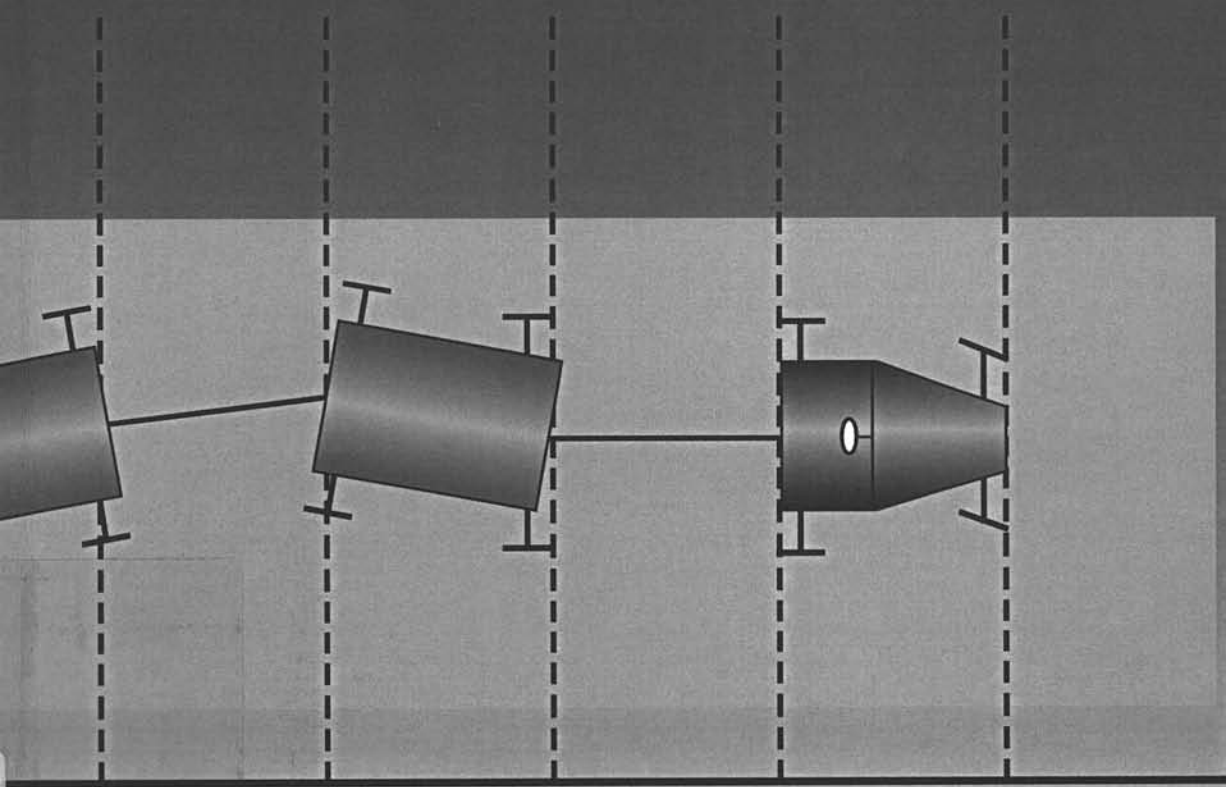
Mathematical Systems Theory

G.J. Olsder
with the collaboration of
J.W. van der Woude



Delft University Press

Mathematical Systems Theory
second edition

Mathematical Systems Theory
second edition

prof.dr. G.J. Olsder

with the collaboration of dr. J.W. van der Woude

Faculty of Information Technology and Systems
Delft University of Technology

DUP Blue Print

First edition 1994
Second edition 1998, 2003

DUP Blue Print is een imprint van:
Delft University Press
P.O. Box 98, 2600 MG Delft, The Netherlands
tel. +31 15 27 85678, telefax +31 15 27 85706, e-mail info@library.tudelft.nl
internet: http://www.library.tudelft.nl/dup

Uitgegeven in opdracht van:
Vereniging voor Studie- en Studentenbelangen te Delft
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 met informatie over dit boek: **http://www.vssd.nl/hlf/a003.htm**

ISBN 90-407-1272-7
NUGI 919
Keywords: mathematical systems theory

Preface

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 (63) than definitions (31) and more examples (56) than 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 Technology. For the preparation of these revised notes, also use was made of 'Cours d'Automatique, Commande Linéaire des Systèmes

Dynamiques' by B. d'Andréa-Novel and M. Cohen dé Lara, Ecole Nationale Supérieure des Mines de Paris. The contents of Chapter 2 have been prepared by dr. 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 ir. P. Twaalfhoven and dr. ir. J.G. Braker.

Delft, April 1994 G.J. Olsder

second edition

The main changes of this second edition over the first one are (i) the addition of a chapter with MATLAB 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

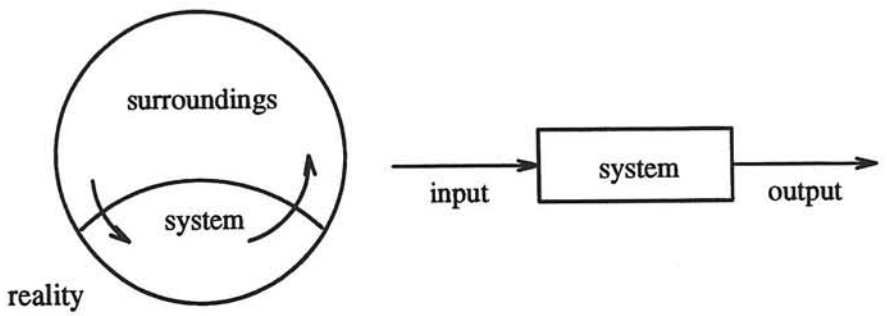
Contents

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 has an influence on the surroundings by means of the output(-functions).



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 behaviour (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.

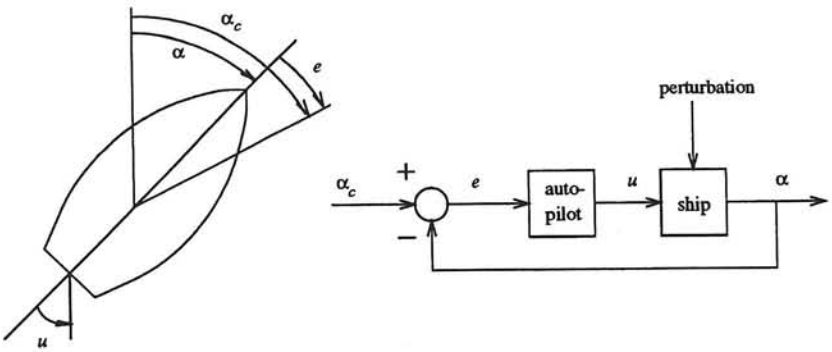
Most modelling uses mathematics. The important features of many physical phenomena can be described numerically and the relations between these features described by equations or inequalities. Particularly in natural sciences and engineering, quantities such as mass, acceleration and force can be described in mathematical terms. To successfully utilize the modelling approach, however, requires a knowledge 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 that model into instructions for a computer, and running the computer, it is possible to model systems larger and more complex 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 behaviour of these phenomena, i.e. how do characteristic features (such as input and output) change in time and what are the relationships, also as functions of time. One tries to design control systems such that a desired behaviour is achieved. In this sense mathematical system(s) theory (and control theory) distinguishes itself from many other branches of mathematics in the sense that it 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 must satisfy a certain requirement as well as possible. In filter theory the interpretation of the input function is observations with measurement errors, the system tries to realize an output which equals the 'ideal' observations, that is, 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, 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 present heading α of a boat (measured by an instrument such as a

magnetic compass or a gyrocompass) and the desired heading α_c (reference point) by the navigator. Using this information, the device automatically yields, as a function of time, the positioning command u of the rudder so as to achieve the smallest possible heading error $e = \alpha_c - \alpha$.



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;

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

Alternatively, it might be proportional;

$$u = Ke,$$

where K is a constant. It is tacitly assumed here that for all e-values of interest, $-u_{\max} \leq Ke \leq u_{\max}$. If this is not the case, some kind of saturation must be introduced. The control law might also consist of a proportional part, an integrating part and a differentiating part;

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

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 integral part and **D** for the differential part. The lower bound of the integral in (1.1) has not been given explicitly; various choices are possible.

Automatic control theory aids 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 its output is the rudder setting u. Thus we see that the output of one system can be the input of a different system. The combination of ship, autopilot and the connection from α to α_c (see the figure) can also be considered as a system; the input is the desired course α_c and the output is the real course α. □

EXERCISE 1.1 *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. Another old example of a control is the steam engine with Watt's centrifugal governor. How does it work? See [Faurre, Depeyrot (1977)].*

EXAMPLE **1.2** [Optimal control problem] The motion of a ship is described by

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

where the state $x = (x_1, x_2)^T \in \mathcal{R}^2$ represents the ship's position with respect to a fixed coordinate system. The vector $u = (u_1, u_2)^T \in \mathcal{R}^2$ represents the control and t is the time. The notation $\dot{x}$ refers to the time derivatives of the two state components. The superscript T refers to 'transposed'; if not explicitly stated differently, vectors are supposed to be column vectors. One control variable to be chosen is the ship's heading u_1; 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, et cetera. Formally, $u = (u_1, u_2)^T$ must be chosen such that

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

is minimized. This criterion describes the fuel used. The function g is the amount of fuel used 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 with the aid of which various alternative NAVSAT concepts and scenarios could be evaluated.

The central idea of satellite based navigation system is the following. A user (such as an airplane or a ship) receives messages from satellites, from which he can estimate his own position. Such a 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 waterwheels and for the control of water levels in Roman aquaducts. Historian Otto Mayr describes the first explicit use of a feedback mechanism as having 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 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]. Kuffler was a temporary of Christian Huygens [1629–1695], who himself 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 flyball governors (see Exercise 1.1) were in use. Soon it was realized that these contraptions gave problems if control was too rigid. Nowadays one realizes that that behaviour was a form of instability due to a high gain in the feedback loop. This problem of bad behaviour was

investigated J.C. Maxwell [1831–1879] – the Maxwell of the electromagnetism – who was the first to study the 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; 'own' results were obtained which were especially related to the 'structure' of the 'box' between input and output. Two developments contributed to that. Firstly there were fundamental theoretical developments in the 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 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

The present chapter, Chapter 1, gives a very superficial overview of what system theory is and discusses the relations with other (mainly: technically oriented) fields. 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.

Chapter 2 discusses modelling techniques and as such it does, strictly speaking, not belong to the area of system theory. Since, however, the starting point in system theory is always 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 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 twenty five 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. In subsequent chapters other views are presented, together with the relationships to the approach of Chapter 6. Chapter 3 deals specifically with linearization and linear differential systems. Chapter 4 deals with structural properties of linear systems. Specifically, various forms of stability and relationships between input, output and the state of the system, such as controllability and observability, are dealt with. Chapter 5 considers feedback issues, both state feedback and output feedback, such as 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 theories of the Laplace transform and of 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 chapter is also suitable 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 decribe 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 in 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 recently introduced 'behavioural 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 are used heavily.

Lastly, Chapter 9 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 nature of this chapter is clearly different from that of the other chapters.

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. For many technical expressions Dutch translations are given in a list before the index.

Chapter 2

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; in this theory one usually starts with a given model, perhaps built by an expert of the field of application concerned.

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 realising) 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 are very much depending 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, but they simply are the result of experiments. There is no reasoning 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 thermodynamic 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 can not be reached*

The second law is often also rephrased as that the entropy of a system can not 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 use, without realizing this, some very important laws and principles. One of these principles, the conservation of energy, is already discussed. Other forms of the conservation principle are also used often. 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 in 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* $F = m\frac{d^2s}{dt^2}$
3. *action = − reaction*

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

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 is applied at a point of a body, and the moment of inertia around that point equals J, then $N = J\frac{d^2\phi}{dt^2}$, where $\frac{d^2\phi}{dt^2}$ denotes the angular acceleration of the body.

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 that 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, completed 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.

$$\text{div } E = \frac{1}{\epsilon}\rho, \text{ rot } E = -\frac{\partial B}{\partial t}, \text{ div } B = 0, \text{ rot } B = \mu(\iota + \epsilon\frac{\partial E}{\partial t}).$$

In these equations all variables depend on the time t, and on the position, in general (x, y, z). Furthermore, E, B and ι are vectorial quantities, whereas ρ is a scalar. The words div and rot stand for the divergence and the rotation, respectively. The first and the 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 magnetic charges).

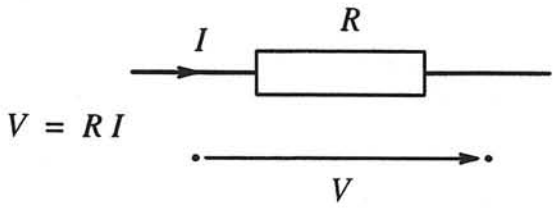
The force F on a particle with a charge q moving with a velocity v in a medium as described above is given by the Lorentz equation

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

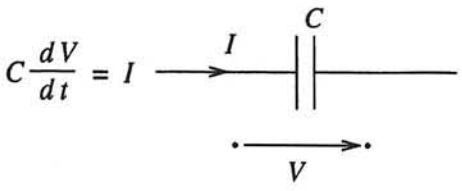
Here $\times$ denotes the cross product. Both F and v are vectors, and q is a scalar. All three will depend on the time t and the position (x, y, z).

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. A number of these laws for electrical networks is discussed below. Amongst others, these networks are built from basic elements such resistors, capacitors and coils. For these elements the following relations have been established.

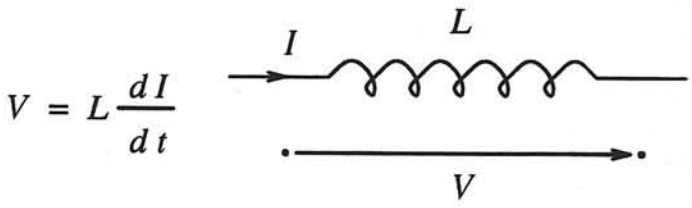
1. If a current of strength I is led through a resistor with value R, then the voltage drop V over the resistor can be computed by Ohm's law as



2. If a current of strength I is sent into a capacitor with capacity C, then the voltage drop V over the capacitor is related to I and C in the following way



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



The variables V and I in the above are functions of time. Often, the values R, C and L are assumed to be time independent.

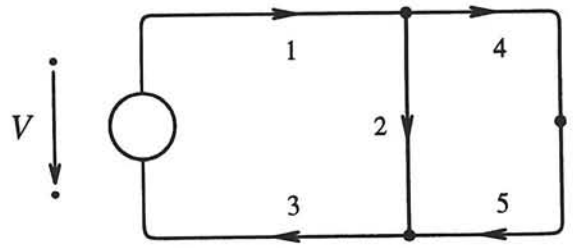
The above laws (rules) are phenomenological in nature. They are the results of experiments. In addition to these laws, two other laws (rules) play an important role in the area of electrical networks. These laws are called the **laws of Kirchhoff**, and can be formulated as follows.

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.*

Note that the Kirchhoff laws are of the conservation type. To explain these two laws we consider the next abstract network with a source over which the voltage

drop is constant V. An arrow in the figure below with 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.



Then there should hold in the four nodes (also the source is considered to be a node) :

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

For the three loops in the network there should hold :

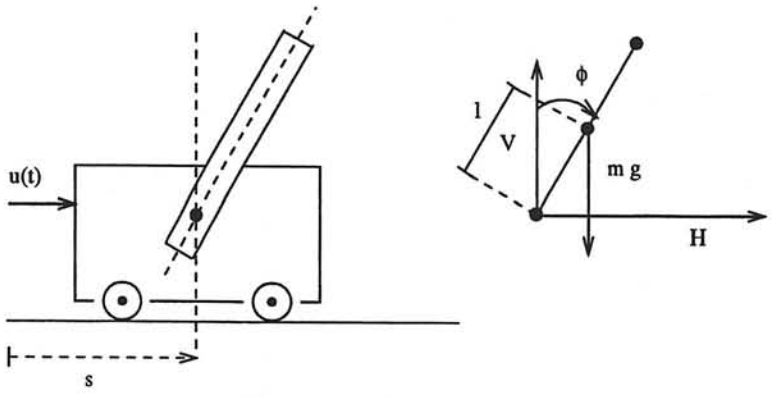
$$V = V_1 + V_2 + V_3, \quad V = V_1 + V_4 + V_5 + V_3, \quad 0 = -V_2 + V_4 + V_5.$$

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 the following figure. The pivot of the pendulum is mounted on a carriage which can move in 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. The distance of the pendulum between pivot and center of gravity is l. In the figure $H(t)$ denotes the horizontal reaction force and $V(t)$ is the vertical reaction force in the pivot. The angle that the pendulum makes with the vertical is indicated by $\phi(t)$. For the center of gravity of the pendulum we have the following equations, that all are in the spirit of Newton's second law.

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

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

The function $s(t)$ denotes the position of the carriage and J is the moment of inertia with respect to the center of gravity. If the pendulum has length $2l$ and a uniform mass distribution of $\frac{m}{2l}$ per unit of length, then the moment of inertia around the 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

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

Elimination of H and V in the above equations leads to

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

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

The above two equations can also be written as a set of four first order differential equations with a vector x defined as $x = (\phi, \dot{\phi}, s, \dot{s})^T$, where T denotes the transpose (see also Exercise 2.1 below). In order to distinguish such differential eqations from partial differential equations, to be introduced shortly, one sometimes calls the differential equations just introduced also 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

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

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

where T, in addition to the kinetic energy of the carriage, consists of the kinetic energy of all the infinitesimal parts of the pendulum $d\sigma$ at a distance σ from the pivot, $0 \leq \sigma \leq 2l$. A similar remark holds with respect to the potential energy. Defining the Lagragian $L = T - V$, it follows after evaluation of the integrals that

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

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

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

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

$$\frac{\partial L}{\partial\dot{\phi}} = ml\dot{s}\cos\phi + \frac{4}{3}ml^2\dot{\phi},$$

and similar for $\frac{\partial L}{\partial\dot{s}}$, $\frac{\partial L}{\partial\phi}$ and $\frac{\partial L}{\partial s}$.

EXERCISE 2.1 *Assume that the angle ϕ of the pendulum with the vertical is measured. Let this measurement be denoted by the variable y. So, $y = \phi$. Note that y, as well as all the other variables $\phi, \dot{\phi}, s, \dot{s}$ and u are functions of time. Consider the vector $x = (\phi, \dot{\phi}, s, \dot{s})^T$, and find functions $f(x, u)$ and $h(x, u)$ such that the inverted pendulum can be described as*

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

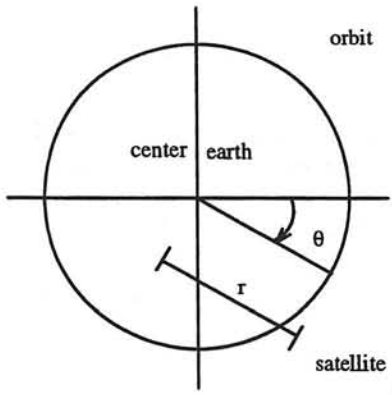
Here $\dot{x} = \frac{d}{dt}x = (\dot{\phi}, \ddot{\phi}, \dot{s}, \ddot{s})^T$.

EXERCISE 2.2 *Take the variable L as in (2.5) and derive the equations of motion of the inverted pendulum by working out the Euler-Lagrange equations.*

EXERCISE 2.3 *In the above example 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 in the above example.*

2.4.2 Satellite dynamics

Consider the motion of a satellite of mass m_s, in a plane through the center of earth. See also the picture below.



As the satellite will orbit around the earth, it is natural to give its position and velocity in terms of polar coordinates r, θ, and their time derivatives $\dot{r}, \dot{\theta}$, with the earth's center located at the origin ($r = 0$).

The velocity of the satellite has a radial[1] component given by $\dot{r}$, and a tangential[2] 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 above 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 to the earth by the gravitational force. This force is radially directed, and its magnitude equals $G\frac{m_e m_s}{r^2}$, 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

[1] 'radial' refers to in the direction of the radius
[2] 'tangential' refers to in the direction of the tangent

to be directed away from the earth. Both F_r and F_θ can be caused by thrust jets that can be mounted on the satellite.

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

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

REMARK **2.1** The above equations also can be obtained from the equations of Euler-Lagrange. Therefore, note that the kinetic energy T and the potential energy V of the satellite is given as follows

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

Now defining the Lagrangian as $L = T - V$, the above equations follow from working out the next equations

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

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

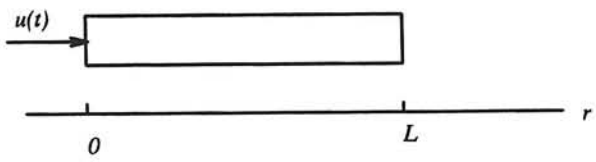
EXERCISE 2.4 *Assume that the distance r is measured and is denoted y. Further, introduce the vectors $x = (r, \theta, \dot{r}, \dot{\theta})^T$ and $u = (\frac{F_r}{m_s}, \frac{F_\theta}{m_s})^T$, and find functions $f(x,u)$ and $h(x,u)$ such that the above model for a satellite can be described as*

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

EXERCISE 2.5 *Starting from the above Lagrangian, work out the equations of Euler-Lagrange to obtain the equations of the motion of the satellite.*

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 $u(t)$.



The temperature of the bar at position r, with $0 \leq r \leq L$, is denoted by $T(t,r)$,

i.e. r is the spatial variable. In order to be able to determine the thermal behaviour of the bar one must know $T(t_0, r)$, $0 \leq r \leq L$, the initial temperature distribution and $u(t)$, $t \geq t_0$. The state of the system is $T(t, \cdot) : [0, L] \to \mathcal{R}$. From physics it is known that T satisfies the 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 we have

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

where A is the surface of the cross section of the bar. At the right hand side of the bar we have

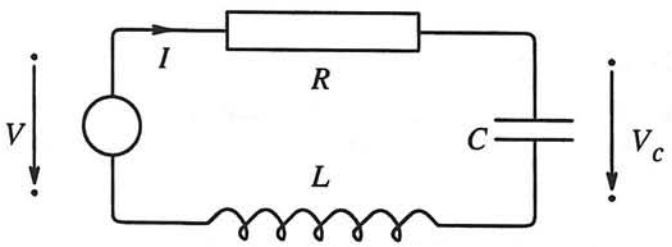
$$\frac{\partial T(t,r)}{\partial r}\,|_{r=L} = 0, \tag{2.9}$$

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. Can you give an interpretation of the partial differential equation

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

2.4.4 Electrical circuit

Consider the following network 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 subsection that

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

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

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

Now rearranging these equations it follows that

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

We define $u = V$, $y = V_C$, and

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

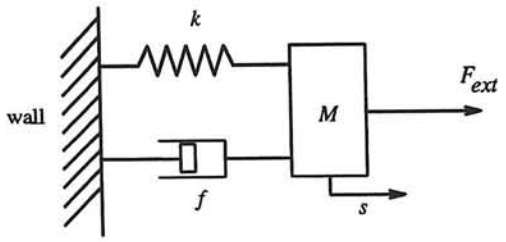
where it must be emphasized that the newly defined C is a matrix (more specifically: here a row vector with two elements). It should not be confused with the capacity C. This is an instance of the same symbol being used for different quantities. With this way of writing, the following description of 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 equations may arise. For instance, when modelling a mechanical structure as depicted below.



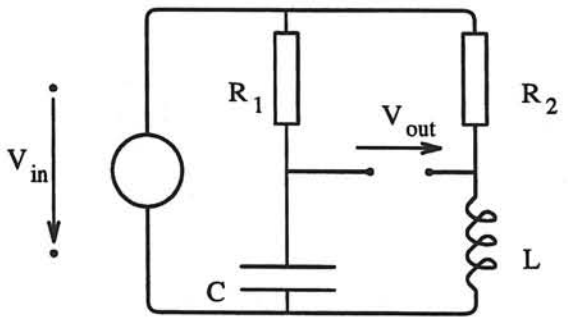
The structure consists of a mass M connected to a vertical wall by means of a spring with constant k and a damper with damping factor f. On the mass an external force F_{ext} may be exerted. As the mass is moving horizontally only, gravity does not play a role. If s denotes the displacement of the mass 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 the modelling of phenomena in disciplines like accoustics, chemistry and hydraulics. □

EXERCISE 2.6 *Consider the following electrical network.*



Take as input V_{in}, and as output V_{out}, and derive a state space model for the network of the above form using the laws introduced in the previous section.

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 varying 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 faster to zero 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,

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

for some constants b and d. This means that

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

Defining $r = b-d$, dividing by δ and taking the limit for δ to zero, it follows that

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

This equation has a 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 then 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 susceptability for deceases. Assuming that the population can not consist of more than $K > 0$ individuals, the above model can be modified as

$$\dot{N}(t) = r(1 - \frac{N(t)}{K})N(t).$$

This 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. It is then reasonable to assume that $r > 0$, and that the previous equation has to be changed into

$$\dot{N}(t) = r(1 - \frac{N(t)}{K})N(t) - \alpha N(t)M(t),$$

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

$$\dot{M}(t) = -cM(t) + \beta N(t)M(t),$$

with $c > 0$ and $\beta > 0$. Together these two equations form a so-called **prey-predator 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 can be unbounded ($K = \infty$). Think of anchovy as prey and of salmon as predator. Assume that due to fishing a fraction $u_1(t)$ of the anchovy is caught, and a fraction $u_2(t)$ of the salmon. The previously derived prey-predator model then has to be changed as follows

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

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

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

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

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

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

EXERCISE 2.7 *For each of the above models 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.*

2.4.6 Age dependent population dynamics

Consider again a population. To express the population size 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 date 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 (i.e. on each finite interval b has at most a finite number of discontinuities, and at points of discontinuity the left and right limits of b exist), then the above integral exists.

Returning to the original integral and assuming that a function g exists such that $P(t,r) = g(t-r)$, 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 interpreted as a time-invariant, strictly causal input/output system. (The notions of time-invariance and (strict) causality will be made precise later (in Sections 3.1 and 3.3). 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 behaviour of a process.) For such a system the probability that somebody is still alive at age r is determined by r only, and not by the date of birth.

EXERCISE 2.8 *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 : $\mu(t,r)drdt$ is the fraction of people in the age class $[r, r+dr]$ who die in the time interval $[t, t+dt]$. Based on the infinitesimal equality*

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

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

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

and the birth rate function as the boundary condition

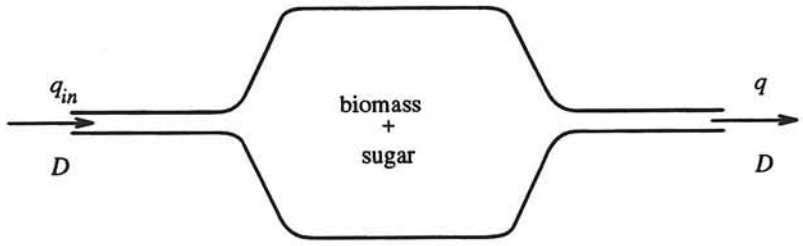
$$p(t,0) = u(t), \quad t \ge 0.$$

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

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

2.4.7 Bioreactor

Consider a bioreactor as depicted in the following.



In the reactor there is biomass (organisms) that is nourished with sugar (nutrition). Further extra nutrition is supplied and products are withdrawn. 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 flow into the reactor (g/l),*

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

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

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

where Dp and Dq stand for the amount the biomass and sugar, respectively, that are withdrawn from the reactor, and Dq_{in} 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 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

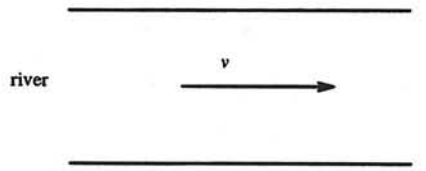
$$\frac{d}{dt}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \mu(q)p - Dp \\ -\nu(q)p - Dq + Dq_{\text{in}} \end{pmatrix}.$$

EXERCISE 2.9 *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 2.10 *The same question as above, but now the sugar concentration q_{in} in the incoming flow is fixed, and the amount of flow D can be controlled.*

2.4.8 Transport of pollution

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



Denote

$\rho(r,t)$ *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

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

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. (If in addition v would be independent of r, a direct resemblance with (2.10) becomes visible.)

2. There is only diffusion. Then ρ and q are related by $q = -\mu \frac{\partial \rho}{\partial r}$, 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 $q = \rho v - \mu \frac{\partial \rho}{\partial r}$. 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

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

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 $A = \mu \frac{\partial^2}{\partial r^2} - v \frac{\partial}{\partial r} - \nu$ 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* $0 \leq m < 1$.
3. *The investment in year* k *depends 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.

$$i(k+1) - \mu c(k+1) = -\mu c(k),$$

$$c(k+1) = my(k) = m(i(k) - \mu c(k)) + m(1+\mu)c(k) + mu(k).$$

If a state vector is defined as $x(k) = (x_1(k), x_2(k))^T$ 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

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

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

EXERCISE 2.11 *Suppose that the government decides to stop its expenditure from the year* $k = 0$ *on. Hence,* $u(k) = 0$ *for all* $k \geq 0$. *Suppose furthermore 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* $k \geq 0$.

EXERCISE 2.12 *For the above model of the economy, 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, when* $u(k) = 1$ *for all* k*.*

Chapter 3

Linear differential systems

3.1 Linearization

In this chapter we will mainly be concerned with linear differential systems, being of the form

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \tag{3.1}$$
$$y(t) = C(t)x(t) + D(t)u(t). \tag{3.2}$$

The treatment of linear difference systems

$$x(k+1) = A(k)x(k) + B(k)u(k), \tag{3.3}$$
$$y(k) = C(k)x(k) + D(k)u(k), \tag{3.4}$$

will be postponed till Chapter 7. There are two reasons for the importance of linear systems. The first one is that they are analytically attractive. These systems can be analyzed much better than nonlinear systems. This is particularly true if the matrices in (3.1), (3.2) are constant with respect to time. In this case the solution, expressed in an initial condition and the input function, can be written down explicitly as we will see later on. The second reason is that many systems are 'almost' linear or can, at least, be approximated by linear systems. Even nonlinear systems may locally be linearized, i.e. in the neighbourhood of a solution small perturbations will behave as solutions of a linear system. It will be assumed that, given an initial condition for (3.1), say $x(0)$, and an input function $u(t)$, $t \geq 0$, that the solution to (3.1) and the function $y(\cdot)$ of (3.2) are well defined. Such initial conditions and input functions are called *admissible*. This is for instance the case if the entries of all matrices concerned and of the input are piecewise continuous. In general we will assume that sets U, $\underline{U}$, Y, $\underline{Y}$, X and $\underline{X}$ exist with $u(t) \in U$ for each t, $u(\cdot) \in \underline{U}$, $y(t) \in Y$ for each t, $y(\cdot) \in \underline{Y}$, $x(t) \in X$ for each t and $x(\cdot) \in \underline{X}$ such that the solution to (3.1) and (3.2) exists for elements of these sets. For simplicity of presentation, these spaces will not

always explicitly be indicated. If the matrices A, B, C and D happen to be constant matrices, i.e. they do not depend on the time t, then we say that the system is **time-invariant.**

We will now make the concept of **linearization** more precise. Consider a nonlinear differential equation

$$\dot{x} = f(x,u)\,, \quad x \in \mathcal{R}^n\,, \quad u \in \mathcal{R}^m. \tag{3.5}$$

(The restriction to a time-invariant system, i.e. time t does not appear explicitly in the function f, is not essential). Given a solution $\tilde{x}(t)$ of (3.5) with initial condition $\tilde{x}(0) = \tilde{x}_0$ and input function $\tilde{u}(t)$, consider another solution $\tilde{x}(t)+z(t)$, with initial condition $\tilde{x}_0 + z_0$ and input function $\tilde{u}(t) + v(t)$, where $\tilde{x} + z$ 'in the neighbourhood of' $\tilde{x}$ and $\tilde{u} + v$ is 'in the neighbourhood of' $\tilde{u}$. This 'in the neighbourhood' will be made more precise later on. We have

$$\dot{\tilde{x}} = f(\tilde{x},\tilde{u})\,, \quad \tilde{x}(0) = \tilde{x}_0, \tag{3.6}$$

$$\frac{d}{dt}(\tilde{x}+z) = f(\tilde{x}+z,\ \tilde{u}+v), \quad \tilde{x}(0)+z(0) = \tilde{x}_0 + z_0\,. \tag{3.7}$$

We assume z and v to be small such that the right-hand side of (3.7) can be expanded into a Taylor series, where the expansion up to the linear terms yields a good approximation:

$$\frac{d}{dt}(\tilde{x}+z) = f(\tilde{x},\tilde{u}) + \frac{\partial f}{\partial x}(\tilde{x},\tilde{u})z + \frac{\partial f}{\partial u}(\tilde{x},\tilde{u})v + \text{ higher order terms.} \tag{3.8}$$

This is a vector equation. Written out in components, the terms are

$$\frac{d}{dt}(\tilde{x}) = \begin{pmatrix} \frac{d\tilde{x}_1}{dt} \\ \vdots \\ \frac{d\tilde{x}_n}{dt} \end{pmatrix}\,, \quad f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}\,, \quad \frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}\,,$$

$$\frac{\partial f}{\partial u} = \begin{pmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{pmatrix}\,, \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}\,, \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}\,.$$

If (3.6) is subtracted from (3.8) and if the higher order terms are ignored, we get

$$\dot{z} = \frac{\partial f}{\partial x}(\tilde{x},\tilde{u})z + \frac{\partial f}{\partial u}(\tilde{x},\tilde{u})v \tag{3.9}$$

for the approximated system. This differential equation is linear, since the coefficients $\frac{\partial f}{\partial x}(\tilde{x},\tilde{u})$ and $\frac{\partial f}{\partial u}(\tilde{x},\tilde{u})$ are *given matrix-functions of time.* Therefore we write for (3.9)

$$\dot{z} = A(t)z + B(t)v \tag{3.10}$$

The output function $y = g(x, u)$ can also be linearized around the pair $(\tilde{x}, \tilde{u})$. If $\tilde{y} = g(\tilde{x}, \tilde{u})$ and $\tilde{y} + w = g(\tilde{x} + z,\ \tilde{u} + v)$, then

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

and therefore, as an approximation,

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

which we write as

$$w(t) = C(t)z(t) + D(t)v(t). \tag{3.11}$$

Equations (3.10), (3.11) together form the linearized system; linearized around the solution $(\tilde{x}(t),\ \tilde{u}(t))$.

EXAMPLE **3.1** Consider the nonlinear systems equations

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = ux_1 + x_2^2 - x_1 - 1, \quad y = x_1^2 + u^3.$$

If $u(t) = \sin t$, show that $x_1 = \sin t$ and $x_2 = \cos t$ is a solution of these equations. Subsequently, linearize the differential equations and the output equation around this solution and write the result in matrix form.

Answer. That the proposed trigoniometric functions are a solution follows by substitution. The linearized equation becomes, after some analysis,

$$\dot{z} = \begin{pmatrix} 0 & 1 \\ -1 + \sin t & 2\cos t \end{pmatrix} z + \begin{pmatrix} 0 \\ \sin t \end{pmatrix} v, \quad w = \begin{pmatrix} 2\sin t & 0 \end{pmatrix} z + (3\sin^2 t)v.$$

This linearized system is not time-invariant because some of the elements of the matrices involved depend on t. □

EXAMPLE **3.2** [Continuation of the inverted pendulum.] We start with the equations of motion (2.4) of Subsection 2.4.1 which are repeated here;

$$\left.\begin{aligned} \tfrac{4l}{3}\ddot{\phi} - g\sin\phi + \ddot{s}\cos\phi = 0, \\ (M+m)\ddot{s} + ml(\ddot{\phi}\cos\phi - \dot{\phi}^2\sin\phi) = u\ . \end{aligned}\right\} \tag{3.12}$$

This system can be written as a set of four first order differential equations where the state vector is defined as $x = (\phi, \dot{\phi}, s, \dot{s})^T$, where T denotes the transpose. We either can linearize that vector differential equation, or we can linearize (2.4) directly and then afterwards we will construct a set of linear differential equations. We will continue with the latter method (the reader should do the first method himself and convince himself that the answers will be the same). Linearization of (2.4) will be shown around the solution

$$\tilde{\phi}(t) = \dot{\tilde{\phi}}(t) = \tilde{s}(t) = \dot{\tilde{s}}(t) = 0,$$

and leads to (i.e. the nonlinear terms in (2.4) are replaced by their Taylor series expansions up to the linear term)

$$\frac{4l}{3}\ddot{\phi} - g\phi + \ddot{s} = 0\,, \quad (M+m)\ddot{s} + ml\ddot{\phi} = u\,, \tag{3.13}$$

which can be viewed as two equations with the two unknowns $\ddot{\phi}$ and $\ddot{s}$. These unknowns are solved and expressed in the other quantities ϕ and u. By defining the state vector as $(\phi, \dot{\phi}, s, \dot{s})^T$, Equations (3.13) can then be rewritten as

$$\frac{dx}{dt} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ a_{41} & 0 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ b_2 \\ 0 \\ b_4 \end{pmatrix} u\,, \quad x = \begin{pmatrix} \phi \\ \dot{\phi} \\ s \\ \dot{s} \end{pmatrix} \tag{3.14}$$

where

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

If we take $M = 0.98$ kg, $m = 0.08$ kg, $l = 0.312$ m and $g = 10$ m/sec^2, then (3.14) becomes

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ -2.4 \\ 0 \\ 1 \end{pmatrix} u \tag{3.15}$$

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

$$y = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} x \tag{3.16}$$

□

EXERCISE 3.1 *This is a continuation of Subsection 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 of $u_2(t)$. 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 (2.6) and take $m_s = 1$ and rewrite Gm_e as g.) Show that, if $u_1(t) = u_2(t) = 0$, $r(t) = \sigma$ (constant), $\theta(t) = \omega t$ (ω is constant) with $\sigma^3\omega^2 = g$ is a

solution and that linearization around this solution leads to (with $x_1 = r(t) - \sigma$; $x_2 = \dot{r}$; $x_3 = \sigma(\theta - \omega t)$; $x_4 = \sigma(\dot{\theta} - \omega)$*)*

$$\frac{dx}{dt} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3\omega^2 & 0 & 0 & 2\omega \\ 0 & 0 & 0 & 1 \\ 0 & -2\omega & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} u \,. \tag{3.17}$$

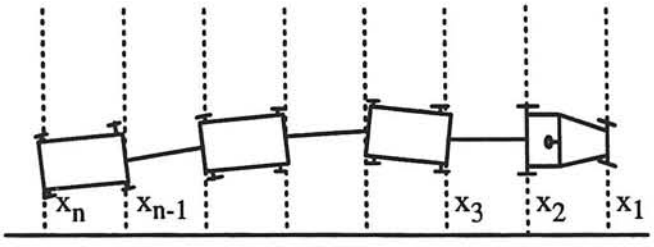
□

EXERCISE 3.2 *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* $u(t) = \cos^2(t)$ *a solution of the differential equations is* $x_1 = \sin t$, $x_2 = \cos t$*. 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 3.3 *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 the figure, 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 frontwheels 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 line* l*. The tractor moves with unit speed forward.*

The (scalar) control is the angle that the frontwheels 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 the distance between two successive middles of axles equals 1. Show that the linearized equations (around $x_i = 0$, $i = 1, \ldots, n$, and $u = 0$) equal

$$\frac{dx}{dt} = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ 0 & \cdots & \cdots & 1 & -1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -1 \end{pmatrix} x + \begin{pmatrix} \epsilon \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{pmatrix} u, \tag{3.18}$$

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

3.2 Solution of linear differential equations

In this section we will consider time-dependent equations

$$\dot{x} = A(t)x + B(t)u \tag{3.19}$$

and time-invariant equations

$$\dot{x} = Ax + Bu \tag{3.20}$$

and consider their solutions. According to the theory of ordinary differential equations we know that the homogeneous differential equation

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

has n independent solutions, to be denoted by $x_1(t), \ldots, x_n(t)$. Note that here x_i is a vector and not a component of x. The matrix Y, the columns of which are x_i;

$$Y(t) = (x_1(t), \ldots, x_n(t))$$

is called a **fundamental matrix.** The matrix $\Phi(t, s) = Y(t)Y^{-1}(s)$ is called the **transition matrix** and it is the unique solution of the matrix differential equation

$$\frac{d}{dt}\Phi(t, s) = A(t)\Phi(t, s) \ , \quad \Phi(s, s) = I \ ,$$

where I is the identity matrix. The i-th column of $\Phi(t, s)$ is the unique solution of $\dot{x} = A(t)x$ with initial condition $x(s) = e_i$, where e_i is the i-th basis vector in $\mathcal{R}^n$.

The solution of $\dot{x}(t) = A(t)x(t)$, $x(t_0) = x_0$, can be expressed as $x(t) = \Phi(t, t_0)x_0$. The transition matrix has the following properties:

$$\begin{array}{lcll} \Phi(t_2, t_0) & = & \Phi(t_2, t_1)\Phi(t_1, t_0) & , \quad \text{for all } t_0, t_1, t_2 \in \mathcal{R} , \\ \Phi^{-1}(t, s) & = & \Phi(s, t) & , \quad \text{for all } s, t \in \mathcal{R} , \end{array}$$

which are called the group properties. The first one expresses the so-called semi-group property. Note that it follows easily from these properties that $\Phi(s, s) = I$. The solution of the inhomogeneous differential equation (3.19) with initial condition $x(t_0) = x_0$ is

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, s)B(s)u(s)ds$$

which can be verified by direct substitution into (3.19).

EXAMPLE **3.3** Consider the time-dependent equation

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{2}{t^2} & \frac{2}{t} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{t^2} \end{pmatrix} u(t) , \quad t > 0.$$

which is equivalent to the following second order differential equation ($x_1(t) = y(t)$, $x_2(t) = \dot{y}(t)$) ;

$$t^2\ddot{y}(t) - 2t\dot{y}(t) + 2y(t) = u(t).$$

First consider the homogeneous equation (i.e. $u(t) = 0$) and substitute a possible solution of the form $y(t) = t^k$, which leads to

$$k^2 - 3k + 2 = 0 \rightarrow k = 1, k = 2 .$$

Therefore $y(t) = t$ and $y(t) = t^2$ are two independent solutions and $(t, 1)^T$ and $(t^2, 2t)^T$ are two independent solutions of

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ -\frac{2}{t^2} & \frac{2}{t} \end{pmatrix} x.$$

The fundamental and transition matrix are respectively

$$Y(t) = \begin{pmatrix} t & t^2 \\ 1 & 2t \end{pmatrix} , \quad \Phi(t, s) = \begin{pmatrix} \frac{2t}{s} - \frac{t^2}{s^2} & -t + \frac{t^2}{s} \\ \frac{2}{s} - \frac{2t}{s^2} & -1 + \frac{2t}{s} \end{pmatrix} .$$

□

We will now confine ourselves to time-invariant equations of the form (3.20). For such systems an explicit expression for the transition matrix exists. Therefore we define

$$e^{At} = I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \cdots \tag{3.21}$$

The quantity e^{At} thus defined (the series is convergent!) is, as A is, an $n \times n$ matrix.

REMARK **3.1** In the following the notation $e^{A(t-s)}$ will be used, which is defined according to the definition in (3.21), i.e. $e^{A(t-s)} = I + A(t-s) + \cdots$. The notation $A(t-s)$ here refers to multiplication of each element of A by the factor $(t-s)$. The notation $e^{(t-s)A}$ is also used. Please note the ambiguity in notation. In (3.19) for instance $A(t)$ stands for a matrix A of which the elements are time-functions. □

THEOREM 3.1 *The matrix $e^{A(t-s)}$ is the transition matrix of $\dot{x} = Ax$.*

Proof The proof is by substitution:

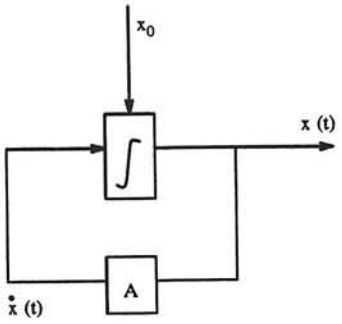
$$\begin{aligned}\frac{d}{dt}e^{A(t-s)} &= \frac{d}{dt}(I + A(t-s) + \frac{1}{2!}A^2(t-s)^2 + \frac{1}{3!}A^3(t-s)^3 + \cdots) \\ &= A + A^2(t-s) + \frac{1}{2!}A^3(t-s)^2 + \cdots \\ &= A(I + A(t-s) + \frac{1}{2!}A^2(t-s)^2 + \cdots) = Ae^{A(t-s)}\,.\end{aligned}$$

Hence $\frac{d}{dt}\Phi(t,s) = A\Phi(t,s)$. Furthermore,

$$\Phi(s,s) = e^{A(s-s)} = I + A.0 + \frac{1}{2!}A^2.0 + \cdots = I\,.$$

□

The solution of $\dot{x} = Ax$ with $x(0) = x_0$ is $x(t) = e^{At}x_0$. This solution can also be obtained by means of the following flow diagram which represents the differential equation:



Going around once in this diagram, starting at x_0, we get

$$x(t) = x_0 + \int_0^t A x_{\sigma_1} d\sigma_1.$$

As x_{σ_1} can be expressed in the same way as $x(t)$ and so on, we get, by going around and around in the diagram,

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

Because e^{At} is a transition matrix, the following properties hold;

1. $e^{A(t_2-t_0)} = e^{A(t_2-t_1)} \cdot e^{A(t_1-t_0)}$, equivalently, $e^{A(t+s)} = e^{At} \cdot e^{As}$;
2. $(e^{At})^{-1} = e^{-At}$.

The exponential e^{At} plays an important role in linear system theory and 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 (3.21). This method works reasonably as long as the eigenvalues of A are close together. For more information and for more reliable methods the reader is referred to [Golub and Van Loan, 1983]. We will continue with an analytical method of calculating e^{At}. The following lemma will be used heavily for that purpose.

LEMMA 3.1 *If P is an invertible matrix, then $e^{At} = Pe^{(P^{-1}AP)}P^{-1}$.*

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

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

□

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

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

In fact, $\{\lambda_i\}$ are the eigenvalues of A and the columns of T are the corresponding eigenvectors. By means of Lemma 3.1 it now follows that

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

The exponential e^{Dt} can easily be obtained by using (3.21). Indeed,

$$e^{Dt} = \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix}.$$

Unfortunately not all square matrices are diagonalizable. Therefore the method described above cannot be used for arbitrary square matrices. Diagonalization is only possible if A has n linearly independent eigenvectors. A sufficient (but not necessary) condition for A to have n linearly independent eigenvectors is that all its eigenvalues are different. A non-diagonalizable matrix of size $n \times n$ has therefore $k(< n)$ different eigenvalues, which are denoted by λ_i, $i = 1, ..., k$. By the (algebraic) **multiplicity** of λ_i is meant the multiplicity of λ_i as a root of the **characteristic polynomial** $\det(\lambda I - A)$; det is determinant. The roots λ_i may be complex.

THEOREM 3.2 *Suppose that the $n \times n$ matrix A has k different eigenvalues λ_i with multiplicity m_i, $i = 1, ..., k$; $\sum_{i=1}^{k} m_i = n$. Define $N_i = \ker[(A - \lambda_i I)^{m_i}]$, then*

1. *the dimension of the linear vector subspace N_i is m_i, $i = 1, ..., k$;*
2. *the n-dimensional linear vector space C^n over the complex numbers is the direct sum of the subspaces N_i, i.e. $C^n = N_1 \oplus N_2 \oplus ... \oplus N_k$.*

For a proof of this theorem and other background material for matrix theory the reader is for instance referred to [Bellman, 1970]. The **kernel** of matrix M, indicated by $\ker[M]$, is defined as all vectors x (here with complex components) for which $Mx = 0$. A linear space N is the **direct sum** of two linear subspaces N_1 and N_2, notation $N = N_1 \oplus N_2$, if each $x \in N$ can uniquely be decomposed as $x = x_1 + x_2$ with $x_1 \in N_1$ and $x_2 \in N_2$. If the $n \times n$ matrix A has n different eigenvalues, then N_i as defined in Theorem 3.2, is a one dimensional subspace, spanned by the eigenvector corresponding to λ_i.

THEOREM 3.3 *For each $n \times n$ matrix A a nonsingular matrix T exists such that*

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

where J, the so-called **Jordan form**, *has a blockdiagonal structure defined as $J = \mathrm{diag}(J_1, ..., J_k)$, equivalently,*

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

Here k is as defined in Theorem 3.2. Each block J_i, $i = 1, ..., k$, corresponding to a unique eigenvalue, also has a block diagonal structure, i.e. $J_i = \mathrm{diag}(J_{i1}, ..., J_{il_i})$, equivalently,

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

where l_i is an integer ≥ 1 and where each sub-block has the following form:

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

EXAMPLE **3.4** If

$$J = \begin{pmatrix} 2 & 1 & & & & & & \\ & 2 & & & & & & \\ & & 2 & 1 & & & & \\ & & & 2 & & & & \\ & & & & -1 & 1 & & \\ & & & & & -1 & 1 & \\ & & & & & & -1 & \\ & & & & & & & -1 \end{pmatrix}, \tag{3.25}$$

with the elements which are not explicitly indicated 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

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

□

If the matrix T of (3.22) is partitioned as $T = [T_1, T_2, ..., T_k]$, conform the partitioning in (3.23), then the column vectors of T_i form a basis for the subspace N_i. Equation (3.22) yields $AT = TJ$. If the individual column vectors of T are denoted by $q_1, ..., q_n$, then the ith column of AT equals Aq_i and the ith column of TJ equals $\lambda q_i + \gamma_i q_{i-1}$, with γ_i either one or zero, depending on the location of the ith row with respect to the Jordan block concerned. Hence

$$Aq_i = \lambda q_i + \gamma_i q_{i-1} \; ; \; i = 1, ..., n, \; (\gamma_1 = 0) \, , \tag{3.26}$$

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. If $\gamma_i = 1$, then q_i is a so-called **generalized eigenvector.**

Now we are in a position to calculate e^{At};

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

Application of the definition of e^{Jt}, see (3.21), gives $e^{Jt} = \text{diag}(e^{J_1 t}, ..., e^{J_k t})$, and for each block, $e^{J_i t} = \text{diag}(e^{J_{i1} t}, ..., e^{J_{il_i} t})$. Finally, for each sub-block,

$$e^{J_{ij}t} = e^{\lambda_i t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{d_{ij}-1}}{(d_{ij}-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{t^2}{2!} \\ & & & \ddots & t \\ 0 & & & & 1 \end{pmatrix} , \tag{3.27}$$

where d_{ij} is the dimension of J_{ij}. See Exercises 3.6 and 3.7 for a proof.

REMARK **3.2** Please note that if $q_{l_{ij}+1}, \ldots, q_{l_{ij}+d_{ij}}$ are the (generalized) eigenvectors belonging to the Jordan block J_{ij} (and this block on its turn corresponds to the eigenvalue λ_i), then $(A - \lambda_i I)^k q_{l_{ij}+k} = 0$, $k = 1, \ldots, d_{ij}$. This

can been proved as follows. For $k = 1$ obviously $(A - \lambda_i I)q_{l_{ij}+1} = 0$ because $q_{l_{ij}+1}$ is an eigenvector (and not a generalized one). For $k = 2$ we can write $(A-\lambda_i I)^2 q_{l_{ij}+2} = (A-\lambda_i I)(A-\lambda_i I)q_{l_{ij}+2} = (A-\lambda_i I)q_{l_{ij}+1} = 0$, where we used (3.26). The proof by induction can be continued for higher values of k. Thus the vectors $q_{l_{ij}+1}, \ldots, q_{l_{ij}+d_{ij}}$ span the linear subspace N_i as introduced in the statement of Theorem 3.2 (or: these vectors span part of this linear subspace if there is more than one Jordan block with the same eigenvalue λ_i). □

EXAMPLE **3.5** This is a continuation of Example 3.2. Calculate the transition matrix for the system given in (3.15). The characteristic polynomial is $\lambda^2(\lambda^2-25)$ and therefore the eigenvalues are $\lambda_{1,2} = 0$, $\lambda_3 = 5, \lambda_4 = -5$. To the eigenvalue 0 only one eigenvector corresponds. So one extra generalized eigenvector for the eigenvalue 0 is needed, which can be computed by (3.26). Therefore

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

The matrices T and T^{-1} can be taken as

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

and

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

□

EXAMPLE **3.6** This is a continuation of Exercise 3.1. Calculate the transition matrix for the system given in (3.17) with $\omega = 1$. The characteristic polynomial is $\lambda^4 + \lambda^2$ and therefore the eigenvalues are $\lambda_{1,2} = 0$, $\lambda_3 = i$, $\lambda_4 = -i$. For the eigenvalue 0 again one eigenvector exists. Again an extra generalized eigenvector for this eigenvalue needs to be computed. Therefore

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

The matrices T and T^{-1} can be taken as

$$T = \begin{pmatrix} 0 & -\frac{2}{3} & 1 & 1 \\ 0 & 0 & i & -i \\ 1 & 0 & 2i & -2i \\ 0 & 1 & -2 & -2 \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} 0 & -2 & 1 & 0 \\ -6 & 0 & 0 & -3 \\ -\frac{3}{2} & -\frac{1}{2}i & 0 & -1 \\ -\frac{3}{2} & \frac{1}{2}i & 0 & -1 \end{pmatrix},$$

and

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

□

EXERCISE 3.4 *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 *If* A_1 *and* A_2 *commute (i.e.* $A_1A_2 = A_2A_1$*), then* $e^{(A_1+A_2)t} = e^{A_1t} \cdot e^{A_2t}$. *Prove this. Give a counterexample to this equality if* A_1 *and* A_2 *do not commute.*

EXERCISE 3.6 *Consider the* $n \times n$ *matrix*

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

So N has zeros everywhere exept for the diagonal directly above the main diagonal,

where it has ones. Using (3.21) prove that

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

EXERCISE 3.7 *Let J_{ij} be as in Theorem 3.3. 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 3.6. Using Exercise 3.5 prove the expression for $e^{J_{ij}}$ in (3.27).*

EXERCISE 3.8 *We are given the n-th order system $\dot{x} = Ax$ with*

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{pmatrix}.$$

Show that the characteristic polynomial of A is

$$\lambda^n + a_{n-1}\lambda^{n-1} + ... + a_1\lambda + a_0.$$

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

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

EXERCISE 3.9 *Show that a Jordan form of the system matrix A of Exercise 3.3 (the tractor example) equals*

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

For diagonalizable A we can write

$$A = TDT^{-1} = (v_1 ... v_n) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}, \qquad (3.28)$$

where $v_1, \ldots, v_n$ are the column vectors of T (formerly we also used the notation q_i for the column vectors), being the eigenvectors of A, and where $w_1, \ldots, w_n$ are the row vectors of T^{-1}. It easily follows that $A = \sum_{i=1}^{n} \lambda_i v_i w_i$. The product of a column vector with a row vector $v_i w_i$ is an $n \times n$ matrix called a **dyad** (a dyad has maximal rank one).

Matrix A is the sum of n dyads. The transition matrix can be written as

$$e^{At} = T \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} T^{-1} = \sum_{i=1}^{n} e^{\lambda_i t} v_i w_i \,. \tag{3.29}$$

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

$$x(t) = e^{At} x_0 = \sum_{i=1}^{n} e^{\lambda_i t} v_i w_i x_0 = \sum_{i=1}^{n} \mu_i e^{\lambda_i t} v_i \,, \tag{3.30}$$

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 (3.28) and (3.29) to the nondiagonalizable 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 (the components of A, σ and ω are real and i denotes the imaginary unit) with corresponding eigenvector $v = r + is$ ($r,\ s \in \mathcal{R}^n$). Because $Av = \lambda v$, $A\bar{v} = \bar{\lambda}\bar{v}$, the upperbar denotes complex conjugate, and therefore $\bar{\lambda} = \sigma - i\omega$ is also an eigenvalue, with eigenvector $r - is$. Suppose that x_0 lies in the subspace spanned by r and s. Then, with $a, b \in \mathcal{R}$,

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

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

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

If μ is written as $\mu = \frac{p}{2i} e^{i\phi}$, with p and ϕ real, then

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

In some applications, the **adjoint system** to $\dot{x} = Ax$, defined as $\dot{\tilde{x}} = -A^T\tilde{x}$, plays a role. It is easily verified, by straightforward differentiation, that $d/dt(x^T\tilde{x}) = 0$ and hence the innerproduct of the vectors $x(t)$ and $\tilde{x}(t)$ does not depend on time. The key property of the adjoint system is that it propagates backward in time as the original system propagates forward in time, as made precise in the following theorem.

THEOREM 3.4 *If $\Phi(t,s)$ is the transition matrix for $\dot{x}(t) = A(t)x(t)$, then $\Phi^T(s,t)$ is the transition matrix for its adjoint system $\dot{\tilde{x}}(t) = -A^T(t)\tilde{x}(t)$.*

Proof Differentiate $I = \Phi^{-1}(t,s)\Phi(t,s)$ to get

$$\begin{aligned} 0 = \frac{d}{dt}I &= \frac{d}{dt}[\Phi^{-1}(t,s)\Phi(t,s)] \\ &= [\frac{d}{dt}\Phi^{-1}(t,s)]\Phi(t,s) + \Phi^{-1}(t,s)\frac{d}{dt}\Phi(t,s) \\ &= [\frac{d}{dt}\Phi^{-1}(t,s) + \Phi^{-1}(t,s)A(t)]\Phi(t,s). \end{aligned}$$

Since $\Phi(t,s)$ is nonsingular, this means

$$\frac{d}{dt}\Phi^{-1}(t,s) = -\Phi^{-1}(t,s)A(t)$$

or

$$\frac{d}{dt}[\Phi^{-1}(t,s)]^T = -A^T[\Phi^{-1}(t,s)]^T,$$

which, because $\Phi^{-1}(t,s) = \Phi(s,t)$, see before, is equivalent to

$$\frac{d}{dt}[\Phi(s,t)]^T = -A^T[\Phi(s,t)]^T.$$

□

3.3 Impulse and step response

The solution of $\dot{x} = A(t)x + B(t)u$ can be described as

$$x(t) = \Phi(t,t_0)x_0 + \int_{t_0}^{t} \Phi(t,s)B(s)u(s)ds\ . \tag{3.31}$$

If an output function of the form

$$y(t) = C(t)x(t) + D(t)u(t)$$

is given, $y(t)$ can be expressed in $u(\cdot)$ as

$$y(t) = C(t)\Phi(t,t_0)x_o + \int_{t_0}^{t} C(t)\Phi(t,s)B(s)u(s)ds + D(t)u(t)\ . \tag{3.32}$$

Define the $p \times m$ matrix $K(t,s)$ by

$$K(t,s) = C(t)\Phi(t,s)B(s)\ . \tag{3.33}$$

Since $\Phi(t,s)$ is continuously differentiable in its arguments and $C(t)$ and $B(s)$ are assumed to be piecewise continuous, the matrix $K(t,s)$ is also piecewise continuous in its arguments.

We assume that a time t_0 exists such that $x_0 = 0$ at this time instant. We are only interested in the system for $t \geq t_0$ and assume $u(s) = 0$ for $s < t_0$. Then (3.32) can be written as

$$y(t) = (Fu)(t) = \int_{-\infty}^{t} K(t,s)u(s)ds + D(t)u(t)\ , \tag{3.34}$$

where F is a mapping which maps an m-dimensional input function $u(\cdot)$, which is supposed to be zero before some time t_0, into a p-dimensional output function $y(\cdot)$. Note that F is a linear mapping and that $K(t,s)$ and $D(t)$ provide a characterization of the **external description** of the system in state space form we started with. We also talk about external behaviour. See Chapter 8 for a discussion on the external description and behaviour. Heuristically speaking, an external description refers to the situation where the input function is directly mapped into an output function, without the 'intermediate' state. This 'intermediate' state has been eliminated. We now assume that $D(t) = 0$, resulting in

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

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 and $\delta(t-t_1)$ is the so-called **delta function**, defined as

$$\int_{-\infty}^{\infty} \delta(s-t_1)\phi(s)ds = \phi(t_1),$$

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

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

The output for such an input is

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

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

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

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

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 $H(t-t_1)e_i$, assuming that the system starts at the origin at a time t_0 far in the past, is given by

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

The $p \times m$ matrix $S(t,t_1) = \int_{t_1}^{t} K(t,s)ds$ is called the **step response matrix.** A relation between $S(t,s)$ and $K(t,s)$ is

$$\frac{d}{ds}S(t,s) = \frac{d}{ds}\int_{s}^{t} K(t,\tau)d\tau = -K(t,s) . \tag{3.36}$$

(The reader is invited to contemplate the $-$ sign at the right-hand side of (3.36).) The Heaviside function and the δ function are related in the same way as the step response matrix and the impuls response matrix. This is not surprising; since the system is linear, it preserves integration. To a δ function as input it is irrelevant in what order the integration and the system are applied.

For time-invariant systems the impulse response matrix becomes

$$K(t,s) = Ce^{A(t-s)}B\,.$$

Since only the difference $t-s$ already defines this matrix, the matrix $K(t,s)$ is usually written as a matrix $G(t)$, depending on only one parameter;

$$G(t-s) = Ce^{A(t-s)}B\,. \tag{3.37}$$

EXERCISE 3.10 *Show that for linear systems with $D(t) \not\equiv 0$, the impulse response can be defined as $K(t,s) + D(t)\delta(t-s)$, where δ is the delta function.*

EXAMPLE **3.7** We start with the linearized equation (3.17) of the satellite dynamics. 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

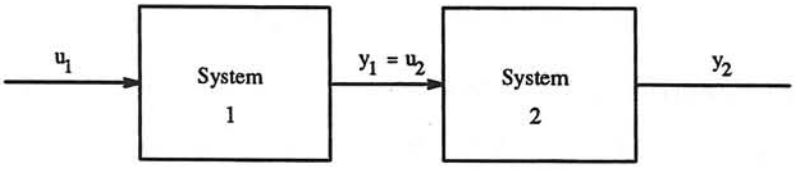
$$y = \begin{pmatrix} r-\sigma \\ \theta-\omega t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma} & 0 \end{pmatrix} x\,. \tag{3.38}$$

Take for the constants $\omega = 1$ and $\sigma = 1$. The impulse response matrix for this system is (for the calculation of e^{At} see Example 3.6):

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

□

EXERCISE 3.11 *Given are two linear differential systems in a series:*



The in- and outputs are scalar functions and the impulse response 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

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

The external description (3.35) does not only hold for (strictly causal (see later for the definition of causality)) linear differential systems as is shown by the following example.

EXAMPLE **3.8** 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

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

This system is not of the form (3.19) as will follow as a direct consequence of the following theorem. If one wants to define a state corresponding to this system, then it will be clear, at least intuitively, that the state $x(t)$ equals the function u on the interval $[t-T, t)$. Hence this system has as its state space a function space, which is infinite dimensional. □

THEOREM 3.5 *A given matrix $K(t,s)$ is realizable as the impulse response matrix of a linear finite dimensional system (i.e. of the form (3.19)) if and only if there exists a decomposition of the form*

$$K(t,s) = H_1(t)H_2(s) \ ,$$

valid for all t and s, with H_1 and H_2 being matrices with finite sizes.

Proof **Sufficiency.** Suppose that the factorization given in the statement of the theorem is possible. Consider the degenerate realization $[A = 0, B = H_2, C = H_1]$, that is, consider the relation

$$\dot{x} = H_2(t)u(t) \ ; \quad y = H_1(t)x \ .$$

Clearly this gives

$$\begin{aligned} y(t) &= H_1(t)x(t_0) + H_1(t) \int_{t_0}^{t} H_2(\sigma)u(\sigma)d\sigma \\ &= H_1(t)x(t_0) + \int_{t_0}^{t} K(t,\sigma)u(\sigma)d\sigma \ . \end{aligned}$$

Necessity. Suppose a linear system of the form (3.19) is given. Then for this system,

$$K(t,s) = C(t)\Phi(t,s)B(s) .$$

However, if t_1 is any constant we have from the composition law for transition matrices

$$\Phi(t,s) = \Phi(t,t_1)\Phi(t_1,s) .$$

If we make the identification

$$H_1(t) = C(t)\Phi(t,t_1) , \quad H_2(s) = \Phi(t_1,s)B(s) ,$$

we see that $H_1(t)H_2(s) = K(t,s)$. □

EXERCISE 3.12 *Show that the impulse response fuction of the moving average example (Example 3.8) cannot be written as a product of the form $H_1(t)H_2(s)$ and hence the moving average system cannot be represented as a finite-dimensional linear system.*

If we write

$$y(t) = \int_{-\infty}^{+\infty} K(t,s)u(s)ds \text{ , or, } y(t) = \int_{-\infty}^{+\infty} G(t-s)u(s)ds , \qquad (3.39)$$

with the upperbound $+\infty$ instead of t, we get in principle a noncausal system.

REMARK **3.3** The formal definition of **strict causality** will be given in Chapter 8. Heuristically it means that the present 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. □

These relations do not define a system according to the definitions given here. The causal systems form a subclass of the class of systems described by (3.39) 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 behaviour of a linear differential system is completely determined by the matrices $K(t,s)$ and $D(t)$. It is possible that different sets of matrices $(A(t), B(t), C(t))$ define the same matrix $K(t,s)$. Let us substantiate this for time-invariant systems only:

$$\dot{x} = Ax + Bu , \quad y = Cx + Du , \quad G(t) = Ce^{At}B . \qquad (3.40)$$

If $S : \mathcal{R}^n \to \mathcal{R}^n$ is an invertible basis transformation in the state space $X = \mathcal{R}^n$, then we get for the transformed state $z = Sx$ the following equations:

$$\dot{z} = S\dot{x} = SAx + SBu = SAS^{-1}z + SBu,$$

$$y = Cx + Du = CS^{-1}z + Du\,.$$

The basis transformation S transforms the set of matrices (A, B, C, D) into $(SAS^{-1}, SB, CS^{-1}, D)$. Calculation of the impulse response matrix for the transformed system yields

$$CS^{-1}e^{SAS^{-1}t}SB = CS^{-1}Se^{At}S^{-1}SB = Ce^{At}B,$$

which shows that $G(t)$ does not change under a basis transformation. This should be clear since the choice of a new basis in the state space should not change the external behaviour of a system.

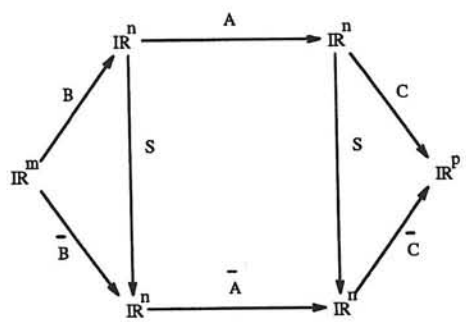
DEFINITION 3.1 *Two linear systems*

$$\begin{array}{rclcrcl} \dot{x} & = & Ax + Bu, & & \dot{\tilde{x}} & = & \tilde{A}\tilde{x} + \tilde{B}u, \\ y & = & Cx + Du, & & y & = & \tilde{C}\tilde{x} + \tilde{D}u, \\ x & \in & \mathcal{R}^n, & & \tilde{x} & \in & \mathcal{R}^n, \end{array}$$

with the same number of inputs and with the same number of outputs are called isomorphic if an invertible linear mapping $S : \mathcal{R}^n \to \mathcal{R}^n$ exists such that

$$\tilde{A} = SAS^{-1}\,, \quad \tilde{B} = SB\,, \quad \tilde{C} = CS^{-1}\,, \quad \tilde{D} = D\,.$$

The relationship between the two systems in the definition above can be given as a commutative diagram:



Just before Definition 3.1 we have proved that two isomorphic systems have the same impulse response matrix function. It may be clear that for a given impulse response function there exist realizations (A, B, C, D) having state vectors of

different dimensions. A trivial example of this is obtained from the system given in (3.40) by adding a vector equation which does not affect the output; e.g.

$$\begin{aligned} \dot{x} &= Ax + Bu \\ \dot{\tilde{x}} &= F\tilde{x} + Gu \end{aligned}, \quad y = Cx + Du\,,$$

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 upperbound on the dimension of a realization of a given impulse response function. However, under reasonable conditions there does exist a lower bound. If a system, given by its matrices (A, B, C, D), realizes the impulse response function $K(t, s)$, it will be called a **minimal realization** if there exists no other realization of $K(t, s)$ having a lower dimensional state vector. The minimum dimension is called the order of the impulse response function.

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 F introduced in (3.34), determine a state space description. For linear time-invariant finite-dimensional differential systems this problem boils down to: given the impulse response matrix function $G(t) + D\delta(t)$ (see Exercise 3.10 for the addition of $D\delta(t)$), find an $n \times n$ matrix A, $n \times m$ matrix B and a $p \times n$ matrix C such that $G(t) = Ce^{At}B$; n is also to be determined. A conclusion is that, even with minimal n, not a unique realization exists; for, if (A, B, C, D) satisfies the conditions, then $(SAS^{-1}, SB, CS^{-1}, D)$, with the $n \times n$ matrix S nonsingular, also satisfies the same conditions.

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 first four following subsections deal with the first mentioned concept of stability, the fifth one deals with input/output stability. Proofs of the theorems related to Routh stability and interval stability are not given here; they make an extensive use of complex function theory and, though not difficult, are lengthy and fall outside the scope of these notes.

4.1.1 Stability in terms of eigenvalues

DEFINITION 4.1 *Given a first order differential equation $\dot{x} = f(x)$, the solution of which, with initial condition $x(0) = x_0$, will be indicated by $x(t, x_0)$. A vector $\bar{x}$ which satisfies $f(\bar{x}) = 0$ is called an* **equilibrium point**. *An equilibrium point $\bar{x}$ is called* **stable** *if for every $\varepsilon > 0$ a $\delta > 0$ exists such that, if $\|x_0 - \bar{x}\| < \delta$, then $\|x(t, x_0) - \bar{x}\| < \varepsilon$ 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 $\lim_{t \to \infty} \|x(t, x_0) - \bar{x}\| = 0$ provided that $\|x_0 - \bar{x}\| < \delta_1$. An equilibrium point x is* **unstable** *if it is not stable.*

In this definition $\| \ \|$ is an arbitrary norm; usually the Euclidean norm is used. Intuitively, stability means that the solution remains in a neighbourhood of the equilibrium point, asymptotic stability means that the solution converges to the equilibrium point (provided the initial point is sufficiently close to this equilibrium point), instability means that there is always a solution, which may start arbitrarily close to an equilibrium point, that 'explodes' or 'diverges away' from the equilibrium point.

For linear differential equations $\dot{x} = Ax$ we will take as equilibrium point $\bar{x} = 0$ (though there will be others if $\det(A) = 0$). We will call $\dot{x} = Ax$, or even A, (asymptotically) stable if $\bar{x} = 0$ is (asymptotically) stable.

THEOREM 4.1 *Given the differential equation $\dot{x} = Ax$, with A an $n \times n$ matrix with the different eigenvalues $\lambda_1, ..., \lambda_k$ $(k \leq n)$. The origin $x = 0$ is asymptotically stable if and only if $\Re\, \lambda_i < 0$ for $i = 1, ..., k$. The origin is stable if $\Re\, \lambda_i \leq 0$ for $i = 1, ..., k$ and moreover if to each eigenvalue with $\Re\, \lambda_i = 0$ there correspond as many linearly independent eigenvectors as the multiplicity of λ_i.*

Proof In the proof use is made of the formula

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

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

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

the factor $e^{\lambda_i t}$ remains bounded (but does not approach zero (with $\Re(\lambda_i) = 0$ it follows that $|\exp \lambda_i t| = 1$)), whereas the elements in the matrix do not all remain bounded. Therefore, if the size of J_{ij} is greater than 1×1, we do not have stability. If the size of all sub-blocks J_{ij} corresponding to eigenvalues with real part zero is 1×1, then stability is guaranteed. The condition given in the statement of the theorem exactly expresses the fact that all such sub-blocks have size 1×1. □

EXAMPLE **4.1** Consider the matrices in Exercise 3.4. The first one is stable, the fifth one is asymptotically stable, the others are unstable. □

EXERCISE 4.1 *Investigate the (asymptotic) stability of the system matrices A corresponding to the inverted pendulum in Example 3.2 and to the satellite in Exercise 3.1.*

EXAMPLE **4.2** The results of Theorem 4.1 do not hold for time-varying systems as shown by the solution of

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

The eigenvalues of the system matrix are $\lambda_1 = a$, $\lambda_2 = 3a$ (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 is (with initial condition $x_1(0) = x_{10}$, $x_2(0) = x_{20}$):

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

which is unstable for any nonzero real a. □

DEFINITION 4.2 *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 N_i (see Theorem 3.2) which corespond to eigenvalues of A in the left half plane (i.e. eigenvalues with real parts less than zero). The* **unstable subspace** *is defined similarly; it corresponds to eigenvalues with nonnegative real parts.*

Remark that in this definition $\Re\lambda_i = 0$ is supposed to correspond to the unstable subspace, though strictly speaking we could have distinguished 'stable subspaces' and 'asymptotically stable subspaces'. It follows from this definition that the state space $\mathcal{R}^n$ is the direct sum of the stable and unstable linear subspace.

EXERCISE 4.2 *Show that of the scalar nonlinear system $\dot{x} = -\varepsilon x + x^2$ the equilibrium point $\bar{x} = 0$ is asymptotically stable for each $\varepsilon > 0$ and unstable for $\varepsilon \leq 0$. The linearized system (linearized around the equilibrium point), however, is stable for $\varepsilon = 0$. How is this explained?*

Suppose that the differential equation $\dot{x} = f(x,t)$ has an equilibrium point $\bar{x} = 0$, i.e. $f(\bar{x},t) = 0$ for all t. Suppose also that this differential equation can equivalently be written as

$$\dot{x} = Ax + h(x,t) , \tag{4.2}$$

where Ax is the linear part of $f(x,t)$ around the origin. Note that the matrix A is assumed to be constant. By construction, $\|h(x,t)\| = o(\|x\|)$, which means

$$\lim_{\|x\|\to 0} \frac{\|h(x,t)\|}{\|x\|} = 0 .$$

THEOREM 4.2 *Let the matrix A in (4.2) be constant (and real) with characteristic roots all having negative real parts. If h in (4.2) is real, continuous for small $\|x\|$ and $t \geq 0$ and moreover $\|h(x,t)\| = o(\|x\|)$ uniformly in t, $t \geq 0$, then the identically zero solution of (4.2) is asymptotically stable.*

The proof of this theorem can for instance be found in [Coddington and Levinson, 1955].

4.1.2 Routh's criterion

The eigenvalues of A are the roots of the characteristic polynomial $det(\lambda I - A) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + ... + a_1\lambda + a_0$ with $a_n = 1$. By means of the so-called **Routh's criterion** the stability of A can be checked directly by considering the coefficients $\{a_i\}$, without calculating the roots of the polynomial explicitly. In terms of number of numerical operations, calculation of the position of the eigenvalues is much more expensive than the Routh criterion, which only checks whether the eigenvalues lie in the left-half plane (and does *not* calculate the positions of the eigenvalues). The criterion works as follows (no proof is given here: it can be found in [Gantmacher, 1959]). Arrange the coefficients $\{a_i\}$ in the following way;

$$\begin{array}{cccc} a_n & a_{n-2} & a_{n-4} & \cdots \\ a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ b_1 & b_2 & b_3 & \cdots \\ c_1 & c_2 & c_3 & \cdots \\ \vdots & \vdots & \vdots & \end{array}$$

where the coefficients $\{b_i\}, \{c_i\}, ...$ are defined as

$$b_1 = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}} \quad , \quad b_2 = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}} \, , ...,$$

$$c_1 = \frac{b_1 a_{n-3} - a_{n-1}b_2}{b_1} \quad , \quad c_2 = \frac{b_1 a_{n-5} - a_{n-1}b_3}{b_1} \, , ...,$$

etcetera. The scheme is continued until only zeroes appear (both to the right and downwards). Thus one has obtained the so-called Routh table. In order to exclude singular cases, it is assumed that the first $n+1$ elements of the first column of the Routh table are well defined and nonzero. Subject to this assumption, Routh's criterion states that the matrix A has eigenvalues with negative real part if and only if the elements in the first column all have the same sign. It is not difficult to find polynomials for which the assumption (on excluding singular cases) is not satisfied; in that case an adapted version of Routh's criterion exists, for which the reader is referred to [Gantmacher, 1959].

EXERCISE 4.3 *For which value(s) of k has the equation $\lambda^3 + 3\lambda^2 + 3\lambda + k = 0$ only roots with negative real parts?*

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 ∞ 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 converge to 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 origin in classical mechanics where stability criteria involving the notion of *energy* are quite useful. A mechanical system was defined to be stable if its energy remained bounded. Lyapunov developed this idea and subsequently 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 $V(x(t))$, not depending on t explicitly now, and defined as $x^T(t)Px(t)$ for some positive-definite[1] matrix P will be regarded as a 'generalized' energy associated with the system. In an asymptotically stable system the energy should decay with time;

$$\frac{d}{dt}V(x(t)) = \dot{x}^T(t)Px(t) + x^T P\dot{x}(t) = x^T(t)[PA + A^T P]x(t),$$

and hence, if $Q \stackrel{\text{def}}{=} -(PA + A^T P)$ is positive-definite, the energy does decrease with time. Indeed, if $Q > 0$, then $\lim_{t\to\infty} V(x(t)) = 0$.

THEOREM 4.3 *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*

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

Proof **Sufficiency:** from the existence of the P matrix we will prove that all eigenvalues of A have negative real parts. Suppose that a matrix $P > 0$ exists such that (4.3) is true, then the argumentation just above has shown that $\lim_{t\to\infty} V(x(t)) = 0$, which can only be true if $\lim_{t\to\infty} x(t) = 0$. This latter can only be true if all eigenvalues of A have negative real part.
Necessity: from the asymptotic stability of A it will be shown that (4.3) has a solution $P > 0$. If A is such that $\Re\lambda_i < 0$ for all eigenvalues λ_i, then we shall prove that a suitable P-matrix is given by

$$P = \int_0^\infty e^{A^T t} Q e^{At} dt.$$

Due to the asymptotic stability of A, this integral will exist. Next, by substitution,

$$A^T P + PA = \int_0^\infty A^T e^{A^T t} Q e^{At} dt + \int_0^\infty e^{A^T t} Q e^{At} A dt$$

[1] A square matrix P is called positive-definite if it is symmetric and if $x^T Px > 0$ for all $x \neq 0$.

$$= \int_0^\infty \frac{d}{dt}(e^{A^T t} Q e^{At}) dt$$
$$= e^{A^T t} Q e^{At} \,|_0^\infty = -Q.$$

□

Equation (4.3) is referred to as the **Lyapunov equation.**

4.1.4 Interval stability

In this section polynomials of the form $p(\lambda) = \sum_{i=0}^n a_i \lambda^i$ will be studied; we are specifically interested in whether their zeros belong to the left-half plane. The novelty here is that the coefficients a_i are not exactly known; it will be assumed that only the lowerbounds a_i^- and upperbounds a_i^+ are known: $a_i \in [a_i^-, a_i^+]$. The central question is: if we know a_i^- and a_i^+, $i = 0, 1, \ldots, n$, and arbitrary a_i subject to $a_i \in [a_i^-, a_i^+]$ are chosen, what can one say about the location of the zeros of $p(\lambda)$? What conditions should be imposed on a_i^- and a_i^+ such that the zeros lie in the 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, $i = 0, 1, \ldots, n$, will not be known; one only knows these values approximately by means of lower- and upperbounds. Sometimes such an 'uncertain' polynomial is called an **interval polynomial** and is written as $p(\lambda, a) = \sum_{i=0}^n [a_i^-, a_i^+] \lambda^i$. Throughout this section it will be assumed that $0 \notin [a_n^-, a_n^+]$ which amounts to saying that the degree of $p(\lambda, a)$ is constant (and equals n), whatever the values of a_i. In this case one says that the family of polynomials $p(\lambda, a)$ has **invariant degree.**

DEFINITION 4.3 *Associated with the interval polynomial*

$$p(\lambda, a) = \sum_{i=0}^n [a_i^-, a_i^+] \lambda^i$$

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

$$\begin{aligned}
p_1(\lambda) &= a_0^- + a_1^- \lambda + a_2^+ \lambda^2 + a_3^+ \lambda^3 + a_4^- \lambda^4 + a_5^- \lambda^5 + a_6^+ \lambda^6 + \cdots, \\
p_2(\lambda) &= a_0^+ + a_1^+ \lambda + a_2^- \lambda^2 + a_3^- \lambda^3 + a_4^+ \lambda^4 + a_5^+ \lambda^5 + a_6^- \lambda^6 + \cdots, \\
p_3(\lambda) &= a_0^+ + a_1^- \lambda + a_2^- \lambda^2 + a_3^+ \lambda^3 + a_4^+ \lambda^4 + a_5^- \lambda^5 + a_6^- \lambda^6 + \cdots, \\
p_4(\lambda) &= a_0^- + a_1^+ \lambda + a_2^+ \lambda^2 + a_3^- \lambda^3 + a_4^- \lambda^4 + a_5^+ \lambda^5 + a_6^+ \lambda^6 + \cdots
\end{aligned}$$

It will turn out that these four Kharitonov polynomials play a crucial role in the stability of $p(\lambda, a)$, with the vector a arbitrary, but subject to $a_i \in [a_i^-, a_i^+]$, as is expressed by

THEOREM 4.4 *Any polynomial $p(\lambda, a)$ with $a_i \in [a_i^-, a_i^+]$ and $0 \notin [a_n^-, a_n^+]$ has all its zeros in the left-half plane if and only if the four Kharitonov polynomials have all their zeros in the left-half plane.*

EXAMPLE **4.3** Suppose we are given the interval polynomial

$$p(\lambda, a) = [15, 19] + [20, 24]\lambda + [2, 3]\lambda^2 + [1, 2]\lambda^3,$$

then the four Kharitonov polynomials are:

$$\begin{aligned} p_1(\lambda) &= 15 + 20\lambda + 3\lambda^2 + 2\lambda^3, \\ p_2(\lambda) &= 19 + 24\lambda + 2\lambda^2 + \lambda^3, \\ p_3(\lambda) &= 19 + 20\lambda + 2\lambda^2 + 2\lambda^3, \\ p_4(\lambda) &= 15 + 24\lambda + 3\lambda^2 + \lambda^3. \end{aligned}$$

To study the stability of these four polynomials, one can for instance use the Routh criterion. If one does so, it turns out that these four polynomials are indeed stable, i.e. they have all their zeros in the left-half plane. From Theorem 4.4 it follows that $p(\lambda, a)$ has all its zeros in the left-half plane for all vectors a with $a_i \in [a_i^-, a_i^+]$, $i = 0, 1, 2, 3$. □

EXERCISE 4.4 *Show, by means of an interval polynomial of degree one, that the condition* $0 \notin [a_n^-, a_n^+]$ *is indeed necessary for Theorem 4.4 to hold.*

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 $u(\cdot)$ is called bounded if a constant c exists such that $\|u(t)\| \leq c$ for all t. One has a similar definition for the boundedness of the output. Let us give the formal definition.

DEFINITION 4.4 *The system*

$$\begin{pmatrix} \dot{x}(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} A(t) & B(t) \\ C(t) & D(t) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix}.$$

is **BIBO stable** *(BIBO stands for bounded input, bounded output) if for all* t_0 *and for 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*, independent of* t_0 *such that for all* t_0 *the statements*

$$x(t_0) = 0, \quad \|u(t)\| \leq 1, \ \forall t \geq t_0$$

imply $\|y(t)\| \leq k$*, for all* $t \geq t_0$*.*

Other types of input-output stability definitions 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 this direction.

4.2 Controllability

Controllability is a fundamental concept in mathematical system theory, as is the concept of observability, the latter to be introduced in the next section. These concepts play an essential role in the design and control of systems as will become clear. We will confine ourselves to lineair, time-invariant differential systems as introduced in previously. Consider therefore

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

with $x \in \mathcal{R}^n$, $u \in \mathcal{R}^m$ and $y \in \mathcal{R}^p$. The constant matrices have appropriate sizes. The space of admissible input functions will be the class of piecewise continuous (vector) functions. The solution of (4.4) with initial condition $x(0) = x_0$ will be written as $x(t, x_0, u)$ and the corresponding output as $y(t, x_0, u)$.

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

System (4.4) will sometimes be indicated by system (A, B, C, D) for the sake of brevity.

DEFINITION 4.5 *System (A, B, C, D) is called* **controllable** *if for each arbitrary two points x_0 and $x_1 \in \mathcal{R}^n$, a finite time $t_1 > 0$ and an admissible input function $u(\cdot)$ exist such that $x(t_1, x_0, u) = x_1$.*

A system is controllable if an arbitrary point x_1 can be reached starting from an arbitrary point x_0 in finite time by means of an admissible input function. Sometimes controllability is only defined with respect to final points x_1 which are the origin. In that case it would be more appropriate to talk about **null controllability**. We will stick to Definition 4.5. The 'reverse' concept of null controllability, i.e. being able to reach an arbitrary point starting from the origin, is called **reachability**. For differential systems (A, B, C, D) the two additional concepts are equivalent to Definition 4.5 as can be proven. If such a system is reachable, then it is also controllable and null-controllable, etc. This 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 and therefore null-controllability is easier fulfilled than 'full' controllability.) Controllability will be characterized in terms of the matrices A and B. It is clear that C and D do not play any role here. Define

$$R = [B|AB|A^2B|\ldots|A^{n-1}B]$$

which is an $n \times nm$ matrix consisting of n blocks A^jB, $j = 0, 1, \ldots, n-1$, and is called the **controllability matrix**. The image of R, Im R, is the linear subspace[2]

[2]Note the difference in notation between Im, 'image' and $\Im$, 'imaginary part'

of $\mathcal{R}^n$ spanned by the column vectors of R and is called the **controllability subspace.** This name will become clear later on.

LEMMA 4.1 *Im $A^kB \subset$ Im R for all k.*

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

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

is the characteristic polynomial of A, then the **theorem of Cayley-Hamilton,** which is well known in matrix theory, states that $p(A) = 0$. This results in

$$A^n = -p_{n-1}A^{n-1} - p_{n-2}A^{n-2} - \ldots - p_1A - p_0I. \tag{4.5}$$

Hence A^n is a linear combination of A^j, $j = 0, 1, \ldots, n-1$. Multiplying (4.5) by A leads to (and substituting A^n of (4.5))

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

Hence A^{n+1} also is a linear combination of A^j, $j = 0, 1, \ldots, n-1$. With induction it can be shown that A^k, $k \geq n$, can be written as such a combination. Consequently, all A^kB can be written as a linear combination of $B, AB, \ldots, A^{n-1}B$ for all $k \geq n$. This is of course also true for all k with $0 \leq k < n$. Therefore Im $A^kB \subset$ Im R for all k. □

THEOREM 4.5 *The following statements are equivalent:*

1. *System (A, B, C, D) is controllable.*
2. *R has rank n.*
3. *Im $R = \mathcal{R}^n$.*

Proof The equivalence of statement 2 and 3 is obvious. We continue with proof $1 \rightarrow 2$. Assuming rank$(R) < n$, we will show that (A, B, C, D) is not controllable. For each input function $u(t)$, $0 \leq t \leq t_1$ $(t_1 > 0)$ we have

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

This formula states that $x(t_1, 0, u)$ is a linear combination of the column vectors of B, AB, $A^2B, \ldots$ and according to Lemma 4.1, $x(t_1, 0, u) \in \text{Im } R$ for each u. If $\text{rank}(R) < n$, then $\text{Im } R \neq \mathcal{R}^n$ (there are points which cannot be reached from $x_0 = 0$) and an n-vector $a \neq 0$ exists such that $a^T R = 0$. Therefore $a^T x(t_1, 0, u) = 0$ which means that the system cannot be steered in the direction of a. Any state reached is always perpendicular to a if the system started at the origin.

Now we will prove the direction $2 \rightarrow 1$. Suppose $\text{rank}(R) = n$. First it will be shown that starting from $x_0 = 0$ each point $x_1 \in \mathcal{R}^n$ can be reached in an arbitrarily short time $t_1 > 0$. Later the initial point being arbitrary will be considered.

Define the symmetric $n \times n$ matrix K as

$$K = \int_0^{t_1} e^{-As} B B^T e^{-A^T s} ds. \tag{4.6}$$

It will be shown in Lemma 4.2 that matrix K is invertible. For arbitrary $x_1 \in \mathcal{R}^n$ and $t_1 > 0$ define

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

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

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

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

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

Hence, for this $\overline{u}$,

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

□

LEMMA 4.2 *The matrix K as defined in (4.6) is invertible.*

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

$$\int_0^{t_1} a^T e^{-As} BB^T e^{-A^T s} a \, ds = 0 \quad \Leftrightarrow$$

$$\int_0^{t_1} \|a^T e^{-As} B\|^2 ds = 0 \quad \Leftrightarrow$$

$$a^T e^{-As} B = 0 \qquad \forall\, s \in [0, t_1].$$

The latter equivalence follows because $a^T e^{-As} B$ is a continuous function of s. Differentiating the latter expression $(n-1)$ times with respect to s and then substituting $s = 0$ gives

$$\begin{aligned} a^T e^{-As} B = 0 \quad &\rightarrow \quad a^T B = 0, \\ a^T A e^{-As} B = 0 \quad &\rightarrow \quad a^T AB = 0, \\ &\vdots \qquad\quad \vdots \\ a^T A^{n-1} e^{-As} B = 0 \quad &\rightarrow \quad a^T A^{n-1} B = 0. \end{aligned}$$

This gives $a^T R = 0$ which is impossible because $\operatorname{rank}(R) = n$. Therefore K is invertible. □

Controllability of a system is determined by the matrices A and B as Theorem 4.5 teaches us. Therefore we will also speak about the controllability of the pair (A, B). The condition $\operatorname{rank} R = n$ 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 identical to $\det R \neq 0$. Please note that Theorem 4.5 does not say anything about $t_1 > 0$. Therefore, 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).

EXAMPLE **4.4** Consider the satellite dynamics of (3.17) 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 $\operatorname{rank} R = 4$ 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. Then (denote the controllability matrix by R_2)

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

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; x_0 and x_1, initial and final point, should be chosen such that the linearized equations (3.17) 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

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

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

EXERCISE 4.5 *Investigate whether system (3.15) -the inverted pendulum- is controllable.*

EXERCISE 4.6 *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. $A = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$

EXAMPLE **4.5** Consider:

$$\dot{x}(t) = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & \dots & -p_{n-1} \end{pmatrix} x(t) + \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} u(t),$$

$$y(t) = (q_0, q_1, \dots, q_{n-1})x(t).$$

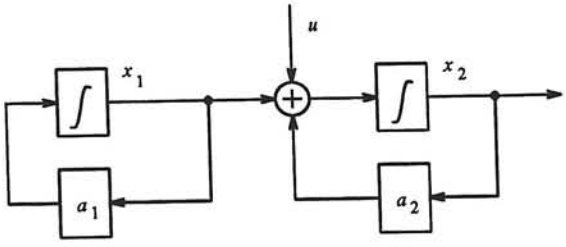
(This system is said to be in the 'standard controllabe form'. Not all systems can be put in this form by means of a coordinate transformation; see the next chapter). For this system

$$R = \begin{pmatrix} 0 & \dots & \dots & 0 & 1 \\ \vdots & & . & 1 & x \\ \vdots & . & . & . & \vdots \\ 0 & 1 & . & . & \vdots \\ 1 & x & \dots & \dots & x \end{pmatrix},$$

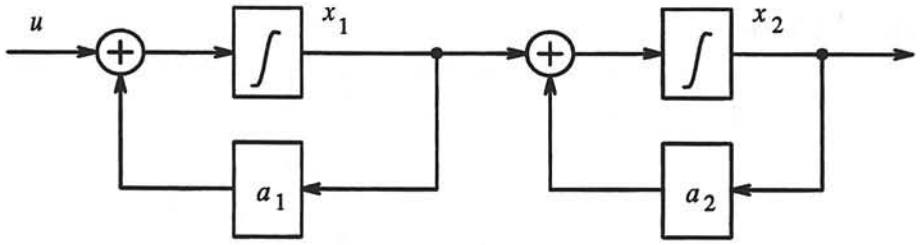
where the x denote possibly nonzero-elements. Since rank $R = n$, this system is always controllable, irrespective of the values of the coefficients q_i and p_i. Hence the name 'standard controllable realization'. □

EXERCISE 4.7 *Are the equations of motion of the tractor and carts (Exercise 3.3) when the combination moves in forward direction controllable? Same question, but now the combination moves in backward direction.*

EXAMPLE **4.6** Controllability can also be studied in terms of flow diagrams. The system



is not controllable since x_1 cannot be influenced by u. The next system,



is controllable. Its state space description is

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

□

If the system (A, B) with state space $\mathcal{R}^n$ is not controllable, then those points of $\mathcal{R}^n$ which are reachable (if the system starts at the origin) are exactly all vectors of Im R.

$$\text{Im } R = \{x_1 \in \mathcal{R}^n | \text{there exists } t_1 > 0 \text{ and } u \in \underline{U} \text{ such that } x_1 = x(t_1, 0, u)\}$$

The space $\underline{U}$ refers to the space of admissible $u(\cdot)$ functions. An algebraic proof of this statement will not be given, instead we will interpret Im R in a geometric way.

DEFINITION 4.6 *A linear subspace $V \in \mathcal{R}^n$ is called* **A-invariant** *if $AV \subset V$.*

THEOREM 4.6 Im R *is the smallest linear subspace of $\mathcal{R}^n$ such that*

1. Im $B \subset$ Im R
2. Im R *is A-invariant*

Proof First it will be shown that Im R satisfies properties 1 and 2. Because $R = [B|AB|\ldots|A^{n-1}B]$, clearly Im $B \subset$ Im R. Furthermore,

$$A(\text{Im } R) = A(\text{Im } [B|AB|\ldots|A^{n-1}B]) = \text{Im } [AB|A^2B|\ldots|A^nB].$$

According to Cayley-Hamilton, A^nB can be expressed as a linear combination of the column vectors of $B, AB, \ldots, A^{n-1}B$. Therefore

$$A(\text{Im } R) \subset \text{Im } [B|AB|\ldots|A^{n-1}B] = \text{Im } R.$$

It remains to be shown that Im R is the smallest subspace that satisfies points 1 and 2. (That such a 'smallest' subspace exists is guaranteed by the fact that

the intersection of two subspaces which both satisfy points 1 and 2, also satisfies these points.) Suppose now that a linear subspace V is given which satisfies points 1 and 2. It will be shown that Im $R \subset V$. Because Im $B \subset V$ and $AV \subset V$ the following inclusion holds:

$$\begin{aligned} \text{Im } AB &= A(\text{Im } B) \subset AV \subset V, \\ \text{Im } A^2B &= A(\text{Im } AB) \subset AV \subset V, \\ &\vdots \quad \vdots \quad \vdots \\ \text{Im } A^{n-1}B &= A(\text{Im } A^{n-2}B) \subset AV \subset V. \end{aligned}$$

Therefore,

$$\text{Im } R = \text{Im } [B|AB|\ldots|A^{n-1}B] = \text{Im } B + \text{Im } AB + \ldots + \text{Im } A^{n-1}B \subset V.$$

□

It is well known (from the theory of matrices) that if a linear subspace $V \subset \mathcal{R}^n$ is A-invariant, then a basis $\{a_1, \ldots, a_n\}$ of $\mathcal{R}^n$ can be found such that $V = \text{span}\{a_1, \ldots, a_k\}$ - $\dim V = k < n$ - and that with respect to this basis the mapping A has the form

$$A = \begin{array}{c} \left(\begin{array}{c|c} A_{11} & A_{12} \\ \ldots \; \ldots & \ldots \\ 0 & A_{22} \end{array} \right) \begin{array}{c} \updownarrow \; k \\ \\ \updownarrow \; n-k \end{array} \\ \begin{array}{cc} \leftrightarrow & \leftrightarrow \\ k & n-k \end{array} \end{array}. \tag{4.7}$$

The matrix $A_{21} = 0$ because $A \begin{pmatrix} x_1 \\ - \\ 0 \end{pmatrix}$ must have the form $\begin{pmatrix} y_1 \\ - \\ 0 \end{pmatrix}$.

Such a basis can for instance be found by means of the Gram-Schmidt procedure. The conclusion here is that if $\dim (\text{Im } R) = k < n$, then a basis $(a_1, \ldots, a_n)$ of $\mathcal{R}^n$ can be found such that Im $R = \text{span}\{a_1, \ldots, a_k\}$ and A has the form (4.7) with respect to this basis. Because Im $B \subset$ Im R and the above mentioned subspace V is identified with Im R, then in the new basis B has the form

$$B = \begin{array}{c} \begin{pmatrix} B_1 \\ \ldots \\ 0 \end{pmatrix} \begin{array}{c} \updownarrow \; k \\ \\ \updownarrow \; n-k \end{array} \\ \begin{array}{c} \leftrightarrow \\ m \end{array} \end{array}. \tag{4.8}$$

In this basis,

$$R = [B|AB|\ldots|A^{n-1}B] = \left[\begin{array}{c|c|c|c|c|c} B_1 & A_{11}B_1 & \ldots & A_{11}^{n-1}B_1 \\ -- \; -- & -- \; -- & -- \; -- & -- \; -- \\ 0 \quad | & 0 \quad | & \ldots \quad | & 0 \end{array} \right],$$

and therefore the pair (A_{11}, B_1) is controllable, since

$$\text{rank } [B_1|A_{11}B_1|\dots|A_{11}^{n-1}B_1] = k,$$

which is the maximum rank for this matrix. The choice of a new basis is equivalent with performing a basis transformation. Therefore, an invertible matrix S exists such that $S^{-1}AS$ and $S^{-1}B$ are of the form (4.7) and (4.8) respectively.

EXAMPLE **4.7** Consider the pair

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

The system corresponding to this (A, B) was considered in Exercise 3.1 and Example 4.4. It represents a satellite with only one input, viz. the thrust in the radial direction. This system is not controllable, rank $R = 3$ according to Example 4.4, which also gives three vectors which span 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 a_1, a_2 and a_3 respectively, then a_4 must be chosen independent of a_1, a_2 and a_3. We choose $a_4 = (2, 0, 0, 1)^T$. Now define

$$S = [a_1|a_2|a_3|a_4] = \begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & -2 & 0 & 1 \end{pmatrix},$$

and hence

$$S^{-1} = 0.1 \times \begin{pmatrix} 0 & 10 & -5 & 0 \\ 2 & 0 & 0 & -4 \\ 0 & 0 & -5 & 0 \\ 4 & 0 & 0 & 2 \end{pmatrix},$$

and calculate $S^{-1}AS$ and $S^{-1}B$:

$$\overline{A} = S^{-1}AS = \left(\begin{array}{ccc|c} 0 & 0 & 0 & 7.5 \\ 1 & 0 & -0.5 & 0 \\ 0 & 1 & 0 & -0.5 \\ \hline 0 & 0 & 0 & 0 \end{array}\right), \overline{B} = S^{-1}B = \left(\begin{array}{c} 1 \\ 0 \\ 0 \\ \hline 0 \end{array}\right).$$

The partitioning in sub-blocks as given by (4.7) and (4.8) is clearly visible. The pair $(\overline{A}_{11}, \overline{B}_1)$, with

$$\overline{A}_{11} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -0.5 \\ 0 & 1 & 0 \end{pmatrix}, \quad \overline{B}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

is controllable since

$$\text{rank } R = \text{rank} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 3.$$

□

EXERCISE 4.8 *Write the noncontrollable pairs of Excercise 4.6 in the form of (4.7) and (4.8).*

Some other tests on controllability exist, to be given next.

THEOREM 4.7 *The pair (A, B) where B is an $n \times 1$ vector and A an $n \times n$ matrix, will be noncontrollable if and only if a vector $q \neq 0$ and a scalar λ exist such that*

$$q^T A = \lambda q^T, \quad q^T B = 0. \tag{4.9}$$

In other words, (A, B) will be controllable if and only if there is no row eigenvector (this in contrast to the more usual column eigenvector) of A that is orthogonal to Im B. *Obviously a row eigenvector of A is a column eigenvector of A^T.*

Proof **Necessity.** If there exists a $q \neq 0$ such that

$$q^T A = \lambda q^T, \quad q^T B = 0,$$

then

$$\begin{aligned} q^T AB &= \lambda q^T B &&= 0, \\ q^T A^2 B &= \lambda q^T AB &&= 0, \\ &\vdots \\ q^T A^{n-1} B &= \lambda q^T A^{n-2} B &&= 0, \end{aligned}$$

so that,

$$q^T [B|AB|\ldots|A^{n-1}B] = q^T R = 0,$$

which means that R has rank less than n, i.e. (A, B) is not controllable.
Sufficiency. We have to show that (A, B) noncontrollable implies the existence of a vector q such that as in (4.9). Assume that the pair (A, B) has been put in the block triangular form of formulas (4.7) and (4.8). Now it is clear that the following vector q is perpendicular to Im B:

$$\begin{matrix} q^T = (& 0 & | & z^T &), \\ & \leftrightarrow & & \leftrightarrow & \\ & k & & n-k & \end{matrix}$$

and it is perhaps not hard to guess that we should choose z^T as a row eigenvector of A_{22}:

$$z^T A_{22} = \lambda z^T,$$

because then

$$q^T A = [0|z^T]A = [0|\lambda z^T] = \lambda^T q.$$

Therefore we have shown how to find a row vector q^T satisfying (4.9) and this completes the proof. □

It is not difficult to verify that the proof above, and therefore Theorem 4.7, is not only valid for single-input systems, but also for multi-input systems.

THEOREM 4.8 *A pair (A, B), possibly representing a multi-input system, is controllable if and only if* $\text{rank}[sI - A|B] = n$ *for all complex s. Here n denotes the size of A. Note that this condition will clearly be met for all s that are not eigenvalues of A because then* $\det[sI - A] \neq 0$.

The proof of the Theorem is that the rank must be n, even when s is an eigenvalue of A.

Proof If $[sI - A|B]$ has rank n, there cannot be a nonzero row vector q^T such that

$$q^T[sI - A|B] = 0,$$

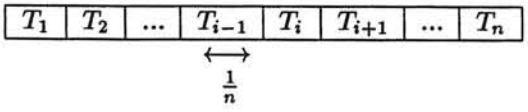
for any s, i.e. such that $q^T B = 0$ and $q^T A = sq^T$. But then by the previous theorem, (A, B) must be controllable. The converse follows easily by reversing the above arguments. □

EXERCISE 4.9 *We are given a single input system $\dot{x} = Ax + Bu$ which is controllable. Suppose that a control is applied of the form $u = Kx + v$, where K is a $1 \times n$ and v the "new" control, which is a scalar also. The new system is the characterized by the pair $(A + BK, B)$. Prove, by using Theorem 4.8, that this new system is also controllable.*

EXAMPLE 4.8 The starting point of this example is equation (2.7) of the heated bar:

$$\frac{\partial T(t,r)}{\partial t} = c\frac{\partial^2 T(t,r)}{\partial r^2}, \quad 0 \leq r \leq 1;\ t \geq 0, \tag{4.10}$$

where we will assume now that $c = 1$. The temperature can be controlled



at both ends of the bar; at $r = 0$ by means of u_1 and 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 $1/n$. The temperature on the interval $(i-1)/n \leq r < i/n$ is indicated by T_i, $i = 1, 2, \ldots, n$. If we use the approximations

$$\begin{aligned}
\frac{\partial^2 T_i}{\partial r^2} &= n^2(T_{i-1} - 2T_i + T_{i+1}),\ i = 2, 3, \ldots, n-1, \\
\frac{\partial^2 T_1}{\partial r^2} &= n^2(u_1 - 2T_1 + T_2), \\
\frac{\partial^2 T_n}{\partial r^2} &= n^2(T_{n-1} - 2T_n + u_2),
\end{aligned}$$

then we get the following finite dimensional model:

$$\frac{1}{n^2}\frac{d}{dt}\begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ \vdots \\ T_n \end{pmatrix} = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 0 & \cdots & 0 & 1 & -2 \end{pmatrix}\begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ \vdots \\ T_n \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

By checking the controllability matrix of this finite dimensional system (not shown here), it is easily verified that this system is controllable; one can steer to any temperature profile. Is controllability maintained if $n \to \infty$, or in other words, is the system (4.10) 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

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

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

4.3 Observability

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

DEFINITION 4.7 *The system* (A, B, C, D) *is* **observable** *if a finite time* $t_1 > 0$ *exists such that for each* $u \in \underline{U}$ *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 state $x(0)$ can be constructed from the knowledge of u and y on the interval $[0, t_1]$. In other words, to the external behaviour of an observable system restricted to the interval $[0, t_1]$ corresponds only one state

vector at time $t = 0$. As with controllability there are several definitions possible for observability. A slightly different one would be that a system is observable if for each pair of states x_0 and x_1, $x_0 \neq x_1$, a control $u \in \underline{U}$ 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 can be distinguished. Definition 4.7, 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

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

LEMMA 4.3 *If*

$$Cx = CAx = \ldots = CA^{n-1}x = 0,$$

then $CA^k x = 0$ *for all* $k \geq 0$.

Proof For $k = 0, 1, \ldots, n-1$ the statement is immediate. According to the Cayley-Hamilton theorem, A^k is a linear combination of A^j, $j = 0, 1, \ldots, n-1$, for $k \geq n$, see also Lemma 4.1. Therefore

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

for certain scalars $\alpha_{i,k}$, and hence

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

□

THEOREM 4.9 *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 $u \in \underline{U}$, 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

$$Ce^{At}x_0 = Ce^{At}x_1 \quad \text{or} \quad Ce^{At}(x_0 - x_1) = 0,$$

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

$$\begin{array}{ccc} Ce^{At}(x_0 - x_1) = 0 & \rightarrow & C(x_0 - x_1) = 0, \\ CAe^{At}(x_0 - x_1) = 0 & \rightarrow & CA(x_0 - x_1) = 0, \\ \vdots & \vdots & \vdots \\ CA^{n-1}e^{At}(x_0 - x_1) = 0 & \rightarrow & CA^{n-1}(x_0 - x_1) = 0. \end{array} \tag{4.11}$$

The result can be written as $W(x_0 - x_1) = 0$. Since rank $W = n$ we have $x_0 = x_1$.

Now we will prove the direction $1 \rightarrow 2$. Suppose rank $W < n$ and it will be shown that the system is not observable. Two different vectors x_0 and x_1 exist such that $x_0 - x_1 \in \ker W$. Then

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

Application of Lemma 4.3 yields $CA^k(x_0 - x_1) = 0$ for all $k \geq 0$. Subsequently,

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

for all t. This now is equivalent with $y(t, x_0, u) = y(t, x_1, u)$ for all t and all adnissible controls. Hence the system is not observable. □

Observability is completely determined by the matrices A and C and therefore we will sometimes speak about 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. Condition rank$(W) = n$ is called the **rank condition** for observability. In case of a single output system W is a square matrix and (C, A) observable is then equivalent to det $W \neq 0$.

EXERCISE 4.10 *A nonsingular coordinate transformation $x = S\bar{x}$, (such that $A \rightarrow S^{-1}AS$, $C \rightarrow CS$) does not destroy observability. Show this. If the observability matrix of the transformed system is denoted by $\bar{W}$, then $WS = \bar{W}$.*

If (C, A) is observable, the interval $[0, t_1]$ in Def. 4.7 may be chosen arbitrarily small, as long as $t_1 > 0$ (the differentiations in (4.11) 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 **4.9** Consider again the satellite dynamics as described in Exercise 3.1 and the output equations given in Example 3.7. 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 current state $x(t)$ can be constructed if we are given the measurement $y(t)$ and the input $u(t)$ of the system on an interval $[t_0, t]$ with $t_0 < t$. (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 on observers). Suppose that only y_2 can be measured. The corresponding observability matrix W_2 is

$$W_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \\ -6 & 0 & 0 & -4 \end{pmatrix},$$

which is nonsingular. Therefore the state is uniquely determined if only y_2 (and u) is available. If only y_1 is available, then

$$W_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & -1 & 0 & 0 \end{pmatrix},$$

and W_1 has rank 3. This system is nonobservable. □

EXERCISE 4.11 *Investigate whether the inverted pendulum, as given by the equations (3.15) and (3.16) is observable. Repeat this investigation if only one of the measurements (i.e. either $y_1(t)$ or $y_2(t)$) is available.*

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

THEOREM 4.10 *(A, B) is controllable if and only if (B^T, A^T) is observable. (C, A) is observable if and only if (A^T, C^T) is controllable.*

Proof (A,B) controllable $\iff$ rank$[B \mid AB \mid .... \mid A^{n-1}B] = n \iff$
$\iff$ rank$[B \mid AB \mid .... \mid A^{n-1}B]^T = n \iff$

$$\iff \text{rank} \begin{pmatrix} B^T \\ - - - \\ B^T A^T \\ - - - \\ \vdots \\ - - - \\ B^T (A^T)^{n-1} \end{pmatrix} = n \iff$$

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

The conclusion of this theorem is that $\dot{x} = Ax + Bu$ is controllable if and only if the system $\dot{z} = A^T z$, $y = B^T z$ is observable. The transposition of matrices $A \to A^T$, $B \to B^T$ is a simple example of the concept of **duality**.

REMARK **4.1** This is a side remark with respect to the notion of duality. The **dual space** of $\mathcal{R}^n$ consists of all linear functions $c : \mathcal{R}^n \to \mathcal{R}$ and this is isomorphic to the set of n-dimensional row-vectors. This dual space is written as $(\mathcal{R}^n)^*$, which in this example can be identified with $\mathcal{R}^n$ itself. If $A : \mathcal{R}^n \to \mathcal{R}^n$ is a linear mapping, then A^T is strictly speaking a mapping from $(\mathcal{R}^n)^*$ to $(\mathcal{R}^n)^*$. □

The mathematical conditions for controllability and observability are dual. Sometimes people speak about duality of these concepts themselves. This duality enables us to formulate results for observability by "dualizing" proven results of controllability. Examples of "dualization" are the following two theorems which will be given without proof.

THEOREM 4.11 ker W *is the largest linear subspace in $\mathcal{R}^n$ that satisfies*

1. ker $W \subset$ ker C,

2. ker W *is A-invariant.*

The linear subspace ker W is called the **nonobservable subspace.** Elements of ker W are exactly those states which cannot be distinguished from the origin. An application of Theorem 4.11 is the following. A basis $(a_1, \ldots . a_n)$ in $\mathcal{R}^n$ exists such that ker W = span $(a_1, \ldots ., a_k)$ and A has the form (with respect to this basis):

$$A = \begin{array}{c} \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array}\right) \\ \begin{array}{cc} \leftrightarrow & \leftrightarrow \\ k & n-k \end{array} \end{array} \quad \begin{array}{c} \updownarrow \; k \\ \updownarrow \; n-k \end{array} . \tag{4.12}$$

Because ker $W \subset$ ker C, in this basis

$$C = \begin{array}{c} \left(\begin{array}{c|c} 0 & C_2 \end{array}\right) \\ \begin{array}{cc} \leftrightarrow & \leftrightarrow \\ k & n-k \end{array} \end{array} . \tag{4.13}$$

Furthermore, the pair (C_2, A_{22}) is observable.

THEOREM 4.12 *A pair* (C, A) *will be observable if and only if*

$$\operatorname{rank} \begin{bmatrix} C \\ sI - A \end{bmatrix} = n \quad \textit{for all complex } s .$$

The connection between input and output is (with $x(0) = 0$ and $D = 0$):

$$y(t) = \int_0^t Ce^{A(t-s)}Bu(s)ds,$$

where $Ce^{At}B$ is the impulse response matrix. Now suppose that (A, B) is not controllable, then a basis in $\mathcal{R}^n$ exists such that

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ 0 \end{pmatrix} . \tag{4.14}$$

In this basis

$$e^{At}B = \begin{pmatrix} B_1 + A_{11}B_1 t + A_{11}^2 B_1 \frac{t^2}{2!} + \ldots . \\ 0 \end{pmatrix} = \begin{pmatrix} e^{A_{11}t}B_1 \\ 0 \end{pmatrix} .$$

The second part of the state space (i.e. the complement of Im R in $\mathcal{R}^n$) does not play any role in the impulse response matrix. The conclusion is that only the controllable subspace of $\mathcal{R}^n$ plays a role in the impulse response matrix. Similarly,

suppose (C, A) is not observable, then a basis in $\mathcal{R}^n$ exists (not necessarily the same as above) such that

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad C = (0 \;\; C_2) . \tag{4.15}$$

Please be aware of the fact that the A_{ij} in (4.14) and (4.15) will in general be different. In this basis

$$Ce^{At} = (0 \;\; C_2 e^{A_{22}t}) .$$

The first part of the state space (i.e. ker W) does not play any role in the impulse response matrix.

4.4 Realization theory and Hankel matrices

In this section we consider single-input single-output system only. The quantities

$$g_i = CA^{i-1}B , \quad i = 1, 2, ... \tag{4.16}$$

which are called the **Markov parameters**, determine the external description of the system

$$\dot{x} = Ax + Bu , \quad y = Cx . \tag{4.17}$$

The Markov parameters appear in the power series expansion of the impulse response

$$G(t) = Ce^{At}B = \sum_{i=0}^{\infty} CA^i B \frac{t^i}{i!} . \tag{4.18}$$

From this latter equation it follows that

$$g_i = \frac{d^{i-1}}{dt^{i-1}} G(t)|_{t=0} .$$

We form the so-called **Hankel matrix** of size $\alpha \times \beta$,

$$H(\alpha, \beta) = \begin{pmatrix} g_1 & g_2 & g_3 & \cdots & g_\beta \\ g_2 & g_3 & g_4 & & g_{\beta+1} \\ g_3 & g_4 & & & \vdots \\ g_4 & \vdots & & & \\ \vdots & \vdots & & & \vdots \\ g_\alpha & g_{\alpha+1} & \cdots & \cdots & g_{\alpha+\beta-1} \end{pmatrix} . \tag{4.19}$$

THEOREM 4.13 *To the sequence* $\{g_1, g_2, ...\}$ *corresponds a finite dimensional realization of the form (4.17) of order* n *(i.e. the state space is* $\mathcal{R}^n$*) if and only if*

$$\det H(n+i, n+i) = 0 \quad \textit{for} \quad i = 1, 2, ...$$

If moreover $\det H(n,n) \neq 0$ *then* n *is the order of the minimal realization of the sequence* $\{g_1, g_2, \ldots\}$

The proof wil 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 therefore coefficients $a_1, ..., a_n$ must exist such that

$$a_n g_j + a_{n-1} g_{j+1} + ... + a_1 g_{j+n-1} + g_{j+n} = 0\,, \quad j = 1, ..., n, n+1\,.$$

Also without proof it is stated that, given $\{g_1, g_2, \ldots\}$ such that the conditions mentioned in Theorem 4.13 are satisfied, a possible realization of the underlying system in state space space form is

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -a_n & \dots & \dots & \dots & -a_1 \end{pmatrix}, \quad B = \begin{pmatrix} g_1 \\ \vdots \\ \vdots \\ \vdots \\ g_n \end{pmatrix}, \quad C^T = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}.$$

Such a realization is not unique; basis transformations will give other realizations.

EXERCISE 4.12 *Suppose that the Hankel matrices* $H(\alpha, \beta)$ *satisfy the conditions given in the statement of Theorem 4.13. Prove that* $H(n,n) = WR$*, where* W *and* R *are the controllability and the obervability matrices respectively, and that , if* $\det H(n,n) \neq 0$*, an* n*-dimensional state space realization is both controllable and observable.*

Chapter 5

State and Output Feedback

5.1 Feedback and stabilizability

In Example 1.1 about the autopilot of a boat, the control u was expressed in known quantities such as to obtain a good steering behaviour of the ship. A possible control law had the form $u = Ke$, where K is a constant, e is the error between the actual and the desired heading. One can imagine that the desired heading has been set by the helmsman and that the actual course is continuously measured (is an output signal 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 (or the measurements) is fed back to the input u Such a control is a form of **feedback control**, or, equivalently, **closed-loop control** (the output is connected to the input; the loop is closed, the system will govern itself). In contrast to closed-loop control there exists **open-loop** control. In a system with open-loop control the control action (the function $u(t)$) is independent of the output.

EXAMPLE **5.1** An automatic toaster (i.e. a toaster that switches off automatically) is an open-loop system because it is controlled by a timer (the function $u(t)$ is an on-off function). The time 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 (color is the output) is removed once the timer has been set. One could design a toaster with a feedback control; the color of the toast is continuously measured and this measurement is connected to the switch of the heating element. □

EXERCISE 5.1 *Consider the one-dimensional model* $\dot{x}(t) = u(t)$, $t \geq 0$, $x(0) = 1$ *with the following two options for the (one-dimensional) control function* u*:*

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 x satisfies $x(t) = e^{-t}$. Which of the two control options would you prefer if there are (unmodelled) disturbances in the system and if the aim of the control design is to have $\lim_{t\to\infty} x(t) = 0$?

We will now turn to a more mathematical treatment of the feedback principle. Suppose we are given a system described by

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

and this system is unstable. Furthermore assume $C = I$ and $D = 0$; i.e. the whole state is observed. To focus the ideas, one could think of the inverted pendulum introduced in Example 3.2. Given an initial perturbation $x_0 \neq 0$ ($x_0 = 0$ corresponds to the unstable equilibrium of a carriage at rest situated at $s = 0$, with a vertical pendulum), one could calculate a time function $u(t) : [0, \infty) \to \mathcal{R}$ 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 F is a $m \times n$ (here 1×4) matrix. The state x then satisfies

$$\dot{x} = Ax + BFx = (A + BF)x . \tag{5.2}$$

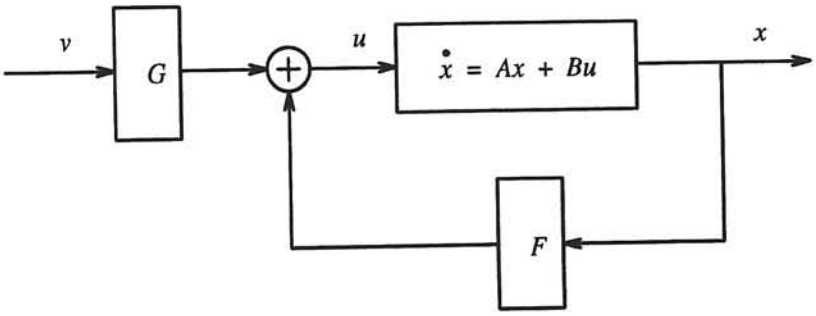
The matrix F must be chosen such that the behaviour of the closed-loop system (5.2) has a desired behaviour (if possible), i.e. for instance being asymptotically stable. A control law of the form (5.1) is called **state feedback.** If the state is not available, one might feed back the output; $u = Hy$ where H is a suitably chosen $m \times p$ matrix. The state x will then satisfy

$$\dot{x} = Ax + BHy = Ax + BHCx = (A + BHC)x$$

(provided $D = 0$). 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 possibility for the control law is then

$$u = Fx + Gv ,$$

where v is the new input, 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 $v(t)$).



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 dynamic behaviour 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 will be the heart of this section, though there are other system properties which can also be influenced by means of a compensator. Conditions on the matrices A, B will be given such that the new matrix $A+BF$ is asymptotically stable if an appropriate matrix F is chosen.

DEFINITION 5.1 *The system* $\dot{x} = Ax + Bu$ *is* **stabilizable** *if an* $m \times n$ *matrix* F *exists such that* $\Re\lambda < 0$ *for all eigenvalues* λ *of* $A + BF$.

THEOREM 5.1 *The system* $\dot{x} = Ax + Bu$ *is controllable if and only if for each polynomial* $r(\lambda) = \lambda^n + r_{n-1}\lambda^{n-1} + \ldots + r_1 s + r_0$, *with real coefficients* $r_{n-1}, \ldots, r_0$, *an* $m \times n$ *matrix* F *exists such that* $\det(\lambda I - (A + BF)) = r(\lambda)$.

Hence, if (A, B) is controllable, the characteristic polynomial of $(A + BF)$ can be chosen arbitrarily for 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. A particular location is the left half of the complex plane, such that if $\dot{x} = Ax + Bu$ is controllable then it is also stabilizable (the converse statement is not necessarily true). Theorem 5.1 is sometimes called the **pole-assignment theorem.**

Proof of Theorem 5.1. It will only be given for single-input systems.

Necessity. In this part we prove that, if the system is not controllable, an F matrix 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 in $\mathcal{R}^n$ can be found such that, see formulas (4.7) and (4.8),

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ 0 \end{pmatrix}. \tag{5.3}$$

If we partition an arbitrary feedback matrix F accordingly as (F_1, F_2), then

$$A+BF = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} + \begin{pmatrix} B_1 \\ 0 \end{pmatrix} (F_1 \;\; F_2)$$

$$= \begin{pmatrix} A_{11}+B_1F_1 & A_{12}+B_1F_2 \\ 0 & A_{22} \end{pmatrix}$$

and the characteristic polynomial is

$$\det(\lambda I - A - BF) = \det \begin{pmatrix} \lambda I - (A_{11}+B_1F_1) & -(A_{12}+B_1F_1) \\ 0 & \lambda I - A_{22} \end{pmatrix} =$$

$$= \det \begin{pmatrix} \lambda I - (A_{11}+B_1F_1) & 0 \\ 0 & I \end{pmatrix} \cdot$$

$$\det \begin{pmatrix} I & (\lambda I - (A_{11}+B_1F_1))^{-1}(A_{12}+B_1F_1)(\lambda I - A_{22})^{-1} \\ 0 & I \end{pmatrix} \cdot$$

$$\det \begin{pmatrix} I & 0 \\ 0 & \lambda I - A_{22} \end{pmatrix} =$$

$$= \det(\lambda I - (A_{11}+B_1F_1)) \cdot \det(\lambda I - A_{22}) .$$

This formula is based on the matrix identity

$$\begin{pmatrix} P & Q \\ 0 & R \end{pmatrix} = \begin{pmatrix} P & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & P^{-1}QR^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & R \end{pmatrix},$$

where it is tacitly assumed that the inverses exist.

Whatever the choice of F is, the polynomial $\det(\lambda I - A_{22})$ forms always a part of the characteristic polynomial of $A+BF$ and 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 can be found if the system is controllable. Hence, 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 so-called *controllability canonical form*, defined as

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & \dots & & -p_{n-1} \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{5.4}$$

That such a coordinate transformation exists will be proved in the next lemma. Choose $F = (p_0 - r_0, p_1 - r_1, ..., p_{n-1} - r_{n-1})$, then

$$A + BF = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & \dots & \dots & -p_{n-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} (p_0 - r_0, ..., p_{n-1} - r_{n-1})$$

$$= \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ -r_0 & -r_1 & \dots & \dots & \dots & -r_{n-1} \end{pmatrix}.$$

and therefore $\det(\lambda I - (A + BF)) = r(\lambda)$ (see Exercise 3.8). It is clear that 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 $S^{-1}(A + BF)S$ where S is an invertible matrix) and the result therefore also holds true for the original system which was possibly not in the *controllability canonical form.* □

In the proof of Theorem 5.1 we used the following lemma.

LEMMA 5.1 *If (A, B), with $m = 1$, is controllable, then a basis transformation S exists,* $\det S \neq 0$, *such that $\overline{A} = S^{-1}AS$ and $\overline{B} = S^{-1}B$ are in the controllability canonical form, as defined in Equation (5.4). The elements p_i are the coefficients in the characteristic polynomial of A (and of $\overline{A}$)* : $\det(\lambda I - A) = \lambda^n + p_{n-1}\lambda^{n-1} + ... + p_1\lambda + p_0$.

Proof A new basis $(q_1, ..., q_n)$ in $\mathcal{R}^n$ is constructed as follows:

$$\begin{aligned} q_n &= B, \\ q_{n-1} &= AB + p_{n-1}B = Aq_n + p_{n-1}q_n, \\ q_{n-2} &= A^2B + p_{n-1}AB + p_{n-2}B = Aq_{n-1} + p_{n-2}q_n, \\ &\vdots \\ q_1 &= A^{n-1}B + p_{n-1}A^{n-2}B + ... + p_1B = Aq_2 + p_1q_n \,. \end{aligned} \tag{5.5}$$

Because (A, B) is controllable, span $\{B, AB, ..., A^{n-1}B\} = \mathcal{R}^n$ and therefore, by construction, also span $\{q_1, ..., q_n\} = \mathcal{R}^n$. Hence $\{q_1, ..., q_n\}$ is a basis. The

corresponding basis transformation is defined by $q_i = Se_i$, $i = 1, ..., n$, with $\{e_1, ..., e_n\}$ being the standard basis of $\mathcal{R}^n$ (i.e. q_i is the ith column of S). In this new coordinate system B is equal to $(0, ..., 0, 1)^T$, since $B = q_n$. Furthermore, from the second till the last equation of (5.6) we obtain

$$\begin{aligned} Aq_n &= q_{n-1} - p_{n-1}q_n \\ Aq_{n-1} &= q_{n-2} - p_{n-2}q_n \\ &\vdots \\ Aq_2 &= q_1 - p_1 q_n \end{aligned}$$

and we can write (using again the last equation of (5.6), and Cayley-Hamilton)

$$\begin{aligned} Aq_1 &= A(A^{n-1}B + p_{n-1}A^{n-2}B + ... + p_1 B) \\ &= A^n B + p_{n-1}A^{n-1}B + ... + p_1 AB \\ &= (A^n + p_{n-1}A^{n-1} + ... + p_1 A + p_0 I - p_0 I)B = -p_0 B = -p_0 q_n \, . \end{aligned}$$

Now Aq_i, $i = 1, .., n$, have been expressed as linear combinations of q_j, $j = 1, ..., n$. From these expressions we see directly that A in the new basis can be written as (by realizing that $S\overline{A} = AS$)

$$\overline{A} = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & 0 & 1 & & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & -p_2 & \dots & \dots & -p_{n-1} \end{pmatrix} .$$

□

EXERCISE 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 its unstable subspace (see Definition 4.2) is contained in its controllable subspace* (Im R, *see text above Lemma 4.1). Hint: assume A and B are given with respect to a basis in $\mathcal{R}^n$ such that they have the form (4.7), (4.8).*

The proof of Theorem 5.1 yields an algorithm for finding an $1 \times n$ matrix F that gives the system desired properties, i.e. a desired characteristic polynomial. Towards the end the system (A, B) is first transformed to $(S^{-1}AS, S^{-1}B)$, the *controllable canonical form.* With respect to this form F is obtained in a trivial way. With respect to the original basis F must be replaced by FS. The following exercise gives another, more direct, algorithm.

EXERCISE 5.3 *Prove that the following algorithm yields a correct feedback matrix F, if A, B and the desired $r(\lambda)$ are given (B is $1 \times n$), as follows:*

$$F = -[0, ..., 0, 1][B|AB|...|A^{n-1}B]^{-1} r(A) ,$$

where $r(A) = A^n + r_{n-1}A^{n-1} + ... + r_1 A + r_0 I$.

EXAMPLE **5.2** Consider the system of the inverted pendulum, Eq. (3.15). Since this system is controllable (Exercise 4.5), it can be made asymptotically stable for a appropriate feedback matrix $F = (f_1, f_2, f_3, f_4)$. Of course, in order for this feedback control to be realisable, all components of the state vector must be known (in the next section we will see what can be done if this is not true, but instead only the output y is known). The matrices A and B are

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix} , \quad B = \begin{pmatrix} 0 \\ -2.4 \\ 0 \\ 1 \end{pmatrix} .$$

A has eigenvalues 0 (2×), 5 and -5 (Example 3.5) 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 $-2 \pm i$; the 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 either use the algorithm in the proof of Theorem 5.1 or the formula in Exercise 5.3. Another more direct method is that $F = (f_1, f_2, f_3, f_4)$ must be chosen such that

$$\det (\lambda I - (A + BF)) = r(\lambda) .$$

This gives

$$\det \begin{pmatrix} \lambda & -1 & 0 & 0 \\ -25 + 2.4f_1 & \lambda + 2.4f_2 & 2.4f_3 & 2.4f_4 \\ 0 & 0 & \lambda & -1 \\ 0.6 - f_1 & -f_2 & -f_3 & \lambda - f_4 \end{pmatrix}$$
$$= \lambda^4 + 7\lambda^3 + 19\lambda^2 + 23\lambda + 10.$$

Hence,

$$\lambda^4 + (2.4f_2 - f_4)\lambda^3 + (-f_3 - 25 + 2.4f_1)\lambda^2 + (25f_4 - 1.44f_4)\lambda +$$
$$(25f_3 - 1.44f_3) = \lambda^4 + 7\lambda^3 + 19\lambda^2 + 23\lambda + 10$$

and the correct feedback components are therefore

$$f_3 = \frac{10}{23.56}; \quad f_4 = \frac{23}{23.56}; \quad f_1 = \frac{1}{2.4}(44 + \frac{10}{23.56}); \quad f_2 = \frac{1}{2.4}(7 + \frac{23}{23.56}).$$

□

EXAMPLE **5.3** We are given the system (in controllable canonical 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 $u(t) = (f_1 \quad f_2 \quad f_3)x(t)$, with f_i 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

$$\lambda^3 + (-1 - f_3)\lambda^2 + (3 - f_2)\lambda + (-2 - f_1) = 0 \, .$$

Since the exact location of the zeros is not important, we will use the criterion of Routh (see Section 4.1) to obtain conditions for f_i which will guarantee asymptotic stability. The scheme is

$$\begin{array}{cc} 1 & 3 - f_2 \quad 0 \\ -1 - f_3 & -2 - f_1 \quad 0 \\ \frac{(-1-f_3)(3-f_2)-(-2-f_2)}{-1-f_3} & 0 \quad 0 \\ -2 - f_1 & 0 \\ 0 & \end{array}$$

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 **5.4** We are 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 (sub)space?
2. Is the system observable? What is the nonobservable (sub)space?

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 which span the intersection of the controllable (sub)space and the nonobservable (sub)space;
   - append to these vectors new basis vectors such that both the controllable (sub)space and the nonobservable (sub)space are spanned;
   - add more basis vectors, if necessary, such that the whole state space ($\mathcal{R}^4$) is spanned.

5. Can you design a control law $u = Fx$ such that the feedback system has its poles in -1, -1, -3 and -4, respectively?. The same question again, but now the poles 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

$$\mathrm{Im}R = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

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 nonobservable subspace is spanned by (two) linearly independent vectors x for which $Wx = 0$. Two such vectors are

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

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 nonobservable subspace is spanned by the vector $v_1 = (1\ 1\ 0\ 1)^T$. The controllable subspace is spanned by v_1 and v_2, with $v_2 \stackrel{\text{def}}{=} (1\ 0\ 1\ 0)^T$; the nonobservable subspace is spanned by v_1 and v_3, with $v_3 \stackrel{\text{def}}{=} (0\ 1\ -1\ 0)^T$. Finally, $\mathcal{R}^4$ is spanned by v_1, v_2, v_3 and v_4, with $v_4 \stackrel{\text{def}}{=} (1\ 0\ 0\ 0)^T$. If we choose the basis transformation $S = (v_1, v_2, v_3, v_4)$, then

$$\tilde{A} = S^{-1}AS = \begin{pmatrix} -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 2 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}, \quad \tilde{B} = S^{-1}B = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$\tilde{C} = CS = (0\ 1\ 0\ 0).$$

Answer question 5. Matrices $\tilde{A}$ and $\tilde{B}$ are of the form

$$\tilde{A} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} B_1 \\ 0 \end{pmatrix}.$$

The eigenvalues of A_{22} will not change if a feedback law is introduced. Therefore, whatever (linear) feedback law is implemented, the poles at locations -3 and -4 (the eigenvalues of A_{22}), cannot be influenced. Therefore, a design with the first set of requirements (poles at -1, -1, -3, -4) is possible; a design subject to the second set of requirements (poles at -1, -1, -2, -3) is not possible. □

EXERCISE 5.4 *Consider the noncontrollable 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?

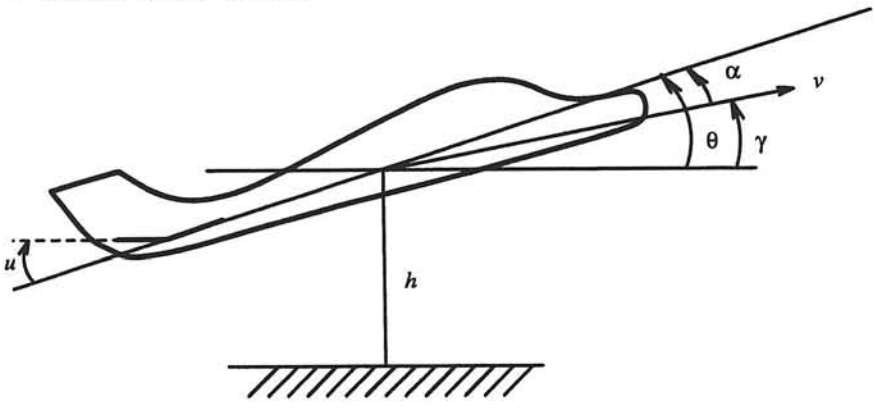
The conclusion of all this is that $A + BF$ can be made asymptotically stable provided (A, B) is stabilizable *and* provided state feedback is possible, i.e. the output y equals the state. If $y = Cx$ and C is not invertible, then the problem of making $A + BHC$ asymptotically stable by means of a suitable 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 constructing the state x out of past measurements y such that the state is indeed available for feedback purposes.

EXERCISE 5.5 *We consider the equations of motion of an airplane in a vertical plane. 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.*



One must design an automatic pilot to keep h constant (and equal to zero) in the presence of all kinds of perturbations such as vertical gusts.

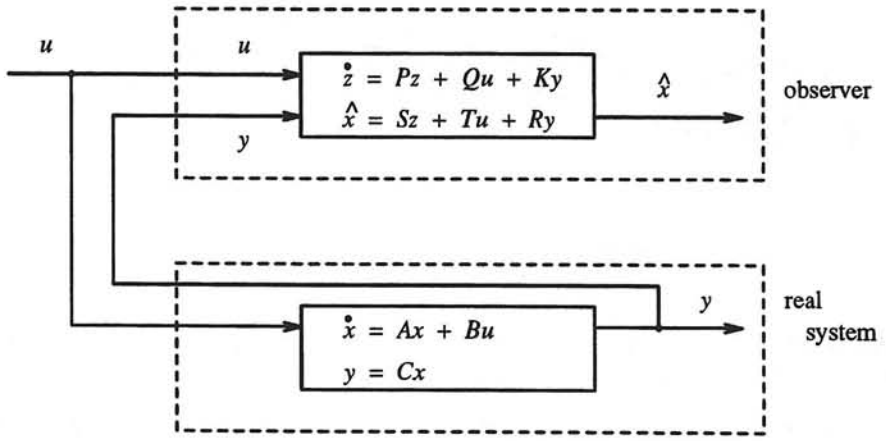
- *Linearize the equations of motion and write them as a set of first order differential equations.*
- *Show that the designer who proposes a feedback of the form $u = kh$, where k is a suitably chosen constant, cannot be successful.*
- *Prove that a feedback of the form $u = k_1 h + k_2\theta$, with suitably chosen constants k_1 and k_2, 'does the job', i.e. the resulting closed-loop system is asymptotically stable.*

5.2 Observers and the separation principle

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; because of the weight problems hardly any measurement equipment (for the temperature for instance) can be built into the satelite. Once in orbit, it is too far away to measure certain quantities from the earth. In all these cases, control must be based on the available information, viz. the output $y = Cx$ (take for simplicity $D = 0$; the case $D \neq 0$ can be handled in the standard case if we interpret $y - Du = Cx$ as the new measurement). An auxiliary system will be built, called the **observer**, which has as input 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 of the form

$$\begin{aligned} \dot{z} &= Pz + Qu + Ky \\ \hat{x} &= Sz + Tu + Ry \end{aligned}$$



In the flow diagram both the observer and 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. Thinking of the real system as a satellite in orbit, where x cannot easily be measured, only a few output variables such as position and distance for instance may be helpful. The observer is an auxiliary system on earth (a computer program for instance) from which all variables are easily obtained.

If at all possible, the observer should at least satisfy the following requirements:

1. If $\hat{x}(t_0) = x(t_0)$ at a certain time instant t_0, then we should have $\hat{x}(t) = x(t)$

for $t \geq t_0$. Once the observer has the correct estimate of the real state vector, then this estimate should remain correct for the future.

2. The difference $x(t) - \hat{x}(t)$ must converge to zero (as $t \to \infty$), irrespective of the initial condition conditions $x(0) = x_0$, $z(0) = z_0$ and the control u.

We will construct an observer for which $S = I$, $T = R = 0$. This yields $\hat{x} = z$ and the state of the observer has the role of an approximation of the state x. We now have

$$\begin{aligned} \frac{d}{dt}(x - \hat{x}) &= Ax + Bu - P\hat{x} - Qu - Ky \\ &= Ax + Bu - P\hat{x} - Qu - KCx \\ &= (A - KC)x - P\hat{x} + (B - Q)u \,. \end{aligned}$$

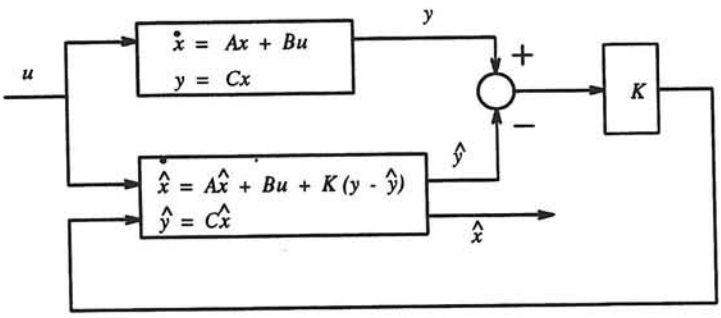
The first requirement formulated above now yields

$$B = Q \,, \quad A - KC = P \,.$$

The observer then has the form

$$\dot{\hat{x}} = A\hat{x} + Bu + K(y - \hat{y}) \quad \text{with} \quad \hat{y} = C\hat{x} \,. \tag{5.6}$$

Apparently it 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 second requirement to be satisfied we consider how the estimate error $e(t) = x(t) - \hat{x}(t)$ behaves as $t \to \infty$. We have here

$$\dot{e} \;=\; \frac{d}{dt}(x - \hat{x}) = (A - KC)e \,.$$

Because $e(t)$ should converge to zero, this requirement now can be translated into: the matrix $(A - KC)$ must be asymptotically stable. Can a matrix K be found such that this is possible? The following theorem states that this is possible if (C, A) is observable.

THEOREM 5.2 *For each polynomial* $w(\lambda) = \lambda^n + w_{n-1}\lambda^{n-1} + \ldots + w_1\lambda + w_0$ *with real coefficients* $w_0, w_1, \ldots, w_{n-1}$, *an* $n \times p$ *matrix* K *exists such that* $\det(\lambda I - (A - KC)) = w(\lambda)$ *if and only if* (C, A) *is observable.*

Proof (C, A) is observable if and only if (A^T, C^T) is controllable (Theorem 4.10). Theorem 5.1 states that (A^T, C^T) is controllable if and only if for each polynomial $w(\lambda)$ as mentioned in the statement of this theorem a matrix F exists such that $\det(\lambda I - (A^T + C^T F)) = w(\lambda)$. Choose $K = -F^T$, then

$$\det(\lambda I - (A - KC)) = \det(\lambda I - (A^T - C^T K^T)) = w(\lambda) .$$

□

EXAMPLE **5.5** This example is a continuation of Example 5.2 of the inverted pendulum. We assume that only measurements of the position of the carriage are made, such that A and C are

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix}, \quad C = (0\ 0\ 1\ 0) .$$

The observability matrix is

$$W = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \\ 0 & -0.6 & 0 & 0 \end{pmatrix} \text{ so rank } W = 4 ,$$

such that Theorem 5.2 can indeed be applied.

Suppose we want to construct an observer such that the poles of $(A - KC)$ are situated in the points -1 $(2\times)$ and $-1 \pm i$. This means that $K = (k_1, k_2, k_3, k_4)^T$ must be constructed 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

$$k_3 = 4\,;\;\; k_4 = 32\,;\;\; k_1 = -\frac{106}{0.6}\,;\;\; k_2 = -\frac{802}{0.6}\,.$$

The observer has the form

$$\frac{d\hat{x}}{dt} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -0.6 & 0 & 0 & 0 \end{pmatrix} \hat{x} + \begin{pmatrix} 0 \\ -2.4 \\ 0 \\ 1 \end{pmatrix} u + \begin{pmatrix} \frac{-1060}{6} \\ \frac{-8020}{6} \\ 4 \\ 32 \end{pmatrix} (y - (0\ 0\ 1\ 0)\hat{x})\,.$$

The solution of this observer satisfies $\lim_{t\to\infty}(\hat{x}(t) - x(t)) = 0$. □

EXERCISE 5.6 *Consider the dynamics of the satellite as given in Example 4.9. 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 poles of the error equation $\dot{e} = (A - KC)e$ are all situated in -1.*

Theorem 5.2 gives a necessary and sufficient condition such that the poles of $(A - KC)$ can be chosen at will. One may, however, be satisfied with all poles in the left half plane (and the poles are not necessarily at prespecified places). This is of course a weaker requirement for which observability is a sufficient but not a necessary condition. Consider for instance the matrix pair

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,,\quad C = (1\ \ 0)\,,$$

which is not observable. The poles of $A - KC$, with $K = (k_1, k_2)^T$ 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 left half plane and an observer has been constructed which converges to the real state vector for $t \to \infty$. One of the poles, viz. $\lambda = -1$, is fixed and cannot be chosen at will.

A necessary and sufficient condition such that the poles of $(A - KC)$ must be situated in the left half plane is most easily given when A and C are expressed with respect to a particular basis such that they have the form (4.12) and (4.13) respectively;

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}\,,\quad C = (0\ \ C_2)$$

where the pair (C_2, A_{22}) is observable. The condition now is that the matrix A_{11} must be asymptotically stable. The property of the pair (C, A) such that a K can be chosen such that the poles of $(A - KC)$ are in the open left half plane is called **detectability.**

EXERCISE 5.7 *Prove that the linear time-invariant system* $\dot{x} = Ax + Bu$, $y = Cx$, *is detectable if and only if its nonobservable subspace is contained in its stable subspace (compare Exercise 5.2).*

EXERCISE 5.8 *Show that detectability is the dual concept of stabilizability, i.e.* (A, B) *is stabilizable if and only if* (B^T, A^T) *is detectable.*

EXERCISE 5.9 *Consider Exercise 3.3 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.*

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 desired properties. We are now going to combine the feedback concept with that of the observer. If $u = Fx$ is a feedback law that makes

$$\dot{x} = Ax + BFx$$

asymptotically stable, then we would like to have the same for the feedback law $u = F\hat{x}$. Now we have to investigate the behaviour of the original system and the observer together:

$$\begin{aligned} \dot{x} &= Ax + BF\hat{x} \\ \dot{\hat{x}} &= A\hat{x} + BF\hat{x} + K(Cx - C\hat{x}) . \end{aligned}$$

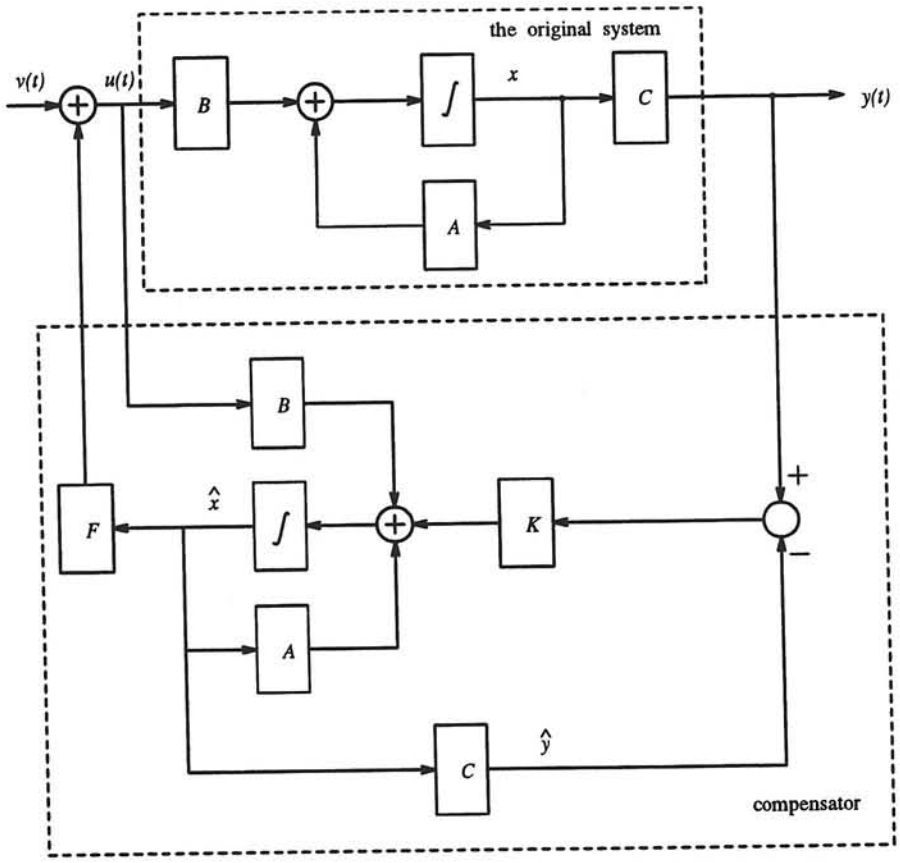
These equations can be simplified somewhat when using $e = x - \hat{x}$;

$$\begin{aligned} \dot{x} &= (A + BF)x - BFe , \\ \dot{e} &= (A - KC)e , \end{aligned} \tag{5.7}$$

equivalently

$$\frac{d}{dt}\begin{pmatrix} x \\ e \end{pmatrix} = \begin{pmatrix} A + BF & -BF \\ 0 & A - KC \end{pmatrix}\begin{pmatrix} x \\ e \end{pmatrix} . \tag{5.8}$$

The eigenvalues of this system are equal to the eigenvalues of $A + BF$ together with those of $A - KC$. Hence the eigenvalues of the overall system are equal to those obtained with state feedback $u = Fx$ and those obtained by constructing the observer. The feedback law $u = Fx$ and the observer can be designed independently! When putting together the original system and observer, with the feedback law $u = Fx$, the eigenvalues do not interfere. This principle is called the **separation principle**. The total system of the original system, observer and feedback-loop is summarized in the following flow-diagram.



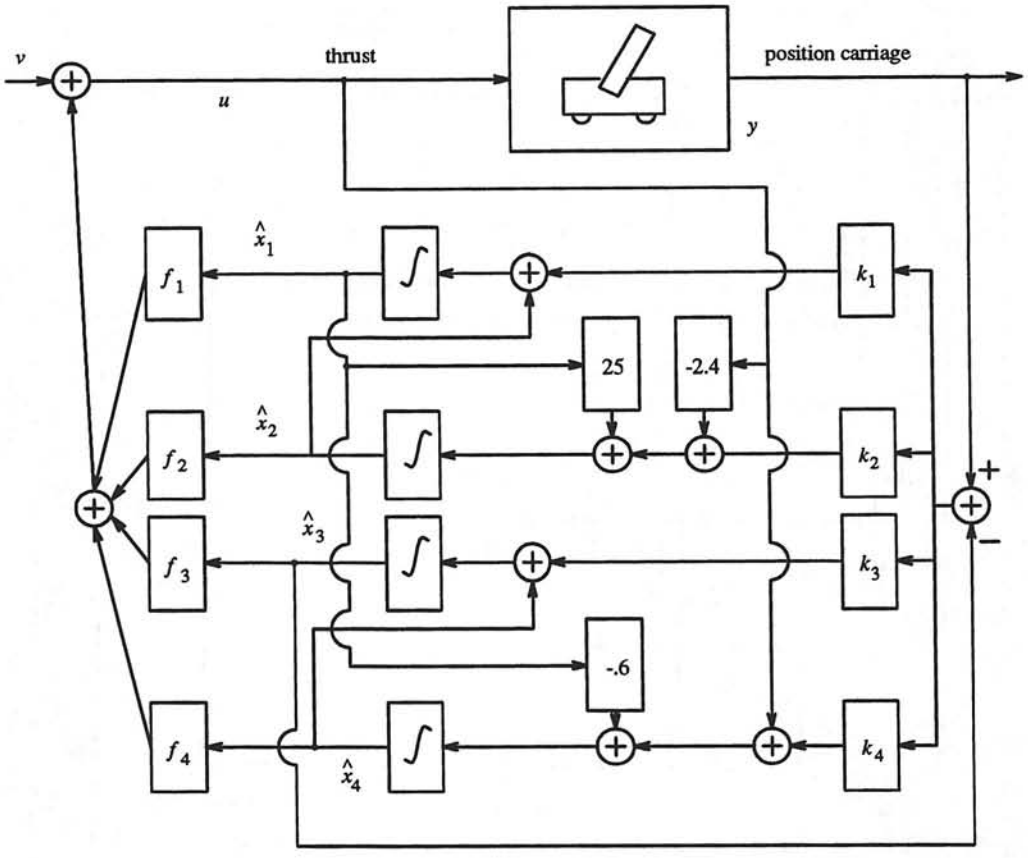
The two subsystems surrounded by a dotted line are the original system and, what is called, the **compensator**. In order to distinguish this compensator from the static compensator introduced earlier, this one is sometimes called a dynamic compensator for obvious reasons.

EXERCISE 5.10 *Show that (5.8) can equivalently be written as*

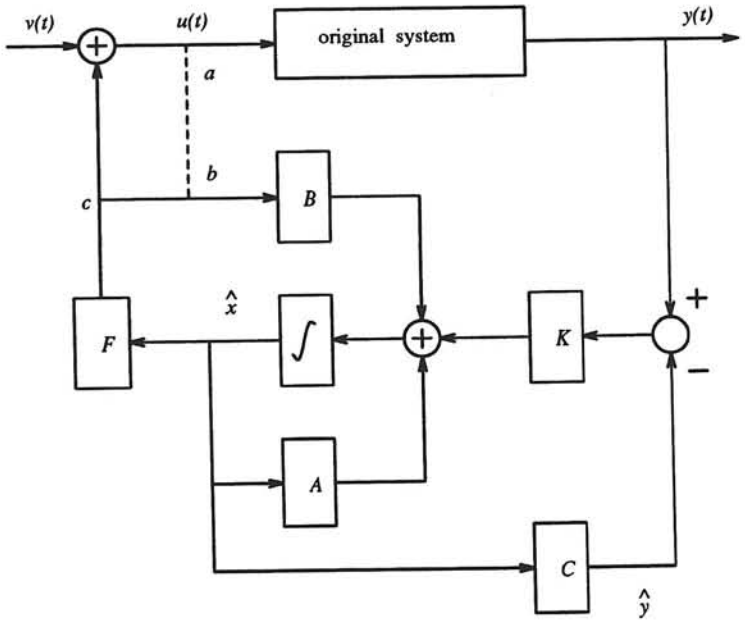
$$\frac{d}{dt}\begin{pmatrix} \hat{x} \\ e \end{pmatrix} = \begin{pmatrix} A+BF & +KC \\ 0 & A-KC \end{pmatrix}\begin{pmatrix} \hat{x} \\ e \end{pmatrix},$$

where the relationship between $\hat{x}$, x and e is as before.

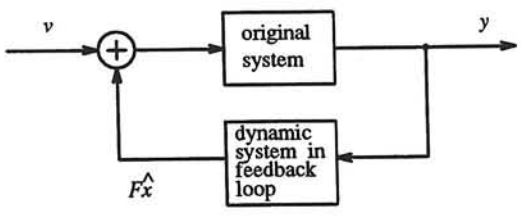
EXAMPLE **5.6** We will now conclude the example of the carriage with the inverted pendulum. A state feedback law was designed in Example 5.2 and an observer in Example 5.5. The combined design is as depicted below. The numerical values for f_i, k_i, $i = 1, 2, 3, 4$, are given in Examples 5.2 and 5.5. □



Sofar we have not considered the new input $v(t)$, which was introduced by defining $u = Fx + v$; Fx is the feedback component, whereas $v(t)$ is a new open-loop input. We slightly change the overal design of the original system and observer in the following sense:



Instead of the connection ab of the previous diagram, we now have the connection bc. The term $v(t)$ does not enter the compensator in this way. This will not change the stability property of the overall system since $v(t)$ is a time-dependent (and not a state dependent) term. The latter diagram can be drawn symbolically as follows:



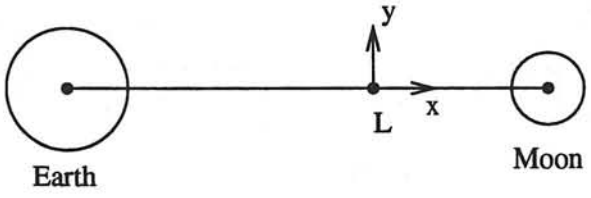
Viewed in this way, stabilization by means of an observer can be interpreted as an output feedback, where now we have a dynamic system in the feedback loop.

EXERCISE 5.11 *On the straight line connecting the earth with the moon a point (in the figure indicated by L) exists where the gravitation force exerted by the earth on a satellite, with mass m, in point L equals (i.e. neutralizes) the gravitation force exerted by the moon and the centrifugal force (due to the rotation of the moon around the earth). The equations of motion of the satellite in the*

neighbourhood of L are

$$\begin{aligned} \ddot{x} - 2\omega\dot{y} - 9\omega^2 &= 0, \\ \ddot{y} + 2\omega\dot{x} + 4\omega^2 &= u, \end{aligned}$$

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,* $\omega = 2\pi/29$ *radians/day.*



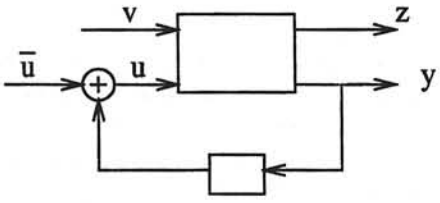
1. *Write the system as a linear dynamical system of order 1 and show that the equilibrium point* $x = \dot{x} = y = \dot{y} = 0$ *is unstable.*
2. *Investigate the controllability and/or stabilizability of this system.*
3. *Determine a linear state feedback such that the poles of the closed-loop system are located in* -3ω, -4ω, $(-3 \pm 3i)\omega$.
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).*

5.3 Disturbance rejection

Consider a linear time-invariant system with $m + l$ inputs partitioned as (u, v) and $p + q$ outputs partitioned as (y, z):

$$\dot{x} = Ax + Bu + Ev, \quad y = C_1 x, \quad z = C_2 x, \tag{5.9}$$

where u is the usual control and v is to be interpreted as a 'disturbance'. For sake of simplicity we assume $C_1 = I$ and hence $y = x$ in this brief section.



The disturbance v cannot be measured directly (one only measures y) and the objective is to design a feedback law

$$u = Fx + \overline{u},$$

(more generally: u is a function of y and of $\overline{u}$) such that v has no effect whatsoever on the output z, no matter what $\overline{u}$ or the initial condition $x(0)$ of (5.9) are.

EXAMPLE **5.7** 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 $u = -x_2 + \overline{u}$, however, one gets

$$z(t) = x_1(0) + \int_0^t \overline{u}(s)ds,$$

which is independent of v. □

In general terms we wish to have that $y(t)$ defined by

$$z(t) = Ce^{t(A+BF)}x(0) + C\int_0^t e^{(t-s)(A+BF)}(B\overline{u}(s) + Ev(s))ds$$

is independent of v for some matrix F. This is equivalent to the requirement that

$$C\int_0^t e^{(t-s)(A+BF)}Ev(s)ds = 0$$

for all functions $v(\cdot)$ and all $t > 0$. This requirement can be shown to be equivalent to the following problem: find a matrix F such that $C(A+BF)^kE = 0$, $\forall k \geq 0$. The latter problem is known as the **disturbance rejection problem.** This equivalence of requirements will not be shown here, but the proof resembles the proof of Lemma 4.2.

EXERCISE 5.12 *Consider the model*

$$\frac{d}{dt}\begin{pmatrix} \phi \\ \dot{\phi} \\ \nu \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -2 & 1 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} \phi \\ \dot{\phi} \\ \nu \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\delta + \begin{pmatrix} 0 \\ 0 \\ f(t) \end{pmatrix},$$

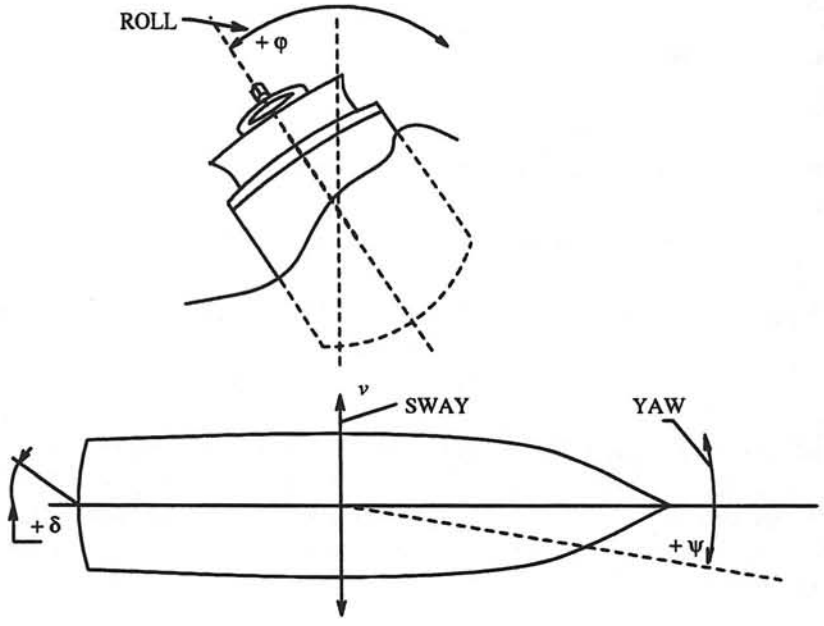
$$z = (1\ 0\ 0)\begin{pmatrix} \phi \\ \dot{\phi} \\ \nu \end{pmatrix},$$

which describes the movement of a ship. The variable ϕ is the roll angle, $\dot{\phi}$ its time derivative and ν is the lateral velocity. The control δ represents the rudder angle and the function $f(t)$ 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 $x^T \stackrel{\text{def}}{=} (\phi, \dot{\phi}, \nu)$, such that ϕ is (completely) independent of the function $f(t)$, 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 $C = (1\ 0\ 0)$.



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 3.3 on impulse response functions or matrices. By means of the function $K(t,s)$ the input function was directly related to the output function. That description was obtained by the elimination of the state vector x. 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

The Laplace transform $\mathcal{L}(f)$ of a piecewise continuous function $f : [0,\infty) \to \mathcal{R}$ is defined as

$$F(s) = \mathcal{L}(f) = \int_0^\infty f(t)e^{-st}dt \,. \tag{6.1}$$

If $f = O(e^{bt})$ for $t \to \infty$, i.e. it grows (at most) at an exponential rate ($b \in \mathcal{R}$ is a constant), then the integral exists for all real $s > b$. If $f = O(e^{bt})$, then the integral also exists for all complex s with $\Re s > b$, because

$$|f(t)e^{-st}| = |f(t)|e^{-(\Re s)t} \,.$$

Therefore the domain of the function $F : (b,\infty) \to \mathcal{R}$ can be extended to all $s \in \mathcal{C}$ with $\Re\, s > b$ and

$$F : \{s \in \mathcal{C} | \Re s > b\} \to \mathcal{C}$$

The function F is complex valued. In this section s wil always denote a complex number. The extension to vector valued functions is straightforward:

$$\mathcal{L}(f) = (\mathcal{L}(f_1), ..., \mathcal{L}(f_n))^T = (F_1(s), ..., F_n(s))^T = F(s) .$$

The extension to matrix valued functions is also componentwise. Consider a linear time-invariant strictly causal differential system given by its impulse response matrix;

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

For simplicity we assume $u(\tau) = 0$ for $\tau \leq 0$ and hence

$$y(t) = \int_{0}^{t} G(t-\tau)u(\tau)d\tau . \tag{6.2}$$

Suppose that the functions $y(\cdot)$, $u(\cdot)$ and $G(\cdot)$ have Laplace transforms, to be denoted by $Y(\cdot)$, $U(\cdot)$ and $H(\cdot)$;

$$Y(s) = \int_{0}^{\infty} y(t)e^{-st}dt; \quad U(s) = \int_{0}^{\infty} u(t)e^{-st}dt; \quad H(s) = \int_{0}^{\infty} G(t)e^{-st}dt$$

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.

If $G(t) = O(e^{bt})$ then the transfermatrix is only defined for $\Re s > b$. The theory of Laplace transforms teaches us that $H(s)$ is analytic for $\Re s > b$ and then complex function theory tells us that a unique analytic continuation of $H(s)$ exists. A unique matrix function exists for all $s \in \mathcal{C}$, analytic in the complex plane except for a number of isolated points, and which is identical to $H(s)$ for $\Re s > b$. In the remainder we will not distinguish $H(s)$ and its analytic continuation.

If $X(s)$ is the Laplace transform of $x(t)$, then

$$\mathcal{L}(\frac{dx}{dt}) = \int_{0}^{\infty} \frac{dx}{dt}e^{-st}dt = x(t)e^{-st}|_{0}^{\infty} + \int_{0}^{\infty} x(t)se^{-st}dt = -x(0) + sX(s) .$$

The Laplace transform of the equation

$$\dot{x} = Ax + Bu , \quad x(0) = x_0,$$

therefore is

$$sX(s) - x_0 = AX(s) + BU(s) \,, \tag{6.4}$$

equivalently

$$X(s) = (sI - A)^{-1}x_0 + (sI - A)^{-1}BU(s) \,.$$

If we also transform the output equation $y = Cx$ to $Y(s) = CX(s)$ and assume that $x_0 = 0$, then

$$Y(s) = C(sI - A)^{-1}BU(s) = H(s)U(s) \,. \tag{6.5}$$

Comparison with

$$y(t) = \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau$$

leads to

$$H(s) = \mathcal{L}(Ce^{At}B) = C(Is - A)^{-1}B \,. \tag{6.6}$$

At this latter equation the theory of analytic continuation can be illustrated clearly. At first instance $H(s)$ is only defined for $\Re s > \max(\Re\lambda_i)$, where λ_i are the eigenvalues of A. The expression $C(Is - A)^{-1}B$, however, is well defined for all $s \in \mathcal{C}$, except for possibly the points $s = \lambda_1, ..., \lambda_n$, 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 $H(s)$. Eq. (6.6) states that the transfer matrix is the Laplace transform of the impulse response matrix.

EXAMPLE **6.1** Given is the system which describes the dynamics of the satellite (see also Examples 3.1 and 3.7);

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

The transfer matrix for this system is

$$H(s) = \mathcal{L}(G(t)) = \mathcal{L}(\begin{pmatrix} \sin t & 2 - 2\cos t \\ -2 + 2\cos t & -3t + 4\sin t \end{pmatrix}) =$$

$$\begin{pmatrix} \frac{1}{s^2+1} & \frac{2}{s} - \frac{2s}{s^2+1} \\ -\frac{2}{s} + \frac{2s}{s^2+1} & -\frac{3}{s^2} + \frac{4}{s^2+1} \end{pmatrix} = \begin{pmatrix} \frac{1}{s^2+1} & \frac{2}{s^3+s} \\ \frac{-2}{s^3+s} & \frac{s^2-3}{s^4+s^2} \end{pmatrix} \,.$$

□

EXERCISE 6.1 *Consider the dynamics of the inverted pendulum as given in eq. (3.15) and assume that only the position of the carriage is measured;*

$$y = (1\ 0\ 0\ 0)x\ .$$

In Example 3.5 the transition matrix was calculated for this problem. Show that the impulse response function and transfer function are given by, respectively,

$$G(t) = -0.48\sinh(5t)\ ;\quad H(s) = \frac{-2.4}{s^2 - 25}\ .$$

A new method has now been found to calculate the transition matrix. The Laplace transforms of $\dot{x} = Ax$ $(x(0) = x_0)$ and $x(t) = e^{At}x_0$ are

$$X(s) = (sI - A)^{-1}x_0\ ,\quad X(s) = \mathcal{L}(e^{At})x_0\ .$$

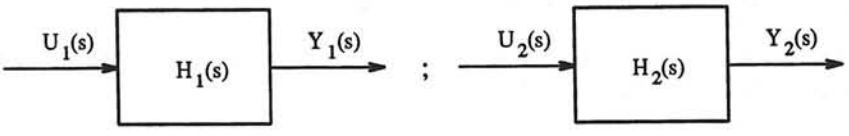
Therefore it follows

$$e^{At} = \mathcal{L}^{-1}((sI - A)^{-1})\ ,$$

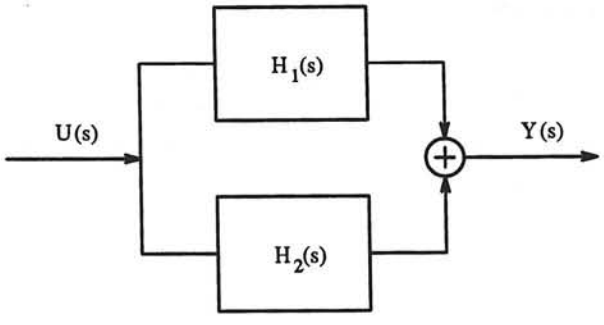
where $\mathcal{L}^{-1}$ is the inverse Laplace transform. The matrix function $(sI - A)^{-1}$ is called the **resolvente** of the matrix A.

6.1.1 Connection of systems

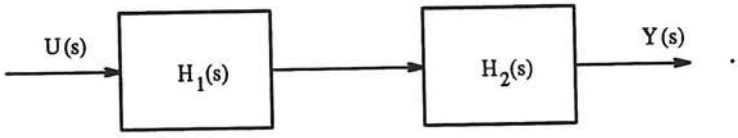
The description of systems by means of transfer matrices is useful if one wants to connect systems. If we are given two systems by means of the transfer matrices $H_1(s)$ and $H_2(s)$ respectively;



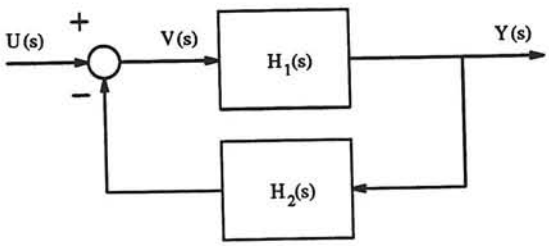
then the **parallel connection** is given by



where $\oplus$ denotes addition. In formula: $H(s)$, the transfer matrix of the parallel connection, equals $H_1(s) + H_2(s)$. The **series connection** is given by



In formula: $H(s) = H_2(s)H_1(s)$. Please note that for multi-input multi-output systems ($p, m > 1$) the product of the two matrix functions $H_2(\cdot)$ and $H_1(\cdot)$ is in general not commutative; $H_1H_2 \neq H_2H_1$, 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, where the starting point is two state space descriptions, is far more difficult. The **feedback connection** is given by



If the signal that enters $H_1(s)$ is called $V(s)$, then the transfer matrix $H(s)$ of the overall system can be calculated as follows.

$$\left.\begin{array}{rcl} V(s) & = & U(s) - H_2(s)Y(s) \\ Y(s) & = & H_1(s)V(s) \end{array}\right\} \quad Y(s) = H_1(s)(U(s) - H_2(s)Y(s)) \; .$$

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 connection considered above it was tacitly assumed that the number of inputs (m) and the number of outputs (p) were such that the connections described made sense.

6.1.2 Rational functions

Let us consider the transfer matrix $H(s) = \mathcal{L}(G(t)) = C(Is - A)^{-1}B$ 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

$$p(s) = s^n + p_{n-1}s^{n-1} + ... + p_1 s + p_0 \ , \quad p_i \in \mathcal{R} \ .$$

For all i, j, the terms $q_{ij}(s)$ are determinants of an $(n-1) \times (n-1)$ submatrix of $(Is - A)$ and are therefore polynomials in s, of at most degree $n-1$. Therefore the elements of $(Is - A)^{-1}$ are rational functions of s; $q_{ij}(s)/p(s)$. A **rational function** is 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 strictly proper is $\lim_{|s|\to\infty} h(s) = 0$. If this limit is finite, but not necessarily zero, then one speaks of a **proper** rational function. Written as a qoutient of two polynomials, a rational function is proper if and only if the degree of the numerator polynomial is not larger than the degree of the denominator polynomial.

It easily follows that the elements of $H(s)$ are strictly proper rational functions; $H(s)$ can be written as $R(s)/p(s)$, where $R(s)$ is an $m \times p$ matrix with polynomials as elements (of degree $< n$) and the degree of the polynomial $p(s)$ 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 are the only candidates for poles (but not necessarily all of them are poles).

EXAMPLE **6.2** If $A = \begin{pmatrix} 0 & -2 \\ 1 & -3 \end{pmatrix}$, $B = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $C = (0 \ \ 1)$, 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, viz. in $s = -1$. □

EXAMPLE **6.3** We are given two linear, asymptotically 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. Suppose $h_i(s) = p_i(s)/q_i(s)$ and that common factors have been deleted, then the poles of system Σ_i are the zeros of $q_i(s) = 0$. Asymptotic stability means that these zeros are located in the left-half plane. The transfer function of the series connection equals $p_1(s)p_2(s)/(q_1(s)q_2(s))$. Perhaps some terms in numerator and denominator will cancel, but the remaining zeros of the denominator are a (sub)set of the zeros of $q_1(s) = 0$ and $q_2(s) = 0$ and therefore lie in the 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

$$h_1(s) = \frac{1}{s+1}, \quad h_2(s) = \frac{-2}{s+1}.$$

The transfer function of the feedback connection is

$$h(s) = \frac{h_1(s)}{1 + h_1(s)h_2(s)} = \frac{(s+1)}{(s+1-\sqrt{2})(s+1+\sqrt{2})},$$

which represents an unstable system.
Answer question 3. The parallel connection is stable again. The proof is 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 $D \neq 0$, then, see Exercise 3.10, and assuming $x_0 = 0$,

$$\begin{aligned} y(t) &= \int_0^t (Ce^{A(t-s)}B + D\delta(t-s))u(s)ds, \\ H(s) &= C(sI - A)^{-1}B + L\{D\delta(t-s)\} = C(sI - A)^{-1}B + D\,. \end{aligned}$$

If we consider this transfer function in detail, it turns out that the elements are proper functions (and not strictly proper) since the degree of the numerator will now in general be equal to the degree of the denominator. 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 3.8 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 this will be proved for transfer functions, to which single-input, single-output systems correspond.

6.2 Transfer functions and 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 is also used for the multi-input multi-output case. In this section we will also assume that $h(s)$ is proper, i.e. degree (numerator) $\leq$ degree (denominator). Without loss of generality $h(s)$ will be written more explicitly as

$$h(s) = \frac{q(s)}{p(s)} = \frac{q_k s^k + q_{k-1}s^{k-1} + ... + q_0}{s^n + p_{n-1}s^{n-1} + ... + p_0}, \tag{6.8}$$

with $k \leq n$ and where the coefficient of the leading term s^n in the denominator equals one. (Polynomials with the coefficient of the leading term equal to one are called **monic** polynomials.) It is well known that a polynomial can be factorized in a number of linear terms equal to the degree of the polynomial. Hence, we can write

$$h(s) = \frac{q(s)}{p(s)} = \frac{c(s-b_1)(s-b_2)...(s-b_k)}{(s-a_1)(s-a_2)...(s-a_n)}, \tag{6.9}$$

with $a_i,\ b_i \in \mathcal{C},\ c \in \mathcal{R}$ and $k \leq n$. We will assume that $q(s)$ and $p(s)$ do not have any common factors. If so, they will be cancelled. The zeros of the denominator, $a_1, ..., a_n$, are called the **poles** of the transfer function and $b_1, ..., b_k$ are the **zeros** of the transfer function. The reason for this terminology is the following. Suppose the input is given by

$$u(t) = \begin{cases} e^{s_0 t} & , \ t \geq 0, \\ 0 & , \ t < 0, \end{cases}$$

then the Laplace transform of the output can be written as

$$Y(s) = \frac{c(s-b_1)...(s-b_k)}{(s-a_1)...(s-a_n)} \cdot \frac{1}{s-s_0} .$$

If $s_0 \neq b_i$, $i = 1, ..., k$, then this expression can be factorized as

$$Y(s) = \frac{A_1}{s-a_1} + \frac{A_2}{s-a_2} + ... + \frac{A_n}{s-a_n} + \frac{A_{n+1}}{s-s_0} , \quad A_i \in \mathcal{C} , \tag{6.10}$$

where, for reason of simplicity, we assumed that all poles a_i have multiplicity one, and moreover, $s_0 \neq a_i$, $i = 1, ..., n$. The inverse transformation of (6.10) yields

$$y(t) = A_1 e^{a_1 t} + ... + A_n e^{a_n t} + A_{n+1} e^{s_0 t} .$$

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

$$\begin{aligned} Y(s) &= \frac{c(s-b_1)....(s-b_k)}{(s-a_1)....(s-a_n)} \frac{1}{(s-b_1)} = \frac{c(s-b_2)....(s-b_k)}{(s-a_1)....(s-a_n)} = \\ &= \frac{A_1}{s-a_1} + ... + \frac{A_n}{s-a_n} , \quad A_i \in \mathcal{C} . \end{aligned}$$

The frequency s_0 of the input signal does not show up in the output signal; only the free modes are excited. *The zeros of a system are those frequencies which do not form part of the output signal.*

EXERCISE 6.2 *Design a system of the form $\dot{x} = Ax + Bu$, $y = Cx$, and find a suitably chosen initial condition, 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 just treated that a possible transfer function is*

$$h(s) = \frac{s+3}{(s+1)(s+2)} .$$

DEFINITION 6.1 *If all eigenvalues λ_i have a negative real part, the* **time constant** *σ of the corresponding system is defined as $\sigma^{-1} = \min_i \{\Re\lambda_i\}$.*

DEFINITION 6.2 *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 **6.5** 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 $y(\infty) = 1/6 > 0$; a positive input leads to a positive output in the long run. For a stabilizing output feedback it is therefore tempting to think of $u(t) = ky(t)$ with $k < 0$. 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 **6.6** Continuation of the satellite example (see Examples 3.1, 3.7 and 6.1). We consider a version of the dynamics where there is only one input variable (u_2) and one output variable (y_2). The matrices involved are (with $\omega = 1$):

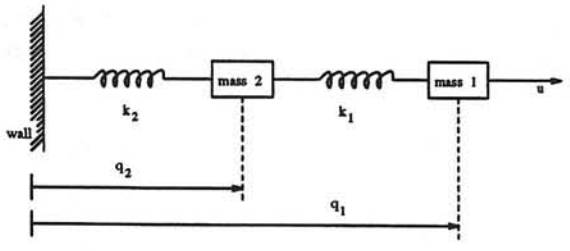
$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad C = (0 \;\; 0 \;\; 1 \;\; 0).$$

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 here; $s = \pm\sqrt{3}$ correspond to exponential functions) cannot form a component of the output signal. However, because the system is not stable, the free modes excited by the input will not die out. □

EXERCISE 6.3 *Two unit masses are connected by springs, characterized by spring constants k_1 and k_2 respectively; see the figure:*



The position of the masses are indicated by 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; $y = q_1$. 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 $\pm i\sqrt{(k_1 + k_2)}$; they correspond to 'real' frequencies!

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)

$$h(s) = C(Is - A)^{-1}B + D \, . \tag{6.11}$$

In the following theorem the reverse will be shown.

THEOREM 6.1 *To each transfer function (which is a proper rational function) corresponds an $n \times n$ matrix A, $n \times 1$ matrix B, $1 \times n$ matrix C and a 1×1 matrix D such that (6.11) will hold.*

Proof We are given a rational function $h(s) = q(s)/p(s)$ with deg $(q(s)) \leq$ deg $(p(s))$. First $D \in \mathcal{R}$ 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

$$\begin{aligned} h(s) &= \frac{q(s)}{p(s)} = \frac{q_n s^n + q_{n-1}s^{n-1} + ... + q_0}{s^n + p_{n-1}s^{n-1} + ... + p_0} = \\ &= \frac{q_n(s^n + p_{n-1}s^{n-1} + ... + p_0)}{p(s)} \\ &+ \frac{(q_{n-1} - q_n p_{n-1})s^{n-1} + ... + (q_0 - q_n p_0)}{p(s)} = q_n + \frac{\tilde{q}(s)}{p(s)} \, , \end{aligned}$$

where deg $(\tilde{q}(s)) <$ deg$(p(s))$. We take $D = q_n$ in this case.

In order not to complicate the notation $\tilde{q}(s)$ will again be written as $q(s)$, such that we can continue with $q(s)/p(s)$, deg$(q(s)) <$ deg$(p(s))$, and with D already defined. Hence we can write

$$p(s) = s^n + p_{n-1}s^{n-1} + ... + p_0 \; ; \; q(s) = q_{n-1}s^{n-1} + ... + q_0 \, .$$

If $Y(s)$ and $U(s)$ are the Laplace transforms of y and u, then they are connected according to $Y(s) = h(s)U(s)$, equivalently,

$$p(s)Y(s) = q(s)U(s) \, ,$$

or

$$s^n Y(s) + p_{n-1}s^{n-1}Y(s) + ... + p_0 Y(s) = q_{n-1}s^{n-1}U(s) + ... + q_0 U(s) \ . \quad (6.12)$$

We start with a special $q(s)$ polynomial, viz. $q(s)$ is a constant, and for this constant we choose 1; $q(s) = q_0 = 1$. Since such a system is different from the original one, we will call the output $z(t)$, with Laplace transform $Z(s)$, instead of $y(t)$ which is preserved for the output of the original system. Then

$$s^n Z(s) + p_{n-1}s^{n-1}Z(s) + ... + p_0 Z(s) = U(s) \ ,$$

which is the Laplace transform of

$$\frac{d^n}{dt^n}z(t) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}z(t) + ... + p_0 z(t) = u(t), \quad (6.13)$$

with initial values $z(0) = \dot{z}(0) = ... = z^{(n-1)}(0) = 0$. Here we used the following properties of Laplace transforms of derivatives:

$$\begin{aligned} \mathcal{L}[f^{'}(t)] &= s\mathcal{L}[f] - f(0), \\ \mathcal{L}[f^{''}(t)] &= s^2\mathcal{L}[f] - sf(0) - f^{'}(0), \\ &\vdots \end{aligned}$$

Eq. (6.13) can be written as a set of first order differential equations:

$$\frac{d}{dt}\begin{pmatrix} z(t) \\ \dot{z}(t) \\ \vdots \\ \vdots \\ z^{(n-1)}(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & -p_{n-2} & -p_{n-1} \end{pmatrix} \begin{pmatrix} z(t) \\ \dot{z}(t) \\ \vdots \\ \vdots \\ z^{(n-1)}(t) \end{pmatrix}$$

$$+ \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} u(t).$$

Thus a linear differential system $\dot{x} = Ax + Bu$, $z = Cx$ with as state $x = (z, \dot{z}, ..., z^{(n-1)})$ has been obtained with

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & -p_{n-2} & -p_{n-1} \end{pmatrix} , B = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} , C = (1 \ 0 \ ... \ 0) \ . \quad (6.14)$$

This is a realization of the transfer function $h(s) = \frac{1}{p(s)}$. Note that the eigenvalues of A are the poles of $h(s)$ because $\det(Is - A) = p(s)$. See Exercise 3.8.

We now consider the general case with an arbitrary numerator polynomial $q(s)$ of degree $< n$. Inverse Laplace transformation of (6.12) yields (with the initial values of all derivatives of u and y equal to zero)

$$\frac{d^n}{dt^n}y(t) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}y(t) + ... + p_0 y(t) = q_{n-1}\frac{d^{n-1}}{dt^{n-1}}u(t) + ... + q_0 u(t) \ . \quad (6.15)$$

The solution $z(t)$ of (6.13) will be related to the solution $y(t)$ of (6.15). To that end, because $z(t)$ satisfies (6.13), $q_0 z(t)$ satisfies

$$\frac{d^n}{dt^n}(q_0 z(t)) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}(q_0 z(t)) + ... + p_0(q_0 z(t)) = q_0 u(t) \ . \quad (6.16)$$

Differentiation of (6.13) and subsequent multiplication by q_1 leads to

$$\frac{q^n}{dt^n}(q_1 \dot{z}(t)) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}(q_1 \dot{z}(t)) + ... + p_0(q_1 \dot{z}(t)) = q_1 \dot{u}(t) \ . \quad (6.17)$$

Continuing this way, we get ultimately

$$\frac{d^n}{dt^n}(q_i z^{(i)}(t) + p_{n-1}\frac{d^{n-1}}{dt^{n-1}}(q_i z^{(i-1)}(t)) + ... + (p_0 q_i z^{(i)}(t)) = q_i u^{(i)}(t),$$

for $i = 0, ..., n-1$. If we add all these n equations, the result is

$$\begin{aligned} &\frac{d^n}{dt^n}(q_0 z + q_1 \dot{z} + ... + q_{n-1} z^{(n-1)}) + ... + \\ &+p_0(q_0 z + q_1 \dot{z} ... + q_{n-1} z^{(n-1)}) = q_0 u + q_1 \dot{u} + ... + q_{n-1} u^{(n-1)} \ . \quad (6.18) \end{aligned}$$

If (6.15) and (6.18) are compared, we see that the unique solution $y(t)$ of (6.15), with $y(0) = \dot{y}(0) = ... = y^{(n-1)}(0) = 0$, equals $q_0 z + q_1 \dot{z} + ... + q_{n-1} z^{(n-1)}$. A realization of $h(s) = q(s)/p(s)$ therefore is, with state variable $x = (z, \dot{z}, ..., z^{(n-1)})^T$,

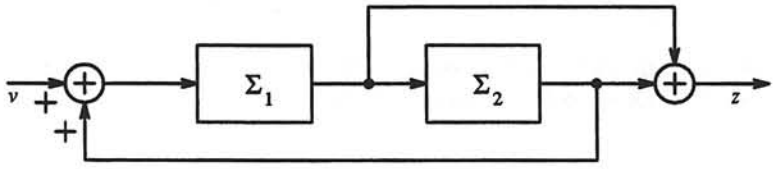
$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & \dots & -p_{n-1} \end{pmatrix} , \ B = \begin{pmatrix} 0 \\ \vdots \\ \\ 0 \\ 1 \end{pmatrix} , \ C = (q_0, ..., q_{n-1}) \ . \quad (6.19)$$

Other realizations exist (other triples of matrices A, B, C are possible which correspond to the same transfer function. Indeed, as explained in section 3.3, a coordinate transformation in the state space does not change the transfer function). □

EXAMPLE **6.7** Consider the two systems ($x,\ u,\ y \in \mathcal{R}$):

$$\begin{array}{lllllll} \Sigma_1: & \dot{x} & = & -x+2u; & y & = & 2x, \\ \Sigma_2: & \dot{x} & = & -2x+3u; & y & = & -x. \end{array}$$

What are the transfer functions of these two systems? Subsequently, these systems are coupled to each other as indicated in the next figure.



The input and output of this combined system are called ν and z respectively. What is the tranfer function which describes the relation between ν and z? Give also a state space description of this combined system.

Answer. The transfer functions of Σ_1 and Σ_2 are respectively

$$h_1(s) = \frac{4}{s+1}, \quad h_2(s) = \frac{-3}{s+2}.$$

In order to determine the transfer function of the coupled system, we define x_i to be the output of system Σ_i. Then we formally get

$$X_2 = h_2 X_1, \quad X_1 = h_1(\nu + X_2), \quad Z = X_1 + X_2,$$

from which

$$Z = \frac{h_1(1+h_2)}{1-h_1h_2}\nu.$$

Substitution of h_1 and h_2 leads to the transfer function

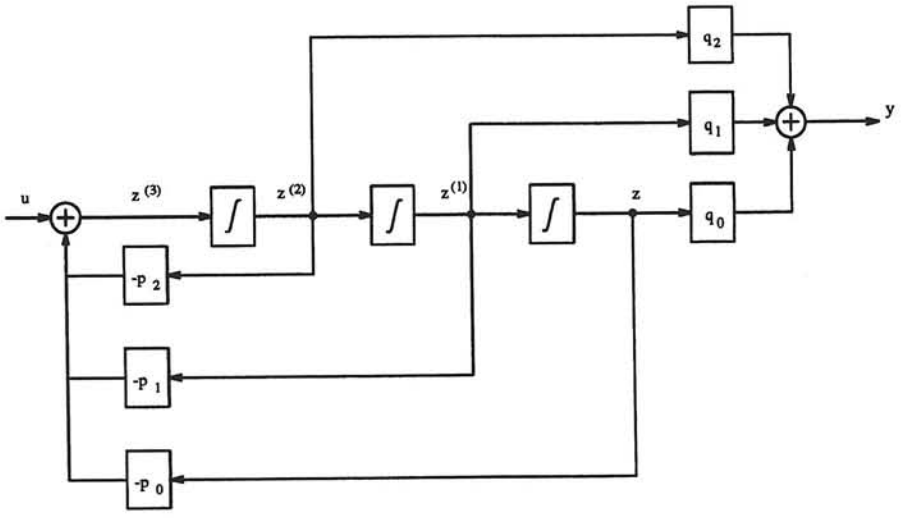
$$\frac{4s-4}{s^2+3s+14}.$$

A state space description is

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ -14 & -3 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \nu; \quad z = (-4 \ \ 4)x.$$

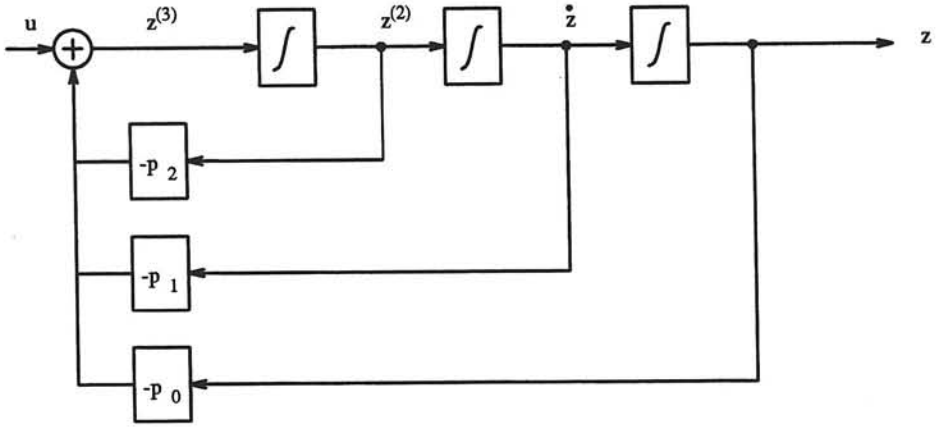
□

The realization in (6.19) has a special name, viz., **standard controllable realization** or the **controllability canonical form** and was already met in (5.4). The procedure given above to obtain a realization can also be given by means of a flow diagram (in this diagram $n = 3$ and the notation $z^{(i)}$ refers to the ith derivative of z).

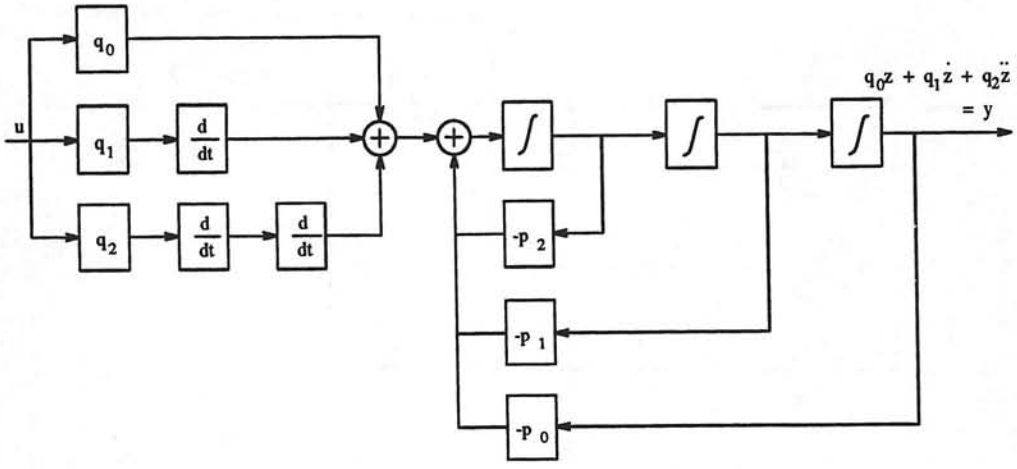


In this diagram the box $\boxed{\int}$ denotes integration which is a shorthand notation of the system $\{\dot{x} = u \ ; \ y = x\}$ with transfer function $1/s$ and the boxes $\boxed{-p_i}$ and $\boxed{q_i}$ denote multiplication with the coefficient inside the box. The diagram also indicates how the 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 an **analog computer** is being used for.

Superficially, we could also implement or build this system by means of **differentiators**. This design or flow diagram could be obtained as follows. The flow diagram between u and z is (take $n = 3$):



and therefore, by superposition,



The latter flow diagram also describes the system characteristized by $h(s) = q(s)/p(s)$. However, now differentiators, the blocks $\boxed{\frac{d}{dt}}$, have been used. As will be explained in Example 6.13, 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.

EXAMPLE **6.8** In Example 6.6 a form of the satellite problem was discussed with transfer function

$$\frac{s^2 - 3}{s^4 + s^2} .$$

A realization of this function is

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} u \; ; \; y = (-3, \; 0, \; 1, \; 0)x \; .$$

Clearly, this realization is different from the one given in Example 6.6. □

EXERCISE 6.4 *Show that the controllability canonical form of the system with the two connected springs in Exercise 6.3 equals*

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -k_1k_2 & 0 & -(2k_1 + k_2) & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} u \; ;$$

$$y = (k_1 + k_2,\ 0,\ 1,\ 0)x\ .$$

Another particular realization carries the name of **observer canonical form.** We will not discuss it here extensively. We only give it here, for sake of completeness, for the transfer function (6.8) with $q_n = 0$;

$$A = \begin{pmatrix} -p_{n-1} & 1 & 0 & \dots & 0 \\ -p_{n-2} & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & 0 & \dots & 0 & 1 \\ -p_0 & 0 & \dots & \dots & 0 \end{pmatrix}, \quad B = \begin{pmatrix} q_{n-1} \\ q_{n-2} \\ \vdots \\ \vdots \\ q_0 \end{pmatrix}, \quad C = (1, 0, \dots, 0)\ . \tag{6.20}$$

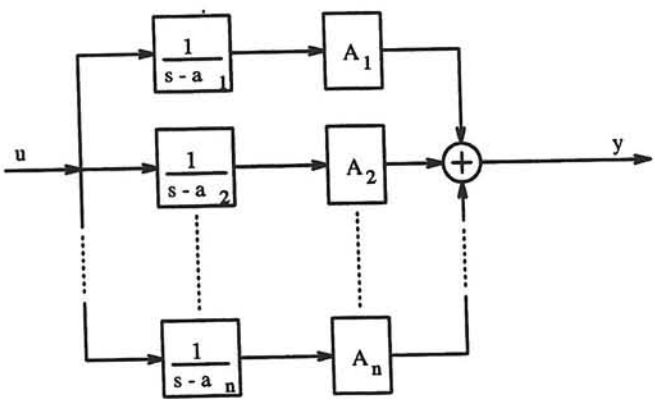
We will conclude this section with one other method which also realizes a rational function $h(s) = q(s)/p(s)$ with deg $(q(s))$ < deg $(p(s))$, as a linear differential system. The method is based on the factorization of $h(s)$;

$$h(s) = \frac{q(s)}{p(s)} = \frac{A_1}{s - a_1} + \frac{A_2}{s - a_2} \dots + \frac{A_n}{s - a_n}\ ,$$

where a_i are the poles of $h(s)$, which are assumed to be real and to have multiplicity one for the time being. A realization of $h(s)$ is given by $x \in \mathcal{R}^n$:

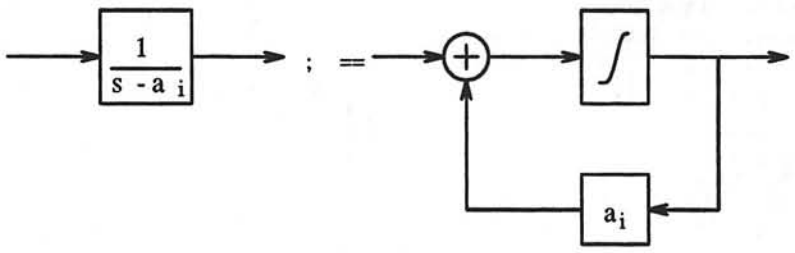
$$\dot{x} = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_n \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} u\ ; \quad y = (A_1, A_2, \dots, A_n)x,$$

which looks in a block diagram as follows:



This realization is called the **diagonal realization.** The original n-th order system has been **decoupled** into n independent subsystems. The blocks with the

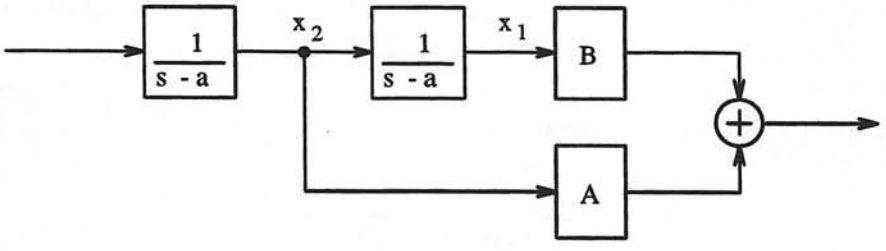
contents $1/(s-a_i)$ are shorthand graphs for the following blocks:



If $p(s) = 0$ has *real* roots of multiplicity larger than one, say $s = a$ has multiplicity two, then factorization leads to

$$h(s) = \frac{A}{s-a} + \frac{B}{(s-a)^2} + \dots .$$

These terms can be realized jointly as follows:



If the outputs of the two blocks with integrators are denoted by x_2 and x_1, as indicated in the figure, then a state space realization of $\frac{A}{s-a} + \frac{B}{(s-a)^2}$ is

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u\,; \quad y = (B \;\; A)x\,.$$

The system matrix is now a Jordan block of size 2×2.

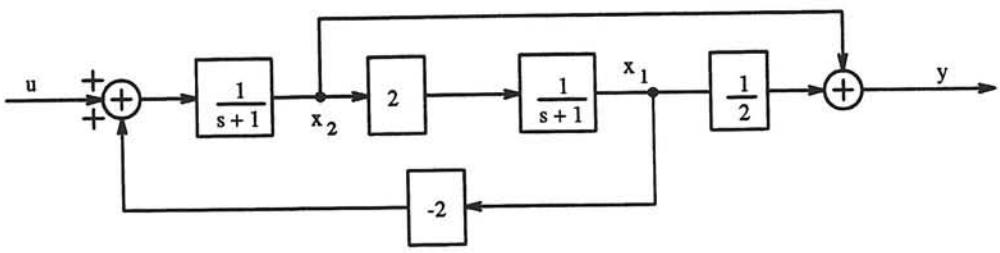
If in $p(s)$ factors exist of the form $s^2 + bs + c$ with $b^2 - 4c < 0$ such that a further decomposition in real factors is impossible, then the following example shows a possible flow diagram. Suppose the transfer function is given by

$$h(s) = \frac{s+2}{s^2+2s+5}\,.$$

The denominator cannot be decomposed any further into real factors. Therefore $h(s)$ is written as

$$h(s) = \frac{(s+1)+1}{(s+1)^2+4} = \frac{\frac{1}{s+1}}{1+\frac{2^2}{(s+1)^2}} + \frac{1}{2}\frac{\frac{2}{(s+1)^2}}{1+\frac{2^2}{(s+1)^2}},$$

and a flow diagram is



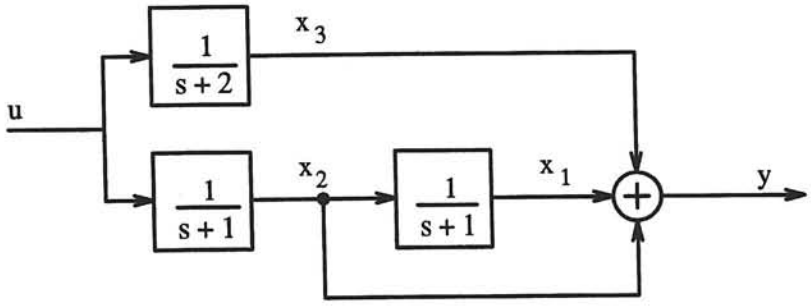
If the output of the two blocks with integrators are denoted by x_2 and x_1, as indicated in the figure, then a state space realization is

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ -2 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u\,; \quad y = \begin{pmatrix} \frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

EXERCISE 6.5 *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 factorize this function. Show that this factorization can be given in a flow diagram as follows:



If the output of the local subsystems are called x_1, x_2 and x_3 as indicated, give a

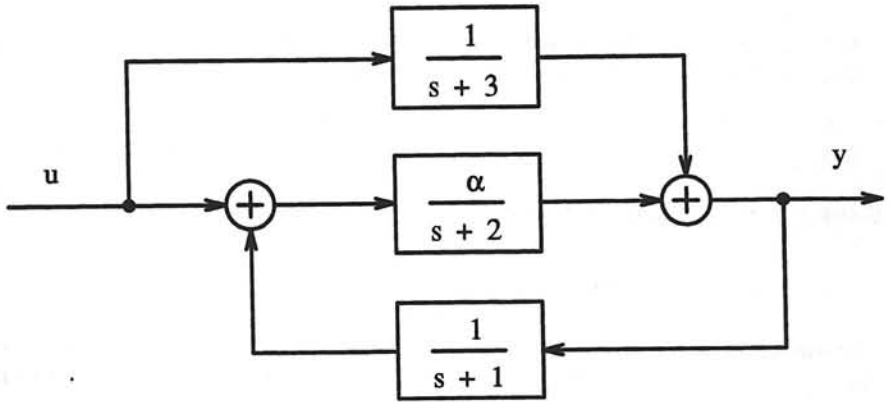
description in state space form with the vector x as state. Prove that

$$\frac{d\tilde{x}}{dt} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -5 & -4 \end{pmatrix} \tilde{x} + \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix} u\,, \quad y = (1\ 0\ 0)\tilde{x}$$

is another state space description of the same system. Show that a nonsingular 3×3 matrix T exists such that with the transformation $x = T\tilde{x}$ one state space description can be obtained from the other. Matrix T can be interpreted as a basis transformation in state space.

EXERCISE 6.6 *If $\{A, B, C\}$, 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 $q(s)/p(s)$, prove that degree $q(s) = m$ if and only if $CA^iB = 0$, $i = 0, 1, ..., n - m - 2$ and $CA^{n-m-1}B \neq 0$.*

EXERCISE 6.7 *Given the flow diagram below,*



determine the transfer function of the overall system. For which value(s) of α is the system stable?

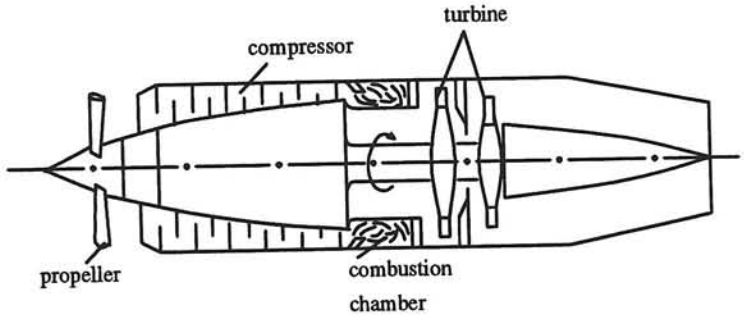
EXAMPLE **6.9** We are given the system

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

This is a model of a turbopropeller engine, 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 nomial value;
- x_3 is the deviation of the fuel rate from its normal value;
- u_1 is the deviation of the propeller blade angle from the normal point;
- u_2 is the time-derivative of the fuel rate.



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 $w \in \mathcal{R}^2$, where $u \in \mathcal{R}^2$, $x \in \mathcal{R}^3$, w and an auxiliary variable $v \in \mathcal{R}^2$ 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).
Answer question 3. With the new input (and output), the system equations can be written as
$$\begin{aligned} \dot{x} &= (A + BGF)x + BGw, \\ y &= Cx. \end{aligned}$$

Write

$$G = \begin{pmatrix} g_1 & g_2 \\ g_3 & g_4 \end{pmatrix}, \quad F = \begin{pmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \end{pmatrix}.$$

Since $y_1 = x_1$ must not depend on w_2, a requirement is $g_2 = 0$. 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, take for instance $f_1 = f_5 = 0$. 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. □

EXERCISE 6.8 *Determine a realization of the transfer matrix*

$$H(s) = \begin{pmatrix} \frac{s}{s^2+2s} & 0 & 0 \\ 0 & \frac{s^2-1}{(s+3)(s+2)} & 0 \\ 0 & 0 & \frac{s-1}{(s+1)^3} \end{pmatrix}$$

DEFINITION 6.3 *If $H(s)$ is a matrix of which the entries are strictly proper rational functions, then we will say that n is the* **McMillan degree** *of $H(s)$ if $H(s)$ has a realization $C(Is - A)^{-1}B$ with A having size $n \times n$ and no realizations with A of dimension k with $k < n$ exist.*

The following theorem, of wich the proof will not be given, provides an algorithm for calculating the McMillan degree.

THEOREM 6.2 *Given any $p \times m$ matrix $H(s)$ of which the entries are strictly proper rational functions, and given that $H(s)$ has the expansion*

$$H(s) = L_0 s^{-1} + L_1 s^{-2} + L_2 s^{-3} + \cdots,$$

then the McMillan degree n of $H(s)$ is given by $n = \text{rank} L(r-1, r-1)$, where

$$L(\alpha, \beta) \stackrel{\text{def}}{=} \begin{pmatrix} L_0 & L_1 & L_2 & \cdots & L_\beta \\ L_1 & L_2 & L_3 & \cdots & L_{\beta+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ L_\alpha & L_{\alpha+1} & L_{\alpha+2} & \cdots & L_{\alpha+\beta+1} \end{pmatrix},$$

with r being the degree of the least common multiple of the denominators of $H(s)$.

6.3 Transfer functions and minimal realizations

THEOREM 6.3 *A state space realization of a transfer function $h(s) = \frac{q(s)}{p(s)}$ (single-input single-output system) is both controllable and observable if and only if $q(s)$ 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 observabibility, 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.

Part 1. Consider the diagonal realization

$$A = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & \lambda_n \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_n \end{pmatrix}, \quad C = (c_1, c_2, ..., c_n), \tag{6.21}$$

with corresponding transfer function

$$h(s) = \frac{q(s)}{p(s)} = \sum_{i=1}^{n} \frac{g_i}{s - \lambda_i}. \tag{6.22}$$

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

$$\det\ R = \det \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \lambda_n^2 & & \lambda_n^{n-1} \end{pmatrix} \times \prod_{i=1}^{n} b_i\ . \tag{6.23}$$

The determinant at the right-hand side of (6.23) 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

$$\prod_{1 \le i < j \le n} (\lambda_j - \lambda_i)\ .$$

Hence $\det\ R \neq 0$ if and only if $\lambda_i \neq \lambda_j$ for all $i \neq j$ 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. Realization (6.21) is controllable if and only if $\lambda_i \neq \lambda_j$ and $b_i \neq 0$. With the same argument it can be shown that the realization is observable if and only if $\lambda_i \neq \lambda_j$ and $c_i \neq 0$. For a controllable and observable realization of the form (6.21), $c_i \neq 0$ and $b_i \neq 0$ and therefore $g_i \neq 0$. This means that there are no common factors in $h(s)$.

Part 2. Now we suppose that $p(s)$ and $q(s)$ have a common factor, say $(s - \lambda_1)$. Therefore we must have $g_1 = 0$ in (6.22), hence

$$h(s) = \sum_{i=1}^{n} \frac{g_i}{s - \lambda_i} = \sum_{i=2}^{n} \frac{g_i}{s - \lambda_i} = h_{\text{red}}(s)\ ,$$

where red stands for reduced. Since $g_1 = 0$, we have $b_1 = 0$ or $c_1 = 0$ (or both). It is easily seen that $b_1 = 0$ refers to non-controllability and $c_1 = 0$ refers to non-observability, which then proves the theorem. □

EXAMPLE **6.10** Consider the system

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & -4 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} u;\ y = (1\ -1\ 1)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{6.24}$$

The R and W matrices 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 has a common factor. The transfer function is

$$h(s) = \frac{(s-1)(s-2)}{(s-1)(s-2)^2} = \frac{1}{s-2}.$$

Hence the input-output behaviour of the system given by the realization (6.24) 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.** □

6.4 Frequency methods

6.4.1 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. If we take the complex valued input function

$$u(t) = \begin{cases} 0 & \text{for } t < 0\,, \\ e^{st}c & \text{for } t \geq 0\,, \end{cases}$$

with $s \in \mathcal{C}$ and c a complex vector. If $x(0) = 0$, then the corresponding output function will be

$$y(t) = \int_0^t G(t-\tau)e^{s\tau}c\,d\tau = \int_0^t G(r)e^{s(t-r)}c\,dr =$$

$$= \left(\int_0^t G(r)e^{-sr}\,dr\right)e^{st}c = \left(\int_0^t G(\tau)e^{-s\tau}\,d\tau\right)u(t)\,.$$

If we consider the limit of $t \to \infty$ and assume that the integral converges to $H(s)$ for $\Re s$ sufficiently large, then

$$y(t) \sim H(s)u(t).$$

This somewhat weird looking expression must be viewed as the approximate equality of two (complex valued) time functions; $H(s)$ 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 $s = \rho + i\omega,\ \rho,\ \omega \in \mathcal{R},\ u(\cdot)$ represents an oscillation. If $\int_0^\infty G_{ij}(\tau)d\tau < \infty$ for each element G_{ij} of the matrix G, then $\int_0^\infty G(\tau)e^{-st}dt$ exists for $\Re s \geq 0$ and more explicitly, for all $s = i\omega$ with $\omega \in \mathcal{R}$. If an input $u(t) = ce^{i\omega t}$ is applied, an output $y(t) \sim H(i\omega)u(t)$ 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 behaviour**. This behaviour approaches zero as $t \to \infty$ if $\int_0^\infty G_{ij}(\tau)d\tau < \infty$ for all i, j. It follows from the section on stability that $\int_0^\infty G_{ij}(\tau)d\tau < \infty$ if $\Re(\lambda_i) < 0$ for all eigenvalues λ_i of A.

REMARK **6.1** If in the definition of the Laplace transform as given by (6.1) one confines oneself to s on the imaginary axis, i.e. $s = i\omega$, 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 concerned is either 0 or $-\infty$.) Thus the theory of Fourier transforms can be viewed as a special case of the theory of Laplace transforms. The next subsection gives more information on the frequency responses.

6.4.2 Nyquist and Bode diagrams

In this section we confine ourselves again to single-input single-output linear differential systems. For the frequency response we can write

$$h(i\omega) = |h(i\omega)|e^{i \arg\ h(i\omega)}\ .$$

The stationary response of $u(t) = u_\omega e^{i\omega t+\phi}$, ω and $u_\omega \in \mathcal{R}$ is

$$y(t) = h(i\omega)u_\omega e^{i\omega t+\phi} = |h(i\omega)|u_\omega e^{i(\omega t+\arg\ h(i\omega)+\phi)}\ . \tag{6.25}$$

Now consider $u_\omega \sin\ (\omega t + \phi)$ as a sinusoidal input signal and treat it as the imaginary part of $u(t)$. Then, if we now take the imaginary parts of (6.25), the stationary response of

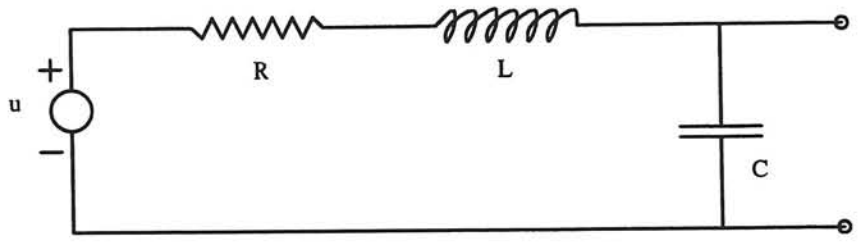
$$\Im(u(t)) = u_\omega \sin\ (\omega t + \phi) = \Im((u_\omega e^{i\phi})e^{i\omega t})$$

equals

$$\begin{aligned} \Im(y(t)) &= \Im(|h(i\omega)|u_\omega e^{i(\omega t+\phi+\arg\ h(i\omega))})\ , \\ &= |h(i\omega)|u_\omega \sin\ (\omega t + \phi + \arg\ h(i\omega)) \end{aligned} \tag{6.26}$$

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 with $|h(i\omega)|$ and increasing the phase with arg $h(i\omega)$.

EXAMPLE **6.11** Consider the following electric network (compare the example of subsection 2.4.4):



For the state we choose $x_1 = q$ (charge of the capacitor) and $x_2 = \phi$ (magnetic flux of the induction coil). If i is the current and v the voltage, then

$$\begin{aligned} \dot{x}_1 &= \dot{q} = i = \frac{1}{L}\phi = \frac{1}{L}x_2, \\ \dot{x}_2 &= \dot{\phi} = v = -Ri - \frac{1}{C}q + u = -\frac{R}{L}\phi - \frac{1}{C}q + u\,. \end{aligned}$$

Thus

$$A = \begin{pmatrix} 0 & \frac{1}{L} \\ -\frac{1}{C} & -\frac{R}{L} \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad C = (\frac{1}{C}\ 0),$$

(do not confuse the output matrix C 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 and hence

$$y(t) \sim |h(i\omega)|u_\omega \sin\,(\omega t + \phi + \arg\ h(i\omega))\,;$$

if an input signal $u_\omega \sin\,(\omega t + \phi)$ is applied. Further,

$$\begin{aligned} |h(i\omega)| &= \frac{1}{\sqrt{\{(1 - LC\omega^2)^2 + R^2C^2, \omega^2\}}} \\ \arg\ h(i\omega) &= \arctan\left(\frac{-RC\omega}{1 - LC\omega^2}\right). \end{aligned}$$

More generally, if a linear combination of sinusoidal signals, possibly of different frequencies, is applied to the system, then the output will be a linear combination of sinusoidal signals with the same frequencies as those of the input signal. □

EXERCISE 6.9 *Can you design a system of the form* $\dot{x} = Ax + Bu$, $y = Cx$, *with a suitably chosen initial condition, 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; but the design requires* $D = 0$.*) If your answer is affirmative (which it should be) what conditions should the transfer function* $h(s)$ *satisfy?*

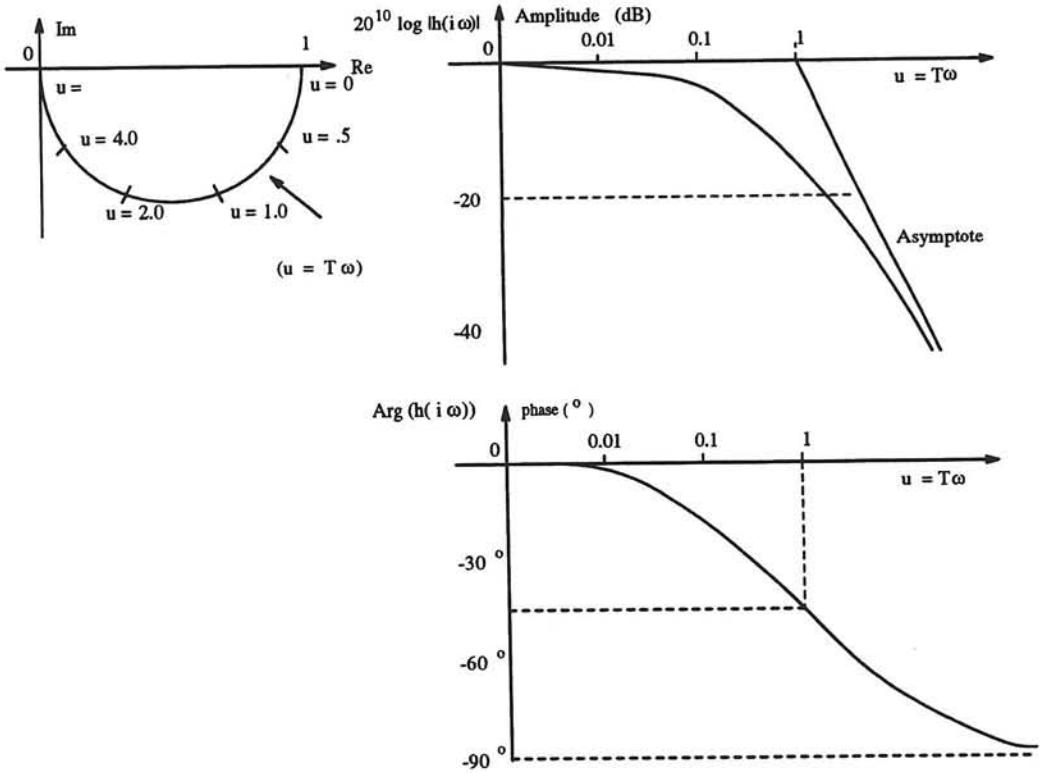
Frequency response functions are used frequently in network analysis, automatic control and acoustics. There are two well known methods to display $h(i\omega)$ graphically (and to get an impression of the properties of the system by studying these graphs). Many design techniques are based on these methods, to be discussed briefly here.

1. The **Nyquist diagram** or **polar plot**. The function $h(i\omega)$ is plotted as a curve in the 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; 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 $\ln \omega$.

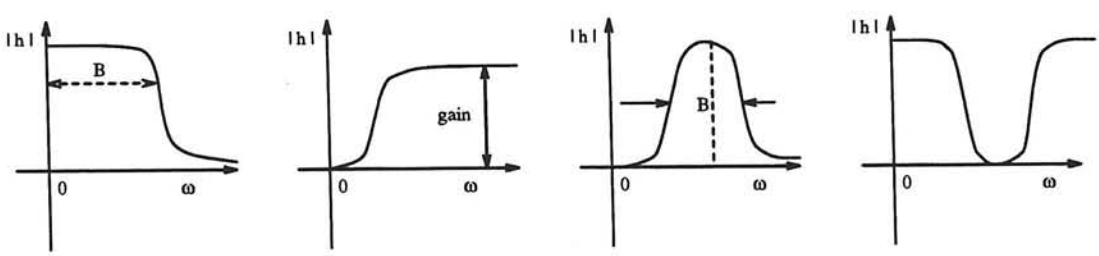
In the following figures the Nyquist diagram and the Bode diagram of the system with transfer function $1/(1 + Ts)$, with $T > 0$ being a constant, are given as an example. Please note that the scale of $\ln |h(i\omega)|$ is expressed in so-called **decibels** (dB). The graph of $|h(i\omega)|$ versus ω indicates which frequencies can pass the system and also with what **gain**. The system can thus be interpretated as a **filter** for the input signals. In the first of the next four figures in a row 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) over which the system will respond satisfactorily.

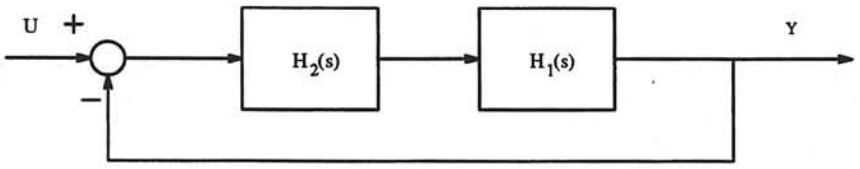
A simple application of a low frequency filter is the following. Noise signals consist usually of high frequency signals. If we want to get rid of this noise, a low frequency filter can be used. As a consequence, those parts of the input signal related to high frequencies will also be cut off.



EXAMPLE **6.12** Consider the feedback system of the following configuration (two blocks in the forward loop and unity feedback):

$H_1(s)$ 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

$$Y(s) = H(s)U(s)\ , \quad H(s) = (I + H_1(s)H_2(s))^{-1}H_1(s)H_2(s)\ .$$

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 sence, since then $(I+H_1(s)H_2(s))^{-1}H_1(s)H_2(s) \sim I$ and subsequently $Y(s) \sim U(s)$. For frequency considerations s is replaced by $i\omega$, and then $S(\omega)$ is defined as

$$S(\omega) = [I + H_1(i\omega)H_2(i\omega)]^{-1},$$

and is called the **sensitivity operator.** A system is said to have good sensitivity characteristics if – now suppose $u(t)$ and $y(t)$ are scalar so that $S(\omega)$ is a 1×1 matrix –

$$|1 + H_1(i\omega)H_2(i\omega)| \geq \phi(\omega)$$

for all $|\omega| \leq \omega_0$, the bandwidth of interest, and where $\phi(\omega)$ is a large positive function. □

EXAMPLE **6.13** [The differentiator] Suppose $y(t) = \frac{du(t)}{dt}$, $t \in \mathcal{R}$. Then

$$Y(s) = \int_0^\infty e^{-st}\frac{du}{dt}dt = ue^{-st}|_0^\infty + s\int_0^\infty ue^{-st}dt = sU(s),$$

if $u(0) = 0$. The transfer function is s, which is a nonproper rational function $(s/1)$. 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 $u(s)$ is known up to time t, then the derivative at the end point $s = t$ does not exist). Furthermore, $|h(i\omega)| = |\omega|$, such that higher frequencies are amplified more and more. For the phase, we get $\arg(i\omega) = \frac{\pi}{2}$ for all frequencies. □

Consider (6.7) with H_2 being the unit feedback, i.e. $H_2 = I$, and with H_1 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 $H_1(s)$. It is assumed

that this transfer function 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 text 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, 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 6.4 *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, equals the number of unstable poles of the closed-loop system minus the number of unstable 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 provided with the unity feedback. One talks about unstable poles if they are located in the right-half plane. This theorem is a simplified version of a slightly 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 6.5 *Assume that h is a rational (or, more generally: a meromorphic) function having no poles or zeros on a simple closed curve C. Assume in addition that C is oriented clockwise. Then*

$$\frac{1}{2\pi i} \int_C \frac{\frac{d}{ds}h(s)}{h(s)} ds$$

equals the number of poles minus the number of zeros, both only counted in the region bounded by C.

The assumption that the feedback system had to be the 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 $h_1(s)h_2(s)(1 + h_1(s)h_2(s))^{-1}$ and $h_2^{-1}(s)$ respectively, provided that both are well defined. The first of these two subsystems itself represents

a system characterized by $h_1(s)h_2(s)$ provided with 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 $h_2^{-1}(s)$).

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:

$$\begin{aligned} x(k+1) &= A(k)x(k) + B(k)u(k), \\ y(k) &= C(k)x(k) + D(k)u(k), \quad k = 0, 1, 2, \ldots \end{aligned} \tag{7.1}$$

Also, for other kinds of linear system descriptions, such as in Chapters 6 and 8.2, discrete analogues exist. Of 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. Some of them are:

1. The analysis must be performed on a digital computer which, because of its discrete-time behaviour, 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). The so-called **sampling periods** will be indicated by

   $$[0, \phi), \ [\phi, 2\phi), \ [2\phi, 3\phi), \ \ldots$$

   The input u is constant in each of these periods.

3. The output can only be measured at time instants ϕ, 2ϕ, ...

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

$$x(k+1) = A(k)x(k), \quad x \in \mathcal{R}^n, \tag{7.2}$$

can be written as

$$x(k) = \phi(k, k_0)x(k_0), \quad k \geq k_0, \tag{7.3}$$

where the **transition matrix** ϕ is given by

$$\phi(k, k_0) = \begin{cases} A(k-1)A(k-2)\dots A(k_0), & k \geq k_0 + 1, \\ I, & k = k_0. \end{cases} \tag{7.4}$$

The transition matrix is the unique solution of the matrix difference equation

$$\phi(k+1, k_0) = A(k)\phi(k, k_0), \quad k \geq k_0; \quad \text{with} \quad \phi(k_0, k_0) = I$$

If the matrices $A(k)$ do not depend on time, i.e.

$$A(k) = A \qquad k = k_0, k_0 + 1, \dots$$

then

$$\phi(k, k_0) = A^{k-k_0}, \qquad k \geq k_0.$$

Please note that the transition matrix ϕ is not necessarily nonsingular, this in contrast to the transition matrix for continuous-time systems. This is related to the fact that (7.2) is not necessarily well-defined in backward time. If A is invertible, then $\phi(k, k_0)$ is invertible and $\phi(k, k_0) = (\phi(k_0, k))^{-1}$. 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 that of the proof of Theorem 4.1).

THEOREM 7.1 *Given is the time-invariant linear difference equation (7.2) with A an $n \times n$ matrix with different eigenvalues $\lambda_i, \dots, \lambda_k$ $(k \leq n)$. The origin $x = 0$ is asymptotically stable if and only if $|\lambda_i| < 1$ for $i = 1, \dots, k$. The origin is stable if $|\lambda_i| \leq 1$ for $i = 1, \dots, k$ and moreover if to each eigenvalue with $|\lambda_i| = 1$ there correspond as many eigenvectors as the multiplicity of λ_i.*

EXAMPLE **7.1** Consider the model of a national economy as developed in Example 2.4.9. The system matrix is

$$A = \begin{pmatrix} 0 & -\mu \\ m & m(1+\mu) \end{pmatrix}. \tag{7.5}$$

The characteristic polynomial is $\lambda^2 - m(1+\mu)\lambda + m\mu$. The system is for instance asymptotically stable for $\mu = 1$ and $m < 1$. It is unstable if $\mu = 1$ and $m > 1$. For $\mu = 1$ and $m = 1$ it is also unstable. □

EXERCISE 7.1 *Let $u(k)$, $k = 0, 1, \dots$, be a sequence of measurements. In order to smoothen these measurements somewhat, a* **moving average** *is defined as*

$$y(k) = \frac{1}{3}(u(k) + u(k-1) + u(k-2)).$$

Determine a system in state space form (Equation (7.1)) for this input/output description.

EXERCISE 7.2 *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 % of sunshine of today is the input and where the forecast for tomorrow is the output. What is the dimension of the state?

The **characteristic polynomial** corresponding to $x(k+1) = Ax(k)$, A being a constant matrix, is $P(z) \stackrel{\text{def}}{=} \det(zI - A) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ with $a_n = 1$. The following theorem gives the **criterion of Jury**, being the discrete-time counterpart of the Routh criterion. A similar criterion is named after **Routh-Schur**, see [Isermann, 1989].

THEOREM 7.2 *The polynomial $P(z)$ has all its zeros strictly within the unit circle if and only if all the following conditions are satisfied:*

$$\begin{aligned}
\sum_{i=0}^{n} a_i &> 0, \\
(-1)^n \sum_{i=0}^{n} (-1)^i a_i &> 0, \\
|a_n| - a_0 &< 0, \\
|a_0^j| - |a_{n-j}^j| &< 0, \quad \text{for} \quad j = 1, 2, \ldots, n-2,
\end{aligned}$$

where the coefficients a_i^j are recursively determined by

$$a_k^{j+1} \stackrel{\text{def}}{=} a_0^j a_k^j - a_{n-j}^j a_{n-j-k}^j,$$

with $a_i^0 \stackrel{\text{def}}{=} a_{n-i}$.

The solution of the inhomogeneous state equation

$$x(k+1) = A(k)x(k) + B(k)u(k), \qquad x(k_0) = x_0$$

can be written as

$$x(k) = \phi(k, k_0)x_0 + \sum_{j=k_0}^{k-1} \phi(k, j+1)B(j)u(j), \quad k \geq k_0 + 1$$

If $x_0 = 0$, then the output vector $y(k) = C(k)x(k) + D(k)u(k)$ can be written as

$$y(k) = \sum_{j=k_0}^{k} K(k, j)u(j), \quad k \geq k_0, \tag{7.6}$$

where the **impulse response matrix** is given by

$$K(k,j) = \begin{cases} C(k)\phi(k,j+1)B(j), & j \leq k-1, \\ D(k) & j = k. \end{cases} \tag{7.7}$$

As far as the actual computation of $\phi(k,j)$ or $K(k,j)$ 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.

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 $x(k)$, $k = 0, 1, 2, \ldots$, is a sequence of (real or complex) numbers. The z-transform of this sequence is defined as

$$X(z) = \sum_{j=0}^{\infty} x(j) z^{-j}, \quad z \in \mathcal{C}, \tag{7.8}$$

where only those z-values will be considered for which this summation converges. If $X(z)$ exists for a value $z = z_0$, then it will exist for all z with $|z| \geq |z_0|$. If $X(z)$ is known, the sequence $x(k)$, $k = 0, 1, 2, \ldots$ can be recovered in several ways. One way is to look it up in a table of z-transforms. Another way is to write $X(z)$ as a power series expansion in z^{-1} and subsequently to identify the coefficients of the terms in this series expansion with $x(j)$, see (7.8). Still another way is provided by the following theorem (no proof is given here).

THEOREM 7.3 *If*

$$X(z) = \sum_{j=0}^{\infty} x(j) z^{-j}$$

converges for $|z| \geq |z_0|$, *then*

$$x(k) = \frac{1}{2\pi i} \int_C X(z) z^{k-1} dz, \quad k = 0, 1, 2, \ldots,$$

where C *is a closed contour in the complex plane in the area* $|z| \geq |z_0|$, *around the origin (take for instance a circle with radius* $r \geq |z_0|$*).*

If $x(k+1) = Ax(k) + Bu(k)$ is the state equation with A and B being constant matrices, then successively multiplying this equation by z^{-j}, $j = 0, 1, 2, \ldots$ and summing the results yields

$$zX(z) - zx(0) = AX(z) + BU(z).$$

If we solve for $X(z)$, the result is

$$X(z) = (zI - A)^{-1} BU(z) + z(zI - A)^{-1} x(0). \tag{7.9}$$

The z-transformation of $y(k) = Cx(k) + Du(k)$ yields

$$Y(z) = CX(z) + DU(z). \tag{7.10}$$

The combination of (7.9) and (7.10), with $x(0) = 0$, gives

$$Y(z) = (C(zI - A)^{-1}B + D)U(z).$$

The matrix $C(zI - A)^{-1}B + D$ is called the **transfer matrix.**

Suppose $u(k)$ is a periodic (complex-valued) input signal of the following form:

$$u(k) = u_0 e^{ik\alpha}, \quad k = 0, 1, 2, \ldots, \text{ with } \alpha \in \mathcal{R}.$$

As is well known from the theory of difference equations, the general solution of

$$x(k+1) = Ax(k) + Bu_0 e^{ik\alpha} \tag{7.11}$$

can be written as the summation of one arbitrary solution of this inhomogeneous difference equation and the general solution of the homogeneous equation $x(k + 1) = Ax(k)$. We will try to construct a solution of (7.11) in the form of

$$x(k) = x_0 e^{ik\alpha}, \quad k = 0, 1, 2, \ldots$$

Substitution into (7.11) gives the following condition on x_0:

$$e^{i\alpha}x_0 = Ax_0 + Bu_0, \text{ rm so } x_0 = (Ie^{i\alpha} - A)^{-1}Bu_0.$$

The general solution of (7.11) therefore is

$$x(k) = (Ie^{i\alpha} - A)^{-1}Bu_0 e^{ik\alpha} + A^k a,$$

with a arbitrary. This vector can be determined by the initial condition for the system. If the system is asymptotically stable, then $\lim_{k\to\infty} A^k = 0$, and

$$(Ie^{i\alpha} - A)^{-1}Bu_0 e^{i\alpha}$$

is called the **stationary response** of the state if the input signal is as given above. The term $A^k a$ is called the **transient behaviour.** The stationary response of the output is

$$y_{\text{stat}}(k) = \left[C(Ie^{i\alpha} - A)^{-1}B + D\right] u_0 e^{ik\alpha} = H(e^{i\alpha})u(k).$$

This formula (for single-input single-output systems) is the discrete time analogue of formula (6.26). Note that here the stationary response is completely determined by the transfer matrix-values on the unit circle (in the continuous-time case: on the imaginary-axis). Confining ourselves to single-input single-output systems, the transfer function can be written as

$$h(z) = d + \frac{q(z)}{p(z)}, \tag{7.12}$$

with

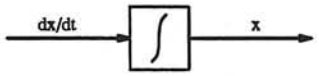
$$q(z) = q_{n-1}z^{n-1} + \ldots + q_1 z + q_0; \qquad p(z) = z^n + p_{n-1}z^{n-1} + \ldots + p_1 z + p_0.$$

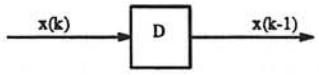
A state space realization corresponding to (7.12) is

$$x(k+1) = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & 0 & 1 & 0 \\ 0 & \dots & & \dots & 0 & 1 \\ -p_0 & -p_1 & \dots & & \dots & -p_{n-1} \end{pmatrix} x(k) + \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} u(k),$$

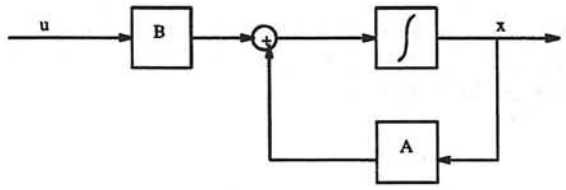
$$y(k) = (q_0, q_0, \dots, q_{n-1})x(k) + du(k).$$

The derivation is exactly the same as in the continuous-time case. Block diagrams for time-discrete time-invariant linear systems can be drawn similarly as in continuous-time case. The only difference is that the integrator $\int$ must be replaced by an operator D defined by $Dx(k) = x(k-1)$. This operator is sometimes called **delay operator** or **backward delay operator.** Its inverse $\sigma \stackrel{\text{def}}{=} D^{-1}$; $\sigma x(k) = x(k+1)$, is called the **(forward) delay operator.**

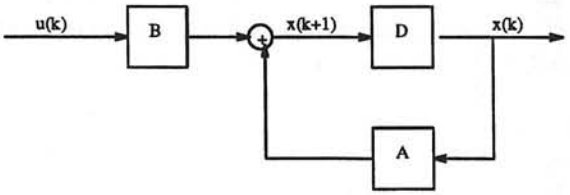




dx/dt =Ax + Bu:



x(k+1) = Ax(k) + Bu(k):



EXAMPLE **7.2** Consider a simplified version of the national economy as given in Example 2.4.9. The model now is

$$\begin{aligned} x(k+1) &= px(k) - ru(k), \\ y(k) &= x(k). \end{aligned} \tag{7.13}$$

The scalar $y(k)$ is the total national income in year k and the scalar $u(k)$ is the expenditure in year k; p and r are constants. The transition matrix is

$$\phi(k, k_0) = p^{k-k_0}.$$

The general solution of (7.13) is

$$x(k) = p^{k-k_0}x(k_0) + \sum_{j=k_0}^{k-1} p^{k-j-1}(-ru(j)). \tag{7.14}$$

The impulse response function is

$$K(k, k_0) = \begin{cases} -rp^{k-k_0-1}, & k \geq k_0 + 1, \\ 0, & k = k_0, \end{cases}$$

and the transfer function is

$$h(z) = \frac{-r}{z-p}, \qquad |z| > p.$$

Suppose that the expenditures are constant, i.e. $u(k) = u_0$ for $k = 0, 1, 2, \ldots$ Then

$$U(z) = \sum_{j=0}^{\infty} u_0 z^{-j} = \frac{1}{1-\frac{1}{z}}u_0 = \frac{z}{z-1}u_0, \qquad |z| > 1.$$

The response of the output (assume $x(0) = 0$) is

$$Y(z) = h(z)U(z) = \frac{-r}{z-p}\frac{z}{z-1}u_0, \qquad |z| > \max(p, 1).$$

In order to find $y(k)$ we will use the second mentioned method given above, i.e. by means of a power series expansion. Factorization of $h(z)$ gives

$$Y(z) = \left(\frac{-p}{z-p} + \frac{1}{z-1}\right)\frac{-r}{1-p}u_0,$$

with

$$\begin{aligned} \frac{-p}{z-p} &= \frac{-p}{z}\left(1 + \frac{p}{z} + (\frac{p}{z})^2 + \ldots\right), \\ \frac{1}{z-1} &= \frac{1}{z}\left(1 + \frac{1}{z} + (\frac{1}{z})^2 + \ldots\right), \qquad |z| > \max(p, 1), \end{aligned}$$

and hence

$$Y(z) = \sum_{j=0}^{\infty} \frac{-r}{p-1}(1-p^j)u_0 z^{-j}.$$

By definition, $Y(z) = \sum_{j=0}^{\infty} y(j)z^{-j}$, and therefore

$$y(k) = \begin{cases} 0 & \text{for} \quad k = 0, \\ -\frac{r-rp^k}{p-1}u_0 & \text{for} \quad k = 1, 2, \ldots \end{cases}$$

This result can also be obtained by using (7.14) directly. □

EXAMPLE **7.3** We are given the discrete-time 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, $\lambda = -1/5$ and $\lambda = -4/5$. If the two corresponding eigenvectors are put together to form the matrix T, and if the two eigenvalues are the diagonal values of the diagonal matrix D, we get for the transisition matrix

$$\Phi(k, k_0) = A^{k-k_0} = TD^{k-k_0}T^{-1}$$

$$= \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} (-\frac{1}{5})^{k-k_0} & 0 \\ 0 & (-\frac{4}{5})^{k-k_0} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{4}{3}(-\frac{1}{5})^{k-k_0} - \frac{1}{3}(-\frac{4}{5})^{k-k_0} & \frac{1}{3}(-\frac{1}{5})^{k-k_0} - \frac{1}{3}(-\frac{4}{5})^{k-k_0} \\ -\frac{4}{3}(-\frac{1}{5})^{k-k_0} + \frac{4}{3}(-\frac{4}{5})^{k-k_0} & -\frac{1}{3}(-\frac{1}{5})^{k-k_0} + \frac{4}{3}(-\frac{4}{5})^{k-k_0} \end{pmatrix}.$$

The impulse response can be calculated according to $K(k,j) = C\Phi(k, j+1)B$ for $j \leq k-1$:

$$K(k,j) = \frac{1}{3}(-\frac{1}{5})^{k-j-1} + \frac{2}{3}(-\frac{4}{5})^{k-j-1}.$$

For $j = k$ we have $K(k,j) = 0$, since there is no direct throughput term in the model. The transfer function follows from $h(z) = C(zI - A)^{-1}B$ and it equals

$$\frac{z + \frac{2}{5}}{z^2 + z + \frac{4}{25}}.$$

□

EXERCISE 7.3 *We are given the time discrete 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 following periodic input signal is applied to the system:

$$u(k) = \begin{cases} 0 & k < 0, \\ (-1)^k, & k \geq 0. \end{cases}$$

What is the output response (take $x(0)$ as the zero state)? Why is the output signal not periodic?

The time-discrete, time-invariant, linear system characterized by matrices (A, B, C, D) is called **controllable** if for each $x_0, x_1 \in \mathcal{R}^n$ a time $k > 0$ and a sequence $u(0), u(1), \ldots$ exist such that $x(k, x_0, u) = x_1$. The meaning of $x(k, x_0, u)$ will be clear; the state at time instant k, starting with initial condition $x(0) = x_0$ and having applied an input sequence u. The system is **observable** if a $k > 0$ exists such that for any sequence of controls u we have:

$$y(j, x_0, u) = y(j, x_1, u), \quad j = 0, 1, \ldots, k, \quad \text{implies} \quad x_0 = x_1$$

The conditions in terms of matrices A, B, C and D for controllability and observability are the same as in the time-continuous case. This will be shown in the next theorem for controllability. Sometimes one distinguishes **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 next theorem); the condition is:

$$\text{rank}\left[B \; AB \; \ldots \; A^{n-1}B\right] = n, \tag{7.15}$$

wheras null controllability is not as strong as "standard" controllability.

EXAMPLE **7.4** 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 7.4 *The discrete-time system characterized by (A, B, C, D) is controllable if (7.15) holds.*

Proof The system is

$$x(k+1) = Ax(k) + Bu(k), \qquad x \in \mathcal{R}^n.$$

Assume for the time being that $x(0) = 0$. Which states can be reached in one time-step? In other words, what is the class of possible $x(1)$-vectors?

$$x(1) = Ax(0) + Bu(0) = Bu(0)$$

and therefore the class of reachable $x(1)$-vectors equals Im B. Within two time-steps we can reach

$$x(2) = Ax(1) + Bu(1) = ABu(0) + Bu(1).$$

Since $u(0)$ and $u(1)$ are arbitrary, the class of reachable $x(2)$ vectors equals Im B + Im AB. Continuing this way we obtain that the class of reachable $x(k)$ vectors (starting from $x(0) = x_0$) is

$$\text{Im } B + \text{Im } AB + \ldots + \text{Im } A^{k-1}B = \text{Im } \left[B \; AB \; \ldots \; A^{k-1}B\right].$$

According to Lemma 4.1

$$\text{Im } A^k B \subset \text{Im } \left[\, B \; AB \; \ldots \; A^{n-1}B\right], \qquad k = 0, 1, \ldots, n, n+1, \ldots$$

and therefore the set of x-values which are reachable at all, can be reached in at most n steps, and this set equals

$$\text{Im } \left[B \; AB \; \ldots A^{n-1}B\right].$$

If (7.15) is fulfilled then any point x_1 in $\mathcal{R}^n$ can be reached from the origin. The final point can always be reached in exactly n time-steps. Sometimes it may also be possible to reach it in fewer time steps. If $x(0) = x_0 \neq 0$, then instead of going from x_0 to the final point x_1, we consider the equivalent problem of going from the origin to $x_1 - A^n x_0$ in exactly n steps and this is possible if (7.15) is fulfilled. The same technique was used in the proof of Theorem 4.5. □

EXERCISE 7.4 *Consider the discrete-time system*

$$x(k+1) = Ax(k) + Bu(k), \qquad x \in \mathcal{R}^n, \qquad u \in \mathcal{R}.$$

For this system

$$\text{rank } \left[\, B \; AB \; \ldots \; A^{n-1}B\right] = r < n.$$

Prove that the state x can be steered from the initial point $x(0) = \theta_0$ to the final point θ_f in at most r steps if it is known that

$$\theta_0, \theta_f \in \text{Im } \left[\, B \; AB \; \ldots \; A^{n-1}B\right].$$

EXERCISE 7.5 *In econometrics, one works a lot with so-called* **ARMA models.** *They will be briefly introduced in this exercise. For the so-called moving average (MA) model,*

$$Y(z) = [q_{n-1}z^{-1} + q_{n-2}z^{-2} + \cdots + q_1 z^{-n+1} + q_0 z^{-n}]U(z),$$

give in matrix form $(\{A, B, C\})$ *and in block diagram form, two different types of realizations. Again the same question, but now with respect to the so-called autoregressive (AR) model;*

$$Y(z)[1 + p_{n-1}z^{-1} + \cdots + p_1 z^{-n+1} + p_0 z^{-n}] = U(z).$$

Given a mixed or ARMA model,

$$\frac{Y(z)}{U(z)} = \frac{q_{n-1}z^{-1} + q_{n-2}z^{-2} + \cdots + q_1 z^{-n+1} + q_0 z^{-n}}{1 + p_{n-1}z^{-1} + p_{n-2}z^{-2} + \cdots + p_1 z^{-n+1} + p_0 z^{-n}}$$

$$= \frac{q_{n-1}z^{n-1} + q_{n-2}z^{n-2} + \cdots + q_1 z^{-n+1} + q_0}{z^n + p_{n-1}z^{n-1} + p_{n-2}z^{n-2} + \cdots + p_1 z^{-n+1} + p_0},$$

show how to merge the block diagrams of the MA and AR models just obtained so as to obtain realizations of the ARMA model. Is it possible to construct realizations with no more than n delay operators?

EXAMPLE **7.5** 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; all components of b are zero except for one which equals one. The user of the system may choose the position of this latter component and this position may be k-dependent. 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), 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 the admissible b vectors controllability is not possible. □

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 **sampling interval**. A choice to be made is how large should Δ 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 7.5 *If the function $x(t)$ is band-limited, i.e. a $W > 0$ exists such that $X(i\omega) = 0$ for $|\omega| > W$, then no information is lost by sampling at a period less than or equal to π/W. (Here we used the notation $X \stackrel{\text{def}}{=} \mathcal{L}(x)$.)*

If one would sample with a certain 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 distinguisable 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**.

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 sometimes simply written as u and y. The time will either be continuous ($t \in T$ with $T = (-\infty, +\infty)$ or $T = [t_0, \infty)$), or be discrete ($t \in T$ with $T = \mathcal{Z}$ or $T = \{t_1, t_2, \ldots, t_n, \ldots\}$). If $t = \mathcal{R}$ we talk about continuous-time systems, if $T = \mathcal{Z}$ we talk about discrete-time systems.

Two ways exist in order to describe the dynamic behaviour of systems; an external and an internal description. The external description considers the system as an input/output map, i.e. $y(t) = f(u(\cdot), t)$. If a system is described by means of the internal or state space form description, another quantity, the state $x(t)$, 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*

$$\sum_{\text{I/O}} = \{T, U, \underline{U}, Y, \underline{Y}, F\},$$

whereby

i) T is the time axis (i.e. $T = \mathcal{R}$ or $\mathcal{Z}$ or a subset of $\mathcal{R}$ or $\mathcal{Z}$).

ii) U is the set of input values; this set is called the input space. Quite often $U = \mathcal{R}^m$, or U is a subset of $\mathcal{R}^m$.

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* $Y = \mathcal{R}^p$; 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.* $F : \underline{U} \to \underline{Y}$.
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* $t \leq t'$ *with* $t' \in T$, *then* $(Fu_1)(t') = (Fu_2)(t')$ *and therefore also* $(Fu_1)(t) = (Fu_2)(t)$ *for all* $t \leq t'$.

DEFINITION 8.2 *The system* $\sum_{I/O}$ *is called* **linear** *if* U, Y, $\underline{U}$ *and* $\underline{Y}$ *are linear vectorspaces (for example* $U = \mathcal{R}^m$, $Y = \mathcal{R}^p$*) and if* $F : \underline{U} \to \underline{F}$ *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* $F(\lambda u_1) = \lambda F u_1$.

DEFINITION 8.3 *The system* $\sum_{I/O}$ *is called* **time-invariant** *(or, equivalently,* **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 *defined by* $(S_\tau u)(t) = u(t + \tau)$, $(S_\tau y)(t) = y(t + \tau)$, *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* $\tau \in T$.

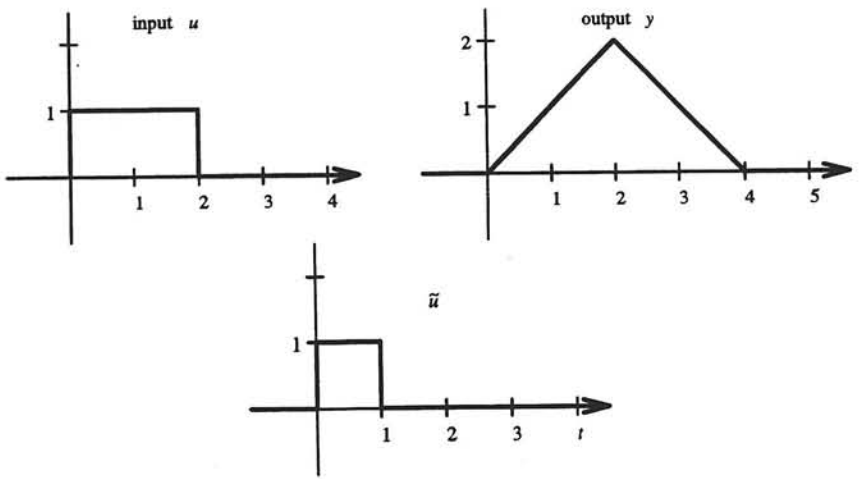
To say it in a simple way, a system is time-invariant if a shift along the time axis yields an equivalent system. If $t \to u(t)$ leads to an output $t \to y(t)$, then $t \to u(t - \tau)$ should result in $t \to y(t - \tau)$. If a signal is applied one hour later, we get the same response, expect for a delay of one hour.

EXERCISE 8.1 *We are given a time-invariant and linear system of which we know that the input* $u(t)$ *yields an output* $y(t)$, *where*

$$u(t) = \begin{cases} 1 & 0 \leq t < 2 \\ 0 & \text{otherwise} \end{cases}, \quad y(t) = \begin{cases} t & 0 \leq t < 2, \\ 4 - t & 2 \leq t < 4, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the output function $\tilde{y}(t)$ *which corresponds to the input* $\tilde{u}(t)$, *where*

$$\tilde{u}(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

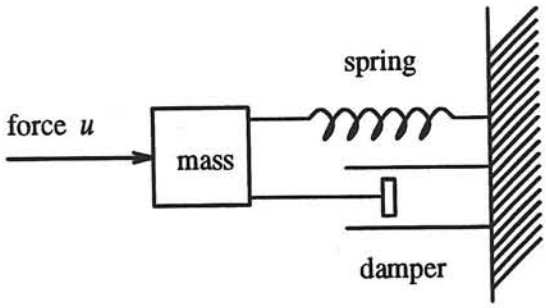


DEFINITION 8.4 *The system* $\sum_{I/O}$ *is called* **memoryless** *or* **static** *if a function* f *exists,* $f : U \times T \to Y$ *such that* $(Fu)(t) = f(u(t), t)$. *This means that* Fu *at time* t *only depends on* $u(t)$ *and not on the past (or future) of* u.

EXAMPLE **8.1** A mass m moves along a straight line and is connected with a spring with characteristic constant k. There is friction which is a function of the speed of the mass. An external force $u(t)$ acts on the mass. 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, that 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.



□

EXAMPLE **8.2** 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 time discrete system. As input at time n we define

$$u(n) = \begin{pmatrix} \text{number of spots on the die at } n\text{-th throw, first person} \\ \\ \text{number of spots on the die at } n\text{-th throw, second person} \end{pmatrix}.$$

The state can be defined as

$$x(n) = \begin{pmatrix} \text{position of first} & \text{person's marker on the board at time } n \\ \\ \text{position of second} & \text{person's marker on the board at time } n \end{pmatrix}.$$

For simplicity 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 8.5 *[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*

$$\sum_{M} = \{T, U, \underline{U}, Y, \underline{Y}, X, \phi, r\},$$

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* $X = \mathcal{R}^n$ *or* X *is a subset of* $\mathcal{R}^n$.

iii) $\phi : T_+^2 \times X \times \underline{U} \to X$, *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* $\phi(t_1, t_0, x_0, u)$ *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. $\phi(t, t, x, u) = x$.

b) satisfy the semi-group property, i.e. $\phi(t_2, t_1, \phi(t_1, t_0, x_0, u), u) = \phi(t_2, t_0, x_0, u)$.

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* $\phi(t_1, t_0, x_0, u_1) = \phi(t_1, t_0, x_0, u_2)$.

iv) $r : X \times U \times T \to Y$ *is the output function (or measurement function or observation function)* $y(t) = r(x(t), u(t), t)$. *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 8.6 *$\sum_M$ is called linear if $U, Y, \underline{U}, \underline{Y}$, and X are linear vectorspaces and if*

i) the mapping $\phi(t_1, t_0, \cdot, \cdot) : X \times \underline{U} \to X$ is jointly linear in both arguments, i.e. if $\phi(t_1, t_0, x_0, u) = x$ and $\phi(t_1, t_0, \tilde{x}_0, \tilde{u}) = \tilde{x}$ then $\phi(t_1, t_0, \lambda x_0, \lambda u) = \lambda x$ and $\phi(t_1, t_0, x_0 + \tilde{x}_0, u + \tilde{u}) = x + \tilde{x}$.

ii) the mapping $r(t, \cdot, \cdot) : X \times U \to$ is jointly linear in both arguments.

DEFINITION 8.7 *$\sum_M$ is called time-invariant if $t_1, t_2 \in T$ then $t_1 + t_2 \in T$, $S_t\underline{U} \subset \underline{U}$, $S_t\underline{Y} \subset \underline{Y}$ for all $t \in T$ and if moreover*

i) $\phi(t_1 + t, t_0 + t, x_0, u) = \phi(t_1, t_0, x_0, S_t u)$ for all $t \in T$

ii) $r(t, x, u)$ is independent of t and therefore often written as $r(x, u)$.

EXAMPLE **8.3** Suppose that the relationship between input $u(\cdot)$ and output $y(\cdot)$ 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, viz. $x(t)$, should at least contain the function $U(\cdot)$: $[t - \theta, t) \to \mathcal{R}$. It turns out that the state equals this function:

$$x(t) = u|_{[t-\theta,t)},$$

where $u|_{[t-\theta,t)}$ denotes the restriction of $u(\cdot)$ to the interval $[t - \theta, t)$. □

DEFINITION 8.8 *The system $\sum_M$ is called autonomous if U consists of only one element. (Therefore no control is possible.)*

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 $\sum_M = \{T, U, \underline{U}, Y, \underline{Y}, X, \phi, r\}$ 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 ϕ- and r-relations. Suppose for simplicity that $\sum_M$ is time invariant. Choose a $t_0 \in T$ and a $x_0 \in X$ (think of initial time and initial state) and define

$$(Fu)(t) = r(\phi(t, t_0, x_0, u), u(t)) \quad \text{for} \quad t \geq t_0.$$

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

$$x(t) = x_0, \quad u(t) = u_0, \quad y(t) = y_0, \quad t < t_0,$$

where u_0 and y_0 are constants in resp. U and Y. 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 $-\infty$ (if $T = (-\infty, +\infty)$) then 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 space X instead of 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 8.9 *A dynamical system is a triple $\Sigma = (T, \mathcal{W}, \mathcal{B})$, where T represents the time-axis, $\mathcal{W}$ is the signal space and $\mathcal{B} \subseteq \mathcal{W}^T$ is the behaviour*

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 $x = (x_1, \ldots, x_n)$. Let us assume that the f_i-functions are defined in such a way that mathematically well defined solutions $x(t)$ to the differential equations exist. In this example, T is the real axis, $\mathcal{W}$ is the set of all $x(\cdot)$ functions and $\mathcal{B}$ 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 ordinary differential equations. The differentiation operator $\frac{d}{dt}$ is represented by the (Laplace) variable s.

DEFINITION 8.10 *A* **polynomial matrix** *(in s) is a matrix of which the entries are polynomials in the variable s.*

DEFINITION 8.11 *A linear time-independent 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 $\overline{n} \times \overline{n}$, $\overline{n} \times m$ and $p \times \overline{n}$ respectively; the vector ξ, having $\overline{n}$ 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) simply are (possibly higher order) differential equations related to each other.

EXAMPLE **8.4** The classical equation of a force F acting on a point mass with mass m is (F is the input, x the output)

$$\ddot{x} = F/m.$$

This equation allows at least two polynomial representations:

$$\begin{cases} s^2\xi_1 &= F/m, \\ y &= \xi_1, \end{cases}$$

with the one-dimensional partial state $\xi_1 = x$, and

$$\begin{cases} \begin{pmatrix} s & -1 \\ 0 & s \end{pmatrix} \xi_2 &= \begin{pmatrix} 0 \\ \frac{1}{m} \end{pmatrix}, \\ y &= (1 \;\; 0)\xi_2, \end{cases}$$

with the two-dimensional partial state $\xi_2 = (x, \dot{x})^T$. □

EXERCISE 8.2 *Argue that a polynomial representation of $\dot{x} = Ax + Bu$, $y = Cx$ equals $(sI - A)\xi = Bu$, $y = C\xi$.*

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 $\overline{n} \leq n$, 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 will not be used explicitly anymore in this text.

DEFINITION 8.12 *The two systems*

$$\Sigma_i : \quad \begin{cases} P_i(s)\xi_i &= Q_i(s)u, \\ y &= R_i(s)\xi_i, \end{cases} \qquad i = 1, 2, \tag{8.2}$$

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 $M_1(s)_{n_1 \times n_2}$, $M_2(s)_{n_2 \times n_1}$, $N_1(s)_{n_1 \times m}$ and $N_2(s)_{n_2 \times m}$ exist such that the following two systems

$$S_1 : \quad \begin{cases} P_1(s)\xi_1 &= Q_1(s)u, \\ \xi_2 &= M_2(s)\xi_1 + N_2(s)u, \\ y &= R_1(s)\xi_1, \end{cases} \tag{8.3}$$

$$S_2 : \quad \begin{cases} \xi_1 &= M_1(s)\xi_2 + N_1(s)u, \\ P_2(s)\xi_2 &= Q_2(s)u, \\ y &= R_2(s)\xi_2, \end{cases} \tag{8.4}$$

have identical solutions (i.e. the same input applied to both systems yields identical outputs, provided that one chooses the initial conditions suitably).

DEFINITION 8.13 *The* **transfer matrix** *of (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 polynomial in s and is 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 **8.5** 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 8.14 *A square polynomial matrix is called* **nonsingular** *if its determinant is a polynomial not identically zero. A square polynomial matrix is called* **unimodular** *if its determinant is a nonzero constant.*

By the Cramer 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 **8.6** 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 8.15 *The rank of a polynomial matrix is the size of a largest square submatrix (of this polynomial matrix) that is invertible.*

DEFINITION 8.16 *Suppose $N(s) = \{N_{ij}(s)\}$ is a polynomial matrix of rank k. It is in the so-called* **Smith form** *if*

- $N_{ij}(s) = 0$ for $i \neq j$;
- $N_{ii}(s) = 0$ for $i \geq k+1$;
- $N_{ii}(s)$ is monic and divides $N_{i+1,i+1}(s)$.

THEOREM 8.1 *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 **8.1** If all polynomials in $N(s)$ would be constants, the theorem above resembles the so-called singular value decomposition, well known in matrix algebra. □

One can construct the Smith form of a polynomial matrix in a way which resembles the conventional column and row operations, as shown in the following example.

EXAMPLE **8.7** Suppose

$$N(s) = \begin{pmatrix} s-a & 1 \\ 0 & s-a \end{pmatrix}.$$

Permutation of the columns yields

$$N_1(s) = \begin{pmatrix} 1 & s-a \\ s-a & 0 \end{pmatrix}$$

and 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 $U(s)$ and $V(s)$ matrices are

$$U(s) = \begin{pmatrix} 1 & 0 \\ s-a & 1 \end{pmatrix}, \quad V(s) = \begin{pmatrix} s-a & 1 \\ -1 & 0 \end{pmatrix}.$$

□

EXERCISE 8.3 *Show that the Smith form of the matrix $N(s)$ introduced in Example 8.5 equals*

$$\begin{pmatrix} 1 & 0 \\ 0 & s(s+s^3) \end{pmatrix}.$$

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 $s = d/dt$ 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; x and u are finite dimensional vectors and f and g are vector, respectively matrix, functions of appropriate size. A typical example is the (simplified) modelling of maneuvring a car.

EXAMPLE **8.8** 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 sytem around any point in $\mathcal{R}^3$, the linearized system turns out to be noncontrollable. However, driving a car is a controllable process, at least to the experience of most people. □

A theory exists which studies controllability directly in terms of 'vector fields' $(f(x),\ g_{\cdot,i}(x)$, the latter expression denoting the ith column of the matrix g, $i = 1, \ldots, m$, 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 other 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 $x(t) \in \mathcal{R}^n$ contains those variables of which the time derivatives appear in the equations; the vector $p(t) \in \mathcal{R}^r$ contains the variables which only appear algebraically. The function $u(t) \in \mathcal{R}^m$ 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 those of the form $\dot{x} = Ax + Bu$. Systems characterized by such 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 $\dot{x} = Ax + Bu$, 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 behaviour 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; 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 probabilisitically. 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 $u(s)$, $s \leq t$, all future quantities within the system are determined in a probabilistic way. That is for instance the case if the probability distribution functions are uniquely determined by $x(t)$ and $u(s)$, $s \leq t$. The future behaviour 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 **8.9** An industrial area can be in two situations: 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 tomorrow	
		G	B
initial condition	G	0.8	0.2
	B	0.4	0.6

$u = 0$

		condition tomorrow	
		G	B
initial condition	G	0.9	0.1
	B	0.6	0.4

$u = 1$

The numbers in these tabular forms denote transformation 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; $X = \{G, B\}$. □

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 **8.10** We consider the following situation of an oversimplified and old-fashioned marriage. The state space has three elements, viz.

- 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 behaviour of the wife:

- u_1: wife is quiet;
- u_2: wife shouts;
- u_3: wife cooks.

As a result of 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

□

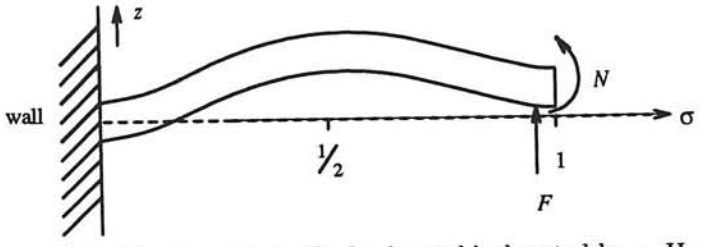
EXERCISE 8.4 *Give a state space form description of a discrete-time system where $U = Y = \{0, 1\}$, such that the output y at time t equals 1 if the input until (and not including) t has shown an even number of 1's and equals 0 otherwise.*

8.3.5 Distributed parameter systems

In this subsection we will briefly talk about a class of systems which is (also) important from a practical point of view, but which will not be 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 ($\mathcal{R}^n$) 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 therefore electric charge, temperature, etc., are not only time dependent, but also location (spatial) dependent. Such 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 are called lumped systems).

EXAMPLE **8.11** Consider the following flexible beam of length one.



The 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 behaviour of the flexible beam subject to a force $F(t)$ and a torque $N(t)$. For the complete description of the behaviour 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 can not 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. □

EXERCISE 8.5 *Assume that* $\rho = 1$ *and* $EI = 1$ *in the above model, 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?*

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 $x \in \mathcal{R}^n$. 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

$$\begin{aligned} x_i(k+1) &= \max(a_{i1} + x_1(k), a_{i2} + x_2(k), \ldots, a_{in} + x_n(k)) \\ &= \max_j (a_{ij} + x_j(k), \quad k = 0, 1, 2, \ldots), \quad i = 1, \ldots, n \ , \end{aligned} \tag{8.11}$$

$k = 0, 1, \ldots$ 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 take

$$A = \begin{pmatrix} 3 & 7 \\ 2 & 4 \end{pmatrix}, \tag{8.12}$$

and as initial condition

$$x_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{8.13}$$

Then the time evolution of (8.11) becomes

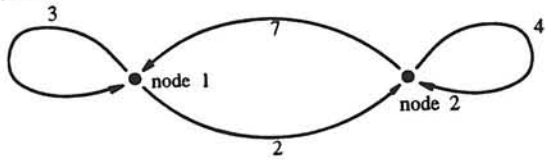
$$x(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ x(1) = \begin{pmatrix} 7 \\ 4 \end{pmatrix},\ x(2) = \begin{pmatrix} 11 \\ 9 \end{pmatrix},\ x(3) = \begin{pmatrix} 16 \\ 13 \end{pmatrix}, \ldots \quad (8.14)$$

We are used to thinking of the argument t in $x(t)$ as a time instant; 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 for each component x_i. Entry a_{ij} corresponds to the arc from node j to node i. 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 (the rather neutral word 'results' is used, it could equally have been messages, ingredients or products,...) to node i, 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 (i.e. time duration of the activity) at node j and the travelling time (the rather neutral 'travelling time' is used rather than for instance 'transportation time' or 'communication time') from node j to node i.

For the example given above, the network has two nodes and four arcs, as given in the figure below.



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 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 of 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, respectively at times 7 and 11). 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**)

$$\int_{t_0}^{t_1} g(t, x(t), u(t))dt + q(x(t_1)) \tag{8.16}$$

is minimized, subject to $u \in \underline{U}$ and (8.15). In this problem

$$\begin{aligned} &x(.) \in \mathcal{R}^n; \qquad u(.) \in \mathcal{R}^m, \\ &f : \mathcal{R} \times \mathcal{R}^n \times \mathcal{R}^m \to \mathcal{R}^n, \\ &g : \mathcal{R} \times \mathcal{R}^n \times \mathcal{R}^m \to \mathcal{R}, \\ &q : \mathcal{R}^n \to \mathcal{R}. \end{aligned}$$

We tacitly assume that such a minimizing u-function, 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 $u(.)$ must be chosen such that a given point x_f is reached at t_1; $x(t_1) = x_f$. This is an additional requirement on u. In the latter case the term $q(x(t_1)) = q(t_f)$ is predetermined and is independent of the control function $u(.)$ used. Another variation is that 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 $\mathcal{R}^n \times \mathcal{R}$ space; 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

$$\int_{t_0}^{t_1} (x^T Q x + u^T R u) dt + x^T(t_1) Q_f x(t_1), \tag{8.18}$$

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 and are assumed to be positive definite. These matrices are weighting matrices, because of the positiveness,

$$x^T Q x \geq 0; \qquad u^T R u \geq 0; \qquad x^T(t_1) Q_f x(t_1) \geq 0,$$

and these terms penalize deviations of x, u, resp. $x(t_1)$, 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 $x^T(t)Qx(t)$ in the cost function), but not at the expense of too much control effort (expressed by the term $u^T(t)Ru(t)$). The term $x^T(t_1)Q_f x(t_1)$ expresses the fact that we would like to have the final point $x(t_1)$ close to the origin also. For this particular 'linear quadratic' problem the solution can be obtained in a straightforward way, by a completing the square argument:

$$\begin{aligned} & \int_{t_0}^{t_1} (x^T Q x + u^T R u) dt + x^T(t_1) Q_f x(t_1) \\ = & \int_{t_0}^{t_1} (x^T Q x + u^T R u + d/dt(x^T P(t) x)) dt + x^T(t_0) P(t_0) x(t_0), \end{aligned}$$

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: $P(t) = P^T(t)$. The cost function becomes

$$\begin{aligned} & \int_{t_0}^{t_1} (x^T Q x + u^T R u + \dot{x}^T P x + x^T \dot{P} x + x^T P \dot{x}) dt + x^T(t_0) P(t_0) x(t_0) \\ = & \int_{t_0}^{t_1} \left[x^T Q x + u^T R u + (Ax + Bu)^T P x + x^T \dot{P} x + x^T P (Ax + Bu) \right] dt \\ & + x^T(t_0) P(t_0) x(t_0) \\ = & \int_{t_0}^{t_1} \left[x^T \left(Q + \dot{P} + A^T P + PA \right) x + u^T R u + u^T B^T P x + x^T P B u \right] dt \\ & + x^T(t_0) P(t_0) x(t_0) \\ = & \int_{t_0}^{t_1} \left[x^T \left(Q + \dot{P} + A^T P + PA \right) x + \left(u + R^{-1} B^T P x \right)^T R \left(u + R^{-1} B^T P x \right) - \right. \\ & \left. x^T P B^T R^{-1} B P x \right] dt + x^T(t_0) P(t_0) x(t_0) \end{aligned}$$

$$= \int_{t_0}^{t_1} \left[x^T \left(Q + \dot{P} + A^T P + PA - PBR^{-1}B^T P \right) x + \right.$$
$$\left. \left(u + R^{-1}B^T Px \right)^T R \left(u + R^{-1}B^T Px \right) \right] dt + x^T(t_0) P(t_0) x(t_0).$$

If we now choose $P(t)$ to satisfy the differential equation

$$\dot{P} = -A^T P - PA - Q + PBR^{-1}B^T P; \quad P(t_1) = Q_f, \tag{8.19}$$

then the cost function becomes

$$\int_{t_0}^{t_1} \left(u + R^{-1}B^T Px \right)^T R \left(u + R^{-1}B^T Px \right) dt + x_0^T P(t_0) x_0. \tag{8.20}$$

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 $R > 0$, it is clear from (8.20) that the minimizing control is

$$u^*(t) = -R^{-1}B^T P(t) x(t), \tag{8.21}$$

and that the value of the cost function will be

$$x_0^T P(t_0) x_0,$$

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; 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); Q_f is symmetric; the right-hand side of the differential equation is symmetric, therefore P also, and in conclusion, $P(t)$ itself must be symmetric. One has also studied the behaviour 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 $D^T D = Q$, the optimal control becomes

$$u^*(t) = -R^{-1}B^T Px(t), \tag{8.22}$$

where now P is the positive definite solution of the **algebraic Riccati equation**

$$-A^T P - PA - Q + PBR^{-1}BP = 0.$$

Note that in both (8.21) and (8.22) the optimal control is given in feedback form; the current control depends on the current state. If (8.22) is substituted in (8.17), the result is

$$\dot{x} = (A - BR^{-1}B^T P)x; \quad x(t_0) = x_0,$$

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

$$\begin{aligned} x(k+1) &= x(k) + Bu(k), \\ y &= Cx + Du, \end{aligned}$$

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 $h(s) = q(s)/p(s)$ 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 a case 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):

$$\begin{aligned} &y(k+n) + p_{n-1}y(k+n-1) + \ldots + p_0 y(k) \\ &\quad = q_{n-1}u(k+n-1) + \ldots + q_0 u(k), \qquad k = 0, 1, 2, \ldots \end{aligned}$$

The parameters $\{p_i\}$ and $\{q_i\}$ 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 $\{p_i\}$ and $\{q_i\}$ parameters. The observations satisfy

$$\begin{aligned} y(k) &= -p_{n-1}y(k-1) - \ldots - p_0 y(k-n) + \\ &\quad q_{n-1}u(k-1) + \ldots + q_0 u(k-n) + \xi(k), \\ &\quad k = n, n+1, \ldots, r, \end{aligned} \tag{8.23}$$

where the quantity $\xi(k)$ denotes a possible perturbation in the system due to measurement errors in $\{y_i\}$ and $\{u_i\}$ for instance. The quantity $\xi(k)$ makes the relation between $y(k)$ and $\{y_i\}$, $\{u_i\}$, $i < k$, again an exact equality. In general $\xi(k)$ will be small. Introduce the following notation:

$$\begin{aligned} \theta &= (p_{n-1}, p_{n-2}, \ldots, p_0, q_{n-1}, q_{n-2}, \ldots, q_0)^T, \\ x(k) &= (-y(k-1), \ldots, -y(k-n), u(k-1), \ldots, u(k-n))^T, \end{aligned}$$

then (8.23) can be written as

$$y(k) = x^T(k)\theta + \xi(k), \quad k = n, n+1, \ldots, r.$$

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;

$$\hat{\theta} = \arg\min_{\theta} \sum_{j=n}^{r} (y(j) - x^T(j)\theta)^2. \tag{8.24}$$

The estimate $\hat{\theta}$ thus defined is called the **least squares estimate.** The summation in (8.24) can be written as

$$S = \sum_{j=n}^{r} (y(j) - x^T(j)\theta)^2 = (Y - X\Theta)^T (Y - X\Theta), \tag{8.25}$$

where

$$\begin{aligned} Y &= (y(n), y(n+1), \ldots, y(r))^T, \\ X &= (x(n), x(n+1), \ldots, x(r))^T. \end{aligned}$$

Note that X is a matrix. Differentiation of (8.25) with respect to θ yields for the minimum

$$(X^T X)\hat{\Theta} = X^T Y.$$

If $X^T X$ is invertible then the least squares estimate can be written as

$$\hat{\Theta} = (X^T X)^{-1} X^T Y.$$

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 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 behaviour. The matrix G is assumed to be known; 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 subtilities. An easier way is to start with a discrete time system:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Gw(k), \\ y(k) &= Cx(k) + v(k). \end{aligned} \tag{8.27}$$

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 equations, are uncorrelated. We also assume that $\{w(k)\}$ and $\{v(k)\}$ 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 $y(k)$;

$$\hat{x}(k+1) = \arg\min_x E\{||x(k+1) - x||^2 | y(0), y(1), \ldots, y(k)\}.$$

$E\{.|.\}$ 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

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + K(k)(y(k) - C\hat{x}(k)). \tag{8.28}$$

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, Eq. (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; $C\hat{x}(k)$ is the predicted value of the output and $y(k)$ 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 $\hat{x}(k+1)$. 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 into 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 behaviour, 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 behaviour of the 'barred' system resembles the behaviour 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 then of the original transfer function. We will only devote a few words on model reduction in state space here.

As an example consider

$$\dot{x} = \begin{pmatrix} -1 & 0 \\ 0 & -10 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix} u, \; y = (1 \;\; 1)x.$$

Intuitively one may approximate this system by the one-dimensonal system

$$\dot{x}_1 = -x_1 + u, \;\; y = x_1;$$

one simply deletes the parts of the system related to the smaller eigenvalue (-10). However, whether it is wise to replace

$$\dot{x} = \begin{pmatrix} -1 & 0 \\ 0 & -10 \end{pmatrix} x + \begin{pmatrix} 1 \\ 10 \end{pmatrix} u, \; y = (1 \;\; 12)x$$

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:**

$$P = \int_0^\infty e^{At} B B^T e^{A^T t} dt,$$

and the **observability Gramian:**

$$Q = \int_0^\infty e^{A^T t} C^T C e^{At} dt.$$

These Gramians are well defined for asymptotically stable systems.

EXERCISE 8.6 *Show that the controllability Gramian P and the observability Gramian Q satisfy, respectively, the Lyapunov equations (compare (4.3))*

$$\begin{aligned} AP + PA^T &= -BB^T, \\ A^T Q + QA &= -C^T C. \end{aligned}$$

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 hardly controllable respectively observable. One can easily show that these eigenvalues are *not* invariant with respect to coordinate transformations. One talks about a **balanced realization** of the

system if the coordinates are chosen in such a way that P and Q are equal, i.e. $P = Q$, and diagonal for the transformed system.

One can also easily show that the eigenvalues of the product PQ are invariant under state space transformations and hence can be viewed as input/output invariants. Assume that $\Re(\lambda_i(A)) < 0$, $\forall i$, then the **Hankel singular values** σ_i of the system are defined as $\sigma_i = \{\lambda_i(PQ)\}^{\frac{1}{2}}$, 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 behaviour 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 fully-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 repect to time (for instance due to aging, 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 these parameters; if these values change markedly from the nominal values, one will change 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 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 δA 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. $||A|| = \{\lambda_{\max}(AA^T)\}^{\frac{1}{2}}$ 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

$$||\Delta A|| \leq a = \frac{1}{\sup_{0 \leq \omega \leq \infty} ||i\omega I - A||},$$

where $i = \sqrt{-1}$ and where I is the unity matrix. The system with feedback, $\dot{x} = (A+\Delta A+BF)x$, is asymptotically stable if $||\Delta A|| \leq (\sup_{0\leq\omega\leq\infty} ||(i\omega I - A - BF)||)^{-1}$. This uncertainty bound can be maximized by choosing an appropriate feedback matrix F;

$$F^* = \arg\min_F(\sup_\omega ||(i\omega I - A - BF)^{-1}||).$$

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 ('Nature') 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 **worst case design** of 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.

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 **(Moving average)**
Consider the moving average exercise from the chapter on linear difference systems. Generate a random sequence of numbers 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 9.2 **(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:

$$P - Q_1 = C\frac{d\Theta}{dt}.$$

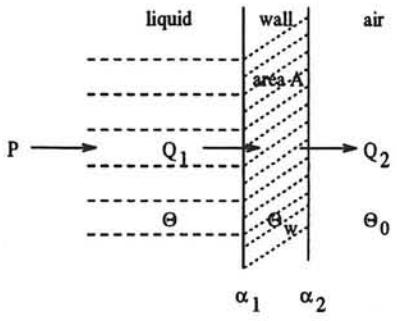


Figure 9.1: Liquid in a barrel.

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 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) $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $A = 3$, $C = 0.4$ en $C_w = 0.2$.
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, both analytically and from the plot, the finally reached equilibrium state x_{eq}.

EXERCISE 9.3 **(Four moving vehicles)**

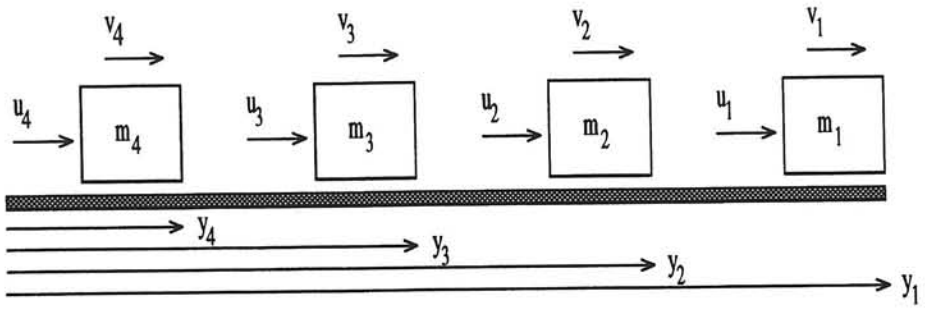


Figure 9.2: Four moving vehicles.

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

$$\begin{aligned} v_i &= \dot{y}_i, \\ u_i &= k v_i + m_i \dot{v}_i. \end{aligned}$$

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

$$\begin{aligned} \overline{y}_{i,i+1}(t) &= y_i(t) - y_{i+1}(t) - h_0, & i &= 1, 2, 3 \\ \overline{v}_i(t) &= v_i(t) - v_0, & i &= 1, 2, 3, 4 \\ \overline{u}_i(t) &= u_i(t) - k v_0, & i &= 1, 2, 3, 4. \end{aligned}$$

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

$$x(t) = (\ \overline{v}_1(t)\ \ \overline{y}_{1,2}(t)\ \ \overline{v}_2(t)\ \ \overline{y}_{2,3}(t)\ \ \overline{v}_3(t)\ \ \overline{y}_{3,4}(t)\ \ \overline{v}_4(t)\ )',$$

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 $x(0) = (\ 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0\ )'$ and simulate over 3 time units. What happens?

EXERCISE 9.4 **(Pole placement of four-vehicle system)**
Consider again Exercise 9.3. 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 $x(0) = (\ 2 \quad 1 \quad 3 \quad 1 \quad 4 \quad 1 \quad 5\ )'$.

EXERCISE 9.5 **(Observer for four-vehicle system)**
Continuation of Exercise 9.3 and Exercise 9.4.

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 $(x - \hat{x})(0) = (\ 2 \quad 1 \quad 3 \quad 1 \quad 4 \quad 1 \quad 5\ )'$.

EXERCISE 9.6 **(From external description to state space)**
Consider the system (from exercise 6.5) 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

$$\frac{d\bar{x}}{dt} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -5 & -4 \end{pmatrix} \bar{x} + \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix} u\,, \quad y = (\ 1 \quad 0 \quad 0\ )\,\bar{x}$$

is identical to the one found in **a**. Determine what quantities the state vector $\bar{x}$ is composed of.

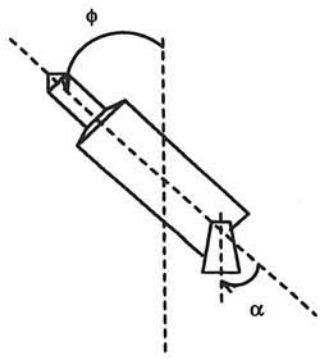


Figure 9.3: Rocket with adjustable engine.

EXERCISE 9.7 **(Rocket)**

For the rocket in Figure 9.3 the simplified equation of motion is

$$I\frac{d^2\phi}{dt^2} = k\alpha,$$

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

$$\frac{\mathcal{L}(\phi)}{\mathcal{L}(\alpha)} = \frac{A}{s^2}.$$

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.

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?

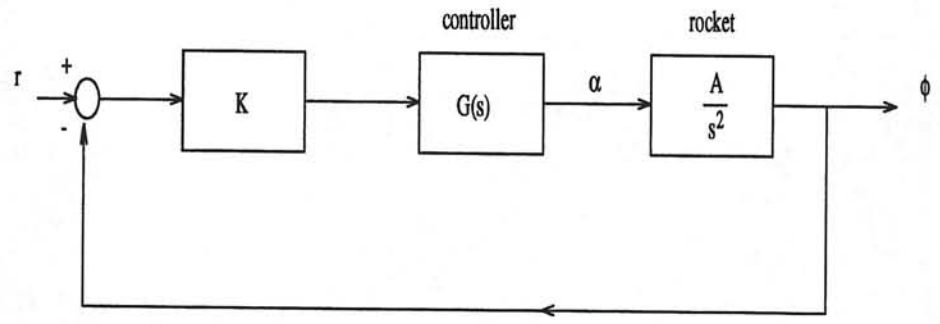


Figure 9.4: Block diagram.

9.2 Solutions

EXERCISE 9.1 **(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) (---)')
```

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 **(Thermal capacity of a wall)**

a.

$$\dot{x} = \begin{pmatrix} -\frac{\alpha_1 A}{C} & \frac{\alpha_1 A}{C} \\ \frac{\alpha_1 A}{C_w} & -\frac{\alpha_1 A}{C_w} - \frac{\alpha_2 A}{C_w} \end{pmatrix} x + \begin{pmatrix} \frac{1}{C} & 0 \\ 0 & \frac{\alpha_2 A}{C_w} \end{pmatrix} u$$

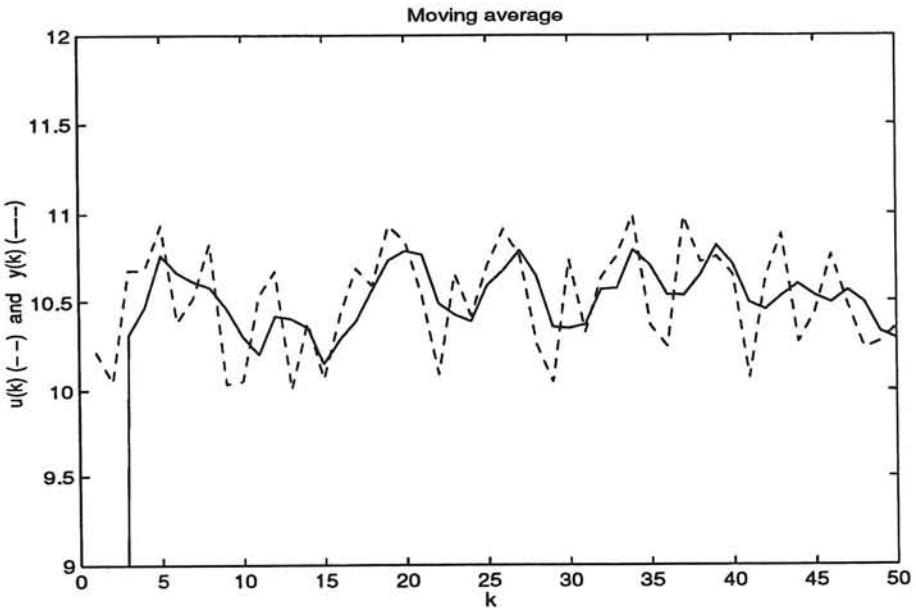


Figure 9.5: Moving average.

$$= \begin{pmatrix} -\frac{3}{4} & \frac{3}{4} \\ \frac{3}{2} & -\frac{9}{2} \end{pmatrix} x + \begin{pmatrix} \frac{5}{2} & 0 \\ 0 & 3 \end{pmatrix} u\,,$$

$$y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} u\,.$$

b.

```
a1=0.1; a2=0.2; area=3; C1=0.4; Cw=0.2;
A=[-a1*area/C1 a1*area/C1 ; a1*area/Cw -a1*area/Cw+
                                        -a2*area/Cw];
B=[1/C1 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 ')
```

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

$$\begin{aligned} \Theta_w = \Theta_o \qquad\qquad\quad & +P(\alpha_2 A)^{-1} = 21\tfrac{2}{3} \\ \Theta = \Theta_o \quad & +P(\alpha_1 A)^{-1} + P(\alpha_2 A)^{-1} = 25 \end{aligned}$$

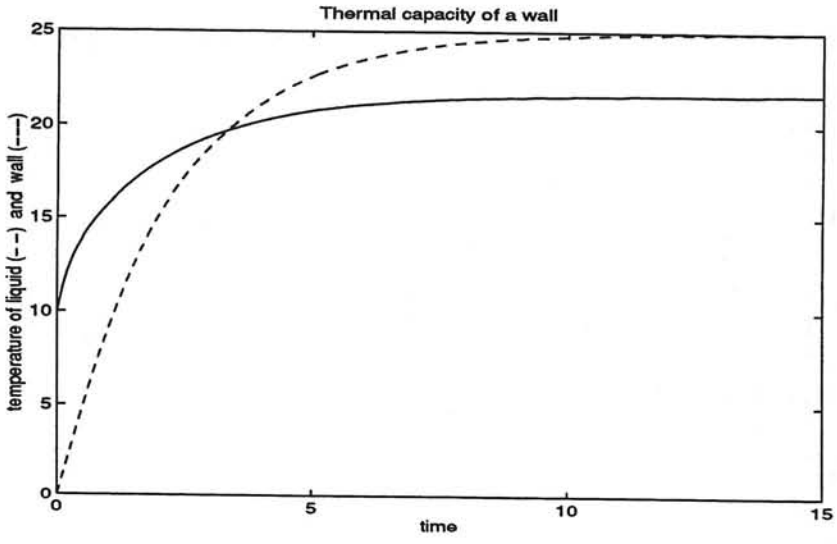


Figure 9.6: Thermal capacity of a wall.

which agrees with the values suggested by the plot in Figure 9.6.

EXERCISE 9.3 **(Four moving vehicles)**

a. The resulting state-space model reads: $\dot{x} = Ax + Bu$, $y = Cx + Du$ with

$$A = \begin{pmatrix} -k/m_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -k/m_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -k/m_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -k/m_4 \end{pmatrix}, \quad (9.1)$$

$$B = \begin{pmatrix} 1/m_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1/m_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/m_3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/m_4 \end{pmatrix}, \quad C = I_7 , \quad D = O_{74} \quad (9.2)$$

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.

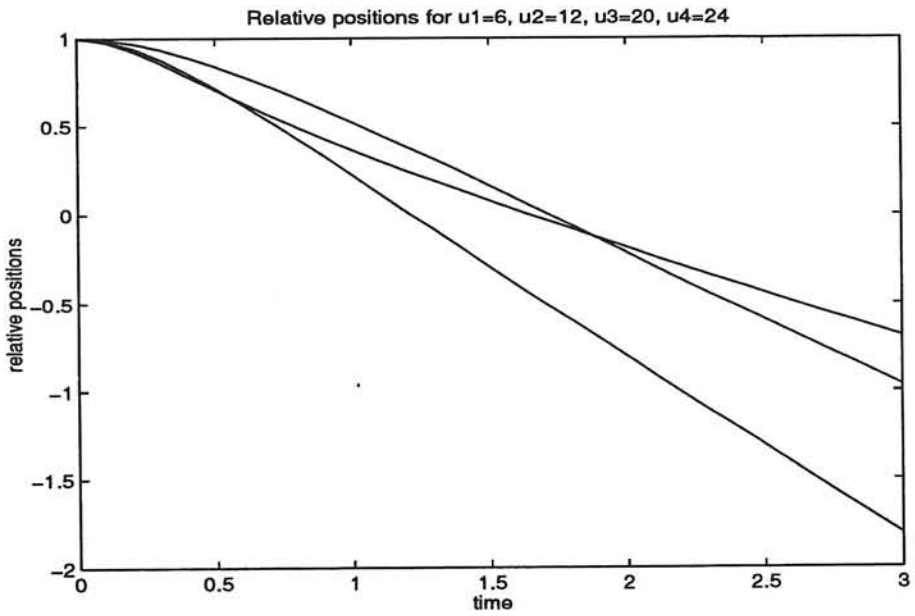


Figure 9.7: Four moving vehicles.

EXERCISE 9.4 **(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
```

```
  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 as a counter anywhere
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.

EXERCISE 9.5 **(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 (4.12) and (4.13):

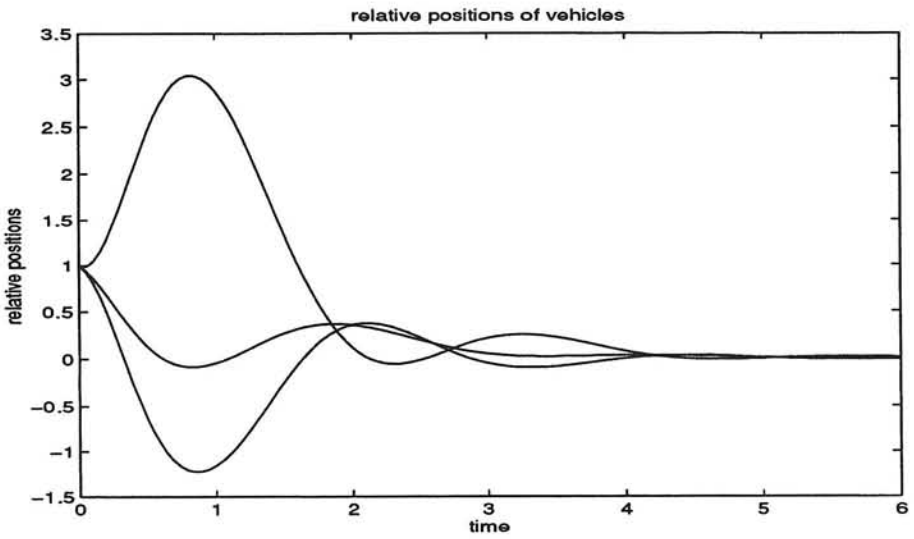


Figure 9.8: Positions in system with feedback, poles at desired locations.

```
[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);
```

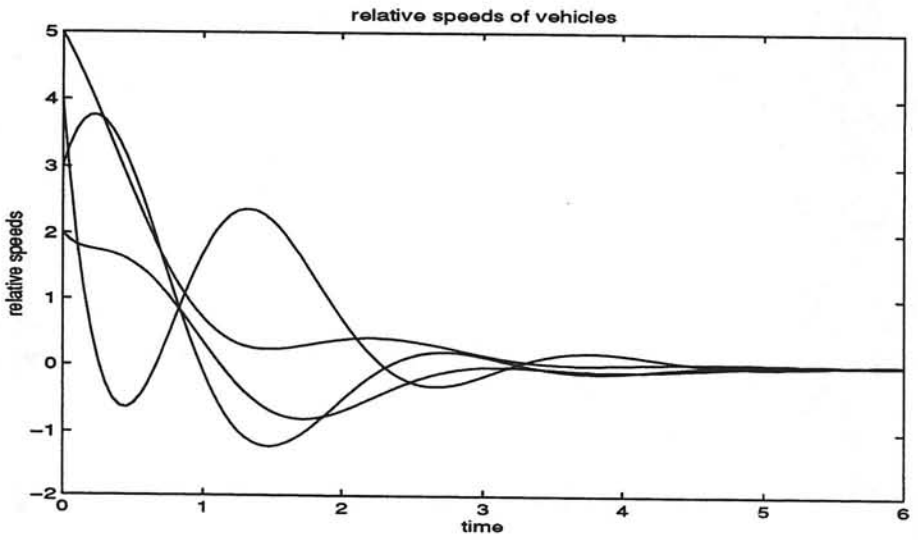


Figure 9.9: Speeds in system with feedback, poles at desired locations.

```
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.

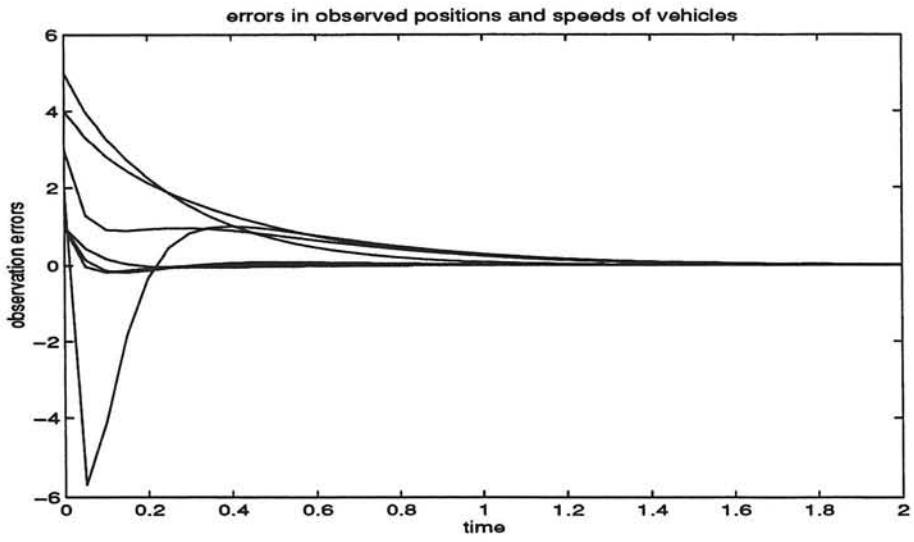


Figure 9.10: Response of four-vehicle system: relative positions observed.

EXERCISE 9.6 **(From external description to state space)**

a. The system is equivalent to $(s^3 + 4s^2 + 5s + 2)Y = (2s^2 + 6s + 5)U$, 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

$$H(s) = \frac{1}{s+2} + \frac{1}{s+1} + \frac{1}{(s+1)^2} \,. \tag{9.3}$$

See the course book (Exercise 6.4) for a flow diagram.

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 description 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 9.7 **(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 $R(s)$:

$$\begin{aligned}
\Phi(s) &= KG(s)\frac{A}{s^2}(R(s) - \Phi(s)) \text{ , so} && (9.8)\\
\Phi(s) &= \frac{KAG(s)/s^2}{1 + KAG(s)/s^2}R(s) \text{ .} && (9.9)
\end{aligned}$$

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. The Routh-Hurwitz criterion applies to the denominator $a_0s^4 + a_1s^3 + a_2s^2 + a_3s + a_4$ of $H(s)$, with $a_0 = 1$, $a_1 = 10$, $a_2 = 20\rho$, $a_3 = 20\rho$ and $a_4 = 5\rho$. Referring to the notation of § 4.1.2 of the course book 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_0, a_1, 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:

```
K=[0,2.7,10,20,28.2,31.4,40,60,100,...   % reference values
   0.4,1,1.8, 4:2:26, 29,30,30.6,31.1,...% and some further
   31.3,31.5,31.8,32.5,35,45,50, 70:10:90]';
                                  % illustrative values
```

```
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')
```

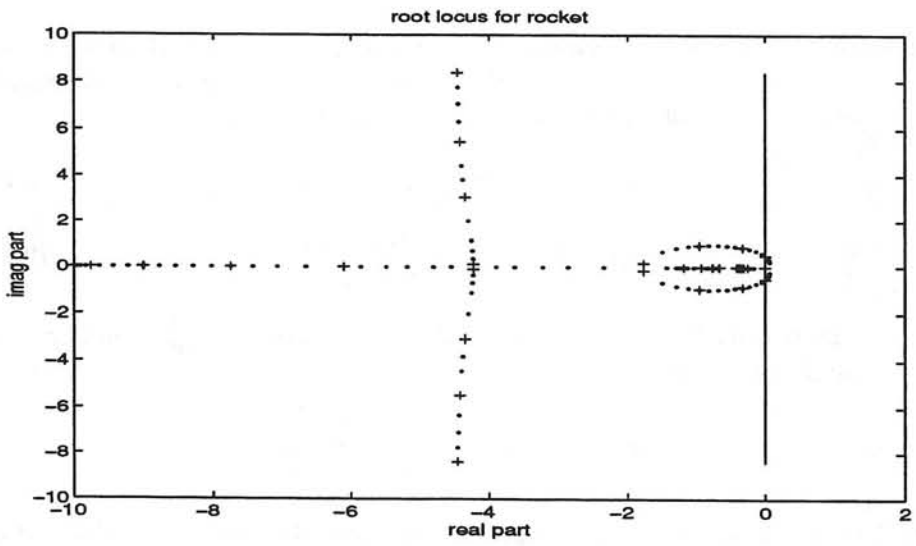


Figure 9.11: Location of closed-loop poles for varying gain K.

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.

Dutch translations of some phrases used

game of goose	ganzenbord
least squares estimate	kleinste kwadraten schatter
white noise	witte ruis
transfer matrix	overdrachtsmatrix
strictly proper rational function	strikt eigenlijk
proper rational function	eigenlijk
pole	pool (polen)
zero	nul (nullen)
factorization	hier: breuksplitsen
least common multiple	kleinste gemene veelvoud
transient behaviour	inschakelgedrag of inloopverschijnsel
gain	versterkingsfactor
low frequency filter	laagdoorlaatfilter
controller	regelaar
track (a signal)	het volgen van een signaal
toaster	broodrooster
stabilizability	stabiliseerbaar
pole-assignment theorem	poolplaatsingsstelling
observer	waarnemer
detectability	detecteerbaarheid
roll angle	rolhoek, de beweging om de lengteas van het schip
lateral velocity	dwarssnelheid
lateral waves	dwarsgolven, golven die van opzij aankomen
equilibrium point	evenwichtspunt
stable	stabiel
controllable	bestuurbaar
reachability	bereikbaarheid
image	beeld
observable	waarneembaar
nonobservable subspace	nietwaarneembare deelruimte

fundamental matrix	fundamentaalmatrix
transition matrix	overgangsmatrix
adjoint system	geadjungeerd systeem
impulse response matrix	impulsrespons(ie) matrix
step response	staprespons(ie)
moving average	bewegend of voortschrijdend gemiddelde
conservation laws	behoudswetten
current	stroomsterkte
torque	moment
moment of inertia	traagheidsmoment
cross product	uitwendig product
prey-predator	prooi-roofdier model
advection	horizontale verplaatsing
diffusion	verspreiding of verstrooiing
linear mapping	lineaire afbeelding
consumer expenditure	consumptieve uitgaven
government expenditure	nationale bestedingen
control	regeling (en niet: contrôle of iets dergelijks)
automatic control	regeltechniek
optimal control	optimale besturingstheorie
filter theory	filtertheorie
yaw	gieren
pitch	stampen
helmsman	roerganger

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] R.E. Bellman. *Introduction to matrix analysis.* McGraw-Hill, 1970.

[4] R.W. Brockett. *Finite dimensional systems.* John Wiley, 1970.

[5] A.E. Bryson and Y.C. Ho. *Applied optimal control.* Blaisdell publishing company, 1969.

[6] C.T. Chen. *Linear system theory and design.* Holt, Rinehart and Winston, 1984.

[7] E.A. Coddington and N. Levinson. *Theory of ordinary differential equations.* McGraw-Hill, 1955.

[8] P. Faurre and M. Depeyrot. *Elements of system theory.* North-Holland, 1977.

[9] Z. Gajíc and M. Lelíc. *Modern Control Systems Engineering.* Prentice Hall, London, 1996.

[10] F.R. Gantmacher. *The theory of matrices.* Chelsea Publishing Company, 1959.

[11] G.H. Golub and C.F. van Loan. *Matrix Calculations.* Johns Hopkins University Press, 1983.

[12] R. Isermann. *Digital Control Systems, Volume I.* Springer Verlag, 1989.

[13] T. Kailath. *Linear systems.* Prentice-Hall, 1980.

[14] D.G. Luenberger. *Introduction to dynamical systems.* John Wiley, 1979.

[15] J.M. Maciejowski. *Multivariable feedback design.* Addison-Wesley Publishing Company, 1989.

[16] H.H. Rosenbrock. *State-space and multivariable theory.* Wiley, New York, 1970.

[17] E.D. Sontag. *Mathematical control theory.* Springer Verlag, 1990.

[18] H.W. Sorenson. *Parameter estimation, principles and problems.* Marcel Dekker, 1980.

[19] J.C. Willems. Paradigms and puzzles in the theory of dynamical systems. *IEEE Transactions on Automatic Control*, 36:259–294, 1991.

[20] W.M. Wonham. *Linear multivariable control: a geometric approach.* Springer Verlag, 1979.

Index

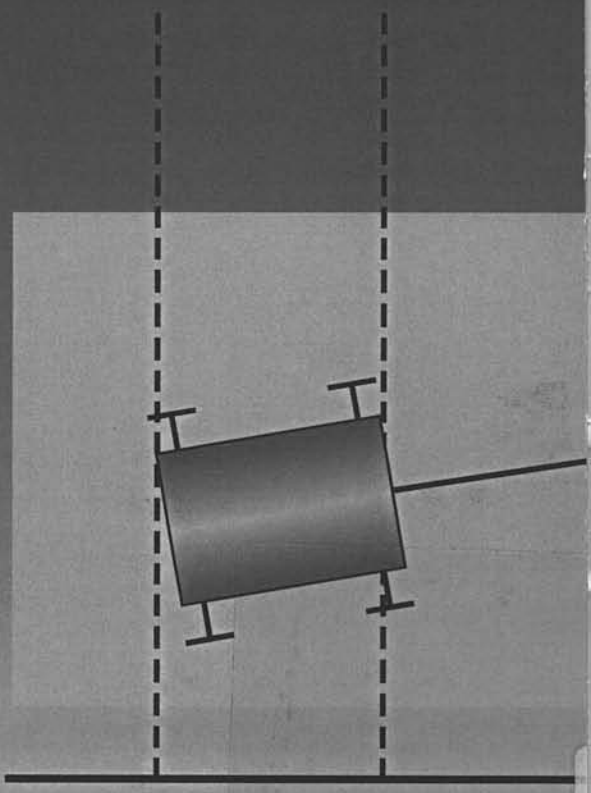
Mathematical Systems Theory

A system is part of reality which we think to be a separate unit within this reality. The reality outside the system is called the surrounding. The interaction between system and surroundings is realized via quantities, which are called input and output. Quite often one wants, through a proper choice of the input, the system to behave in a desired way.

Mathematical Systems Theory is concerned with the study and control of input/output phenomena. The emphasis is on the dynamic behaviour of these phenomena, i.e. how do characteristic features change in time and what are the relationships. These course notes are intended for use at the undergraduate level and form the basis for other courses such as optimal control and filter theory.

URL on this book: http://www.vssd.nl/hlf/a003.htm

Published on behalf of the Vereniging voor Studie- en Studentenbelangen te Delft by Delft University Press.
ISBN 90-407-1272-7



ISBN 90-407-1272-7
9 789040 712722

A003