. This problem of bad behavior was investigated J.C. Maxwell [1831–1879] – the Maxwell of the electromagnetism – who was the first to perform a mathematical analysis of stability problems. His paper 'On Governors' can be viewed as the first mathematical article devoted to control theory.

The next important development started in the period before the Second World War, in the Bell Labs in the USA. The invention of the electronic amplification by means of feedback started the design and use of feedback controllers in communication devices. In the theoretical area, frequency-domain techniques were developed for the analysis of stability and sensitivity. H. Nyquist [1889–1976] and H.W. Bode [1905–1982] are the most important representatives of this direction.

Norbert Wiener [1894–1964] worked on the fire-control of anti-aircraft defence during the Second World War. He also advocated control theory as some kind of artificial intelligence as an independent discipline which he called 'Cybernetics' (this word was already used by A.M. Ampere [1775–1836]).

Mathematical system theory and automatic control, as known nowadays, found their feet in the 1950s. Classic control theory played a stimulating role. Initially mathematical system theory was more or less a collection of concepts and techniques from the theory of differential equations, linear algebra, matrix theory, probability theory, statistics, and, to a lesser extent, complex function theory. Later on (around 1960) system theory got its own face, i.e., 'own' results were obtained which were especially related to the 'structure' of the 'box' between input and output, see the right-hand side picture in Figure 1.1. Two developments contributed to that. Firstly, there were fundamental theoretical developments in the nineteen fifties. Names attached to these developments are R. Bellman (dynamic programming), L.S. Pontryagin (optimal control) and R.E. Kalman (state space models and recursive filtering). Secondly, there was the invention of the chip at the end of the nineteen sixties and the subsequent development of micro-electronics. This led to cheap and fast computers by means of which control algorithms with a high degree of complexity could really be used.

1.3 Brief description of contents

In the present chapter a very superficial overview is given of what system theory is and the relations with other (mainly: technically oriented) fields are discussed. One could say that in this chapter the 'geographical map' is unfolded and that in the subsequent chapters parts of the map are studied in (more) detail.

In Chapter 2 modelling techniques are discussed and as such the chapter, strictly speaking, does not belong to the area of system theory. Since, however, the starting point in system theory always is a model or a class of models, it is important to know about modelling techniques and the principles underlying such models. Such principles are for instance the conservation of mass and of energy. A classification of the variables involved into input (or: control) variables, output (or: measurement) variables, and variables which describe dependencies within the model itself, will become apparent.

In Chapters 3, 4 and 5 the theory around the important class of linear differential systems is dealt with. The reason for studying such systems in detail is twofold. Firstly, many systems in practice can (at least: approximately) be described by linear differential systems. Secondly, the theory for these systems has been well developed and has matured during the last forty years or so. Many concepts can be explained quite naturally for such systems.

The view on systems is characterized by the 'state space approach' and the main mathematical technique used is that of linear algebra. Besides linear algebra one also encounters matrix theory and the theory of differential equations. Chapter 3 deals specifically with linearization and linear differential systems. Chapter 4 deals with structural properties of linear systems. Specially, various forms of stability and relationships between the input, output and state of the system, such as controllability and observability, are dealt with. Chapter 5 considers feedback issues, both state feedback and output feedback, in order to obtain desired system properties. The description of the separation principle is also part of this chapter.

Chapter 6 also deals with linear systems, but now from the input/output point of view. One studies formulas which relate inputs to outputs directly. Main mathematical tools are the theory of the Laplace transform and complex function theory. The advantage of this kind of system view is that systems can easily be viewed as 'blocks' and that one can build larger systems by combining subsystems. A possible disadvantage is that this way of describing systems is essentially limited to linear time-invariant systems, whereas the state space approach of the previous chapters is also suited as a means of describing nonlinear and/or time-dependent systems.

In Chapters 3, 4, 5 and 6 'time' was considered to flow continuously. In Chapter 7 one deals with 'discrete-time' models. Rather than differential equations one now has difference equations which describe the model from the state space point of view. The most crucial concepts of Chapters 4 and 5 are repeated here for such systems. The role of the Laplace transform is taken over by the so-called z-transform. The theories of continuous-time systems and of discrete-time systems are equivalent in many aspects, and therefore Chapter 7 has been kept rather brief. Some modelling pitfalls when approximating a continuous-time system by a discrete-time one are briefly indicated.

Chapter 8 shows some avenues towards related fields. There is an abstract point of

view on systems, characterizing them in terms of input space, output space, and maybe state space, and the mappings between these spaces. Also the more recently introduced 'behavioral approach' towards system theory is briefly mentioned. In this approach no distinction is made between inputs and outputs. It is followed by a brief introduction of polynomial matrices used to represent linear systems algebraically. Some remarks on nonlinear systems – a class many times larger than the class of linear systems – will be made together with some progress in this direction. Also other types of systems are mentioned such as descriptor systems, stochastic systems, finite state systems, distributed parameter systems and discrete event systems. Brief introductions to optimal control theory, filter theory, model reduction, and adaptive and robust control will be given. In those fields system theoretical notions introduced earlier are used heavily.

Lastly, Chapter 9 contains a collection of problems and their solutions that can be used for a course on system theory. The problems are solved using the software package MATLAB. For most of them also the MATLAB *Control Toolbox* must be used. The nature of this chapter is clearly different from that of the others.

Books mentioned in the text and some 'classics' in the field of systems theory are given in the bibliography. This book ends with an index.

1.4 Exercises

Exercise 1.4.1 *The water clock ('clepsydra') invented by Ktesibios, a Greek of the third century before Christ, is an old and very well known example of* ***feedback control*** *(i.e., the error is fed back in order to make corrections). Look this up and give a schematic drawing of the water clock with control.*

Exercise 1.4.2 *Another example of an old control mechanism is Watt's centrifugal governor for the control of a steam engine. Consult the literature and find out how this governor works. See for instance [Faurre and Depeyrot, 1977].*

Exercise 1.4.3 *Determine how a float in the water reservoir of a toilet operates.*

Exercise 1.4.4 *Investigate the working of a thermostat in the central heating of a greenhouse. Specify the controls and the measurements.*

Exercise 1.4.5 *Describe how feedback plays a role when riding a bicycle. What are the inputs/controls and what are the outputs/measurements.*

Exercise 1.4.6 *Investigate the mechanism of your body to control its temperature. What is the control action?*

Chapter 2

Some Modelling Principles

In this chapter we present some tools that can be used in the modelling of dynamical phenomena. This chapter does not give an exhaustive treatment of such tools, but it is meant as an introduction to some of the underlying principles. One could argue that modelling principles do not belong to the domain of mathematical system theory. Indeed, in the latter theory one usually starts with a given model, perhaps built by an expert in the field of application.

2.1 Conservation laws

One of the most fundamental modelling principles is the notion of conservation. The laws derived from this notion follow from natural reasoning and can be applied everywhere.

For instance, when modelling physical phenomena, one often uses (even without realizing) conservation of matter, conservation of electrical charge, conservation of energy, and so on. But also in disciplines that are not so much physically oriented conservation principles are used. For instance, in describing the evolution of a population, it can be assumed that there is conservation of individuals, simply because no individuals can be created or lost without reason. Similarly in economy, there always has to be conservation of assets in one sense or the other.

Hence, conservation laws can be seen as laws based on reasoning and on counting.

2.2 Phenomenological principles

In addition to the conservation laws discussed above, often also so-called phenomenological laws are used. These laws are obtained in an empirical way and depend very much on the nature of the phenomenon that has to be modelled.

One example of such a law is Ohm's law $V = RI$ relating the voltage V over a resistor of value R with the current I that goes through the resistor. Ohm's law is of importance in modelling electrical networks. However, laws with a similar form occur in other disciplines like Fourier's law on heat conductivity and Fick's law on light diffusion. It is not by reasoning that laws like Ohm's law are derived; they are simply the result of experiments. There is no reason why the voltage, the current and the resistance should be related as they do in Ohm's law. Nevertheless, it turns out to be part of the physical reality and therefore it can be used in the modelling of dynamic phenomena. Many more phenomenological laws exist, some of which are discussed in the next section.

2.3 Physical principles and laws

In this section we briefly discuss some of the most important laws and principles that hold in (parts of) the physical reality.

2.3.1 Thermodynamics

When modelling a thermodynamical phenomenon we can make use of three very fundamental laws and principles.

1. *Conservation of energy.*
2. *The irreversibility of the behavior of a macroscopic system.*
3. *The absolute zero temperature cannot be reached.*

The second law is often also expressed by saying that the entropy of a system cannot decrease. The entropy is a measure for the disorder in a system.

We note that the first law is based on reasoning. If the law were not satisfied, then some form of energy would be missing, and the law could be made to hold by simply introducing the missing type of energy. The second and third law are based on experiments and describe phenomenological properties.

2.3.2 Mechanics

When modelling mechanical phenomena we often, without realizing this, use some very important laws and principles. One of these principles, the conservation of energy, has already been discussed. Other forms of the conservation principle are also often used. Furthermore, the following three laws (postulates) of Newton are very useful.

1. *If there is no force acting on a point mass, then this mass will stay at rest, or it will move with a constant speed along a straight line.*
2. *The force F on a point mass m and its position s are related by* $F = m\dfrac{d^2s}{dt^2}$.
3. *action = − reaction.*

The first law was already known to Galileo, as the result of experiments that he had carried out. The second law was formulated by Newton, using the differential calculus he had developed.

Newton's laws, especially the first one, are inspired by experiments. Originally, the laws were developed for point masses and rectilinear movements. Gradually, versions of his laws were developed for continuous media, rotational motions, in fluids, in gasses, and so on. For instance, if a torque N with respect to some axis is applied to a body, and the moment of inertia of the body around the axis is J, then $N = J\dfrac{d^2\varphi}{dt^2}$, where $\dfrac{d^2\varphi}{dt^2}$ denotes the angular acceleration of the body around the used axis.

After Newton's laws were available, also other approaches to describe the general motion of mechanical structures were developed. One of these approaches, using the concepts of kinetic and potential energy, leads to equations of motion which are known as the Euler-Lagrange equations.

2.3.3 Electromagnetism

When modelling electromagnetic phenomena, versions of laws that are expressed by the four Maxwell equations can be used, complemented by the Lorentz equation.

In a medium with dielectric constant ε and magnetic susceptibility μ, the Maxwell equations relating an electric field E, a magnetic field B, a charge density ρ and a current density ι are the following

$$\operatorname{div} E = \frac{1}{\varepsilon}\rho, \quad \operatorname{rot} E = -\frac{\partial B}{\partial t}, \quad \operatorname{div} B = 0, \quad \operatorname{rot} B = \mu\left(\iota + \varepsilon\frac{\partial E}{\partial t}\right).$$

In these equations all variables depend on time t and, in general, position (x,y,z). Furthermore, E, B and ι are vectorial quantities, whereas ρ is a scalar. The words 'div' and 'rot' stand for divergence and rotation, respectively. The first and third equation in the above Maxwell equations express in a sense the conservation of electrical charge and 'magnetic charge', respectively. In fact, div $B = 0$ can be related to the fact that there do not exist magnetic monopoles (isolated charges).

The force F on a particle with charge q moving with velocity v in a medium as described above, with an electric field E and a magnetic field B, is given by the Lorentz equation

$$F = q\,(E + v \times B)\,.$$

Here $\times$ denotes the cross product. Both F and v are vectors, and q is a scalar. All three quantities will depend on time t and position (x,y,z).

The above equations are very general in nature and are often too general for our purposes. Therefore, other (more simplified) laws have been obtained from these equations. Some of these laws for electrical networks are discussed below. These networks are built, amongst others, from basic elements like resistors, capacitors and coils. For these elements the following relations have been established.

1. If a current of strength I is led through a resistor with value R, then the voltage drop V over the resistor can be computed by Ohm's law as illustrated in Figure 2.1.

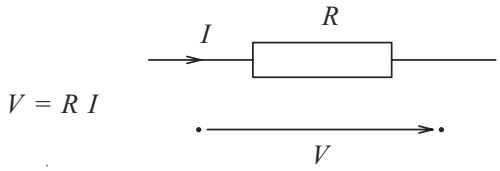

Figure 2.1 Ohm's law.

2. If a current of strength I flows into a capacitor with capacity C, the voltage drop V over the capacitor is related to I and C in the way shown in Figure 2.2.

3. Finally, if a current of strength I goes through a coil with inductance L, the voltage drop V over the coil can be obtained as depicted in Figure 2.3.

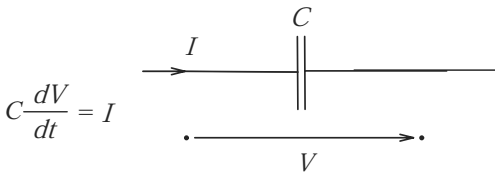

Figure 2.2 Law for capacitor.

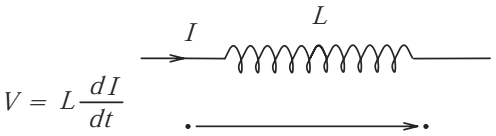

Figure 2.3 Law for inductor.

The variables V and I in Figures 2.1, 2.2 and 2.3 are functions of time. The values R, C and L are assumed to be time independent.

The above laws (rules) are phenomenological in nature. They are the results of experiments. In addition to these laws, two other laws (rules) play an important role in the area of electrical networks. These laws are called the 'laws of Kirchhoff', and can be formulated as follows.

4. *In any node of the network the sum of all the currents is zero.*

5. *In any loop of the network the sum of all the voltage drops is zero.*

In both laws the direction of currents and voltage drops have to be taken into account. Note that the Kirchhoff laws are of the conservation type. To explain these two laws we consider the abstract network in Figure 2.4, with a source over which the voltage drop is a constant equal to V. An arrow in the figure with an index i stands for an element through which a current I_i floats that induces a voltage drop V_i, both in the direction of the arrow.

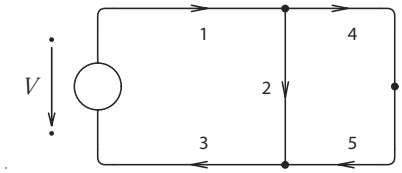

Figure 2.4 Electrical network.

Then in the four nodes (also the source is considered to be a node) the following holds

$$-I_1+I_2+I_4=0, \quad -I_2-I_5+I_3=0, \quad -I_4+I_5=0, \quad I_1-I_3=0.$$

For the three loops in the network it follows that

$$-V+V_1+V_2+V_3=0, \quad -V+V_1+V_4+V_5+V_3=0, \quad -V_2+V_4+V_5=0.$$

2.4 Examples

In this section we give some examples of systems. The models underlying the examples can be derived using the physical principles and laws discussed in the previous.

2.4.1 Inverted pendulum

Consider the inverted pendulum in Figure 2.5. The pivot of the pendulum is mounted on a carriage which can move in the horizontal direction. The carriage is driven by a small motor that at time t exerts a force $u(t)$ on the carriage. This force is the input variable to the system. The mass of the carriage will be indicated by M, that of the pendulum by m.

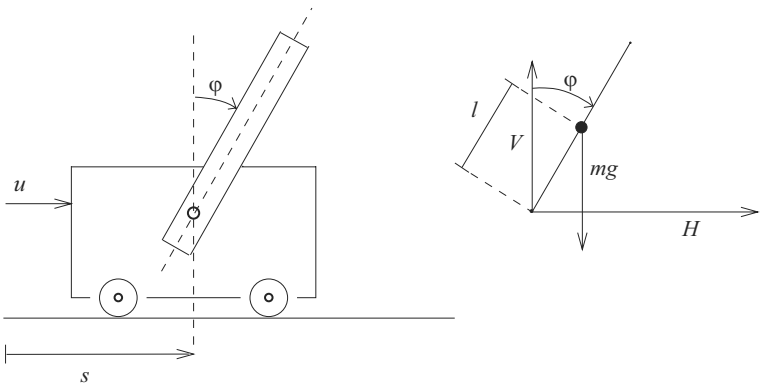

Figure 2.5 Inverted pendulum.

In the pendulum the distance between the pivot and the center of gravity is l. In Figure 2.5 the variable H denotes the horizontal reaction force and V is the vertical reaction force in the pivot. The angle that the pendulum makes with the vertical is indicated by φ. For the center of gravity of the pendulum we have the following equations, which are in the spirit of Newton's second law.

$$m\frac{d^2}{dt^2}(s+l\sin\varphi)=H, \quad m\frac{d^2}{dt^2}(l\cos\varphi)=V-mg, \tag{2.1}$$

$$J\frac{d^2\varphi}{dt^2}=Vl\sin\varphi-Hl\cos\varphi. \tag{2.2}$$

The function s denotes the position of the carriage and J is the moment of inertia of the pendulum with respect to the center of gravity. Clearly, φ, s, H and V depend on time t, whereas m, l, g and J are constant. The pendulum has length $2l$ and if it has a uniform mass distribution of $\frac{m}{2l}$ per unit of length, then the moment of inertia around its center of gravity is given by

$$J=\frac{m}{2l}\int_{-l}^{l}\sigma^2 d\sigma=\frac{1}{3}ml^2.$$

The equation which describes the motion of the carriage is

$$M\frac{d^2s}{dt^2}=u-H, \tag{2.3}$$

where u may depend on t, while M is constant. Elimination of H and V in the above equations leads to

$$\begin{aligned} &\tfrac{4l}{3}\ddot{\varphi} - g\sin\varphi + \ddot{s}\cos\varphi = 0, \\ &(M+m)\ddot{s} + ml(\ddot{\varphi}\cos\varphi - \dot{\varphi}^2\sin\varphi) = u, \end{aligned} \tag{2.4}$$

where $\dot{}$ denotes the first derivative with respect to time, and $\ddot{}$ the second derivative. So, $\dot{s} = \dfrac{ds}{dt}$ and $\ddot{\varphi} = \dfrac{d^2\varphi}{dt^2}$.

The above two equations can also be written as a set of four first order differential equations in $\varphi, \dot{\varphi}, s$ and $\dot{s}$.

In order to distinguish the above type differential equations from partial differential equations, to be introduced shortly, one refers to the above type of differential equations also as ordinary differential equations.

The equations of motion of the inverted pendulum can also be obtained as the Euler-Lagrange equations using the following expressions for the total kinetic energy T and the potential energy V

$$T = \frac{1}{2}M\dot{s}^2 + \frac{1}{2}\frac{m}{2l}\int_0^{2l}\left((\dot{s}+\sigma\dot{\varphi}\cos\varphi)^2 + (\sigma\dot{\varphi}\sin\varphi)^2\right)d\sigma,$$

$$V = \frac{m}{2l}g\int_0^{2l}\sigma\cos\varphi\, d\sigma = mgl\cos\varphi,$$

where T, in addition to the kinetic energy of the carriage, consists of the kinetic energy of all the infinitesimal parts $d\sigma$ of the pendulum at a distance σ from the pivot, $0 \le \sigma \le 2l$. A similar remark holds with respect to the potential energy.

With the Lagragian L, defined as $L = T - V$, it follows after evaluation of the integrals that

$$L = \frac{1}{2}M\dot{s}^2 + \frac{1}{2}m\dot{s}^2 + ml\dot{s}\dot{\varphi}\cos\varphi + \frac{2}{3}ml^2\dot{\varphi}^2 - mgl\cos\varphi. \tag{2.5}$$

The Euler-Lagrange equations describing the motion of the inverted pendulum can now be obtained by working out the next equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\varphi}}\right) - \frac{\partial L}{\partial\varphi} = 0, \quad \frac{d}{dt}\left(\frac{\partial L}{\partial\dot{s}}\right) - \frac{\partial L}{\partial s} = u.$$

In these equations the variable L is considered to depend on $\varphi, \dot{\varphi}, s$ and $\dot{s}$, whereas the latter variables depend on t. For instance, with T and V as above, this means that

$$\frac{\partial L}{\partial\dot{\varphi}} = ml\dot{s}\cos\varphi + \frac{4}{3}ml^2\dot{\varphi}, \quad \frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\varphi}}\right) = ml\ddot{s}\cos\varphi - ml\dot{s}\dot{\varphi}\sin\varphi + \frac{4}{3}ml^2\ddot{\varphi},$$

and similarly for the other (partial) derivatives.

2.4.2 Model of a satellite

Consider the motion of a satellite with mass m_s in a plane through the center of earth. See also the picture in Figure 2.6. As the satellite will orbit around the earth, it is natural

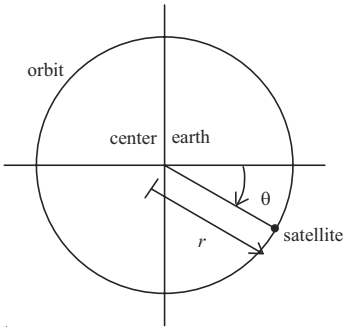

Figure 2.6 Satellite.

to give its position and velocity in terms of the polar coordinates r, θ, and their time derivatives $\dot{r}, \dot{\theta}$, with the earth's center located at the origin. Clearly, $r, \theta, \dot{r}$ and $\dot{\theta}$ depend on time t.

The velocity of the satellite has a radial component given by $\dot{r}$, and a tangential component equal to $r\dot{\theta}$. To apply Newton's laws, also the radial and tangential components of the acceleration of the satellite are required. The radial component of the acceleration is given by $\ddot{r} - r\dot{\theta}^2$, and the tangential component equals $2\dot{r}\dot{\theta} + r\ddot{\theta}$. The previous expressions for the radial and tangential components of the velocity and acceleration are elementary and can be found in any textbook on mechanics.

When in orbit the satellite is attracted by the earth by the gravitational force. This force has a radial direction and its magnitude equals $G\dfrac{m_e m_s}{r^2}$, where m_e denotes the mass of the earth and G stands for the gravitational constant. Assume that, in addition to gravity, the satellite is also subjected to a radially directed force F_r, and a tangentially directed force F_θ. The force F_r is assumed to be directed away from the earth. Both F_r and F_θ are thought to be caused by thrust jets mounted on the satellite.

Application of Newton's second law in the radial direction and the tangential direction results in

$$m_s\left(\ddot{r} - r\dot{\theta}^2\right) = -G\frac{m_e m_s}{r^2} + F_r, \quad m_s\left(2\dot{r}\dot{\theta} + r\ddot{\theta}\right) = F_\theta. \tag{2.6}$$

Remark 2.1 The above equations also can be obtained from the Euler-Lagrange equations. For that purpose, note that the kinetic energy T and the potential energy V of the satellite are given as follows

$$T = \frac{1}{2}m_s\left(\dot{r}^2 + (r\dot{\theta})^2\right), \quad V = -G\frac{m_e m_s}{r}.$$

Now define the Lagrangian as $L = T - V$, then the equations in (2.6) follow by working out the next equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}}\right) - \frac{\partial L}{\partial r} = F_r, \quad \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = rF_\theta,$$

where rF_θ must be interpreted as a torque due to the tangential force F_θ. □

2.4.3 Heated bar

Consider a metal bar of length L which is insulated from its environment, except at the left side where the bar is heated by a jet with heat transfer u at time t. For a picture, see Figure 2.7. The temperature of the bar at time t and position r, with $0 \leq r \leq L$, is denoted

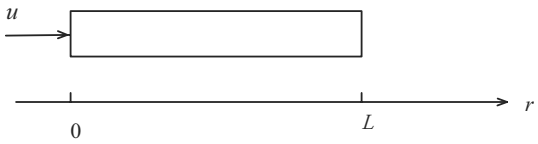

Figure 2.7 Heated bar.

by $T(t,r)$, i.e., r is the spatial variable. In order to be able to determine the thermal behavior of the bar one must know $T(t_0,r)$, $0 \leq r \leq L$, the initial temperature distribution at time $t = t_0$, and the heat transfer $u(t)$ for $t \geq t_0$. The state of the system is the function $T(t,\cdot) : [0,L] \to \mathbb{R}$. From physics it is known that T satisfies a partial differential equation

$$\frac{\partial T(t,r)}{\partial t} = c\frac{\partial^2 T(t,r)}{\partial r^2}, \tag{2.7}$$

where c is a characteristic constant of the bar. At the left side of the bar we have

$$-A\frac{\partial T(t,r)}{\partial r}\bigg|_{r=0} = u(t), \tag{2.8}$$

where A is a measure for the area of the cross section of the bar. At the right side we have

$$\frac{\partial T(t,r)}{\partial r}\bigg|_{r=L} = 0, \tag{2.9}$$

because of the insulation there. The evolution of the state is described by the partial differential equation (2.7) with boundary conditions (2.8) and (2.9). In this example the input enters the problem only via the boundary conditions. In other problems the input can also be distributed; see Exercise 2.5.10.

2.4.4 Electrical circuit

Consider the electrical network depicted in Figure 2.8, consisting of a resistor R, a capacitor C and a coil L. The network is connected to a source with constant voltage drop V and the voltage drop over the capacity is measured. The current is denoted by I. If V_R, V_C and V_L denote the voltage drops over the resistor, the capacitor and the coil, respectively, then it follows from the laws of electricity mentioned in the previous section, that

$$V_R = RI, \quad V_C = \frac{1}{C}Q, \quad V_L = L\frac{dI}{dt},$$

where the variable Q denotes the electrical charge on the capacitor, which satisfies $I = \frac{dQ}{dt}$. According to the Kirchhoff laws, $V = V_R + V_C + V_L$. Hence,

$$V = RI + \frac{1}{C}Q + L\frac{dI}{dt}, \quad I = \frac{dQ}{dt}.$$

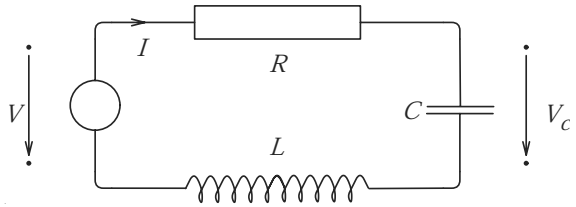

Figure 2.8 RLC network.

Now rearranging these equations, it follows that

$$\frac{d}{dt}\begin{pmatrix} Q \\ I \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{1}{LC} & -\frac{R}{L} \end{pmatrix}\begin{pmatrix} Q \\ I \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{L} \end{pmatrix} V,\ V_C = \begin{pmatrix} \frac{1}{C} & 0 \end{pmatrix}\begin{pmatrix} Q \\ I \end{pmatrix}.$$

Define $u = V$, $y = V_C$ and

$$x = \begin{pmatrix} Q \\ I \end{pmatrix},\ A = \begin{pmatrix} 0 & 1 \\ -\frac{1}{LC} & -\frac{R}{L} \end{pmatrix},\ B = \begin{pmatrix} 0 \\ \frac{1}{L} \end{pmatrix},\ C = \begin{pmatrix} \frac{1}{C} & 0 \end{pmatrix},$$

where it must be emphasized that the newly defined C is a matrix (more specifically, here a row matrix with two elements). It should not be confused with the capacity C. This is an instance of the same symbol being used for different quantities.

With the above way of writing, the following description of the system is obtained

$$\dot{x} = Ax + Bu, \quad y = Cx.$$

Remark 2.2 Elimination of I from the equations above yields the following ordinary linear differential equation with constant coefficients

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q = V.$$

This type of equation not only occurs in the modelling of electrical networks. Also in other disciplines this type of equation may arise. For instance, when modelling a mechanical structure as depicted in Figure 2.9. The structure consists of a mass M connected

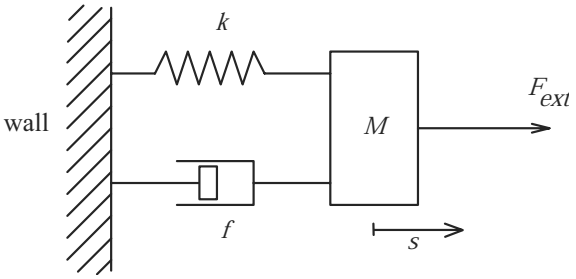

Figure 2.9 Mass-damper-spring system.

to a vertical wall by means of a spring with constant k and a damper with damping factor

f. On the mass an external force F_{ext} may be exerted. As the mass is moving horizontally only, gravity does not play a role. If s denotes the displacement of the mass from its equilibrium position, it follows from Newton's second law that $M\ddot{s} = -ks - f\dot{s} + F_{\text{ext}}$. Hence,

$$M\ddot{s} + f\dot{s} + ks = F_{\text{ext}}.$$

This equation is similar to the one derived for the electrical network above. Other examples of equations of this type can be found in modelling phenomena in disciplines like acoustics, chemistry and hydraulics. □

2.4.5 Population dynamics

Consider a closed population of humans in a country, or animals or organisms in nature. Let $N(t)$ denote the number of individuals in the population at time t. Assume that $N(t)$ is so large that it can be thought of as being a continuously changing variable. If $B(t,t+\delta)$ and $D(t,t+\delta)$ denote the number of births and deaths, respectively, in the interval $(t,t+\delta]$, then conservation of individuals means that

$$N(t+\delta) - N(t) = B(t,t+\delta) - D(t,t+\delta).$$

Let

$$B(t,t+\delta) = b(t)\delta + o(\delta), \quad D(t,t+\delta) = d(t)\delta + o(\delta),$$

where $o(\delta)$ stands for a function that tends to zero faster than δ. The functions $b(t)$ and $d(t)$ are called the birth rate and death rate, respectively. Moreover, assume that $b(t)$ and $d(t)$ depend on $N(t)$ in a proportional way, independent of time. Hence,

$$b(t) = bN(t), \quad d(t) = dN(t),$$

for some constants b and d. This means that

$$N(t+\delta) - N(t) = (b-d)N(t)\delta + o(\delta).$$

Define $r = b - d$, divide by δ and take the limit for δ to zero. Then it follows that

$$\dot{N}(t) = rN(t).$$

This equation has as solution $N(t) = N(t_0)e^{r(t-t_0)}$. Hence, the number of individuals is increasing (decreasing) when $r > 0$ ($r < 0$).

In general, the growth rate of a population depends on more factors than the above mentioned birth and death rates alone. In particular, it often depends on how the internal interaction is. For instance, if a country is densely populated, then the death rate may increase due to the effects of competition for space and resources, or due to the high susceptibility for deceases. Assuming that the population cannot consist of more than $K > 0$ individuals, the above model might be modified as

$$\dot{N} = r\left(1 - \frac{N}{K}\right)N,$$

where in the equation the dependency of N on t is omitted. The equation is also known as the 'logistic equation'.

The model can further be modified in the following way. Assume that the species of the above population are the prey for a second population of predators consisting of $M(t)$ individuals at time t. It is then reasonable to assume that $r > 0$, and that the previous equation has to be changed into

$$\dot{N} = r\left(1 - \frac{N}{K}\right)N - \alpha NM,$$

with $\alpha > 0$. The modification means that the rate of decrease of prey due to the presence of predators is proportional to the number of predators, but also to the number of prey itself. As a model for the predators the following can be used

$$\dot{M} = -cM + \beta NM,$$

with $c > 0$ and $\beta > 0$. Together these two equations form a so-called 'predator-prey model'. Note that $r > 0$ means that the population of the prey has a natural tendency to increase, whereas because of $c > 0$ the population of predators has a natural tendency to decrease.

Now assume that the number of prey is unbounded ($K = \infty$). Think of anchovy as prey and of salmon as predator. Assume that due to fishing at time t a fraction $u_1(t)$ of the anchovy is caught, and a fraction $u_2(t)$ of the salmon. The previously derived predator-prey model then has to be changed as follows

$$\dot{N} = rN - \alpha NM - Nu_1 = (r - \alpha M - u_1)N,$$

$$\dot{M} = \beta NM - cM - Mu_2 = (\beta N - c - u_2)M.$$

This type of model is well-known, and is also called a **Volterra-Lotka model**. If the number of salmon is monitored in some way and is denoted $y(t)$, then the above model can be described as a system

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

with $x = (x_1,x_2)^\top = (N,M)^\top, u = (u_1,u_2)^\top$ and $y = M$, and functions

$$f(x,u) = \begin{pmatrix} (r - \alpha x_2 - u_1)x_1 \\ (\beta x_1 - c - u_2)x_2 \end{pmatrix}, \quad h(x,u) = x_2.$$

2.4.6 Age dependent population dynamics

Consider again a population and let its size be denoted by N. To express N as a function of the birth rate b, let $P(t,r)$ be the probability that somebody, born at time $t - r$, is still alive at time t (at which he/she has an age of r). Then

$$N(t) = \int_{-\infty}^{t} P(t,t-s)b(s)ds,$$

where s represents the time of birth. Assume that the functions P and b are such that this integral is well defined. It is reasonable to assume that $P(t,r) = 0$ for $r > L$ for some L

(nobody will become older than L). Then

$$N(t) = \int_{t-L}^{t} P(t, t-s)b(s)ds.$$

If P is continuous in its arguments and if b is piecewise continuous (a description of piecewise continuity is given later), then the above integral exists.

Returning to the original integral and assuming that a function g exists such that $g(r) = P(t,r)$, it follows that

$$N(t) = \int_{-\infty}^{t} g(t-s)b(s)ds.$$

If this integral exists for all admissible birth rates b, then it will be shown later that it can be associated with a time-invariant, strictly causal input/output system. (The notions of time-invariance and (strict) causality will be made precise later (in Sections 3.2 and 3.4). Heuristically, time-invariance means that the absolute (calendar) time does not play any role and causality means that the future does not influence the current behavior.) For such a system the probability that somebody is still alive at age r is determined by r only and not by the time of birth.

2.4.7 Bioreactor

Consider a bioreactor as depicted in Figure 2.10. In the reactor there is biomass (or-

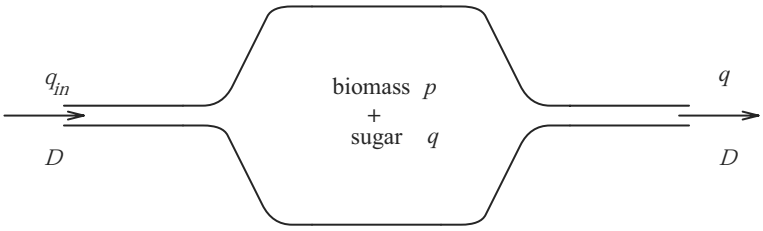

Figure 2.10 Bioreactor.

ganisms) that is nourished with sugar (nutrition). Further, extra nutrition is supplied and products are withdrawn. At time t denote

$p(t)$ *for the concentration of biomass in the reactor (g/l),*

$q(t)$ *for the concentration of sugar in the reactor (g/l),*

$q_{\text{in}}(t)$ *for the concentration of sugar in the flow into the reactor (g/l),*

$D(t)$ *for the flow of 'sugar water' through the reactor (1/sec), i.e., the fraction of its contents that flows through the reactor per second.*

The equations that govern the reaction inside the reactor are given as follows

$$\frac{d}{dt}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \text{natural growth } -Dp \\ -\text{ natural consumption } -Dq + Dq_{\text{in}} \end{pmatrix}.$$

Note that Dp and Dq are products (in mathematical sense). They stand for the amounts the biomass and sugar, respectively, that are withdrawn from the reactor. The product Dq_{in} stands for the amount of sugar that is supplied to the reactor. To complete the mathematical description some empirical laws (or rules of thumb) on the relation between biomass and sugar concentration will be used. Here these laws state that the growth of biomass is proportional to its concentration and that its consumption of sugar is also proportional to its concentration. Furthermore, it is assumed that these proportionalities only depend on the sugar concentration. Hence, there are functions μ and ν, depending on the sugar concentration, that determine the rate of growth of biomass and the consumption rate of sugar, respectively, it the following way

$$\frac{d}{dt}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \mu(q)p - Dp \\ -\nu(q)p - Dq + Dq_{\text{in}} \end{pmatrix}.$$

2.4.8 Transport of pollution

Consider a 'one dimensional' river, contaminated by organic material that is dissolved in the water, see Figure 2.11. Once in the water, the material is degraded by the action of bacteria. Denote

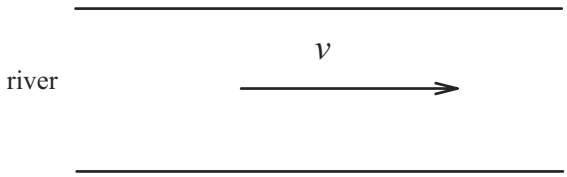

Figure 2.11 Pollution in river.

$\rho(r,t)$ *for the density of pollutant in the river at place r and at time t (kg/m),*

$v(r,t)$ *for the speed of pollutant and water in the river at place r and at time t (m/sec),*

$q(r,t)$ *for the flux of pollutant in the river at place r and at time t (kg/sec),*

$k(r,t)$ *for the rate of change by which the density of the pollutant is increased in the river at place r and at time t (kg/(m sec)).*

Conservation of mass can be expressed as

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial r} = k,$$

which has been obtained by considering the infinitesimal equality

$$\rho(r,t+dt)dr = \rho(r,t)dr + q(r,t)dt - q(r+dr,t)dt + k(r,t)dtdr.$$

Now two extreme cases can be considered.

1. There is only advection. Then ρ, q and v are related by $q = \rho v$. This means that the flux of pollutant is only due to transportation phenomena.

2. There is only diffusion. Then ρ and q are related by $q = -\mu \dfrac{\partial \rho}{\partial r}$, where μ is some constant depending on the place r and the time t. Diffusion means that everything is smoothed.

When both diffusion and advection are taken into account then $q = \rho v - \mu \dfrac{\partial \rho}{\partial r}$. Assuming that μ is a constant, independent of r and t, and that v does not depend on r, but only on t, the conservation of mass equation can be written as

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial r}\left(\rho v - \mu \frac{\partial \rho}{\partial r}\right) + k = \mu \frac{\partial^2 \rho}{\partial r^2} - v\frac{\partial \rho}{\partial r} + k.$$

To model the action of bacteria that degrade the pollution, and to model the role of industry, assume that $k = -\nu\rho + \beta$ with ν independent of r and t, and with β a measure for the pollution in the river caused by the industry. Then it follows that

$$\frac{\partial \rho}{\partial t} = \mu \frac{\partial^2 \rho}{\partial r^2} - v\frac{\partial \rho}{\partial r} - \nu\rho + \beta.$$

Remark 2.3 With μ, v and ν constant the last equation can also formally be written as

$$\dot{x} = Ax + \beta,$$

where $x = \rho$ and $A = \mu \dfrac{\partial^2}{\partial r^2} - v\dfrac{\partial}{\partial r} - \nu$ is a linear mapping between appropriate function spaces. □

2.4.9 National economy

Consider the following simplified model of the national economy of a country. Let

$y(k)$ *be the total national income in year k,*

$c(k)$ *be the consumer expenditure in year k,*

$i(k)$ *be the investments in year k,*

$u(k)$ *be the government expenditure in year k.*

For the model of the national economy the following assumptions are made.

1. $y(k) = c(k) + i(k) + u(k)$,

2. *The consumer expenditure is a fixed fraction of the total income of the previous year:* $c(k) = my(k-1)$ *with* $0 \le m < 1$,

3. *The investments in year k depend on the increase in consumer expenditure from year* $k-1$ *to year k:* $i(k) = \mu(c(k) - c(k-1))$, *where* μ *is some positive constant.*

Note the first assumption is of the conservation type, whereas the other two assumptions may be based on observations.

With the above assumptions the evolution of the national economy can be described as follows:

$$\begin{aligned} &i(k+1)-\mu c(k+1) = -\mu c(k),\\ &c(k+1) = my(k) = m(i(k)-\mu c(k))+m(1+\mu)c(k)+mu(k). \end{aligned}$$

If a state vector is defined as $x(k)=(x_1(k),x_2(k))^\top$, with $x_1(k)=i(k)-\mu c(k)$ and $x_2(k)=c(k)$, then the state evolution equation is given by

$$\begin{pmatrix} x_1(k+1)\\ x_2(k+1)\end{pmatrix} = \begin{pmatrix} 0 & -\mu\\ m & m(1+\mu)\end{pmatrix}\begin{pmatrix} x_1(k)\\ x_2(k)\end{pmatrix}+\begin{pmatrix} 0\\ m\end{pmatrix}u(k),$$

and the output equation by

$$y(k) \;=\; \begin{pmatrix} 1 & 1+\mu\end{pmatrix}\begin{pmatrix} x_1(k)\\ x_2(k)\end{pmatrix} + u(k).$$

Thus, a linear time-invariant discrete-time system has been obtained as a model for the national economy.

2.5 Exercises

Exercise 2.5.1 *Consider the inverted pendulum in Section 2.4.1. Assume that the angle φ of the pendulum with the vertical is measured. Let this measurement be denoted by the variable y. So, $y=\varphi$. Note that y as well as all the other variables $\varphi,\dot\varphi,s,\dot s$ and u are functions of time. Consider the vector $x=(\varphi,\dot\varphi,s,\dot s)^\top$, and find functions $f(x,u)$ and $h(x,u)$ such that the inverted pendulum can be described as*

$$\dot x = f(x,u), \quad y = h(x,u).$$

Here $\dot x = \dfrac{d}{dt}x = (\dot\varphi,\ddot\varphi,\dot s,\ddot s)^\top$.

Exercise 2.5.2 *Take the variable L as in (2.5) and derive the equations of motion of the inverted pendulum in Section 2.4.1 by working out the given Euler-Lagrange equations.*

Exercise 2.5.3 *In the above exercise, the pendulum is assumed to be able to rotate around its end point. Now assume that the pendulum can rotate around a given point somewhere on its longitudinal axis, not necessarily the end point. Derive the equations of motion of this modified inverted pendulum. Start with the (direct) approach of Section 2.4.1 and verify your results with the approach using the Euler-Lagrange equations.*

Exercise 2.5.4 *In the inverted pendulum example of Section 2.4.1 the input is a force exerted on the carriage. Now assume that the input is a torque exerted on the pendulum around its pivot. Determine how the equations change with respect to those in Section 2.4.1.*

Exercise 2.5.5 *In Section 2.4.1 the carriage moves horizontally. Now assume that the carriage moves only in the vertical direction and that only vertical forces can be exerted, while the gravity remains to act vertically. Investigate how the equations change with respect to those in Section 2.4.1.*

Exercise 2.5.6 *Consider the model of the satellite in Section 2.4.2. Assume that the distance r is measured and is denoted y. Further, introduce the vectors* $x = (r, \theta, \dot{r}, \dot{\theta})^\top$ *and* $u = (\frac{F_r}{m_s}, \frac{F_\theta}{m_s})^\top$, *and find functions* $f(x,u)$ *and* $h(x,u)$ *such that the model of the satellite can be described as*

$$\dot{x} = f(x,u), \quad y = h(x,u).$$

Exercise 2.5.7 *In Section 2.4.2, starting from the Lagrangian* $L = T - V$, *work out the given Euler-Lagrange equations to obtain the equations of the motion of the satellite.*

Exercise 2.5.8 *Consider the electrical network depicted in Figure 2.12. Take* V_{in} *as input and* V_{out} *as output, and derive a state space model for the network using the laws introduced in the Section 2.3.3. Note that* V_{out} *can be seen as a voltage drop in the 'loop' containing just the two resistors. Clearly, there are more such loops containing* V_{out}.

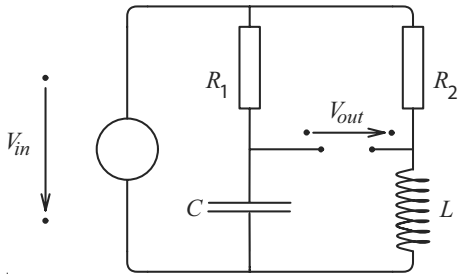

Figure 2.12 Bridge network.

Exercise 2.5.9 *Consider the electrical network in Figure 2.13. Take the source voltage as input u, the voltage over the most right capacitor as output y and derive a state space model for the network using the methods of the Section 2.3.3.*

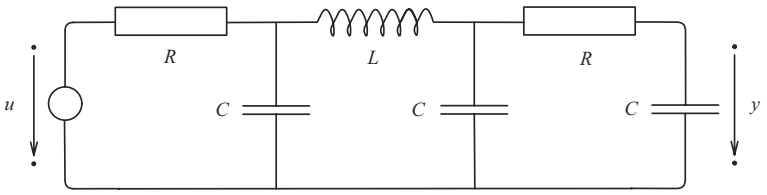

Figure 2.13 Electrical network.

Exercise 2.5.10 *In the context of Section 2.4.3, consider the partial differential equation*

$$\frac{\partial T(t,r)}{\partial t} = c\frac{\partial^2 T(t,r)}{\partial r^2} + u(t,r),$$

and give an interpretation of $u(t,r)$, *seen as a distributed input/control function.*

Exercise 2.5.11 *For each of the models in Section 2.4.5, find the stationary situations. These are situations in which the variables remain at a constant level and therefore have (time) derivatives that are identically equal to zero.*

Exercise 2.5.12 *Let p denote the population density, and let it depend on time t and age r. The number of people of ages between r and $r+dr$ at a certain time t is given by $p(t,r)dr$. Define the mortality rate $\mu(t,r)$ in the following way: $\mu(t,r)p(t,r)drdt$ is the number of people in the age class $[r,r+dr]$ who die in the time interval $[t,t+dt]$. Prove the infinitesimal equality*

$$p(t+dt,r+dt)dr - p(t,r)dr = -\mu(t,r)\,p(t,r)\,drdt,$$

and show that p satisfies the following partial differential equation

$$\frac{\partial p}{\partial r} + \frac{\partial p}{\partial t} = -\mu p. \tag{2.10}$$

Let the initial age distribution be given as

$$p(0,r) = p_0(r),\ 0 \le r \le 1,$$

and the birth rate function as the boundary condition

$$p(t,0) = u(t),\ t \ge 0.$$

Here it assumed that the age r is scaled in such a way that nobody reaches an age $r > 1$. One can consider $u(t)$ as the input to the system and as output $y(t)$ for instance the number of people in the working age, say between the ages a and b, $0 < a < b < 1$. This means that

$$y(t) = \int_a^b p(t,r)dr.$$

Exercise 2.5.13 *In Section 2.4.7, assume that the flow D of 'sugar water' into the reactor is fixed, but that the sugar concentration q_{in} in this flow can be controlled. Further, assume that the concentration of sugar of the outgoing flow is measured. Now describe the above process as a system with state, input and output.*

Exercise 2.5.14 *The same as the above question, but now the sugar concentration q_{in} in the incoming flow is fixed, and the amount of flow D can be controlled.*

Exercise 2.5.15 *In Section 2.4.9, suppose that the government decides to stop its expenditure from the year $k = 0$ on. Hence, $u(k) = 0$ for all $k \ge 0$. Furthermore, suppose that in the year $k = 0$ the consumers do not spend any money and that the investments are 1 (scaled). So, $c(0) = 0$, $i(0) = 1$. Investigate how the total national income changes for $k \ge 0$.*

Exercise 2.5.16 *For the same model of the economy as in the above question, find the stationary situations when $u(k) = 1$ for all k, i.e., find those situations that will not change anymore as the years pass by, when $u(k) = 1$ for all k.*

Chapter 3

Linear Differential Systems

3.1 Input-State-Output Descriptions

In Chapter 2 we have presented some tools to obtain a dynamical system model. In general, such model can be formulated as a set of first-order differential and algebraic equations of the form

$$\dot{x}(t) = f\big(x(t), u(t), t\big), \tag{3.1}$$
$$y(t) = g\big(x(t), u(t), t\big), \tag{3.2}$$

where $x(t) \in \mathbb{R}^n$ represents the **state** (vector) of the system, $u(t) \in \mathbb{R}^m$ represents the **input** (vector), and $y(t) \in \mathbb{R}^p$ the **output** (vector). The functions $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^p$ denote the system vector field and output vector field, respectively.

As indicated in Chapter 1, the input $u(t)$ can be seen as the variable by which the state of the system can be influenced from the outside world, whereas the output $y(t)$ can be regarded as the variable by which information on the state of the system becomes available to the outside world. For that reason, the input $u(t)$ is often referred to as the control, while the output $y(t)$ is interpreted as the measurement. The set of equations (3.1)–(3.2) is referred to as an **input-state-output system**.

A particular case of (3.1)–(3.2) arises when the vector functions f and g do not explicitly depend on time, i.e.,

$$\dot{x}(t) = f\big(x(t), u(t)\big), \tag{3.3}$$
$$y(t) = g\big(x(t), u(t)\big). \tag{3.4}$$

Systems of the form (3.3)–(3.4) are said to be **time-invariant**, whereas systems of the form (3.1)–(3.2) are **time-variant** systems. For a precise definition of time-invariance, we refer to Chapter 8.

In general, the functions f and g are non-linear. For linear systems, the input-state-output system (3.1)–(3.2) takes the special form

$$\dot{x}(t) = A(t)x(t) + B(t)u(t),$$
$$y(t) = C(t)x(t) + D(t)u(t),$$

where $A(t)$ is a $n \times n$ matrix, referred to as the **system matrix**, $B(t)$ is a $n \times m$ matrix, referred to as the **input matrix**, the **output matrix** $C(t)$ has the dimensions $p \times n$, and $D(t)$ is the **feed-through matrix** having dimensions $p \times m$. This type of systems is referred to as **linear** and **time-variant (LTV)** systems. Hence, a **linear** and **time-invariant (LTI)** system takes the form

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{3.5}$$
$$y(t) = Cx(t) + Du(t), \tag{3.6}$$

where A, B, C, and D are matrices with constant coefficients.

Linear systems form an important class of systems. First of all, linear systems can be analyzed much easier than nonlinear systems. This is particularly true for linear and time-invariant systems of the form (3.5)–(3.6) since, as we show in Section 3.3, the solution can be expressed analytically in terms of the initial conditions and the input function. Based on this solution, many important properties of the system can be studied. The second reason for studying linear systems is that many practical systems can be accurately modeled as linear systems or can be approximated by a linear system using a **linearization** procedure as discussed in Section 3.2.

Remark 3.1 For ease of notation, we will often omit the time-dependency of the input, state, and output functions, and simply write u, x, and y, instead of $u(t)$, $x(t)$, and $y(t)$, respectively.

3.2 Linearization

The linearization process of a non-linear time-invariant system is as follows. The pair (x^*,u^*), with constant vectors $x^* \in \mathbb{R}^n$ and $u^* \in \mathbb{R}^m$, is said to be an **equilibrium pair** for (3.3) and (3.4), if

$$f(x^*,u^*) = 0. \tag{3.7}$$

Clearly, with (x^*,u^*) an equilibrium pair, the solution $\tilde{x}(t)$ of (3.3) for the initial condition $\tilde{x}(0) = x^*$ and input function $\tilde{u}(t) = u^*, \forall t \geq 0$, is given by $\tilde{x}(t) = x^*, \forall t \geq 0$. Let $\tilde{x}(t) + z(t) (= x^* + z(t))$ be another solution of (3.3) for the initial condition $\tilde{x}_0 + z_0 (= x^* + z_0)$ and input function $\tilde{u}(t) + v(t) (= u^* + v(t)), \forall t \geq 0$. Hence,

$$\frac{d}{dt}(\tilde{x}+z) \;=\; f(\tilde{x}+z,\tilde{u}+v), \qquad \tilde{x}(0)+z(0) \;=\; x^*+z_0. \tag{3.8}$$

It then follows that

$$\frac{d}{dt}z \;=\; f(x^*+z,u^*+v), \qquad z(0) \;=\; z_0. \tag{3.9}$$

We assume that f is sufficiently smooth (for instance, f has continuous partial derivatives up to order two) such that, according to the theorem of Taylor, the right-hand side $f(x^* + z, u^* + v)$ in Equation (3.9) can be expanded as

$$f(x^*+z,u^*+v) = f(x^*,u^*) + \frac{\partial f}{\partial x}(x^*,u^*)z + \frac{\partial f}{\partial u}(x^*,u^*)v + \text{ higher order terms.} \tag{3.10}$$

Note that this is a vectorial expression. Written out in components, the terms in the above are

$$f = \begin{pmatrix} f_1 \\ \vdots \\ \vdots \\ f_n \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ \vdots \\ \vdots \\ z_n \end{pmatrix}, \quad \frac{d}{dt}z = \begin{pmatrix} \frac{dz_1}{dt} \\ \vdots \\ \frac{dz_n}{dt} \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ \vdots \\ \vdots \\ v_m \end{pmatrix},$$

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}, \quad \frac{\partial f}{\partial u} = \begin{pmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{pmatrix}.$$

If $f(x^*+z,u^*+v)$ in Equation (3.9) is replaced by Equation (3.10), and if z and v are chosen to be 'small', so that the higher order terms can be ignored, we get the differential equation and initial condition

$$\dot{z} = \frac{\partial f}{\partial x}(x^*,u^*)z + \frac{\partial f}{\partial u}(x^*,u^*)v, \qquad z(0) \;=\; z_0. \tag{3.11}$$

which describes approximately the relation between z and v, seen as deviations from x^* and u^*, respectively. The coefficients of the matrices $\frac{\partial f}{\partial x}(x^*,u^*)$ and $\frac{\partial f}{\partial u}(x^*,u^*)$ are constant because the linearization is done around a fixed equilibrium pair. Hence, the differential equation (3.11) is linear and of the form

$$\dot{z} = Az + Bv, \tag{3.12}$$

with

$$A = \frac{\partial f}{\partial x}(x^*,u^*), \quad B = \frac{\partial f}{\partial u}(x^*,u^*).$$

The output function

$$y = g(x,u), \quad y \in \mathbb{R}^p,$$

can also be linearized around the pair (x^*,u^*), assuming that g is sufficiently smooth. If $\tilde{y} = g(\tilde{x},\tilde{u})$ and $\tilde{y}+w = g(\tilde{x}+z,\tilde{u}+v)$, then, with $y^* = g(x^*,u^*)$, it follows from the Taylor series expansion that

$$y^* + w = g(x^*,u^*) + \frac{\partial g(x^*,u^*)}{\partial x}z + \frac{\partial g(x^*,u^*)}{\partial u}v + \text{ higher order terms},$$

and, therefore, as an approximation, we get

$$w = \frac{\partial g(x^*,u^*)}{\partial x}z + \frac{\partial g(x^*,u^*)}{\partial u}v,$$

which we write as

$$w = Cz + Dv, \tag{3.13}$$

with

$$C = \frac{\partial g}{\partial x}(x^*,u^*), \quad D = \frac{\partial g}{\partial u}(x^*,u^*).$$

Equations (3.12) and (3.13) together form the linearized system, i.e., the system linearized around the equilibrium pair (x^*,u^*).

Remark 3.2 Note that, instead of linearizing around a fixed point, it is also possible to linearize around a given **solution pair** $(\tilde{x}(t),\tilde{u}(t))$, satisfying $\dot{\tilde{x}}(t) = f(\tilde{x}(t),\tilde{u}(t))$, and $\tilde{y}(t) = g(\tilde{x}(t),\tilde{u}(t))$. In that case, the linearized system becomes time-varying, i.e.,

$$\begin{aligned} \dot{z}(t) &= A(t)z(t) + B(t)v(t), \\ w(t) &= C(t)z(t) + D(t)v(t), \end{aligned}$$

with

$$A(t)=\frac{\partial f}{\partial x}\big(\tilde{x}(t),\tilde{u}(t)\big),\quad B(t)=\frac{\partial f}{\partial u}\big(\tilde{x}(t),\tilde{u}(t)\big),$$
$$C(t)=\frac{\partial g}{\partial x}\big(\tilde{x}(t),\tilde{u}(t)\big),\quad D(t)=\frac{\partial g}{\partial u}\big(\tilde{x}(t),\tilde{u}(t)\big).$$

Example 3.3 Consider the nonlinear system equations

$$\dot{x}=f(x,u),\quad y=g(x,u),\quad x\in\mathbb{R}^2,\ u,y\in\mathbb{R},$$

with

$$x=\begin{pmatrix}x_1\\x_2\end{pmatrix},\quad f(x,u)=\begin{pmatrix}x_2\\ux_1+x_2^2+x_1x_2+6\end{pmatrix},\quad g(x,u)=x_1^3-x_2+u^2.$$

Suppose $\tilde{u}(t)=u^*=-3,\forall t\geq 0$, then $\tilde{x}_1(t)=x_1^*=2$ and $\tilde{x}_2(t)=x_2^*=0,\forall t\geq 0$, is an equilibrium solution. Subsequently, the linearization around the equilibrium pair $((2,0)^\top,-3)$ is computed as follows. Note that

$$\frac{\partial f}{\partial x}(x,u)=\begin{pmatrix}0&1\\x_2+u&x_1+2x_2\end{pmatrix},\quad \frac{\partial f}{\partial u}(x,u)=\begin{pmatrix}0\\x_1\end{pmatrix},$$

so that

$$A=\frac{\partial f}{\partial x}(x^*,u^*)=\begin{pmatrix}0&1\\-3&2\end{pmatrix},\quad B=\frac{\partial f}{\partial u}(x^*,u^*)=\begin{pmatrix}0\\2\end{pmatrix}.$$

Likewise, it follows that

$$C=\frac{\partial g}{\partial x}(x^*,u^*)=\begin{pmatrix}12&-1\end{pmatrix},\quad D=\frac{\partial g}{\partial u}(x^*,u^*)=\begin{pmatrix}-6\end{pmatrix}.$$

Hence, the linearized system becomes

$$\dot{z}=\begin{pmatrix}0&1\\-3&2\end{pmatrix}z+\begin{pmatrix}0\\2\end{pmatrix}v,\quad w=\begin{pmatrix}12&-1\end{pmatrix}z-6v.$$

This system is time-invariant because all system matrices are independent of t. □

Remark 3.4 Note that the starting point in Example 3.3 was a first order description and an equilibrium pair (x^*,u^*) around which the linearization was done. In many cases the first order description is coming from a higher order description. Then, it is also possible to first linearize the higher order description to obtain a linear higher order approximation that subsequently can be transferred into a first order description. For instance, the system in Example 3.3 may have been obtained from

$$\ddot{\zeta}=u\zeta+\dot{\zeta}^2+\zeta\dot{\zeta}+6,\quad y=\zeta^3-\dot{\zeta}+u^2,$$

by introducing $x_1=\zeta$ and $x_2=\dot{\zeta}$. Note that $\tilde{\zeta}(t)=2$ and $\tilde{u}(t)=-3,\forall t\geq 0$, consequently with $\dot{\tilde{\zeta}}(t)=\ddot{\tilde{\zeta}}(t)=0,\forall t\geq 0$, satisfy the above differential equation, i.e.,

$$\ddot{\tilde{\zeta}}=\tilde{u}\tilde{\zeta}+\dot{\tilde{\zeta}}^2-\tilde{\zeta}\dot{\tilde{\zeta}}+6.$$

Assume that also $\tilde{\zeta}+s$ and $\tilde{u}+v$ satisfy the differential equation, where s, v, and all their derivatives are small. Hence,

$$\dddot{\tilde{\zeta}}+\ddot{s}=(\tilde{u}+v)(\tilde{\zeta}+s)+(\dot{\tilde{\zeta}}+\dot{s})^2+(\tilde{\zeta}+s)(\dot{\tilde{\zeta}}+\dot{s})+6.$$

Then, ignoring products and powers of s, v and their derivatives, it follows after some manipulation that

$$\ddot{s}=\tilde{u}s+\tilde{\zeta}v+\tilde{\zeta}\dot{s}=-3\,s+2\,\dot{s}+2\,v.$$

Introduction of $z_1=s$ and $z_2=\dot{s}$ yields

$$\dot{z}_1=z_2,\quad \dot{z}_2=-3\,z_1+2\,z_2+2\,v,$$

or

$$\dot{z}=Az+Bv,$$

with

$$z=\begin{pmatrix} z_1 \\ z_2 \end{pmatrix},\quad A=\begin{pmatrix} 0 & 1 \\ -3 & 2 \end{pmatrix},\quad B=\begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

The linearization of the output equation can be obtained similarly.

Remark 3.5 In the previous the distinction between variables like the state, input and output, and their deviations from the given solutions around which the linearization is done, is clearly indicated by the use of different symbols. Indeed, the state, input and output are denoted by x, u and y, respectively, whereas the deviations from the given solutions are denoted by z, v and w, respectively. However, in practice, this distinction is often not supported by the notation. Often, both the state and its deviation from a given solution are denoted by one and the same symbol, mostly x, and similarly for the input and output. In those cases it should be clear from the context which meaning should be attached to the variable x, being the true state or being the deviation from some given solution.

Example 3.6 [Continuation of the inverted pendulum.] We start with the equations of motion in Equations (2.4) of Section 2.4.1 which are repeated here.

$$\begin{aligned} &\tfrac{4l}{3}\ddot{\varphi}-g\sin\varphi+\ddot{s}\cos\varphi=0,\\ &(M+m)\ddot{s}+ml(\ddot{\varphi}\cos\varphi-\dot{\varphi}^2\sin\varphi)=u. \end{aligned} \tag{2.4}$$

This system can be written as a set of four first order differential equations where the state vector is defined as $x=(\varphi,\dot{\varphi},s,\dot{s})^\top$; see Exercise 2.5.1. As indicated in Remark 3.4, we either can linearize these first order differential equations as described above, or we can first linearize Equations (2.4) and then afterwards construct a set of linear first order differential equations. We will continue with the latter method. In Exercise 3.5.1 the reader is asked to do the first method and to convince him/herself that the outcome is the same. Linearization of (2.4) around the equilibrium pair (x^*,u^*) with $x^*=0\in\mathbb{R}^n$ and $u^*=0\in\mathbb{R}$, i.e.,

$$\tilde{\varphi}(t)=\dot{\tilde{\varphi}}(t)=\tilde{s}(t)=\dot{\tilde{s}}(t)=0,\quad \tilde{u}(t)=0,\quad \forall t\geq 0,$$

leads to (i.e., the nonlinear terms in (2.4) are replaced by their Taylor series expansion around the chosen solutions up to and including the linear term)

$$\frac{4l}{3}\ddot{\varphi} - g\varphi + \ddot{s} = 0, \quad (M+m)\ddot{s} + ml\ddot{\varphi} = u, \tag{3.14}$$

which can be viewed as two equations with the unknowns $\ddot{\varphi}$ and $\ddot{s}$. These unknowns can be solved and expressed in the other quantities φ, $\dot{\varphi}$, s, $\dot{s}$ and u. This is left as an exercise to the reader.

As in Remark 3.5, note the difference in meaning of the variables φ, $\dot{\varphi}$, s, $\dot{s}$ and u in Equations (2.4) and in Equations (3.14). In (2.4) the variables have the physical meaning as described in Section 2.4.1, whereas in (3.14) the variables should be seen as deviations from the chosen solutions. In general, these two meanings will be different. See also the distinction between x, u in Equations (3.3), and z, v in Equations (3.11). However, since here the zero solutions are chosen, the two meanings of the variables coincide.

Defining the state vector $x = (\varphi, \dot{\varphi}, s, \dot{s})^\top$, Equations (3.14) can be rewritten as

$$\frac{dx}{dt} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ a_{41} & 0 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ b_2 \\ 0 \\ b_4 \end{pmatrix} u, \quad x = \begin{pmatrix} \varphi \\ \dot{\varphi} \\ s \\ \dot{s} \end{pmatrix}, \tag{3.15}$$

where

$$a_{21} = \frac{3g(M+m)}{l(4M+m)}, \quad a_{41} = \frac{-3gm}{4M+m}, \quad b_2 = \frac{-3}{l(4M+m)}, \quad b_4 = \frac{4}{4M+m}.$$

If we take $M = 0.98$ kg, $m = 0.08$ kg, $l = 0.312$ m and $g = 9.8$ m/sec^2, then Equation (3.15) becomes

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ -2.4 \\ 0 \\ 1 \end{pmatrix} u. \tag{3.16}$$

If s and φ are the measured quantities, then the output function is

$$y = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} x. \tag{3.17}$$

□

3.3 Solution of a system of linear differential equations

In this section we briefly list the well-known results about the solution of a system of linear differential equations. A more detailed treatment can be found in a text-book on the theory of linear differential equations such as [3]. We restrict our attention to time-invariant systems

$$\dot{x} = Ax + Bu. \tag{3.18}$$

To describe the solution we need the definition of the exponential of a matrix. For a given $n \times n$ matrix A and a scalar t the exponential e^{At} is defined as the following power series

$$e^{At} = I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \cdots = \sum_{k=0}^{\infty} \frac{A^k t^k}{k!}. \tag{3.19}$$

It can be shown that the series is convergent for all A and t, so that e^{At} is a well-defined $n \times n$ matrix for all A and t.

The solution of the homogeneous system $\dot{x} = Ax$ with the initial condition $x(0) = x_0$ is given by

$$x(t) = e^{At}x_0. \tag{3.20}$$

This solution can be explained by means of the flow diagram in Figure 3.1, representing

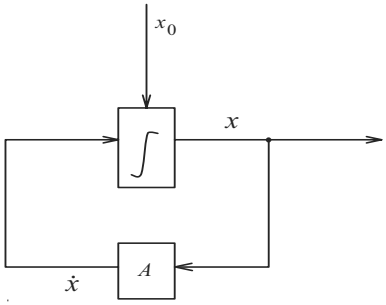

Figure 3.1 Flow diagram of $\dot{x} = Ax$.

the differential equation with initial condition. The box with the integral sign represents the integration of the incoming signal, here $\dot{x}$, starting with the initial condition x_0 at $t = 0$, resulting in the outgoing signal, here x. Hence, this box represents $x(t) = x_0 + \int_0^t \dot{x}(\sigma_1)d\sigma_1$. The box with the matrix A represents the multiplication of the incoming signal with A to yield the outgoing signal. Hence, here this box represents $\dot{x} = Ax$. Now going around once in this diagram, we get

$$x(t) = x_0 + \int_0^t Ax(\sigma_1)d\sigma_1.$$

As $x(\sigma_1)$ can be expressed in the same way as $x(t)$, it follows that, by going around and around in the diagram,

$$\begin{aligned} x(t) &= x_0 + \int_0^t Ax_0 d\sigma_1 + \int_0^t A \int_0^{\sigma_1} Ax_0 d\sigma_2 d\sigma_1 + \\ &\qquad \int_0^t A \int_0^{\sigma_1} A \int_0^{\sigma_2} Ax_0 d\sigma_3 d\sigma_2 d\sigma_1 + \cdots \\ &= \left(I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \cdots \right) x_0 = e^{At}x_0. \end{aligned}$$

The matrix e^{At} is called a transition matrix. The formula $x(t) = e^{At}x(0)$ shows that the state $x(0)$ is transferred to the state $x(t)$ by the matrix e^{At}. It is easy to derive from this that the transfer from the state $x(t_0)$ to the state $x(t_1)$ is accomplished by the matrix $e^{A(t_1-t_0)}$.

Now let's have a look at the non-homogeneous system (3.18). The solution (3.20) of the homogeneous system is modified by the addition of an integral over the interval $[0,t]$ caused by the input signal u. The solution of the system $\dot{x} = Ax + Bu$ with the initial condition $x(0) = x_0$ is given by

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}Bu(s)ds. \tag{3.21}$$

The exponential e^{At} obviously plays an important role in linear system theory. Many papers have been published about what would be a good numerical procedure to calculate this exponential. A possible procedure would be to use a finite number of the terms in the series expansion in the definition in (3.19). This method works reasonably well as long as the eigenvalues of A are close together. For more information and for more reliable methods the reader is referred to [8]. We will now summarize the analytical method of calculating e^{At}, which makes use of several concepts and theorems from linear algebra.

We start with the following lemma.

Lemma 3.7 *If P is an invertible matrix, then $e^{At} = Pe^{(P^{-1}AP)t}P^{-1}$.*

Proof We will show that $e^{P^{-1}APt} = P^{-1}e^{At}P$.

$$\begin{aligned} e^{P^{-1}APt} &= I + P^{-1}APt + \frac{1}{2!}(P^{-1}AP)^2t^2 + \frac{1}{3!}(P^{-1}AP)^3t^3 + \cdots \\ &= P^{-1}P + P^{-1}APt + \frac{1}{2!}P^{-1}A^2Pt^2 + \frac{1}{3!}P^{-1}A^3Pt^3 + \cdots \\ &= P^{-1}(I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \cdots)P = P^{-1}e^{At}P. \end{aligned}$$

□

Suppose that A is diagonalizable, i.e., suppose an invertible matrix T exists such that $T^{-1}AT = D$, where

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

The elements $\lambda_1, \ldots, \lambda_n$ in D are the eigenvalues of A and the columns of T are the corresponding eigenvectors. By means of Lemma 3.7 it now follows that

$$e^{At} = Te^{(T^{-1}AT)t}T^{-1} = Te^{Dt}T^{-1}.$$

The exponential e^{Dt} can easily be obtained by using the definition in (3.19) as

$$e^{Dt} = \sum_{k\geq 0}\frac{1}{k!}D^kt^k = \begin{pmatrix} \sum_{k\geq 0}\frac{\lambda_1^kt^k}{k!} & & 0 \\ & \ddots & \\ 0 & & \sum_{k\geq 0}\frac{\lambda_n^kt^k}{k!} \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix}.$$

Unfortunately not all square matrices are diagonalizable and therefore the method described above cannot be used for arbitrary square matrices. Diagonalization of an $n \times n$ matrix is only possible if the matrix has n linearly independent eigenvectors. A sufficient (but not necessary) condition for the latter to be the case is that all n eigenvalues of the matrix are different. A non-diagonalizable matrix of size $n \times n$ has therefore $k(< n)$ different eigenvalues.

For the sequel the following notions are useful. The **kernel** (or null space) of a matrix M, indicated by $\ker M$, is defined as the set of all vectors x for which $Mx = 0$. The **image** (or column space) of a matrix M, indicated by $\mathrm{im}\, M$, is the set of all linear combinations of the columns of M. Clearly, both $\ker M$ and $\mathrm{im}\, M$ are linear spaces. A linear space $\mathcal{N}$ is the **direct sum** of two linear subspaces $\mathcal{N}_1$ and $\mathcal{N}_2$, notation $\mathcal{N} = \mathcal{N}_1 \oplus \mathcal{N}_2$, if each $x \in \mathcal{N}$ can be uniquely decomposed as $x = x_1 + x_2$ with $x_1 \in \mathcal{N}_1$ and $x_2 \in \mathcal{N}_2$. If M is a square matrix its **determinant** will be denoted by $\det M$.

For each eigenvalue λ_i of a square matrix A two multiplicities are defined. Namely, the **algebraic multiplicity**, which is the usual multiplicity of λ_i as a root of the **characteristic polynomial** $\det(\lambda I - A)$, and the **geometric multiplicity** of λ_i, which is the dimension of the eigenspace $\ker(\lambda_i I - A)$. It can be shown for each eigenvalue that its geometric multiplicity is less than or equal to its algebraic multiplicity.

Returning to the point of diagonalization, it is a well-known and fundamental result from linear algebra that a square matrix A is diagonalizable if and only if for each eigenvalue of A the algebraic multiplicity is equal to the geometric multiplicity. Further, if a square matrix A is not diagonaizable, the matrix can be transformed into a form, the so-called **Jordan form** of A, that is close to a diagonal form. The latter result is stated in the following theorems.

Theorem 3.8 *Suppose that the $n \times n$ matrix A has k different eigenvalues λ_i with algebraic multiplicity m_i, $i = 1, \ldots, k$. Then $\sum_{i=1}^{k} m_i = n$. Define $\mathcal{N}_i = \ker(A - \lambda_i I)^{m_i}$, then*

1. *the dimension of the linear vector subspace $\mathcal{N}_i$ is equal to m_i, $i = 1, \ldots, k$,*
2. *the n dimensional linear vector space $\mathbb{C}^n$ over the complex numbers is the direct sum of the subspaces $\mathcal{N}_i$, i.e., $\mathbb{C}^n = \mathcal{N}_1 \oplus \mathcal{N}_2 \oplus \cdots \oplus \mathcal{N}_k$.*

For a proof of this theorem and other background material on matrix theory the reader is, for instance, referred to [Lancaster and Tismenetsky, 1985]. If the $n \times n$ matrix A has n different eigenvalues, then each $\mathcal{N}_i$, as defined in Theorem 3.8, is a one dimensional subspace spanned by the eigenvector corresponding to λ_i.

The following theorem is a consequence of Theorem 3.8.

Theorem 3.9 *Suppose that the $n \times n$ matrix A has k different eigenvalues λ_i with algebraic multiplicity m_i, $i = 1, \ldots, k$. Then a nonsingular matrix T exists such that*

$$T^{-1} A T = J, \tag{3.22}$$

where J, the so-called Jordan form of A, has a block-diagonal structure defined as $J =$

$\mathrm{diag}(J_1,\ldots,J_k)$*, or equivalently*

$$J = \begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_k \end{pmatrix}. \tag{3.23}$$

For each $i = 1,\ldots,k$*, matrix* J_i *is of size* $m_i \times m_i$*, has* λ_i *as its only eigenvalue and possibly also has a block diagonal structure of its own, i.e., possibly* $J_i = \mathrm{diag}(J_{i1},\ldots,J_{il_i})$*, or equivalently*

$$J_i = \begin{pmatrix} J_{i1} & & & 0 \\ & J_{i2} & & \\ & & \ddots & \\ 0 & & & J_{il_i} \end{pmatrix}, \tag{3.24}$$

where l_i *is a positive integer, and where each subblock has the following form*

$$J_{ij} = \begin{pmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \ldots & \ldots & 0 & \lambda_i \end{pmatrix}. \tag{3.25}$$

Example 3.10 Consider the matrix (already in Jordan form)

$$J = \begin{pmatrix} 2 & 1 & & & & & & & \\ & 2 & & & & & & & \\ & & 2 & 1 & & & & & \\ & & & 2 & & & & & \\ & & & & -1 & 1 & & & \\ & & & & & -1 & 1 & & \\ & & & & & & -1 & & \\ & & & & & & & -1 & \\ & & & & & & & & -1 \end{pmatrix}, \tag{3.26}$$

where the elements that are not explicitly specified are equal to zero. Then

$$J_1 = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{12} \end{pmatrix} = \left(\begin{array}{cc|cc} 2 & 1 & & \\ & 2 & & \\ \hline & & 2 & 1 \\ & & & 2 \end{array}\right),$$

and

$$J_2 = \begin{pmatrix} J_{21} & 0 & 0 \\ 0 & J_{22} & 0 \\ 0 & 0 & J_{23} \end{pmatrix} = \left(\begin{array}{ccc|cc} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & & \\ \hline & & & -1 & \\ & & & & -1 \end{array}\right).$$

Hence, the matrix has two different eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -1$ with algebraic multiplicities $m_1 = 4$ and $m_2 = 5$, respectively. Note that the geometric multiplicities of λ_1 and λ_2 are 2 and 3, respectively. Further, $l_1 = 2$ and $l_2 = 3$ and the $J_{11}, J_{12}, J_{21}, J_{22}$ and J_{23} are as indicated above. □

If the matrix T of Equation (3.22) is partitioned as $T = \left(T_1, T_2, \ldots, T_k\right)$, conform the partitioning in Equation (3.23), then the columns of submatrix T_j form a basis for the subspace $\mathcal{N}_j$. Equation (3.22) yields that $AT = TJ$. If the individual columns of T are denoted by $q_1, \ldots, q_n$, then the i-th column of AT equals Aq_i. The i-th column of TJ equals $\lambda q_i + \gamma_i q_{i-1}$, with γ_i either 0 or 1, depending on the location of the i-th column in J in relation to some appropriate Jordan block corresponding to eigenvalue λ. Hence

$$Aq_i = \lambda q_i + \gamma_i q_{i-1}, \quad \forall i = 1, \ldots, n, \quad \text{with} \quad \gamma_i \in \{0, 1\}, \tag{3.27}$$

where λ is an eigenvalue and where γ_i is either zero or one. If $\gamma_i = 0$, then q_i is an eigenvector of A and can be obtained by solving $(A - \lambda_i I)q_i = 0$. If $\gamma_i = 1$, then q_i is a so-called **generalized eigenvector** and can be obtained by solving $(A - \lambda_i I)q_i = q_{i-1}$, where q_{i-1} is a (generalized) eigenvector obtained earlier in a similar way.

Now we are in a position to calculate e^{At}, namely by

$$e^{At} = Te^{Jt}T^{-1}.$$

Application of the definition of e^{Jt}, see (3.19), gives $e^{Jt} = \operatorname{diag}(e^{J_1 t}, \ldots, e^{J_k t})$, and for each block, $e^{J_i t} = \operatorname{diag}(e^{J_{i1} t}, \ldots, e^{J_{il_i} t})$. Finally, for each subblock,

$$e^{J_{ij}t} = e^{\lambda_i t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \cdots & \frac{t^{d_{ij}-1}}{(d_{ij}-1)!} \\ & \ddots & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \frac{t^2}{2!} & \vdots \\ & & & \ddots & t & \frac{t^2}{2!} \\ & & & & 1 & t \\ 0 & & & & 0 & 1 \end{pmatrix}, \tag{3.28}$$

where d_{ij} is the dimension of J_{ij}. See Exercises 3.5.12 and 3.5.13 for a proof.

Remark 3.11 Please note that if $\tilde{q}_1, \ldots, \tilde{q}_{d_{ij}}$ are the (generalized) eigenvectors belonging to the Jordan block J_{ij} (and this block on its turn corresponds to the eigenvalue λ_i), then $(A - \lambda_i I)^k \tilde{q}_k = 0$ for $k = 1, \ldots, d_{ij}$. This can been proved as follows. For $k = 1$ obviously $(A - \lambda_i I)\tilde{q}_1 = 0$ because $\tilde{q}_1$ is an eigenvector (and not a generalized one). For $k = 2$ we can write $(A - \lambda_i I)^2 \tilde{q}_2 = (A - \lambda_i I)(A - \lambda_i I)\tilde{q}_2 = (A - \lambda_i I)\tilde{q}_1 = 0$, where we used Equation (3.27), with $\lambda = \lambda_i$, $q_i = \tilde{q}_2$, $q_{i-1} = \tilde{q}_1$ and $\gamma_i = 1$, yielding that $\tilde{q}_1 = (A - \lambda_i I)\tilde{q}_2$. The proof by induction can be continued for higher values of k. Thus the vectors $\tilde{q}_1, \ldots, \tilde{q}_{d_{ij}}$ span the linear subspace $\mathcal{N}_i$ as introduced in the statement of Theorem 3.8, in case there is one Jordan block corresponding to the eigenvalue λ_i. The vectors span part of this subspace if there is more than one Jordan block corresponding to λ_i. □

Example 3.12 This is a continuation of Example 3.6. Calculate the transition matrix for the system given in Equation (3.16). The characteristic polynomial is $\lambda^2(\lambda^2-25)$ and therefore the eigenvalues are $\lambda_{1,2}=0$, $\lambda_3=5, \lambda_4=-5$. The eigenspace corresponding to eigenvalue 0 is one dimensional, because the matrix $A-\lambda_{1,2}I$ has rank 3. So, in addition to an eigenvector in the usual sense, an extra generalized eigenvector for the eigenvalue 0 is needed, which can be computed using Equation (3.27). The result of these computations is given below.

$$J=\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & -5 \end{pmatrix}, \quad e^{Jt}=\begin{pmatrix} 1 & t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{5t} & 0 \\ 0 & 0 & 0 & e^{-5t} \end{pmatrix}.$$

The matrices T and T^{-1} are

$$T=\begin{pmatrix} 0 & 0 & -125 & 125 \\ 0 & 0 & -625 & -625 \\ 1 & 0 & 3 & -3 \\ 0 & 1 & 15 & 15 \end{pmatrix}, \quad T^{-1}=\frac{1}{1250}\begin{pmatrix} 30 & 0 & 1250 & 0 \\ 0 & 30 & 0 & 1250 \\ -5 & -1 & 0 & 0 \\ 5 & -1 & 0 & 0 \end{pmatrix}$$

and

$$e^{At}=Te^{Jt}T^{-1}=\begin{pmatrix} \cosh 5t & \frac{1}{5}\sinh 5t & 0 & 0 \\ 5\sinh 5t & \cosh 5t & 0 & 0 \\ \frac{3}{125}(1-\cosh 5t) & \frac{3}{625}(5t-\sinh 5t) & 1 & t \\ -\frac{3}{25}\sinh 5t & \frac{3}{125}(1-\cosh 5t) & 0 & 1 \end{pmatrix},$$

where $\cosh 5t=\frac{1}{2}(e^{5t}+e^{-5t})$ and $\sinh 5t=\frac{1}{2}(e^{5t}-e^{-5t})$. Observe that the first column of T is an eigenvector of A in the usual sense and that the second column of T is a generalized eigenvector of A, both for the eigenvalue $\lambda_{1,2}=0$. □

Example 3.13 This is a continuation of Exercise 3.5.2. Calculate the transition matrix for the system given in Equation (3.43) with $\omega=1$. The characteristic polynomial is $\lambda^4+\lambda^2$ and therefore the eigenvalues are $\lambda_{1,2}=0$, $\lambda_3=i$, $\lambda_4=-i$. Like above, for the eigenvalue 0 the eigenspace is one dimensional. Again an extra generalized eigenvector for this eigenvalue needs to be computed. The result of the computations yields

$$J=\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix}, \quad e^{Jt}=\begin{pmatrix} 1 & t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{it} & 0 \\ 0 & 0 & 0 & e^{-it} \end{pmatrix}.$$

The matrices T and T^{-1} are

$$T=\begin{pmatrix} 0 & -2 & 1 & 1 \\ 0 & 0 & i & -i \\ 3 & 0 & 2i & -2i \\ 0 & 3 & -2 & -2 \end{pmatrix}, \quad T^{-1}=\begin{pmatrix} 0 & -\frac{2}{3} & \frac{1}{3} & 0 \\ -3 & 0 & 0 & -1 \\ -\frac{3}{2} & -\frac{1}{2}i & 0 & -1 \\ -\frac{3}{2} & \frac{1}{2}i & 0 & -1 \end{pmatrix},$$

and

$$e^{At} = Te^{Jt}T^{-1} = \begin{pmatrix} 4-3\cos t & \sin t & 0 & 2-2\cos t \\ 3\sin t & \cos t & 0 & 2\sin t \\ -6t+6\sin t & -2+2\cos t & 1 & -3t+4\sin t \\ -6+6\cos t & -2\sin t & 0 & -3+4\cos t \end{pmatrix},$$

where $\cos t = \frac{1}{2}(e^{it}+e^{-it})$ and $\sin t = \frac{1}{2i}(e^{it}-e^{-it})$. It follows that the first column of T is an eigenvector of A in the usual sense and the second column of T is a generalized eigenvector of A, both for the eigenvalue $\lambda_{1,2} = 0$. □

For diagonalizable matrices A we can write

$$A = TDT^{-1} = \begin{pmatrix} v_1 \cdots v_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}, \tag{3.29}$$

where $v_1,\ldots,v_n$ are the columns of T (formerly we also used the notation q_i for the columns), being the eigenvectors of A, and where $w_1,\ldots,w_n$ are the rows of T^{-1}. It easily follows that $A = \sum_{i=1}^{n} \lambda_i v_i w_i$. The product $v_i w_i$, of a column v_i with a row w_i, is an $n \times n$ matrix, called a **dyad** (a dyad has at most rank one). Then the matrix A is the sum of n dyads. The same applies to the transition matrix, since it can be written as

$$e^{At} = T \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} T^{-1} = \sum_{i=1}^{n} e^{\lambda_i t} v_i w_i. \tag{3.30}$$

The solution of $\dot{x} = Ax$ with $x(0) = x_0$ can therefore be written as

$$x(t) = e^{At}x_0 = \sum_{i=1}^{n} e^{\lambda_i t} v_i w_i x_0 = \sum_{i=1}^{n} \mu_i e^{\lambda_i t} v_i, \tag{3.31}$$

where $\mu_i = w_i x_0$ is a scalar quantity. The solution of $\dot{x} = Ax$ (or of $\dot{x} = Ax + Bu$ with $u = 0$, the reason why this solution is sometimes called the **free response**) is thus decomposed along the eigenvectors, i.e., it is a linear combination of terms with exponential coefficients. The solution corresponding to only one eigenvector (i.e., x_0 is such that $\mu_i \neq 0$ for some i and $\mu_k = 0$ for $k \neq i$) is called a **mode** of the system. If the initial vector is aligned with one eigenvector, then the corresponding solution is completely situated in the one dimensional space spanned by this eigenvector. Generalizations of Equations (3.29) and (3.30) to the non-diagonalizable case exist, but will not be treated here.

Tacitly we assumed that λ_i and therefore v_i were real in the treatment above. For complex λ_i and v_i the formulation above can be adjusted as follows. Suppose $\lambda = \sigma + i\omega$ is an eigenvalue of A (σ and ω are real and i now denotes the imaginary unit) with a corresponding eigenvector $v = r + is$, where $r,s \in \mathbb{R}^n$. It is clear that $\sigma \in \mathbb{R}$ and $r \in \mathbb{R}^n$ are the real parts of $\lambda \in \mathbb{C}$ and $v \in \mathbb{C}^n$, respectively, and $\omega \in \mathbb{R}$ and $s \in \mathbb{R}^n$ are the imaginary parts of $\lambda \in \mathbb{C}$ and $v \in \mathbb{C}^n$, respectively. Denoting Re for the real part and Im for the imaginary part of a variable or expression, it follows that $\mathrm{Re}\,\lambda = \sigma, \mathrm{Im}\,\lambda = \omega, \mathrm{Re}\,v = r$

and $\mathrm{Re}\, v = s$. Because A is real and $Av = \lambda v$, also $A\bar{v} = \bar{\lambda}\bar{v}$, where the upperbar denotes the complex conjugate. Therefore, $\bar{\lambda} = \sigma - i\omega$ is also an eigenvalue of A, with eigenvector $r - is$. Suppose that x_0 lies in the subspace spanned by r and s. Then there exist $a, b \in \mathbb{R}$ such that

$$x_0 = ar + bs = \frac{1}{2}(a - ib)(r + is) + \frac{1}{2}(a + ib)(r - is) = \mu v + \bar{\mu}\bar{v},$$

where $\mu = \frac{1}{2}(a - ib) \in \mathbb{C}$. The corresponding free response is

$$x(t) = \mu e^{\lambda t} v + \bar{\mu} e^{\bar{\lambda} t} \bar{v}.$$

If μ is written in polar form as $\mu = \frac{1}{2i} p e^{i\varphi}$, with p and φ real, then

$$\begin{aligned} x(t) &= \frac{1}{2i} p \left(e^{\lambda t + i\varphi} v - e^{\bar{\lambda} t - i\varphi} \bar{v}\right) \\ &= p\,\mathrm{Im}\left(e^{\lambda t + i\varphi} v\right) = p\,\mathrm{Im}\left(e^{\sigma t + i(\omega t + \varphi)}(r + is)\right) \\ &= p e^{\sigma t}\left(r \sin(\omega t + \varphi) + s \cos(\omega t + \varphi)\right). \end{aligned}$$

3.4 Impulse response and step response

The solution of $\dot{x} = Ax + Bu$, with $x(t_0) = x_0$ is given by

$$x(t) = e^{A(t - t_0)} x_0 + \int_{t_0}^{t} e^{A(t-s)} B u(s) ds. \tag{3.32}$$

Now let be given an output function of the form

$$y(t) = Cx(t) + Du(t).$$

Then we find the following relation between the output function $y = y(t)$, the initial state x_0 and the input function $u = u(t)$:

$$y(t) = C e^{A((t - t_0)} x_0 + \int_{t_0}^{t} C e^{A(t-s)} B u(s) ds + Du(t). \tag{3.33}$$

Let the $p \times m$ matrix $K(t,s)$ be defined by

$$K(t,s) = C e^{A(t-s)} B. \tag{3.34}$$

This is called the **impulse response matrix** of the system. We shall explain this terminology below.

Suppose a time t_0 exists such that $x_0 (= x(t_0)) = 0$. We are only interested in the system for $t \geq t_0$ and assume $u(s) = 0$ for $s < t_0$. Then Equation (3.33) can be written as

$$y(t) = (\mathscr{F} u)(t) = \int_{-\infty}^{t} K(t,s) u(s) ds + Du(t), \tag{3.35}$$

where $\mathscr{F}$ is a mapping which maps an m-dimensional input function u, which is supposed to be zero before some time t_0, to a p-dimensional output function y. Note that $\mathscr{F}$ is a linear mapping, and that $K(t,s)$ and D provide a characterization of the **external description** of the system. This is also referred to as the external behavior of the system. See Chapter 8 for a discussion on the external description and behavior. Heuristically speaking, an external description refers to the situation where the input function is directly mapped to an output function, without the use of an internal state x. This 'intermediate' state has been eliminated.

Now let us assume that $D = 0$. Then Equation (3.34) becomes

$$y(t) = \int_{-\infty}^{t} K(t,s)\,u(s)ds. \tag{3.36}$$

The matrix function $K(t,s)$ has the following interpretation. Suppose the input function is $u(t) = \delta(t-t_1)\,e_i$, where e_i is the i-th basis vector in $\mathbb{R}^m$ (the i-th column of the $m \times m$ identity matrix) and $\delta(t-t_1)$ is the so-called **delta function** defined as

$$\int_{-\infty}^{\infty} \varphi(t)\,\delta(t-t_1)dt = \varphi(t_1),$$

for any continuous function φ. Heuristically, the $\delta(t-t_1)$ function can be defined as the limit for $n \to \infty$ of the sequence of functions

$$f_n(t-t_1) = \begin{cases} \dfrac{n}{2} & \text{for} \quad |t-t_1| < \dfrac{1}{n}, \\ \\ 0 & \text{for} \quad |t-t_1| \geq \dfrac{1}{n}. \end{cases}$$

The output for such an input for $t \geq t_1$ is given by

$$y(t) = \int_{-\infty}^{t} K(t,s)\,\delta(s-t_1)\,e_i ds = K(t,t_1)e_i = i\text{th column of } K(t,t_1).$$

The columns of $K(t,t_1)$ can be interpreted as the response of the system (being the output) at time $t \geq t_1$ caused by an impulse shaped input function (i.e., a δ function) at time t_1. This is the reason why $K(t,s)$ is called the impulse response matrix.

Note that the matrix $K(t,s) = Ce^{A(t-s)}B$ does in fact depend on a single variable, viz. the variable $t-s$. Instead of $K(t,s)$ one often writes $G(\tau)$, where τ is understood to be equal to $t-s$, i.e.,

$$G(\tau) = Ce^{A\tau}B. \tag{3.37}$$

Often t is used instead of τ, which of course is not the same variable as the t in $K(t,s)$.

Another important response matrix is the so-called step response matrix. Instead of an impulse shaped input function now a step shaped function will be applied. Such a step function, or **Heaviside function**, $H(t-t_1)$ is defined as

$$H(t-t_1) = \begin{cases} 1 & \text{for} \quad t \geq t_1, \\ \\ 0 & \text{for} \quad t < t_1. \end{cases}$$

Note that $H(t-t_1)$ does belong to the class of admissible input functions (piecewise continuous functions), whereas one has to be very careful with impulse functions (strictly speaking, the delta function does not satisfy the conventional definition of a function). Also note that a step function is an integrated version of the impulse function, i.e.,

$$H(t-t_1) = \int_{-\infty}^{t} \delta(s-t_1)ds.$$

The output corresponding to the step function $H(t-t_1)e_i$ is, assuming that the system starts at the origin at a time t_0 in the past, is for $t \geq t_1$ given by

$$y(t) = \int_{-\infty}^{t} K(t,s)H(s-t_1)e_i ds = \int_{t_1}^{t} K(t,s)e_i ds = (\int_{t_1}^{t} K(t,s)ds)e_i.$$

The matrix that appears on the right hand side between brackets is called the **step response matrix** of the system, and will be denoted by $S(t,t_1)$. Denoting the second variable by s we obtain the following definition of the step response matrix:

$$S(t,s) = \int_{s}^{t} K(t,\tau)d\tau. \tag{3.38}$$

The two partial derivatives of S are easily seen to satisfy

$$\frac{\partial}{\partial t}S(t,s) = K(t,s), \tag{3.39}$$

$$\frac{\partial}{\partial s}S(t,s) = -K(t,s). \tag{3.40}$$

Example 3.14 We start with the linearized Equation (3.6) of the satellite dynamics; see Exercise 3.5.2. Assume that both the angle θ and the distance r are measured and processed to yield $r-\sigma$ and $\theta-\omega t$ (σ and ω are constants, r and θ are functions of time.) Hence,

$$y = \begin{pmatrix} r-\sigma \\ \theta-\omega t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma} & 0 \end{pmatrix} x. \tag{3.41}$$

Take for the constants $\omega = 1$ and $\sigma = 1$. The impulse response matrix for this system is (see Example 3.13 for the calculation of e^{At})

$$G(t) = Ce^{At}B = \begin{pmatrix} \sin t & 2-2\cos t \\ -2+2\cos t & -3t+4\sin t \end{pmatrix}.$$

□

The external description (3.36) does not only hold for strictly causal linear differential systems, as is shown by the following example (for a description of (strict) causality, see Remark 3.16).

Example 3.15 Consider a single-input single-output system of the form

$$y(t) = \frac{1}{T} \int_{t-T}^{t} u(s)ds,$$

which is sometimes called a **moving average**. This system is linear, time-invariant and the impulse response function is

$$G(\tau) = \begin{cases} \dfrac{1}{T} & \text{for} \quad 0 \leq \tau \leq T, \\ \\ 0 & \text{for} \quad \tau > T \text{ and } \tau < 0. \end{cases}$$

This is an example of an infinte-dimensional linear system. The state space of this system is a function space. □

In Equation (3.36) the upper bound of the integral is t. However, if a given system description requires that we have to write

$$y(t) = \int_{-\infty}^{+\infty} K(t,s)u(s)ds, \text{ or } y(t) = \int_{-\infty}^{+\infty} G(t-s)u(s)ds, \tag{3.42}$$

with the upper bound $+\infty$ instead of t, then we are dealing with a so-called non-causal system, see also below.

Remark 3.16 The formal definition of **causality** will be given in Chapter 8. Heuristically, it means that the current evolution of a system cannot depend on phenomena which will happen in the future. For state space descriptions (strict) causality can be characterized as follows. For a **strictly causal** system the present state only depends on the past states and past inputs. If a system is only **causal** (and not strictly causal), then the present state is only allowed to depend on the past states and the past and *present* input. A system that is not causal is called a **non-causal** system. In such a system the present state is allowed to depend on, for instance, future inputs. □

The relation in Equation (3.42) does not define a system according to the definitions given here. The causal systems form a subclass of the class of systems described by (3.42) by requiring

$$K(t,s) = 0 \quad \text{for} \quad t < s, \quad \text{or} \quad G(\tau) = 0 \quad \text{for} \quad \tau < 0\,.$$

The external behavior of a linear differential system is completely determined by the matrices $K(t,s)$ and D. We show that different triples (A,B,C) of system matrices can produce the same impulse response matrix $K(t,s)$.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible basis transformation of the state space $\mathbb{R}^n$ such that $x = Tz$. Then the transformed coordinate state vector $z = T^{-1}x$ satisfies the equations

$$\dot{z} = T^{-1}\dot{x} = T^{-1}Ax + T^{-1}Bu = T^{-1}ATz + T^{-1}Bu,$$

$$y = Cx + Du = CTz + Du.$$

Hence the triple (A,B,C) is sent by T to the triple $(T^{-1}AT, T^{-1}B, CT)$. A straightforward computation of the impulse response matrix for the transformed system yields

$$CTe^{T^{-1}ATt}T^{-1}B = CTT^{-1}e^{At}TT^{-1}B = Ce^{At}B,$$

which shows that the impulse response matrix does not change under a basis transformation. This of course should be expected since the choice of a new basis in the state space should not change the external behavior of a system. These considerations are formalized in the following definition.

Definition 3.17 *Two linear systems*

$$\left\{\begin{array}{rcl} \dot{x} &=& Ax+Bu \\ y &=& Cx+Du, \end{array}\right. \qquad \left\{\begin{array}{rcl} \dot{\tilde{x}} &=& \tilde{A}\tilde{x}+\tilde{B}u, \\ y &=& \tilde{C}\tilde{x}+\tilde{D}u, \end{array}\right.$$

are called equivalent if an invertible linear mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ exists such that

$$\tilde{A} = T^{-1}AT, \quad \tilde{B} = T^{-1}B, \quad \tilde{C} = CT, \quad \tilde{D} = D.$$

Obviously this notion of equivalence is concerned with the triples (A,B,C) only (the matrix D plays no role). Note that $\tilde{C}e^{\tilde{A}t}\tilde{B} = Ce^{At}B$ for all equivalent triples $(\tilde{A},\tilde{B},\tilde{C})$ and (A,B,C).

Equivalent systems are defined on a state space of the same dimension. It is also possible for two systems on state spaces of different dimensions to have the same impulse response matrix. A trivial example of this is obtained by adding a vector equation which does not affect the output. For instance,

$$\left\{\begin{array}{rcl} \dot{x} &=& Ax+Bu, \\ \dot{\hat{x}} &=& F\hat{x}+Gu, \end{array}\right. \quad \text{with} \quad y = Cx+Du,$$

written as

$$\frac{d}{dt}\begin{pmatrix} x \\ \hat{x} \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & F \end{pmatrix}\begin{pmatrix} x \\ \hat{x} \end{pmatrix} + \begin{pmatrix} B \\ G \end{pmatrix}u,$$

$$y = \begin{pmatrix} C & 0 \end{pmatrix}\begin{pmatrix} x \\ \hat{x} \end{pmatrix} + Du.$$

Apparently, there can be no upper-bound on the dimension of a realization of a given impulse response matrix. However, under reasonable conditions a lower bound does exist. If a triple (A,B,C) realizes a given impulse response matrix $G(t)$, it will be called a **minimal realization** if there exists no realization of $G(t)$ on a state space of lower dimension. The minimum dimension is called the order of the impulse response matrix.

A well-known branch of system theory is concerned with the realization problem: given the external description of a system, such as for instance determined by the mapping $\mathscr{F}$ introduced in Equation (3.35), determine a state space description. For linear time-invariant finite dimensional differential systems this problem boils down to the following: given the impulse response matrix $G(t) + D\delta(t)$ (see Exercise 3.5.16 for the extra term $D\delta(t)$) find an $n \times n$ matrix A, $n \times m$ matrix B and a $p \times n$ matrix C such that $G(t) = Ce^{At}B$, where also n is to be determined. Note that a realization of an impulse response matrix $G(t)$, if it exists, is not unique (any equivalent system will do).

3.5 Exercises

Exercise 3.5.1 *Note that $\tilde{x}(t) = 0 \in \mathbb{R}^4$ and $\tilde{u}(t) = 0$, for all $t \geq 0$, is a solution pair of the system description obtained in Exercise 2.5.1. Follow the approach in the beginning of the present chapter to obtain the linearization around the above solution pair and compare the result with Example 3.6.*

Exercise 3.5.2 *This is a continuation of Section 2.4.2. Consider a satellite of unit mass in earth orbit specified by its position and velocity in polar coordinates $r, \dot{r}, \theta, \dot{\theta}$. The input functions are a radial thrust $u_1(t)$ and a tangential thrust $u_2(t)$. Newton's laws yield*

$$\ddot{r} = r\dot{\theta}^2 - \frac{g}{r^2} + u_1, \quad \ddot{\theta} = -\frac{2\dot{\theta}\dot{r}}{r} + \frac{1}{r}u_2.$$

(Compare Equation (2.6) with $m_s = 1, u_1 = F_r, u_2 = F_\theta$ and $Gm_e = g$.) Show that, if $\tilde{u}_1(t) = \tilde{u}_2(t) = 0$, for all $t \geq 0$, then $\tilde{r}(t) = \sigma$ (constant), $\tilde{\theta}(t) = \omega t$ (ω is constant), with $\sigma^3\omega^2 = g$, is a solution, and that linearization around this solution (with $z_1(t) = r(t) - \sigma, z_2(t) = \dot{r}(t), z_3(t) = \sigma(\theta(t) - \omega t), z_4(t) = \sigma(\dot{\theta}(t) - \omega), v_1(t) = u_1(t), v_2(t) = u_2(t)$) leads to

$$\frac{dz}{dt} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3\omega^2 & 0 & 0 & 2\omega \\ 0 & 0 & 0 & 1 \\ 0 & -2\omega & 0 & 0 \end{pmatrix} z + \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} v. \tag{3.43}$$

□

Exercise 3.5.3 *Given the state description*

$$\begin{aligned} \dot{x}_1 &= x_2(x_2 - 1) + u_1 \cos u_2 \\ \dot{x}_2 &= u_1 \sin x_2 \\ y &= (x_1 + u_2)^2 - e^{x_2} \end{aligned}$$

Show that for $u_1 \equiv 0$, $u_2 \equiv 1$ a solution is given by $x_1 \equiv 1$, $x_2 \equiv 0$, $y \equiv 3$. Linearize the state equations and the output equation around this solution and write the result in matrix form ($\dot{z} = Az + Bv$, $w = Cz + Dv$).

Exercise 3.5.4 *Consider the state-space system*

$$\begin{aligned} \dot{x}_1 &= -x_1^6 - x_2 \\ \dot{x}_2 &= x_1 + u \\ y_1 &= x_2^2 \\ y_2 &= x_1, \end{aligned}$$

with input u, outputs $y = (y_1, y_2)^\top$, and state $x = (x_1, x_2)^\top$.

1. *Determine the equilibrium point(s) of the state-space system for $u^* = 1$.*
2. *Linearize the system for $u^* = 1$. Take v as input and w as output of the linearized system.*

Exercise 3.5.5 *Consider the state-space system*

$$\begin{aligned} \dot{x}_1 &= -x_2^2 u_1 \\ \dot{x}_2 &= -x_1 + u_2 \\ \dot{x}_3 &= 1 - x_2 \\ y &= x_1, \end{aligned}$$

with inputs $u = (u_1, u_2)^\top$, *output* y, *and state* $x = (x_1, x_2, x_3)^\top$.

1. *Determine the equilibrium point of the state-space system for* $u_1^* = 0$ *and* $u_2^* = 1$.
2. *Linearize the system around the equilibrium point found under a). Take v as input and w as output of the linearized system.*

Exercise 3.5.6 *Given the differential equations*

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= -x_1(t) - x_2^2(t) + u(t) \end{aligned}$$

and the output function $y(t) = x_1(t)$. *Show that for* $\tilde{u}(t) = \cos^2(t)$ *a solution of the differential equations is* $\tilde{x}_1(t) = \sin t$, $\tilde{x}_2(t) = \cos t$. *Linearize the state equations and the output function around this solution and write the result in matrix form. Is the linearized system time-invariant?*

Exercise 3.5.7 *For* $t \geq 1$, *the following nonlinear system is given.*

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

1. *Show that if* $u(t) = \sin t$, *then* $x_1(t) = t \sin t$ *en* $x_2(t) = -\cos t$ *form a solution of the state equations.*
2. *Linearize the system, including the equation for* $y(t)$, *around the above solution.*

Exercise 3.5.8 *A tractor with* $n-2$ *axles connected to it (if n is even then these axles can be interpreted as* $(n-2)/2$ *wagons), see Figure 3.2, follows a linear track, i.e., the middles of all axles (including the two axles of the tractor) are approximately on one line l. Each wagon is connected by means of a pole to the hook-up point of the preceding wagon. This hook-up point is exactly in the middle of the rear axle of this preceding wagon. The distances of the middles of all axles to line l are not exactly zero (due to perturbations) and are indicated by* $x_1, \ldots, x_n$; *the distance of the midpoint of the two front wheels of the tractor to the line is* x_1 *and the distance of the middle of the last axle, furthest away from the tractor, to the line is* x_n. *With these 'distances' is meant the distance vertical to the line l. The tractor moves with unit speed forward. The (scalar) control is the angle that the front wheels of the tractor make with respect to the symmetry axis of the tractor (with* $u = 0$ *the tractor moves in a straight line (not necessarily line l). It is assumed that*

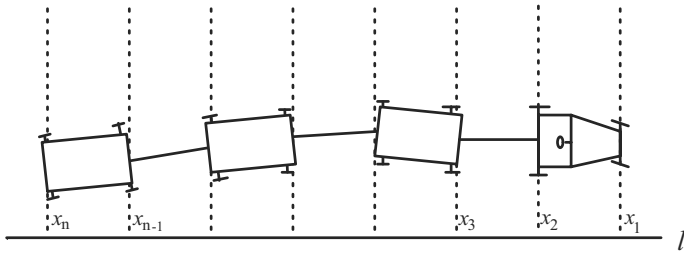

Figure 3.2 Tractor with carts.

the distance between two successive middles of axles equals 1. Show that the linearized equations (around $x_i = 0, i = 1, \ldots, n$, and $u = 0$) are given by

$$\frac{dx}{dt} = \begin{pmatrix} 1 & -1 & 0 & \ldots & \ldots & 0 \\ 1 & -1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ 0 & \ldots & \ldots & 1 & -1 & 0 \\ 0 & \ldots & \ldots & 0 & 1 & -1 \end{pmatrix} x + \begin{pmatrix} \varepsilon \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{pmatrix} u, \tag{3.44}$$

for some appropriate ε (depending on the scaling of the steering wheels). What are the linearized equations of motion if the tractor moves with speed 1 in the backward direction? □

Exercise 3.5.9 *Consider the following system (with only scalar quantities).*

$$\begin{aligned} \dot{x}(t) &= e^t(-x(t) + u(t)), \\ y(t) &= x(t). \end{aligned}$$

1. *Is the system time invariant?*
2. *Apply the transformation $\tau = e^t$. Hence, the time scale is changed. Is the transformed system time invariant? How are the two answers related?*

Exercise 3.5.10 *Calculate e^{At} if*

1. $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$,
2. $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$,
3. $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$,
4. $A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$,
5. $A = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix}$,
6. $A = \begin{pmatrix} 5 & 9 \\ -1 & -1 \end{pmatrix}$.

Exercise 3.5.11 *If A_1 and A_2 commute (i.e., $A_1A_2 = A_2A_1$), then $e^{(A_1+A_2)t} = e^{A_1t}e^{A_2t}$. Prove this. As a hint, prove first that $\frac{d}{dt}e^{A_1t}e^{A_2t} = (A_1+A_2)e^{A_1t}e^{A_2t}$. Give a counterexample if A_1 and A_2 do not commute.*

Exercise 3.5.12 *Consider the $n \times n$ matrix*

$$N = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & 0 & 1 \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}.$$

So N has zeros everywhere except for the diagonal directly above the main diagonal, where it has ones. Using Equation (3.19) prove that

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

Exercise 3.5.13 *Let matrix J_{ij} be as in Theorem 3.9 and assume that it has size $d_{ij} \times d_{ij}$. Note that $J_{ij} = \lambda_i I + N$, where I is the $d_{ij} \times d_{ij}$ identity matrix and where N is a $d_{ij} \times d_{ij}$ matrix as in Exercise 3.5.12. Using Exercise 3.5.11 prove the expression for $e^{J_{ij}t}$ in Equation (3.28).*

Exercise 3.5.14 *Let be given the n-th order system $\dot{x} = Ax$ with*

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -p_0 & -p_1 & \cdots & -p_{n-2} & -p_{n-1} \end{pmatrix}.$$

Show that the characteristic polynomial of A, i.e., $\det(sI - A)$, is given by

$$s^n + p_{n-1}s^{n-1} + \ldots + p_1 s + p_0.$$

If λ is an eigenvalue of A, then prove that the corresponding eigenvector is

$$(1, \lambda, \lambda^2, \ldots, \lambda^{n-1})^\top.$$

Exercise 3.5.15 *Show that a Jordan form of the system matrix A of Exercise 3.5.8 (the tractor example) equals*

$$J = \left(\begin{array}{cc|ccccc} 0 & 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & \dots & \dots & 0 \\ \hline 0 & 0 & -1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & -1 & 1 & & 0 \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & & 0 & -1 & 1 \\ 0 & 0 & \dots & \dots & \dots & 0 & -1 \end{array}\right).$$

Exercise 3.5.16 *Show that for linear systems with $D \not\equiv 0$, the impulse response can be defined as $K(t,s)+D\delta(s-t)$, where δ is the delta function.*

Exercise 3.5.17 *Let be given two linear differential systems in a series interconnection, as depicted in Figure 3.3 The in- and outputs are scalar functions and the impulse re-*

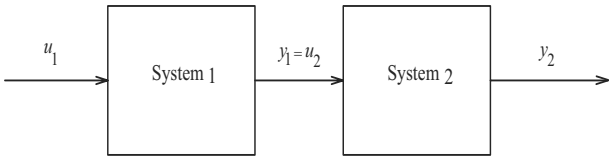

Figure 3.3 Series interconnection of systems.

sponse functions of the two systems are $K_i(t,s)$, $i = 1,2$. Prove that the impulse response function of the series connection is given by

$$K(t,\tau) = \int_{\tau}^{t} K_2(t,v)K_1(v,\tau)dv.$$

Exercise 3.5.18 *Verify the equalities in Equations (3.39) and (3.40).*

Exercise 3.5.19 *Consider a real $n \times n$ matrix A with characteristic polynomial $p(s) = \det(sI-A)$, written as $p(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$. The purpose of this exercise is to prove the theorem of Cayley-Hamilton, i.e., to show that $p(A) = 0$, where $p(A) = A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I$, and I and 0 denote the $n \times n$ identity and zero matrix, respectively. Therefore, let A have Jordan form J, as in Theorem 3.9. Show that $p(s)$ also is the characteristic polynomial of J. Consider the $d_{ij} \times d_{ij}$ subblock matrix J_{ij} as in (3.25) and let $q_{ij}(s)$ be its characteristic polynomial, i.e., $q_{ij}(s) = \det(sI - J_{ij})$. Clearly, $q_{ij}(s) = (s-\lambda_i)^{d_{ij}}$. Prove that $q_{ij}(J_{ij}) = 0$ and that $q_{ij}(s)$ divides $p(s)$. Next show that $p(J_{ij}) = 0$. Finally, conclude that $p(J) = 0$ and, consequently, that $p(A) = 0$.*

Exercise 3.5.20 *Let $\dot{x} = Ax + Bu$, $y = Cx + Du$ and $\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}u$, $y = \tilde{C}\tilde{x} + \tilde{D}u$ be two systems. Show that if $AT = T\tilde{A}$, $B = T\tilde{B}$, $CT = \tilde{C}$ and $D = \tilde{D}$, for some matrix T, then the two systems are not necessarily isomorphic, but do have the same impulse response matrix.*

Exercise 3.5.21 *Below you will find a number of statements. For each of statements determine whether it is true or false. Make your answer plausible by means of a simple reasoning or (counter)example.*

1. *Every system has a finite dimensional state space.*
2. *The linearization of a system around a time-varying solution always results in a time-dependent linearization.*
3. *Every system can only have a finite number of equilibrium points.*
4. *The next two matrices have the same Jordan form*

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} i & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -i \end{pmatrix}.$$

5. *The next two matrices have the same Jordan form*

$$\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

6. *The next two matrices have the same Jordan form*

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

7. *The next two matrices have the same Jordan form*

$$\begin{pmatrix} 0 & 1 & 0 & 3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 & 3 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

8. *If A_1, A_2 are two $n \times n$ matrices such that $(A_1 + A_2)(A_1 - A_2) = A_1^2 - A_2^2$, then $e^{A_1 t} e^{A_2 t} = e^{(A_1 + A_2)t}$.*

Chapter 4

System Properties

4.1 Stability

Several concepts of stability for differential equations exist. They can be distinguished according to stability corresponding to autonomous systems (related to the state vector) and to stability corresponding to systems with inputs and outputs (where the stability is defined in terms of these inputs and outputs). The next four sections deal with the first mentioned concept of stability, the subsequent fifth section deals with input/output stability. Proofs of the theorems related to Routh's criterion and interval stability are not given here. They make an extensive use of complex function theory and fall outside the scope of this book. However, references to papers with short/simple proofs of Routh's criterion and interval stability are included.

4.1.1 Stability in terms of eigenvalues

Definition 4.1 *Consider the first order differential equation $\dot{x} = f(x)$, with $x \in \mathbb{R}^n$, and write $x(t,x_0)$ for its solution at time t, given the initial condition $x(0) = x_0$.*

- *A vector x^* which satisfies $f(x^*) = 0$ is called an* **equilibrium point**.
- *An equilibrium point x^* is called* **stable** *if for every $\varepsilon > 0$ a $\delta > 0$ exists such that if $\|x_0 - x^*\| < \delta$ then $\|x(t,x_0) - x^*\| < \varepsilon$ for all $t \geq 0$.*
- *An equilibrium point x^* is called* **asymptotically stable** *if it is stable and, moreover, a $\delta_1 > 0$ exists such that $\lim_{t\to\infty} \|x(t,x_0) - x^*\| = 0$, provided that $\|x_0 - x^*\| < \delta_1$.*
- *An equilibrium point x^* is* **unstable** *if it is not stable.*

In this definition $\|\cdot\|$ is an arbitrary norm; usually the Euclidean norm is used. Intuitively, stability means that the solution remains in a neighborhood of the equilibrium point and asymptotic stability means that in addition the solution converges to the equilibrium point, provided the initial point is sufficiently close to this equilibrium point. Instability means that, no matter how close starting to the equilibrium point, there always exists at least one solution that 'diverges' away from this equilibrium point.

In the definition of asymptotic stability, both requirements make sense. Indeed, there do exist examples (though not straightforward) of differential equations for which x^* is an unstable equilibrium point, while a $\delta_1 > 0$ exists such that $\lim_{t\to\infty} \|x(t,x_0) - x^*\| = 0$ for $\|x_0 - x^*\| < \delta_1$. In these examples, although there is convergence to the equilibrium point x^*, this convergence is at the expense of large deviations from x^*.

For the linear differential equation $\dot{x} = Ax$, we will take as equilibrium point the origin $x^* = 0$ (though there will be others if $\det A = 0$). We will call the linear differential equation $\dot{x} = Ax$, or even the $n \times n$ matrix A, asymptotically stable, stable or unstable, if

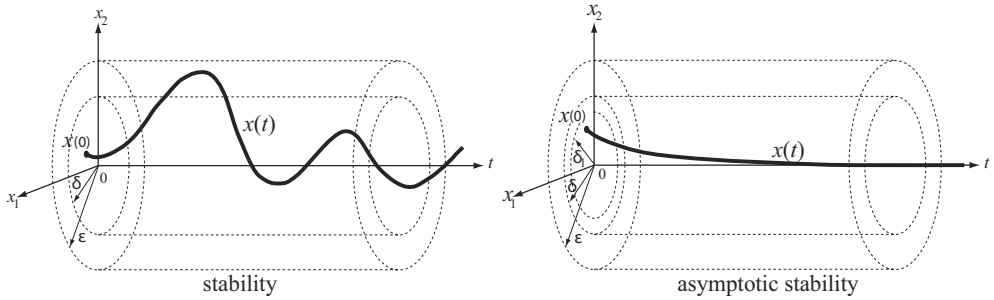

Figure 4.1 Stability and asymptotic stability.

the origin $x^* = 0$, seen as an equilibrium point, is asymptotically stable, stable or unstable, respectively.

Note that the notions of asymptotic stability, stability and instability do not depend on the chosen basis. Hence, if a differential equation is asymptotically stable with respect to one basis, it is asymptotically stable with respect to any basis, and similarly for stability and instability. Therefore, to investigate stability issues, one possibility is to move to a basis on which the description is as simple as possible. This is in fact done in the proof of the next theorem.

For the next theorem, recall the notions of algebraic and geometric multiplicity of an eigenvalue from page 33. Also recall that $\mathrm{Re}\,\lambda$ denotes the real part of the complex number λ.

Theorem 4.2 *Consider the differential equation $\dot{x} = Ax$, with A an $n \times n$ matrix having k distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, implying that $k \leq n$. The origin $x^* = 0$ is*

- *Asymptotically stable if and only if $\mathrm{Re}\,\lambda_i < 0$ for all $i = 1, \ldots, k$.*
- *Stable if and only if $\mathrm{Re}\,\lambda_i \leq 0$ for all $i = 1, \ldots, k$, and for each eigenvalue λ_i on the imaginary axis, i.e., with $\mathrm{Re}\,\lambda_i = 0$, the algebraic multiplicity and the geometric multiplicity are the same.*
- *Unstable if and only if $\mathrm{Re}\,\lambda_i > 0$ for some $i = 1, \ldots, k$, or there is an eigenvalue λ_i on the imaginary axis for which the algebraic multiplicity is larger than the geometric multiplicity.*

Proof In the proof use is made of the formula

$$e^{At} = Te^{Jt}T^{-1}, \tag{4.1}$$

where J is Jordan form of A. It is easily verified that if all eigenvalues have real parts less than zero, all the elements of e^{Jt} converge to zero for $t \to \infty$. Therefore, in that situation, also the elements of e^{At} approach zero and subsequently the solution $x(t) = e^{At}x_0$ also approaches zero. If some eigenvalues have real part zero, the situation is slightly more subtle. The subblocks J_{ij} in J with $\mathrm{Re}\,\lambda_i < 0$ still do not cause any problem (since $e^{J_{ij}t} \to 0$

as $t \to \infty$), but the subblocks with $\operatorname{Re}\lambda_i = 0$ may disturb stability. In the latter case, in the matrix

$$e^{J_{ij}t} = e^{\lambda_i t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \\ 0 & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{t^2}{2!} \\ & & & \ddots & t \\ 0 & & & 0 & 1 \end{pmatrix}, \tag{4.2}$$

the factor $e^{\lambda_i t}$ remains bounded (but does not approach zero, because with $\operatorname{Re}\lambda_i = 0$ it follows that $|e^{\lambda_i t}| = 1$), whereas the elements in the matrix do not all remain bounded, due to entries like $t, \frac{1}{2!}t^2$, etc, which appear when the size of J_{ij} is 'greater' than 1×1. In that case initial conditions do exist such that the resulting solution becomes unbounded. Therefore, if the size of some J_{ij} corresponding to an eigenvalue on the imaginary axis is 'greater' than 1×1, there is no stability. If the size of all subblocks J_{ij} corresponding to eigenvalues with real part zero is 1×1, then stability is guaranteed. The condition given in the second statement of the theorem exactly expresses the fact that all such subblocks have size 1×1. □

Example 4.3 Consider the matrices in Exercise 3.5.10. The first one is stable, the fifth one is asymptotically stable, the others are unstable. □

Example 4.4 The results of Theorem 4.2 do not hold for time-varying systems as is shown by the solution of the next differential equation

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4a & -3ae^{8at} \\ ae^{-8at} & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where a is a real parameter. The eigenvalues of the 'system' matrix are $\lambda_1 = a$ and $\lambda_2 = 3a$ (they happen to be constants, i.e., they do not depend on time) and hence for $a < 0$ both eigenvalues have real parts less than zero. However, the exact solution, with initial condition $x_1(0) = x_{10}$ and $x_2(0) = x_{20}$, is given by

$$\begin{aligned} x_1(t) &= \frac{3}{2}(x_{10}+x_{20})e^{5at} - \frac{1}{2}(x_{10}+3x_{20})e^{7at}, \\ x_2(t) &= \frac{1}{2}(x_{10}+3x_{20})e^{-at} - \frac{1}{2}(x_{10}+x_{20})e^{-3at}, \end{aligned}$$

which is unstable for any nonzero a. Indeed, take, for instance, $x_{10} = 1, x_{20} = -1$. Then $x_1(t) = e^{7at}$ and $x_2(t) = -e^{-at}$. Hence, there exists an initial condition for which the solution diverges away from the origin, being the only equilibrium point of the differential equation. This conclusion holds for both $a > 0$ and $a < 0$. If $a = 0$ the 'system' matrix equals the zero matrix and any point is equilibrium point and is also stable. □

Definition 4.5 *Consider the n dimensional system $\dot{x} = Ax$. The* **stable subspace** *for this system is the (real) subspace of the direct sum of those linear subspaces $\mathcal{N}_i$ (see Theorem 3.8) that correspond to eigenvalues of A in the open left half-plane (i.e., eigenvalues with real parts less than zero). The* **unstable subspace** *is defined similarly, then corresponding to eigenvalues with nonnegative real parts.*

Note that it follows from the above definition that the state space $\mathbb{R}^n$ is the direct sum of the stable subspace and the unstable subspace.

4.1.2 Routh's criterion

The eigenvalues of A are the zeros of its characteristic polynomial. This polynomial will be denoted here by $\det(sI-A) = s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0$. By means of a criterion, known as **Routh's criterion**, the asymptotic stability of A can be checked directly by considering the coefficients $p_i, i = 0,1,\ldots,n-1$, without calculating the zeros of the polynomial explicitly. In terms of numbers of numerical operations, calculation of the location of the eigenvalues is much more expensive than Routh's criterion, which only checks whether the eigenvalues lie in the open left half-plane (and does *not* calculate the precise location of the eigenvalues).

For a polynomial $a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$, with $a_n \neq 0$, the criterion works as follows (no proof is given here; a simple proof can be found in [Meinsma, 1995]). First arrange the coefficients $a_i, i = 0,1,\ldots,n$, into two rows in the following way.

$$\begin{array}{cccc} a_n & a_{n-2} & a_{n-4} & \cdots \\ a_{n-1} & a_{n-3} & a_{n-5} & \cdots \end{array}$$

where, if needed, a_{-1} is defined to be zero. Next, compute subsequent rows to obtain the following table

$$\begin{array}{cccc} a_n & a_{n-2} & a_{n-4} & \cdots \\ a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ \hline b_{n-2} & b_{n-4} & b_{n-6} & \cdots \\ c_{n-3} & c_{n-5} & c_{n-7} & \cdots \\ d_{n-4} & d_{n-6} & d_{n-8} & \cdots \\ \vdots & \vdots & \vdots & \end{array}$$

where the coefficients b_i, c_i, d_i, etc., are defined as follows

$$\begin{array}{llll} b_{n-2} = \dfrac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}, & b_{n-4} = \dfrac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}}, & \cdots \\ c_{n-3} = \dfrac{b_{n-2}a_{n-3} - a_{n-1}b_{n-4}}{b_{n-2}}, & c_{n-5} = \dfrac{b_{n-2}a_{n-5} - a_{n-1}b_{n-6}}{b_{n-2}}, & \cdots \\ d_{n-4} = \dfrac{c_{n-3}b_{n-4} - b_{n-2}c_{n-5}}{c_{n-3}}, & d_{n-6} = \dfrac{c_{n-3}b_{n-6} - b_{n-2}c_{n-7}}{c_{n-3}}, & \cdots \\ \vdots & \vdots & \end{array}$$

Like for a_{-1}, if in the above computations coefficients show up that have a negative index, then these coefficients are defined to have a zero value.

Clearly, the computation of a next row breaks down if the first element of the lastly computed row is a zero. Therefore, the scheme is just continued until a zero in the first column has been encountered. It can be shown easily that this certainly will happen when the $(n+2)$-nd row is to be computed. However, observe that this may also happen earlier, possibly even in the second row, when $a_{n-1} = 0$. Note that the first row always starts

with a nonzero, because it is assumed that $a_n \neq 0$. The so-called **Routh table** is the table consisting of all the rows obtained/computed in the way as described above. Hence, the Routh table consists of at most $n+1$ rows, and the first column of the Routh table only contains nonzero elements!

Routh's criterion
The roots of the polynomial $a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0$, with $a_n \neq 0$, all have a negative real part if and only if the Routh table consists of $n+1$ rows and all the elements in the first column of the table have the same sign, i.e., all elements of this column are either positive or negative.

Example 4.6

1. Consider the polynomial $p(s) = a_3 s^3 + a_2 s^2 + a_1 s + a_0$. Then $n = 3$ and the associated Routh table looks like

$$\begin{array}{cc} a_3 & a_1 \\ a_2 & a_0 \\ b_1 & \\ c_1 & \end{array}$$

with $b_1 = \dfrac{a_2 a_1 - a_3 a_0}{a_2}$ and $c_1 = \dfrac{b_1 a_0}{b_1} = a_0$. According to Routh's criterion the roots of $p(s)$ all have negative real part if and only if a_3, a_2, b_1 and c_1 are nonzero and have the same sign.

2. Consider the polynomial $p(s) = s^5 + 3s^4 + 5s^3 + 5s^2 + 4s + 2$. The polynomial can be factorized as $(s^2+1)(s+1)(s^2+2s+2)$, showing that it has roots located at $s = \pm i, s = -1$ and $s = -1 \pm i$. Hence, not all roots of the polynomial have a negative real part. This also follows from the associated Routh table. This table looks like

$$\begin{array}{ccc} 1 & 5 & 4 \\ 3 & 5 & 2 \\ \frac{10}{3} & \frac{10}{3} & \\ 2 & 2 & \end{array}$$

Since the degree of $p(s)$ is 5, i.e., $n = 5$, it follows that the Routh table cannot be developed far enough. Hence, the associated polynomial does not have all its roots in the open left half-plane.

3. Consider the polynomial $p(s) = s^4 + s^3 + s^2 - 3s + 4$. It follows from numerical methods that $p(s)$ has roots at $s = +0.758 \pm 0.701i$. Hence, not all roots of the polynomial lie in the open left half-plane. Again, the associated Routh table can be used to show this. This table looks like

$$\begin{array}{ccc} 1 & 1 & 4 \\ 1 & -3 & \\ 4 & 4 & \\ -4 & & \\ 4 & & \end{array}$$

Although the full table can be computed not all elements in the first column have the same sign. Hence, the associated polynomial does not have all its roots in the open left half-plane.

4.1.3 Lyapunov stability

Determining whether or not all solutions of a particular linear differential equation (time-invariant or time-dependent) remain bounded, or go to zero as t tends to infinity, can be quite difficult. It is possible to derive some useful sufficiency conditions which, if satisfied, guarantee that all solutions will be bounded, or even approach zero. To this end, we will introduce certain scalar functions of x and t, and study their evolution in time. The basic idea has its roots in classical mechanics, where stability criteria involving the scalar notion of *energy* are quite useful. A mechanical system can defined to be stable if its energy remains bounded. Lyapunov developed this idea and, consequently, the corresponding theory bears his name.

Let us concentrate here on time-invariant linear differential equations of the form $\dot{x} = Ax$. The scalar function $V(x(t))$, defined as $x^\top(t)Px(t)$, for some positive-definite matrix P, will be regarded as a 'generalized' energy associated with the system. Recall that a square matrix P is called positive-definite if it is symmetric and if $a^\top Pa > 0$ for all $a \neq 0$. Furthermore, note that because of the time-invariance the function $V(x(t))$ does not need to depend on t explicitly. In a system which is asymptotically stable the energy should strictly decay with time, i.e., the next derivative should be negative

$$\frac{d}{dt}V(x(t)) = \dot{x}^\top(t)Px(t) + x^\top P\dot{x}(t) = x^\top(t)\left(PA + A^\top P\right)x(t),$$

and hence, if $Q \stackrel{\text{def}}{=} -(PA + A^\top P)$ is positive-definite, the energy does decrease with time as long as $x(t) \neq 0$. Indeed, it will be shown below that if $Q > 0$, then $\lim_{t\to\infty} V(x(t)) = 0$.

Above, the starting point was a positive-definite matrix P such that hopefully the matrix Q, defined as $-PA - A^\top P$, is also positive definite. This order is in contrast with the next theorem, where the starting point is a positive-definite matrix Q and a positive-definite matrix P is looked for that satisfies $A^\top P + PA = -Q$.

Theorem 4.7 *All eigenvalues of the matrix A have negative real part if and only if for any given positive-definite matrix Q there exists a positive-definite matrix P that satisfies*

$$A^\top P + PA = -Q. \tag{4.3}$$

Proof *Sufficiency.* From the existence of the matrix P we will prove that all eigenvalues of A have negative real parts. Suppose that a matrix $P > 0$ exists such that (4.3) holds and let $Ax = \lambda x$ for some non-zero vector x. Multiplication of equation (4.3) with $\bar{x}^T$ on the left and x on the right yields

$$\begin{aligned} \bar{x}^T A^T Px + \bar{x}^T PAx &= -\bar{x}^T Qx \\ \bar{\lambda}\bar{x}^T Px + \lambda\bar{x}^T Px &= -\bar{x}^T Qx \\ (\bar{\lambda} + \lambda)\bar{x}^T Px &= -\bar{x}^T Qx. \end{aligned}$$

Since $\bar{x}^T Px > 0$ and $-\bar{x}^T Qx < 0$, $\bar{\lambda} + \lambda = 2\mathrm{Re}\,\lambda$ must be strictly negative.

Necessity. From the asymptotic stability of A it will be shown that (4.3) has a solution $P>0$. If A is such that $\mathrm{Re}\,\lambda_i<0$ for all eigenvalues λ_i, then it can be shown that a suitable matrix P is given by

$$P=\int_0^\infty e^{A^\top t}Qe^{At}dt.$$

Due to the asymptotic stability of A this integral will exist, and due to the fact that Q is positive-definite, it follows that P is also positive-definite. Finally, by substitution, it follows that

$$\begin{aligned}A^\top P+PA &= \int_0^\infty A^\top e^{A^\top t}Qe^{At}dt+\int_0^\infty e^{A^\top t}Qe^{At}Adt\\ &= \int_0^\infty \frac{d}{dt}(e^{A^\top t}Qe^{At})dt\\ &= \left[e^{A^\top t}Qe^{At}\right]_0^\infty=-Q.\end{aligned}$$

□

Equation (4.3) is referred to as a **Lyapunov equation**. Note that the equation is a linear equation and can, in principle, be treated by elementary techniques from linear algebra. Once a solution has been found, it can be tested for its positiveness (and uniqueness).

4.1.4 Interval stability

In this section polynomials of the form $p(s)=a_0+a_1s+\cdots+a_{n-1}s^{n-1}+s^n$ will be studied, and again the interest is whether the zeros belong to the open left half-plane. The novelty here is that the coefficients a_i are not exactly known. Indeed, it will be assumed that only the lower bounds a_i^- and upper bounds a_i^+ for each a_i are known, i.e., $a_i\in[a_i^-,a_i^+]$, $i=0,1,\ldots,n-1$. The central question is: if a_i^- and a_i^+, $i=0,1,\ldots,n-1$ are known, and arbitrary coefficients a_i subject to $a_i\in[a_i^-,a_i^+]$ are chosen, what can be said about the location of the roots of $p(s)$? What conditions should be imposed on a_i^- and a_i^+ such that the roots lie in the open left half-plane? These questions are related to *robustness* issues of linear systems, since quite often the exact numerical values of the coefficients a_i, $i=0,1,\ldots,n-1$, will not be known, but these values are only known approximately by means of upper and lower bounds. Sometimes an **uncertain polynomial** as above will be denoted as $p(s,a)$, where a is a vector containing the coefficients a_i of the polynomial. If $a_i\in[a_i^-,a_i^+]$, then $p(s,a)$ can be seen as an element of an **interval polynomial**, that will be denoted by $p(s,[a^-,a^+])$, where a^- and a^+ are vectors that contain the lower and upper bound of each a_i, i.e., $p(s,[a^-,a^+])=[a_0^-,a_0^+]+[a_1^-,a_1^+]s+\cdots+[a_{n-1}^-,a_{n-1}^+]s^{n-1}+s^n$.

Definition 4.8 *Associated with the interval polynomial*

$$p(s,[a^-,a^+])=[a_0^-,a_0^+]+[a_1^-,a_1^+]s+\cdots+[a_{n-1}^-,a_{n-1}^+]s^{n-1}+s^n$$

are the following four polynomials, the so-called **Kharitonov polynomials**

$$\begin{aligned}p_{--}(s) &= a_0^-+a_1^-s+a_2^+s^2+a_3^+s^3+a_4^-s^4+a_5^-s^5+a_6^+s^6+\cdots+s^n,\\ p_{+-}(s) &= a_0^++a_1^-s+a_2^-s^2+a_3^+s^3+a_4^+s^4+a_5^-s^5+a_6^-s^6+\cdots+s^n,\\ p_{-+}(s) &= a_0^-+a_1^+s+a_2^+s^2+a_3^-s^3+a_4^-s^4+a_5^+s^5+a_6^+s^6+\cdots+s^n,\\ p_{++}(s) &= a_0^++a_1^+s+a_2^-s^2+a_3^-s^3+a_4^+s^4+a_5^+s^5+a_6^-s^6+\cdots+s^n.\end{aligned}$$

It will turn out that the four Kharitonov polynomials play a crucial role in the stability of $p(s,a)$, with the vector a arbitrary, but subject to $a_i \in [a_i^-, a_i^+]$, $i = 0,1,\ldots,n-1$. Indeed, the following can be shown (a simple proof can be found in [Minnichelli, Anagnost and Desoer, 1989])

Theorem 4.9 *Let $p(s,[a^-,a^+])$ be an interval polynomial as described above. Then for any vector a with $a_i \in [a_i^-, a_i^+]$, $i = 0,1,\ldots,n-1$, the polynomial $p(s,a)$ has all its zeros in the open left half-plane if and only if the four Kharitonov polynomials have all their zeros in the open left half-plane.*

Example 4.10 Suppose that the next interval polynomial is given

$$p(s,[a^-,a^+]) = [15,19] + [20,24]s + [2,3]s^2 + s^3.$$

Then the four Kharitonov polynomials are

$$\begin{aligned} p_{--}(s) &= 15 + 20s + 3s^2 + s^3, \\ p_{+-}(s) &= 19 + 20s + 2s^2 + s^3, \\ p_{-+}(s) &= 15 + 24s + 3s^2 + s^3, \\ p_{++}(s) &= 19 + 24s + 2s^2 + s^3. \end{aligned}$$

To study the stability of these four polynomials, we can for instance use Routh's criterion. If we do so, it turns out that these four polynomials are indeed stable, i.e., they have all their zeros in the open left half-plane. From Theorem 4.9 it follows that $p(s,a)$ has all its zeros in the open left half-plane for all vectors $a = (a_0,a_1,a_2)$ with $a_0 \in [15,19], a_1 \in [20,24], a_2 \in [2,3]$. □

4.1.5 Input-output stability

This type of stability refers to the effects of input functions. It centers around the idea that every bounded input should produce a bounded output provided that the underlying system can be regarded stable. Such a stability is called input-output stability. An input function u is called bounded if a constant c exists such that $||u(t)|| \leq c$ for all t. One has a similar definition for the bounded-ness of the output function y. Let us give the formal definition.

Definition 4.11 *The system*

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t), \\ y(t) &= C(t)x(t) + D(t)u(t), \end{aligned}$$

is **BIBO stable** *(BIBO stands for bounded input, bounded output) if for all t_0, with zero initial conditions at $t = t_0$, every bounded input defined on $[t_0,\infty)$ gives rise to a bounded output on $[t_0,\infty)$. The system is called* **uniformly BIBO stable** *if there exists a constant k such that for all t_0, if $x(t_0) = 0$ and $||u(t)|| \leq 1$ for all $t \geq t_0$, then $||y(t)|| \leq k$ for all $t \geq t_0$. Then k clearly is independent of x_0.*

BIBO stability is often also referred to as external stability, in contrast to asymptotic stability of $\dot{x}(t) = A(t)x(t)$, which is often referred to as internal stability. For time-invariant systems, i.e., linear systems with constant matrices A, B, C and D, it can be shown that a system is BIBO stable if and only if $\int_0^\infty ||G(t)||dt < \infty$, where $G(t) = Ce^{At}B$, i.e., the impulse response of the system apart from the additional term $D\delta(t)$, and $||\cdot||$ denotes some appropriate matrix norm. Note that the matrix D does not play a role, because its contribution cannot result in an unbounded output when the input is bounded. Further, it can be shown that if a system is internally stable then it is also externally stable. Without additional requirements, the converse need not to be true; see Exercise 4.5.16. Actually, for the converse to be true, the concepts introduced in the following two sections play an important role.

Other types of input-output stability exist, for instance, related to the requirement that input and output functions must be L_2-functions (functions which are measurable and square-integrable), but we will not continue further in these directions.

4.2 Controllability

Controllability is a fundamental concept in mathematical system theory, as is the concept of observability. Controllability will be treated in this section and observability will be introduced in the next section. The two concepts play an essential role in the design and control of systems, as will become clear in the sequel. We will confine ourselves to linear, time-invariant differential systems, as introduced in Chapter 3. Consider therefore

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx + Du, \end{aligned} \tag{4.4}$$

with $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$. The constant matrices have appropriate sizes. The space of admissible input functions will be the class of piecewise continuous (vector-) functions. This space will occasionally be denoted as $\underline{U}$. The solution of (4.4) at time t, for the initial condition $x(0) = x_0$ and input function u, will be written as $x(t, x_0, u)$ and the corresponding output as $y(t, x_0, u)$. Then, it can be shown that (see Exercise 4.5.15)

$$\begin{aligned} x(t, x_0, u) &= e^{At}x_0 + \int_0^t e^{A(t-s)}Bu(s)ds, \\ y(t, x_0, u) &= Ce^{At}x_0 + \int_0^t Ce^{A(t-s)}Bu(s)ds + Du(t). \end{aligned} \tag{4.5}$$

The system (4.4) will sometimes be referred to as the system (A, B, C, D), for the sake of brevity.

Definition 4.12 *The system (A, B, C, D) is called* **controllable** *if for any two states $x_0, x_1 \in \mathbb{R}^n$, a finite time $t_1 > 0$ and an admissible input function u exist such that $x(t_1, x_0, u) = x_1$.*

Hence, a system is controllable if an arbitrary state $x_1 \in \mathbb{R}^n$ can be reached starting from an arbitrary state $x_0 \in \mathbb{R}^n$, in finite time t_1, by means of the application of a suitable admissible input function u. Sometimes controllability is only defined for final states x_1

being equal to the origin. In that case, it would be more appropriate to talk about **null-controllability**. The 'reverse' concept of null-controllability, i.e., being able to reach an arbitrary state starting from the origin is called **reachability**. For differential systems (A,B,C,D), the two additional controllability concepts are equivalent to controllability as defined in Definition 4.12; see Exercise 4.5.28. Hence, if a differential system is reachable, then it is also controllable and null-controllable, etc. As all three controllability concepts are equivalent, we will stick to Definition 4.12. The previous equivalence does not hold for discrete-time systems; see Chapter 7. (The essence is that the transition matrix for discrete-time systems does not necessarily have full rank, consequently yielding that null-controllability is easier fulfilled than 'full' controllability.)

Controllability will be characterized in terms of the matrices A and B. From the expression for $x(t,x_0,u)$ in (4.5) it is clear that C and D do not play any role. Define

$$R = \begin{pmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{pmatrix} \tag{4.6}$$

which is an $n \times nm$ matrix consisting of n blocks A^jB, $j = 0,1,\ldots,n-1$, and which is called the **controllability matrix**. The image of R, denoted as $\mathrm{im}\,R$ (see page 33 for this notation), is called the **controllable subspace**. This name will become clear later on.

The next lemma is useful in the development of conditions for the controllability of a system.

Lemma 4.13 $\mathrm{im}\,A^kB \subset \mathrm{im}\,R$ *for all* $k \geq 0$.

Proof The assertion for $k = 0,1,\ldots,n-1$ follows from the definition of R. If

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + p_{n-1}\lambda^{n-1} + \cdots + p_0,$$

is the characteristic polynomial of A, then the **theorem of Cayley-Hamilton**, which is well known in matrix theory, states that $p(A) = 0$ (see also Exercise 3.5.19), resulting in

$$A^n = -p_{n-1}A^{n-1} - p_{n-2}A^{n-2} - \cdots - p_1A - p_0I. \tag{4.7}$$

Hence, A^n is a linear combination of A^j with scalar weights $-p_j$, $j = 0,1,\ldots,n-1$. Multiplying (4.7) by A and substitution of A^n as in (4.7) leads to

$$\begin{aligned} A^{n+1} &= -p_{n-1}\left(-p_{n-1}A^{n-1} - p_{n-2}A^{n-2} - \cdots - p_1A - p_0I\right) \\ &\quad -p_{n-2}A^{n-1} - \cdots - p_1A^2 - p_0A \end{aligned}$$

Therefore, A^{n+1} also is a linear combination of A^j with scalar weights, $j = 0,1,\ldots,n-1$. With induction it can be shown that A^k, for all $k \geq n$, can be written as such a combination. Consequently, A^kB, for all $k \geq n$, can be written as a linear combination of $B,AB,\ldots,A^{n-1}B$. Hence, for all $k \geq n$, the columns of A^kB can be written as a linear combination of the columns of $B,AB,\cdots,A^{n-1}B$. So, also for all $k \geq n$, it follows that $\mathrm{im}\,A^kB \subset \mathrm{im}\,R$. □

Theorem 4.14 *The following statements are equivalent.*

1. *The system* (A,B,C,D) *is controllable.*

2. *R has rank n.*

3. $\operatorname{im} R = \mathbb{R}^n$.

Proof The equivalence of statements 2 and 3 is well known from linear algebra. We continue with the proof of the implication $1 \Rightarrow 2$. Assuming that rank $R < n$, we will show that the system (A,B,C,D) is not controllable. For each admissible input function $u(t), 0 \le t \le t_1$, with $t_1 > 0$, we have that

$$\begin{aligned} x(t_1,0,u) &= \int_0^{t_1} e^{A(t_1-s)}Bu(s)ds \\ &= \int_0^{t_1} \left(I + A(t_1-s) + \frac{A^2}{2!}(t_1-s)^2 + \cdots \right) Bu(s)ds \\ &= B\int_0^{t_1} u(s)ds + AB\int_0^{t_1}(t_1-s)u(s)ds + \\ &\qquad A^2B\int_0^{t_1}\frac{(t_1-s)^2}{2!}u(s)ds + \cdots \end{aligned}$$

Note that all integrals in the last expression are constant (weight) vectors. Hence, the above expression shows that $x(t_1,0,u)$ is a linear combination of the columns of the matrices $B, AB, A^2B, \cdots$. According to Lemma 4.13, it follows that $x(t_1,0,u) \in \operatorname{im} R$ for each input function u. If rank $R < n$, then $\operatorname{im} R \neq \mathbb{R}^n$ (there are points which cannot be reached from $x_0 = 0$) and an n-vector $a \neq 0$ exists such that $a^\top R = 0$. Therefore, $a^\top x(t_1,0,u) = 0$ for all admissible input functions u, which means that the system cannot be steered in the direction of a. Any reached state is always perpendicular to a if the system started at the origin. Hence, the system is not controllable.

Now we will prove the implication $2 \Rightarrow 1$. Suppose rank $R = n$. First, it will be shown that starting from $x_0 = 0$ each point $x_1 \in \mathbb{R}^n$ can be reached in an arbitrarily short time $t_1 > 0$. Later, the case of an arbitrary initial point x_0 will be considered.

For an arbitrary time $t_1 > 0$, define the symmetric $n \times n$ matrix K as

$$K = \int_0^{t_1} e^{-As}BB^\top e^{-A^\top s}ds. \tag{4.8}$$

It will be shown in Lemma 4.15 that matrix K is invertible. Now take an arbitrary $x_1 \in \mathbb{R}^n$ and define the input function

$$\overline{u}(t) = B^\top e^{-A^\top t}K^{-1}e^{-At_1}x_1.$$

If this input is applied to the system with initial condition $x_0 = 0$, then

$$\begin{aligned} x(t_1,0,\overline{u}) &= \int_0^{t_1} e^{A(t_1-s)}BB^\top e^{-A^\top s}K^{-1}e^{-At_1}x_1 ds \\ &= e^{At_1}\left(\int_0^{t_1} e^{-As}BB^\top e^{-A^\top s}ds\right)K^{-1}e^{-At_1}x_1 \\ &= e^{At_1}KK^{-1}e^{-At_1}x_1 \\ &= x_1. \end{aligned}$$

Lastly, if x_0 is arbitrary, the input function $\tilde{u}$ will be constructed as follows. Consider the state $x_1 - e^{At_1}x_0 \in \mathbb{R}^n$. According to the previous part of this proof a control $\tilde{u}$ exists, which steers the system from the origin to $x_1 - e^{At_1}x_0 \in \mathbb{R}^n$, i.e.,

$$x(t_1, 0, \tilde{u}) = \int_0^{t_1} e^{A(t_1-s)}B\tilde{u}(s)ds = x_1 - e^{At_1}x_0.$$

Hence, for this input function $\tilde{u}$ it follows that

$$x_1 = e^{At_1}x_0 + \int_0^{t_1} e^{A(t_1-s)}B\tilde{u}(s)ds.$$

So, for time $t_1 > 0$ and arbitrary chosen states x_0, x_1, a control input $\tilde{u}$ has been found that steers the system from x_0 at time $t = 0$ to x_1 at time $t = t_1$, which clearly implies that the system is controllable. □

Lemma 4.15 *Assume that A and B are such that the matrix R defined in (4.6) has full (row) rank. Then the matrix K as defined in (4.8) is invertible.*

Proof Suppose that matrix K is not invertible. Then $Ka = 0$ for an n-vector $a \neq 0$, and hence also $a^\top Ka = 0$, or equivalently

$$\int_0^{t_1} a^\top e^{-As}BB^\top e^{-A^\top s}a\, ds = 0 \quad \Leftrightarrow$$
$$\int_0^{t_1} \|a^\top e^{-As}B\|^2 ds = 0 \quad \Leftrightarrow \quad a^\top e^{-As}B = 0 \qquad \forall\, s \in [0, t_1].$$

The last equivalence follows because $a^\top e^{-As}B$ is a continuous function of s. In fact, it is easy to see that the function $a^\top e^{-As}B$ can be differentiated with respect to s infinitely many times. Differentiating the function $(n-1)$ times, and subsequently substituting $s = 0$, gives

$$\begin{array}{ccc} a^\top e^{-As}B = 0 & \longrightarrow & a^\top B = 0, \\ a^\top Ae^{-As}B = 0 & \longrightarrow & a^\top AB = 0, \\ \vdots & \vdots & \vdots \\ a^\top A^{n-1}e^{-As}B = 0 & \longrightarrow & a^\top A^{n-1}B = 0. \end{array}$$

This gives that $a^\top R = 0$ with $a \neq 0$, which is impossible because rank $R = n$. Therefore, K must be invertible. □

Controllability of a system is determined by the matrices A and B, as Theorem 4.14 tells us. Therefore, we will also speak of the controllability of the pair (A, B). The condition rank $R = n$ is called the **rank condition** for controllability. In case $m = 1$, i.e., the input is a scalar, the matrix R is a square $n \times n$ matrix and controllability is equivalent to $\det R \neq 0$. Please note that Theorem 4.14 does not say anything about $t_1 > 0$. It just follows that the final point can be reached in arbitrarily short time if it can be reached at all (of course, for smaller t_1, the norm of the input function will increase). See also Exercises 4.5.26 and 4.5.27.

Example 4.16 Consider the satellite dynamics of (3.43) and take $\omega = 1$. The controllability matrix is

$$R = \left(\begin{array}{cc|cc|cc|cc} 0 & 0 & 1 & 0 & 0 & 2 & -1 & 0 \\ 1 & 0 & 0 & 2 & -1 & 0 & 0 & -2 \\ 0 & 0 & 0 & 1 & -2 & 0 & 0 & -4 \\ 0 & 1 & -2 & 0 & 0 & -4 & 2 & 0 \end{array}\right),$$

and rank $R = 4$ by inspection. Hence, the satellite system is controllable. Suppose now that $u_1 = 0$ and the controllability-question is asked with respect to u_2 only. Denote the related controllability matrix by R_2. Then

$$R_2 = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & -2 \\ 0 & 1 & 0 & -4 \\ 1 & 0 & -4 & 0 \end{pmatrix}.$$

By inspection rank $R_2 = 4$ and hence u_2 on its own is able to manoeuver the satellite to arbitrary positions (from a pragmatic point of view, the initial and final point x_0 and x_1 should be chosen such that the linearized Equations (3.43) make sense for these and intermediate points). Suppose now that $u_2 = 0$ and the question is whether u_1 alone is able to take care of the controllability. For that purpose consider the related controllability matrix, denoted by R_1,

$$R_1 = \begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & -2 & 0 & 2 \end{pmatrix},$$

which has rank 3. Hence, the system with $u_2 = 0$ is not controllable! □

Example 4.17 To study a system it is often useful to try to bring the system into a standard form, i.e., an elementary form that the system can be put in by changing the basis in the state space. Below two standard forms will be presented.

(i) A system in controllability form, with $\tilde{x} \in \mathbb{R}^n, u \in \mathbb{R}$, is a system that is described by

$$\begin{aligned} \dot{\tilde{x}}(t) &= \begin{pmatrix} 0 & 0 & \dots & & \dots & -p_0 \\ 1 & \ddots & & & & -p_1 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & & 1 & 0 & -p_{n-2} \\ 0 & 0 & \dots & 0 & 1 & -p_{n-1} \end{pmatrix} \tilde{x}(t) + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \end{pmatrix} u(t), \\ y(t) &= \begin{pmatrix} q_0 & q_1 & \dots & \dots & \dots & q_{n-1} \end{pmatrix} \tilde{x}(t). \end{aligned} \tag{4.9}$$

For this system the controllability matrix equals the $n \times n$ identity matrix. Hence, controllability of the system is immediate, also explaining the name 'controllability form'.

(ii) A system in controller form, with $\bar{x} \in \mathbb{R}^n, u \in \mathbb{R}$, is a system that is described by

$$\begin{aligned} \dot{\bar{x}}(t) &= \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & & \\ 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & \cdots & -p_{n-2} & -p_{n-1} \end{pmatrix} \bar{x}(t) + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} u(t), \\ y(t) &= \begin{pmatrix} q_0 & q_1 & \cdots & \cdots & q_{n-1} \end{pmatrix} \bar{x}(t). \end{aligned} \tag{4.10}$$

For this system the controllability matrix looks like

$$\begin{pmatrix} 0 & \dots & \dots & 0 & 1 \\ \vdots & & & 1 & * \\ \vdots & & & & * \\ 0 & 1 & & & \vdots \\ 1 & * & * & \dots & * \end{pmatrix},$$

where the $*$'s represent numbers that are not relevant in the present context. Since the controllability matrix clearly has rank n, this system is always controllable, irrespective of the values of the coefficients q_i and p_i. The name 'controller form' comes from the fact that the form is very useful in the design of state feedback controllers.

It can be shown that any controllable system $\dot{x} = Ax + Bu$, $y = Cx$, with A an $n \times n$ matrix, B an $n \times 1$ matrix and C a $1 \times n$ matrix, always can be brought into 'controllability form' or 'controller form', by applying an appropriate basis transformation. In fact, see Exercise 4.5.22 for the transformation into the controllability form, and the transformation into the controller form will be treated in Lemma 5.4. □

Example 4.18 Controllability can also be studied in terms of flow diagrams. For instance, consider the flow diagram in Figure 4.2. This system is not controllable because x_1 cannot be influenced by u. Its state space description is (see also Exercise 4.5.20)

$$\dot{x} = \begin{pmatrix} a_1 & 0 \\ 1 & a_2 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u.$$

Also from its controllability matrix it follows that the system is not controllable.
The system in Figure 4.3 is controllable. Its state space description is

$$\dot{x} = \begin{pmatrix} a_1 & 0 \\ 1 & a_2 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u.$$

Clearly, its controllability matrix has rank 2, so the system is controllable. □

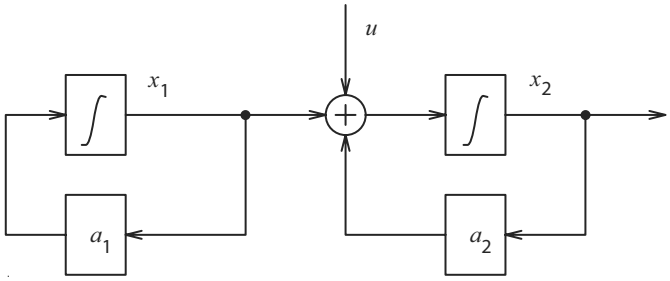

Figure 4.2 Uncontrollable system.

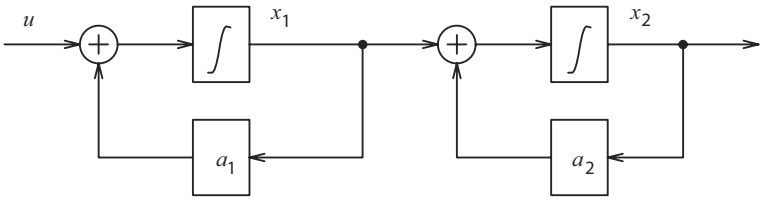

Figure 4.3 Controllable system.

If the system (A,B) with state space $\mathbb{R}^n$ is not controllable, then those points of $\mathbb{R}^n$ which are reachable (starting from the origin) are exactly all vectors of $\operatorname{im} R$. Indeed,

$$\operatorname{im} R = \{x_1 \in \mathbb{R}^n \,|\, \text{there exist } t_1 > 0 \text{ and } u \in \underline{U} \text{ such that } x_1 = x(t_1, 0, u)\},$$

where $\underline{U}$ denotes a set of admissible control functions, like, for instance, the set of piecewise continuous (vector-)functions. Because of the above characterization, $\operatorname{im} R$ is generally also referred to as the **reachable subspace**, implying that the names 'controllable subspace' and 'reachable subspace' actually refer to the same subspace.

An algebraic proof of the above characterization will not be given, instead $\operatorname{im} R$ will be interpreted in a geometric way. Therefore, first define

Definition 4.19 *A linear subspace $\mathcal{V} \subset \mathbb{R}^n$ is called* **A-invariant** *if $A\mathcal{V} \subset \mathcal{V}$, i.e., $Av \in \mathcal{V}$ for all $v \in \mathcal{V}$.*

Then the next theorem follows.

Theorem 4.20 $\operatorname{im} R$ *is the smallest linear subspace of* $\mathbb{R}^n$ *such that*

1. $\operatorname{im} B \subset \operatorname{im} R$,
2. $\operatorname{im} R$ *is A-invariant.*

Proof First it will be shown that $\operatorname{im} R$ satisfies properties 1 and 2. Clearly $\operatorname{im} B \subset \operatorname{im} R$, because $R = (B\ AB\ \cdots\ A^{n-1}B)$. Furthermore,

$$A \operatorname{im} R = A \operatorname{im} (B\ AB\ \cdots\ A^{n-1}B) = \operatorname{im} (AB\ A^2B\ \cdots\ A^nB).$$

A consequence of the Cayley-Hamilton theorem is that the columns of A^nB can be expressed as a linear combination of the columns of $B, AB, \ldots, A^{n-1}B$. Therefore,

$$A\,\mathrm{im}\,R \subset \mathrm{im}\,(B\;AB\;\cdots\;A^{n-1}B) = \mathrm{im}\,R.$$

It remains to be shown that $\mathrm{im}\,R$ is the smallest subspace that satisfies points 1 and 2. Suppose now that a linear subspace $\mathscr{V}$ is given which satisfies points 1 and 2. It will be shown that $\mathrm{im}\,R \subset \mathscr{V}$. Because $\mathrm{im}\,B \subset \mathscr{V}$ and $A\mathscr{V} \subset \mathscr{V}$, the next inclusions follow by induction.

$$\begin{aligned}
\mathrm{im}\,AB &= A(\mathrm{im}\,B) \subset A\mathscr{V} \subset \mathscr{V},\\
\mathrm{im}\,A^2B &= A(\mathrm{im}\,AB) \subset A\mathscr{V} \subset \mathscr{V},\\
&\vdots \quad \vdots \quad \vdots\\
\mathrm{im}\,A^{n-1}B &= A(\mathrm{im}\,A^{n-2}B) \subset A\mathscr{V} \subset \mathscr{V},
\end{aligned}$$

Therefore,

$$\mathrm{im}\,R = \mathrm{im}\,(B\;AB\;\cdots\;A^{n-1}B) = \mathrm{im}\,B + \mathrm{im}\,AB + \cdots + \mathrm{im}\,A^{n-1}B \subset \mathscr{V}.$$

Hence, $\mathrm{im}\,R$ is contained in every linear subspace that satisfies points 1 and 2 of the theorem statement. Therefore, in the sense of subspace inclusions, $\mathrm{im}\,R$ is the smallest linear subspace that is A-invariant and contains $\mathrm{im}\,B$. □

It is well known from linear algebra that if a linear subspace $\mathscr{V}$ is A-invariant, it can be used to transform matrix A into a block upper triangular form. To obtain this, assume that $\mathscr{V}$ is an A-invariant subspace in $\mathbb{R}^n$. To rule out some trivial cases, assume that $0 < k < n$, where $k = \dim \mathscr{V}$. Then there exist k vectors $q_1, \ldots, q_k$ in $\mathbb{R}^n$ that form a basis for $\mathscr{V}$. This basis can be extended by $n-k$ additional vectors $q_{k+1}, \ldots, q_n$ in $\mathbb{R}^n$ to form a basis for $\mathbb{R}^n$. Now write $T = (q_1, \ldots, q_k, q_{k+1}, \ldots, q_n)$, then T is an $n \times n$ matrix that is invertible. Next consider the matrix $\tilde{A}$ defined by $\tilde{A} = T^{-1}AT$, or $AT = T\tilde{A}$. Then it easily follows that

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix},$$

with $\tilde{A}_{11}, \tilde{A}_{12}$ and $\tilde{A}_{22}$ matrices of size $k \times k, k \times (n-k)$ and $(n-k) \times (n-k)$, respectively. The 0 represents a zero matrix of size $(n-k) \times k$ and is a consequence of the A-invariance of $\mathscr{V}$.

Next, let G be an $n \times l$ matrix such that $\mathrm{im}\,G \subset \mathscr{V}$ and define $\tilde{G} = T^{-1}G$, with T as above. Then,

$$\tilde{G} = \begin{pmatrix} \tilde{G}_1 \\ 0 \end{pmatrix},$$

where $\tilde{G}_1$ is an $k \times l$ matrix.

Dually, let H be an $t \times n$ matrix such that $\mathscr{V} \subset \ker H$ (see page 33 for the meaning of $\ker H$) and define $\tilde{H} = HT$, with T as above. Then,

$$\tilde{H} = \begin{pmatrix} 0 & \tilde{H}_2 \end{pmatrix},$$

where $\tilde{H}_2$ is an $t \times (n-k)$ matrix.

Now assume that $\mathscr{V} = \mathrm{im}R$, i.e., $\mathscr{V}$ is the controllable subspace and as such $\mathscr{V}$ is the smallest A-invariant subspace that contains $\mathrm{im}B$. Again, to rule out trivial cases, assume that $0 < k < n$, where $\dim \mathrm{im}R = k$ or, equivalently, rank $R = k$. Let T be the basis transformation matrix as described before, i.e., T is made up of the vectors of a basis for $\mathscr{V}$ and the extension towards a basis for $\mathbb{R}^n$. Then

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} \tilde{B}_1 \\ 0 \end{pmatrix}, \tag{4.11}$$

with $\tilde{A}_{11}, \tilde{A}_{12}, \tilde{A}_{22}$ and $\tilde{B}_1$ matrices of size $k \times k, k \times (n-k), (n-k) \times (n-k)$ and $k \times m$, respectively. The 0's represent zero matrices of suitable sizes. It then follows that $\tilde{R} = T^{-1}R$, where R is the controllability matrix of the pair (A,B) and $\tilde{R}$ is the same for the pair $(\tilde{A},\tilde{B})$, i.e., $R = \begin{pmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{pmatrix}$ and $\tilde{R} = \begin{pmatrix} \tilde{B} & \tilde{A}\tilde{B} & \tilde{A}^2\tilde{B} & \cdots & \tilde{A}^{n-1}\tilde{B} \end{pmatrix}$. Moreover, it follows from (4.11) that

$$\tilde{R} = \begin{pmatrix} \tilde{B}_1 & \tilde{A}_{11}\tilde{B}_1 & \cdots & \tilde{A}_{11}^{n-1}\tilde{B}_1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

Because T is an invertible matrix it is clear that rank R = rank $\tilde{R} = k$, and, consequently, rank $\begin{pmatrix} \tilde{B}_1 & \tilde{A}_{11}\tilde{B}_1 & \cdots & \tilde{A}_{11}^{n-1}\tilde{B}_1 \end{pmatrix} = k$. The Cayley Hamilton theorem applied to $\tilde{A}_{11}$ yields that rank $\begin{pmatrix} \tilde{B}_1 & \tilde{A}_{11}\tilde{B}_1 & \cdots & \tilde{A}_{11}^{k-1}\tilde{B}_1 \end{pmatrix} = k$, implying that the pair $(\tilde{A}_{11}, \tilde{B}_1)$ is controllable.

From the above it follows that a system of the form

$$\begin{array}{rcl} \dot{x} &=& Ax + Bu, \\ y &=& Cx + Du, \end{array}$$

which is not controllable, by a change of basis given by $x = T\tilde{x}$ (or $\tilde{x} = T^{-1}x$), with matrix T based on $\mathscr{V} = \mathrm{im}R$ as described above, can be transformed into a system of the form

$$\left.\begin{array}{rcl} \dot{\tilde{x}}_1 &=& \tilde{A}_{11}\tilde{x}_1 + \tilde{A}_{12}\tilde{x}_2 + \tilde{B}_1 u, \\ \dot{\tilde{x}}_2 &=& \tilde{A}_{22}\tilde{x}_2, \\ y &=& \tilde{C}_1\tilde{x}_1 + \tilde{C}_2\tilde{x}_2 + Du, \end{array}\right\} \tag{4.12}$$

where the pair $(\tilde{A}_{11}, \tilde{B}_1)$ is controllable. In the above description the submatrices in (4.11) show up, together with $\tilde{C} = CT = \begin{pmatrix} \tilde{C}_1 & \tilde{C}_2 \end{pmatrix}$ and D.

Example 4.21 Consider the pair (A,B) with

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

The system corresponding to the pair was considered in Example 4.16. It represents a satellite with only one input, namely the thrust in the radial direction. According to its

controllability matrix R, in Example 4.16 denoted by R_1, the system is not controllable. The columns of R also give the next three vectors that span $\operatorname{im} R$

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -2 \\ 0 \end{pmatrix}.$$

If these vectors are identified with q_1, q_2 and q_3, respectively, then q_4 must be chosen independent of q_1, q_2 and q_3. We choose $q_4 = (2,0,0,1)^\top$. Now define

$$T = (q_1\ q_2\ q_3\ q_4) = \begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & -2 & 0 & 1 \end{pmatrix},$$

and hence

$$T^{-1} = \frac{1}{10} \begin{pmatrix} 0 & 10 & -5 & 0 \\ 2 & 0 & 0 & -4 \\ 0 & 0 & -5 & 0 \\ 4 & 0 & 0 & 2 \end{pmatrix}.$$

Calculating $T^{-1}AT$ and $T^{-1}B$ gives

$$\tilde{A} = T^{-1}AT = \left(\begin{array}{ccc|c} 0 & 0 & 0 & 7.5 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -0.5 \\ \hline 0 & 0 & 0 & 0 \end{array}\right), \quad \tilde{B} = T^{-1}B = \left(\begin{array}{c} 1 \\ 0 \\ 0 \\ \hline 0 \end{array}\right).$$

The partitioning in blocks as given by (4.11) is clearly visible. The pair $(\tilde{A}_{11}, \tilde{B}_1)$, with

$$\tilde{A}_{11} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \tilde{B}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

is controllable since

$$\operatorname{rank}(\tilde{B}_1\ \ \tilde{A}_{11}\tilde{B}_1\ \ \tilde{A}_{11}^2\tilde{B}_1) = \operatorname{rank} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 3.$$

□

Instead of checking controllability by computing the rank of the controllability matrix R, some other tests on controllability can be applied. One of these tests is described below. However, first a preliminary result is stated and proved. In the following theorem row (eigen)vectors are used instead of the more common column (eigen)vectors.

Theorem 4.22 *The pair (A,B), where A is a real $n\times n$ vector and B an $n\times m$ matrix, is not controllable if and only if a nonzero row vector q and a scalar λ exist such that*

$$qA=\lambda q,\quad qB=0. \tag{4.13}$$

In other words, (A,B) will be controllable if and only if there is no row eigenvector of A that is orthogonal to $\operatorname{im}B$.

Proof *Sufficiency.* If there exists a vector $q\neq 0$ such that $qA=\lambda q$, $qB=0$, then

$$\begin{array}{rcccl} qAB & = & \lambda qB & = & 0,\\ qA^2B & = & \lambda qAB & = & 0,\\ & & \vdots & & \\ qA^{n-1}B & = & \lambda qA^{n-2}B & = & 0, \end{array}$$

so that,

$$qR=q(B\ AB\ \cdots\ A^{n-1}B)=0,$$

which means that R has linearly dependent rows. It then follows that the rank of R is less than n, implying that (A,B) is not controllable.

Necessity. We have to show that (A,B) not controllable implies the existence of a nonzero row vector q satisfying (4.13). Denote the rank of R by k, then we have here that $k<n$. Assume that the pair (A,B) has been put in the block form of (4.11). Then it is clear that the following vector q is perpendicular to $\operatorname{im}B$

$$q\ =(\ \underset{\longleftrightarrow}{0}_{k}\ \Big|\ \underset{\longleftrightarrow}{z}_{n-k}\ ),$$

where z is an arbitrary row vector consisting of $n-k$ components. It is perhaps not hard to guess that we should choose z as a row eigenvector of A_{22}

$$zA_{22}=\lambda z,$$

because then

$$qA=(0\ z)A=(0\ zA_{22})=(0\ \lambda z)=\lambda(0\ z)=\lambda q.$$

Therefore, we have shown how to find a row vector q satisfying (4.13) and this completes the proof. □

With the use of Theorem 4.22 the next important result can be proved.

Theorem 4.23 *Consider the pair (A,B), where A is an $n\times n$ matrix and B an $n\times m$ matrix. Then the following statements are equivalent.*

1. *The pair (A,B) is controllable.*
2. $\operatorname{rank}(sI-A\ ,\ B)=n$ *for all* $s\in\mathbb{C}$.

3. $\operatorname{rank}(\lambda I - A \;,\; B) = n$ *for all eigenvalues* λ *of matrix* A.

Proof The equivalence of statements 2. and 3. follows straightforwardly. Namely, if the rows of $(sI - A \quad B)$ are linearly dependent, so are the rows of $sI - A$, implying that s is an eigenvalue of A. In that case, there is an eigenvalue λ of A such that the rows of $(\lambda I - A \quad B)$ are linearly dependent, just take $\lambda = s$. In terms of ranks it means that if rank $(sI - A \quad B) < n$ for some $s \in \mathbb{C}$, then rank $(\lambda I - A \quad B) < n$ for some eigenvalue λ of matrix A. Since the converse immediately follows, the equivalence of statements 2. and 3. has been proved.

If $(sI - A \quad B)$ has rank n for all $s \in \mathbb{C}$, then for any $s \in \mathbb{C}$ there doesn't exist a nonzero row vector v such that $v(sI - A \;,\; B) = 0$, i.e., such that $vA = sv$ and $vB = 0$. But then, by Theorem 4.22, the pair (A, B) must be controllable. So, statement 2. implies statement 1.

Assume that rank $(\lambda I - A \quad B) < n$ for some eigenvalue λ of A. Then there exists a nonzero row vector q such that $q(\lambda I - A \quad B) = 0$. Hence, $qA = q\lambda$ and $qB = 0$. By Theorem 4.22 it follows that the pair (A, B) is not controllable. Hence, statement 1. implies statement 3. □

Theorem 4.23 is known as the **Hautus test** for controllability. It provides a test for controllability which requires the computation of the eigenvalues of matrix A and (at most) n rank tests of matrices of size $n \times (n+m)$. Testing the controllability by means of the controllability matrix requires the construction of this matrix and the test of its rank as a matrix of size $n \times nm$. Clearly, depending on n, m, and the fact whether or not the eigenvalues of A are available, and partially coincide, the test resulting from Theorem 4.23 may be more efficient than the test via the controllability matrix. This is also true if the computation of powers of the matrix A is numerically troublesome. This is, for instance, the case when the eigenvalues of A are not all of the same order.

Example 4.24 The starting point here is Equation (2.7) of the heated bar

$$\frac{\partial T(t,r)}{\partial t} = c\frac{\partial^2 T(t,r)}{\partial r^2}, \tag{4.14}$$

with $0 \le r \le L$ and $t \ge 0$, where we will assume now that $c = 1$ and $L = 1$. In this example the temperature can be controlled at both ends of the bar, i.e., at $r = 0$ by means of u_1, and

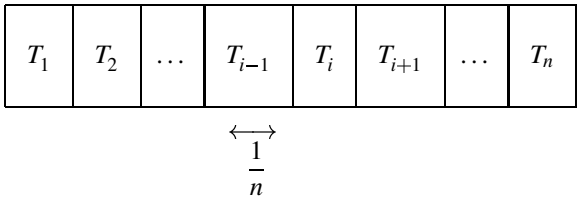

Figure 4.4 Discretization of heated bar.

at $r = 1$ by u_2. We are going to discretize the interval of the location parameter r into n discrete subintervals, each of length $1/n$, see Figure 4.4. The temperature on the interval $(i-1)/n \le r < i/n$ is assumed to depend on t only and is indicated by T_i, $i = 1, 2, \ldots, n$.

If we use standard approximations for the second order derivative with respect to r with step size $\frac{1}{n}$, we obtain the following model, consisting of ordinary differential equations

$$\begin{aligned}
\frac{dT_1}{dt} &= n^2(u_1 - 2T_1 + T_2),\\
&\vdots\\
\frac{dT_i}{dt} &= n^2(T_{i-1} - 2T_i + T_{i+1}), \quad i = 2,3,\ldots,n-1,\\
&\vdots\\
\frac{dT_n}{dt} &= n^2(T_{n-1} - 2T_n + u_2).
\end{aligned}$$

Then we get the following finite dimensional model

$$\frac{1}{n^2}\frac{d}{dt}\begin{pmatrix} T_1 \\ \vdots \\ T_i \\ \vdots \\ T_n \end{pmatrix} = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots\ 2 & \ddots \\ 0 & \cdots & 0 & 1 & -2 \end{pmatrix}\begin{pmatrix} T_1 \\ \vdots \\ T_i \\ \vdots \\ T_n \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}. \quad (4.15)$$

By checking the controllability matrix of this finite dimensional system (see Exercise 4.5.29), it is easily verified that the system is controllable. Hence, in theory one can steer to any temperature profile. (Is controllability maintained if $n \to \infty$, or in other words, is the system (4.14) controllable? Formally, controllability for partial differential equations has not yet been defined and we will not give the definition.) Intuitively, however, it should be clear that a temperature profile with a discontinuity, for instance,

$$T(t,r) = 0 \text{ for } 0 \le r \le 1/2, \quad \text{and} \quad T(t,r) = 1 \text{ for } 1/2 < r \le 1,$$

at a certain time t, can never be achieved in practice. □

4.3 Observability

We now turn to another fundamental concept in system theory, namely observability.

Definition 4.25 *The system* (A,B,C,D) *is* **observable** *if a finite time* $t_1 > 0$ *exists such that for each admissible input function u, it follows from* $y(t,x_0,u) = y(t,x_1,u)$ *for all* $t \in [0,t_1]$, *that* $x_0 = x_1$.

A system is called observable if the initial state x_0 can be constructed from the knowledge of u and y on the interval $[0,t_1]$ for some finite $t_1 > 0$. Because u is given, once x_0 is known, the state x on the whole interval $[0,t_1]$ can be determined. In other words, the external behavior of an observable system restricted to some interval of positive length uniquely determines the state on this time interval.

As with controllability there are several definitions possible for observability. A slightly different one would be that a system is observable if for any two states x_0, x_1, with

$x_0 \neq x_1$, an admissible input u and a time $t_1 > 0$ exist such that $y(t,x_0,u)$ and $y(t,x_1,u)$ are not the same on $[0,t_1]$. The latter definition means that a control can be found such that x_0 and x_1 (ultimately) can be distinguished from each other through the output. Definition 4.25, however, assumes that x_0 and x_1 can be distinguished for any control if $x_0 \neq x_1$. It turns out that for linear systems both definitions are equivalent (no proof).

It will be shown that observability of a linear system (A,B,C,D) can be completely characterized by the matrices A and C. Define the $np \times n$ matrix W, called the **observability matrix**, as

$$W = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}.$$

Lemma 4.26 *Let the vector $x \in \mathbb{R}^n$ be such that $Cx = CAx = \cdots = CA^{n-1}x = 0$. Then $CA^k x = 0$ for all $k \geq 0$.*

Proof For $k = 0,1,\ldots,n-1$ the statement is immediate. A consequence of the Cayley-Hamilton theorem is that, for each $k \geq n$, the kth power of A is a linear combination of A^j, $j = 0,1,\ldots,n-1$, with scalar weights, see also Lemma 4.13. Therefore, for all $k \geq n$

$$CA^k = \alpha_{0,k}C + \alpha_{1,k}CA + \cdots + \alpha_{n-1,k}CA^{n-1},$$

for certain scalars $\alpha_{j,k}$, and hence it follows for all $k \geq 0$ that

$$CA^k x = \alpha_{0,k}Cx + \alpha_{1,k}CAx + \cdots + \alpha_{n-1,k}CA^{n-1}x = 0.$$

□

Theorem 4.27 *The following statements are equivalent.*

1. *System (A,B,C,D) is observable.*
2. *W has rank n.*
3. $\ker W = 0$.

Proof The equivalence of statement 2 and 3 is obvious. We continue with proof $2 \Rightarrow 1$. Let rank $W = n$. Given an arbitrary time $t_1 > 0$ and an arbitrary admissible control u, assume that $y(t,x_0,u) = y(t,x_1,u)$ for all $t \in [0,t_1]$. We will show that these assumptions will lead to $x_0 = x_1$. The equality $y(t,x_0,u) = y(t,x_1,u)$ implies

$$Ce^{At}x_0 + \int_0^t Ce^{A(t-s)}Bu(s)ds + Du(t) = Ce^{At}x_1 + \int_0^t Ce^{A(t-s)}Bu(s)ds + Du(t),$$

and hence $Ce^{At}x_0 = Ce^{At}x_1$, i.e.,

$$Ce^{At}(x_0 - x_1) = 0,$$

for all $t \in [0,t_1]$. If the latter expression is differentiated $(n-1)$ times with respect to t, and if then $t=0$ is substituted, we get subsequently

$$\left.\begin{array}{ccc} Ce^{At}(x_0-x_1)=0 & \longrightarrow & C(x_0-x_1)=0 \\ CAe^{At}(x_0-x_1)=0 & \longrightarrow & CA(x_0-x_1)=0 \\ \vdots & \vdots & \vdots \\ CA^{n-1}e^{At}(x_0-x_1)=0 & \longrightarrow & CA^{n-1}(x_0-x_1)=0 \end{array}\right\} \tag{4.16}$$

The result can be written as $W(x_0-x_1)=0$. Since W is an $np \times n$ matrix and rank $W=n$, we have that x_0-x_1 is the zero vector, implying that $x_0=x_1$.

Now we will prove the direction $1 \Rightarrow 2$. Suppose rank $W<n$. Then it will be shown that the system is not observable. Under the condition that rank $W<n$, the n columns of the matrix W are linearly dependent. Hence, vectors x_0, x_1 exist, with $x_0 \neq x_1$, such that $W(x_0-x_1)=0$. This means that

$$C(x_0-x_1)=CA(x_0-x_1)=\ldots=CA^{n-1}(x_0-x_1)=0.$$

Application of Lemma 4.26 yields $CA^k(x_0-x_1)=0$ for all $k \geq 0$. Subsequently, it follows for all t that

$$Ce^{At}(x_0-x_1)=\sum_{k=0}^{\infty}\frac{t^k}{k!}CA^k(x_0-x_1)=0, \quad \text{or} \quad Ce^{At}x_0=Ce^{At}x_1.$$

Now adding to both sides $\int_0^t Ce^{A(t-s)}Bu(s)ds+Du(t)$, where u is an admissible control, it follows that $y(t,x_0,u)=y(t,x_1,u)$ for all t and all admissible controls. Hence, the system is not observable. □

Observability is completely determined by the matrices A and C and therefore we will sometimes speak of the observability of the pair (C,A). The system (A,B,C,D) is observable if and only if the system $(A,0,C,0)$, i.e., $\dot{x}=Ax$, $y=Cx$, is observable. The condition rank $W=n$ is called the **rank condition** for observability. In case of a single-output system W is a square $n \times n$ matrix, and (C,A) being observable is then equivalent to $\det W \neq 0$.

If (C,A) is observable, the interval $[0,t_1]$ in Definition 4.25 may be chosen arbitrarily small, as long as $t_1>0$ (the differentiations in (4.16) should be possible). Furthermore, if knowledge of u and y on $[0,t_1]$ leads to unique determination of x_0, then also $x(t)$ for all $t \in [0,t_1]$ is known. Hence, given the initial conditions x_0 and the known input u, the solution of the state equation is uniquely determined.

Example 4.28 Consider again the satellite dynamics described in Exercise 3.5.2 and the output equations given in Example 3.14. Take $\omega=\sigma=1$, then

$$A=\begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \end{pmatrix}, \quad C=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The observability matrix is

$$W = \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 3 & 0 & 0 & 2 \\ 0 & -2 & 0 & 0 \\ \hline 0 & -1 & 0 & 0 \\ -6 & 0 & 0 & -4 \end{array}\right),$$

which has rank 4. Therefore, the system is observable and the state $x(t)$, $t \in [0,t_1]$, can be constructed if we are given the measurement y and the input u of the system on the interval $[0,t_1]$ with $0 < t_1$. (We have not considered the question of how the actual construction of the state should take place; we have proved, however, that it is unique. The actual construction is the subject of Section 5.2, which deals with observers). Suppose that only y_2 can be measured. The corresponding observability matrix W_2 is

$$W_2 = \left(\begin{array}{cccc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \\ -6 & 0 & 0 & -4 \end{array}\right),$$

which is nonsingular. Therefore, the state is uniquely determined if only y_2, together with u, is available over the interval $[0,t_1]$, for an arbitrary $t_1 > 0$. If only y_1 is available, then

$$W_1 = \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & -1 & 0 & 0 \end{array}\right),$$

and W_1 has rank 3. This system is non-observable! Hence, no matter how large the interval $[0,t_1]$ is taken, from output y_1 alone, together with input u, the whole state cannot fully be constructed. □

Though the properties of controllability and observability are different, the rank conditions are rather similar, which is expressed in the following theorem.

Theorem 4.29

- (A,B) *is controllable if and only if* $(B^\top, A^\top)$ *is observable.*
- (C,A) *is observable if and only if* $(A^\top, C^\top)$ *is controllable.*

Proof (A,B) controllable $\iff$ $\operatorname{rank}(B\ AB \cdots A^{n-1}B) = n \iff$

$$\operatorname{rank}(B\ AB \cdots A^{n-1}B)^\top = n \iff \operatorname{rank} \begin{pmatrix} B^\top \\ B^\top A^\top \\ \vdots \\ B^\top (A^\top)^{n-1} \end{pmatrix} = n \iff$$

$(B^\top, A^\top)$ is observable.
The proof of the second assertion is similar. □

The statements in this theorem are symbolic statements which express the fact that the properties of controllability and observability are **dual properties**. Let's have a closer look at what this actually means in terms of the systems involved.

Consider the system $\dot{x} = Ax + Bu$ with $x(t_0) = 0$. The state at time t_1 is given by

$$x(t_1) = \int_{t_0}^{t_1} e^{A(t_1 - s)} Bu(s) ds.$$

Now think of this formula as the definition of a linear mapping α, say, which sends the function u to the state $x(t_1)$. It then follows that controllability of (A,B) means that this mapping α is onto (surjective).

Secondly, consider the system $\dot{x} = Ax, y = Cx$, which generates the output signal

$$y(t) = Ce^{A(t-t_0)}x_0.$$

This defines a linear mapping β, say, which sends the state x_0 to the function y and observability of (C,A) means that this mapping is one-to-one (injective).

Now the adjoint α^* of α is the linear mapping that sends a state vector v to the function $B^T e^{A^T(t_1 - t)}v = B^T e^{-A^T(t - t_1)}v$. By comparing this with the definition of the mapping β we find that (A,B) is controllable if and only if $(B^T, -A^T)$ is observable. Note that the minus sign in $-A^T$ has not been included in the statements of Theorem 4.29. Although this minus sign is not important in the algebraic context of Theorem 4.29 it does have a specific meaning in the context of so-called adjoint systems, which we shall now explain.

Let be given the system $\dot{x} = Ax$ on the state space $\mathbb{R}^n$. The **adjoint** of this system is defined as $\dot{z} = -A^T z \Longleftrightarrow \dot{z^T} = -z^T A$, which has the dual space $(\mathbb{R}^n)^*$ as its state space. The solutions of the two systems with $x(t_0) = x_0$ and $z^T(t_1) = z_1^T$ are given by

$$\begin{aligned} x(t) &= e^{A(t-t_0)}x_0 \\ z^T(t) &= z_1^T e^{A(t_1 - t)}, \end{aligned}$$

from which we see that the solution of the adjoint system propagates backward in time. The dual properties of controllability and observability deal with processes that propagate forward in time (construcion of state x_1 in future) and backward in time (reconstruction of state x_0 in past), respectively.

Theorem 4.29 enables us to formulate results for observability by dualizing results that have already been proved for controllability. An example of this process of dualization is given by the following Theorem (dualization of Theorem 4.20).

Theorem 4.30 $\ker W$ *is the largest linear subspace in* $\mathbb{R}^n$ *such that*

1. $\ker W \subset \ker C$,
2. $\ker W$ *is A-invariant*.

In the context of the above theorem, $\ker W$ is the largest linear subspace that satisfies points 1 and 2 of the theorem statement, in the sense that every linear subspace that is A-invariant and is contained in $\ker C$, must be contained in $\ker W$.

The linear subspace $\ker W$ is called the **non-observable subspace**. Elements of $\ker W$ are exactly those states that cannot be distinguished from the origin by only looking at the output. An application of Theorem 4.30 is the following. A basis $\{q_1,\ldots,q_k,q_{k+1},\ldots,q_n\}$ in $\mathbb{R}^n$ exists such that $\ker W = \text{span}\{q_1,\ldots,q_k\}$. With respect to this basis A has the form (take $T=(q_1,\ldots,q_k,q_{k+1},\ldots,q_n)$):

$$\bar{A} = T^{-1}AT = \begin{pmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{pmatrix}, \tag{4.17}$$

with $\bar{A}_{11},\bar{A}_{12}$ and $\bar{A}_{22}$ matrices of size $k\times k, k\times(n-k)$ and $(n-k)\times(n-k)$, respectively. Because $\ker W \subset \ker C$, with respect to this basis,

$$\bar{C} = CT = \begin{pmatrix} 0 & \bar{C}_2 \end{pmatrix}, \tag{4.18}$$

with $\bar{C}_2$ a matrix of size $p\times(n-k)$, Furthermore, the pair $(\bar{C}_2,\bar{A}_{22})$ is observable.

It now follows that a system of the form

$$\begin{array}{rcl} \dot{x} &=& Ax + Bu, \\ y &=& Cx + Du, \end{array}$$

which is not observable, by a change of basis given by $x = T\bar{x}$ (or $\bar{x} = T^{-1}x$), with matrix T based on $\ker W$ as above, can be transformed into a system of the form

$$\left.\begin{array}{rcl} \dot{\bar{x}}_1 &=& \bar{A}_{11}\bar{x}_1 + \bar{A}_{12}\bar{x}_2 + \bar{B}_1u, \\ \dot{\bar{x}}_2 &=& \bar{A}_{22}\bar{x}_2 + \bar{B}_2u, \\ y &=& \bar{C}_2\bar{x}_2 + Du, \end{array}\right\} \tag{4.19}$$

where the pair $(\bar{C}_2,\bar{A}_{22})$ is observable. In the above description the submatrices in (4.17) and (4.18) show up, together with $\bar{B} = T^{-1}B = \begin{pmatrix} \bar{B}_1 \\ \bar{B}_2 \end{pmatrix}$ and D. Compare the above form with the form obtained in (4.12).

The next alternative conditions for observability easily follow by dualizing the results in Theorem 4.23, yielding the Hautus test for observabililty.

Theorem 4.31 *Given an $n\times n$ matrix A and a $p\times n$ matrix C, the following statements are equivalent.*

- *The pair (C,A) is observable.*
- $\text{rank}\begin{pmatrix} sI-A \\ C \end{pmatrix} = n$ *for all $s\in\mathbb{C}$.*
- $\text{rank}\begin{pmatrix} \lambda I-A \\ C \end{pmatrix} = n$ *for all eigenvalues λ of matrix A.*

The connection between the input and output (with $x(0) = 0$ and $D = 0$) is given by

$$y(t) = \int_0^t Ce^{A(t-s)}Bu(s)ds,$$

where $Ce^{At}B$ is the impulse response matrix. Now suppose that (A,B) is not controllable, then a basis in $\mathbb{R}^n$ exists such that

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} \tilde{B}_1 \\ 0 \end{pmatrix}, \quad \tilde{C} = \begin{pmatrix} \tilde{C}_1 & \tilde{C}_2 \end{pmatrix}, \tag{4.20}$$

see (4.12). On this basis, $\tilde{A}^k\tilde{B} = \begin{pmatrix} \tilde{A}_{11}^k\tilde{B}_1 \\ 0 \end{pmatrix}$ for all $k \geq 0$, and consequently

$$e^{\tilde{A}t}\tilde{B} = \sum_{k=0}^{\infty} \frac{\tilde{A}^k t^k}{k!}\tilde{B} = \begin{pmatrix} \sum_{k=0}^{\infty} \frac{\tilde{A}_{11}^k t^k}{k!}\tilde{B}_1 \\ 0 \end{pmatrix} = \begin{pmatrix} e^{\tilde{A}_{11}t}\tilde{B}_1 \\ 0 \end{pmatrix}.$$

The second part of the state space (i.e., the complement of $\operatorname{im} R$ in $\mathbb{R}^n$) will not be influenced by the control, and particularly not by an impulsive control. The conclusion is that only the controllable subspace of $\mathbb{R}^n$ will play a role in the impulse response matrix. This also follows from the fact that the impulse response of the system (A,B,C,D) is the same as the impulse response of the system $(\tilde{A}_{11},\tilde{B}_1,\tilde{C}_2,D)$. This statement follows from the next equality, which is a direct consequence of (4.20).

$$Ce^{At}B = \tilde{C}_1 e^{\tilde{A}_{11}t}\tilde{B}_1$$

Recall that $Ce^{At}B = \tilde{C}e^{\tilde{A}t}\tilde{B}$ as consequence of Definition 3.17. Hence, the impulse response is completely determined by the submatrices corresponding to the controllable part of the original system.

Similarly, suppose (C,A) is not observable, then a basis in $\mathbb{R}^n$ exists (not necessarily the same as above) such that

$$\bar{A} = \begin{pmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{pmatrix}, \quad \bar{C} = \begin{pmatrix} 0 & \bar{C}_2 \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} \bar{B}_1 \\ \bar{B}_2 \end{pmatrix}, \tag{4.21}$$

see (4.19). Please be aware of the fact that the $\tilde{A}_{ij}$ in (4.20) and the $\bar{A}_{ij}$ in (4.21) will in general be different. In this basis

$$Ce^{At} = \begin{pmatrix} 0 & \bar{C}_2 e^{\bar{A}_{22}t} \end{pmatrix} \text{ and } Ce^{At}B = \bar{C}_2 e^{\bar{A}_{22}t}\bar{B}_2$$

The first part of the state space (i.e., $\ker W$) does not play any role in the impulse response matrix. Now it follows that the impulse response is completely determined by the submatrices corresponding to the observable part of the original system.

Combining the above two observations, it can be shown that the impulse response is completely determined by the submatrices corresponding to the part of the original system that is both controllable and observable, see also Exercise 4.5.33.

4.4 Realization theory and Hankel matrices

In this section we consider single-input single-output systems only. The quantities

$$g_i = CA^{i-1}B, \quad i = 1,2,3,\ldots, \tag{4.22}$$

which are called the **Markov parameters**, determine the external description of the system

$$\dot{x} = Ax + Bu, \quad y = Cx. \tag{4.23}$$

The Markov parameters show up in the power series expansion of the impulse response

$$G(t) = Ce^{At}B = \sum_{i=0}^{\infty} CA^iB\frac{t^i}{i!}. \tag{4.24}$$

From this latter equation it follows that

$$g_i = \frac{d^{i-1}}{dt^{i-1}}G(t)\Big|_{t=0}.$$

We form the so-called **Hankel matrix** of size $\alpha \times \beta$,

$$H(\alpha,\beta) \stackrel{\text{def}}{=} \begin{pmatrix} g_1 & g_2 & g_3 & g_4 & \cdots & g_\beta \\ g_2 & g_3 & g_4 & \cdots & \cdots & g_{\beta+1} \\ g_3 & g_4 & & & & \vdots \\ g_4 & \vdots & & & & \vdots \\ \vdots & \vdots & & & & \vdots \\ g_\alpha & g_{\alpha+1} & \cdots & \cdots & \cdots & g_{\alpha+\beta-1} \end{pmatrix}. \tag{4.25}$$

Theorem 4.32 *Given the sequence $\{g_1,g_2,g_3,\ldots\}$ there exists a finite-dimensional realization of the form (4.23) of order n (i.e., the state space is $\mathbb{R}^n$) if and only if*

$$\det H(n+i,n+i) = 0 \quad \textit{for all} \quad i = 1,2,\ldots.$$

If, moreover, $\det H(n,n) \neq 0$*, then n is the order of any minimal realization of the sequence $\{g_1,g_2,g_3,\ldots\}$.*

The proof will not be given here; though not difficult, it is somewhat tedious. It can for instance be found in [Chen, 1984]. The last column of $H(n+1,n+1)$ is a linear combination of the first n columns, and, consequently, coefficients $p_0,p_1,\ldots,p_{n-1}$ must exist such that for the jth component, $j = 1,\ldots,n,n+1$, it follows that

$$p_0g_j + p_1g_{j+1} + \cdots + p_{n-1}g_{j+n-1} + g_{j+n} = 0.$$

Also without proof it is stated that, given $\{g_1,g_2,g_3,\ldots\}$ such that the conditions mentioned in Theorem 4.32 are satisfied, a possible realization of the underlying system in

state space space form is

$$A=\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & \cdots & -p_{n-2} & -p_{n-1} \end{pmatrix}, \quad B=\begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ \vdots \\ g_n \end{pmatrix}, \quad C^\top=\begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}.$$

The realization is not unique, as basis transformations will give other realizations.

4.5 Exercises

Exercise 4.5.1 *Investigate the (asymptotic) stability of the system matrices A corresponding to the inverted pendulum in Example 3.6, and the satellite in Exercise 3.5.2.*

Exercise 4.5.2 *Show that the equilibrium point $\bar{x}=0$ of the scalar nonlinear system $\dot{x}=-\varepsilon x+x^2$ is asymptotically stable for each $\varepsilon>0$ and unstable for $\varepsilon\leq 0$. The linearized system (linearized around the equilibrium point $\bar{x}=0$), however, is stable for $\varepsilon=0$. How is this explained?*

Exercise 4.5.3 *Following Definition 4.5 compute the stable and unstable subspace of the matrix*

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

Exercise 4.5.4 *Using (3.28) give the general solution of the equations*

$$\begin{cases} \dot{x}_1 &= \alpha x_1+x_2, \\ \dot{x}_2 &= \alpha x_2+x_3, \\ \dot{x}_3 &= \alpha x_3, \end{cases}$$

where α is a real parameter. From the obtained general solution, derive statements on the (asymptotic) stability or instability of the origin for various values of α, i.e., $\alpha<0$, $\alpha=0$ and $\alpha>0$, and compare the results with Theorem 4.2.

Exercise 4.5.5 *Let A be a real $n\times n$ matrix and I the $n\times n$ identity matrix. Prove that, if*

$$M=\begin{pmatrix} A & I \\ 0 & A \end{pmatrix}, \quad \textit{then} \quad e^{Mt}=\begin{pmatrix} e^{At} & te^{At} \\ 0 & e^{At} \end{pmatrix}.$$

Exercise 4.5.6 *Consider the equations*

$$\begin{cases} \dot{x}_1 &= -x_2+x_3, \\ \dot{x}_2 &= x_1+x_4, \\ \dot{x}_3 &= -x_4, \\ \dot{x}_4 &= x_3. \end{cases}$$

Using Exercise 4.5.5, determine the general solution of the above equations. Next derive results on the (in)stability of the origin and compare them with Theorem 4.2.

Exercise 4.5.7 *Consider the Example 4.6.1 and assume that $a_3 > 0$. Prove that all the roots of $p(s)$ have a negative real part if and only if a_3, a_2, a_1, a_0 are positive and $a_2 a_1 > a_3 a_0$.*

Exercise 4.5.8 *Apply the criterion of Routh to $p(s) = \sum_{k=1}^{6} k s^{6-k}$. What is your conclusion?*

Exercise 4.5.9 *For which value(s) of k has the equation $\lambda^3 + 3\lambda^2 + 3\lambda + k = 0$ only roots with negative real parts?*

Exercise 4.5.10 *For which $k \in \mathbb{R}$ does the polynomial $p(s) = s^4 + 2s^3 + ks^2 + s + 3$ have all its roots in the open left half-plane?*

Exercise 4.5.11 *Modify the criterion of Routh to obtain a criterion to test whether or not a given polynomial has all its roots in the open right half-plane.*

Exercise 4.5.12 *Let A and Q be given $n \times n$ matrices. Assume that all eigenvalues of A have negative real parts. Then prove that the matrix P given by*

$$P = \int_0^\infty e^{A^\top t} Q e^{At} dt$$

is well defined. (Hint: first show that, without loss of generality, it may be assumed that A is in Jordan form and next consider a general element P_{ij} of P.) Further show that if Q is a positive-definite matrix, so is P.

Exercise 4.5.13 *Consider the interval polynomial $[1,2] + [3,4]\lambda + 3\lambda^2 + \lambda^3$ and investigate its stability. Note that the term $3\lambda^2$ can be interpreted as $[3,3]\lambda^2$.*

Exercise 4.5.14 *Consider the polynomial $k + 4\lambda + 2\lambda^2 + \lambda^3$. Investigate if the polynomial has all its roots in the open left half-plane for all $k \in [1,9]$. Do this by means of Routh's criterion, but also by means of Kharitonov's criterion, where terms like $2\lambda^2$ should be interpreted as $[2,2]\lambda^2$, etc.*

Exercise 4.5.15 *Verify by substitution that the expression for $x(t, x_0, u)$ in (4.5) is the solution of $\dot{x} = Ax + Bu$, given the initial condition $x(0) = x_0$ and the input function u.*

Exercise 4.5.16 *Consider the system described by*

$$\dot{x} = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u, \quad y = \begin{pmatrix} 1 & 1 \end{pmatrix} x.$$

Show that the system is externally stable (i.e., BIBO stable), but not internally stable (i.e., asymptotically stable). Explain this by investigating the controllability and observability of the system.

Exercise 4.5.17 *Investigate whether the system described in Equation (3.16) - the inverted pendulum - is controllable.*

Exercise 4.5.18 *Investigate whether the following pairs of matrices are controllable.*

1. $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$,
2. $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$,
3. $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$,
4. $A = \begin{pmatrix} a_1 & 0 \\ a_2 & 0 \end{pmatrix}, B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$,
5. $A = \begin{pmatrix} 0 & l \\ -l & 0 \end{pmatrix}, B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$,
6. $A = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$,
7. $A = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$.

Exercise 4.5.19 *Are the equations of motion of the tractor and cart (Exercise 3.5.8) controllable when the combination moves in forward direction? Same question, but now the combination moves in backward direction.*

Exercise 4.5.20 *Consider the flow diagrams in Figures 4.2 and 4.3. Note that the initial conditions are not specified/included in the diagrams. Verify that the diagrams indeed yield the two systems given in Example 4.18. See also Figure 3.1 and the text that surrounds this figure.*

Exercise 4.5.21 *Consider the matrix pair*

$$A = \begin{pmatrix} 4 & -4 & 2 \\ 3 & -3 & 2 \\ -3 & 2 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 5 \\ 2 \\ -2 \end{pmatrix}.$$

1. *Is the pair (A,B) controllable?*
2. *Which vectors span the controllable subspace?*
3. *Show that the controllable subspace is A-invariant.*

Exercise 4.5.22 *Consider a controllable system $\dot{x} = Ax + Bu$, $y = Cx$, with A an $n \times n$ matrix, B an $n \times 1$ matrix and C a $1 \times n$ matrix. Let T be the corresponding controllability matrix, i.e., $T = \begin{pmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{pmatrix}$. Then T is square and invertible. Show that the triple $(T^{-1}AT, T^{-1}B, CT)$ represents the system in controllability form, defined in (4.9).*

Exercise 4.5.23 *Write the non-controllable pairs of Exercise 4.5.18 in the form of (4.11).*

Exercise 4.5.24 *Using a suitable basis transformation, bring the following pair into the form of (4.11)*

$$A = \begin{pmatrix} -1 & 1 & -1 & -1 \\ -2 & 3 & -1 & 0 \\ -2 & 3 & -1 & 2 \\ 0 & -1 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Exercise 4.5.25 *We are given a scalar input system $\dot{x} = Ax + Bu$, i.e., $u \in \mathbb{R}$, that is controllable. Suppose that a control of the form $u = Kx + v$ is applied, where K is an $1 \times n$ matrix and v is a 'new' control, which again is a scalar. The new system is then characterized by the pair $(A + BK, B)$. Prove that this new system is also controllable.*

Exercise 4.5.26 *On page 60, after the proof of Lemma 4.15, it is claimed that a smaller t_1 leads to a large u (in norm). Can you make this plausible?*

Exercise 4.5.27 *Consider the system described by $\dot{x} = Ax + Bu$, with $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$. Following the proof of Theorem 4.14, prove that controllability, as defined in Definition 4.12, is equivalent to the following: given any $t_1 > 0$, for any two states $x_0, x_1 \in \mathbb{R}^n$, a control function u can be found such that $x_1 = x(t_1, x_0, u)$, i.e., x_1 is the state at time t_1, starting from state x_0 at time $t = 0$ and applying input function u. Please note the difference with Definition 4.12.*

Exercise 4.5.28 *Using the alternative characterization of controllability given in Exercise 4.5.27, show that controllability, reachability and null-controllability are equivalent for linear time-invariant continuous-time systems.*

Exercise 4.5.29 *Verify the controllability of the heated bar model in (4.15) and show that the controllability is independent of the chosen discretization step $\frac{1}{n}$.*

Exercise 4.5.30 *A nonsingular coordinate transformation $x = T\bar{x}$, (sending A to $T^{-1}AT$ and C to CT) does not destroy observability. Show this. If the observability matrix of the transformed system is denoted by $\bar{W}$, then $WT = \bar{W}$.*

Exercise 4.5.31 *Investigate whether the inverted pendulum, as given by the Equations (3.16) and (3.17) is observable. Repeat this investigation if only one of the measurements (i.e., either $y_1(t)$ or $y_2(t)$) is available.*

Exercise 4.5.32 *Consider the pair*

$$A = \begin{pmatrix} -2 & 0 & 0 \\ 1 & -2 & 0 \\ 3 & -1 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & -1 & 1 \end{pmatrix}$$

and, using a suitable basis transformation, write it in the form of (4.17) and (4.18).

Exercise 4.5.33 *Consider the system $\dot{x} = Ax + Bu$, $y = Cx$, given by the matrices*

$$A = \begin{pmatrix} 3 & 1 & -3 & 2 \\ 2 & 2 & 0 & 2 \\ 7 & 0 & 3 & 7 \\ 1 & -1 & 3 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} -1 \\ -3 \\ 0 \\ 1 \end{pmatrix}, \quad C = \begin{pmatrix} -1 & -2 & 0 & -1 \end{pmatrix}.$$

For this system

1. *compute the controllable subspace,*
2. *compute the non-observable subspace,*
3. *determine a* **Kalman decomposition** *(see also Example 5.7), i.e.,*
   - *start with the basis vectors which span the intersection of the controllable subspace and the non-observable subspace,*
   - *append to these vectors new basis vectors such that the controllable subspace is spanned,*
   - *next add to first set of basis vectors new basis vectors such that the non-observable subspace is spanned,*
   - *add more basis vectors, if necessary, such that the whole state space is spanned,*
   - *finally, write the system with respect to the obtained basis.*
4. *compute the impulse response.*

Exercise 4.5.34 *Suppose that the Hankel matrices $H(\alpha,\beta)$ satisfy the conditions given in the statement of Theorem 4.32. Prove that $H(n,n) = WR$, where W and R are the observability and the controllability matrices, respectively, and that, if $\det H(n,n) \neq 0$, any n dimensional state space realization is both controllable and observable.*

Exercise 4.5.35 *Consider the system described by*

$$\dot{x} = Ax + Bu, \quad y = Cx,$$

with $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times 1}$ and $C \in \mathbb{R}^{1\times n}$. Let $H(\alpha,\beta)$ be the Hankel matrix, defined in (4.25), made up of the Markov parameters $g_i = CA^{i-1}B$, $i \geq 1$, see (4.22). Now prove that rank $H(i,n) =$ rank $H(i,n+1)$ for all $i \geq 1$.

Exercise 4.5.36 *Below you will find a number of statements. For each of statements determine whether it is true or false. Make your answer plausible by means of a simple reasoning or (counter)example.*

1. *If a non-linear system $\dot{x} = f(x,u)$ has two different solution pairs $(\tilde{x}_1,\tilde{u}_1)$ en $(\tilde{x}_2,\tilde{u}_2)$, then the linearization around both solution pairs must be either stable or unstable.*
2. *Asymptotic stability can be lost under linearization.*

3. *The linear system $\dot{x} = A(t)x$, with $A(t)$ a time-varying $n \times n$ matrix, is asymptotically stable if and only if all ('time-varying') eigenvalues of $A(t)$ lie in the open left half-plane.*

4. *The roots of $-5 - 2s^3 + 6s + 2s^2 - s^4$ all have a negative real part.*

5. *The eigenvalues of the matrix*

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4 & -8 & -2 \end{pmatrix}$$

all have a negative real part.

6. *The roots of $s^3 + 4s^2 + 2s + K$ all have a negative real part for all K satisfying $0 < K < 8$.*

7. *If (A,B) is a controllable pair, so is the pair $(A+BF,B)$ for any suitable matrix F.*

8. *If the pair (A,B) is controllable, then also the pair (A^2,B) is controllable.*

9. *If the pair (A^2,B) is controllable, then also the pair (A,B) is controllable.*

10. *If (A_1,B) and (A_2,B) are controllable pairs, then so is the pair (A_1+A_2,B).*

11. *If (A,B_1) and (A,B_2) are controllable pairs, then so is the pair (A,B_1+B_2).*

12. *The controllable subspace corresponding to the pair (A,B) is A^2-invariant.*

13. *The linear system $\dot{x} = Ax$, with A a constant $n \times n$ matrix, is stable if and only is all eigenvalues of A lie in the closed left half-plane.*

14. *If A contains identical eigenvalues, then there is no suitable matrix B such that the pair (A,B) is controllable.*

15. *Let A be an $n \times n$ matrix and B an $n \times m$ matrix with $m > 2$. If b is a column of B and the pair (A,b) is controllable, then also the pair (A,B) is controllable.*

16. *The controllable subspace of the system $\dot{x} = Ax + Bu$, $y = Cx$ is the largest A-invariant subspace that contains the image of B.*

17. *Controllability of a linear time-invariant system is basis dependent.*

18. *Let $C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ be a $2 \times n$ matrix with c_1 and c_2 two $1 \times n$ rows. If the pair (C,A) is observable, then at least one of the pairs (c_1,A) and (c_2,A) is also observable.*

19. *The non-observable subspace of the system $\dot{x} = Ax + Bu$, $y = Cx$ is the smallest A-invariant subspace in the kernel of C.*

20. *Let A be an $n \times n$ matrix such that $\operatorname{rank} A < n-1$ and let C be a $1 \times n$ matrix. Then the pair (C,A) can not be observable.*

21. *Observability can be lost under linearization.*

22. *Let $\dot{x} = Ax + Bu$, $y = Cx$ be a linear system with $x \in \mathbb{R}^n$, $u, y \in \mathbb{R}$. If the controllable subspace is contained in the non-observable subspace, the impulse response of the system is identically equal to zero.*

23. *A system can not be simultaneously unstable, controllable and observable.*

24. *A system can not be simultaneously uncontrollable and unobservable.*

25. *A linear time-invariant system whose step response function $g(t)$ satisfies $\lim_{t\to\infty} g(t) = 0$ can not be asymptotically stable.*

26. *Let $H(n,n)$ be the $n \times n$ Hankel matrix corresponding to the impulse response of the system $\dot{x} = Ax + Bu$, $y = Cx$ with $x \in \mathbb{R}^n$ and $u, y \in \mathbb{R}$. The element of $H(n,n)$ in row i and column j is given by $CA^{i-1}A^{j-1}B$.*

27. *The 'maximal' realization problem is a meaningful problem.*

Chapter 5

State and Output Feedback

5.1 Feedback and stabilizability

In Example 1.1, on the autopilot of a boat, the control function u was expressed in known quantities such as to obtain a good steering behavior of the ship. A possible control law had the form $u = Ke$, where K is a constant and e is the difference between the actual and the desired heading. One can imagine that the desired heading has been set by the helmsman and that the actual heading is continuously measured (can then be seen as an output of the ship dynamics). Also, with manual control by the helmsman (when the autopilot is not in use) the helmsman is aware of the current heading and makes corrections if this heading deviates from the desired heading. In both situations the output (the measurement/observation) is fed back to the input (the control). Such a control mechanism is a form of **feedback control**, or, equivalently, **closed-loop control** (the output is connected to the input, so that the 'loop' is closed, and the system governs itself). In contrast to closed-loop control there also exists **open-loop** control. In a system with open-loop control the control action (the function u) is independent of the output.

Example 5.1 An automatic toaster (i.e., a toaster that switches off automatically) is a system with open-loop control, because it is controlled by a timer (the function u is an on-off function). The time required to make 'good' toast must be estimated by the user, who is not a part of the system. Control over the quality (say color) of toast (with color seen as the output) is removed once the timer has been set. One could design a toaster with a feedback control, where the color of the toast is continuously measured and this measurement is connected to the switch of the heating element. □

We will now turn to a more mathematical treatment of the feedback principle. Suppose we are given a system described by

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

with $x \in \mathbb{R}^n, u \in \mathbb{R}^m, y \in \mathbb{R}^p$, and A, B, C real matrices of appropriate sizes. Suppose furthermore that the system is unstable and that the whole state is measured/observed, i.e., $C = I$. To focus the ideas, think of the inverted pendulum introduced in Section 2.4.1, and worked out in Example 3.6, with every state component being measured. Given an initial perturbation $x_0 \neq 0$ ($x_0 = 0$ corresponds to the unstable equilibrium of the carriage situated at $s = 0$ and the pendulum vertically upwards with both carriage and pendulum at rest), one could calculate a time function $u : [0, \infty) \to \mathbb{R}$ such that the solution of $\dot{x} = Ax + Bu$, $x(0) = x_0$ will converge to 0 as $t \to \infty$. Such an open-loop control will be not very practical, since future perturbations are not taken into account.

Instead, one could think of a feedback control and, more specifically, of a linear feedback control

$$u(t) = Fx(t), \tag{5.1}$$

where, in general, F is an $m \times n$ matrix (in the context of Example 3.6, F is an 1×4 matrix). The state x then satisfies

$$\dot{x} = Ax + BFx = (A + BF)x. \tag{5.2}$$

The matrix F must be chosen such that the closed-loop system (5.2) has a desired behavior (if possible), for instance, being asymptotically stable. A control law of the form (5.1) is called **state feedback**.

If the state is not available, one might feed back the output, i.e., $u = Hy$, where H is a suitably chosen $m \times p$ matrix. The state x will then satisfy

$$\dot{x} = Ax + BHy = Ax + BHCx = (A + BHC)x.$$

Such a control is called **output feedback**. It is clear that state feedback is at least as powerful as output feedback.

Sometimes one would like to have the possibility of influencing the system after (state) feedback has been applied. A candidate for the control law is then

$$u = Fx + Gv,$$

where v is the new input and G is a matrix of appropriate size. One could, for instance, think of stabilizing the inverted pendulum (keeping the pendulum vertical), while the carriage must be moved from one position to the other (by means of v).

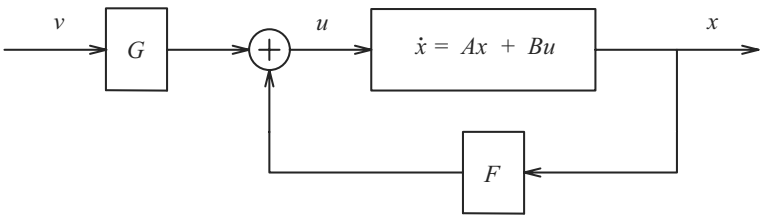

Figure 5.1 System with state feedback and new controls.

The input $u = Fx$ is a **control law**. If it is viewed as a (static) system itself with x as input and u as output, the control law is called a (static) **compensator**. The word 'static' is occasionally added to stipulate that the control at time t only depends on the state at time t.

The dynamic behavior of a system can be influenced by means of a compensator. We want to use this influence to **stabilize** the system around an unstable equilibrium point. This stabilization issue will be the main topic of this chapter, though there are other system properties that can also be influenced by means of a compensator. Conditions on the matrices A and B will be given such that the new matrix $A + BF$ is asymptotically stable when an appropriate matrix F is chosen. Therefore, we first define the following.

Definition 5.2 *The system $\dot{x} = Ax + Bu$, or also the pair (A,B), is* **stabilizable** *if there exists a real $m \times n$ matrix F such that $\mathrm{Re}\,\lambda < 0$ for all eigenvalues λ of $A+BF$.*

We recall that controllability of a system is equivalent to the ability that the system can be steered from any initial state to any final state in some finite time by applying an appropriate control. Null-controllability is the ability to steer a system from an arbitrary initial state to zero in some finite time. Stabilizability can be shown to be equivalent to the ability that any initial state can be steered to zero by applying an appropriate control that however may need to be defined over an infinitely long time interval.

The following theorem provides a sufficient condition for a system to be stabilizable. The theorem is one of the milestones in the history of system theory and has played a fundamental role in its development.

Theorem 5.3 *Consider the linear system $\dot{x} = Ax + Bu$, with $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ and A, B real matrices of appropriate sizes. Then, the system is controllable if and only if for each polynomial $r(\lambda) = \lambda^n + r_{n-1}\lambda^{n-1} + \cdots + r_1\lambda + r_0$, with real coefficients $r_{n-1}, \ldots, r_1, r_0$, there exists a real $m \times n$ matrix F such that $\det(\lambda I - (A+BF)) = r(\lambda)$.*

Hence, if (A,B) is controllable, the characteristic polynomial of $A+BF$ can be assigned arbitrarily by a suitable choice of F. Therefore, the zeros of the characteristic polynomial, which are identical to the eigenvalues of $A+BF$, can be placed at any location (provided that complex eigenvalues occur in conjugate pairs). A particular location is the open left half of the complex plane, implying that when $\dot{x} = Ax + Bu$ is controllable, the system is also stabilizable (the converse statement is not necessarily true). Theorem 5.3 is sometimes called the **pole-assignment theorem**.

Proof of Theorem 5.3. (The proof will only be given for single-input systems.)

Necessity. In this part we prove that, if the system is not controllable, a matrix F with the required property does not exist. First, assume that for each arbitrary $r(\lambda)$ of the form as given in the statement of the theorem a matrix F exists such that $\det(\lambda I - (A+BF)) = r(\lambda)$ and that the system $\dot{x} = Ax + Bu$ is not controllable. Then, a basis transformation matrix T can be found such that (see Equation (4.11) on page 65),

$$\tilde{A} = T^{-1}AT = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}, \quad \tilde{B} = T^{-1}B = \begin{pmatrix} \tilde{B}_1 \\ 0 \end{pmatrix}. \tag{5.3}$$

If we transform an arbitrary feedback matrix F correspondingly to $\tilde{F} = FT$ and partition the latter conveniently as $(\tilde{F}_1 \;\; \tilde{F}_2)$, then

$$\begin{aligned} \tilde{A} + \tilde{B}\tilde{F} &= \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} + \begin{pmatrix} \tilde{B}_1 \\ 0 \end{pmatrix} (\tilde{F}_1 \;\; \tilde{F}_2) \\ &= \begin{pmatrix} \tilde{A}_{11} + \tilde{B}_1\tilde{F}_1 & \tilde{A}_{12} + \tilde{B}_1\tilde{F}_2 \\ 0 & \tilde{A}_{22} \end{pmatrix}, \end{aligned}$$

and the characteristic polynomial is

$$\begin{aligned} \det(\lambda I - \tilde{A} - \tilde{B}\tilde{F}) &= \det \begin{pmatrix} \lambda I - (\tilde{A}_{11} + \tilde{B}_1\tilde{F}_1) & -(\tilde{A}_{12} + \tilde{B}_1\tilde{F}_2) \\ 0 & \lambda I - \tilde{A}_{22} \end{pmatrix} \\ &= \det(\lambda I - (\tilde{A}_{11} + \tilde{B}_1\tilde{F}_1)) \cdot \det(\lambda I - \tilde{A}_{22}). \end{aligned}$$

The latter is based on the well-known identity from linear algebra

$$\det\begin{pmatrix} P & Q \\ 0 & R \end{pmatrix} = \det P \cdot \det R,$$

where P and R are square matrices and Q has an appropriate size.

It follows that whatever the choice of F is, the polynomial $\det(\lambda I - \tilde{A}_{22})$ always is a factor of the characteristic polynomial of $\tilde{A}+\tilde{B}\tilde{F}$, and cannot be chosen arbitrarily. Since the characteristic polynomials of $\tilde{A}+\tilde{B}\tilde{F}$ and $A+BF$ are the same, the characteristic polynomial $A+BF$ also cannot be chosen arbitrarily. Hence, a contradiction has been obtained and therefore $\dot{x} = Ax + Bu$ is controllable.

Sufficiency. In this part we prove that an F with the required properties can be found if the system is controllable. Now we assume that (A,B) is controllable and we will show that for each $r(\lambda)$ a unique $1 \times n$ matrix F exists such that $\det(\lambda I - (A+BF)) = r(\lambda)$. Towards this end, we assume that by means of a coordinate transformation A and B can be brought in the *controller form*, introduced on page 62, and defined as

$$\bar{A} = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & \dots & \dots & -p_{n-1} \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{5.4}$$

It will be proved in the next lemma that such a coordinate transformation exists. Take $\bar{F} = (p_0 - r_0, p_1 - r_1, \dots, p_{n-1} - r_{n-1})$, then $\bar{A} + \bar{B}\bar{F} =$

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & \dots & -p_{n-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} (p_0 - r_0, p_1 - r_1, \dots, p_{n-1} - r_{n-1})$$

$$= \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ -r_0 & -r_1 & \dots & \dots & \dots & -r_{n-1} \end{pmatrix}$$

and therefore $\det(\lambda I - (\bar{A} + \bar{B}\bar{F})) = r(\lambda)$ (see Exercise 3.5.14). It is clear that $\bar{F}$ is unique. A coordinate transformation does not change the eigenvalues (i.e., eigenvalues of $A+BF$ are exactly the same as the eigenvalues of $T^{-1}(A+BF)T$, where T is an invertible matrix,

for instance such that $\bar{A} = T^{-1}AT$ and $\bar{B} = T^{-1}B$) and the result therefore also holds true for the original system, which was possibly not in the *controller form.* □

In the proof of Theorem 5.3 we used the following lemma.

Lemma 5.4 *If (A,B), with $m = 1$, is a controllable pair, then a basis transformation T exists,* $\det T \neq 0$, *such that $\bar{A} = T^{-1}AT$ and $\bar{B} = T^{-1}B$ are in the controller form, defined in Equation (5.4). The elements p_i in matrix $\bar{A}$, $0 \leq i \leq n-1$, are the coefficients in the characteristic polynomial of A, i.e.,* $\det(\lambda I - A) = \lambda^n + p_{n-1}\lambda^{n-1} + \cdots + p_1\lambda + p_0$.

Proof A new basis $\{q_1,\ldots,q_n\}$ in $\mathbb{R}^n$ is constructed as follows:

$$\begin{aligned} q_n &= B, \\ q_{n-1} &= AB + p_{n-1}B = Aq_n + p_{n-1}q_n, \\ q_{n-2} &= A^2B + p_{n-1}AB + p_{n-2}B = Aq_{n-1} + p_{n-2}q_n, \\ &\vdots \\ q_1 &= A^{n-1}B + p_{n-1}A^{n-2}B + \ldots + p_1B = Aq_2 + p_1q_n. \end{aligned} \tag{5.5}$$

Because the pair (A,B) is controllable, $\operatorname{span}\{B,AB,\cdots,A^{n-1}B\} = \mathbb{R}^n$, and therefore, by construction, also $\operatorname{span}\{q_1,q_2,\ldots,q_n\} = \mathbb{R}^n$. Indeed, by induction it follows that $\operatorname{span}\{q_{n-k},\ldots,q_{n-1},q_n\} = \operatorname{span}\{B,AB,\cdots,A^kB\}$, for all $k = 0,1,\ldots,n-1$. Hence, $\{q_1, q_2,\ldots,q_n\}$ is a basis. Let T be the corresponding basis transformation matrix, then $T = (q_1,q_2,\ldots,q_n)$ and $T^{-1}B = T^{-1}q_n = (0,0,\ldots,0,1)^\top = \bar{B}$, see (5.4). Furthermore, from the second till the last equation of (5.5), we obtain

$$\begin{aligned} Aq_n &= q_{n-1} - p_{n-1}q_n, \\ Aq_{n-1} &= q_{n-2} - p_{n-2}q_n, \\ &\vdots \\ Aq_2 &= q_1 - p_1q_n, \end{aligned}$$

and we can write (using again the last equation of (5.5), and Cayley-Hamilton)

$$\begin{aligned} Aq_1 &= A(A^{n-1}B + p_{n-1}A^{n-2}B + \cdots + p_1B) \\ &= A^nB + p_{n-1}A^{n-1}B + \cdots + p_1AB \\ &= (A^n + p_{n-1}A^{n-1} + \cdots + p_1A + p_0I - p_0I)B = -p_0B = -p_0q_n. \end{aligned}$$

Now, for $i = 1,2,\ldots,n$, the vectors Aq_i have been expressed as linear combinations of the vectors $q_j, j = 1,2,\ldots,n$. From the expressions we see directly that $AT = T\bar{A}$ or $\bar{A} = T^{-1}AT$, with $\bar{A}$ given as

$$\bar{A} = \begin{pmatrix} 0 & 1 & 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & 1 & & & \vdots \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & & & 1 \\ -p_0 & -p_1 & -p_2 & \ldots & \ldots & -p_{n-1} \end{pmatrix}.$$

Since expressions for $\bar{A}$ and $\bar{B}$ have been found, as in (5.4), the proof is done. □

The proof of Theorem 5.3 provides an algorithm for finding an $1 \times n$ matrix $\bar{F}$ that gives the system the desired properties, i.e., the desired characteristic polynomial. Towards this end, the system (A,B) is first transformed to $(\bar{A},\bar{B}) = (T^{-1}AT, T^{-1}B)$, the *controller form*, defined in (5.4). With respect to this form $\bar{F}$ is obtained in a trivial way. With respect to the original basis, F can be obtained by computing $\bar{F}T^{-1}$. Exercise 5.5.3 gives an alternative algorithm.

Example 5.5 Consider the system of the inverted pendulum described by Equation (3.16). Since this system is controllable (see Exercise 4.5.17), it can be made asymptotically stable using an appropriate feedback matrix $F = (f_1, f_2, f_3, f_4)$. Of course, in order for the associated feedback control to be realizable, all components of the state vector must be known (in the next section we will see what can be done if, for example, only the output y is known). The matrices A and B are given by

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ -2.4 \\ 0 \\ 1 \end{pmatrix}.$$

Matrix A has eigenvalues $0\,(2\times), 5$ and -5 (see Example 3.12) and the uncontrolled system $(u \equiv 0)$ is not stable. The characteristic polynomial of A is $\lambda^4 - 25\lambda^2$. Suppose we want to choose F in such a way that the eigenvalues of $A+BF$ are $-1, -2$ and $-2 \pm i$. Then the associated desired characteristic polynomial $r(\lambda)$ is $(\lambda+1)(\lambda+2)(\lambda^2+4\lambda+5) = \lambda^4+7\lambda^3+19\lambda^2+23\lambda+10$. In order to construct F we could use the algorithm in proof of Theorem 5.3, or the formula in Exercise 5.5.3. Another more direct method is that F, written as (f_1, f_2, f_3, f_4), must be chosen such that $\det(\lambda I - (A+BF)) = r(\lambda)$. This gives the equation

$$\det \begin{pmatrix} \lambda & -1 & 0 & 0 \\ -25+2.4f_1 & \lambda+2.4f_2 & 2.4f_3 & 2.4f_4 \\ 0 & 0 & \lambda & -1 \\ 0.6-f_1 & -f_2 & -f_3 & \lambda-f_4 \end{pmatrix} = \lambda^4+7\lambda^3+19\lambda^2+23\lambda+10.$$

Hence, it follows that

$$\lambda^4 + (2.4f_2 - f_4)\lambda^3 + (-f_3 - 25 + 2.4f_1)\lambda^2 + 23.56f_4\lambda + 23.56f_3 = \lambda^4+7\lambda^3+19\lambda^2+23\lambda+10,$$

and the associated feedback components are therefore

$$f_3 = \frac{10}{23.56}, \quad f_4 = \frac{23}{23.56}, \quad f_1 = \frac{1}{2.4}\left(44 + \frac{10}{23.56}\right), \quad f_2 = \frac{1}{2.4}\left(7 + \frac{23}{23.56}\right).$$

□

Example 5.6 Let be given the system (in controller form)

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -3 & 1 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} u.$$

If the input is chosen as $u(t) = (f_1\ f_2\ f_3)x(t)$, with $f_i, i = 1,2,3$, constants, for what values of the f_i is the closed-loop system asymptotically stable? Substitution of the feedback law results in

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2+f_1 & -3+f_2 & 1+f_3 \end{pmatrix} x.$$

The characteristic polynomial of this system matrix is (see Exercise 3.5.14)

$$\lambda^3 + (-1-f_3)\lambda^2 + (3-f_2)\lambda + (-2-f_1) = 0.$$

Since the exact location of the zeros is not important, we will use the criterion of Routh (see Section 4.1.2) to obtain conditions for $f_i, i = 1,2,3$, which will guarantee asymptotic stability. The resulting Routh table is

$$\begin{array}{cc} 1 & +3-f_2 \\ -1-f_3 & -2-f_1 \\ \psi & \\ -2-f_1 & \end{array}$$

where $\psi = \dfrac{1}{(-1-f_3)}\big((-1-f_3)(3-f_2) - (-2-f_2)\big)$. Necessary and sufficient conditions for asymptotic stability are therefore

$$\begin{aligned} &-1-f_3 > 0, \\ &-2-f_1 > 0, \\ &(-1-f_3)(3-f_2) > (-2-f_1). \end{aligned}$$

□

Example 5.7 Let be given the system $\dot{x} = Ax + Bu,\ y = Cx$ with

$$A = \begin{pmatrix} -1 & 2 & 0 & -3 \\ 0 & -2 & 0 & 0 \\ 2 & 1 & -3 & -3 \\ 0 & 2 & 0 & -4 \end{pmatrix}, \quad B = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad C = (0\ 1\ 1\ -1).$$

1. Is the system controllable? What is the controllable subspace?
2. Is the system observable? What is the non-observable subspace?
3. Is the system stabilizable?
4. Write the system in terms of basis vectors which are chosen according to the following rules (and in the order specified):
   - start with the vectors that span the intersection of the controllable subspace and the non-observable subspace,

- append to these vectors new basis vectors such that the controllable subspace is spanned,
- next add to first set of basis vectors new basis vectors such that the non-observable subspace is spanned,
- add more basis vectors, if necessary, such that the whole state space $\mathbb{R}^n$, with $n = 4$, is spanned,
- finally, write the system with respect to the obtained basis.

5. Can you design a control law $u = Fx$ such that the feedback system has its eigenvalues in $-1, -1, -3$ and -4, respectively? The same question again, but now the eigenvalues must be located in $-1, -1, -2$ and -3, respectively.

Answer question 1. The controllability matrix equals

$$R = \begin{pmatrix} 2 & -3 & 5 & -9 \\ 1 & -2 & 4 & -8 \\ 1 & -1 & 1 & -1 \\ 1 & -2 & 4 & -8 \end{pmatrix},$$

which has rank 2. Hence, the system is not controllable. The controllable subspace is spanned by the first two columns of R, which is equivalent to

$$\operatorname{im} R = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Answer question 2. The observability matrix equals

$$W = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 2 & -3 & -3 & 1 \\ -8 & 9 & 9 & -1 \\ 26 & -27 & -27 & 1 \end{pmatrix},$$

which has rank 2. Hence, the system is not observable. The non-observable subspace is spanned by two linearly independent vectors x for which $Wx = 0$. Two such vectors are $(1\ 1\ 0\ 1)^\top$ and $(0\ 1\ -1\ 0)^\top$, so that

$$\ker W = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \right\}.$$

Answer question 3. If one calculates the eigenvalues of A, they turn out to be $-1, -2, -3$ and -4 and, hence, A is (asymptotically) stable. The system is therefore (trivially) stabilizable.

Answer question 4. The intersection of the controllable subspace and the non-observable subspace is spanned by the vector $v_1 \stackrel{\text{def}}{=} (1\ 1\ 0\ 1)^\top$. The controllable subspace

is spanned by v_1 and v_2, with $v_2 \stackrel{\text{def}}{=} (1\,0\,1\,0)^\top$, and the non-observable subspace is spanned by v_1 and v_3, with $v_3 \stackrel{\text{def}}{=} (0\,1\,-1\,0)^\top$. Finally, $\mathbb{R}^4$ is spanned by v_1, v_2, v_3 and v_4, with $v_4 \stackrel{\text{def}}{=} (1\,0\,0\,0)^\top$. If we choose the basis transformation matrix $T = (v_1, v_2, v_3, v_4)$, then

$$\hat{A} = T^{-1}AT = \left(\begin{array}{cc|cc} -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 2 \\ \hline 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & -3 \end{array}\right), \quad \hat{B} = T^{-1}B = \left(\begin{array}{c} 1 \\ 1 \\ \hline 0 \\ 0 \end{array}\right),$$

$$\hat{C} = CT = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \end{array}\right).$$

Answer question 5. Matrices $\hat{A}$ and $\hat{B}$ are of the form

$$\hat{A} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} B_1 \\ 0 \end{pmatrix}.$$

The eigenvalues of A_{22} will not change if a feedback law is introduced. Therefore, whatever (linear) feedback law is implemented, the eigenvalues at locations -3 and -4 (the eigenvalues of A_{22}), cannot be influenced. Hence, a design with the second set of requirements (eigenvalues at $-1,-1,-2,-3$) is impossible, whereas a design subject to the first set of requirements (eigenvalues at $-1,-1,-3,-4$) is possible. Indeed, in the latter case, first determine the 1×2 matrix F_1 such that the eigenvalues of $A_{11} + B_1F_1$ are equal to $-1\,(2\times)$. It is easily seen that $F_1 = (1 \quad 0)$. Next, take any arbitrary 1×2 matrix F_2, and compute $F = (F_1 \quad F_2)T^{-1}$. Then F is such that $A + BF$ has eigenvalues at $-1,-1,-3,-4$. □

The conclusion of the previous is that $A + BF$ can be made asymptotically stable if (A,B) is controllable *and* provided state feedback is possible, i.e., the output y equals the state. If $y = Cx$ and C is not invertible, then the problem of making $A + BHC$ asymptotically stable by means of a suitable choice of H is far more difficult. Hardly any general theorem exists with respect to this situation. In the next section we will consider the problem of reconstructing the state x out of past measurements y such that the reconstructed state will be available for feedback purposes.

We conclude this section with two remarks. The first remark provides necessary and sufficient conditions for a system $\dot{x} = Ax + Bu$, or simply the pair (A,B), to be stabilizable. A proof of the conditions can be found in [Trentelman, Stoorvogel and Hautus, 2001], Chapter 3. The second remark is concerned with the uniqueness of an eigenvalue assigning feedback in case $m = 1$.

Remark 5.8 *Consider the pair (A,B), where A is a real $n \times n$ matrix and B a real $n \times m$ matrix. Then the following statements are equivalent.*

1. *The pair (A,B) is stabilizable.*
2. $\operatorname{rank}(sI - A \;,\; B) = n$ *for all* $s \in \mathbb{C}$ *with* $\operatorname{Re} s \geq 0$.
3. $\operatorname{rank}(\lambda I - A \;,\; B) = n$ *for all eigenvalues* λ *of matrix A with* $\operatorname{Re}\lambda \geq 0$.

Remark 5.9 Consider the controllable pair (A,B), where A is an $n \times n$ matrix and B an $n \times m$ matrix. Then, the requirement that F is such that $A+BF$ has a certain desired characteristic polynomial $r(\lambda)$ results in n constraints (the n coefficients of $r(\lambda)$) that need to be satisfied. If $m = 1$, then F contains n parameters. Hence, in that case there are as many degrees of freedom as there are constraints, and it turns out that F is uniquely determined. See also the sufficiency part of the proof of Theorem 5.3. If $m > 1$, then F contains $nm(> n)$ parameters. In that case there are more degrees of freedom than constraints, implying that F will not be uniquely determined.

5.2 Observers and state reconstruction

Many procedures for the control of systems are based on the assumption that the whole state vector can be observed. In such procedures the control law is of the form $u = Fx$ (or $u = Fx+Gv$). In many systems, however, not the whole state vector can be observed. Sometimes very expensive measurement equipment would be necessary to observe the whole state, specifically in physical systems. In economic systems very extensive, statistical measurement procedures would be necessary. Sometimes, also, it is simply impossible to obtain measurements of the whole state if some internal variables cannot be reached. Think for instance of a satellite, where, because of the weight problems, hardly any measurement equipment (for the temperature, for instance) can be built in the satellite. Once in orbit, the satellite is too far away to measure certain quantities from the earth. In all these cases, control must be based on the available information, i.e., the output $y = Cx$. An auxiliary system will be built, called the **observer**, which has as input both the control u and the output y of the real system, and which has as output an approximation $\hat{x}$ of the state vector x of the real system. An observer for the system $\dot{x} = Ax+Bu$, $y = Cx$, is assumed to be a system itself, of the form

$$\begin{aligned} \dot{z} &= Pz+Qu+Ky, \\ \hat{x} &= Sz+Tu+Ry. \end{aligned}$$

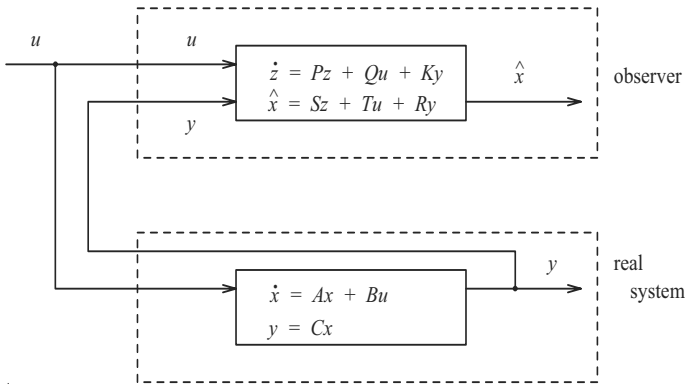

Figure 5.2 System and general observer.

In the flow diagram in Figure 5.2 the observer, the real system and the connections between these systems have been drawn. The vector z is the state of the observer. The matrices P,Q,K,S,T and R are to be determined. Think of a real system as the satellite in orbit, where x cannot easily be measured, and only measurements of a few state variables, such as position and distance, are available. The observer is an auxiliary system on earth (a computer program, for instance) from which all variables are easily accessible.

Recall that we want an observer for the unknown state x. It turns out that we can construct such an observer by taking $S = I, T = R = 0$. This yields that $\hat{x} = z$ and that the state of the observer has the role of an approximation of the unknown state x, i.e., $\dot{\hat{x}} = P\hat{x} + Qu + Ky$.

If at all possible, the observer should at least satisfy the following requirements:

1. The difference $x(t) - \hat{x}(t)$ must converge to zero as $t \to \infty$, irrespective of the initial condition conditions $x(0) = x_0$, $\hat{x}(0) = \hat{x}_0$, and the applied control function u.
2. If $\hat{x}(t_0) = x(t_0)$ at a certain time instant t_0, then it should hold that $\hat{x}(t) = x(t)$ for all $t \geq t_0$ and every control function u. Hence, once the observer has the correct estimate of the real state, then this estimate should remain correct for the future, no matter which control is applied.

We now have

$$\begin{aligned} \frac{d}{dt}(x-\hat{x}) &= Ax + Bu - P\hat{x} - Qu - Ky \\ &= Ax + Bu - P\hat{x} - Qu - KCx \\ &= (A-KC)x - P\hat{x} + (B-Q)u. \end{aligned}$$

The second requirement formulated above now yields that

$$B = Q, \quad A - KC = P.$$

The observer then has the form

$$\dot{\hat{x}} = A\hat{x} + Bu + K(y - \hat{y}) \quad \text{in which} \quad \hat{y} = C\hat{x}. \tag{5.6}$$

Apparently, the observer very much looks like the original system. It is a duplicate of the real system, apart from an additional input term $K(y-\hat{y})$, which can be interpreted as a correction term.

In order for the first requirement to be satisfied we consider how the difference $e(t) \stackrel{\text{def}}{=} x(t) - \hat{x}(t)$ behaves as $t \to \infty$. We have

$$\dot{e} = \frac{d}{dt}(x-\hat{x}) = (A-KC)e,$$

and so the requirement that $\lim_{t\to\infty} e(t) = 0$, for any initial $e(0)$, means that the matrix $A - KC$ must be asymptotically stable. This brings us to the following definition, where A,B and C are real matrices of size $n \times n, n \times m$ and $p \times n$, respectively.

Definition 5.10 *The system $\dot{x} = Ax + Bu$, $y = Cx$, or also the pair (C,A), is* **detectable** *if there exists a real $n \times p$ matrix K such that $\operatorname{Re}\lambda < 0$ for all eigenvalues λ of $A - KC$.*

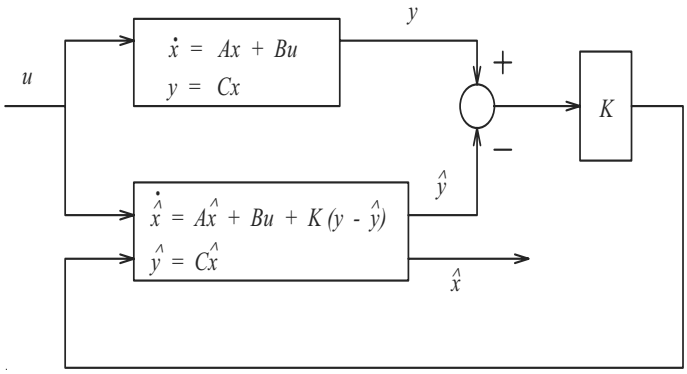

Figure 5.3 System and state observer.

The following theorem provides a sufficient condition for the pair (C,A) to be detectable.

Theorem 5.11 *For each polynomial* $w(\lambda) = \lambda^n + w_{n-1}\lambda^{n-1} + \cdots + w_1\lambda + w_0$, *with real coefficients* $w_0, w_1, \ldots, w_{n-1}$, *there exists a real* $n \times p$ *matrix* K *such that* $\det(\lambda I - (A - KC)) = w(\lambda)$ *if and only if the pair* (C,A) *is observable.*

Proof By Theorem 4.29 the pair (C,A) is observable if and only if the pair $(A^\top, C^\top)$ is controllable. Theorem 5.3 states that the pair $(A^\top, C^\top)$ is controllable if and only if for each polynomial $w(\lambda)$, as mentioned in the statement of this theorem, there exists a matrix F such that $\det(\lambda I - (A^\top + C^\top F)) = w(\lambda)$. Choose $K = -F^\top$, then

$$\det(\lambda I - (A - KC)) = \det(\lambda I - (A^\top - C^\top K^\top)) = w(\lambda).$$

□

Theorem 5.11 gives a necessary and sufficient condition such that the eigenvalues of $A - KC$ can be chosen at will. However, in observer design one is satisfied when all eigenvalues are in the open left half-plane (and the eigenvalues are not necessarily at arbitrarily prescribed places). This is of course a weaker requirement for which observability is a sufficient, but not a necessary condition. For instance, consider the matrix pair

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad C = (1\ 0),$$

which is not observable. The eigenvalues of $A - KC$, with $K = (k_1, k_2)^\top$, are the zeros of

$$\det(\lambda I - (A - KC)) = (\lambda - 1 + k_1)(\lambda + 1)\,.$$

If we choose $k_1 > 1$, both zeros are in the open left half-plane and an observer can be constructed whose state converges to the real state vector for $t \to \infty$. One of the eigenvalues, i.e., $\lambda = -1$, is fixed and cannot be chosen at will.

Example 5.12 This example is a continuation of Example 5.5 of the inverted pendulum. We assume that only measurements of the position of the carriage are available, such that A, B and C are given by

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ -2.4 \\ 0 \\ 1 \end{pmatrix}, \quad C = (0 \quad 0 \quad 1 \quad 0).$$

The observability matrix is

$$W = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \\ 0 & -0.6 & 0 & 0 \end{pmatrix}.$$

Clearly, rank $W = 4$, so that Theorem 5.11 can indeed be applied.

Suppose we want to construct an observer such that the eigenvalues of $A - KC$ are situated in the points -1 $(2\times)$ and $-1 \pm i$. This means that K, written as $(k_1, k_2, k_3, k_4)^\top$, must be determined such that

$$\det(\lambda I - (A - KC)) = \det \begin{pmatrix} \lambda & -1 & +k_1 & 0 \\ -25 & \lambda & +k_2 & 0 \\ 0 & 0 & \lambda + k_3 & -1 \\ +0.6 & 0 & +k_4 & \lambda \end{pmatrix} =$$
$$(\lambda+1)^2(\lambda+1-i)(\lambda+1+i) = \lambda^4 + 4\lambda^3 + 7\lambda^2 + 6\lambda + 2.$$

Hence,

$$\lambda^4 + k_3\lambda^3 + (-25 + k_4)\lambda^2 + (-25k_3 - 0.6k_1)\lambda + (-0.6k_2 - 25k_4) = \lambda^4 + 4\lambda^3 + 7\lambda^2 + 6\lambda + 2,$$

which gives the solution

$$k_3 = 4, \quad k_4 = 32, \quad k_1 = -\frac{106}{0.6} \approx -176.67, \quad k_2 = -\frac{802}{0.6} \approx -1336.67.$$

The observer has the form

$$\dot{\hat{x}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix} \hat{x} + \begin{pmatrix} 0 \\ -2.4 \\ 0 \\ 1 \end{pmatrix} u + \begin{pmatrix} -176.67 \\ -1336.67 \\ 4 \\ 32 \end{pmatrix} \Big(y - (0 \;\; 0 \;\; 1 \;\; 0)\hat{x} \Big).$$

The solution of this observer, in combination with the solution of the original system, satisfies $\lim_{t\to\infty} \big(\hat{x}(t) - x(t)\big) = 0$. □

Theorem 5.11 states that observability is a sufficient condition for a pair (C, A) to be detectable. A necessary and sufficient condition for a pair (C, A) to be detectable is most

easily given when A and C are expressed with respect to a particular basis such that they have a form as in (4.17) and (4.18), respectively, i.e.,

$$\bar{A} = \begin{pmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{pmatrix}, \quad \bar{C} = (0 \ \bar{C}_2),$$

where the pair $(\bar{C}_2, \bar{A}_{22})$ is observable. Then, detectability of the pair (C,A) is equivalent to condition that the matrix $\bar{A}_{11}$ is asymptotically stable. See also Exercise 5.5.7.

Detectability can alternatively be investigated as in the following remark, see [Trentelman, Stoorvogel and Hautus, 2001], Chapter 3.

Remark 5.13 *Consider the pair (C,A), where A is a real $n \times n$ matrix and C a real $p \times n$ matrix. Then the following statements are equivalent.*

1. *The pair (C,A) is detectable.*
2. $\operatorname{rank}\begin{pmatrix} sI-A \\ C \end{pmatrix} = n$ *for all* $s \in \mathbb{C}$ *with* $\operatorname{Re} s \geq 0$.
3. $\operatorname{rank}\begin{pmatrix} \lambda I-A \\ C \end{pmatrix} = n$ *for all eigenvalues* λ *of matrix* A *with* $\operatorname{Re}\lambda \geq 0$.

5.3 Separation principle and compensators

Observers were introduced because of lack of knowledge of the whole state vector. This state vector was used in a feedback loop such as to give the system certain desired properties. We are now going to combine the feedback concept with that of the observer. Let $u = Fx$ be a feedback law that makes

$$\dot{x} = Ax + BFx = (A+BF)x$$

asymptotically stable. For the implementation of this feedback knowledge of the whole state is required. However, in many cases the whole state is not fully known, but only a partial state in the form of $y = Cx$. Assume further that an estimate $\hat{x}$ by an observer is available, then we have to be content with the control law $u = F\hat{x}$, instead of $u = Fx$. In this section we are going to see where such a control law leads to. Therefore, consider the system

$$\begin{array}{rcl} \dot{x} & = & Ax + Bu, \\ y & = & Cx, \end{array} \tag{5.7}$$

with $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$. As always the matrices A, B, C are real and of appropriate sizes.

Now let F be a real $m \times n$ matrix such that the eigenvalues of $A+BF$ are at desired locations and use the matrix in the control law

$$u = F\hat{x}, \tag{5.8}$$

where $\hat{x}$ is the state of an observer of the form

$$\begin{array}{rcl} \dot{\hat{x}} & = & A\hat{x} + Bu + K(y - \hat{y}), \\ \hat{y} & = & C\hat{x}, \end{array} \tag{5.9}$$

with $\hat{x} \in \mathbb{R}^n$ and with K a real $n \times p$ matrix such that the eigenvalues of $A - KC$ are at desired locations.

Then the combination of control law (5.8) and observer (5.9) results in a system of the form

$$\begin{array}{rcl} \dot{\hat{x}} & = & (A + BF - KC)\hat{x} + Ky, \\ u & = & F\hat{x}. \end{array} \tag{5.10}$$

The obtained system is called a (dynamic) **compensator**, or also (dynamic) **controller**. It has state $\hat{x}$ and is fed by measurements y of the system (5.7) and produces controls u for the system (5.7). Hence, such a compensator closes the 'loop' and makes that system and compensator together form a new system with no inputs and outputs anymore, see also Figure 5.4.

In order to distinguish the compensator (5.10) from the static compensator introduced earlier, the one in (5.10) is sometimes called a dynamic compensator, to indicate that the control produced at time t not only depends on output fed into the compensator at time t, but also on previous output values.

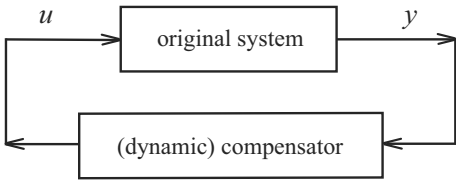

Figure 5.4 Schematic closed-loop system.

Connecting compensator (5.10) to system (5.7), using that $u = F\hat{x}$ and $y = Cx$, it follows that

$$\begin{array}{rcl} \dot{x} & = & Ax + BF\hat{x}, \\ \dot{\hat{x}} & = & KCx + (A + BF - KC)\hat{x}. \end{array}$$

As indicated above, the combined system has no inputs and outputs anymore, and has a state that is the combination of the state of the original system and the state of the compensator. Writing

$$x_{\mathrm{c}} = \begin{pmatrix} x \\ \hat{x} \end{pmatrix}, \quad A_{\mathrm{c}} = \begin{pmatrix} A & BF \\ KC & A + BF - KC \end{pmatrix},$$

with the subscript 'c' standing for 'combined', it follows that

$$\dot{x}_{\mathrm{c}} = A_{\mathrm{c}} x_{\mathrm{c}}.$$

To see how the combined system behaves, observe that

$$A_{\mathrm{c}} = \begin{pmatrix} I & 0 \\ I & -I \end{pmatrix} \begin{pmatrix} A + BF & -BF \\ 0 & A - KC \end{pmatrix} \begin{pmatrix} I & 0 \\ I & -I \end{pmatrix},$$

where I and 0 denote the identity and zero matrix, respectively, of size $n \times n$. Also note that left and right matrices, in the above right hand side, are each others inverse. It then follows that the characteristic polynomial of the matrix A_c satisfies

$$\begin{aligned} \det(\lambda I - A_c) &= \det\begin{pmatrix} \lambda I - (A+BF) & BF \\ 0 & \lambda I - (A-KC) \end{pmatrix} \\ &= \det(\lambda I - (A+BF)) \cdot \det(\lambda I - (A-KC), \end{aligned}$$

implying that the set of eigenvalues of A_c is the union of the sets of eigenvalues of $A+BF$ and $A-KC$, where in each set the multiplicities of the eigenvalues are taken into account.

The latter aspect can also be shown by noting that

$$\begin{pmatrix} x \\ e \end{pmatrix} = \begin{pmatrix} I & 0 \\ I & -I \end{pmatrix} \begin{pmatrix} x \\ \hat{x} \end{pmatrix},$$

with $e = x - \hat{x}$. Then expressing the equations in terms of x and e, instead of x and $\hat{x}$, it follows that

$$\begin{aligned} \dot{x} &= (A+BF)x - BFe, \\ \dot{e} &= (A-KC)e, \end{aligned} \tag{5.11}$$

or, equivalently,

$$\frac{d}{dt}\begin{pmatrix} x \\ e \end{pmatrix} = \begin{pmatrix} A+BF & -BF \\ 0 & A-KC \end{pmatrix} \begin{pmatrix} x \\ e \end{pmatrix}. \tag{5.12}$$

The set of eigenvalues of this system is equal to the union of the set of eigenvalues of $A+BF$ and the set of eigenvalues of $A-KC$. Since only a coordinate transformation has been performed, it once more follows that the eigenvalues of A_c are obtained by combining the eigenvalues of $A+BF$ and $A-KC$.

The previous illustrates that the eigenvalues of the overall system are equal to those obtained with a state feedback and those obtained by constructing a state observer. It is important to note that the feedback law and the observer can be designed independently! When putting together the original system and observer, with a feedback of the observer state, the eigenvalues do not interfere. This principle is called the **separation principle**.

The total system of the original system, observer and feedback law is summarized in the flow-diagram depicted in Figure 5.5. The two subsystems surrounded by a dotted line are the original system and the compensator.

Example 5.14 We will now conclude with the example of the carriage with the inverted pendulum. A state feedback law was designed in Example 5.5 and an observer in Example 5.12. Hence, the eigenvalues of $A+BF$ are as given in Example 5.5 and the eigenvalues of $A-KC$ are as in Example 5.12. However, these eigenvalues were chosen more or less at will.

In order to investigate the behavior of the combined system and the influence of the choice of the eigenvalues a number of simulations have been done. In the pictures in Figure 5.6 some of the results are depicted. In all experiments the (unknown) initial state

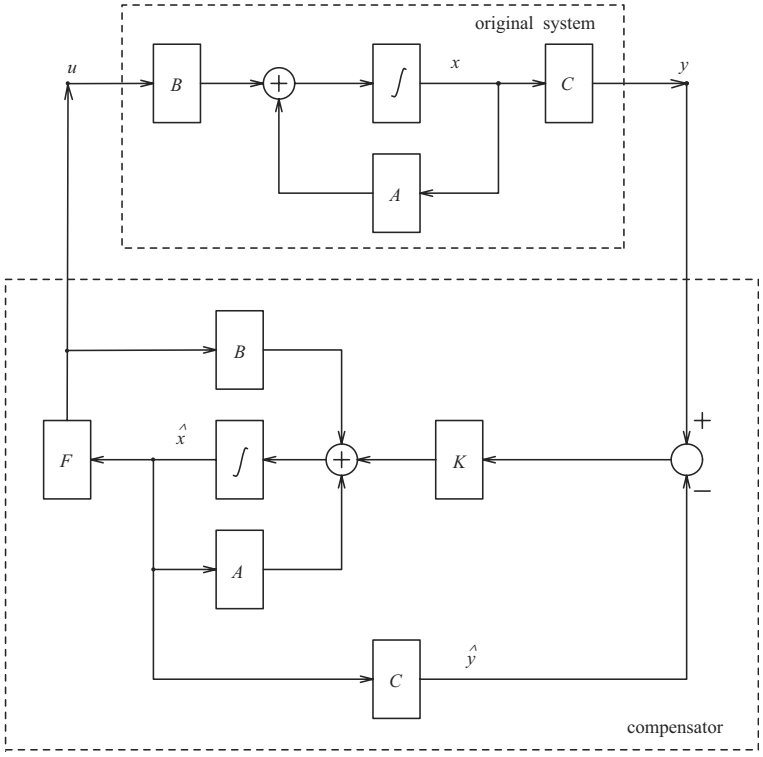

Figure 5.5 System and compensator in closed-loop.

and the initial value for the state estimate are taken to be

$$x(0) = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{x}(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

respectively. Hence, the pendulum does not start in its equilibrium position and needs to be controlled in order not to fall. Indeed, the angle of the pendulum with the vertical does not start at zero, as $x_1(0) = 2$. The control u will be based on the measurement y, being only x_3. To see the effect of the compensator, Figure 5.6 contains plots against time of x_1 and its estimate $\hat{x}_1$, for three cases.

Recall from Example 3.6 that x is the state in the linearized version of the inverted pendulum model. Hence, x must been seen as the deviation from the solution in the original nonlinear model around which the linearization has been done. As such $x_1(0) = 2$ must been seen as an initial deviation of φ of 2 'units' from $\tilde{\varphi}$, which has a stationary value equal to zero. Because of the linear character of the linearized model the magnitude of a 'unit' is irrelevant. However, in the scope of the present example it may be natural to choose degree $^\circ$ as unit for x_1.

- In the first experiment the eigenvalues of $A+BF$ and $A-KC$ are taken as specified

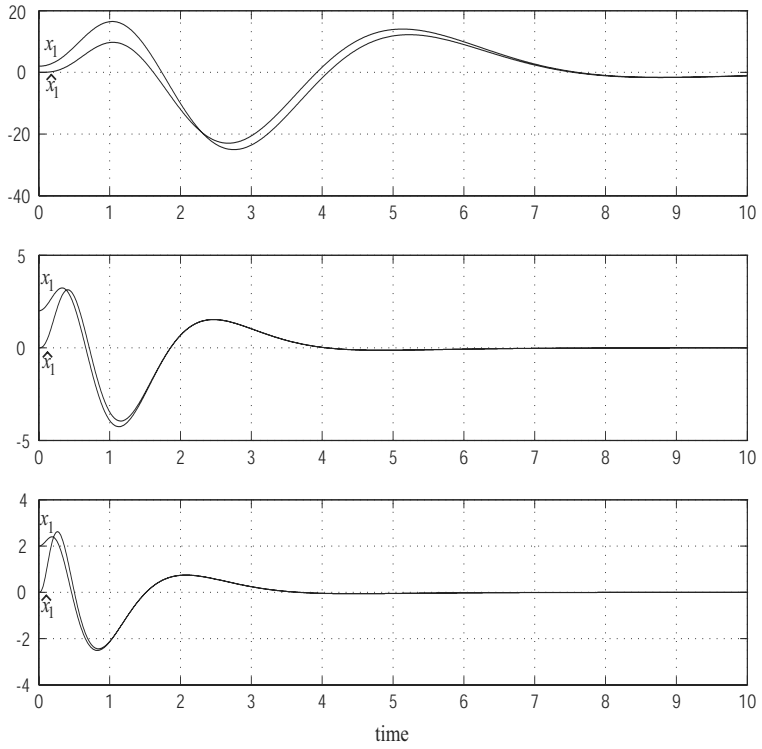

Figure 5.6 Simulation results for various choices of eigenvalues.

in Examples 5.5 and 5.12, respectively, i.e., $\sigma(A+BF) = \{-1,-2,-2\pm i\}$ and $\sigma(A-KC) = \{-1,-1,-1\pm i\}$, where $\sigma(M)$ denotes the set of eigenvalues of the square matrix M, taking the multiplicities into account. From the top plot of Figure 5.6 it is clear that $\hat{x}_1$ tends to look like x_1 as time passes by. From time 7 both signals more or less coincide. Also it can be seen that both signals tend to zero as time passes by. However, note that, compared to their initial values, both signals undergo large deviations, approximately ranging from -25 to $+20$. Also x_1 starts off in the wrong direction as the signal initially starts with increasing. Apparently, the control action is poor because it is based on an estimate of the state that initially is far from perfect. The control action becomes better once the estimate of the state starts to improve.

- In the second experiment the eigenvalues of $A+BF$ are unchanged, but the eigenvalues of $A-KC$ are shifted to the left over a distance of 5 by computing a new matrix K. Hence, $\sigma(A+BF) = \{-1,-2,-2\pm i\}$ and $\sigma(A-KC) = \{-6,-6,-6\pm i\}$. From the middle plot of Figure 5.6 it is clear that $\hat{x}_1$ tends to look like x_1 as time passes by. Now before time 2 both signals more or less coincide. It can also be seen that both signals go to zero as time passes by. Now note that both signals undergo smaller deviations, approximately ranging from -4 to $+4$. Again x_1 starts off in the

wrong direction, but is corrected faster than above as the state estimation is done quicker. The control action has improved because the estimate of the state on which it is based has become better sooner than in the first experiment.

- In the last experiment the eigenvalues for $A+BF$ are again unchanged, but the eigenvalues of $A-KC$ are now shifted to the left over a distance of 10 by using a suitable matrix K. Hence, $\sigma(A+BF)=\{-1,-2,-2\pm i\}$ and $\sigma(A-KC)=\{-11,-11,-11\pm i\}$. From the bottom plot of Figure 5.6 it is clear that $\hat{x}_1$ tends to look like x_1 as time passes by. Now before time 1 both signals more or less coincide. Also it can be seen that both signals go to zero as time passes by. Now note that both signals undergo again smaller deviations, approximately ranging from -3 to $+3$. Clearly, the control action has again improved because the estimate of the state on which it is based has become better even sooner than in the second experiment.

5.4 Disturbance rejection

Consider a linear time-invariant system with $m+l$ inputs, partitioned as (u,v), and $p+q$ outputs, partitioned as (y,z), described by

$$\dot{x}=Ax+Bu+Ev, \quad y=C_1x, \quad z=C_2x, \tag{5.13}$$

where $u\in\mathbb{R}^m$ is the usual control and $v\in\mathbb{R}^l$ is to be interpreted as a 'disturbance'. Further, $y\in\mathbb{R}^p$ is the usual measurement, while $z\in\mathbb{R}^q$ can be seen as an output that has to be regulated. For the sake of simplicity, we assume that $C_1=I$ and, hence, $y=x$ in this brief section. The disturbance cannot be measured directly (one only measures y) and the

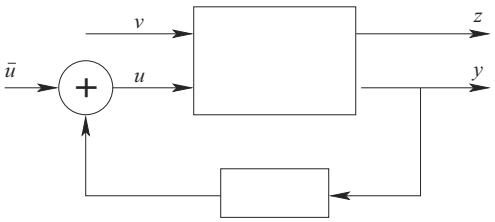

Figure 5.7 Closed loop system with disturbances.

objective is to design a feedback law

$$u=Fx+\bar{u},$$

(more generally, u is a function of y and of $\bar{u}$) such that v **has no effect whatsoever on the output** z, no matter what $\bar{u}$ or the initial condition $x(0)$ of (5.13) are. In Figure 5.7 the closed-loop system is depicted.

Example 5.15 Consider the system

$$\begin{aligned}\dot{x}_1 &= x_2+u,\\ \dot{x}_2 &= v,\\ z &= x_1.\end{aligned}$$

The disturbance is not decoupled from the output as is easily seen (take for instance $u \equiv 0$). If one applies the feedback law $u = -x_2 + \bar{u}$, however, one gets

$$z(t) = x_1(0) + \int_0^t \bar{u}(s)ds,$$

which is independent of v. □

In general terms, we wish to have that $z(t)$, defined by

$$z(t) = Ce^{(A+BF)t}x(0) + C\int_0^t e^{(A+BF)(t-s)}\big(B\bar{u}(s) + Ev(s)\big)ds,$$

is independent of v for some matrix F. This is equivalent to the requirement that

$$C\int_0^t e^{(A+BF)(t-s)}Ev(s)ds = 0,$$

for all functions v and all times $t \geq 0$. This requirement can shown to be equivalent to the requirement that F is such that $C(A+BF)^kE = 0$ for all $k \geq 0$. The problem of the existence and the computation of such F is known as the **disturbance rejection problem**. The equivalence of the requirements will not be shown here, but the proof resembles the proof of Lemma 4.15. More details can, for instance, be found in [Wonham, 1985].

5.5 Exercises

Exercise 5.5.1 *Consider the one dimensional model $\dot{x}(t) = u(t)$, $t \geq 0$, with $x(0) = 1$ and the following two options for the (one dimensional) control function:*

1. $u(t) = -e^{-t}$,
2. $u(t) = -x(t)$.

The first option refers to an open-loop control, the second one to a closed-loop control. Show that in both cases the state satisfies $x(t) = e^{-t}$. Which of the two control options would you prefer if there are disturbances in the initial condition of the system, i.e., $x(0) = 1 + \varepsilon$ with $\varepsilon \neq 0$, and if the aim of the control is to have $\lim_{t\to\infty} x(t) = 0$?

Exercise 5.5.2 *Show that the linear time-invariant system $\dot{x} = Ax + Bu$ (u is not necessarily a scalar, i.e., $m \geq 1$) is stabilizable if and only if its unstable subspace (see Definition 4.5) is contained in its controllable subspace ($\mathrm{im}R$, see text above Lemma 4.13). Hint: assume A and B are given with respect to a basis in $\mathbb{R}^n$ such that they have a form as in Equation (4.11).*

Exercise 5.5.3 *Consider the linear time-invariant system $\dot{x} = Ax + Bu$, with A an $n \times n$ matrix and B an $n \times 1$ matrix. Assume that the system is controllable and let $r(\lambda) = \lambda^n + r_{n-1}\lambda^{n-1} + \cdots + r_1\lambda + r_0$, with $r_i \in \mathbb{R}$, for $i = 0, 1, \ldots, n-1$. Prove that the feedback matrix F such that $\det(\lambda I - (A+BF)) = r(\lambda)$ can be determined by means of the next expression:*

$$F = -[0, \ldots, 0, 1]\left(B\ AB\ \cdots\ A^{n-1}B\right)^{-1} r(A),$$

where $r(A) = A^n + r_{n-1}A^{n-1} + \cdots + r_0 I$.

Exercise 5.5.4 *Consider the non-controllable realization*

$$\dot{x} = \begin{pmatrix} -2 & 1 & 0 & 0 \\ 0 & +2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} u.$$

Is this realization stabilizable? Is it possible to find a vector F such that the feedback law $u = Fx$ causes the eigenvalues of the feedback system to be situated at $-2,-2,-1,-1$, or at $-2,-2,-2,-1$, or at $-2,-2,-2,-2$?

Exercise 5.5.5 *Consider the equations of motion of an airplane in a vertical plane. See Figure 5.8. If the units are scaled appropriately (forward speed equal to one, for instance), then these equations are approximately*

$$\begin{aligned} \dot{\gamma} &= \sin\alpha, \\ \ddot{\theta} &= -(\alpha - u), \\ \dot{h} &= \sin\gamma, \end{aligned}$$

where

- *h is the height of the airplane with respect to a certain reference height,*
- *$\gamma = \theta - \alpha$ is the flight angle,*
- *θ is the angle between the reference axis of the airplane and the horizontal,*
- *u is the rudder control.*

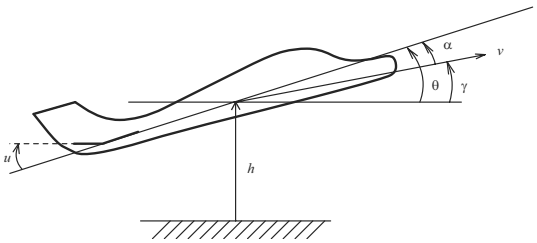

Figure 5.8 Airplane model.

One must design an automatic pilot to keep h constant (h should be kept stationary) in the presence of all kinds of perturbations such as vertical gusts.

- *Determine the stationary value of the other relevant variables.*
- *Linearize the equations of motion and write them as a set of first order differential equations.*
- *Show that the designer who proposes a feedback of the form $u = kh$, where k is a suitably chosen constant, cannot be succesful.*

- *Prove that a feedback of the form $u = k_1 h + k_2 \theta$, with suitably chosen constants k_1 and k_2, 'does the job', i.e., the resulting closed-loop system is asymptotically stable.*

Exercise 5.5.6 *Consider the dynamics of the satellite as given in Example 4.28. If only a scalar measurement is allowed (i.e., either y_1 or y_2) which one would you choose such that observability holds? Construct an observer for this measurement such that the eigenvalues of the matrix $A-KC$ are all situated in -1.*

Exercise 5.5.7 *Prove that the linear time-invariant system $\dot{x} = Ax + Bu$, $y = Cx$, is detectable if and only if its non-observable subspace is contained in its stable subspace (compare Exercise 5.5.2).*

Exercise 5.5.8 *Show that detectability is the dual concept of stabilizability, i.e., (A,B) is stabilizable if and only if $(B^\top, A^\top)$ is detectable.*

Exercise 5.5.9 *Consider Exercise 3.5.8 of the tractor. Show that if the combination of tractor and wagons moves in forward direction (with constant speed), then one has detectability if x_1 is observed (whereas the other x_i-values are not observed). If this combination would move in backward direction, then detectability is assured if of all state components only x_n is observed.*

Exercise 5.5.10 *Show that (5.12) can equivalently be written as*

$$\frac{d}{dt}\begin{pmatrix} \hat{x} \\ e \end{pmatrix} = \begin{pmatrix} A+BF & +KC \\ 0 & A-KC \end{pmatrix}\begin{pmatrix} \hat{x} \\ e \end{pmatrix},$$

where the relationship between $\hat{x}$, x and e is defined as usual, i.e., $e = x - \hat{x}$.

Exercise 5.5.11 *On the straight line connecting the earth with the moon a point (in Figure 5.9 indicated by L) exists where the gravitational force exerted by the earth on a satellite, with mass m, equals (i.e., neutralizes) the gravitational force exerted by the moon and the centrifugal force (due to the rotation of the satellite around the earth). The equations of motion of the satellite in the neighborhood of L are*

$$\begin{aligned} \ddot{x} - 2\omega\dot{y} - 9\omega^2 x &= 0, \\ \ddot{y} + 2\omega\dot{x} + 4\omega^2 y &= u, \end{aligned}$$

where $u = F/(m\omega^2)$. On its turn, F is the force, exerted by a rocket, on the satellite in the y-direction. Moreover, $\omega = \frac{2\pi}{29}$ radians/day.

1. *Write the system as a linear dynamical system in first order form and show that the equilibrium point $x = \dot{x} = y = \dot{y} = 0$ is unstable.*
2. *Investigate the controllability and/or stabilizability of this system.*
3. *Determine a linear state feedback such that the eigenvalues of the closed-loop system are located in $-3\omega, -4\omega, -3\omega \pm 3\omega i$.*

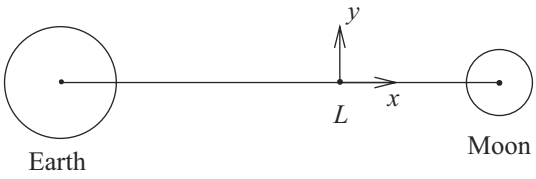

Figure 5.9 Earth-moon model.

4. *Suppose that only y is available for measurements. Is it possible to stabilize the system by means of an output feedback $u(t) = \alpha y(t)$? (Answer: no, it is not possible).*

Exercise 5.5.12 [Disturbance rejection] *Consider the model*

$$\frac{d}{dt}\begin{pmatrix} \varphi \\ \dot{\varphi} \\ \nu \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -2 & 1 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} \varphi \\ \dot{\varphi} \\ \nu \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\delta + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} f,$$

$$y = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} \varphi \\ \dot{\varphi} \\ \nu \end{pmatrix},$$

which describes the movement of a ship. See also Figure 5.10. The variable φ is the roll angle, $\dot{\varphi}$ its time derivative and ν is the lateral velocity. The control δ represents the rudder angle and the function f represents the (unknown) influence of the lateral waves on the ship movement. Please note that other possible movements of the ship, such as pitching and yawing, are not included in this simple model.

A time-varying roll angle φ causes sea-sickness and one wants to design a feedback law $\delta = Fx$, where $x^\top \stackrel{\text{def}}{=} (\varphi, \dot{\varphi}, \nu)$, such that φ is (completely) independent of the function f, whatever its values may be. Is it possible to construct such a matrix F? To this end, parametrize F and investigate whether the controllable subspace characterized by the matrix pair $(A+BF, E)$, where

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -2 & 1 \\ 0 & 0 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad E = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

is contained in the kernel of C, where $C = (1\ 0\ 0)$.

Exercise 5.5.13 *The following system is given.*

$$\begin{aligned} \dot{x} &= \begin{pmatrix} -1 & 0 & 2 \\ 0 & -3 & 0 \\ 1 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} u, \\ y &= \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} x. \end{aligned}$$

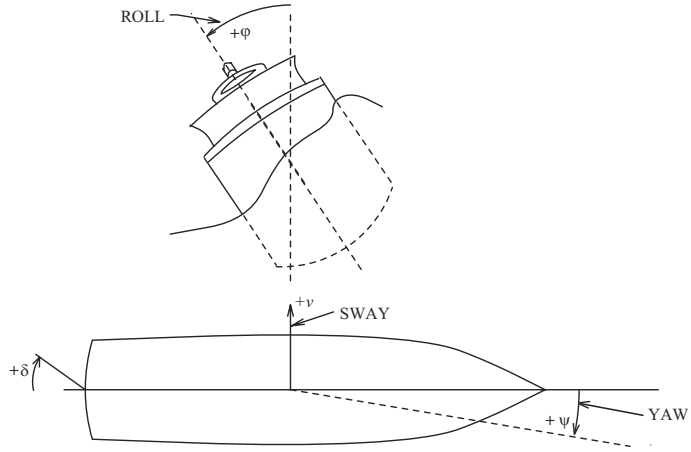

Figure 5.10 Ship model.

1. *Show that the system with $u = 0$ is not stable.*
2. *Show that the system is stabilizable.*
3. *Compute a feedback $u = Fx$ such that all eigenvalues of the closed-loop system are located in -1, -2, en -3.*
4. *Show that the system is not observable.*
5. *Is the system detectable? Explain your answer.*

Exercise 5.5.14 *Consider the linear time-invariant system $\dot{x} = Ax + Bu$, $y = Cx$ with*

$$A = \begin{pmatrix} -2 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -2 \end{pmatrix}, B = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, C = \begin{pmatrix} 2 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

In the following use a computer package like MATLAB, MAPLE®[1], *or so.*

1. *Compute the eigenvalues of A. Is the system (asymptotically) stable?*
2. *Compute a matrix F such that the eigenvalues of $A + BF$ are located at $-1, -2 \pm i, -3 \pm 2i$. You can use Exercise 5.5.3 for this. Explain how controllability plays a role in these computations.*
3. *Compute a matrix K such that the eigenvalues of $A - KC$ are located at $-3, -4, -5, -5 \pm 2i$. How can you use Exercise 5.5.3 now? What is the role of observability?*
4. *Give a state observer for the system using the above computed matrix K.*

[1] MAPLE is a registered trademark of Waterloo Maple Inc

5. *Combine the state observer with the above computed matrix F to obtain a dynamic compensator for the system.*

6. *Illustrate the separation principle by computing the eigenvalues of the matrix describing the interconnection of the system and the dynamic compensator.*

Exercise 5.5.15 *Below you will find a number of statements. For each of statements determine whether it is true or false. Make your answer plausible by means of a simple reasoning or (counter)example.*

1. *If the pair (A,B) is a stabilizable, then so is the pair $(A+BF,B)$ for any suitable matrix F.*

2. *If the sets of eigenvalues of A and $A+BF$ are completely disjoint, the pair (A,B) is controllable.*

3. *If $-A$ is asymptotically stable and the pair (A,B) is stabilizable, then the pair (A,B) is also controllable.*

4. *If B is an $n \times m$ matrix and A is an $n \times n$ matrix such that rank $A < n-m$, then the pair (A,B) can not be stabilizable.*

5. *Let A be an $n \times n$ matrix and C a $p \times n$ matrix. Then the pair (C,A) is detectable if and only if the pair (C,A^3) is detectable.*

6. *Let A be an $n \times n$ matrix and C a $p \times n$ matrix. Then (C,A) is a detectable pair if and only if the eigenvalues of $A-KC$ can be placed everywhere in the open left half-plane by a suitable choice of the $n \times p$ matrix K.*

7. *Every system that is stabilizable, is also detectable.*

8. *Every system that is detectable, is also observable.*

9. *Every system that is controllable, is also stabilizable.*

10. *Every system that is stabilizable, is also asymptotically stable.*

11. *Every system that is asymptotically stable, is also detectable.*

12. *Stabilizability and detectability for linear time-invariant systems are dual notions.*

13. *The minimal realization of a system is stabilizable and detectable.*

14. *A system that is stabilizable and observable, does not have to be a minimal system.*

15. *If the pair (A,B) is controllable, then so is the pair $(A+BKC,B)$ for any suitable matrix K.*

16. *If the pair (C,A) is detectable, then so is the pair $(C,A-BKC)$ for any suitable matrix K.*

Chapter 6

Input/Output Representations

The input/output representation of a system refers to a description where the input is directly related to the output, without other intermediate functions or variables such as the 'state'. We already have encountered such a description in Section 3.4 on impulse response functions or matrices. By means of the function $K(t,s)$ the input function was directly related to the output function. The description was obtained by the elimination of the state vector. In this chapter, and in Chapter 8.2, other useful input/output representations of systems will be discussed.

6.1 Laplace transforms and their use for linear time-invariant systems

Until now most of the systems were described by means of differential equations and algebraic equations with time as the underlying independent parameter. For this reason such systems are also said to be described in the time domain. Especially, linear time-invariant systems were studied in the previous chapters. However, for linear time-invariant systems also other ways of description exist as well. For instance, using the Laplace transform the linear equations of such systems can be transformed into new linear equations in the Laplace domain. There the underlying independent parameter can be interpreted as a complex-valued frequency and for that reason the description in the Laplace domain is also called a description in the frequency domain. In the current section the details of the description of a linear time-invariant system in the frequency domain will be presented

The Laplace transform of a piecewise continuous function $f : [0,\infty) \to \mathbb{R}$, denoted as $F = \mathscr{L}(f)$, is defined as

$$F(s) = \mathscr{L}(f(t)) = \int_0^\infty f(t)e^{-st}dt. \tag{6.1}$$

If $f = O(e^{bt})$ for $t \to \infty$, i.e., f grows (at most) at an exponential rate ($b \in \mathbb{R}$ is a constant), then the integral exists, not only for all real $s > b$, but also for all complex s with $\mathrm{Re}\, s > b$. The latter is due to the identity $|f(t)e^{-st}| = |f(t)|e^{-(\mathrm{Re}\, s)t}$. Therefore, the domain of the function F can be extended to all $s \in \mathbb{C}$ with $\mathrm{Re}\, s > b$, yielding that F then is a complex valued function such that

$$F : \{s \in \mathbb{C} | \mathrm{Re}\, s > b\} \to \mathbb{C}.$$

In this chapter the parameter s will always be complex valued. The extension of the previous to vector valued functions $f : [0,\infty) \to \mathbb{R}^n$ is straightforward.

$$\mathscr{L}(f(t)) = \big(\mathscr{L}(f_1(t)),\ldots,\mathscr{L}(f_n(t))\big)^\top = \big(F_1(s),\ldots,F_n(s)\big)^\top = F(s).$$

The extension to matrix valued functions is also componentwise.

Consider a linear time-invariant strictly causal differential system given by its impulse response matrix G, see Section 3.4. Then the relation between input u and output y is given by

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

For simplicity we assume $u(\tau) = 0$ for $\tau \leq 0$ and hence

$$y(t) = \int_{0}^{t} G(t-\tau)u(\tau)d\tau. \tag{6.2}$$

Hence, the output y can be seen as the **convolution** of the impulse response G and the input u, both possibly matrix/vector valued. Suppose that y, u and G have Laplace transforms, to be denoted by Y, U and H, respectively, i.e.,

$$Y(s) = \int_{0}^{\infty} y(t)e^{-st}dt, \quad U(s) = \int_{0}^{\infty} u(t)e^{-st}dt, \quad H(s) = \int_{0}^{\infty} G(t)e^{-st}dt,$$

then the transformation of (6.2) yields

$$Y(s) = H(s)U(s). \tag{6.3}$$

The $p \times m$ matrix $H(s)$ is called the **transfer matrix** of the system. It gives a very simple description of the system. The property that (6.3) is the Laplace transform of (6.2) is called the **convolution theorem**. It is assumed that the reader is familiar with this property and, more generally, with the theory of Laplace transforms, for more details, see [3].

If $G(t) = O(e^{bt})$ for $t \to \infty$, the transfer matrix is only defined for $\mathrm{Re}\, s > b$. The theory of Laplace transforms tells us that $H(s)$ is analytic for $\mathrm{Re}\, s > b$ and then complex function theory tells us that a unique analytic continuation of $H(s)$ exists. A unique matrix exists for all $s \in \mathbb{C}$, that is analytic in the complex plane, except for a number of isolated points, and that is identical to $H(s)$ for $\mathrm{Re}\, s > b$. In the remainder we will not distinguish between $H(s)$ and its analytic continuation.

If $X(s)$ is the Laplace transform of $x(t)$, then

$$\mathscr{L}(\dot{x}(t)) = \int_{0}^{\infty} \dot{x}(t)e^{-st}dt = \left[x(t)e^{-st}\right]_{0}^{\infty} + \int_{0}^{\infty} x(t)se^{-st}dt = -x(0) + sX(s).$$

The Laplace transform of the equation

$$\dot{x} = Ax + Bu, \quad x(0) = x_0,$$

therefore is

$$sX(s) - x_0 = AX(s) + BU(s), \tag{6.4}$$

yielding

$$X(s) = (sI - A)^{-1}x_0 + (sI - A)^{-1}BU(s).$$

If we also transform the output equation $y = Cx$ to $Y(s) = CX(s)$, and assume that $x_0 = 0$, then

$$Y(s) = C(sI - A)^{-1}BU(s) = H(s)U(s). \tag{6.5}$$

Recall that the impulse response equals $G(t) = Ce^{At}B$. Comparison with

$$y(t) = \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau,$$

see Equation (6.2), leads to

$$H(s) = \mathscr{L}(Ce^{At}B) = C(Is - A)^{-1}B. \tag{6.6}$$

The theory of analytic continuation can be illustrated clearly with respect to the latter equation. At first instance $H(s)$ is only defined for $\operatorname{Re} s > \max(\operatorname{Re}\lambda_i)$, where λ_i are the eigenvalues of A. The expression $C(Is - A)^{-1}B$, however, is well defined for all $s \in \mathbb{C}$, except for possibly the points $s = \lambda_1, \lambda_2, \ldots, \lambda_n$, where $Is - A$ is singular. Please note that it is not necessarily true that all eigenvalues of A cause $H(s)$ to be singular, since, by multiplying $(Is - A)^{-1}$ with C and B, some factors may cancel. In system theory, points where $H(s)$ does not exist are called the **poles** of the transfer function $H(s)$. Equation (6.6) states that the transfer matrix is the Laplace transform of the impulse response matrix.

Example 6.1 Consider the system which describes the dynamics of the satellite (see also Example 3.14 and Exercise 3.5.2). The system is given by

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The transfer matrix for this system is

$$H(s) = \mathscr{L}(G(t)) = \mathscr{L}\left(\begin{pmatrix} \sin t & 2 - 2\cos t \\ -2 + 2\cos t & -3t + 4\sin t \end{pmatrix}\right) =$$

$$\begin{pmatrix} \dfrac{1}{s^2+1} & \dfrac{2}{s} - \dfrac{2s}{s^2+1} \\ -\dfrac{2}{s} + \dfrac{2s}{s^2+1} & -\dfrac{3}{s^2} + \dfrac{4}{s^2+1} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{s^2+1} & \dfrac{2}{s^3+s} \\ \dfrac{-2}{s^3+s} & \dfrac{s^2-3}{s^4+s^2} \end{pmatrix}.$$

□

A new method has now been found to calculate the transition matrix. The Laplace transforms of $\dot{x} = Ax$ with $x(0) = x_0$ and $x(t) = e^{At}x_0$ are

$$X(s) = (sI - A)^{-1}x_0, \quad X(s) = \mathscr{L}(e^{At})x_0,$$

respectively, for any $x_0 \in \mathbb{R}^n$. Therefore, it follows

$$e^{At} = \mathscr{L}^{-1}((sI - A)^{-1}),$$

where $\mathscr{L}^{-1}$ denotes the inverse Laplace transform. The matrix function $(sI - A)^{-1}$ is called the **resolvente** of the matrix A.

6.2 Connection of systems

The description of systems by means of transfer matrices is useful if one wants to connect systems. If we are given two systems by means of the transfer matrices $H_1(s)$ and $H_2(s)$, respectively, as depicted in Figure 6.1, then the **parallel connection** is given as shown

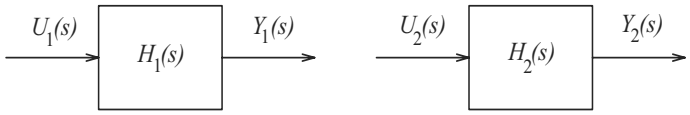

Figure 6.1 Two subsystems with transfer matrices $H_1(s)$ and $H_2(s)$.

in Figure 6.2, where the symbol $\oplus$ denotes addition of all incoming signals. Note that

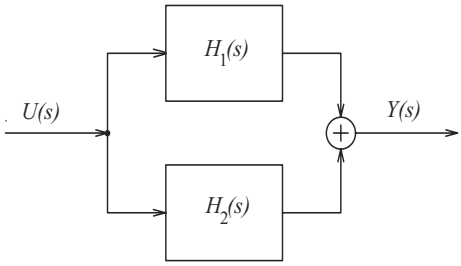

Figure 6.2 Parallel connection of systems described by $H_1(s)$ and $H_2(s)$.

in the parallel connection both subsystems have the same input, and that the output of the parallel connection is the sum of the outputs of its two subsystems. In formula, the transfer matrix of the parallel connection, denoted by $H(s)$, is equal to $H_1(s)+H_2(s)$.

The **series connection** is as depicted in Figure 6.3. Note that in the series connection the output of one subsystem acts as the input to the other subsystem. Hence, these input and output need to be of the same size. In formula, $H(s) = H_2(s)H_1(s)$, where $H(s)$ now

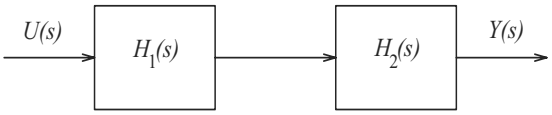

Figure 6.3 Series connection of systems described by $H_1(s)$ and $H_2(s)$.

denotes the transfer matrix of the series connection. Please note that for multi-input multi-output systems, the product of the *matrices* $H_2(s)$ and $H_1(s)$ is in general not commutative, i.e., in general $H_1(s)H_2(s) \neq H_2(s)H_1(s)$, such that the order in which the systems are connected is important. The reader may convince him/herself that the description of a series connection, starting from two state space descriptions, is far more difficult.

The **feedback connection** is given as depicted in Figure 6.4. In frequency domain terms, if the signal that enters $H_1(s)$ is called $V(s)$, then the transfer matrix of the overall

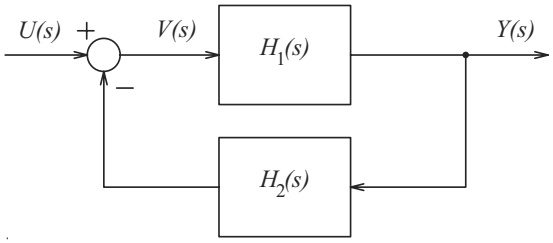

Figure 6.4 Feedback connection of systems described by $H_1(s)$ and $H_2(s)$.

system, again denoted by $H(s)$, can be calculated as follows.

$$\left.\begin{array}{rcl} V(s) & = & U(s)-H_2(s)Y(s) \\ Y(s) & = & H_1(s)V(s) \end{array}\right\} \Rightarrow Y(s)=H_1(s)(U(s)-H_2(s)Y(s)).$$

Solving for $Y(s)$ yields

$$Y(s)=(I+H_1(s)H_2(s))^{-1}H_1(s)U(s),$$

and therefore

$$H(s)=(I+H_1(s)H_2(s))^{-1}H_1(s). \tag{6.7}$$

In the connections considered above it was tacitly assumed that the number of inputs and the number of outputs were such that the described connections made sense.

6.3 Rational functions

Let us consider the transfer matrix $H(s)=\mathscr{L}(G(t))=C(Is-A)^{-1}B$ in more detail. The inverse $(Is-A)^{-1}$ can in principle be obtained by applying Cramer's rule, the result of which is

$$(Is-A)^{-1}=\frac{1}{p(s)}\begin{pmatrix} q_{11}(s) & \cdots & q_{1n}(s) \\ \vdots & & \vdots \\ q_{n1}(s) & \cdots & q_{nn}(s) \end{pmatrix},$$

where $p(s)$ is the characteristic polynomial of A. We write $p(s)$ as

$$p(s)=s^n+p_{n-1}s^{n-1}+\cdots+p_1s+p_0$$

with $p_0,p_1,\ldots,p_{n-1}\in\mathbb{R}$. For all $i,j=1,2,\ldots,n$, the elements $q_{ij}(s)$ are determinants of $(n-1)\times(n-1)$ submatrices of $Is-A$, and consequently are polynomials in s of degree at most $n-1$. Therefore, the elements of $(Is-A)^{-1}$ are rational functions of s, i.e., functions of the form $\dfrac{q_{ij}(s)}{p(s)}$.

In general, a **rational function** is defined as the quotient of two polynomials. It is called **strictly proper** if the degree of the numerator polynomial is smaller than the degree of the denominator polynomial. If the rational function is given by $h(s)$, then an equivalent definition of being strictly proper is that $\lim_{|s|\to\infty}h(s)=0$. If this limit is finite,

but not necessarily zero, then one speaks of a **proper** rational function. Written as a quotient of two polynomials, a rational function is proper if and only if the degree of the numerator polynomial is less than or equal to the degree of the denominator polynomial.

It easily follows that the elements of a transfer matrix $H(s)$ are strictly proper rational functions. Indeed, from the above it follows that $H(s)$ can be written as $\frac{1}{p(s)}R(s)$, where $R(s)$ is an $p \times m$ matrix with as elements polynomials of degree at most $n-1$, and with $p(s)$ a scalar polynomial of which the degree is n. As defined earlier, **poles** of $H(s)$ are points where $H(s)$ has a singularity, i.e., points s_0 where $\lim_{s\to s_0} H(s)$ does not exist. The eigenvalues of A, being the roots of $p(s)$, are the only candidates for poles, but not necessarily all of them are poles.

Example 6.2 If

$$A = \begin{pmatrix} 0 & -2 \\ 1 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 \end{pmatrix},$$

then

$$(Is-A)^{-1} = \begin{pmatrix} s & 2 \\ -1 & s+3 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{s+3}{(s+1)(s+2)} & \frac{-2}{(s+1)(s+2)} \\ \frac{1}{(s+1)(s+2)} & \frac{s}{(s+1)(s+2)} \end{pmatrix}.$$

The matrix $(Is-A)^{-1}$ has poles in $s=-1$ and $s=-2$. However,

$$C(Is-A)^{-1}B = \frac{1}{s+1}$$

and has only one pole, namely in $s=-1$. □

In the context of transfer matrices, the underlying systems are said to be stable if all of the associated poles have a negative real part. In fact, compared with notions of stability in Section 4.1, it would have been more consistent to call these systems asymptotically stable. However, it is common practice to say that a system described by a transfer matrix is stable if all its poles have a negative real part. Note that this notion of stability is closely related to the BIBO stability introduced in Section 4.1.

Example 6.3 We are given two linear, stable, single-input single-output systems Σ_1 and Σ_2, with transfer matrices (actually transfer functions) $h_1(s)$ and $h_2(s)$, respectively. Prove, or, if not true, give a counterexample for, each of the following three assertions.

1. The series connection is stable.
2. The feedback connection is stable.
3. The parallel connection is stable.

Answer question 1. For $i=1,2$, suppose that $h_i(s) = \frac{q_i(s)}{p_i(s)}$, and that common factors have been deleted, so that the poles of system Σ_i are the roots of $p_i(s)$. Stability now

means that these roots are located in the open left half-plane, see also the above remark. The transfer function of the series connection equals $\frac{q_1(s)q_2(s)}{p_1(s)p_2(s)}$. Perhaps some factors in numerator and denominator will cancel, but the roots of the remaining denominator form a subset of the set of roots of $p_1(s)$ and $p_2(s)$, and therefore lie in the open left half-plane. Hence, the series connection is stable.

Answer question 2. The feedback connection is not necessarily stable as shown by the (counter)example. Consider

$$h_1(s) = \frac{1}{s+1}, \quad h_2(s) = \frac{-4}{s+1}.$$

The transfer function of the feedback connection is

$$h(s) = \frac{h_1(s)}{1+h_1(s)h_2(s)} = \frac{(s+1)}{(s+1)^2-4} = \frac{(s+1)}{(s+3)(s-1)},$$

which represents an unstable system, since it has a pole with nonnegative real part at $s=1$.

Answer question 3. The parallel connection is stable again. A proof can be given along the same lines as the proof of the stability of the series connection. □

So far we have concentrated on strictly causal linear systems, i.e., with $D=0$. If $D \neq 0$, and assuming $x_0 = 0$, then (see also Exercise 3.5.16),

$$\begin{aligned} y(t) &= \int_0^t \Big(Ce^{A(t-\tau)}B + D\delta(t-\tau) \Big) u(\tau) d\tau, \\ H(s) &= C(sI-A)^{-1}B + \mathscr{L}\{D\delta(t)\} = C(sI-A)^{-1}B + D. \end{aligned}$$

If we consider this transfer matrix in detail, it turns out that its elements are proper functions, which are not all strictly proper, because the degree of the numerator of at least one element will now be equal to the degree of the denominator (otherwise we would have that $D=0$).

The following example shows that also transfer matrices exist of which the elements are not rational functions at all.

Example 6.4 The transfer function for the moving average system, treated in Example 3.15, is

$$H(s) = \int_0^\infty G(t)e^{-st}dt = \frac{1}{T}\int_0^T 1 \cdot e^{-st}dt = \frac{1-e^{-sT}}{sT}.$$

This is not a rational function. □

It can be shown that for all proper rational transfer matrices $H(s)$, matrices A, B, C and D exist such that $H(s) = C(Is-A)^{-1}B + D$. Hence, to such transfer matrices linear time-invariant differential systems correspond. In the next section the latter will be proved for transfer functions, to which single-input single-output systems can be associated.

6.4 Transfer functions and transfer matrices

In this section we will mainly consider single-input single-output linear differential systems. The transfer matrix is therefore a scalar function, called the transfer function, and it will be indicated by $h(s)$ instead of the more general $H(s)$, which will be used later on to indicate the multi-input multi-output case. In this section we will also assume that $h(s)$ is proper, i.e., the degree of the numerator is less than or equal to the degree of the denominator. Without loss of generality $h(s)$ can be written more explicitly as

$$h(s) = \frac{q(s)}{p(s)} = \frac{q_k s^k + q_{k-1} s^{k-1} + \cdots + q_0}{s^n + p_{n-1} s^{n-1} + \cdots + p_0}, \tag{6.8}$$

with $k \leq n$, and where the coefficient of the highest power of s in the denominator is equal to one. (Polynomials with the coefficient of the highest power equal to one are called **monic** polynomials.) It is well known that a polynomial can be factorized in a number of linear factors equal to the degree of the polynomial. Hence, we can write

$$h(s) = \frac{q(s)}{p(s)} = \frac{c(s-b_1)(s-b_2)\cdots(s-b_k)}{(s-a_1)(s-a_2)\cdots(s-a_n)}, \tag{6.9}$$

with $c \in \mathbb{R}$ and $a_i, b_i \in \mathbb{C}$ for $i = 1,2,\ldots,k$, where $k \leq n$. We will assume that $q(s)$ and $p(s)$ do not have any common factors. If so, they can be cancelled. The roots of the denominator $p(s)$, i.e., $a_1, a_2, \ldots, a_n$, are called the **poles** of the transfer function, and $b_1, b_2, \ldots, b_k$, i.e., the roots of $q(s)$, are called the **zeros** of the transfer function. Suppose the input is given by

$$u(t) = \begin{cases} e^{s_0 t} & \text{for} \quad t \geq 0, \\ 0 & \text{for} \quad t < 0, \end{cases}$$

then the Laplace transform of the output can be written as

$$Y(s) = \frac{c(s-b_1)\cdots(s-b_k)}{(s-a_1)\cdots(s-a_n)} \cdot \frac{1}{(s-s_0)}.$$

If $s_0 \neq b_i$, for $i = 1,\ldots,k$, then a partial fraction decomposition of $Y(s)$ yields

$$Y(s) = \frac{\gamma_1}{s-a_1} + \frac{\gamma_2}{s-a_2} + \cdots + \frac{\gamma_n}{s-a_n} + \frac{\gamma_{n+1}}{s-s_0}, \quad \gamma_i \in \mathbb{C}, \tag{6.10}$$

where, for reason of simplicity, we assumed that all poles a_i have multiplicity one, and, moreover, that $s_0 \neq a_i$, for all $i = 1,2\ldots,n$. The inverse Laplace transform of (6.10) yields

$$y(t) = \gamma_1 e^{a_1 t} + \cdots + \gamma_n e^{a_n t} + \gamma_{n+1} e^{s_0 t}.$$

The first n terms of the right-hand side of this expression are the free modes of the system. The last term is a consequence of the input.

If now $s_0 = b_i$, for some i, say $i = 1$, then

$$\begin{aligned} Y(s) &= \frac{c(s-b_1)\cdots(s-b_k)}{(s-a_1)\cdots(s-a_n)} \cdot \frac{1}{(s-b_1)} = \frac{c(s-b_2)\cdots(s-b_k)}{(s-a_1)\cdots(s-a_n)} = \\ &= \frac{\gamma_1}{s-a_1} + \cdots + \frac{\gamma_n}{s-a_n}, \quad \gamma_i \in \mathbb{C}. \end{aligned}$$

The frequency s_0 of the input signal does not show up in the output signal; only the free modes are excited. So it follows that *the zeros of a system are those frequencies in the input signal which do not form part of the output signal*.

Definition 6.5 *If all eigenvalues λ_i have a negative real part, the* **time constant** σ *of the corresponding system is defined as* $\sigma^{-1} = -\max_i\{\operatorname{Re}\lambda_i\}$.

Definition 6.6 *The single-input single-output system $\dot{x} = Ax + Bu$, $y = Cx$ is said to be a* **non-minimum phase** *system if at least one of its zeros has positive real part.*

Example 6.7 Consider the system with transfer function

$$\frac{-s+1}{s^2+5s+6} = \frac{3}{s+2} + \frac{-4}{s+3}.$$

This is a non-minimum phase system. If the Heaviside function is applied to the system, which was at rest for $t \leq 0$, then it is straightforward to show that the output is

$$y(t) = \frac{3}{2}(1-e^{-2t}) - \frac{4}{3}(1-e^{-3t}), \quad t \geq 0.$$

Of course $y(0) = 0$, and one sees that $y(\infty) = 1/6 > 0$. So, a positive input leads to a positive output in the long run. For a stabilizing output feedback it is therefore tempting to consider $u(t) = ky(t)$, with $k < 0$, to counteract the output with the input. However, one also has $\dot{y}(0) = -1 < 0$. Hence, the sign of $y(t)$ for small values of t is different from the sign of $y(t)$ for large values of t. This is sometimes felt to be counter-intuitive and leads to problems if one wants to apply an output feedback control of the form $u(t) = ky(t)$. Hence, non-minimum phase systems require careful attention if one wants to apply such an output feedback. □

Example 6.8 Continuation of the satellite example (see Examples 3.14 and 6.1, and Exercise 3.5.2). We consider a version of the dynamics where there is only one input variable and one output variable, namely u_2 and y_2, respectively. The matrices involved are (with $\omega = 1$):

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad C = (0\ 0\ 1\ 0).$$

The transfer function of this system is (see Example 6.1):

$$\frac{s^2-3}{s^4+s^2}$$

The zeros of this system are $s = +\sqrt{3}$ and $s = -\sqrt{3}$. These 'frequencies' (strictly speaking, there is no oscillation at all, since $s = \pm\sqrt{3}$ correspond to real exponential functions) cannot appear in a component of the output signal. However, because the system is not stable, see Exercise 4.5.1, the free modes excited by the input will not die out. □

We already know that a single-input single-output system $\dot{x} = Ax + Bu$, $y = Cx + Du$, gives rise to a transfer function (a 1×1 matrix)

$$h(s) = C(Is - A)^{-1}B + D, \tag{6.11}$$

which is a proper rational function. In the following theorem it will be shown that the reverse also holds.

Theorem 6.9 *Consider a transfer function $h(s)$ and assume that it is a proper rational function. Then there exists an $n \times n$ matrix A, an $n \times 1$ matrix B, a $1 \times n$ matrix C and a 1×1 matrix D such that (6.11) is satisfied.*

Proof Assume a rational function $h(s) = \dfrac{q(s)}{p(s)}$ is given with $\deg q(s) \leq \deg p(s)$, where $\deg q(s)$ denotes the **degree** of the polynomial $q(s)$, and similarly for $\deg p(s)$. First $D \in \mathbb{R}$ will be constructed. There are two possibilities:

1. if $\deg q(s) < \deg p(s)$, then take $D = 0$.
2. if $\deg q(s) = \deg p(s)$, then

$$\begin{aligned} h(s) &= \frac{q(s)}{p(s)} = \frac{q_n s^n + q_{n-1}s^{n-1} + \cdots + q_0}{s^n + p_{n-1}s^{n-1} + \cdots + p_0} \\ &= \frac{q_n(s^n + p_{n-1}s^{n-1} + \cdots + p_0)}{p(s)} + \\ &\qquad \frac{(q_{n-1} - q_n p_{n-1})s^{n-1} + \cdots + (q_0 - q_n p_0)}{p(s)} \\ &= q_n + \frac{\bar{q}(s)}{p(s)}, \end{aligned} \tag{6.12}$$

where $\deg \bar{q}(s) < \deg p(s)$. Take $D = q_n$ in this case.

In order not to complicate the notation, $\bar{q}(s)$ will again be written as $q(s)$, so that we can continue with $\dfrac{q(s)}{p(s)}$, $\deg q(s) < \deg p(s)$, and with D already defined. Hence, we will write

$$p(s) = s^n + p_{n-1}s^{n-1} + \cdots + p_0, \quad q(s) = q_{n-1}s^{n-1} + \cdots + q_0.$$

If Y and U are the Laplace transforms of y and u, respectively, then they are connected according to $Y(s) = h(s)U(s)$, or, equivalently,

$$p(s)Y(s) = q(s)U(s),$$

which is a shorthand notation for

$$s^n Y(s) + p_{n-1}s^{n-1}Y(s) + \cdots + p_0 Y(s) = q_{n-1}s^{n-1}U(s) + \cdots + q_0 U(s). \tag{6.13}$$

We start with a special polynomial $q(s)$, namely $q(s)$ is a constant and for this constant we choose 1. Hence, $q(s) = q_0 = 1$. Since such a system is different from the original one,

we will denote its output by z, with Laplace transform Z, instead of y, which is preserved for the output of the original system. Then

$$s^n Z(s) + p_{n-1}s^{n-1}Z(s) + \cdots + p_0 Z(s) = U(s),$$

which is the Laplace transform of

$$\frac{d^n}{dt^n}z(t) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}z(t) + \cdots + p_0 z(t) = u(t), \tag{6.14}$$

with initial values $z(0) = \dot{z}(0) = \cdots = z^{(n-1)}(0) = 0$. Here we used the following properties of Laplace transforms of derivatives.

$$\begin{aligned} \mathscr{L}(f'(t)) &= s\mathscr{L}(f(t)) - f(0), \\ \mathscr{L}(f''(t)) &= s^2\mathscr{L}(f(t)) - sf(0) - f'(0), \\ &\vdots \end{aligned}$$

Equation (6.14) can be written as a set of first order differential equations

$$\begin{pmatrix} \dot{z}(t) \\ \ddot{z}(t) \\ \vdots \\ \vdots \\ z^{(n)}(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & \cdots & -p_{n-2} & -p_{n-1} \end{pmatrix} \begin{pmatrix} z(t) \\ \dot{z}(t) \\ \vdots \\ \vdots \\ z^{(n-1)}(t) \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} u(t).$$

Thus, a linear differential system $\dot{x} = Ax + Bu$, $y = Cx$, with state $x = (z, \dot{z}, \dots, z^{(n-1)})^\top$, has been obtained with

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & \cdots & -p_{n-2} & -p_{n-1} \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad C = (1\ 0\ \cdots\ 0). \tag{6.15}$$

The latter is a realization of the transfer function $h(s) = \dfrac{1}{p(s)}$. Note that the eigenvalues of A are the poles of $h(s)$, because $\det(Is - A) = p(s)$ (see Exercise 3.5.14).

We now consider the general case with an arbitrary numerator polynomial $q(s)$ of degree $< n$. Inverse Laplace transformation of (6.13) yields (with the initial values of all appropriate derivatives of u and y equal to zero)

$$\frac{d^n}{dt^n}y(t) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}y(t) + \cdots + p_0 y(t) = q_{n-1}\frac{d^{n-1}}{dt^{n-1}}u(t) + \cdots + q_0 u(t). \tag{6.16}$$

The solution $z(t)$ of (6.14) will now be related to the solution $y(t)$ of (6.16). To that end, because $z(t)$ satisfies (6.14), $q_0 z(t)$ satisfies

$$\frac{d^n}{dt^n}(q_0 z(t)) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}(q_0 z(t)) + \cdots + p_0(q_0 z(t)) = q_0 u(t). \tag{6.17}$$

Differentiation of (6.14) and subsequent multiplication by q_1 leads to

$$\frac{d^n}{dt^n}(q_1\dot{z}(t))+p_{n-1}\frac{d^{n-1}}{dt^{n-1}}(q_1\dot{z}(t))+\cdots+p_0(q_1\dot{z}(t))=q_1\dot{u}(t). \tag{6.18}$$

Continuing this way, we get ultimately

$$\frac{d^n}{dt^n}(q_iz^{(i)}(t))+p_{n-1}\frac{d^{n-1}}{dt^{n-1}}(q_iz^{(i)}(t))+\cdots+p_0(q_iz^{(i)}(t))=q_iu^{(i)}(t),$$

for $i=0,\ldots,n-1$. If we add all these n equations, the result is

$$\begin{aligned}\frac{d^n}{dt^n}\left(q_0z+q_1\dot{z}+\cdots+q_{n-1}z^{(n-1)}\right)+\cdots+p_0\left(q_0z+q_1\dot{z}+\cdots+q_{n-1}z^{(n-1)}\right)\\ =q_0u+q_1\dot{u}+\cdots+q_{n-1}u^{(n-1)}.\end{aligned} \tag{6.19}$$

If (6.16) and (6.19) are compared, we see that the unique solution $y(t)$ of (6.16), with $y(0)=\dot{y}(0)=\ldots=y^{(n-1)}(0)=0$, is equal to $q_0z+q_1\dot{z}+\ldots+q_{n-1}z^{(n-1)}$. A realization of $h(s)=\dfrac{q(s)}{p(s)}$ therefore is

$$A=\begin{pmatrix}0&1&0&\ldots&0\\ \vdots&0&\ddots&\ddots&\vdots\\ \vdots&\vdots&\ddots&\ddots&0\\ 0&0&\ldots&0&1\\ -p_0&-p_1&\cdots&\cdots&-p_{n-1}\end{pmatrix},\quad B=\begin{pmatrix}0\\ \vdots\\ \vdots\\ 0\\ 1\end{pmatrix},\quad C=(q_0\ q_1\ \cdots\ q_{n-1}), \tag{6.20}$$

with state variable $x=(z,\dot{z},\ldots,z^{(n-1)})^\top$.

Other realizations exist, i.e., other triples of matrices A,B,C are possible that correspond to the same transfer function. Indeed, as explained in section 3.4, a coordinate transformation in the state space does not change the transfer function. □

Example 6.10 In Example 6.8 a particular part of the satellite model is discussed that has transfer function

$$\frac{s^2-3}{s^4+s^2}.$$

According the ideas above, a realization of this function is

$$\dot{x}=\begin{pmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&-1&0\end{pmatrix}x+\begin{pmatrix}0\\0\\0\\1\end{pmatrix}u,\quad y=(-3\ 0\ 1\ 0)x.$$

Clearly, this realization is different from the one given in Example 6.8. □

Example 6.11 Consider the two systems ($x,u,y\in\mathbb{R}$):

$$\begin{array}{lrcllrcl}\Sigma_1: & \dot{x} & = & -x+2u, & \qquad y & = & 2x,\\ \Sigma_2: & \dot{x} & = & -2x+3u, & \qquad y & = & -x.\end{array}$$

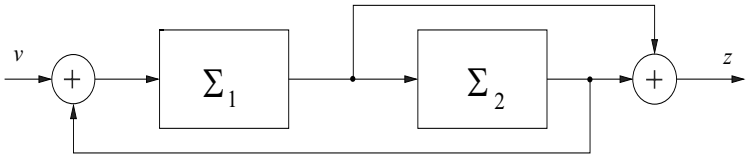

Figure 6.5 Scheme of two coupled systems.

What are the transfer functions of these two systems? Subsequently, these systems are coupled to each other as indicated in Figure 6.5. The input and output of this combined system are called v and z, respectively. What is the transfer function that describes the relation between v and z? Give also a state space description of this combined system.

Answer. The transfer functions of Σ_1 and Σ_2 are

$$h_1(s) = \frac{4}{s+1}, \quad h_2(s) = \frac{-3}{s+2},$$

respectively. In order to determine the transfer function of the coupled system, we define $y_i(t)$ to be the output of system Σ_i. Then we formally get

$$Y_2(s) = h_2(s)Y_1(s), \quad Y_1(s) = h_1(s)\big(V(s)+Y_2(s)\big), \quad Z(s) = Y_1(s)+Y_2(s),$$

from which it follows that

$$Z(s) = \frac{h_1(s)\big(1+h_2(s)\big)}{1-h_1(s)h_2(s)}V(s).$$

Substitution of $h_1(s)$ and $h_2(s)$ leads to the transfer function

$$\frac{4s-4}{s^2+3s+14}.$$

A state space description is

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ -14 & -3 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} v, \quad z = \begin{pmatrix} -4 & 4 \end{pmatrix} x.$$

□

6.5 More on realizations

6.5.1 Flow diagrams

The realization in (6.20) has a special name, namely the **standard controllable realization**, or the **controller form**, and was already met in (5.4). The procedure given in the previous section to obtain a realization can also be visualized by means of a flow diagram, as depicted in Figure 6.6, (in this diagram $n = 3$ and the notation $z^{(i)}$ refers to the i-th derivative of z). In the diagram the box $\boxed{\int}$ denotes integration, which is a shorthand notation for the system $\dot{x} = u$, $y = x$, with transfer function $\frac{1}{s}$, and the boxes $\boxed{-p_i}$ and

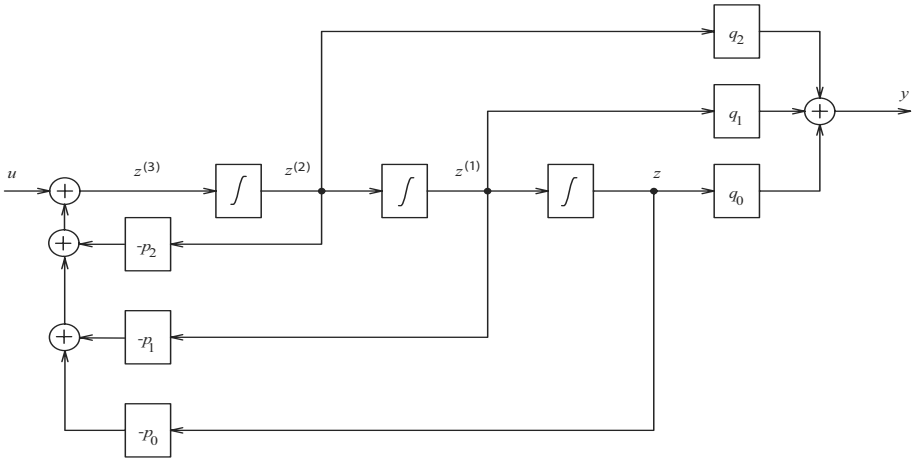

Figure 6.6 Flow diagram of realization.

$\boxed{q_i}$ denote multiplication with the coefficient inside the box. The diagram also indicates how the system could be realized in practice (i.e., be built) if we have devices (building blocks) at hand which integrate, add and multiply. This is exactly what is done in an **analog computer**.

Superficially, we could also implement or build this system by means of **differentiators**. The starting point would then be the design or flow diagram as depicted in Figure 6.7. The flow diagram between u and z is (take $n = 3$) and therefore, by superposition,

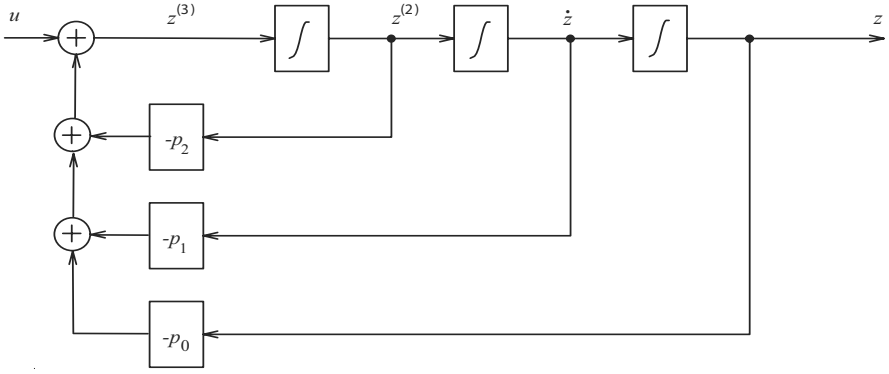

Figure 6.7 Flow diagram of input part.

a diagram as in Figure 6.8 results. This flow diagram also describes the system characterized by $h(s) = \dfrac{q(s)}{p(s)}$. However, now differentiators, i.e., the blocks $\boxed{\frac{d}{dt}}$, have been used. As will be explained in Example 6.21, differentiators are technically difficult to build. Because integrators can be realized much easier, a flow diagram with integrators instead of differentiators is to be preferred.

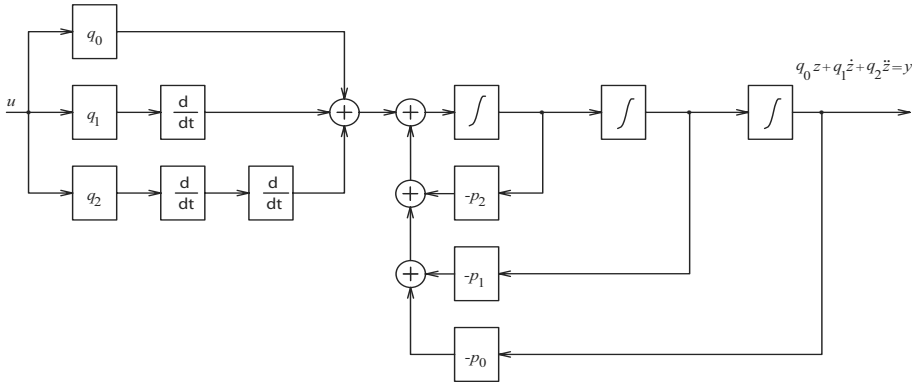

Figure 6.8 Flow diagram of combination.

6.5.2 Alternative realizations

In addition to the above given standard controllable realization, other useful realizations exist. One of them carries the name of **standard observable realization**. We will not discuss it here extensively. We only give it here for sake of completeness for the transfer function (6.8) with $q_n = 0$. Then this realization is given by

$$A = \begin{pmatrix} 0 & 0 & \cdots & \cdots & -p_0 \\ 1 & 0 & 0 & & -p_1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 1 & 0 & -p_{n-2} \\ 0 & \cdots & \cdots & 1 & -p_{n-1} \end{pmatrix}, \quad B = \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_{n-2} \\ q_{n-1} \end{pmatrix}, \quad C = (0\ 0\ \cdots\ 0\ 1). \tag{6.21}$$

We will continue this section with yet another realization that also realizes a rational function $h(s) = \dfrac{q(s)}{p(s)}$, with $\deg q(s) < \deg p(s)$, as a linear differential system. The method is based on the following partial fraction decomposition of $h(s)$, where the a_i are the poles of $h(s)$, which, for the time being, are assumed to be real and to have multiplicity one,

$$h(s) = \frac{q(s)}{p(s)} = \frac{\gamma_1}{s-a_1} + \frac{\gamma_2}{s-a_2} + \cdots + \frac{\gamma_n}{s-a_n}.$$

A realization of $h(s)$ is then given by

$$\dot{x} = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_n \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} u, \quad y = (\gamma_1\ \gamma_2\ \cdots\ \gamma_n)x,$$

which can be depicted in a block diagram as shown in Figure 6.9. This realization is called a **diagonal realization**. The original n-th order system has been **decoupled** into n

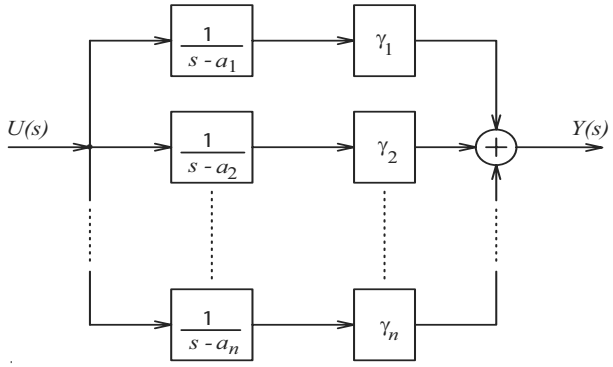

Figure 6.9 Diagonal realization.

independent subsystems. The blocks with the contents $\frac{1}{s-a_i}$ are shorthand notation for the flow diagram depicted in the right-hand side of Figure 6.10.

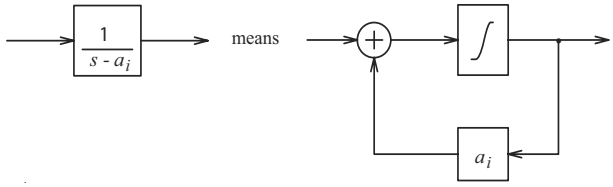

Figure 6.10 Implementation of an elementary building block.

If, instead of the above, $p(s)$ has real roots of multiplicity larger than one, say $s=a$ has multiplicity two, then partial fraction decomposition leads to

$$h(s) = \frac{\gamma}{s-a} + \frac{\delta}{(s-a)^2} + \cdots$$

These terms can be realized jointly as shown in Figure 6.11. If the outputs of the two

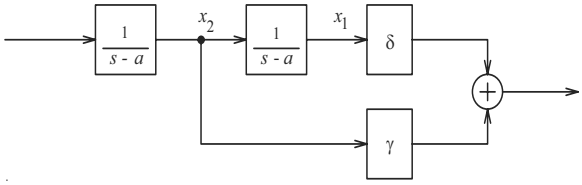

Figure 6.11 Realization of more a general element.

blocks with integrators are denoted by x_2 and x_1, as indicated in the figure, then a state space realization of $\frac{\gamma}{s-a} + \frac{\delta}{(s-a)^2}$ is

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u, \quad y = (\delta \;\; \gamma)x.$$

The system matrix A is now a Jordan block of size 2×2 for the eigenvalue a. Similar results hold if a is a real root of $p(s)$ with a multiplicity higher than two.

If $p(s)$ contains factors of the form s^2+bs+c with $b^2-4c<0$, such that a further decomposition in real factors is impossible, then the following shows a possible way to obtain a flow diagram. As an example, consider the transfer function given by

$$h(s)=\frac{s+2}{s^2+2s+5}.$$

The denominator cannot be decomposed any further into real factors. Then $h(s)$ can be written as

$$h(s)=\frac{s+2}{(s+1)^2+4}=\frac{\frac{1}{s+1}}{1+\frac{2^2}{(s+1)^2}}+\left(\tfrac{1}{2}\right)\frac{\frac{2}{(s+1)^2}}{1+\frac{2^2}{(s+1)^2}},$$

and a flow diagram can be given as in Figure 6.12. If the output of the two blocks

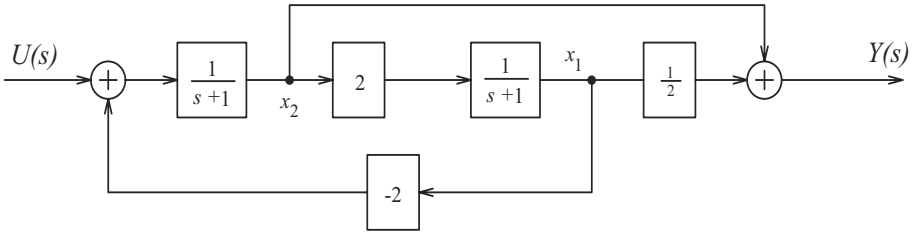

Figure 6.12 Flow diagram of an irreducible transfer function.

with integrators are denoted by x_1 and x_2, as indicated in the figure, then a state space realization is

$$\frac{d}{dt}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}-1 & 2\\-2 & -1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}0\\1\end{pmatrix}u,\quad y=\begin{pmatrix}\frac{1}{2} & 1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$

6.5.3 Example

Example 6.12 We are given the system

$$\frac{d}{dt}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}-1 & 0 & 1\\0 & -1 & 1\\0 & 0 & 0\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}+\begin{pmatrix}1 & 0\\2 & 1\\0 & 1\end{pmatrix}\begin{pmatrix}u_1\\u_2\end{pmatrix},$$

$$\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}.$$

This is a model of a turbo-propeller engine, see also Figure 6.13, where

- $x_1=y_1$ is the deviation of the rotating speed from its nominal value at the desired steady-state operating point,
- $x_2=y_2$ is the deviation of the turbine-inlet temperature from its nominal value,

- x_3 is the deviation of the fuel rate from its nominal value,
- u_1 is the deviation of the propeller blade angle from the nominal value,
- u_2 is the time-derivative of the fuel rate.

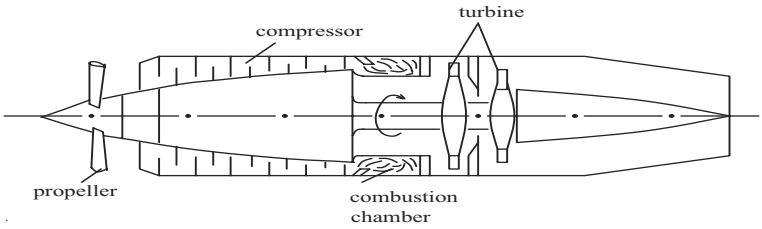

Figure 6.13 Turbo-propeller engine model.

Consider the following questions.

1. Determine the transfer matrix of the system.
2. One wants to **decouple** the inputs and the outputs. That is, the first input should only influence the first output and the second input should only influence the second output. For the decoupling to be true, what properties must the transfer matrix satisfy?
3. Instead of with u, we are going to control the system by means of $w \in \mathbb{R}^2$, where $u \in \mathbb{R}^2$, $x \in \mathbb{R}^3$, $w \in \mathbb{R}^2$ and an auxiliary variable $v \in \mathbb{R}^2$ are related to each other as

$$u = Gv, \quad v = Fx + w.$$

Determine constant matrices G and F such that the new system, with input w and output y, is decoupled.

Answer question 1. The transfer matrix is calculated from $H(s) = C(sI - A)^{-1}B$. It equals

$$H(s) = \begin{pmatrix} \frac{1}{s+1} & \frac{1}{s(s+1)} \\ \frac{2}{s+1} & \frac{1}{s} \end{pmatrix}.$$

Answer question 2. $H(s)$ must be a diagonal matrix (which is not the case here).

Answer question 3. With the new input (and output), the system equations can be written as

$$\begin{aligned} \dot{x} &= (A + BGF)x + BGw, \\ y &= Cx. \end{aligned} \tag{6.22}$$

Write

$$G = \begin{pmatrix} g_1 & g_2 \\ g_3 & g_4 \end{pmatrix}, \quad F = \begin{pmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \end{pmatrix}.$$

From the impulse response of system (6.22) at time $t = 0$, it follows that since $y_1 = x_1$ must not depend on w_2, a requirement is that $g_2 = 0$. Similarly, $y_2 = x_2$ must not depend on w_1, which leads to $2g_1 + g_3 = 0$. The element g_4 can be chosen freely. Hence, a possible choice for G is

$$G = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}.$$

With this G we get

$$A + BGF = \begin{pmatrix} -1+f_1 & f_2 & 1+f_3 \\ f_4 & -1+f_5 & 1+f_6 \\ -2f_1+f_4 & -2f_2+f_5 & -2f_3+f_6 \end{pmatrix}.$$

By choosing $f_2 = 0$ and $f_3 + 1 = 0$, the requirement that x_1 is not influenced by x_2 and x_3 has been taken care of. Similarly, by choosing $f_4 = 0$ and $f_6 + 1 = 0$, x_2 is not influenced by x_1 and x_3. The remaining elements f_1 and f_5 can be chosen freely. For instance, take $f_1 = f_5 = 0$. Hence,

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

With this choice of F and G, the system becomes

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix},$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

which is clearly decoupled. □

6.6 Transfer functions and minimal realizations

6.6.1 Realizations of single-input single-output systems

In this first subsection some further results on realizations of single-input single-output systems are presented. The next subsection deals with realizations of multiple-input multiple-output systems.

Theorem 6.13 *Let $p(s)$ be a polynomial of degree n and let $q(s)$ be a polynomial whose degree is at most n. Then a realization with state space $\mathbb{R}^n$ of the transfer function*

$$h(s) = \frac{q(s)}{p(s)}$$

is both controllable and observable if and only if the polynomials $q(s)$ and $p(s)$ do not have common factors.

Proof The proof will only be given for transfer functions which allow a diagonal realization, i.e., the system matrix A is diagonal. The proof consists of two parts. First, we will prove that, given controllability and observability, there are no common factors. Subsequently, we will prove that, if the system is not controllable and/or not observable, there are common factors. Together these two parts then prove the theorem.

Necessity. Consider the diagonal realization

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \quad C = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix}, \tag{6.23}$$

with corresponding transfer function

$$h(s) = \frac{q(s)}{p(s)} = \sum_{i=1}^{n} \frac{g_i}{s - \lambda_i}. \tag{6.24}$$

The scalars b_i and c_i satisfy $b_i c_i = g_i$, but are otherwise arbitrary. The controllability matrix is

$$R = \begin{pmatrix} b_1 & b_1\lambda_1 & b_1\lambda_1^2 & \dots & b_1\lambda_1^{n-1} \\ b_2 & b_2\lambda_2 & b_2\lambda_2^2 & \dots & b_2\lambda_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ b_n & b_n\lambda_n & b_n\lambda_n & \dots & b_n\lambda_n^{n-1} \end{pmatrix},$$

and

$$\det R = \det \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \dots & \lambda_n^{n-1} \end{pmatrix} \cdot \prod_{i=1}^{n} b_i. \tag{6.25}$$

The determinant in the right-hand side of (6.25) is the so-called determinant of **Van der Monde** and it can be shown (the proof is by induction with respect to the size of the matrix; the proof will not be given here) that this determinant is equal to

$$\prod_{1 \le i < j \le n} (\lambda_j - \lambda_i). \tag{6.26}$$

(Note that, for instance, $\prod_{1 \le i < j \le 3} (\lambda_j - \lambda_i) = (\lambda_2 - \lambda_1) \cdot (\lambda_3 - \lambda_1) \cdot (\lambda_3 - \lambda_2)$, and that $\prod_{1 \le i < j \le n} (\lambda_j - \lambda_i)$ consists of $\frac{1}{2}n(n-1)$ factors.)

Hence, $\det R \neq 0$ if and only if $\lambda_i \neq \lambda_j$ for all i, j with $i \neq j$, and $b_i \neq 0$ for all i. The latter requirement is quite obvious. If $b_i = 0$ for some i, then the i-th component of the state is not excited by the input, and cannot belong to the controllable subspace.

Hence, realization (6.23) is controllable if and only if $\lambda_i \neq \lambda_j$ for all i, j with $i \neq j$, and $b_i \neq 0$ for all i. With the same argument it can be shown that the realization is observable if and only if $\lambda_i \neq \lambda_j$ for all i, j with $i \neq j$, and $c_i \neq 0$ for all i. For a controllable and observable realization of the form (6.23), $c_i \neq 0$ and $b_i \neq 0$ for all i, and therefore $g_i \neq 0$ for all i. This implies that there are no common factors in $h(s)$.

Sufficiency. Now suppose that the realization is not controllable. Then, according to (6.25) and (6.26), either $\lambda_i = \lambda_j$ for some i, j with $i \neq j$, or $b_i = 0$ for some i, or both. Without loss of generality, assume that either $\lambda_1 = \lambda_2$, or $b_1 = 0$, or both. In all three cases, $h(s)$ can be rewritten as follows

$$h(s) = \sum_{i=1}^{n} \frac{g_i}{s-\lambda_i} = \sum_{i=2}^{n} \frac{\tilde{g}_i}{s-\lambda_i} = h_{\text{red}}(s),$$

where 'red' stands for reduced. Indeed, in all three cases $\tilde{g}_i = g_i$ for $3 \leq i \leq n$, and $\tilde{g}_2 = g_1 + g_2$ when $\lambda_1 = \lambda_2$, and $\tilde{g}_2 = g_2$ otherwise. Hence, $h(s)$ can be written as a proper rational function with a denominator of degree (at most) $n-1$, implying that the numerator and denominator of $h(s)$ contain some common factor. A similar statement follows starting from a realization that is not observable.

Hence, an n dimensional realization of $h(s) = \dfrac{q(s)}{p(s)}$, with $\deg p(s) = n \geq \deg q(s)$, that is not controllable or not observable, implies that $p(s)$ and $q(s)$ contain a common factor. □

Example 6.14 Consider the system

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & -4 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} u, \quad y = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{6.27}$$

The matrices R and W for this system are

$$R = \begin{pmatrix} 0 & -1 & -4 \\ 0 & 0 & 0 \\ 1 & 3 & 8 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & -1 & 1 \\ 2 & -3 & 2 \\ 4 & -7 & 4 \end{pmatrix}.$$

Both R and W are singular! Even one of them being singular would be sufficient to conclude that the transfer function contains some common factor. Indeed, based on the non-controllability, it follows that the numerator and denominator of $h(s)$ have a common factor $(s-1)$, and the non-observability implies that the numerator and denominator of $h(s)$ have a common factor $(s-2)$. The transfer function is

$$h(s) = \frac{(s-1)(s-2)}{(s-1)(s-2)^2} = \frac{1}{s-2}.$$

Hence, the input-output behavior of the system given by the realization (6.27) with a three dimensional state space can also be realized by a realization with a one dimensional state space. Such a realization is (x is one dimensional)

$$\dot{x} = 2x + u, \quad y = x.$$

A realization of which the dimension of the state is minimal is called a **minimal realization**. □

6.6.2 Realizations of multiple-input multiple-output systems

Until now most of the realizations concerned single-input single-output systems. In this subsection a generalization of (6.20) towards multiple-input multiple-output systems is discussed and some additional results are presented.

To present a generalization of (6.20), assume that $H(s)$ is the transfer matrix of a system with m inputs and p outputs. Hence, $H(s)$ is an $p \times m$ matrix. Each entry of $H(s)$ can be written in the form of (6.12). Leaving the $\bar{}$ notation, it follows that all obtained expressions can be combined in such a way that

$$H(s) = D + \frac{1}{p(s)} Q(s)$$

with $D \in \mathbb{R}^{p \times m}$ and

$$\begin{aligned} p(s) &= s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0 \\ Q(s) &= Q_{n-1}s^{n-1} + \cdots + Q_1 s + Q_0, \end{aligned}$$

where $p_0, p_1, \ldots, p_{n-1} \in \mathbb{R}$ and $Q_0, Q_1, \ldots, Q_{n-1} \in \mathbb{R}^{p \times m}$. Hence, all the p_i's are scalars and all the Q_i's are $p \times m$ matrices. Without proof we now state that the following matrices A, B, C, together with D, form a realization of $H(s)$, i.e., $H(s) = D + C(sI - A)^{-1}B$,

$$A = \begin{pmatrix} 0 & I & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & I \\ -p_0 I & -p_1 I & \ldots & \ldots & -p_{n-1} I \end{pmatrix}, B = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ I \end{pmatrix}, C = (Q_0, \ldots, Q_{n-1}), \qquad (6.28)$$

where all the 0's denote $m \times m$ zero matrices and the I's are $m \times m$ identity matrices. Hence, the matrix A consists on nm rows and nm columns. Matrix B has nm rows and m columns, and matrix C has p rows and nm columns. Note that the bottom 'block' row of A consists of matrices of the form $p_i I$.

As an alternative to the above method, all the individual nonzero entries in the transfer matrix $H(s)$ can first be seen as single-input single-output systems and each of them can be realized according to one of the methods described before. Next all obtained realizations can be combined together in a large realization of the original multiple-input multiple-output system. This approach is illustrated by means of the following example.

Example 6.15 *Consider the transfer matrix*

$$H(s) = \begin{pmatrix} h_{11}(s) & 0 \\ h_{21}(s) & h_{22}(s) \\ 0 & h_{32}(s) \end{pmatrix}$$

with

$$h_{11}(s)=\frac{s^2+4s+6}{s^2+3s+2},\qquad h_{21}(s)=\frac{1}{(s+3)(s+2)},$$
$$h_{22}(s)=\frac{3s^2+4s-7}{(s+1)^3},\qquad h_{32}(s)=\frac{s+2}{s^2+4s+5}.$$

Note that

$$h_{11}(s)=1+\frac{s+4}{s^2+3s+2},\qquad h_{21}(s)=\frac{1}{(s+2)}-\frac{1}{(s+3)},$$
$$h_{22}(s)=\frac{3s^2+4s-7}{s^3+3s^2+3s+1},\qquad h_{32}(s)=\frac{\frac{1}{(s+2)}}{1+\frac{1}{(s+1)^2}}.$$

Hence, $h_{11}(s)$ can be realized by means of the standard controllable realization

$$A_{11}=\begin{pmatrix}0&1\\-2&-3\end{pmatrix},\quad B_{11}=\begin{pmatrix}0\\1\end{pmatrix},\quad C_{11}=\begin{pmatrix}4&1\end{pmatrix},\quad D_{11}=1.$$

For $h_{21}(s)$ a diagonal realization can be obtained with

$$A_{21}=\begin{pmatrix}-2&0\\0&-3\end{pmatrix},\quad B_{21}=\begin{pmatrix}1\\1\end{pmatrix},\quad C_{21}=\begin{pmatrix}1&-1\end{pmatrix},\quad D_{21}=0.$$

A standard observable realization for $h_{22}(s)$ is given by

$$A_{22}=\begin{pmatrix}0&0&-1\\1&0&-3\\0&1&-3\end{pmatrix},\quad B_{22}=\begin{pmatrix}-7\\4\\3\end{pmatrix},\quad C_{22}=\begin{pmatrix}0&0&1\end{pmatrix},\quad D_{22}=0.$$

Finally, $h_{32}(s)$ can be realized, in the way as described on page 125, by

$$A_{32}=\begin{pmatrix}-2&-1\\1&-2\end{pmatrix},\quad B_{32}=\begin{pmatrix}1\\0\end{pmatrix},\quad C_{32}=\begin{pmatrix}1&0\end{pmatrix},\quad D_{32}=0.$$

With the realizations for the individual nonzero entries of the transfer matrix $H(s)$, the latter can be realized by means of A,B,C and D made up of the obtained realizations. After some book keeping, it follows that here these matrices may look like

$$A=\begin{pmatrix}A_{11}&0&0&0\\0&A_{21}&0&0\\0&0&A_{22}&0\\0&0&0&A_{32}\end{pmatrix},\quad B=\begin{pmatrix}B_{11}&0\\B_{21}&0\\0&B_{22}\\0&B_{32}\end{pmatrix},$$

$$C=\begin{pmatrix}C_{11}&0&0&0\\0&C_{21}&C_{22}&0\\0&0&0&C_{32}\end{pmatrix},\quad D=\begin{pmatrix}D_{11}&0\\D_{21}&D_{22}\\0&D_{32}\end{pmatrix},$$

where all 0's are zero matrices of suitable sizes. Indeed, with these A,B,C and D, it follows that $D+C(sI-A)^{-1}B=$

$$\begin{pmatrix} D_{11}+C_{11}(sI-A_{11})^{-1}B_{11} & 0 \\ D_{21}+C_{21}(sI-A_{21})^{-1}B_{21} & D_{22}+C_{22}(sI-A_{22})^{-1}B_{22} \\ 0 & D_{32}+C_{32}(sI-A_{32})^{-1}B_{32} \end{pmatrix} = H(s),$$

where all I's stand for identity matrices of suitable sizes. Hence, it follows that $H(s) = D+C(sI-A)^{-1}B$ with

$$A = \left(\begin{array}{cc|cc|ccc|cc} 0 & 1 & & & & & & & \\ -2 & -3 & & & & & & & \\ \hline & & -2 & 0 & & & & & \\ & & 0 & -3 & & & & & \\ \hline & & & & 0 & 0 & -1 & & \\ & & & & 1 & 0 & -3 & & \\ & & & & 0 & 1 & -3 & & \\ \hline & & & & & & & -2 & -1 \\ & & & & & & & 1 & -2 \end{array}\right), \quad B = \left(\begin{array}{c|c} 0 & \\ 1 & \\ \hline 1 & \\ 1 & \\ \hline & -7 \\ & 4 \\ & 3 \\ \hline & 1 \\ & 0 \end{array}\right),$$

$$C = \left(\begin{array}{cc|cc|ccc|cc} 4 & 1 & & & & & & & \\ \hline & & 1 & -1 & 0 & 0 & 1 & & \\ & & & & & & & 1 & 0 \end{array}\right), \quad D = \left(\begin{array}{c|c} 1 & \\ \hline 0 & 0 \\ \hline & 0 \end{array}\right),$$

where the unspecified entries in A,B,C and D are zero.

Both methods presented above to obtain a realization of a multiple-input multiple-output system, in general, will yield a realization with a state space that is too large. As in the single-input single-output case, by restricting an obtained realization to the part that is controllable and observable a minimal realization results of smallest dimension. This smallest dimension is clearly of interest, and can also be defined and determined as follows.

Definition 6.16 *If $H(s)$ is a matrix of which the entries are proper rational functions, then we will say that the* **McMillan degree** *of $H(s)$ is n if $H(s)$ has a realization $D+C(Is-A)^{-1}B$ with A having size $n\times n$, and no realizations exist with matrices A having a size $k\times k$ where $k<n$.*

The following theorem, of wich the proof will not be given, provides an algorithm for calculating the McMillan degree.

Theorem 6.17 *Given any $p\times m$ matrix $H(s)$ of which the entries are proper rational functions, and given that $H(s)$ has the expansion*

$$H(s) = L_0 + L_1 s^{-1} + L_2 s^{-2} + L_3 s^{-3} + \cdots,$$

then the McMillan degree n of $H(s)$ is given by $n = \operatorname{rank} L(r,r)$, where

$$L(\alpha,\beta) \stackrel{\text{def}}{=} \begin{pmatrix} L_1 & L_2 & L_3 & \cdots & L_\beta \\ L_2 & L_3 & L_4 & \cdots & L_{\beta+1} \\ L_3 & L_4 & L_5 & \cdots & L_{\beta+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_\alpha & L_{\alpha+1} & L_{\alpha+2} & \cdots & L_{\alpha+\beta-1} \end{pmatrix},$$

with r being the degree of the least common multiple of all denominators of $H(s)$.

6.7 Frequency methods

6.7.1 Oscillations

So far we assumed that input and output functions are real functions (vectors). From the control point of view it turns out to be useful to admit complex functions. Therefore, we take now the complex valued input function

$$u(t) = \begin{cases} 0 & \text{for} \quad t < 0, \\ e^{st}c & \text{for} \quad t \geq 0, \end{cases}$$

with $s \in \mathbb{C}$ and c a complex vector. If $x(0) = 0$, then the corresponding output function will be

$$y(t) = \int_0^t G(t-\tau)e^{s\tau}c\,d\tau = \int_0^t G(r)e^{s(t-r)}c\,dr =$$

$$\Big(\int_0^t G(r)e^{-sr}dr\Big)e^{st}c = \Big(\int_0^t G(\tau)e^{-s\tau}d\tau\Big)u(t).$$

If we consider the limit of $t \to \infty$, and assume that the integral converges to $H(s)$, for $\operatorname{Re} s$ sufficiently large, then

$$y(t) \approx H(s)u(t).$$

This somewhat weird looking expression must be viewed as the approximate equality of two (complex valued) time functions, where $H(s)$ is a proportionality factor in which s has a specific numerical value.

Since $u(t) = e^{st}c = ce^{\rho t}(\cos \omega t + i \sin \omega t)$, with $s = \rho + i\omega$ and $\rho, \omega \in \mathbb{R}$, the input function u represents an oscillation. If $\int_0^\infty G_{ij}(\tau)d\tau < \infty$ for each element G_{ij} of the matrix G, then $\int_0^\infty G(\tau)e^{-s\tau}d\tau$ exists for $\operatorname{Re} s \geq 0$ and, more explicitly, for all $s = i\omega$ with $\omega \in \mathbb{R}$. If an input $u(t) = ce^{i\omega t}$ is applied, an output $y(t) \approx H(i\omega)u(t)$ results for large t. The function $e^{i\omega t}c = c(\cos \omega t + i \sin \omega t)$ is called a **harmonic oscillation** and $H(i\omega)e^{i\omega t}c$ is the **stationary response** on the harmonic oscillation $e^{i\omega t}c$. The matrix $H(i\omega)$ is called the **frequency response matrix**. The difference between $y(t)$ and the stationary response is called the **transient behavior**. If $\int_0^\infty G_{ij}(\tau)d\tau < \infty$ for all i, j, then this behavior tends to zero as $t \to \infty$. It follows from the section on stability that $\int_0^\infty G_{ij}(\tau)d\tau < \infty$ for all i, j, when $\operatorname{Re} \lambda_i < 0$ for all eigenvalues λ_i of A.

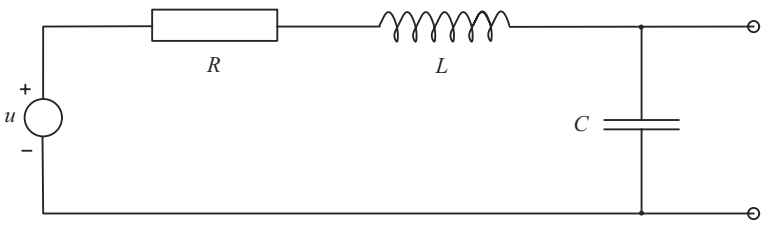

Figure 6.14 Basic electrical network.

Remark 6.18 If, in the definition of the Laplace transform as given by (6.1), one confines oneself to s on the imaginary axis, i.e., $s = i\omega$, with ω being real, then one obtains a version of the Fourier transform. (Two different versions of Fourier transforms exist in the sense that, depending on the application, the lower bound of the integral is either 0 or $-\infty$.)

The next section gives more information on the frequency responses.

6.7.2 Nyquist and Bode diagrams

In this subsection we confine ourselves to single-input single-output linear differential systems with a transient behavior that goes to zero as $t \to \infty$. For the frequency response we can write

$$h(i\omega) = |h(i\omega)|e^{i\arg h(i\omega)}.$$

The stationary response of $u(t) = u_\omega e^{i(\omega t+\varphi)}$, with $\omega, \varphi, u_\omega \in \mathbb{R}$ is

$$y(t) = h(i\omega)u_\omega e^{i(\omega t+\varphi)} = |h(i\omega)|u_\omega e^{i(\omega t+\arg h(i\omega)+\varphi)}. \tag{6.29}$$

Now consider $u_\omega \sin(\omega t+\varphi)$ as a sinusoidal input signal and treat it as the imaginary part of $u(t)$. Then, if we take the imaginary parts of (6.29), the stationary response of

$$\mathrm{Im}(u(t)) = \mathrm{Im}(u_\omega e^{i(\omega t+\varphi)}) = u_\omega \sin(\omega t+\varphi)$$

equals

$$\begin{aligned} \mathrm{Im}(y(t)) &= \mathrm{Im}(|h(i\omega)|u_\omega e^{i(\omega t+\varphi+\arg h(i\omega))}) \\ &= |h(i\omega)|u_\omega \sin(\omega t+\varphi+\arg h(i\omega)). \end{aligned} \tag{6.30}$$

The stationary response is also sine-shaped, with **amplitude** $|h(i\omega)|u_\omega$. The **phase** of the oscillation is increased with $\arg h(i\omega)$. A linear time-invariant system with transfer function $h(s)$ transforms a sinusoidal signal with frequency ω into another sinusoidal signal with frequency ω by multiplying the amplitude by $|h(i\omega)|$, called the **gain**, and increasing the phase by $\arg h(i\omega)$, called the **phase-shift**.

Example 6.19 Consider the electric network depicted in Figure 6.14 (compare the example of Section 2.4.4). For the state we choose $x_1 = q$ (charge of the capacitor) and $x_2 = \varphi$

(magnetic flux of the induction coil). If i is the current and v the voltage, then it follows that

$$\dot{x}_1 = \dot{q} = i = \tfrac{1}{L}\varphi,$$
$$\dot{x}_2 = \dot{\varphi} = v = -Ri - \tfrac{1}{C}q + u = -\tfrac{R}{L}\varphi - \tfrac{1}{C}q + u.$$

Thus

$$A = \begin{pmatrix} 0 & \frac{1}{L} \\ -\frac{1}{C} & -\frac{R}{L} \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad C = \begin{pmatrix} \frac{1}{C} & 0 \end{pmatrix},$$

(do not confuse the output matrix with the capacity C) and

$$h(s) = C(Is - A)^{-1}B = \frac{1}{LCs^2 + RCs + 1}, \quad h(i\omega) = \frac{1}{-LC\omega^2 + 1 + iRC\omega}.$$

The poles of $h(s)$ are the zeros of $s^2 + \frac{R}{L}s + \frac{1}{LC}$. It is straightforward to show that both poles have a negative real part. Hence, if an input signal $u_\omega \sin(\omega t + \varphi)$ is applied, then

$$y(t) \approx |h(i\omega)| u_\omega \sin(\omega t + \varphi + \arg h(i\omega)).$$

Further, the gain and the phase-shift are given by

$$|h(i\omega)| = \frac{1}{\sqrt{\left((1 - LC\omega^2)^2 + R^2C^2\omega^2\right)}},$$
$$\arg\, h(i\omega) = \arctan\left(\frac{-RC\omega}{1 - LC\omega^2}\right).$$

□

In general, if a linear combination of sinusoidal signals, possibly of different frequencies, is applied to the system, then the output will be linear combination of sinusoidal signals with the same frequencies as in the input signal.

Frequency response functions are used frequently in network analysis, automatic control and acoustics. There are two well-known methods to display $h(i\omega)$ graphically and to get an impression of the properties of the system by studying these graphs. The methods will be discussed briefly here. Many design techniques are based on these methods.

1. The **Nyquist diagram** or **polar plot**. The function $h(i\omega)$ is plotted as a curve in the complex plane, parametrized by ω (varying from 0 to $+\infty$). If we think of $h(s)$ as a function mapping from the complex plane into the complex plane, then the Nyquist diagram is the image of the positive imaginary axis under h.

2. The **Bode diagram** or the **logarithmic diagram**. In this case h is represented by two graphs. Namely, the amplitude plot: $\ln|h(i\omega)|$ as a function of $\ln\omega$, and the phase plot: $\arg h(i\omega)$ as a function of $\ln\omega$.

In Figure 6.15 the Nyquist diagram and the Bode diagram of the system with transfer function $1/(1+Ts)$, with $T > 0$ being a constant, are given as an example. Please note that the scale of $\ln|h(i\omega)|$ is expressed in so-called **decibels** (dB). The graph of $|h(i\omega)|$ versus ω (or $\ln\omega$) indicates which frequencies can pass the system and also with what gain. The system can thus be interpreted as a **filter** for the input signals. In the first of the

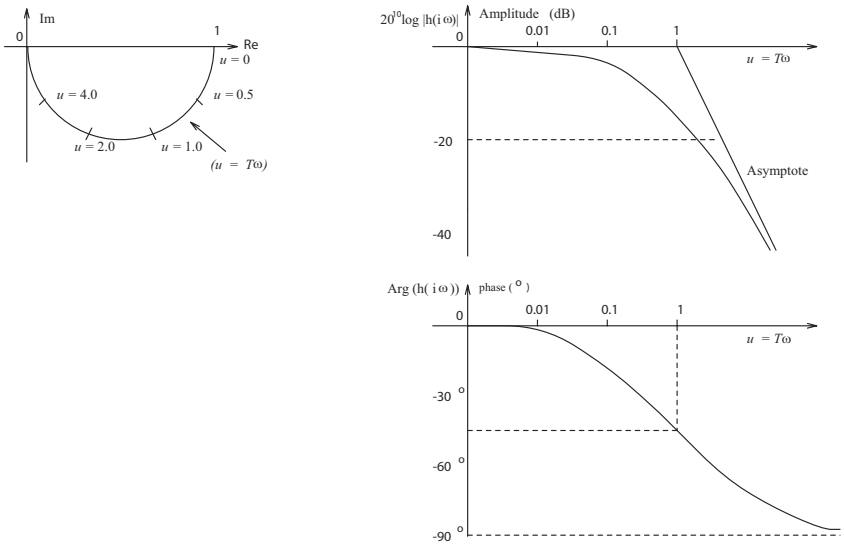

Figure 6.15 Nyquist diagram (upper left) and Bode diagrams (right) of $\frac{1}{1+Ts}$.

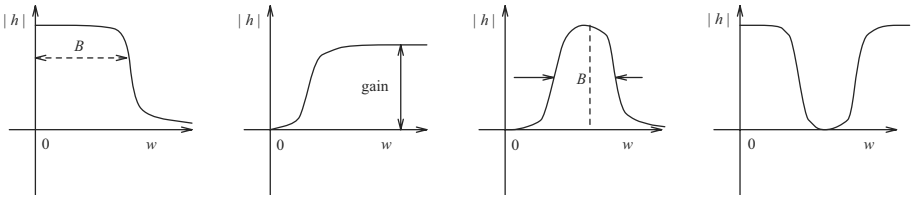

Figure 6.16 Four types of filters.

four plots on a row in Figure 6.16 only the low frequencies will pass the system, whereas the higher frequencies are cut off. Such a filter is called a **low frequency filter**. The other figures show other kinds of filters; they are self-explanatory. The **bandwidth** B of a system is defined as that range of frequencies (of the input signal) for which the system will respond satisfactorily.

A simple application of a low frequency filter is the following. Noise signals consist usually of high frequency signals. If we want to get rid of this noise, a low frequency filter can be used. As a consequence, those parts of the input signal related to high frequencies will be cut off.

Example 6.20 Consider the feedback system of the configuration (two blocks in the forward loop and unity feedback) depicted in Figure 6.17. In the figure $H_1(s)$ is the transfer function of a given system (in practice sometimes also called **plant**). We want to design a **controller** $H_2(s)$ such that the overall feedback system has pleasant characteristics. The controller is characterized by its transfer function, which can be chosen by the designer. It is easily shown that the transfer function of the overall system is given by

$$Y(s) = H(s)U(s), \text{ with } H(s) = \big(I + H_1(s)H_2(s)\big)^{-1} H_1(s)H_2(s).$$

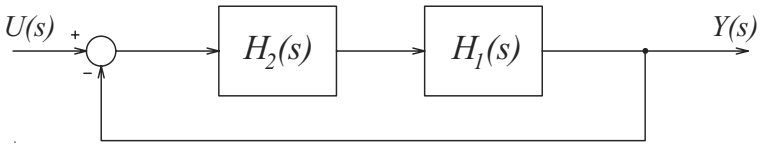

Figure 6.17 Two blocks in forward loop.

A possible design criterion could be that $Y(s)$ must be as close as possible to $U(s)$, i.e., the output tries to follow the input. This is called **tracking**. A possibility to achieve a good tracking system is to design $H_2(s)$ in such a way that $H_1(s)H_2(s)$ is 'large' in some sense, since then $(I+H_1(s)H_2(s))^{-1}H_1(s)H_2(s) \approx I$ and subsequently $Y(s) \approx U(s)$. For frequency considerations the variable s is replaced by $i\omega$, and then $S(\omega)$ is defined as

$$S(\omega) = \big(I + H_1(i\omega)H_2(i\omega)\big)^{-1},$$

and is called the **sensitivity operator**. If $u(t)$ and $y(t)$ are scalar, such that $S(\omega)$ is a 1×1 matrix, then the corresponding system is said to have good sensitivity characteristics if

$$\big|1 + H_1(i\omega)H_2(i\omega)\big| \geq \varphi(\omega),$$

for all $|\omega| \leq \omega_0$, with ω_0 the bandwidth of interest, and where $\varphi(\omega)$ is a (large) positive function. □

Example 6.21 [The differentiator] Suppose $y(t) = \dfrac{du(t)}{dt}$, $t \in \mathbb{R}$. Then for $u(0) = 0$

$$Y(s) = \int_0^\infty e^{-st}\frac{du}{dt}dt = \Big[u(t)e^{-st}\Big]_0^\infty + s\int_0^\infty u(t)e^{-st}dt = sU(s).$$

The transfer function is s, which is a non-proper rational function, because it can be interpreted as $\dfrac{s}{1}$. Since the degree of the numerator is larger than that of the denominator, this is a **non-causal** system. Such a system cannot be realized technically (if $u(\tau)$ is known up to time t, then the derivative at the end point $\tau = t$ does not exist). Furthermore, it is clear that $|h(i\omega)| = |\omega|$, such that higher frequencies are amplified more than lower frequencies. For the phase, we get $\arg(i\omega) = \frac{\pi}{2}$ for all frequencies. □

Consider (6.7) with $H_1(s)$ representing a single-input single-output system of which we write the transfer matrix (i.e., transfer function) now as $h_1(s)$ rather than $H_1(s)$, and with $H_2(s)$ being the unit feedback, i.e., $H_2(s) = 1$. It is assumed that $h_1(s)$ is strictly proper and that it does not have poles on the imaginary axis (the latter assumption is not very essential, but it simplifies the analysis to come). Equation (6.7) then becomes

$$h(s) = \frac{h_1(s)}{1 + h_1(s)}.$$

Consider the mapping $\omega \to h(i\omega)$, where ω runs from $-\infty$ to $+\infty$, and where $h(i\omega)$ then describes a curve in the complex domain (compare this with the Nyquist diagram). For

$\omega = -\infty$ this curve, to be called Γ, starts at the origin, and for $\omega = +\infty$ it ends at the origin again. Therefore, we will include the origin in Γ such that it becomes a closed curve. Assume that the real point -1 in the complex plane is not part of the curve Γ.

Theorem 6.22 *Under the assumptions formulated above, the number of encirclements of the real point* -1 *in the complex plane by the curve* Γ*, if this curve is traversed clockwise, is equal to the number of unstable poles of the closed-loop system minus the number of poles of the open-loop system.*

The open-loop system here refers to the system characterized by $h_1(s)$, and the closed-loop system, characterized by $h(s)$, refers to this system controlled by means of the unit feedback. One speaks of unstable poles if they are located in the right half-plane. This theorem is a simplified version of a more general theorem which is known as the **Nyquist criterion**. It can be used for checking whether the closed-loop system is stable.

The proof of the Nyquist criterion will not be given here, but it is based on the following theorem in complex function theory (known as **Cauchy's theorem**).

Theorem 6.23 *Assume that h is a rational (or, more generally, a meromorphic) function having no poles or zeros on a simple closed curve* $\mathscr{C}$*. Assume in addition that* $\mathscr{C}$ *is oriented clockwise. Then, with* $h'(s)$ *denoting* $\frac{d}{ds}h(s)$*, the expression*

$$\frac{1}{2\pi i}\int_{\mathscr{C}} \frac{h'(s)}{h(s)}ds$$

is equal to the number of poles of h minus the number of zeros of h, both only counted in the region bounded by $\mathscr{C}$*.*

The assumption that the feedback system had to be a unit system, as made above, is not as limited as it might seem at first hand. Assuming only single-input single-output systems, we write for (6.7),

$$h(s) = \frac{h_1(s)}{1+h_1(s)h_2(s)} = \frac{h_1(s)h_2(s)}{1+h_1(s)h_2(s)}h_2^{-1}(s),$$

which can thus be viewed as a system in series, where the two subsystems are characterized by $h_1(s)h_2(s)\big(1+h_1(s)h_2(s)\big)^{-1}$ and $h_2^{-1}(s)$, respectively, provided that both are well defined. The first of these two subsystems represents a system characterized by $h_1(s)h_2(s)$ controlled be means of a unit feedback. Thus the stability study of a system with a general feedback function can be transformed to a stability study of a system with unity feedback (and some additional requirements such as the existence of the system characterized by $h_2^{-1}(s)$).

6.8 Exercises

Exercise 6.8.1 *Consider the dynamics of the inverted pendulum as given in Equation (3.16) and assume that the position of the carriage is measured, i.e.,*

$$y = (1\ 0\ 0\ 0)x.$$

In Example 3.12 the transition matrix was calculated for this problem. Show that the impulse response function and transfer function are given by, respectively,

$$G(t) = -0.48\sinh(5t), \quad H(s) = \frac{-2.4}{s^2 - 25}.$$

Exercise 6.8.2 *Design a system of the form $\dot{x} = Ax + Bu$, $y = Cx$, with a suitably chosen initial condition $x(0)$, such that the input $u(t) = e^{-3t}, t \geq 0$, yields the output $y(t) = e^{-t} + 2e^{-2t}, t > 0$. Hint: it follows from the theory treated in Section 6.4 that a possible transfer function is*

$$h(s) = \frac{s+3}{(s+1)(s+2)}.$$

Exercise 6.8.3 *Two unit masses are connected by springs, characterized by constants k_1 and k_2 respectively, as shown in Figure 6.18. The position of the masses are indicated by*

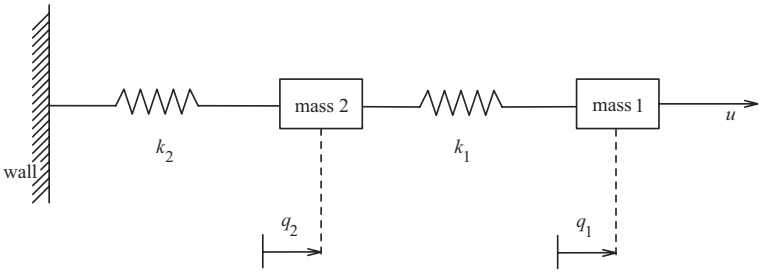

Figure 6.18 System of two masses and two springs.

q_1 and q_2, respectively. To mass 1 we can apply a force u (the input). The output is the position of mass 1, i.e., $y = q_1$. The equations describing this system are

$$\begin{aligned} \ddot{q}_1 &= u - k_1(q_1 - q_2), \\ \ddot{q}_2 &= k_1(q_1 - q_2) - k_2 q_2. \end{aligned}$$

Show that the zeros of this system are $\pm i\sqrt{(k_1 + k_2)}$, i.e., they correspond to 'real' frequencies!

Exercise 6.8.4 *In Example 6.7, show that application of an input in the form of a Heaviside step function indeed yields the output function given in the example.*

Exercise 6.8.5 *Show that the controller form of the system with the two connected springs in Exercise 6.8.3 equals*

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -k_1 k_2 & 0 & -(2k_1 + k_2) & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} u,$$

$$y = \begin{pmatrix} k_1 + k_2 & 0 & 1 & 0 \end{pmatrix} x.$$

Exercise 6.8.6 *Consider the system given by the external description*

$$\frac{d^3y}{dt^3}+4\frac{d^2y}{dt^2}+5\frac{dy}{dt}+2y=2\frac{d^2u}{dt^2}+6\frac{du}{dt}+5u.$$

Determine the transfer function (take all necessary initial conditions equal to zero) and give a partial fraction decomposition of this function. Show that the decomposition can be depicted in a flow diagram as in Figure 6.19. If the output of the local subsystems are

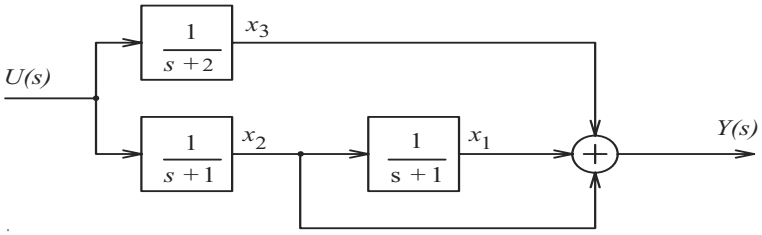

Figure 6.19 Interconnected system build of elementary blocks.

called x_1,x_2 and x_3, as indicated, give a description in state space form with the vector x as state. Prove that

$$\frac{d\bar{x}}{dt}=\begin{pmatrix}0&1&0\\0&0&1\\-2&-5&-4\end{pmatrix}\bar{x}+\begin{pmatrix}2\\-2\\3\end{pmatrix}u,\quad y=\begin{pmatrix}1&0&0\end{pmatrix}\bar{x},$$

is another state space description of the same system. Show that a nonsingular 3×3 matrix T exists such that with the transformation $x=T\bar{x}$ one state space description can be obtained from the other. Matrix T can be interpreted as a basis transformation in state space.

Exercise 6.8.7 *If the triple (A,B,C), with A an $n\times n$ matrix, B an $n\times 1$ matrix and C an $1\times n$ matrix, is a realization of the transfer function $\dfrac{q(s)}{p(s)}$, prove that the degree of $q(s)$ equals k if and only if $CA^iB=0, i=0,1,\ldots,n-k-2$ and $CA^{n-k-1}B\neq 0$.*

Exercise 6.8.8 *Consider the transfer function*

$$h(s)=\frac{s+a}{(s+b)^2+c^2}$$

with $a,b,c\in\mathbb{R}$. Follow the ideas on page 125 to obtain a flow diagram that realizes the transfer function.

Exercise 6.8.9 *Given the flow diagram depicted in Figure 6.20, determine the transfer function of the overall system. For which value(s) of α is the system BIBO stable?*

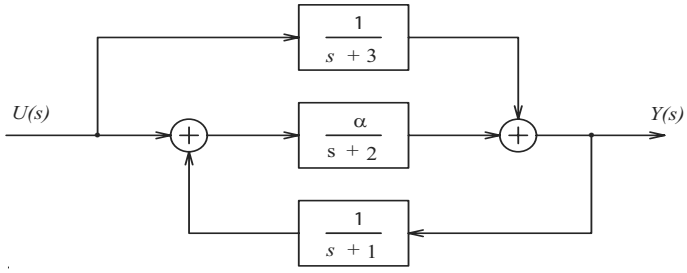

Figure 6.20 Interconnected system build of elementary blocks.

Exercise 6.8.10 *Determine a realization of the transfer matrix*

$$H(s) = \begin{pmatrix} \frac{s}{s^2+2s} & 0 & 0 \\ 0 & 0 & \frac{s-1}{(s+1)^3} \\ 0 & \frac{s^2-1}{(s+3)(s+2)} & 0 \end{pmatrix}.$$

Exercise 6.8.11 *Can you design a system of the form $\dot{x} = Ax + Bu$, $y = Cx$, with a suitably chosen initial condition $x(0)$, such that the input $u(t) = \sin t$ yields the output $y(t) = \sin t$? (Note that if the output equation would have been of the form $y = Cx + Du$, the answer would be affirmative, almost trivially. However, the design requires here that $D = 0$). If your answer is affirmative (which it should be) what conditions should the transfer function $h(s)$ satisfy?*

Exercise 6.8.12

Consider the transfer matrix

$$H(s) = \begin{pmatrix} \frac{1}{(s+2)(s+3)(s+4)} & \frac{1}{(s+3)(s+4)(s+5)} \end{pmatrix}.$$

- *Determine a realization $\dot{x} = Ax + Bu$, $y = Cx$ of $H(s)$ with $x \in \mathbb{R}^6, u \in \mathbb{R}^2$ en $y \in \mathbb{R}$ by realizing each component of $H(s)$ separately and by combining the results.*
- *Determine an alternative realization of $H(s)$ with $x \in \mathbb{R}^4, u \in \mathbb{R}^2$ en $y \in \mathbb{R}$ by observing that $H(s) = \frac{1}{(s+3)(s+4)} \begin{pmatrix} \frac{1}{(s+2)} & \frac{1}{(s+5)} \end{pmatrix}$, by realizing separate factors and by combining the results appropriately.*
- *Does there exist a realization $\dot{x} = Ax + Bu$, $y = Cx$ of $H(s)$ with $x \in \mathbb{R}^3, u \in \mathbb{R}^2$ and $y \in \mathbb{R}$? If so, determine such a realization. If not, explain why not.*

Exercise 6.8.13 *In the electrical network in Figure 6.21 the variable u denotes source voltage and y denotes the voltage drop over the most right capacitor. Determine the transfer function describing the relation between u and y, and give a state space description, i.e., give a realization, of the network. The two resistors both have value R, the value of the coil is L and the three capacitors each have value C. Note that the network can be seen as an interconnection of copies of the network in Figure 6.14 with $L = 0$ or $R = 0$.*

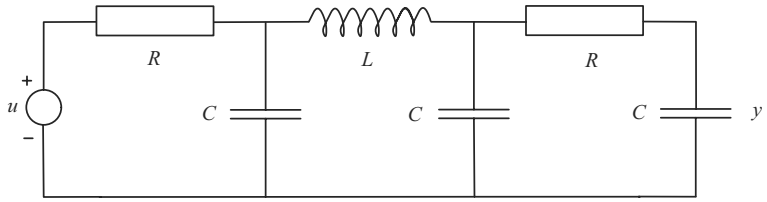

Figure 6.21 Electrical network.

Exercise 6.8.14 *Consider the PID controller described in (1.1) and determine its transfer function. Show that the controller is a non-causal system for $K'' \neq 0$. Next assume that $K'' = 0$ and determine a realization of the transfer function of the PID controller (actually a PI controller, because $K'' = 0$).*

Exercise 6.8.15 *Below you will find a number of statements. For each of statements determine whether it is true or false. Make your answer plausible by means of a simple reasoning or (counter)example.*

1. *If $H(s)$ is the transfer matrix of a linear time-invariant system, then*

$$H(s) = \frac{G(s)}{s}$$

   where $G(s)$ is the Laplace transform of the step response of the system.

2. *The gain of the system $\dot{x} = -2x + 4u$, $y = 3x$ is given by $\dfrac{12}{\sqrt{\omega^2 + 4}}$.*

3. *The poles of the continuous-time system $\dot{x} = Ax + Bu$, $y = Cx$, with A, B and C matrices of suitable sizes, not always coincide with the eigenvalues of the matrix A.*

4. *The series connection of a stable system and an unstable system can result in a stable system.*

5. *The parallel connection of a stable system and an unstable system always results in a stable system.*

6. *It is possible that the linear system $\dot{x} = Ax + bu$, $y = cx$, with A, b, c constant matrices, $x \in \mathbb{R}^n$ and $u, y \in \mathbb{R}$, does contain poles, but no zeros.*

7. *The zeros of the transfer matrix of the system $\dot{x} = Ax + Bu$, $y = Cx$, with A an $n \times n$ matrix, B an $n \times m$ matrix and C a $p \times n$ matrix, coincide with the eigenvalues of A.*

8. *The gain $|h(i\omega)|$ of a stable single-input single-output system is not greater than one, for any frequency ω.*

9. *The feedback interconnection of two stable systems can be unstable.*

10. *The linear system $\dot{x} = Ax + bu$, $y = cx$, with A, b, c constant matrices, $x \in \mathbb{R}^n$ and $u, y \in \mathbb{R}$, always contains more poles than zeros.*

11. *There exists a system* $\dot{x} = Ax + Bu$, $y = Cx$, *with* A *an* $n \times n$ *matrix,* B *an* $n \times 1$ *matrix,* C *a* $1 \times n$ *matrix and a suitable initial state* $x(0) = x_0 \in \mathbb{R}^n$, *such that* $y(t) = e^{-2t}$, *while* $u(t) = e^{-t}$.

Chapter 7

Linear Difference Systems

For most of the theory in Chapters 3, 4 and 5 for linear differential equations, an analogue exists for discrete-time systems of the form

$$\begin{array}{rcl} x(k+1) & = & A(k)x(k)+B(k)u(k), \\ y(k) & = & C(k)x(k)+D(k)u(k), \end{array} \qquad k=0,1,2,\ldots, \tag{7.1}$$

where $x(k) \in \mathbb{R}^n, u(k) \in \mathbb{R}^m$ and $y(k) \in \mathbb{R}^p$ denote the state, the input and the output, respectively, at time k. Further, $A(k), B(k), C(k)$ and $D(k)$ are matrices of suitable sizes. The index/counter k can in the context of discrete-time systems be interpreted as the time variable and is usually assumed to be integer valued and to start from 0. If the matrices $A(k), B(k), C(k)$ and $D(k)$ do not depend on time k, i.e., $A(k) = A, B(k) = B, C(k) = C$ and $D(k) = D$ for all $k \geq 0$, then the time-invariant version of (7.1) is given by

$$\begin{array}{rcl} x(k+1) & = & Ax(k)+Bu(k), \\ y(k) & = & Cx(k)+Du(k), \end{array} \qquad k \geq 0, \tag{7.2}$$

where $x(k), u(k)$ and $y(k)$ are as before, and where A, B, C and D are constant matrices of dimensions $n \times n, n \times m, p \times n$ and $p \times m$, respectively.

For other kinds of linear system descriptions, such as in Chapter 6 and Section 8.2, also discrete-time analogues exist. Concerning such analogues only the z transform, as the counterpart of the Laplace transform, will be dealt with here.

Linear difference equations often arise by discretizing linear differential equations. The reasons for such a discretization can be many. Here are some examples.

1. The analysis must be performed on a digital computer which, because of its discrete-time behavior, is more apt to discrete-time systems than to continuous-time systems.

2. One does not want to control the system by a continuous varying input function. Instead, one wants to keep the input function constant for intervals of fixed length (easier to implement). These so-called **sampling periods** will be indicated by

   $$[0,\tau),\ [\tau,2\tau),\ [2\tau,3\tau),\ \ldots$$

   The input u is constant on each of these periods.

3. The output can only be measured at time instants $0,\ \tau,\ 2\tau,\ \ldots$.

In this chapter we will - very briefly - show what the discrete-time analogues are of some of the concepts already introduced for continuous-time systems.

The solution of the homogeneous difference equation

$$x(k+1) = A(k)x(k), \quad x(k) \in \mathbb{R}^n, \quad k \geq 0, \tag{7.3}$$

can be written as

$$x(k) = \Phi_d(k,0)x(0), \quad k \geq 0, \tag{7.4}$$

where the **transition matrix** Φ_d is given by

$$\Phi_d(k,j) = \begin{cases} A(k-1)A(k-2)\cdots A(j), & k > j, \\ I, & k = j. \end{cases} \tag{7.5}$$

This transition matrix is the unique solution of the matrix difference equation

$$\Phi_d(k+1,j) = A(k)\Phi_d(k,j), \quad k \geq j, \quad \text{with} \quad \Phi_d(j,j) = I.$$

If the matrices $A(k)$ do not depend on time, i.e., $A(k) = A$ for all $k \geq 0$, then (7.3) goes over in

$$x(k+1) = Ax(k), \quad x(k) \in \mathbb{R}^n, \quad k \geq 0, \tag{7.6}$$

resulting in

$$\Phi_d(k,j) = A^{k-j}, \qquad k \geq j. \tag{7.7}$$

Please note that the transition matrix Φ_d is not necessarily nonsingular, the latter in contrast to the transition matrix for continuous-time systems. This is related to the fact that (7.3) and (7.6) are not necessarily well defined in backward time. For instance, if A in (7.6) is invertible, then $\Phi_d(k,j)$ is invertible and $\Phi_d(k,j) = (\Phi_d(j,k))^{-1}$. See also Exercise 7.1.1.

Conditions for stability (the definition of stability is much the same as in Definition 4.1) are given in the following theorem (no proof is given since the line of thought is the same as in the proof of Theorem 4.2).

Theorem 7.1 *Given is the time-invariant linear difference equation (7.6) with A an $n \times n$ matrix with different eigenvalues $\lambda_1, \ldots, \lambda_k$ $(k \leq n)$.*

- *The origin $x = 0$ is asymptotically stable if and only if $|\lambda_i| < 1$ for all $i = 1, \ldots, k$.*
- *The origin $x = 0$ is stable if and only if $|\lambda_i| \leq 1$ for all $i = 1, \ldots, k$, and for each eigenvalue λ_i on the unit circle, i.e., with $|\lambda_i| = 1$, the algebraic multiplicity and the geometric multiplicity are the same.*
- *The origin $x = 0$ is unstable if and only if $|\lambda_i| > 1$ for some $i = 1, \ldots, k$, or there is an eigenvalue λ_i on the unit circle for which the algebraic multiplicity is larger than the geometric multiplicity.*

Example 7.2 Consider the model of a national economy as developed in Example 2.4.9. The system matrix is

$$A = \begin{pmatrix} 0 & -\mu \\ m & m(1+\mu) \end{pmatrix}. \tag{7.8}$$

The characteristic polynomial is $\lambda^2 - m(1+\mu)\lambda + m\mu$. The system is for instance asymptotically stable for $\mu = 1$ and $0 < m < 1$. It is unstable if $\mu = 1$ and $m > 1$. For $\mu = 1$ and $m = 1$ it is also unstable. □

The characteristic polynomial corresponding to $x(k+1) = Ax(k)$, see (7.6), is given by $\det(zI - A) = z^n + p_{n-1}z^{n-1} + \cdots + p_1 z + p_0$. To see whether the roots of this polynomial all have modulus less than one, i.e., are all located in the open unit disc in the complex plane, the following simplified version of the **criterion of Jury** can be used, see [Gajíc and Lelíc, 1996]. This criterion can be seen as a discrete-time counterpart of the Routh criterion.

To introduce the criterion of Jury, let $p(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ be a given polynomial and make the following table, which, in principle, consists of $n+1$ rows. Determine/compute

$$\begin{array}{ccccccc} a_n & a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \\ b_{n-1} & b_{n-2} & \cdots & \cdots & b_1 & b_0 & \\ c_{n-2} & c_{n-3} & \cdots & \cdots & c_0 & & \\ d_{n-3} & d_{n-4} & \cdots & & & & \\ \vdots & \vdots & & & & & \end{array}$$

where the coefficients b_i, c_i, d_i, etc, are computed as described below.

$$b_i = a_{i+1} - \gamma_1 a_{n-1-i}, \quad \text{for } i = 0,1,\ldots,n-1, \quad \text{with } \gamma_1 = \tfrac{a_0}{a_n}$$

$$c_i = b_{i+1} - \gamma_2 b_{n-2-i}, \quad \text{for } i = 0,1,\ldots,n-2, \quad \text{with } \gamma_2 = \tfrac{b_0}{b_{n-1}}$$

$$d_i = c_{i+1} - \gamma_3 c_{n-3-i}, \quad \text{for } i = 0,1,\ldots,n-3, \quad \text{with } \gamma_3 = \tfrac{c_0}{c_{n-2}}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

Similarly as for the Routh table in Section 4.1.2, the computation of a next row in the above table breaks down if the first element of the lastly computed row is zero. Therefore, the above scheme is just continued until a row starting with a zero has been encountered. If such a row is not encountered all $n+1$ rows can be determined. The table consisting of all the rows obtained in the way just described will be referred to as the **simplified Jury table**. Now the next theorem can be stated.

Jury's criterion

The roots of the polynomial $a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$, with $a_n \neq 0$, all have a modulus less than one, i.e., all are located in the open unit disc in the complex plane, if and only if the simplified Jury table consists of $n+1$ rows and all the elements in the first column of the table have the same sign.

Example 7.3

1. Consider $p(z) = \alpha z + \beta$ with $\alpha, \beta \in \mathbb{R}$ and $\alpha \neq 0$. Then $z = -\frac{\beta}{\alpha}$ is the root of $p(z)$, and it is located in the open unit disc in the complex plane if and only if $-1 < -\frac{\beta}{\alpha} < +1$. The latter is easily seen to be equivalent to $\beta^2 < \alpha^2$. Now observe

that the simplified Jury table is given by

$$\begin{array}{cc} \alpha & \beta \\ \alpha - \dfrac{\beta^2}{\alpha} & \end{array}$$

where $\gamma_1 = \frac{\beta}{\alpha}$. Again it is easily seen that the elements in the first column of this table both have the same sign if and only if $\beta^2 < \alpha^2$.

2. Consider $p(z) = z^2 + 2z + 5$. The simplified Jury table is given by

$$\begin{array}{ccc} 1 & 2 & 5 \\ -24 & -8 & \\ -21\frac{1}{3} & & \end{array}$$

where $\gamma_1 = 5$ and $\gamma_1 = \frac{-8}{-24}$. Clearly, not all elements in the first column of this table have the same sign, implying that not all roots of $p(z)$ are inside the open unit disc. This also follows from direct calculation. Indeed, $p(s) = (z+1)^2 + 4$, so that the roots of $p(z)$ are $-1 \pm 2i$, which are both located outside the open unit disc.

3. Consider $p(z) = z^2 + 2\delta z + 1$, with $\delta \in \mathbb{R}$, and determine the simplified Jury table of $p(z)$. Note that in fact only the first two rows of this table can be established. Indeed, the table is given by

$$\begin{array}{ccc} 1 & 2\delta & 1 \\ 0 & 0 & \end{array}$$

where $\gamma_1 = 1$. Hence, $p(z)$ does not have all its roots inside the open unit disc. In fact, the roots of $p(z)$ are $-\delta \pm \sqrt{\delta^2 - 1}$, so that for $|\delta| > 1$ both roots are real, with one root is strictly inside and the other is strictly outside the unit disc. For $\delta \in [-1, +1]$, $p(z)$ has its roots on the boundary of the unit disc.

4. Consider $p(z) = -8z^3 + 12z^2 - 6z + 1$. Because $p(z) = (1-2z)^3$ all roots of $p(z)$ are located in the open unit disc. This conclusion also follows by Jury's criterion. Indeed the simplified Jury table looks like (approximately)

$$\begin{array}{cccc} -8.0000 & 12.0000 & -6.0000 & 1.0000 \\ -7.8750 & 11.2500 & -4.5000 & \\ -5.3036 & 4.8214 & & \\ -0.9205 & & & \end{array}$$

where $\gamma_1 = -0.1250, \gamma_2 = 0.5714$ and $\gamma_3 = -0.9091$. Clearly, the whole table can be computed and the elements of the first column all have the same sign, so that by Jury's criterion all roots of $p(z)$ are located in the open unit disc.

The solution of the inhomogeneous state equation in (7.1)

$$x(k+1) = A(k)x(k) + B(k)u(k), \quad k \geq 0,$$

for $x(0) = x_0$ can be written as

$$x(k) = \Phi_d(k,0)x_0 + \sum_{j=0}^{k-1} \Phi_d(k, j+1)B(j)u(j), \quad k \geq 0. \tag{7.9}$$

Note that this expression is comparable to the similar expression for continuous-time systems, see also Exercise 7.1.3. If $x_0 = 0$, then the output vector

$$y(k) = C(k)x(k) + D(k)u(k)$$

in (7.1) can be written as

$$y(k) = \sum_{j=0}^{k} K_d(k,j)u(j), \quad k \geq 0, \tag{7.10}$$

where the **impulse response matrix** is given by

$$K_d(k,j) = \begin{cases} C(k)\Phi_d(k,j+1)B(j), & k > j, \\ D(k), & k = j. \end{cases}$$

For (7.2), i.e., for the time-invariant version of (7.1), the general solution of the state equation reads

$$x(k) = A^k x(0) + \sum_{j=0}^{k-1} A^{k-j-1}Bu(j), \quad k \geq 0, \tag{7.11}$$

see also Exercise 7.1.4, and the impulse response matrix is given by

$$K_d(k,j) = \begin{cases} CA^{k-j-1}B, & k > j, \\ D, & k = j. \end{cases}$$

As far as the actual computation of $\Phi_d(k,j)$ or $K_d(k,j)$ is concerned, for constant A (and B, C and D), this can be done by writing A in its Jordan normal form by means of a coordinate transformation. This will be illustrated later on in Example 7.7.

The role of the Laplace transformation for continuous-time systems is played by the so-called ***z*-transformation** for discrete-time (time-invariant and linear) systems. Suppose $v(k), k = 0,1,2,\ldots$, is a sequence of (real or complex) numbers. The z-transform of this sequence is defined as

$$V(z) = \sum_{k=0}^{\infty} v(k)z^{-k}, \quad z \in \mathbb{C}, \tag{7.12}$$

where only those values of z will be considered for which this sum converges.

If $V(z)$ exists for a value $z = z_0$, then it will exist for all z with $|z| \geq |z_0|$. If $V(z)$ is known, the sequence $v(k)$, $k = 0,1,2,\ldots$, can be recovered in several ways. One way is to look it up in a table of z-transforms. Another way is to write $V(z)$ as a power series expansion in z^{-1} and subsequently to identify the coefficients of the terms in this series expansion with $v(k)$, see (7.12). Still another way is provided by the following theorem (no proof is given here).

Theorem 7.4 *If*

$$V(z) = \sum_{k=0}^{\infty} v(k)z^{-k}$$

converges for $|z| \geq |z_0|$, *then*

$$v(k) = \frac{1}{2\pi i} \int_{\mathscr{C}} V(z) z^{k-1} dz, \quad k \geq 0,$$

where $\mathscr{C}$ *is a closed contour in the complex plane, to be traversed counterclockwise, in the area* $|z| \geq |z_0|$, *around the origin* (*take for instance a circle with radius* $r \geq |z_0|$).

Example 7.5 Let $a \neq 0$ be a given number and consider the sequence

$$v(k) = \begin{cases} a^k & k \geq 2, \\ 0 & k = 0, 1. \end{cases}$$

Then define for $z \in \mathbb{C}$

$$V(z) = \sum_{k=0}^{\infty} v(k) z^{-k} = \sum_{k=2}^{\infty} a^k z^{-k}.$$

Hence, formally

$$V(z) = \frac{a^2}{z^2} + \left(\frac{a}{z}\right) V(z),$$

so that for all $|z| > |a|$

$$V(z) = \frac{a^2}{z} \frac{1}{(z-a)}.$$

Conversely, let $V(z)$ be as given in the last expression. Note that by partial fraction it follows that

$$V(z) = \frac{a}{z-a} - \frac{a}{z}.$$

Observe that for $|z| > |a|$

$$\frac{a}{z-a}$$

is a compact expression for the series

$$\frac{a}{z} + \frac{a^2}{z^2} + \frac{a^3}{z^3} + \frac{a^4}{z^4} + \cdots,$$

so that for $|z| > |a|$

$$V(z) = \frac{a^2}{z^2} + \frac{a^3}{z^3} + \frac{a^4}{z^4} + \cdots = \sum_{k=2}^{\infty} \frac{a^k}{z^k},$$

implying that indeed

$$v(k) = \begin{cases} a^k & k \geq 2, \\ 0 & k = 0, 1. \end{cases}$$

□

Consider the state equation $x(k+1) = Ax(k) + Bu(k)$ in (7.2). For successive values of k multiply both sides of the equation by z^{-k}, $k \geq 0$, and add the results of both sides together. The right-hand side then simply yields $AX(z) + BU(z)$, where $X(z) = \sum_{k \geq 0} x(k) z^{-k}$ and $U(z) = \sum_{k \geq 0} u(k) z^{-k}$. The left hand side equals

$$x(1) + x(2) z^{-1} + x(3) z^{-2} + \cdots,$$

which formally can be written as $zX(z) - zx(0)$. Hence,

$$zX(z) - zx(0) = AX(z) + BU(z).$$

If we solve for $X(z)$, the result is

$$X(z) = (zI - A)^{-1}BU(z) + (zI - A)^{-1}zx(0). \tag{7.13}$$

The z-transformation of $y(k) = Cx(k) + Du(k)$ yields

$$Y(z) = CX(z) + DU(z), \tag{7.14}$$

where $Y(z) = \sum_{k\geq 0} y(k)z^{-k}$. The combination of (7.13) and (7.14), with $x(0) = 0$, gives

$$Y(z) = \Big(C(zI - A)^{-1}B + D\Big)U(z).$$

The matrix $C(zI - A)^{-1}B + D$ is called the **transfer matrix** of the (discrete-time) system described in (7.2).

Suppose $u(k)$ is a periodic (complex-valued) input signal of the form

$$u(k) = u_0 e^{ik\alpha}, \quad k \geq 0, \quad \text{with } \alpha \in \mathbb{R}.$$

As is well known from the theory of difference equations, the general solution of

$$x(k+1) = Ax(k) + Bu_0 e^{ik\alpha}, \quad k \geq 0, \tag{7.15}$$

can be written as the combination of any (arbitrary) solution of the inhomogeneous difference equation (7.15) and the general solution of the homogeneous equation $x(k+1) = Ax(k), k \geq 0$. Because of the special form of the inhomogeneity, it is well known that we can assume a solution of (7.15) to be of the form

$$x(k) = x_0 e^{ik\alpha}, \quad k \geq 0, \tag{7.16}$$

assuming that $e^{i\alpha}$ is no eigenvalue of A. (If the latter is not the case, the choice for the form of a solution of (7.15) has to be modified, but the technique to follow remains the same.) Then substitution of (7.16) into (7.15) gives the following condition on x_0

$$e^{i\alpha}x_0 = Ax_0 + Bu_0, \quad \text{so that} \quad x_0 = (Ie^{i\alpha} - A)^{-1}Bu_0.$$

The general solution of (7.15) therefore is given by

$$x(k) = (Ie^{i\alpha} - A)^{-1}Bu_0 e^{ik\alpha} + A^k \tilde{x}_0,$$

with $\tilde{x}_0$ an arbitrary vector in $\mathbb{R}^n$. The vector $\tilde{x}_0$ can be determined by the initial condition of the system. If the system is asymptotically stable, then $\lim_{k\to\infty} A^k \tilde{x}_0 = 0$, and for large values of k

$$x(k) \approx (Ie^{i\alpha} - A)^{-1}Bu_0 e^{i\alpha}.$$

The above right hand side, i.e., $(Ie^{i\alpha}-A)^{-1}Bu_0e^{i\alpha}$, $k\geq 0$, is called the **stationary response** of the state, when the input signal is $u_0e^{ik\alpha}$, $k\geq 0$. The expression $A^k\tilde{x}_0$, $k\geq 0$, is called the **transient behavior**. The stationary response of the output is

$$\big(C(Ie^{i\alpha}-A)^{-1}B+D\big)\,u_0e^{ik\alpha}=H(e^{i\alpha})u_0e^{ik\alpha}.$$

This formula (for single-input single-output systems) is the discrete-time analogue of formula (6.29). Note once more that here the stationary response is completely determined by values of the transfer matrix on the unit circle (in the continuous-time case: on the imaginary-axis). Note that for the stationary response to show up the system must be asymptotically stable!

Confining ourselves to single-input single-output systems, the transfer function can be written as

$$h(z)=d+\frac{q(z)}{p(z)}, \tag{7.17}$$

with

$$\begin{aligned} q(z) &= q_{n-1}z^{n-1}+\cdots+q_1z+q_0,\\ p(z) &= z^n+p_{n-1}z^{n-1}+\cdots+p_1z+p_0. \end{aligned}$$

A state space realization corresponding to (7.17) is

$$x(k+1) = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0\\ \vdots & \ddots & \ddots & \ddots & & \vdots\\ \vdots & & \ddots & \ddots & \ddots & \vdots\\ \vdots & & & 0 & 1 & 0\\ 0 & \dots & & & 0 & 1\\ -p_0 & -p_1 & \dots & \dots & \dots & -p_{n-1} \end{pmatrix} x(k)+\begin{pmatrix} 0\\ \vdots\\ \vdots\\ \vdots\\ 0\\ 1 \end{pmatrix} u(k),$$

$$y(k) = \begin{pmatrix} q_0 & q_1 & \dots & \dots & \dots & q_{n-1} \end{pmatrix} x(k)+d\,u(k).$$

The derivation is exactly the same as for the continuous-time case. Block diagrams for discrete-time time-invariant linear systems can be drawn similarly as in the continuous-time case. See Figure 7.1. The only difference is that the integrator $\int$ must be replaced by an operator Δ defined by $\Delta x(k)=x(k-1)$. This operator is sometimes called the **delay operator** or the **backward delay operator**. Its inverse $\sigma\stackrel{\text{def}}{=}\Delta^{-1}$, i.e., $\sigma x(k)=x(k+1)$, is called the **forward delay operator**.

Example 7.6 Consider a simplified version of the national economy that is studied in Example 2.4.9. The model now is

$$\begin{aligned} x(k+1) &= px(k)-ru(k),\\ y(k) &= x(k). \end{aligned} \tag{7.18}$$

The scalar $y(k)$ is the total national income in year k and the scalar $u(k)$ is the expenditure in year k, with p and r constants. It is assumed here that $0<p<1$. The transition matrix is

$$\Phi_d(k,j)=p^{k-j}.$$

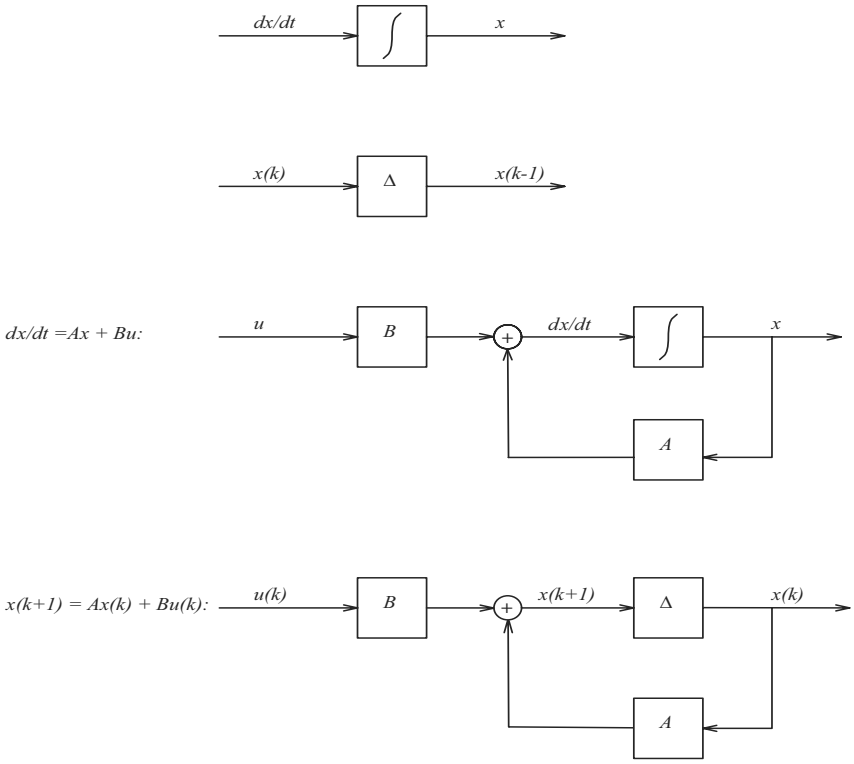

Figure 7.1 Difference continuous-time and discrete-time systems.

The general solution of (7.18) for $x(0) = x_0$ is

$$x(k) = p^k x_0 + \sum_{j=0}^{k-1} p^{k-j-1}(-r)u(j). \tag{7.19}$$

The impulse response function is

$$K_d(k,j) = \begin{cases} p^{k-j-1}(-r), & k > j, \\ 0, & k = j, \end{cases}$$

and the transfer function, for $|z| > p$, is given by

$$h(z) = \frac{-r}{z-p}.$$

Suppose that the expenditures are equal to some constant u_0, i.e., $u(k) = u_0$ for $k = 0,1,2,\ldots$. Then, for $|z| > 1$,

$$U(z) = \sum_{k=0}^{\infty} u_0 z^{-k} = \frac{1}{1-\frac{1}{z}} u_0 = \frac{z}{z-1} u_0.$$

The z-transform of the response of the output (assuming $x_0 = 0$) is, for $|z| > \max(p,1)$,

$$Y(z) = h(z)U(z) = \frac{-r}{z-p}\frac{z}{z-1}u_0.$$

In order to find $y(k), k \geq 0$, we will use the second method mentioned earlier, i.e., by means of a power series expansion. Factorization of $h(z)U(z)$ gives

$$Y(z) = \left(\frac{-p}{z-p} + \frac{1}{z-1}\right)\frac{-r}{1-p}u_0,$$

with, for $|z| > \max(p,1)$,

$$\begin{aligned}\frac{-p}{z-p} &= \frac{-p}{z}\left(1 + \frac{p}{z} + (\frac{p}{z})^2 + \ldots\right),\\ \frac{1}{z-1} &= \frac{1}{z}\left(1 + \frac{1}{z} + (\frac{1}{z})^2 + \ldots\right),\end{aligned}$$

and hence, for $|z| > \max(p,1)$,

$$Y(z) = \sum_{k=1}^{\infty}\frac{-r}{1-p}(1-p^k)u_0 z^{-k}.$$

By definition, $Y(z) = \sum_{k=0}^{\infty} y(k)z^{-k}$, and therefore

$$y(k) = \begin{cases} 0 & \text{for} \quad k = 0,\\ -r\dfrac{1-p^k}{1-p}u_0 & \text{for} \quad k \geq 1.\end{cases}$$

This result can also be obtained by using (7.19) directly; see Exercise 7.1.12. □

Example 7.7 We are given the discrete-time linear time-invariant system

$$\begin{aligned}x(k+1) &= \frac{1}{5}\begin{pmatrix} 0 & 1\\ -4 & -5\end{pmatrix}x(k) + \begin{pmatrix} 0\\ 1\end{pmatrix}u(k),\\ y(k) &= \begin{pmatrix} 2 & 1\end{pmatrix}x(k).\end{aligned}$$

Determine the transition matrix, impulse response function and the transfer function of this system.

Answer. The system matrix A has two eigenvalues, $\lambda_1 = -\frac{1}{5}$ and $\lambda_2 = -\frac{4}{5}$. Hence, it follows that A is invertible and, consequently, so is the requested transition matrix. If the two eigenvalues are the diagonal values of the diagonal matrix Λ, i.e., $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$, and if two corresponding eigenvectors v_1 and v_2 are put together to form the matrix T, i.e., $T = [v_1, v_2]$, we get for the transition matrix

$$\Phi_d(k,j) = A^{k-j} = T\Lambda^{k-j}T^{-1} =$$

$$\begin{pmatrix} 1 & -1\\ -1 & 4\end{pmatrix}\begin{pmatrix} (-\frac{1}{5})^{k-j} & 0\\ 0 & (-\frac{4}{5})^{k-j}\end{pmatrix}\begin{pmatrix} 1 & -1\\ -1 & 4\end{pmatrix}^{-1} =$$

$$\begin{pmatrix} \frac{4}{3}(-\frac{1}{5})^{k-j} - \frac{1}{3}(-\frac{4}{5})^{k-j} & \frac{1}{3}(-\frac{1}{5})^{k-j} - \frac{1}{3}(-\frac{4}{5})^{k-j} \\ -\frac{4}{3}(-\frac{1}{5})^{k-j} + \frac{4}{3}(-\frac{4}{5})^{k-j} & -\frac{1}{3}(-\frac{1}{5})^{k-j} + \frac{4}{3}(-\frac{4}{5})^{k-j} \end{pmatrix}.$$

The impulse response can be calculated according to $K_d(k,j) = C\Phi_d(k,j+1)B$ for $k \geq j$, yielding

$$K_d(k,j) = \frac{1}{3}(-\frac{1}{5})^{k-j-1} + \frac{2}{3}(-\frac{4}{5})^{k-j-1}.$$

For $j = k$ we have $K_d(k,j) = 0$, since there is no direct throughput term in the system.

The transfer function follows from $h(z) = C(zI - A)^{-1}B$ and it equals

$$\frac{z + \frac{2}{5}}{z^2 + z + \frac{4}{25}}.$$

□

Remark 7.8 Note that with Λ and T as above, the impulse response and the transfer function can easily be obtained through $K_d(k,j) = (CT)\Lambda^{k-j-1}(T^{-1}B)$ and $h(z) = (CT)(zI - \Lambda)^{-1}(T^{-1}B)$, respectively. In these expressions, CT simply is the product of C and T, whereas $T^{-1}B$ can be seen as the solution of the equation $T\xi = B$ for the unknown ξ. Hence, not the complete inverse T^{-1} needs to be determined, only the solution of the equation $T\xi = B$. Clearly, Λ^{k-j-1} and $(zI - \Lambda)^{-1}$ can be determined very easily.

The time-discrete, time-invariant, linear system (7.2), characterized by the matrices (A,B,C,D), is called **controllable** if for each $x_0, x_1 \in \mathbb{R}^n$ a time $k > 0$ and a sequence $u(0), u(1), \ldots$ exist, such that $x(k, x_0, u) = x_1$, where $x(k, x_0, u)$ stands for the state at time k, starting with initial condition $x(0) = x_0$ and having applied an input sequence u. The system is called **observable** if a time $l > 0$ exists, such that for any sequence of controls u it follows that

$$y(k, x_0, u) = y(k, x_1, u) \text{ for } k = 0, 1, \ldots, l, \quad \text{implies} \quad x_0 = x_1.$$

The conditions for controllability and observability, in terms of matrices A, B, C and D, are the same as in the time-continuous case. This will be shown in Theorem 7.10 and 7.11 below.

Sometimes one distinguishes between **null-controllability** $(x_1 = 0)$ and **reachability** $(x_0 = 0)$. It can be shown that 'standard' controllability, i.e., with arbitrary x_0 and x_1, is as strong as reachability (see also the proof of Theorem 7.10), whereas null-controllability is not as strong as 'standard' controllability.

Example 7.9 The system

$$x(k+1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x(k) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u(k)$$

is null-controllable, but not controllable. □

Theorem 7.10 *The time-discrete, time-invariant, linear system (7.2), in the context of controllability issues also referred to as the pair (A,B), is controllable if and only if*

$$\operatorname{rank}\left(B \; AB \; \cdots \; A^{n-1}B\right) = n. \tag{7.20}$$

Proof Consider the state equation in (7.2) and recall that the solution at time k, given initial state $x(0)$ and control sequence u, is given by

$$x(k) = A^k x(0) + \sum_{j=0}^{k-1} A^{k-j-1} B u(j).$$

It follows that $x(k) - A^k x(0)$ can be written as a linear combination of the columns of $B, AB, A^2B, \ldots, A^{k-1}B$. From Lemma 4.13 it then follows that, for any $k \geq 0$, the vector $x(k) - A^k x(0)$ is an element of the column space of

$$R = \begin{pmatrix} B & AB & \ldots & A^{n-1}B \end{pmatrix}.$$

If the system is controllable then for any two vectors $x_0, x_1 \in \mathbb{R}^n$ there is a time $k = \tau$ and a sequence of controls such that $x(0) = x_0$ and $x(\tau) = x_1$. This especially holds for $x_0 = 0$. Then, the vector $x(\tau) - A^\tau x(0) (= x(\tau))$ can be chosen to have any value, implying that the column space of R must be $\mathbb{R}^n$, or, equivalently, $\operatorname{rank} R = n$.

Conversely, if the column space of R is $\mathbb{R}^n$, and $x_0, x_1 \in \mathbb{R}^n$ are two arbitrarily chosen vectors, then there exist controls $u(0), u(1), \ldots, u(n-1)$, such that

$$x(n) - A^n x(0) = \sum_{j=0}^{n-1} A^{n-j-1} B u(j).$$

The obtained sequence of control brings the state from x_0 at time $k = 0$ to x_1 at time $k = n$. Hence, the system is controllable. □

Theorem 7.11 *The time-discrete, time-invariant, linear system (7.2), in the context of observability issues also referred to as the pair (C, A), is observable if and only if*

$$\operatorname{rank} \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} = n. \tag{7.21}$$

Proof Consider the equations in (7.2). Recall that the output at time k, given initial state x_0 and control sequence u, is described by

$$y(k, x_0, u) = CA^k x_0 + \sum_{j=0}^{k-1} CA^{k-j-1} B u(j) + Du(k).$$

Hence, $y(k, x_0, u) = y(k, x_1, u)$ for $0 \leq k \leq l$ if and only if $CA^k x_0 = CA^k x_1$ for $0 \leq k \leq l$, which in turn is equivalent to

$$\begin{pmatrix} C \\ CA \\ \vdots \\ CA^l \end{pmatrix} (x_0 - x_1) = 0.$$

Note that the latter is only possible for $x_0 = x_1$ (or $x_0 - x_1 = 0$) if and only if the columns of the matrix

$$\begin{pmatrix} C \\ CA \\ \vdots \\ CA^l \end{pmatrix}$$

are linearly independent. By an elementary reasoning, using the theorem of Cayley Hamilton, it follows that the last statement is equivalent to the rank condition in the statement of the current theorem.

□

Example 7.12 We are given the single-input single-output system

$$x(k+1) = Ax(k) + b(k)u(k), \quad y(k) = c(k)x(k).$$

Note that the vectors b and c may depend on k. At each time instant k only one component of the state vector can be controlled, i.e., all components of b are zero except for one, which equals 1. The user of the system may choose the position of this latter component and this position may be k-dependent. Extend the notion of controllability in the obvious way to systems of the form (7.1) and consider the following questions.

1. Give an example (at least three dimensional) such that
   - if the user chooses the same $b(k)$ vector for each k, then the system is not controllable and in addition,
   - if $b(k)$ does depend on k in a suitable way (the component which equals 1 is not always the same one), the system is controllable.
2. Does for each matrix A a suitable sequence of $b(k)$ vectors exist such that the system is controllable?

Answer question 1. Consider

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

This system is not controllable. If one chooses, however,

$$b(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad b(1) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad b(2) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

then each point $x(3)$ can be reached from any $x(0)$ and thus the system is controllable.

Answer question 2. No, if the state has dimension 3 and A is the zero matrix, it is easily seen that with any sequence of b vectors controllability is not possible. □

Because the conditions for controllability and observability for time-invariant linear discrete-time systems and continuous-time systems are the same, many of the results for the latter type of systems also hold for the former type of systems. For instance, the alternative tests for controllability and observability in Theorems 4.23 and 4.31 for continuous-time systems also hold for discrete-time systems. Also the following eigenvalue assignment results directly follow from Chapter 5.

Theorem 7.13 *Consider the time-discrete system (7.2), or specifically the pair (A,B), with A a real $n\times n$ matrix and B a real $n\times m$ matrix. Then, the system (7.2), or the pair (A,B), is controllable if and only if for each polynomial $r(\lambda)=\lambda^n+r_{n-1}\lambda^{n-1}+\cdots+r_1\lambda+r_0$, with real coefficients $r_{n-1},\ldots,r_1,r_0$, there exists a real $m\times n$ matrix F such that* $\det(\lambda I-(A+BF))=r(\lambda)$.

Theorem 7.14 *Consider the time-discrete system (7.2), or specifically the pair (C,A), with A a real $n\times n$ matrix and C a real $p\times n$ matrix. Then, the system (7.2), or the pair (C,A), is observable if and only if for each polynomial $w(\lambda)=\lambda^n+w_{n-1}\lambda^{n-1}+\cdots+w_1\lambda+w_0$, with real coefficients $w_{n-1},\ldots,w_1,w_0$, there exists a real $n\times p$ matrix K such that* $\det(\lambda I-(A-KC))=w(\lambda)$.

Clearly, when a system is controllable then the eigenvalues of $A+BF$ can be placed anywhere in the complex plane and when it is observable the eigenvalues of $A-KC$ can be placed anywhere in the complex plane. In particular, in both cases the eigenvalues can be placed in the open unit disc. However, if a system is not controllable it may be still possible that the eigenvalues of $A+BF$ can be placed in the open unit disc. Therefore, in the context of discrete-time systems, the system is called **discrete-time stabilizable** if by a suitable F all the eigenvalues of $A+BF$ can be placed in the open unit disc. Dually, a discrete-time system is called **discrete-time detectable** if by a suitable K all the eigenvalues of $A-KC$ can be placed in the open unit disc.

Now the following discrete-time versions of Remarks 5.8 and 5.13 can be proved. The proofs are, however, omitted here.

Remark 7.15 *Consider the pair (A,B), where A is a real $n\times n$ matrix and B a real $n\times m$ matrix. Then the following statements are equivalent.*

1. *The pair (A,B) is discrete-time stabilizable.*
2. $\operatorname{rank}(zI-A\ ,\ B)=n$ *for all* $z\in\mathbb{C}$ *with* $|z|\geq 1$.
3. $\operatorname{rank}(\lambda I-A\ ,\ B)=n$ *for all eigenvalues* λ *of matrix* A *with* $|\lambda|\geq 1$.

Remark 7.16 *Consider the pair (C,A), where A is a real $n\times n$ matrix and C a real $p\times n$ matrix. Then the following statements are equivalent.*

1. *The pair (C,A) is discrete-time detectable.*
2. $\operatorname{rank}\begin{pmatrix} zI-A \\ C \end{pmatrix}=n$ *for all* $z\in\mathbb{C}$ *with* $|z|\geq 1$.
3. $\operatorname{rank}\begin{pmatrix} \lambda I-A \\ C \end{pmatrix}=n$ *for all eigenvalues* λ *of matrix* A *with* $|\lambda|\geq 1$.

We will conclude this chapter with some remarks on the discretization of continuous-time systems resulting in discrete-time systems. Quite often the phenomenon one wants to study is continuous-time. It may happen that one only has measurements at discrete-time instants, which might be a reason to model the phenomenon as a discrete-time system. Also, for numerical purposes, a discrete-time model of a continuous-time phenomenon has often advantages over a continuous-time model.

Sampling consists in replacing a continuous-time signal $x(t)$, $-\infty < t < +\infty$, by the series of values $x(i\Delta)$, $i = \ldots,-2,-1,0,1,2,\ldots$, where $\Delta > 0$ is the length of the **sampling interval**. A choice to be made is how large Δ should be. A very large Δ will definitely lead to loss of information (it will be difficult to get an idea of the original continuous-time signal by solely observing the sampled signal). A very small Δ does not seem very efficient from a numerical point of view. The following theorem, called **Shannon's sampling theorem**, though it is also named after Nyquist, tells how large Δ can be chosen without loosing information.

Theorem 7.17 *If the function $x(t)$ is band-limited, i.e., a number $W > 0$ exists such that $X(i\omega) = 0$ for $|\omega| > W$, then no information is lost by sampling with a period less than or equal to π/W. (Here X denotes the 'two-sided' Laplace transform of x, i.e., $X(s) = \int_{-\infty}^{\infty} x(t)e^{-st}dt$. See also Remark 6.18.)*

If one would sample with some period Δ notwithstanding the fact that high-frequency components are present in the continuous-time signal (i.e., frequencies greater than π/Δ), then the high-frequency components are not distinguishable from low-frequency components. Therefore in the calculations, effects of these high-frequency components, not accounted for because of the sampling period chosen, will be attributed to low-frequency components. This phenomenon is called **aliasing**.

7.1 Exercises

Exercise 7.1.1 *Show that the transition matrix $\Phi_d(k,j)$ corresponding to (7.3) is not invertible if and only if $A(l)$ is not invertible for some $l = j, j+1,\ldots,k-1$.*

Exercise 7.1.2 *For each of the following matrices A, investigate if (the origin for) $x(k+1) = Ax(k)$, $k \geq 0$, is stable, asymptotically stable or unstable. Do this not only by application of Theorem 7.1, but also by solving the corresponding equations for a suitably chosen initial $x(0)$.*

$$A = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{3} \\ 0 & \frac{1}{3} & 0 \end{pmatrix}, \quad A = \begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & 4 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix},$$

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

Exercise 7.1.3 *Show by substitution that the expression in (7.9) satisfies the inhomogeneous state equation in (7.1), i.e., $x(k+1) = A(k)x(k) + B(k)u(k)$.*

Exercise 7.1.4 *Prove that the expression in (7.11) is the solution of the state equation in (7.2), i.e., $x(k+1) = Ax(k) + Bu(k)$, given initial state $x(0)$ and sequence of controls $u(k)$, $k \geq 0$. For instance, use induction with respect to k.*

Exercise 7.1.5 *In a certain country the weather forecast takes place as follows. The percentage of sunshine per day is measured. At day k there has been a_k% sunshine. The forecast for the day thereafter is made according to*

$$\hat{a}_{k+1} = (6a_k + 3a_{k-1} + a_{k-2})/10,$$

where $\hat{a}_{k+1}$ is the forecast. Write the system in state space form for this forecast where the percentage of sunshine today is the input and where the forecast for tomorrow is the output. What is the dimension of the state?

Exercise 7.1.6 *We are given the discrete-time system*

$$\begin{aligned} x(k+1) &= \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} x(k) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(k), \\ y(k) &= \begin{pmatrix} 2 & 1 \end{pmatrix} x(k). \end{aligned}$$

Determine the transition matrix, impulse response function and the transfer function of this system. Suppose the 'periodic' input signal $u(k) = (-1)^k$, $k \geq 0$, is applied to the system. What is the output response (take $x(0)$ as the zero state)? Why is the output signal not periodic?

Exercise 7.1.7 *Consider the discrete-time system*

$$x(k+1) = Ax(k) + Bu(k), \qquad x \in \mathbb{R}^n, \qquad u \in \mathbb{R}.$$

For this system

$$\operatorname{rank}\left(B\ AB\ \cdots\ A^{n-1}B\right) = r < n.$$

Prove that the state x can be steered from the initial point $x(0) = x_0$ to the final point x_1 in at most r steps if it is known that

$$x_0, x_1 \in \operatorname{im}\left(B\ AB\ \cdots\ A^{n-1}B\right).$$

Exercise 7.1.8 *In econometrics one works a lot with so-called* **ARMA models**. *They will be briefly introduced in this exercise.*

- *Consider $U(z) = \sum_{k\geq 0} u(k)z^{-k}$ and $Y(z) = \sum_{k\geq 0} y(k)z^{-k}$, where $\{u(k), k \geq 0\}$ and $\{y(k), k \geq 0\}$ are given sequences. Show that if $Y(z) = z^{-1}U(z)$ then $y(k) = u(k-1)$ $= \Delta u(k)$ for all $k \geq 1$ and $y(0) = 0$.*
- *For the so-called moving average (MA) model*

  $$Y(z) = \left(q_{n-1}z^{-1} + q_{n-2}z^{-2} + \cdots + q_1 z^{-n+1} + q_0 z^{-n}\right) U(z),$$

  derive two different types of realizations; one in state space form with matrices A, B and C, and the other in block diagram form with blocks as in Figure 7.1, where, if necessary, initial values can be taken zero.

- *The same question as above, but now with respect to the so-called autoregressive (AR) model*

$$\left(1+p_{n-1}z^{-1}+\cdots+p_1z^{-n+1}+p_0z^{-n}\right)Y(z)=U(z).$$

- *Given a mixed or ARMA model*

$$\frac{Y(z)}{U(z)}=\frac{q_{n-1}z^{-1}+q_{n-2}z^{-2}+\cdots+q_1z^{-n+1}+q_0z^{-n}}{1+p_{n-1}z^{-1}+p_{n-2}z^{-2}+\cdots+p_1z^{-n+1}+p_0z^{-n}}=$$

$$\frac{q_{n-1}z^{n-1}+q_{n-2}z^{n-2}+\cdots+q_1z+q_0}{z^n+p_{n-1}z^{n-1}+p_{n-2}z^{n-2}+\cdots+p_1z+p_0},$$

show how to merge the block diagrams of the MA and AR models just obtained, so as to obtain a realization of the ARMA model. Is it possible to construct realizations with no more than n delay-operators?

Exercise 7.1.9 *Consider the linear discrete-time time-invariant system*

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

with

$$A=\begin{pmatrix} -\frac{2}{3} & 1 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 1 & -1 & 1\\ 0 & 0 & 0 & \frac{2}{3}\end{pmatrix},\qquad B=\begin{pmatrix}0\\1\\0\\0\end{pmatrix},\quad C=\begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}.$$

Using eigenvalues of A, and their multiplicities, if necessary, determine whether

- *the system is asymptotically stable, just stable or unstable.*
- *the system is controllable,*
- *the system is observable,*
- *the system is discrete-time stabilizable,*
- *the system is discrete-time detectable.*

Exercise 7.1.10 *Using Jury's criterion determine if the following polynomial has all its roots inside the open unit disc in the complex plane.*

$$p(z)=4z^4-6z^3+2z^2+z-1.$$

Exercise 7.1.11 *Consider the polynomial*

$$p(z)=4z^2+8\rho z+1$$

with ρ a real parameter. Apply Jury's criterion to determine for which values of ρ the polynomial has all its roots inside the open unit disc in the complex plane. Verify your results by means of the known formula for the roots of $p(z)$.

Exercise 7.1.12 *Consider the simplified version of the national economy given in Example 7.6. Assume that $x(0)=0$ and $u(k)=u_0$ for all $k\geq 0$. Then show by application of (7.10) that $y(0)=0$ and $y(k)=-r\frac{1-p^k}{1-p}u_0$ for all $k\geq 1$.*

Exercise 7.1.13 *Compute the z-transform of the sequence $v(k)$ defined by*

$$v(k)=\begin{cases}\left(\frac{1}{2}\right)^k-(-1)^k & k\geq 3,\\ \left(\frac{1}{2}\right)^k & k=0,1,2.\end{cases}$$

Exercise 7.1.14 *Compute the discrete-time sequence $\{v(k),\, k\geq 0\}$, corresponding to the z-transform given by*

$$V(z)=\frac{1}{(z-a)(z-b)},$$

with $a,b\in\mathbb{R}$, $a\neq b$.

Exercise 7.1.15 *Consider the discrete-time system $x(k+1)=Ax(k)+Bu(k)$, $y(k)=Cx(k)$ with*

$$A=\begin{pmatrix}\frac{1}{2} & 0\\ 0 & -\frac{1}{4}\end{pmatrix},\quad B=\begin{pmatrix}1\\ 1\end{pmatrix},\quad C=\begin{pmatrix}1 & 1\end{pmatrix}.$$

Compute the stationary response of the output of the system for the 'periodic' input sequence $u(k)=e^{ik\alpha}$ with $\alpha=\pi$.

Exercise 7.1.16 *Determine a realization of the transfer matrix*

$$h(z)=\frac{z^3}{z^3+3z^2+2z+1}.$$

Exercise 7.1.17 *Consider the discrete-time system $x(k+1)=Ax(k)+Bu(k)$, $y(k)=Cx(k)$ with*

$$A=\frac{1}{6}\begin{pmatrix}0 & -5\\ 1 & -6\end{pmatrix},\quad B=\begin{pmatrix}1\\ 1\end{pmatrix},\quad C=\begin{pmatrix}2 & 1\end{pmatrix}.$$

Compute the transition matrix, the impulse response and the transfer function of the system.

Exercise 7.1.18 *Let z,s be complex numbers such that $z=\frac{s+1}{s-1}$. Then prove that $|z|<1$ if and only if $\operatorname{Re}s<0$. Next show that a polynomial $p(z)$ with real coefficients has all its roots in $|z|<1$ if and only if $p(\frac{s+1}{s-1})$, as function of s, has all its roots in $\operatorname{Re}s<0$. Finally, show that if*

$$p(z)=a_nz^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0,$$

with real coefficients $a_i, i=0,1,\ldots,n$, and $a_n\neq 0$, then $p(z)$ has all its roots in $|z|<1$ if and only if the polynomial

$$a_n(s+1)^n+a_{n-1}(s+1)^{n-1}(s-1)+\cdots+a_1(s+1)(s-1)^{n-1}+a_0(s-1)^n$$

has all its roots in $\operatorname{Re}s<0$, which can be verified by means of Routh's criterion. Using this equivalence, verify the outcome of Exercise 7.1.10.

Exercise 7.1.19 *Below you will find a number of statements. For each of statements determine whether it is true or false. Make your answer plausible by means of a simple reasoning or (counter)example.*

1. *The Z-transform of the series* $\{u(k), k \geq 0\}$*, with* $u(k) = (-\frac{1}{2})^k$ *for all* $k \geq 0$*, is given by* $\dfrac{2z}{2z-1}$.
2. *The discrete-time system* $x(k+1) = Ax(k) + Bu(k)$, $y(k) = Cx(k)$*, with* A, B *and* C *matrices of suitable sizes, is unstable if and only if the modulus of every eigenvalue of* A *is larger than one.*
3. *There exists no discrete-time system* $x(k+1) = Ax(k) + Bu(k)$*, with* A *an* $n \times n$ *matrix and* B *an* $n \times m$ *matrix, for which controllability and null-controllability are equivalent.*
4. *For a discrete-time system* $x(k+1) = Ax(k) + Bu(k)$*, with* A *and* B *matrices of suitable sizes, null-controllability and discrete-time stabilizability are equivalent.*
5. *If* A *has all eigenvalues inside the open unit disc and* $A - KC$ *has all eigenvalues outside the open unit disc, with* A, C *and* K *matrices of suitable sizes, then the pair* (C, A) *is observable.*
6. *There is a real* $n \times n$ *matrix* A *such that simultaneously the continuous-time system* $\dot{x}(t) = Ax(t)$ *and the discrete-time system* $x(k+1) = Ax(k)$ *both are stable, but not asymptotically stable.*

Chapter 8

Extensions and Some Related Topics

8.1 Abstract system descriptions

The input at time t will be denoted by $u(t)$ and the output by $y(t)$. For the input function, resp. output function, as functions of time we write $u(\cdot)$ and $y(\cdot)$. If no misunderstanding is possible these functions are simply written as u and y. The time will either be continuous ($t \in T$ with $T = \mathbb{R}$ or $T = [t_0, \infty)$), or be discrete ($t \in T$ with $T = \mathbb{Z}$ or $T = \{t_1, t_2, \ldots, t_n, \ldots\}$). If $T = \mathbb{R}$ we talk about continuous-time systems, if $T = \mathbb{Z}$ we talk about discrete-time systems.

Two ways exist in order to describe the dynamic behavior of systems, namely an external and an internal description. The external description considers the system as an input/output map, i.e., $y(t) = f(u,t)$, where $y(t)$ denotes the output at time t when the input function u has been applied to the system. If a system is described by means of the internal, or state space form, description, another quantity, the state x, is introduced. Later on in this section we will see the usefulness of this concept.

Definition 8.1 *[of the external description.] A system in input/output form is defined as*

$$\sum_{\mathrm{I/O}} = \{T, U, \underline{U}, Y, \underline{Y}, F\},$$

where

i) T is the time axis (i.e., $T = \mathbb{R}$ or $\mathbb{Z}$ or a subset of $\mathbb{R}$ or $\mathbb{Z}$).

ii) U is the set of input values; this set is called the input space. Quite often $U = \mathbb{R}^m$, or U is a subset of $\mathbb{R}^m$.

iii) $\underline{U}$ is a set of functions from $T \to U$; $\underline{U}$ is the set of admissible input functions; clearly $\underline{U} \subset \{f \,|\, f : T \to U\}$.

iv) Y is the set of output values. Usually $Y = \mathbb{R}^p$; Y is called the output space.

v) $\underline{Y}$ is the set of functions from $T \to Y$.

vi) F is a mapping from $\underline{U}$ to $\underline{Y}$, i.e., $F : \underline{U} \to \underline{Y}$. The mapping F defines the relation between input and output functions. If $u \in \underline{U}$, then Fu is the resulting output function. Its value at time t is denoted by $(Fu)(t)$. The mapping F is called the input/output function or the system function. It is assumed that F is **causal***, i.e., if $u_1, u_2 \in \underline{U}$ and $u_1(t) = u_2(t)$ for $t \le t'$ with $t' \in T$, then $(Fu_1)(t') = (Fu_2)(t')$ and therefore also $(Fu_1)(t) = (Fu_2)(t)$ for all $t \le t'$.*

Definition 8.2 *The system $\sum_{\mathrm{I/O}}$ is called* **linear** *if U, Y, $\underline{U}$ and $\underline{Y}$ are linear vector spaces (for example $U = \mathbb{R}^m$, $Y = \mathbb{R}^p$) and if $F : \underline{U} \to \underline{Y}$ is a linear mapping. The latter requirement means that if $u_1, u_2 \in \underline{U}$, then $F(u_1 + u_2) = Fu_1 + Fu_2$ and $F(\lambda u_1) = \lambda F u_1$ for all $\lambda \in \mathbb{R}$.*

Definition 8.3 *The system* $\sum_{I/O}$ *is called* **time-invariant** *(or* **stationary***) if*

i) T *is closed with respect to addition, i.e., if* $t_1, t_2 \in T$ *then also* $t_1 + t_2 \in T$,

ii) $\underline{U}$ *and* $\underline{Y}$ *are invariant with respect to the shift operator* S_τ, $\tau \in T$, *defined by* $(S_\tau u)(t) = u(t+\tau)$ *and* $(S_\tau y)(t) = y(t+\tau)$ *for all* $t \in T$, *i.e.,* $S_\tau \underline{U} \subset \underline{U}$ *and* $S_\tau \underline{Y} \subset \underline{Y}$ *for all* $\tau \in T$.

iii) $S_\tau F = F S_\tau$ *for all* $\tau \in T$.

To say it in a simple way, a system is time-invariant if a shift along the time axis yields an equivalent system. If $t \to u(t)$ leads to an output $t \to y(t)$, then $t \to u(t-\tau)$ should result in $t \to y(t-\tau)$. If a signal is applied one hour later, we get the same response, except for a delay of one hour.

Definition 8.4 *The system* $\sum_{I/O}$ *is called* **memoryless** *(or* **static***) if a function* f *exists,* $f : U \times T \to Y$ *such that* $(Fu)(t) = f(u(t), t)$. *This means that* Fu *at time* t *only depends on* $u(t)$ *and not on the past (or future) of* u.

Example 8.5 A mass m moves along a straight line and is connected to a wall by means of a spring with characteristic constant k. There is friction which is a function of the speed of the mass. The friction is modelled as a damper with characteristic f. An external force $u(t)$ acts on the mass. See Figure 8.1. Classical mechanics tells us that, if we want to describe the motion of the mass from a time instant t_1 onwards, while the force $u(t)$, $t \geq t_1$, is being exerted, the position and velocity of the mass at time t_1 should be known. The state of this system therefore is the vector

$$x(t) = \begin{pmatrix} q(t) \\ v(t) \end{pmatrix},$$

where q denotes the position and v the velocity. □

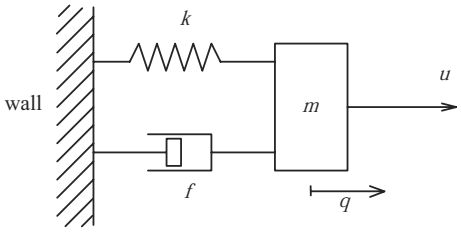

Figure 8.1 Mass-spring-damper system.

Example 8.6 Two persons play the game of goose. As time variable we denote the number of times n that both persons have thrown the die (n is increased by 1 after both persons had a turn). This is a discrete-time system. As input at time n we define

$$u(n) = \begin{pmatrix} \text{number of spots on the die at } n\text{-th throw of first person} \\ \\ \text{number of spots on the die at } n\text{-th throw of second person} \end{pmatrix}.$$

The state can be defined as

$$x(n) = \begin{pmatrix} \text{position of first person's marker on the board at time } n \\ \\ \text{position of second person's marker on the board at time } n \end{pmatrix}.$$

For simplicity reasons we have assumed that the rule 'pass your turn' does not exist. If this rule would be allowed, what could then be defined as the state of the system? □

Definition 8.7 *[of the internal description of a system] (or, equivalently, of a system in state space form). A system in state space form is defined as*

$$\sum_{\mathrm{M}} = \{T, U, \underline{U}, Y, \underline{Y}, X, \varphi, r\},$$

where

i) T, U, $\underline{U}$, Y and $\underline{Y}$ are the same as in the definition of the external description.

ii) X is the state space. Quite often $X = \mathbb{R}^n$, or X is a subset of $\mathbb{R}^n$.

iii) $\varphi : T_+^2 \times X \times \underline{U} \to X$, whereby $T_+^2 = \{(t_1, t_0) \in T^2 \text{ with } t_1 \geq t_0\}$. The mapping φ is called the state evolution function. The quantity $\varphi(t_1, t_0, x_0, u)$ denotes the state at time t_1, which was obtained by applying the input $u \in \underline{U}$ and starting from the state x_0 at time t_0. The function φ must:

a) be consistent, i.e., $\varphi(t, t, x, u) = x$.

b) satisfy the semi-group property, i.e., $\varphi(t_2, t_1, \varphi(t_1, t_0, x_0, u), u) = \varphi(t_2, t_0, x_0, u)$.

c) be determinate, i.e., if $u_1, u_2 \in \underline{U}$ and $u_1(t) = u_2(t)$, $t_0 \leq t \leq t_1$, then $\varphi(t_1, t_0, x_0, u_1) = \varphi(t_1, t_0, x_0, u_2)$.

iv) $r : T \times X \times U \to Y$ is the output function (or measurement function or observation function) $y(t) = r(t, x(t), u(t))$. It is the value of the output at time t if the system is in state $x(t)$ and $u(t)$ is the input at time t. The function $r(\cdot, x(\cdot), u(\cdot))$, must belong to $\underline{Y}$.

Definition 8.8 $\sum_{\mathrm{M}}$ *is called linear if U, Y, $\underline{U}$, $\underline{Y}$, and X are linear vector spaces and if*

i) the mapping $\varphi(t_1, t_0, \cdot, \cdot) : X \times \underline{U} \to X$ is jointly linear in both arguments, i.e., if $\varphi(t_1, t_0, x_0, u) = x$ and $\varphi(t_1, t_0, \tilde{x}_0, \tilde{u}) = \tilde{x}$ then $\varphi(t_1, t_0, \lambda x_0, \lambda u) = \lambda x$ for all $\lambda \in \mathbb{R}$, and $\varphi(t_1, t_0, x_0 + \tilde{x}_0, u + \tilde{u}) = x + \tilde{x}$.

ii) the mapping $r(t, \cdot, \cdot) : X \times U \to$ is jointly linear in both arguments.

Definition 8.9 $\sum_{\mathrm{M}}$ *is called time-invariant if $t_1 + t_2 \in T$ for any $t_1, t_2 \in T$, $S_t \underline{U} \subset \underline{U}$, $S_t \underline{Y} \subset \underline{Y}$ for all $t \in T$, and if, moreover,*

i) $\varphi(t_1 + t, t_0 + t, x_0, u) = \varphi(t_1, t_0, x_0, S_t u)$ for all $t \in T$,

ii) $r(t,x,u)$ is independent of t, and therefore can be written as $r(x,u)$.

Example 8.10 Suppose that the relationship between input function u and output function y is the following: $y(t) = u(t-\theta)$, where θ is a positive constant. The state x for this system should be such that, given x at time t and $u(s)$ with $s \geq t$, the future states x at times $s \geq t$ and future outputs y at times $s \geq t$ are uniquely determined. The function y is only determined for $s \geq t+\theta$ if $u(s)$, $s \geq t$ is given. Therefore the state must contain enough information such as to determine $y(s)$ during the interval $[t,t+\theta)$. Therefore the state at time t, i.e., $x(t)$, should at least contain the part of the function u defined on the interval $[t-\theta,t)$. It turns out that the state equals this function:

$$x(t) = u|_{[t-\theta,t)},$$

where $u|_{[t-\theta,t)}$ denotes the restriction of u to the interval $[t-\theta,t)$. □

Definition 8.11 *The system $\sum_M$ is called autonomous if U consists of only one element.*

The set U in the above definition is never empty as 'no control' can be interpreted as 'the only control'.

So far we talked about the external description and the state space form description of a system. Some words will be devoted now as to how one description can be derived from the other. Suppose $\sum_M = \{T,U,\underline{U},Y,\underline{Y},X,\varphi,r\}$ is a description in state space form. In order to obtain the corresponding $\sum_{I/O}$ the essential idea is to eliminate x from the mappings φ and r. Suppose for simplicity that $\sum_M$ is time invariant. Choose a $t_0 \in T$ and a $x_0 \in X$ (think of an initial time and an initial state) and define

$$(Fu)(t) = r(\varphi(t,t_0,x_0,u),u(t)) \quad \text{for all } t \geq t_0.$$

Thus we obtained a system

$$\sum_{I/O} = \{T \cap [t_0,\infty),U,\underline{U},Y,\underline{Y},F\}.$$

The time axis can be extended to the whole T by defining

$$x(t) = x_0,\ u(t) = u_0,\ y(t) = y_0, \quad \text{for all } t < t_0,$$

where u_0 and y_0 are constants in U and Y, respectively. For every choice we get in principle another F. The state x_0 will usually be interpreted as an equilibrium for the system. A natural choice for x_0 is the zero element of X. Similarly choices for u_0 and y_0 are the zero elements of U resp. Y. If, in addition, t_0 is chosen close to $-\infty$ (in case $T = \mathbb{R} = (-\infty,+\infty)$), we say that the system is in equilibrium or at rest at '$t = -\infty$'.

The reverse problem as how to obtain $\sum_M$ from $\sum_{I/O}$ is far more difficult. Now one has to create a state $x \in X$, instead of to eliminate x. For linear systems this problem has been solved satisfactory. A whole theory has been built around the 'creation' of the state space X and it is called realization theory.

8.1.1 Behavioral modelling

Recently a new modelling philosophy has been developed which states, in an abstract way, that signals, rather than the equations which generate these signals, is the essential result of a modelling procedure. One looks at systems as devices or 'black boxes'. Instead of trying to understand how a device is put together and how its components work in detail, we are told to concentrate on its behavior, on how it interacts with its environment.

Definition 8.12 *A dynamical system is a triple* $\Sigma = (T, \mathscr{W}, \mathscr{B})$*, where* T *represents the time-axis,* $\mathscr{W}$ *is the signal space and* $\mathscr{B} \subseteq \mathscr{W}^T$ *is the behavior, where* $\mathscr{W}^T$ *denotes the set of all functions from* T *to* $\mathscr{W}$*.*

Suppose one has a set of m scalar equations $f_i(x(t), \dot{x}(t), \ddot{x}(t), \ldots) = 0$, $i = 1, 2, \ldots, m$, where $x = (x_1, \ldots, x_n)$. Assume that the f_i-functions are defined in such a way that mathematically well defined solutions x of the differential equations exist. In this example, T is the real axis, $\mathscr{W}$ is the set of all possible x values and $\mathscr{B}$ is the set of all solutions to the differential equation. Instead of having a description by means of differential and/or algebraic equations only, one could also add inequalities.

Based on this philosophy, many concepts introduced in the earlier chapters, are phrased in a more general setting. For a neat introduction, the reader is referred to [Willems, 1991].

8.2 Polynomial representations

Chapter 6 is mainly devoted to systems descriptions in the Laplace domain. The emphasis has largely been on single-input single-output systems. The polynomials (either in the denominator or in the numerator) in the transfer matrix determine the system. As such one can also speak about 'polynomial representations' of systems. This view turns out to be particularly useful for systems with multiple inputs and outputs. The belief is that it is easier to work with polynomials (of varying degree) than with state space descriptions in which the dimensions of the states differ. One can show that the 'modelling power', which we will not formally define here, of state space representations and of polynomial representations of systems are equivalent, i.e., phenomena which can be described in one setting, can also be described in the other.

Polynomial matrices are a means of representing linear ordinary differential equations with constant coefficients. The differentiation operator $\frac{d}{dt}$ is then represented by the (Laplace) variable s. Also linear time-invariant discrete-time system can be represented by means of polynomial matrices. Then the variable s stands for the delay operator σ introduced Chapter 7.

Definition 8.13 *A* **polynomial matrix** *(in* s*) is a matrix of which the entries are polynomials in the variable* s*.*

Definition 8.14 *A linear time-invariant system is said to be described in polynomial form if the relation between the input vector* u*, of dimension* m*, and the output vector* y*, of dimension* p*, is of the form*

$$\begin{cases} P(s)\xi &= Q(s)u, \\ y &= R(s)\xi, \end{cases} \tag{8.1}$$

where P, Q and R are polynomial matrices of sizes $\bar{n}\times\bar{n}$, $\bar{n}\times m$ and $p\times\bar{n}$, respectively. The vector ξ, having $\bar{n}$ components, is called the **partial state**.

It should be emphasized that ξ, u and y in (8.1) are considered to be suitably defined vector functions of time. They are *not* vector functions in the Laplace domain. Equations (8.1) simple are (possibly higher order) differential or difference equations, that are related to each other.

Example 8.15 The classical equation of a force F acting on a point mass with mass m is (F is the input, x the output)

$$m\ddot{x} = F.$$

This equation allows at least two polynomial representations. For instance,

$$\begin{cases} ms^2\xi_1 &= F, \\ y &= \xi_1, \end{cases}$$

with the one dimensional partial state $\xi_1 = x$, and

$$\begin{cases} \begin{pmatrix} s & -1 \\ 0 & ms \end{pmatrix}\xi_2 &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}F, \\ y &= \begin{pmatrix} 1 & 0 \end{pmatrix}\xi_2, \end{cases}$$

with the two dimensional partial state $\xi_2 = (x,\dot{x})^\top$. □

If one compares the dimension of the partial state with the dimension n of the state vector in the state space description, one can conveniently restrict oneself to $\bar{n}\leq n$, hence the name *partial state*. From the example we see that two in some sense equivalent polynomial representations do not necessarily have partial states with the same dimension. For sake of completeness we now give a formal definition of 'equivalence', but warn the reader that it is somewhat technical and that it will not be used explicitly anymore in this text.

Theorem 8.16 *The two systems*

$$\Sigma_i : \begin{cases} P_i(s)\xi_i &= Q_i(s)u, \\ y &= R_i(s)\xi_i, \end{cases} \qquad i = 1,2, \tag{8.2}$$

with the same number of inputs and the same number of outputs, and of which the partial states ξ_1 and ξ_2 have dimension n_1 and n_2, respectively, are equivalent if polynomial matrices M_1, M_2, N_1 and N_2 of sizes $n_1\times n_2$, $n_2\times n_1$, $n_1\times m$ and $n_2\times m$, respectively, exist such that the following two systems

$$S_1 : \begin{cases} P_1(s)\xi_1 &= Q_1(s)u, \\ \xi_2 &= M_2(s)\xi_1 + N_2(s)u, \\ y &= R_1(s)\xi_1, \end{cases} \tag{8.3}$$

$$S_2 : \left\{ \begin{array}{rcl} \xi_1 & = & M_1(s)\xi_2 + N_1(s)u, \\ P_2(s)\xi_2 & = & Q_2(s)u, \\ y & = & R_2(s)\xi_2, \end{array} \right. \tag{8.4}$$

have identical solutions (i.e., the same input applied to both systems yields identical outputs, provided that one chooses the initial conditions suitably).

Definition 8.17 *The* **transfer matrix** *of system (8.1) is defined by* $H(s) = R(s)P^{-1}(s)Q(s)$.

The transfer matrix can always be expressed as

$$H(s) = \frac{N(s)}{d(s)}, \tag{8.5}$$

where $N(s)$ is a polynomial matrix, and $d(s)$ is a scalar polynomial in s equal to the least common multiple of the denominators appearing in $H(s)$. It has tacitly been assumed here that factors common to $d(s)$ and all entries of $N(s)$ have been cancelled.

Example 8.18 Consider the satellite example of Example 6.1. The transfer matrix can be written as

$$H(s) = \frac{N(s)}{d(s)} = \frac{1}{s^2(1+s^2)} \begin{pmatrix} s^2 & 2s \\ -2s & s^2-3 \end{pmatrix}.$$

□

Definition 8.19 *A square polynomial matrix is called* **nonsingular** *if its determinant is a polynomial that is not identically equal to zero. A square polynomial matrix is called* **unimodular** *if its determinant is a nonzero constant.*

By Cramer's rule, for instance, it follows that the inverse of a unimodular polynomial matrix is a polynomial matrix again. In general, the inverse of an invertible polynomial matrix is a rational matrix.

Example 8.20 The polynomial matrix

$$P_1(s) = \begin{pmatrix} s+1 & s+3 \\ s^2+3s+2 & s^2+5s+4 \end{pmatrix}$$

is nonsingular since $\det P_1(s) = -2s-2$. The polynomial matrix

$$P_2(s) = \begin{pmatrix} s+1 & s+3 \\ s^2+3s+2 & s^2+5s+6 \end{pmatrix}$$

is singular since $\det P_2(s) \equiv 0$. The polynomial matrix

$$P_3(s) = \begin{pmatrix} s+1 & s+3 \\ s^2+3s+3 & s^2+5s+7 \end{pmatrix}$$

is unimodular since $\det P_3(s) \equiv -2$. □

Definition 8.21 *The rank of a polynomial matrix is the size of the largest square submatrix (of this polynomial matrix) that is invertible.*

Definition 8.22 *Suppose $N(s)$, with entries $N_{ij}(s)$, is a polynomial matrix of rank k. Then $N(s)$ is in the so-called* **Smith form** *if*

- $N_{ij}(s) = 0 \quad \text{for } i \neq j,$
- $N_{ii}(s) = 0 \quad \text{for } i \geq k+1,$
- $N_{ii}(s)$ is monic and divides $N_{i+1,i+1}(s)$ for $1 \leq i < k$.

Theorem 8.23 *If $N(s)$ is a polynomial matrix, there exist unimodular polynomial matrices $U(s)$ and $V(s)$ such that $N(s) = U(s)\Gamma(s)V(s)$, where $\Gamma(s)$ has the Smith form (it is called the Smith form of $N(s)$).*

Remark 8.24 When all polynomials in $N(s)$ are constants, the theorem above is closely related to the so-called singular value decomposition, well known in matrix algebra. □

One can construct the Smith form of a polynomial matrix in a way that resembles the conventional column and row operations, as shown in the following example.

Example 8.25 Suppose

$$N(s) = \begin{pmatrix} s-a & 1 \\ 0 & s-a \end{pmatrix}.$$

Permutation of the columns yields

$$N_1(s) = \begin{pmatrix} 1 & s-a \\ s-a & 0 \end{pmatrix}.$$

Then adding the first column multiplied by $-(s-a)$ to the second column gives

$$N_2(s) = \begin{pmatrix} 1 & 0 \\ s-a & -(s-a)^2 \end{pmatrix}.$$

Next one multiplies the second column by -1, and then adds the first row multiplied by $-(s-a)$ to the second row, so as to obtain the Smith form:

$$\Gamma(s) = \begin{pmatrix} 1 & 0 \\ 0 & (s-a)^2 \end{pmatrix}.$$

It is not difficult to show that the matrices $U(s)$ and $V(s)$ are

$$U(s) = \begin{pmatrix} 1 & 0 \\ s-a & 1 \end{pmatrix}, \quad V(s) = \begin{pmatrix} s-a & 1 \\ -1 & 0 \end{pmatrix}.$$

□

The concepts of stability, controllability, observability, dynamic output feedback, poles, zeros, etc., introduced in the previous chapters, all have their natural imbedding in the theory of polynomial representations, see [Rosenbrock, 1970] or [Maciejowski, 1989].

The contents of this section remains by and large also valid in the discrete-time setting, provided one makes the assumption that the differential operator $s = \frac{d}{dt}$ is replaced by, and interpreted as, the delay operator σ defined by $\sigma x(k) \stackrel{\text{def}}{=} x(k+1)$.

8.3 Examples of other kinds of systems

These course notes have mainly dealt with linear differential (and difference) systems. Fortunately many practical phenomena can be modelled (at least approximately) by such linear systems. Many phenomena are, however, strictly speaking, nonlinear and it is not always easy, or not even always desired, to come up with an approximate linearization. For specific classes of nonlinear systems mathematical tools are available. For each of these classes a huge literature exists and the interested reader should consult the library for more information. In this section we will, very briefly, touch upon a few such classes.

8.3.1 Nonlinear systems

All systems that are not linear are by definition nonlinear. Mathematical system theory has been well developed for linear systems, but also theory exists for nonlinear systems, specifically with respect to certain classes of nonlinear systems. One such class of systems is given by

$$\dot{x} = f(x) + g(x)u, \tag{8.6}$$

where the control u appears linearly. Further, x and u are finite dimensional vectors and f and g are vector and matrix functions, respectively, of appropriate size. A typical example is the (simplified) modelling of manoeuvring a car.

Example 8.26 Suppose we can directly control the speed (by means of u_1) and the direction (by means of the steering wheel of which the position is given by u_2) of a car, then we obtain the nonlinear system

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \sin x_3 \\ \cos x_3 \\ 0 \end{pmatrix} u_1 + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} u_2,$$

which is of the form (8.6). The variables x_1 and x_2 refer to the position and the angle x_3 to the direction of the car. If one would linearize this system around any point in $\mathbb{R}^3$, the linearized system turns out to be non-controllable. However, driving a car is a controllable process, at least to the experience of most of the people. □

A theory exists which studies controllability directly in terms of the 'vector fields' $f(x)$ and $g_i(x)$, $i = 1,\ldots,m$, with g_i denoting the i-th column of the matrix g and m being the number of (scalar) controls. Toward this end one must construct the so-called **Lie brackets** of each combination of two such vector fields. Such a Lie bracket itself is also a vector field, which is added as a new member to the original set of vector fields. This augmented class of vector fields is again used to construct new Lie brackets, which are added again to the set. In this way one continues until no new vector fields are found anymore. If a rank condition on the ultimate set of vector fields obtained is fulfilled, then one has controllability of the nonlinear system.

Also general methods to study specific aspects of nonlinear systems exist, such as the concept of Lyapunov stability.

8.3.2 Descriptor systems

When modelling, especially in network theory, one sometimes encounters equations of the form

$$T\dot{x}(t) = Mx(t)+Np(t)+Pu(t), \tag{8.7}$$
$$0 = Qx(t)+Rp(t)+Su(t). \tag{8.8}$$

The corresponding system is referred to as a **differential algebraic system**. The vector $x(t) \in \mathbb{R}^n$ contains those variables of which the time derivatives appear in the equations, while the vector $p(t) \in \mathbb{R}^r$ contains the variables which only appear algebraically. The function $u(t) \in \mathbb{R}^m$ is, as usual, the input. The matrices M, N, P, Q, R and S have appropriate sizes such that the equations are well defined. If T and R happen to be square nonsingular matrices, then the equations can be written in the form (by eliminating p)

$$\dot{x} = Ax+Bu,$$

where now $A = T^{-1}M - T^{-1}NR^{-1}Q$ and $B = T^{-1}P - T^{-1}NR^{-1}S$.

Equations of the form $T\dot{x} = Mx+Pu$ are more general than equations of the form $\dot{x} = Ax+Bu$. Systems characterized by the former type of equations are referred to as **descriptor systems**. Descriptor systems allow us, for instance, to model x (or a component of x) as a time derivative of the input u (provided of course that this derivative exists). Consider

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} u,$$

then $x_1 = \dot{u}$. When considered of the form $\dot{x} = Ax+Bu$, the equation $x_1 = \dot{u}$ has x_1 as input and u as state. The notion of eigenvalue is taken over by λ's which satisfy $\det(\lambda T - M) = 0$.

8.3.3 Stochastic systems

The system which we have considered so far are all deterministic. Once the initial condition and input function are known, the future behavior is uniquely determined. There are many systems in practice in which the future is (partly) determined by processes of a stochastic/probabilistic nature. The winner of the game of goose is not determined at the outset of the game, because the evolution of the game depends on the outcomes of the die, which usually are modelled in probabilistic way. In principle, it may be possible to describe the throwing of a die in a deterministic way, but such a model would be extremely complicated and it is preferred to describe the outcome of a die probabilistically. If random influences determine the future of a system, it is called a stochastic system. A quantity $x(t)$, within a stochastic system, could be interpreted as the state at time t if, given $x(t)$ and $u(s)$, $s \geq t$, all future quantities within the system are determined in a probabilistic way. That is, for instance, the case when the probability distribution functions are uniquely determined by $x(t)$ and $u(s)$, $s \geq t$. The future behavior is then characterized by probabilistic laws, but the actual outcome of the system (who will win the game of goose) is not known before the evolution has really taken place.

Example 8.27 An industrial area can be in two situations: either the atmosphere is good (G) or the atmosphere is bad (B). In both situations two possible actions exist: start the alarm phase $(u = 1)$ or not $(u = 0)$. Depending on the atmospheric condition and the action, the atmosphere of the next day will be good or bad according to the following probabilistic rule.

condition today \ condition tomorrow	G	B
G	0.8	0.2
B	0.4	0.6

$u = 0$

condition today \ condition tomorrow	G	B
G	0.9	0.1
B	0.6	0.4

$u = 1$

The numbers in these tabular forms denote transition probabilities. If it is assumed that the transition probabilities are independent (i.e., there is no correlation with respect to time), then the state of this stochastic system is the atmospheric situation, with state 'space' $X = \{G, B\}$. □

See also Section 8.6 for other stochastic systems.

8.3.4 Automata

An automaton (plural: automata) is a special case of a discrete-time system in which the input space U and output space Y are finite. The state space X can be either finite or countably infinite. Because of the finite character of input and output spaces, they are sometimes referred to as **alphabets**, because alphabets have a finite number of elements.

Example 8.28 We consider the following situation of an oversimplified and oldfashioned marriage. The state space has three elements, namely

- x_1: husband is angry,
- x_2: husband is bored,
- x_3: husband is happy.

The input space also has three elements and consists of the behavior of the wife

- u_1: wife is quiet,
- u_2: wife shouts,
- u_3: wife cooks.

As a result of the current state and input the new state is given in the following table. The top row gives the input, the left column indicates the current state and the matrix in the 'south-east' denotes the new states.

	u_1	u_2	u_3
x_1	x_1	x_1	x_3
x_2	x_2	x_1	x_3
x_3	x_3	x_2	x_3

The output, consisting of two elements,

- y_1: husband shouts,
- y_2: husband is quiet,

is related to the current input and current state as indicated by the following table.

	u_1	u_2	u_3
x_1	y_2	y_1	y_2
x_2	y_2	y_2	y_2
x_3	y_2	y_2	y_2

□

8.3.5 Distributed parameter systems

In this section we will briefly talk about a class of systems which is (also) important from a practical point of view, but which is not discussed in these notes (apart from some examples in Section 2.4 and in this section). In all examples so far the state space X was either finite dimensional ($\mathbb{R}^n$) or even finite. In the physical examples a finite dimensional state space could be constructed because physical quantities as mass, velocity, electric charge, temperature were thought to be concentrated in one point. For some problems such a simplification may lead to inadmissible conclusions, and then electric charge, temperature, etc., not only have to be time dependent, but also location (spatial) dependent. These quantities are then elements of a function space and the state space is infinite dimensional. Such systems are called distributed systems (this in contrast to systems with finite dimensional state spaces, which sometimes are called lumped systems).

Example 8.29 Consider the flexible beam of length one, depicted in Figure 8.2. The

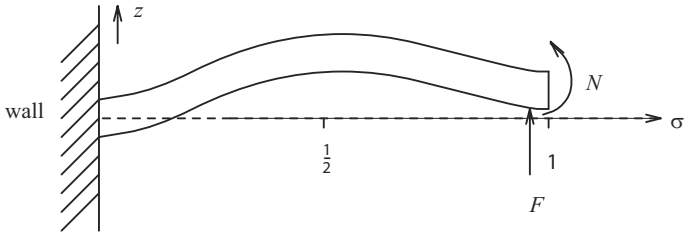

Figure 8.2 Flexible beam system.

displacement of the beam from the horizontal is denoted by z. Hence, $z(\sigma,t)$ denotes the vertical displacement of the beam at place σ and at time t. The beam is fixed horizontally into the wall (at $\sigma = 0$). This means that $z(0,t) = 0$ and $\frac{\partial z}{\partial \sigma}(0,t) = 0$ for all t. At the right end of the beam (at $\sigma = 1$) the motions of the beam are controlled by means of a

force $F(t)$ and a torque $N(t)$. Assume that the displacement in the middle of the beam (at $\sigma = 0.5$) is measured, and that this measurement is denoted $y(t)$. If gravity is not taken into account, and if the beam has uniform mass density ρ and a uniform stiffness EI, then the energy in the beam at time t equals

$$E(t) = \frac{1}{2}\rho \int_0^1 (\frac{\partial z}{\partial t})^2 d\sigma + \frac{1}{2} EI \int_0^1 (\frac{\partial^2 z}{\partial \sigma^2})^2 d\sigma.$$

The first term in this expression can be seen as the kinetic energy due to the motion of the beam, and the second term as the potential energy due to the deflection of the beam from the horizontal. If there is no loss of energy then $E(t)$ is constant for all t. Using this, it follows from $\frac{dE}{dt} = 0$, with some nontrivial mathematics (not explained here), that

$$\rho \frac{\partial^2 z}{\partial t^2} + EI \frac{\partial^4 z}{\partial \sigma^4} = 0.$$

The boundary conditions for the beam at $\sigma = 0$ are $z(0,t) = 0$, $\frac{\partial z}{\partial \sigma}(0,t) = 0$ for all t, and at $\sigma = 1$ there should hold $EI\frac{\partial^2 z}{\partial \sigma^2}(1,t) = N(t)$, $-EI\frac{\partial^3 z}{\partial \sigma^3}(1,t) = F(t)$ for all t.

The above constitutes a model for the dynamical behavior of the flexible beam subject to a force $F(t)$ and a torque $N(t)$. For the complete description of the behavior it remains to specify the initial conditions. These are the deflection and the velocity of beam at time $t = 0$, i.e., $z(\sigma,0)$, $\frac{\partial z}{\partial t}(\sigma,0)$ for all σ, $0 \leq \sigma \leq 1$.

As already can be seen from these initial conditions, the beam cannot be described by a finite number of time functions, but by an infinite number of time functions parametrized by σ, $0 \leq \sigma \leq 1$. This means that the beam cannot be described by means of a finite number of ordinary differential equations. □

8.3.6 Discrete event systems

The starting point is the difference equation

$$x(t+1) = Ax(t), \quad t = 0,1,2,\ldots, \tag{8.9}$$

with $x \in \mathbb{R}^n$. Written out in scalar equations it becomes

$$x_i(t+1) = \sum_{j=1}^{n} a_{ij} x_j(t), \quad i = 1,\ldots,n, \quad t = 0,1,\ldots \tag{8.10}$$

The only operations used in (8.9) or (8.10) are multiplication ($a_{ij} \times x_j(t)$) and addition (the $\sum$ symbol). The theory of discrete event (dynamic) systems can be considered as a study of formulas of the form (8.9) in which the operations are changed. Suppose that the two operations in (8.10) are changed in the following way. Addition becomes maximization and multiplication becomes addition. Then (8.10) becomes

$$\begin{aligned} x_i(k+1) &= \max\big(a_{i1} + x_1(k), a_{i2} + x_2(k), \ldots, a_{in} + x_n(k)\big) \\ &= \max_{j=1,\ldots,n} \big(a_{ij} + x_j(k)\big), \quad i = 1,\ldots,n, \quad k = 0,1,2,\ldots \end{aligned} \tag{8.11}$$

If an initial condition is given for both (8.9) and (8.11), then the time evolutions of (8.9) and (8.11) are completely determined. Of course the time evolutions of (8.10) and (8.11) will be different in general. Equation (8.11), as it stands, is a nonlinear difference equation. As an example conider

$$A = \begin{pmatrix} 3 & 7 \\ 2 & 4 \end{pmatrix}, \tag{8.12}$$

and take as initial condition

$$x_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{8.13}$$

Then the time evolution of (8.11) becomes

$$x(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, x(1) = \begin{pmatrix} 7 \\ 4 \end{pmatrix}, x(2) = \begin{pmatrix} 11 \\ 9 \end{pmatrix}, x(3) = \begin{pmatrix} 16 \\ 13 \end{pmatrix}, \ldots \tag{8.14}$$

We are used to thinking of the argument t in $x(t)$ as a time instant, i.e., at time instant t the state is $x(t)$. With respect to (8.11) we will introduce a different meaning for this argument. In order to emphasize this different meaning, the argument t has already been replaced by k. For a practical motivation we need to think of a network, which consists of a number of nodes and some arcs connecting these nodes. The network corresponding to (8.11) has n nodes; one node for each component x_i. Entry a_{ij} corresponds to the arc from node j to node i (this is no typo!). In terms of graph theory such a network is called a directed graph ('directed' because the individual arcs between the nodes are one-way arrows). Therefore the arcs corresponding to a_{ij} and a_{ji}, if both exist, are considered to be different.

The nodes in the network can perform certain activities; each node has its own kind of activity. Such activities take a finite time, called holding time, to be performed. These holding times may be different for different nodes. It is assumed that an activity at a certain node can only start when all preceding ('directly upstream') nodes have finished their activities and have sent the results of these activities along the arcs to the current node. Thus the arc corresponding to a_{ij} can be interpreted as an output channel for node j and, simultaneously, as an input channel for node i. Suppose that this node i starts its activity as soon as all preceding nodes have sent their results to node i (the rather neutral word 'results' is used, it could equally have been messages, ingredients or products, ...), then (8.11) describes when the activities take place. The interpretation of the quantities used is:

- $x_i(k)$: is the earliest time instant at which node i becomes active for the k-th time;
- a_{ij}: is the sum of the holding time at node j (i.e., time duration of the activity) and the travelling time from node j to node i (the rather neutral 'travelling time' is used rather than for instance 'transportation time' or 'communication time').

For the example given above, the network has two nodes and four arcs, as given in Figure 8.3. The interpretation of the number 3 in this figure is that if node 1 has started an activity, the next activity cannot start within the next 3 time units. Similarly, the time between two subsequent activities of node 2 is at least 4 time units. Node 1 sends its results

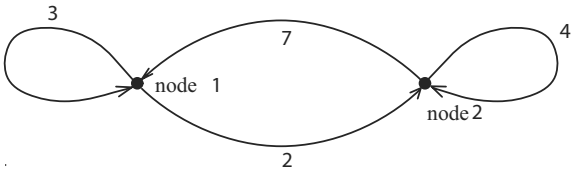

Figure 8.3 Discrete event system.

also to node 2 and once an activity starts in node 1, it takes 2 time units before the result of this activity reaches node 2. Similarly, it takes 7 time units after the initiation of an activity in node 2 for the result of that activity to reach node 1.

If we now look at the sequence (8.14) again, the interpretation of the vectors $x(k)$ is different from the initial one. The argument k is not a time instant anymore, but a counter which states how many times the various nodes have been active. At time 14 node 1 has been active twice (more precisely, node 1 has started two activities at times 7 and 11, respectively). At the same time 14, node 2 has been active three times (it started activities at times 4, 9 and 13). The counting of the activities is such that it coincides with the argument of the x vector. The initial condition is henceforth considered to be the 0-th activity.

8.4 Optimal control theory

In optimal control theory problems of the following kind are considered. A system is described by an ordinary differential equation with an input u

$$\dot{x} = f(t,x,u), \quad x(t_0) = x_0. \tag{8.15}$$

It is assumed that the conditions on f are such that a solution of this differential equation exists on a given interval $[t_0,t_1]$ for any $u \in \underline{U}$. The function u must be chosen such that a given functional (called **cost function**)

$$\int_{t_0}^{t_1} g(t,x,u)dt + q(x(t_1)) \tag{8.16}$$

is minimized, subject to $u \in \underline{U}$ and (8.15). In this problem x and u are functions with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and f, g and q are functions as follows

$$\begin{aligned} f &: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, \\ g &: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \\ q &: \mathbb{R}^n \to \mathbb{R}. \end{aligned}$$

We tacitly assume that a minimizing function u, indicated by u^*, will exist. Such existence questions also belong to the theory of optimal control. There are many variations on the problem stated above. Sometimes the function u must be chosen such that a given point x_f is reached at t_1, i.e., $x(t_1) = x_f$. This is an additional requirement on u. In the latter case the term $q(x(t_1)) = q(x_f)$ is predetermined and is independent of the used control function

u. Another variation is that the end time t_1 is not fixed explicitly, but only implicitly by means of, for instance,

$$t_1 = \min\{t \mid (t, x(t)) \in \Delta\},$$

where Δ is a given set in the $\mathbb{R}\times\mathbb{R}^n$ space. Then t_1 is the first time that the area Δ is entered. Obviously, for different input functions one may have different final times. As a specific example of problem (8.15) and (8.16) consider a linear differential equation

$$\dot{x} = Ax + Bu, \quad x(t_0) = x_0, \tag{8.17}$$

and a quadratic cost function

$$\int_{t_0}^{t_1} (x^\top Qx + u^\top Ru)dt + x^\top(t_1)Q_f x(t_1), \tag{8.18}$$

with t_1 fixed. The constant matrices Q, Q_f and R have sizes $n\times n$, $n\times n$ and $m\times m$, respectively. It is assumed that the matrices Q and Q_f are positive semi-definite, and the matrix R is positive definite. The matrices are weighting matrices, since from their (semi)-positiveness it follows that

$$x^\top Qx \geq 0, \quad u^\top Ru \geq 0, \quad x^\top(t_1)Q_f x(t_1) \geq 0,$$

for all values of x, u and $x(t_1)$. Hence, the above terms penalize deviations of x, u, and $x(t_1)$, respectively, from the zero vector. The interpretation is that system (8.17) must be controlled such that the state stays near the origin (expressed by the term $x^\top(t)Qx(t)$ in the cost function), but not at the expense of too much control effort (expressed by the term $u^\top(t)Ru(t)$). The term $x^\top(t_1)Q_f x(t_1)$ expresses the fact that we would like to have the final point $x(t_1)$ close to the origin as well. For this particular 'linear quadratic' problem the solution can be obtained in a straightforward way by a 'completing the square' argument.

$$\int_{t_0}^{t_1} \Big(x^\top Qx + u^\top Ru\Big)dt + x^\top(t_1)Q_f x(t_1) =$$
$$\int_{t_0}^{t_1} \Big(x^\top Qx + u^\top Ru + \frac{d}{dt}(x^\top P(t)x)\Big)dt + x^\top(t_0)P(t_0)x(t_0),$$

with $P(t_1) = Q_f$. The $n\times n$ matrix $P(t)$ is not completely specified yet. The only requirement sofar is $P(t_1) = Q_f$ and that it is continously differentiable. We also assume it to be symmetric: $P(t) = P^\top(t)$. The cost function becomes

$$\int_{t_0}^{t_1} \Big(x^\top Qx + u^\top Ru + \dot{x}Px + x^\top\dot{P}x + x^\top P\dot{x}\Big)dt + x^\top(t_0)P(t_0)x(t_0) =$$
$$\int_{t_0}^{t_1} \Big(x^\top Qx + u^\top Ru + (Ax+Bu)^\top Px + x^\top\dot{P}x + x^\top P(Ax+Bu)\Big)dt +$$
$$x^\top(t_0)P(t_0)x(t_0) =$$
$$\int_{t_0}^{t_1} \Big(x^\top\big(Q + \dot{P} + A^\top P + PA\big)x + u^\top Ru + u^\top B^\top Px + x^\top PBu\Big)dt +$$
$$x^\top(t_0)P(t_0)x(t_0) =$$
$$\int_{t_0}^{t_1} \Big(x^\top\big(Q + \dot{P} + A^\top P + PA\big)x + \big(u + R^{-1}B^\top Px\big)^\top R\big(u + R^{-1}B^\top Px\big) -$$
$$x^\top PBR^{-1}B^\top Px\Big)dt + x^\top(t_0)P(t_0)x(t_0) =$$

$$\int_{t_0}^{t_1} \Big(x^\top \big(Q + \dot{P} + A^\top P + PA - PBR^{-1}B^\top P\big)x + \\ \big(u + R^{-1}B^\top Px\big)^\top R\big(u + R^{-1}B^\top Px\big)\Big)dt + x^\top(t_0)P(t_0)x(t_0).$$

If we now choose $P(t)$ to satisfy the differential equation

$$\dot{P} = -A^\top P - PA - Q + PBR^{-1}B^\top P, \quad P(t_1) = Q_f, \tag{8.19}$$

then the cost function becomes

$$\int_{t_0}^{t_1} \Big(u + R^{-1}B^\top Px\Big)^\top R\Big(u + R^{-1}B^\top Px\Big)\,dt + x_0^\top P(t_0)x_0. \tag{8.20}$$

It can be shown (no proof here) that the solution to the matrix differential equation (8.19), with the indicated final condition, will exist on the interval $[t_0, t_1]$ and is unique. Because we assumed that $R > 0$, it is clear from (8.20) that the minimizing control is

$$u^*(t) = -R^{-1}B^\top P(t)x(t), \tag{8.21}$$

and that the value of the cost function will be

$$x_0^\top P(t_0)x_0,$$

when the optimal control $u^*(t)$ is applied. The matrix differential equation (8.19) plays such a fundamental role that it is named after one of its first investigators, namely it is called the **Riccati differential equation**. The requirement of $P(t)$ being symmetric is automatically fulfilled as is easily seen from studying (8.19). Indeed, if $P(t)$ is a solution of (8.19), then so is $P^\top(t)$. As the solution (at least locally) exists and is unique, it follows from the fact that Q_f is symmetric that $P(t)$ must be symmetric. One has also studied the behavior of the solution when $t_1 \to \infty$. It turns out that, if the pair (A,B) is controllable and the pair (D,A) observable, where D is defined through $DD^\top = Q$, the optimal control becomes

$$u^*(t) = -R^{-1}B^\top Px(t), \tag{8.22}$$

where now P is the smallest positive semi-definite solution of the **algebraic Riccati equation**

$$-A^\top P - PA - Q + PBR^{-1}B^\top P = 0.$$

Note that in both (8.21) and (8.22) the optimal control is given in feedback form, i.e., the current control depends on the current state. If (8.22) is substituted in (8.17), the result is

$$\dot{x} = (A - BR^{-1}B^\top P)x, \quad x(t_0) = x_0,$$

and it can be proved, subject to the conditions mentioned, that this is an asymptotically stable system. For a textbook on optimal control theory with many applications the reader is referred to [Bryson and Ho, 1969].

8.5 Parameter estimation

Sofar, we always assumed that the parameters in the models are known, i.e., we assumed that the matrices A, B, C and D in

$$\begin{aligned} \dot{x} &= Ax+Bu, \\ y &= Cx+Du, \end{aligned}$$

or

$$\begin{aligned} x(k+1) &= Ax(k)+Bu(k), \\ y(k) &= Cx(k)+Du(k), \end{aligned}$$

have known entries, or that the coefficients in the transfer function $h(s)=\dfrac{q(s)}{p(s)}$ are known. Also, it was tacitly assumed that the order n of the model was known. In some physical models all these assumptions may be reasonable. In many other models, for instance econometric models, (some of) the parameters must be estimated; they do not follow from the modelling itself. In such cases it can happen that a dependence between two variables is assumed (e.g. a linear dependence) and the coefficients specifying this dependence must be estimated given measurements of the input and output values of the system. This will be illustrated by means of the following dependence between input and output variables (in discrete-time).

$$\begin{aligned} &y(k+n)+p_{n-1}y(k+n-1)+\cdots+p_0y(k) \\ &\qquad = q_{n-1}u(k+n-1)+\cdots+q_0u(k), \qquad k=0,1,2,\ldots \end{aligned}$$

The parameters p_i and q_i, $i=0,1,\ldots,n-1$, are not known. What is known, however, is the applied input sequence and the resulting output sequence. Suppose $y(0),\ldots,y(r)$ and $u(0),\ldots,u(r)$ are known for some $r \geq 2n$. (Note that n is assumed to be fixed here.) Given these values we will try to estimate the parameters p_i and q_i, $i=0,1,\ldots,n-1$. The observations satisfy

$$\begin{aligned} y(k) &= -p_{n-1}y(k-1)-\cdots-p_0y(k-n)+ \\ &\qquad q_{n-1}u(k-1)+\cdots+q_0u(k-n)+\xi(k), \\ &\qquad\qquad k=n,n+1,\ldots,r, \end{aligned} \tag{8.23}$$

where the quantity $\xi(k)$ denotes a possible perturbation in the system due to measurement errors in the $y(i)$'s and $u(i)$'s, for instance. The quantity $\xi(k)$ makes the relation between $y(k)$ and $y(i), u(i)$, for $k-n \leq i < k$, again an exact equality. In general $\xi(k)$ will be small. Introduce the following notation.

$$\begin{aligned} \theta &= (p_{n-1},p_{n-2},\ldots,p_0,q_{n-1},q_{n-2},\ldots,q_0)^\top, \\ x(k) &= (-y(k-1),\ldots,-y(k-n),u(k-1),\ldots,u(k-n))^\top, \end{aligned}$$

then (8.23) can be written as

$$y(k)=x^\top(k)\theta+\xi(k), \quad k=n,n+1,\ldots,r.$$

The estimate of θ, denoted by $\hat{\theta}$, is defined here as that value of θ which minimizes the sum of the squares of the perturbations, i.e.,

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=n}^{r} (y(k) - x^{\top}(k)\theta)^2. \tag{8.24}$$

The estimate $\hat{\theta}$ thus defined is called the **least squares estimate**. The summation in (8.24) can be written as

$$S = \sum_{k=n}^{r} (y(k) - x^{\top}(k)\theta)^2 = (Y - X\theta)^{\top}(Y - X\theta), \tag{8.25}$$

where

$$\begin{aligned} Y &= (y(n), y(n+1), \ldots, y(r))^{\top}, \\ X &= (x(n), x(n+1), \ldots, x(r))^{\top}. \end{aligned}$$

Note that X is a matrix. Differentiation of (8.25) with respect to θ yields for the minimum

$$(X^{\top}X)\hat{\theta} = X^{\top}Y.$$

If $X^{\top}X$ is invertible then the least squares estimate can be written as

$$\hat{\theta} = (X^{\top}X)^{-1}X^{\top}Y.$$

A general introduction to parameter estimation is given in [Sorenson, 1980].

8.6 Filter theory

For linear systems filtering theory can be considered as a stochastic extension of the (deterministic) theory of observers as treated in section 5.2. It is assumed that the model is not exactly known, but that it has the form

$$\begin{aligned} \dot{x} &= Ax + Bu + Gw, \\ y &= Cx + v. \end{aligned} \tag{8.26}$$

The new terms, Gw in the system equation and v in the measurement equation, are meant to make up for errors in the system model and for measurement errors, respectively. These errors are not known a priori, but are assumed to have a certain stochastic behavior. The matrix G is assumed to be known, whereas the processes w and v will in general vary with time in an unpredictable way (quite often it is assumed that w and v are so-called **white noises**). Given the measurements $y(s), 0 \leq s \leq t$, we want to construct an estimate $\hat{x}(t)$ of the current value of $x(t)$. Before we can continue, equations (8.26) must be studied in more detail. If w is a stochastic process, drawn from a known sample space, the solution $x(t)$ of the stochastic differential equation, will also be a stochastic vector. This gives rise to many mathematical subtleties. An easier way is to start with a discrete-time system as given next.

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Gw(k), \\ y(k) &= Cx(k) + v(k). \end{aligned} \tag{8.27}$$

We now assume that $w(k)$ and $v(k)$ are independent random vectors, and also $w(k)$ and $w(l)$ are independent for $k \neq l$. Similarly, $v(k)$ and $v(l)$ are independent. Essentially, all uncertainties which enter the system and measurement equation, are uncorrelated. We also assume that $\{w(k), k \geq 0\}$ and $\{v(k), k \geq 0\}$ are zero mean, Gaussian processes with known covariances of R and Q, respectively. The matrices R and Q are assumed to be positive definite. The input $u(k)$ to the system is assumed to be deterministic (we know what we put into the system). It can be shown that the solution $x(k)$ to the difference equation (8.27) is also a Gaussian vector. We now define the estimate $\hat{x}(k+1)$ of $x(k+1)$ – the latter vector is only known probabilistically – by minimizing the conditional minimum variance, given the measurements up to time instant k, as follows.

$$\hat{x}(k+1) = \arg\min_x E\{\|x(k+1) - x\|^2 \big| y(0), y(1), \ldots, y(k)\}.$$

$E\{\cdot|\cdot\}$ denotes a conditional expectation. Other definitions of the estimate are possible, but the above turns out to be an attractive one. It says that the squared distance between the estimate and the actual value of the state must be as small as possible given all the past measurements. It turns out that $\hat{x}(k+1)$ can be determined recursively by

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + K(k)\big(y(k) - C\hat{x}(k)\big) \tag{8.28}$$

The matrix $K(k)$ can be expressed in the known matrices A, B, C, G, Q, and R (will not be shown here). In the literature, Equation (8.28) is often referred to as the **Kalman filter**. The reader should note the resemblance between the observer form in (5.6) and (8.28)! Both equations have a correction term, in which $C\hat{x}$ is the predicted value of the output and y is the actual measurement. If these two values differ, a correction appears in (8.28) (and in (5.6)) for the update from $\hat{x}(k)$ to $\hat{x}(k+1)$ (and something similarly holds with respect to (5.6)). Formulas exist which give the accuracy of $\hat{x}(k+1)$. The estimate $\hat{x}(k+1)$ is also a stochastic vector. In fact, it is Gaussian and the accuracy of $\hat{x}(k+1)$ is expressed in terms of its mean and covariance. For an excellent introduction to this subject see [Anderson and Moore, 1979].

8.7 Model reduction

In the theory of model reduction one replaces a model by a simpler one, which still catches the essential behavior, in order to get a better insight and/or to get numerical (or analytical) results faster, with less effort. In the state space description one could try to replace $\dot{x} = Ax + Bu$, $y = Cx$ by $\dot{\bar{x}} = \bar{A}\bar{x} + \bar{B}u$, $y = \bar{C}\bar{x}$, where the new state $\bar{x}$ has fewer elements than the original state x, and where the behavior of the 'barred' system resembles the behavior of the original system in some way. One thus speaks of **model reduction** since the dimension of the state space has been reduced. If the starting point would have been a transfer function, one could try to replace this transfer function by another one of which the degree of the numerator and of the denominator are smaller than of the original transfer function. We will only devote a few words on model reduction in state space here.

As an example consider

$$\dot{x} = \begin{pmatrix} -1 & 0 \\ 0 & -10 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix} u, \quad y = (1 \;\; 1)x.$$

Intuitively one may approximate this system by the one dimensional system

$$\dot{x}_1 = -x_1 + u, \quad y = x_1,$$

one simply deletes the parts of the system related to the smaller eigenvalue (-10). However, whether it is wise to replace

$$\dot{x} = \begin{pmatrix} -1 & 0 \\ 0 & -10 \end{pmatrix} x + \begin{pmatrix} 1 \\ 10 \end{pmatrix} u, \quad y = (1 \;\; 12)x$$

by the same reduced system $\dot{x}_1 = -x_1 + u,\ y = x_1$ is not so clear anymore.

For a more fundamental approach to model reduction one starts with the so-called **controllability Gramian**

$$P = \int_0^\infty e^{At} B B^\top e^{A^\top t} dt, \tag{8.29}$$

and the **observability Gramian**

$$Q = \int_0^\infty e^{A^\top t} C^\top C e^{At} dt. \tag{8.30}$$

These Graminans are only well defined for asymptotically stable systems.

The eigenvalues of P provide a measure for controllability and the eigenvalues of Q provide a measure for observability. If some of these (nonnegative) eigenvalues are close to zero, then the system is poorly controllable, respectively, observable. One can easily show that these eigenvalues are *not* invariant with respect to coordinate transformations. One speaks of a **balanced realization** of the system if the coordinates are chosen in such a way that for the transformed system P and Q are equal, i.e., $P = Q$, and diagonal. One can also easily show that the eigenvalues of the product PQ are invariant under state space transformations and hence they can be viewed as input/output invariants.

Below the following notation is used. Given a square matrix M with the eigenvalues in some order, write $\lambda_i(M)$ for the ith eigenvalue of M. Assume that $\operatorname{Re}\lambda_i(A) < 0$, for all $i = 1,2,\ldots,n$, then the **Hankel singular values** σ_i of the system are defined as $\sigma_i = \{\lambda_i(PQ)\}^{\frac{1}{2}}$, for $i = 1,2,\ldots,n$, where by convention one orders these values in such a way that $\sigma_i \geq \sigma_{i+1}$. If one wants to reduce the dimension of the system one could disregard those 'parts' of the system which correspond to the smaller Hankel singular values. It turns out that such reduced models still capture the controllability and observability behavior of the original system (provided one keeps those 'parts' of the system which correspond to the larger Hankel singular values).

8.8 Adaptive and robust control

The areas of adaptive control and of robust control both are full-grown scientific areas. In adaptive control one considers systems characterized by some parameters. In case of linear systems these parameters are, for instance, some elements of the A and/or B matrix, which slowly change their values with respect to time (for instance, due to aging or changes in the environmental conditions). For the analysis and design of the feedback control one considers these parameters as constants. One, however, monitors the values

of the parameters. If these values change markedly from their nominal values, one will adjust the feedback control to these changed circumstances. One 'resets' the parameter values and calculates the new control. One 'adapts' the design to the new parameter values, hence the name **adaptive control**.

The theory of robust control yields (usually simple) controllers which maintain the 'stability robustness' of the overall system and/or the 'performance robustness', in spite of uncertain parameters. One assumes that upper bounds on these uncertainties are known and given. There are basically two approaches for solving the robust control problem: the frequency domain approach and the time domain approach. In many cases the most important stability robustness measure is the maximum bound of the tolerable perturbation for maintaining stability. Consider the model $\dot{x} = Ax + Bu$ for which a feedback law $u = Fx$ has been designed such that the closed-loop system is asymptotically stable. One will use the same feedback law for the system $\dot{x} = (A + \Delta A(t))x + Bu$, where the matrix $\Delta A(t)$ satisfies $\|\Delta A(t)\| \leq a$. Will this perturbed system with the feedback law based on the nominal model still be asymptotically stable? It is assumed here that the notation $\|A\|$ refers to the **spectral norm** of the matrix A, i.e., $\|A\|^2 = \big(\lambda_{\max}(AA^\top)\big)$ and that a is a positive constant. For an asymptotically stable A and a constant ΔA, asymptotic stability of $\dot{x} = (A + \Delta A)x$ is assured if

$$\|\Delta A\| \leq a = \Big(\sup_{0 \leq \omega \leq \infty} \|i\omega I - A\|\Big)^{-1}.$$

Hence, the system with feedback, $\dot{x} = (A + \Delta A + BF)x$, is asymptotically stable if $\|\Delta A\| \leq \big(\sup_{0 \leq \omega \leq \infty} \|i\omega I - A - BF\|\big)^{-1}$. This uncertainty bound can be maximized by choosing an appropriate stabilizing feedback matrix F.

Robust control is sometimes also approached from another side. The system is supposed to be given by $\dot{x} = Ax + Bu + Gv$, where the term Gv incorporates everything one is uncertain about or which is unknown. This term consists of a known matrix G and an unknown control v. This control v is supposed to be chosen by Nature which accidentally might try to upset our own goal as much as possible. The question is whether we can still control the system in an appropriate way in spite of the fact that another decision maker interferes in an unpredictable manner. If so, one also speaks of **robust control**. If one assumes that Nature tries to counteract our goals as much as possible, one speaks of a **worst case scenario** for finding the control law for u. The corresponding theory belongs to the field of **differential games** in which one deals with systems in which more decision makers interact with opposite goals.

8.9 Exercises

Exercise 8.9.1 *We are given a time-invariant, linear and causal system of which we know that the following input $u(t)$ yields the given output $y(t)$*

$$u(t) = \begin{cases} 1 & 0 \leq t < 2, \\ 0 & \text{otherwise}, \end{cases} \qquad y(t) = \begin{cases} t & 0 \leq t < 2, \\ 4 - t & 2 \leq t < 4, \\ 0 & \text{otherwise}. \end{cases}$$

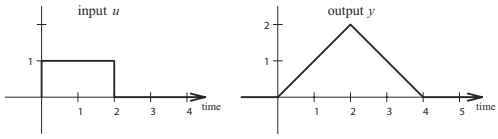

Figure 8.4 Input-output pair for a time-invariant, linear, causal system.

Determine the output function $\tilde{y}(t)$ which corresponds to the input $\tilde{u}(t)$, where

$$\tilde{u}(t) = \begin{cases} 1 & 0 \le t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

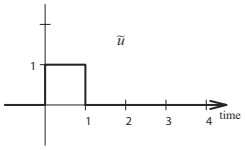

Figure 8.5 Which output corresponds to this input?

Exercise 8.9.2 *Argue that a polynomial representation of $\dot{x} = Ax + Bu$, $y = Cx$ equals $(sI - A)\xi = Bu$, $y = C\xi$.*

Exercise 8.9.3 *Show that the Smith form of the matrix $N(s)$ introduced in Example 8.18 equals*

$$\begin{pmatrix} 1 & 0 \\ 0 & s^2(1+s^2) \end{pmatrix}.$$

Exercise 8.9.4 *Give a state space description of a discrete-time system with sets $U = Y = \{0,1\}$, such that the output y at time t equals 1 if the input until (and not including) t has shown an even number of 1's, and equals 0 otherwise.*

Exercise 8.9.5 *Assume that $\rho = 1$ and $EI = 1$ in the model in Section 8.3.5, i.e.,*

$$\frac{\partial^2 z}{\partial t^2} + \frac{\partial^4 z}{\partial \sigma^4} = 0.$$

To see that there exists an infinite number of solutions to this equation, check that for any real λ and μ both $e^{-\lambda\sigma}\cos(\lambda^2 t)$ and $\cos(\mu\sigma)\cos(\mu^2 t)$ are independent solutions. Here initial and boundary conditions are not yet taken into account. Can you find additional solutions?

Exercise 8.9.6 *Show that the controllability Gramian P in (8.29) and the observability Gramian Q in (8.30) satisfy the Lyapunov equations*

$$\begin{aligned} AP + PA^\top &= -BB^\top, \\ A^\top Q + QA &= -C^\top C, \end{aligned}$$

respectively, where it is recalled that A is asymptotically stable. Compare also with (4.3).

Chapter 9

MATLAB Exercises

This chapter contains a collection of problems and their solutions that can be used for this course on system theory. The problems are solved using the software package MATLAB. For most of them also the MATLAB *Control Toolbox* must be used.

The goal of these exercises is twofold: first of all they serve as an illustration of the theory covered by this book, and secondly they show the usefulness of MATLAB for solving larger control problems. Most of the problems in this book have moderate sizes, but it will be clear that for larger systems it becomes hard, if not impossible, to do the necessary calculations by hand.

In the first section the exercises are given. Two of them come from the previous chapters. In the second section solutions using MATLAB are presented.

9.1 Problems

Exercise 9.1.1 (Moving average) Let $u(k), k = 0, 1, \ldots$, be a sequence of measurements. In order to smoothen these measurements somewhat, a moving average (of three measurements) is defined as

$$y(k) = \frac{1}{3}\left(u(k) + u(k-1) + u(k-2)\right).$$

Generate a random sequence of measurements using `rand`. Then compute the moving average of this sequence. Plot the sequence and the moving average. What effect has raising the number of samples to be averaged?

Exercise 9.1.2 (Thermal capacity of a wall) Consider a barrel in which a liquid is heated. We are interested in the temperature evolution of the liquid as well as that of the wall of the barrel. In Figure 9.1 part of the liquid and the wall has been represented schematically.

For the liquid the following relation holds:

$$P - Q_1 = C\frac{d\Theta}{dt}.$$

Here P is the power added by electrical heating, Q_1 is the heat transfer to the wall, C is the heat capacity of the liquid and Θ is the temperature of the liquid. For the heat transfer to the wall it holds that

$$Q_1 = \alpha_1 A(\Theta - \Theta_w),$$

where α_1 is the heat transfer coefficient of the liquid to the wall, A is the total wall area and Θ_w is the wall temperature. The thermal conductivity of the wall is supposed to be

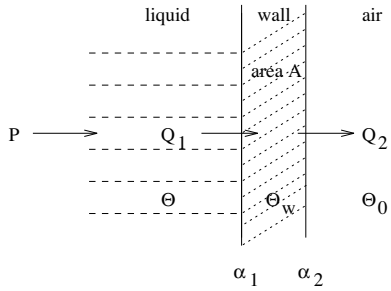

Figure 9.1 Liquid in a barrel.

infinitely large, hence the wall temperature can be regarded as homogeneous. For the wall it holds that

$$Q_1 - Q_2 = C_w \frac{d\Theta_w}{dt},$$

where Q_2 is the heat transfer to the air and C_w represents the heat capacity of the wall. For the heat transfer to the air we have that

$$Q_2 = \alpha_2 A(\Theta_w - \Theta_0),$$

where α_2 is the heat transfer coefficient of the wall to the air and Θ_0 is the air temperature.

a. Determine a two dimensional state space representation, where

$$x = (\Theta \quad \Theta_w)'; \quad u = (P \quad \Theta_0)'; \quad y = x.$$

b. Suppose (in appropriate units) $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $A = 3$, $C = 0.4$ and $C_w = 0.2$. Plot the temperature evolution of both the liquid and the wall, over a time span of 15 time units, when the air temperature is constant and equal to 20, and $x_0 = (\Theta(0) \quad \Theta_w(0))' = (0 \quad 10)'$. A continuous heating of level 1 is being supplied to the system. Choose an appropriate time step.

Note: `lsim` expects a matrix with a row vector for every time step.

c. Determine, analytically and from the plot, the finally reached equilibrium state x_{eq}.

Exercise 9.1.3 (Four moving vehicles)

Consider four vehicles moving in a single lane as shown in Figure 9.2. Let y_i, v_i, m_i and u_i be the position, velocity, mass of and the applied force to the i-th vehicle, respectively. Let k be the viscous friction coefficient, the same for all four vehicles. Then we have, for $i = 1,2,3,4$, that

$$\begin{aligned} v_i &= \dot{y}_i, \\ u_i &= k v_i + m_i \dot{v}_i. \end{aligned}$$

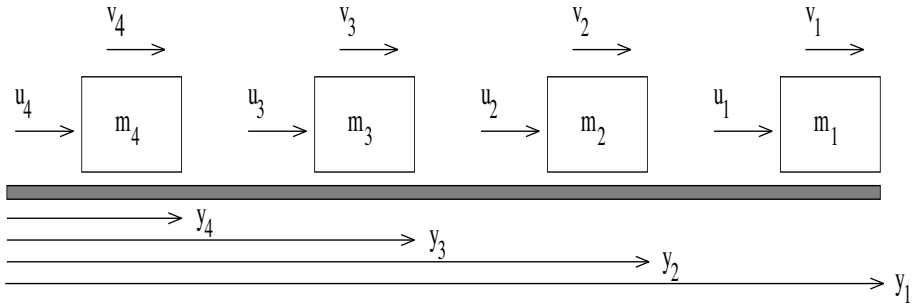

Figure 9.2 Four moving vehicles.

The purpose of this problem is to maintain the distance between adjacent vehicles at a predetermined value h_0 and to maintain the velocity of each vehicle as close as possible to a desired velocity v_0. Define

$$\begin{array}{rcll} \overline{y}_{i,i+1}(t) & = & y_i(t)-y_{i+1}(t)-h_0, & i = 1,2,3 \\ \overline{v}_i(t) & = & v_i(t)-v_0, & i = 1,2,3,4 \\ \overline{u}_i(t) & = & u_i(t)-kv_0, & i = 1,2,3,4. \end{array}$$

The term kv_0 is the force needed to overcome the friction for the vehicles to maintain their velocity at v_0. Now the problem reduces to finding $\overline{u}_i(t)$ such that $\overline{y}_{i,i+1}(t)$ and $\overline{v}_i(t)$ are as close as possible to zero for all t.

a. Derive the state-space description of the system with

$$x(t) = (\ \overline{v}_1(t)\quad \overline{y}_{1,2}(t)\quad \overline{v}_2(t)\quad \overline{y}_{2,3}(t)\quad \overline{v}_3(t)\quad \overline{y}_{3,4}(t)\quad \overline{v}_4(t)\ )',$$

the input consisting of the $\overline{u}_i(t)$ and as output the state of the system. What do you notice about h_0 and v_0?

b. Choose $m_1 = 5$, $m_2 = 4$, $m_3 = 3$, $m_4 = 2$ and $k = 8$. Plot $\overline{y}_{i,i+1}$ $(i = 1,2,3)$ when the applied forces are: $u_1(t) = 6$, $u_2(t) = 12$, $u_3(t) = 20$ and $u_4(t) = 24$ for all $t \geq 0$. Take $x(0) = (0\ 1\ 0\ 1\ 0\ 1\ 0)'$ and simulate over 3 time units. What happens?

Exercise 9.1.4 (Pole placement of four-vehicle system) Consider again Exercise 9.1.3. Use `place` to determine the feedback matrix F that places the poles of the system with feedback at -1, -2, -3, $-1 \pm i\sqrt{7}$ and $-2 \pm i\sqrt{5}$. Simulate the system with zero external input using this feedback matrix if $x(0) = (2\ 1\ 3\ 1\ 4\ 1\ 5)'$.

Exercise 9.1.5 (Observer for four-vehicle system) Continuation of Exercise 9.1.3 and Exercise 9.1.4.

a. Is the system observable in the case where the velocities $\overline{v}_i$ are considered to be the outputs? Is it detectable?

b. And what if the relative positions $\overline{y}_{i,i+1}$ are the outputs?

c. In the case of **b**, heuristically construct an observer such that $|x-\hat{x}| < 0.05$ within two time units, if no input is applied. Take as initial condition $(x-\hat{x})(0) = (2\ 1\ 3\ 1\ 4\ 1\ 5)'$.

Exercise 9.1.6 (From external description to state space description) Consider the system (from Exercise 6.8.6) given by the external description

$$\frac{d^3y}{dt^3}+4\frac{d^2y}{dt^2}+5\frac{dy}{dt}+2y=2\frac{d^2u}{dt^2}+6\frac{du}{dt}+5u\,.$$

a. Determine the transfer function.

b. Use `residue` to obtain the partial-fraction expansion of the transfer function.

c. Apply `tf2ss` to get a state-space realization.

d. Use `ss2tf` to check that the transfer function of the system

$$\frac{d\bar{x}}{dt}=\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -5 & -4 \end{pmatrix}\bar{x}+\begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}u\,, \quad y=\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}\bar{x}$$

is identical to the one found in **a**. Determine what quantities the state vector $\bar{x}$ is composed of.

Exercise 9.1.7 (Rocket) For the rocket in Figure 9.3 the simplified equation of motion is

$$I\frac{d^2\varphi}{dt^2}=k\alpha,$$

where I is the moment of inertia around the centre of gravity, φ is the course angle relative to a fixed coordinate system, α is the angle between the engine and the rocket axis, and k is a constant depending on the thrust power of the engine.

With $k/I = A$ it follows that

$$\frac{\mathscr{L}(\varphi)}{\mathscr{L}(\alpha)}=\frac{A}{s^2}.$$

In Figure 9.4 the block diagram of the rocket course control is depicted. Here the gain is K and the transfer function of the controller is

$$G(s)=\left(1+\frac{1}{2s}\right)\frac{2s+1}{0.1s+1}.$$

a. Determine the transfer function $H(s)$ of r to φ.

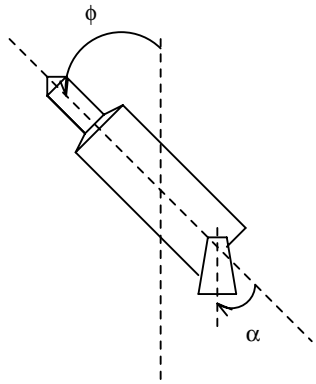

Figure 9.3 Rocket with adjustable engine.

b. Determine the conditions under which the system is stable, using the *Routh* criterion.

c. The question for which values of K the system is stable (for fixed A) can also be answered using the *Root Locus* method. This means that for $0 \leq K \leq \infty$ the positions of the poles of $H(s)$ are plotted, such that the so-called *root locus* is obtained; inspection gives the stabilizing values of K. Plot the poles of $H(s)$ for $0 \leq K \leq 100$, where $A = 0.05$. For which values of K is the system stable?

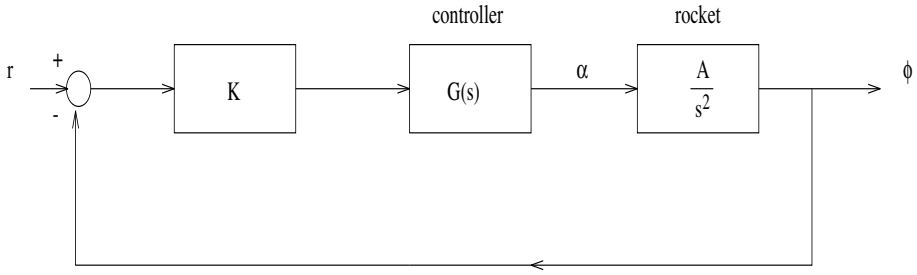

Figure 9.4 Rocket course controller.

9.2 Solutions

Exercise 9.2.1 (Moving average) The program below computes a vector `u` of measurements. The moving average `y` is computed. Both are plotted in Figure 9.5.

```
N = 50;
u = 10 + rand(N,1);            % random measurements
y=zeros(N,1);
```

```
for k = 3:N,
  y(k) = (u(k) + u(k-1) + u(k-2))/3;
end;                              % the moving average
plot(u,'--');
axis([0 N 9 12]);
hold;
plot(y,'-');
title('Moving average');
xlabel('k')
ylabel('u(k) (- -)  and  y(k) (---)')
```

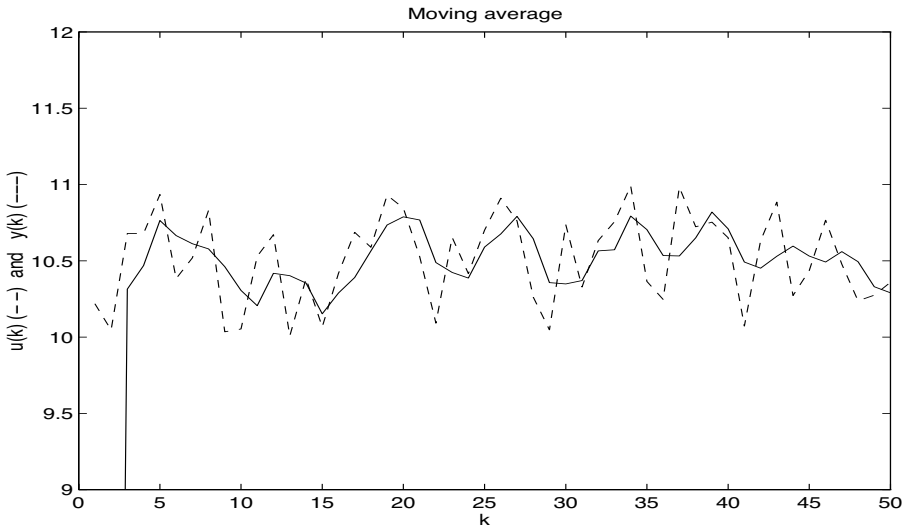

Figure 9.5 Moving average.

Three effects of raising the number of samples to be averaged are:

- lowering the variance of y,
- raising the autocorrelation of y, and
- raising the time lag of y with respect to u.

Exercise 9.2.2 (Thermal capacity of a wall)

a.

$$\dot{x} = \begin{pmatrix} -\frac{\alpha_1 A}{C} & \frac{\alpha_1 A}{C} \\ \frac{\alpha_1 A}{C_w} & -\frac{\alpha_1 A}{C_w} - \frac{\alpha_2 A}{C_w} \end{pmatrix} x + \begin{pmatrix} \frac{1}{C} & 0 \\ 0 & \frac{\alpha_2 A}{C_w} \end{pmatrix} u$$
$$= \begin{pmatrix} -\frac{3}{4} & \frac{3}{4} \\ \frac{3}{2} & -\frac{9}{2} \end{pmatrix} x + \begin{pmatrix} \frac{5}{2} & 0 \\ 0 & 3 \end{pmatrix} u,$$

$$y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} u.$$

b.

```
a1=0.1; a2=0.2; area=3; Cl=0.4; Cw=0.2;
A=[-a1*area/Cl a1*area/Cl ;
    a1*area/Cw -a1*area/Cw-a2*area/Cw];
B=[1/Cl 0 ; 0 a2*area/Cw];
C=eye(2); D=zeros(2,2);
x0=[0 10]; t=[0:0.1:15]';
u=ones(size(t))*[1 20];
[y,x]=lsim(A,B,C,D,u,t,x0);
plot(t,y(:,1),'--');
hold;
plot(t,y(:,2),'-');
xlabel('time'),
ylabel('temperature of liquid (- -) and wall (---)')
title('Thermal capacity of a wall ')
```

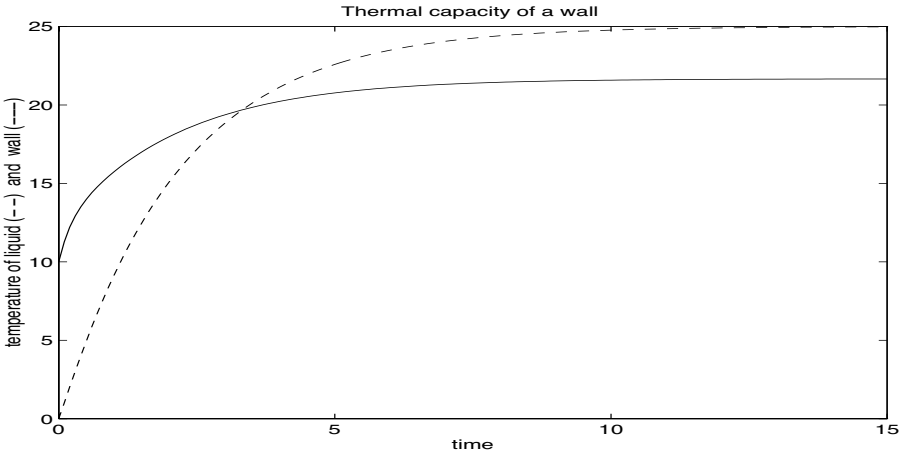

Figure 9.6 Thermal capacity of a wall.

c. The equilibrium can be found by setting $\frac{d\Theta}{dt}$ and $\frac{d\Theta_w}{dt}$ to zero, giving $Q_2 = Q_1 = P = 1$ and $x_{eq} = (\Theta \quad \Theta_w)'$ with

$$\begin{aligned} \Theta_w = \Theta_o \qquad\qquad\quad + P(\alpha_2 A)^{-1} &= 21\tfrac{2}{3} \\ \Theta = \Theta_o + P(\alpha_1 A)^{-1} + P(\alpha_2 A)^{-1} &= 25 \end{aligned}$$

which agrees with the values suggested by the plot in Figure 9.6.

Exercise 9.2.3 (Four moving vehicles)

a. The resulting state-space model reads: $\dot{x} = Ax + Bu$, $y = Cx + Du$ with

$$A = \begin{pmatrix} -k/m_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -k/m_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -k/m_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -k/m_4 \end{pmatrix}, \quad (9.1)$$

$$B = \begin{pmatrix} 1/m_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1/m_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/m_3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/m_4 \end{pmatrix}, \; C = I_7, \; D = O_{74}. \quad (9.2)$$

Notice that h_0 and v_0 have disappeared from the equations.

b.

```
k=8; m1=5; m2=4; m3=3; m4=2;

A=diag([-k/m1 0 -k/m2 0 -k/m3 0 -k/m4])+...
                                  % tridiagonal matrix
  diag([1 0 1 0 1 0],-1)+diag([0 -1 0 -1 0 -1],1);
B=[1/m1 0 0 0 ; 0 0 0 0 ; 0 1/m2 0 0 ; 0 0 0 0 ; ...
   0 0 1/m3 0 ; 0 0 0 0 ; 0 0 0 1/m4];
C=eye(7);
D=zeros(7,4);
x0=[0 ; 1 ; 0 ; 1 ; 0 ; 1 ; 0];
t=[0:0.1:3]';
u=ones(size(t))*[6 12 20 24];
[y,x]=lsim(A,B,C,D,u,t,x0);
plot(t,y(:,2:2:6));               % relative positions
title(['Relative positions for u1=6, '...
                            'u2=12, u3=20, u4=24']);
xlabel('time'), ylabel('relative positions')
```

See plot in Figure 9.7: relative positions get negative: vehicle $i+1$ passes vehicle i.

Exercise 9.2.4 (Pole placement of four-vehicle system)

```
k=8; m1=5; m2=4; m3=3; m4=2;
A=diag([-k/m1 0 -k/m2 0 -k/m3 0 -k/m4])+...
                                  % tridiagonal matrix
```

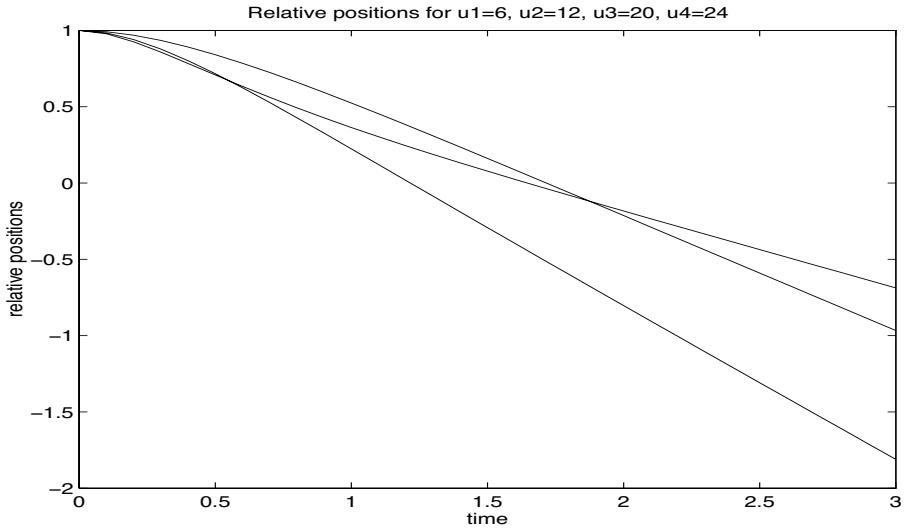

Figure 9.7 Four moving vehicles.

```
  diag([1 0 1 0 1 0],-1)+diag([0 -1 0 -1 0 -1],1);
B=[1/m1 0 0 0 ; 0 0 0 0 ; 0 1/m2 0 0 ; 0 0 0 0 ; ...
   0 0 1/m3 0 ; 0 0 0 0 ; 0 0 0 1/m4];
C=eye(7);
D=zeros(7,4);
x0=[2 ; 1 ; 3 ; 1 ; 4 ; 1 ; 5];

i=sqrt(-1);       % in case i has been used elsewhere
p=[-1 -2 -3 -1+i*sqrt(7) -1-i*sqrt(7)
                            -2+i*sqrt(5) -2-i*sqrt(5)];
F=place(A,B,p);                 % poles to be placed

t=0:0.05:6;
u=zeros(length(t),4);
[y,x]=lsim((A-B*F),B,C,D,u,t,x0);

plot(t,y(:,2:2:6));             % relative positions
title('relative positions of vehicles');
xlabel('time'), ylabel('relative positions')

pause % Press a key to see speeds

plot(t,y(:,1:2:7));             % relative speeds
title('relative speeds of vehicles');
xlabel('time'), ylabel('relative speeds')
```

See Figures 9.8 and 9.9 for plots of relative positions and speeds, respectively.

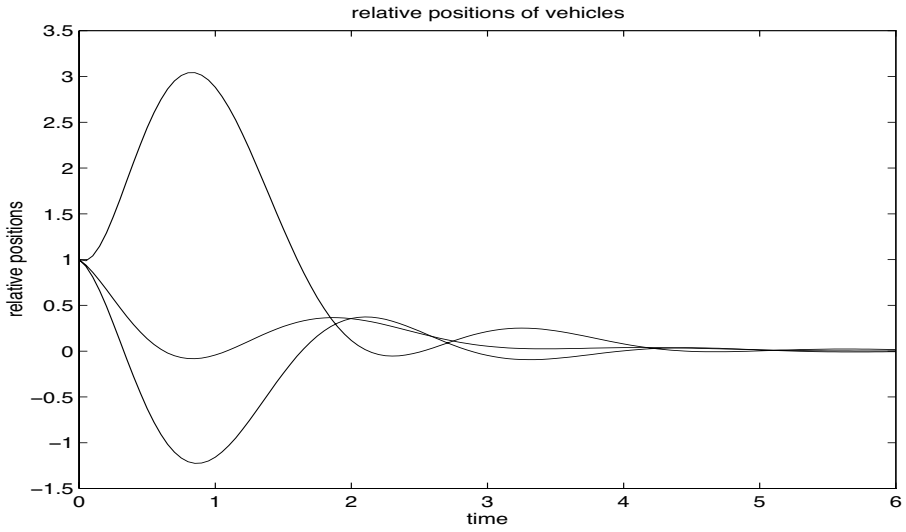

Figure 9.8 Positions in system with feedback, poles at desired locations.

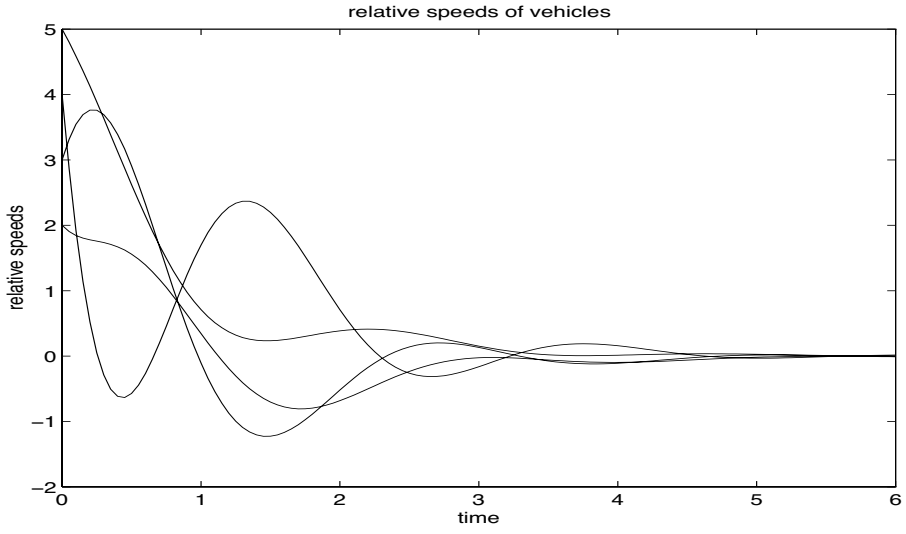

Figure 9.9 Speeds in system with feedback, poles at desired locations.

Exercise 9.2.5 (Observer for four-vehicle system)

a. The system is not observable:

```
k=8; m1=5; m2=4; m3=3; m4=2;
A=diag([-k/m1 0 -k/m2 0 -k/m3 0 -k/m4])+...
                                % tridiagonal matrix
  diag([1 0 1 0 1 0],-1)+diag([0 -1 0 -1 0 -1],1);
C=eye(7);
Cv=C(1:2:7,:);                  % speed measurements
rank(obsv(A,Cv))
ans =
     4
```

The system is also not detectable: for every K the matrix $A-KC$ has three poles equal to zero (hence not in the open left half-plane, as required). This can be seen by computing with MATLAB the decomposition as in (4.17) and (4.18):

```
[Abar,Bbar,Cbar,T,S]=obsvf(A,zeros(7,1),Cv);
                              % the B-matrix is irrelevant
Abar,Cbar
Abar =
 0         0         0         0    1.0000   -1.0000         0
 0         0         0   -1.0000    1.0000         0         0
 0         0         0         0         0   -1.0000    1.0000
 0         0         0   -4.0000         0         0         0
 0         0         0         0   -2.6667         0         0
 0         0         0         0         0   -2.0000         0
 0         0         0         0         0         0   -1.6000
Cbar =
 0        0         0         0         0         0         1
 0        0         0         0         0         1         0
 0        0         0         0         1         0         0
 0        0         0         1         0         0         0
```

b. The system is observable:

```
Cy=C(2:2:6,:);                  % position measurements
rank(obsv(A,Cy))
ans =
     7
```

c.

```
i=sqrt(-1);
p=[-3 -10+sqrt(7)*i -10-sqrt(7)*i -4 -20+sqrt(5)*i ...
                                     -20-sqrt(5)*i -3];
K=place(A',Cy',p)';             % duality
place: ndigits= 17

B=[1/m1 0 0 0 ; 0 0 0 0 ; 0 1/m2 0 0 ; 0 0 0 0 ; ...
   0 0 1/m3 0 ; 0 0 0 0 ; 0 0 0 1/m4];
```

```
D=zeros(7,4);
Dy=D(2:2:6,:);                    % position measurements
e0=[2 ; 1 ; 3 ; 1 ; 4 ; 1 ; 5]; % observation errors
t=0:0.05:2;
u=zeros(length(t),4);
[z,e]=lsim((A-K*Cy),B,Cy,Dy,u,t,e0);% z is not needed

plot(t,e);                        % observation errors
title(['errors in observed positions and speeds '...
                                        'of vehicles']);
xlabel('time'), ylabel('observation errors')

n=zeros(length(t),1);
for k=1:length(t)                 % for every time step
  n(k)=norm(e(k,:));              % calculate 2-norm
end
t(max(find(n>=0.05)))             % last time that norm
ans =                             % of observation error
    1.7500                        % vector is too large
```

See Figure 9.10 for the response of the system.

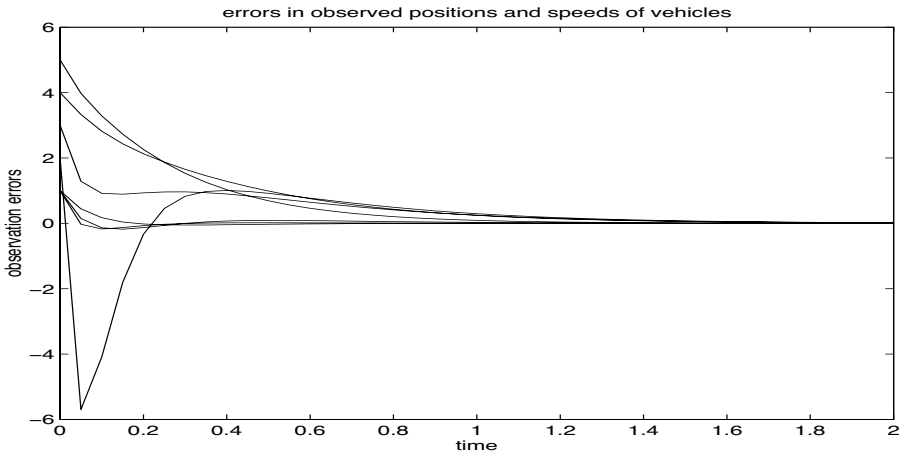

Figure 9.10 Response of four-vehicle system: relative positions observed.

Exercise 9.2.6 (From external description to state space description)

a. The system is equivalent to $(s^3+4s^2+5s+2)Y(s)=(2s^2+6s+5)U(s)$, so

```
num=[0 2 6 5];
```

```
den=[1 4 5 2];
printsys(num,den);
num/den =
        2 s^2 + 6 s + 5
   ---------------------
   s^3 + 4 s^2 + 5 s + 2
```

b.
```
[R,P,K]=residue(num,den)
R =
    1.0000
    1.0000
    1.0000
P =
   -2.0000
   -1.0000


   -1.0000
K =
     []
```

Notice that the pole at -1 has multiplicity 2. So the factorization is such that

$$H(s) = \frac{1}{s+2} + \frac{1}{s+1} + \frac{1}{(s+1)^2}. \tag{9.3}$$

c. Make a state space realization:

```
[A,B,C,D]=tf2ss(num,den);
printsys(A,B,C,D);
a =
                        x1          x2          x3
            x1     -4.00000    -5.00000    -2.00000
            x2      1.00000           0           0
            x3            0     1.00000           0
b =
                        u1
            x1      1.00000
            x2            0
            x3            0
c =
                        x1          x2          x3
            y1      2.00000     6.00000     5.00000
d =
                        u1
            y1            0
```

d. Transfer function of the second system:

```
A2=[0 1 0 ; 0 0 1 ; -2 -5 -4];
B2=[2 ; -2 ; 3];
C2=[1 0 0];
D2=0;
[num2,den2]=ss2tf(A2,B2,C2,D2);
printsys(num2,den2);
num/den =
        2 s^2 + 6 s + 5
   --------------------
   s^3 + 4 s^2 + 5 s + 2
```

This is the same as before. So it is another state space realization of the original system.

From the output equation it follows that $\bar{x}_1 = y$. The first two state equations show that $\bar{x}_2 = \dot{\bar{x}}_1 - 2u = \dot{y} - 2u$ and $\bar{x}_3 = \dot{\bar{x}}_2 + 2u = \ddot{y} - 2\dot{u} + 2u$ respectively. To show that the third state equation is correct we differentiate the last relation and use the differential equation:

$$\begin{aligned}
\dot{\bar{x}}_3 &= \frac{d^3y}{dt^3} - 2\ddot{u} + 2\dot{u} && (9.4)\\
&= (2\ddot{u} + 6\dot{u} + 5u - 4\ddot{y} - 5\dot{y} - 2y) - 2\ddot{u} + 2\dot{u} && (9.5)\\
&= -2y - 5(\dot{y} - 2u) - 4(\ddot{y} - 2\dot{u} + 2u) + 3u && (9.6)\\
&= -2\bar{x}_1 - 5\bar{x}_2 - 4\bar{x}_3 + 3u && (9.7)
\end{aligned}$$

which accounts for the third state equation.

Exercise 9.2.7 (Rocket)

a. Denote the Laplace transforms of φ and r with $\Phi(s)$ and $R(s)$ respectively. The block diagram (replacing all time signals by their Laplace transforms) leads to the following relation between $\Phi(s)$ and $R(s)$:

$$\begin{aligned}
\Phi(s) &= KG(s)\frac{A}{s^2}(R(s) - \Phi(s)) \text{ , so} && (9.8)\\
\Phi(s) &= \frac{KAG(s)/s^2}{1 + KAG(s)/s^2}R(s)\,. && (9.9)
\end{aligned}$$

The transfer function of the controlled system, $H(s) = \frac{\Phi(s)}{R(s)}$ (denoting KA by ρ) can be written as

$$H(s) = \frac{20\rho(s + \frac{1}{2})^2}{s^4 + 10s^3 + 20\rho s^2 + 20\rho s + 5\rho}$$

b. Routh's criterion applies to the denominator $a_4s^4 + a_3s^3 + a_2s^2 + a_1s + a_0$ of $H(s)$, with $a_4 = 1$, $a_3 = 10$, $a_2 = 20\rho$, $a_1 = 20\rho$ and $a_0 = 5\rho$. Referring to the notation of Section 4.1.2 (p. 52) the remaining non-zero coefficients are: $b_1 = 18\rho$, $b_2 = 5\rho$, $c_1 = 20\rho - \frac{25}{9}$ and $d_1 = b_2$. To ensure asymptotic stability of the controlled system the values of a_4, a_3, b_1, c_1 and d_1 must have the same (positive) sign. This results in two conditions on ρ: $18\rho > 0$ and $20\rho - \frac{25}{9} > 0$, which amounts to $\rho > \frac{5}{36}$ or $K > \frac{5}{36A}$.

c. See Figure 9.11 for the plot, resulting from the following commands:

```
% reference values and some further illustrative values
K=[0,2.7,10,20,28.2,31.4,40,60,100,...
   0.4,1,1.8, 4:2:26, 29,30,30.6,31.1,...
   31.3,31.5,31.8,32.5,35,45,50, 70:10:90]';


N=size(K,1);
A=0.05;
den=[ones(N,1)*[1,10],A*K*[20,20,5]];
                              % denomin. polynomials

R=zeros(size(den,2)-1,N);
for k=1:N
  R(:,k)=roots(den(k,:));     % poles
end % k %

plot(R(:,1: 9),'y+'), hold  % reference values of K
plot(R(:,10:N),'w.')        % all other values of K
plot([-max(imag(R))*i,max(imag(R))*i],'r-')
                              % imag. axis
title('root locus for rocket')
xlabel('real part'), ylabel('imag part')
```

In the plot the imaginary axis is crossed for K somewhat greater than 2.7, in agreement with the value from **b**, $\frac{5}{36A} = \frac{25}{9}$, that must be surpassed for stability.

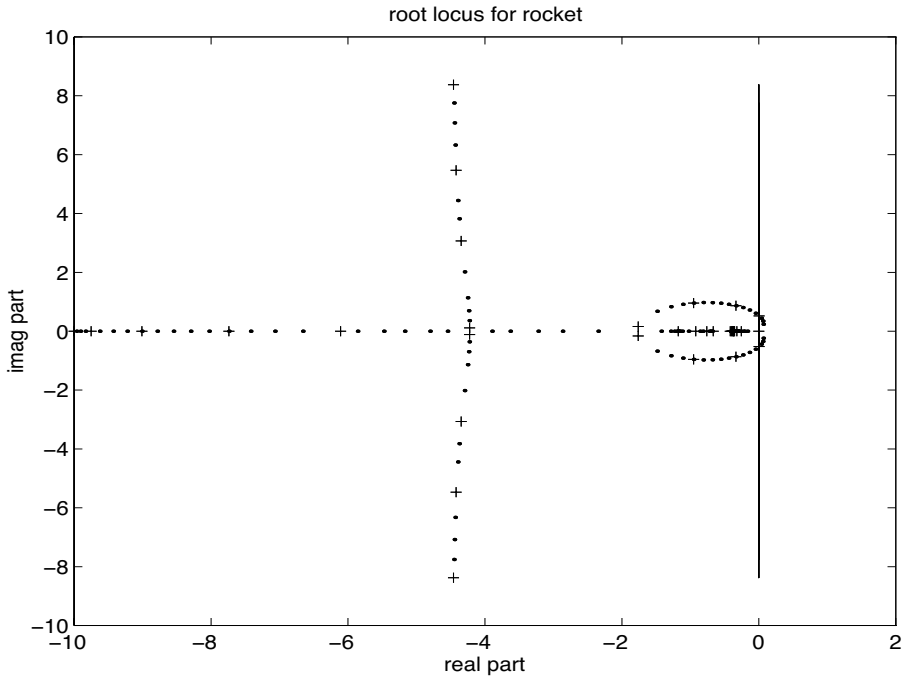

Figure 9.11 Location of closed-loop poles for varying gain K.

Bibliography

[1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.

[2] B.R. Barmish. *New Tools for Robustness of Linear Systems*. McMillan Publishing Company, 1994.

[3] W.E. Boyce and R.C. DiPrima. *Elementary Differential Equations and Boundary Value Problems*. Wiley & Sons, 2001.

[4] C.T. Chen. *Linear system theory and design*. Oxford University Press, 3rd edition, 1998.

[5] R.C. Dorf and R.H. Bishop. *Modern Control Systems*. Prentice Hall, 9th edition, 2001.

[6] P. Faurre and M. Depeyrot. *Elements of system theory*. North-Holland, 1977.

[7] Z. Gajíc and M. Lelíc. *Modern Control Systems Engineering*. Prentice Hall, London, 1996.

[8] G.H. Golub and C.F. van Loan. *Matrix Calculations*, 3rd ed. Johns Hopkins University Press, 1996.

[9] M.I. Kamien and N.L. Schwartz. *Dynamic Optimization* North Holland, 1991.

[10] P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, 1985.

[11] J.M. Maciejowski. *Multivariable feedback deisgn*. Addison-Wesley Publishing Company, 1989. (Recently available through the "Print on demand" program of Pearson Education.)

[12] G. Meinsma. Elementary proof of the Routh-Hurwitz test. *Systems & Control Letters*, 25:237-242, 1995.

[13] R.J. Minnichelli, J.J. Anagnost and C.A. Desoer. An elementary proof of Kharitonov's stability theorem with extensions. *IEEE Transactions on Automatic Control*, 34:995-998, 1989.

[14] H.H. Rosenbrock. *State-space and multivariable theory*. Wiley, New York, 1970.

[15] J.-J.E. Slotine and W. Li. *Applied Nonlinear Control*. Prentice Hall, 1991.

[16] E.D. Sontag. *Mathematical control theory*. Springer Verlag, 1990.

[17] H.L. Trentelman, A.A. Stoorvogel and M.L.J Hautus. *Control Theory for Linear System*. Springer Verlag, 2001.

[18] J.C. Willems. Paradigms and puzzles in the theory of dynamical systems. *IEEE Transactionns on Automatic Control*, 36:259-294, 1991.

[19] W.M. Wonham. *Linear multivariable control: a geometric approach*. Springer Verlag, 1985.

Index