Systems and Control Theory
An Introduction

A. Astolfi

First draft – September 2006

Contents

Preface

These lecture notes are intended as a reference for a 3rd year, 20 hours "Control Engineering" course at the EEE Department of Imperial College London. In addition, they can be partly used as support material for a 50 hours "Automatic Control" course at the University of Rome Tor Vergata. (The reason why similar topics are taught in such a diverse time range is, and will probably remain, a mistery to me, even after various years of experience with students and courses from both institutions!)

In both cases, this is the second control-like course taken by the students. The main goal of these notes is to provide a self-contained and rigorous background on systems theory and an introduction to state space analysis and design methods for linear systems. In preparing these notes I was deeply influenced by the approach pursued in the book "Teoria dei sistemi", by A. Ruberti and A. Isidori (Boringheri, 1985) and by my research experience on nonlinear control theory. Different approaches can be pursued, and the so-called behavioural approach, developed by J. Willems, is (in my opinion) the best alternative.

These notes are organized as follows. I start giving an abstract mathematical description of the notion of system, illustrated by several pratical examples (Chapter 1). I then move to the study of properties of autonomous systems (Chapter 2), and then to the study of the input-to-state, the state-to-output, and the input-to-output interactions (Chapter 3). Finally, I discuss a few basic design tools (Chapter 4). I try to deal at the same time with continuous- and discrete-time systems, pointing out similarities and differences. The reader should be familiar with standard calculus and linear algebra.

I hope the reader will find these notes a valuable starting point to proceed in more advanced areas of systems and control theory. In particular, these notes should provide the necessary tools for the 4th year control courses and the Control M.Sc. course at Imperial College London. I am aware that there are several excellent books where the same topics are dealt with in detail. The idea of these notes is to provide a condensed, yet precise, introduction to systems and control theory amenable for a short second (even first) undergraduates systems and control course, and not to be a substitute for more in-depth study.

Chapter 1

Introduction

1.1 Introduction

Aim of this chapter is to introduce the notion of system. In the (abstract) definition of system, we follow a simple and natural approach, the so-called input-output approach, which is motivated by the study of simple systems. Among the various mathematical representations we give particular emphasys to the one based on the introduction of an auxiliary variable, the state variable, which is denoted as *state space representation.* This representation plays a fundamental role in systems and control theory, hence we discuss in detail its properties and the conditions under which it can be introduced.

1.2 Examples

In this section we consider a few motivating examples, which are instrumental to introduce the notion of (abstract) system.

1.2.1 Growth of a family of rabbits

The number of pairs of rabbits n months after a single pair begins breeding (and newly born bunnies are assumed to begin breeding when they are two months old) is given by the so-called Fibonacci numbers, which are recursively defined as

$$F_0 = 0 \qquad F_1 = 1 \qquad F_n = F_{n-1} + F_{n-2}. \tag{1.1}$$

Interestingly

- the Fibonacci number F_{n+1} gives the number of ways for 2×1 dominoes to cover a $2 \times n$ checkerboard;
- the Fibonacci number F_{n+2} gives the number of ways of picking a set (including the empty set) from the numbers 1, 2, ... n, without picking two consecutive numbers;
- the probability of not getting two heads in a row in n tosses of a coin is $\frac{F_{n+2}}{2^n}$.

Finally, given a resistor network of 1Ω resistors, each incrementally connected in series or parallel to the preceeding resistors, the net resistance is a rational number having maximum possible denominator equal to F_{n+1}.

This example shows that the same mathematical object models several physical situations or properties. This justifies therefore the study of the feature of the abstract object (1.1) without reference to the real situation that it represents.

1.2.2 Model of an infectious disease

There are several mathematical models which describe the interactions between HIV and immunocytes in human body. The most commonly used model for long-term excitement of the immune response, and hence for medication purposes, is described by the equations

$$\dot{x} = \lambda - dx - \eta\beta xy \qquad \dot{y} = \eta\beta xy - ay - yI, \tag{1.2}$$

where x denotes the population of uninfected CD4 T-helper cell (in a unit volume of blood), y denotes the population of infected CD4 T-helper cell (in a unit volume of blood), I denotes the action of the immune system, and λ, d, β, η and a are positive parameters.

The population of healthy (uninfected) CD4 T-helper cells (produced by thymus) increases at a rate λ, and decreases at a rate dx (since a cell dies naturally). The healthy CD4 T-helper cells are a target of HIV, hence its population decreases proportionally to x and y, because infected CD4 T-helper cells produce the virus, i.e. when a cell is infected it generates new virus. The infected cells die out at a rate ay, increase at a rate proportional to x and y and are affected by the immune system.

In general, and without any medication, the model (1.2) has three main operating conditions. One which corresponds to a healthy patient, one which represents a patient with HIV but not with AIDS, and one which represents a patient in which AIDS dominates. It can be (rigorously) shown that the first two operating conditions are unstable, whereas the third one is stable (formal definitions of stability and instability will be given in Chapter 2). This justifies the difficulty in treating HIV infected patients.

1.2.3 A scholastic population

Consider a three-year course and the problem of modelling the number of students in each year. Let

- $u(k)$ be the number of incoming first year students at time k;
- $y(k)$ be the number of graduated students at time k;
- $x_i(k)$ be the number of students in the i-th year at time k;
- $\alpha_i(k) \in [0, 1]$ be the rate of promotion in the i-th year at time k.

The *behaviour* of the students' population can be described by the equations

$$\begin{array}{rcl} x_1(k+1) & = & (1-\alpha_1(k))x_1(k) + u(k) \\ x_2(k+1) & = & (1-\alpha_2(k))x_2(k) + \alpha_1(k)x_1(k) \\ x_3(k+1) & = & (1-\alpha_3(k))x_3(k) + \alpha_2(k)x_2(k) \\ y(k) & = & \alpha_3(k)x_3(k). \end{array} \tag{1.3}$$

In the ideal situation in which $\alpha_i(k) = 1$ for all k we have that

$$y(k) = u(k-3),$$

which clearly shows that all incoming students are graduated after three years. Finally, in the *extreme* situation in which, $\alpha_i(k) = 0$ for all k and for some i, we have

$$\lim_{k\to\infty} y(k) = 0.$$

1.2.4 A tractor-trailer system

Consider a vehicle (see Figure 1.1) consisting of a wheeled tractor with two rear-drive wheels and a front-steering wheel, towing a trailer, possibly with off-axle hitching. The off-axle length c has to be regarded as a variable with sign, being negative when the joint is in front of the wheel axle, and positive otherwise. L_1 and L_2 are constants depending on the geometry of the vehicle. The longitudinal speed v_1 and the steering angle δ of the tractor can be (independently) manipulated so that the guide-point P_1 follows a desired path with an assigned velocity.

Suppose that the vehicle has to follow, at a given speed, a circular path of radius R_1.

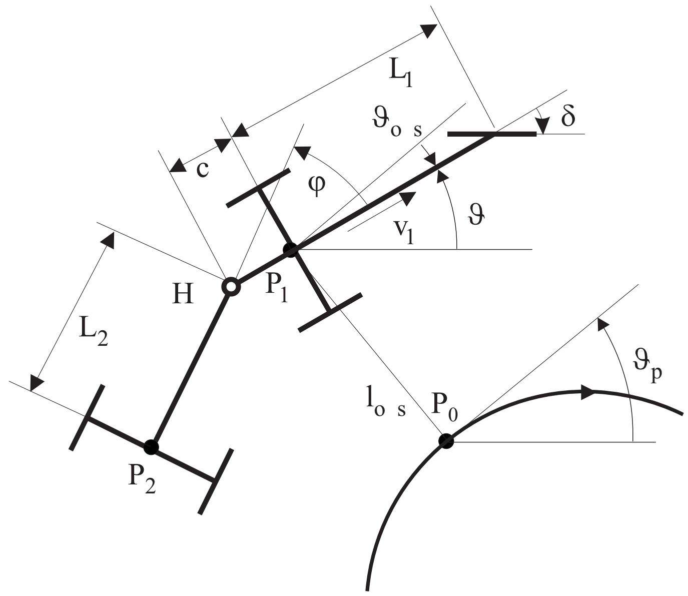


Figure 1.1: Vehicle's geometry and path-tracking offsets l_{os} and ϑ_{os}.

To describe the motion of the vehicle, define l_{os} and ϑ_{os} as the tractor lateral offset and its orientation offset, respectively. They are measured with reference to the projection of the point P_1 of the tractor onto the path. Moreover, let $\varphi_{os} = \varphi - \varphi_p$ be the difference between the current angle φ between tractor and trailer and its steady state value φ_p along the prescribed path. Path-tracking can be viewed as the task of driving these offsets asymptotically to zero.

The offsets are such that

$$\begin{aligned}
\dot{l}_{os} &= -\sigma\,|v_1|\sin\vartheta_{os} \\
\dot{\vartheta}_{os} &= v_1\frac{u}{L_1} - \sigma\,|v_1|\,\frac{\cos\vartheta_{os}}{R_1 + l_{os}} \\
\dot{\varphi}_{os} &= -\frac{v_1}{L_2}\sin\left(\varphi_{os} + \varphi_p\right) - \frac{v_1}{L_1 L_2}\left(c\cos\left(\varphi_{os} + \varphi_p\right) + L_2\right)u,
\end{aligned}$$

where $u = \tan\delta$ is the manipulated variable, and the parameter σ is used to distinguish between counterclockwise ($\sigma = 1$) or clockwise ($\sigma = -1$) directions.

In many applications, the absolute value of the steering angle δ is bounded by a saturation value $\delta_M < \pi/2$.

1.2.5 A simplified atmospheric model

Consider a rectangular slice of air heated from below and cooled from above by edges kept at constant temperatures. This is our atmosphere in its simplest description. The bottom is heated by the earth and the top is cooled by the void of outer space. Within this slice, warm air rises and cool air sinks. In the model, as in the atmosphere, convection cells develop, transferring heat from bottom to top.

The state of the atmosphere in this model can be completely described by three variables, namely the convective flow x, the horizontal temperature distribution y, and the vertical temperature distribution z; by three parameters, namely the ratio of viscosity to thermal conductivity σ, the temperature difference between the top and bottom of the slice ρ, and the width to height ratio of the slice β, and by three differential equations describing the appropriate laws of fluid dynamics, namely

$$\dot{x} = \sigma(y - x) \qquad \dot{y} = \rho x - y - xz \qquad \dot{z} = xy - \beta z. \tag{1.4}$$

These equations were introduced by E.N. Lorenz in 1963, to model the *strange* behaviour of the atmosphere and to justify why weather forecast can be erroneous, and have been recently shown to play an important role on models of lasers and electrical generators. Note that the Lorenz equations are still at the basis of modern weather forecast algorithms.

1.2.6 A perspective vision system

A classical problem in machine vision is to determine the position of an object moving in the three-dimensional space by observing the motion of its projected feature on the two-dimensional image space of a charge-coupled device camera. In this case, the problem of determining the object space coordinates reduces to the problem of estimating the depth (or range) of the object.

The motion of an object undergoing rotation, translation and linear deformation can be described by the equation

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \tag{1.5}$$

where $(x_1, x_2, x_3) \in I\!R^3$ are the coordinates of the object in an inertial reference frame, with x_3 being perpendicular to the camera image space, as shown in Figure 1.2. The parameters a_{ij}, b_i, known as motion parameters, are time-varying and known.

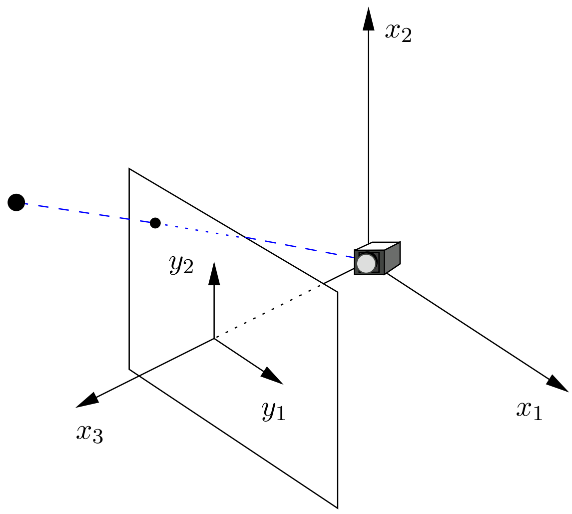


Figure 1.2: Diagram of the perspective vision system.

Using the perspective (or "pinhole") model for the camera, the measurable coordinates on the image space are given by

$$y = \left[\begin{array}{cc} y_1, & y_2 \end{array}\right]' = \epsilon \left[\begin{array}{cc} \frac{x_1}{x_3}, & \frac{x_2}{x_3} \end{array}\right]', \tag{1.6}$$

where ϵ is the focal length of the camera, i.e. the distance between the camera and the origin of the image-space axes.

The perspective estimation problem consists in reconstructing the coordinates x_1, x_2, x_3 from measurements of the image-space coordinates y_1, y_2.

1.2.7 The ABS system

Electronic Anti-lock Braking Systems (ABS) have recently become a standard for all modern cars. ABS can greatly improve the safety of a vehicle in extreme circumstances, as it maximizes the longitudinal tire-road friction while keeping large lateral forces, which guarantee vehicle steerability.

For the preliminary modelling of braking systems, the so-called quarter-car model is used (see Figure 1.3). The model is described by

$$J\dot{\omega} = rF_x - T_b \qquad m\dot{v} = -F_x \tag{1.7}$$

where

- ω is the angular speed of the wheel;
- v is the longitudinal speed of the vehicle body;

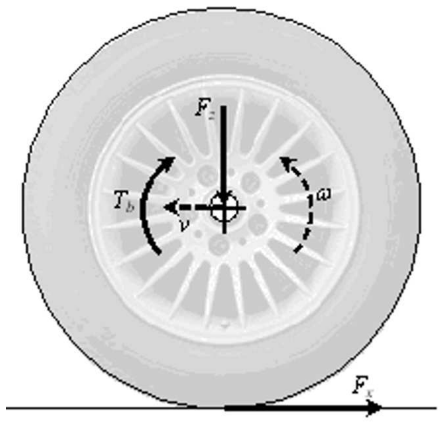


Figure 1.3: Quarter car vehicle model.

- T_b is the braking torque;
- F_x is the longitudinal tire-road contact force;
- J, m and r are the moment of inertia of the wheel, the quarter-car mass, and the wheel radius, respectively.

The dynamic behavior is *hidden* in the expression of F_x, which depends on the variables v and ω, and can be approximated as follows

$$F_x = F_z \mu(\lambda, \beta_t, \theta_r),$$

where

- F_z is the vertical force at the tire-road contact point;
- λ is the longitudinal slip, defined as[1]

$$\lambda = \frac{v - \omega r}{\max\{\omega r, v\}};$$

- β_t is the wheel side-slip angle;
- θ_r is a set of parameters which characterize the shape of the static function $\mu(\lambda, \beta_t; \theta_r)$ and which depend upon the road conditions.

[1]By definition, $\lambda \in [-1, 1]$; during braking, though, as $\omega r \le v$, the wheel slip is defined as $\lambda = \frac{v - \omega r}{v}$ and $\lambda \in [0, 1]$.

1.2.8 A simplified guitar string

Consider a guitar string and assume that it can be modelled by n identical segments (linear lumped springs of unity mass) which interact by means of elastic forces (depending on a *tension* parameter k). Let x_i, $\dot{x}_i$ and $\ddot{x}_i$ be the position, velocity, and acceleration, respectively, of the i-th segment. The string can be described by

$$\begin{array}{rcl} \ddot{x}_1 & = & -k(x_1 - x_2) \\ \ddot{x}_2 & = & -k(x_2 - x_1) - k(x_2 - x_3) \\ & \vdots & \\ \ddot{x}_i & = & -k(x_i - x_{i-1}) - k(x_i - x_{i+1}) \\ & \vdots & \\ \ddot{x}_n & = & -k(x_n - x_{n-1}). \end{array} \tag{1.8}$$

This can be rewritten in compact form as

$$\ddot{x} = \begin{bmatrix} -k & k & 0 & 0 & \dots & 0 \\ k & -2k & k & 0 & \dots & 0 \\ 0 & k & -2k & k & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & k & -2k & k \\ 0 & \dots & 0 & 0 & k & -k \end{bmatrix} x = Ax,$$

where $x = [x_1, x_2, \dots, x_n]'$. The main frequency of oscillation of the string, hence the tune of the string, is a function of the square root of the largest nonzero eigenvalue of A, and this depends on k, hence on the tension on the string.

Remark. A more precise model of a guitar string of length L is given by the so-called one-dimensional wave equation

$$\frac{\partial^2 x(y,t)}{\partial y^2} = \frac{\rho}{T}\frac{\partial^2 x(y,t)}{\partial t^2}, \tag{1.9}$$

where $y \in [0, L]$ denotes the position of a point on the string, $x(y,t)$ the deformation of the point y at time t with respect to the rest position, ρ the mass per unit of length, and T the tension of the string. From this equation it is possible to obtain the given approximate model by considering the finite difference approximation of the derivative, namely

$$\frac{\partial^2 x(y,t)}{\partial y^2} \approx \frac{x(y+h,t) - 2x(y,t) + x(y-h,t)}{h^2}$$

defining the variables

$$x_1 = x(0,t) \quad x_2 = x(h,t) \quad \cdots \quad x_n = x(L,t),$$

and using the constraints

$$x(L+h) - x(L) = 0 \qquad x(0) - x(-h) = 0.$$

Note finally, that the wave equation admits a close solution given by

$$x(y,t) = A \sin \omega_n t \sin \frac{n\pi y}{L},$$

where

$$\omega_n^2 = \frac{n^2\pi^2}{L^2}\frac{T}{\rho}.$$

◇

1.2.9 Approximate discrete-time models

Consider the differential equation

$$\dot{x} = f(x), \tag{1.10}$$

with $x \in I\!R^n$, together with the initial condition $x(0) = x_0$, for some given x_0, and the problem of obtaining a solution $x(t)$, for $t \geq 0$, of such equation. With the exception of very specific examples, it is in general not possible to compute a closed form solution $x(t)$. This implies that $x(t)$ has to be computed numerically, i.e. the idea is to select a sequence of time instants $0 < t_1 < t_2 \cdots$ and to construct a numerical algorithm yielding values $x_1,\ x_2,\ \cdots$ which approximate $x(t_1),\ x(t_2),\ \cdots$. To this end, the simplest possible approach is to consider an equally spaced sequence of time instants, namely

$$\{0, \tau, 2\tau, \cdots, k\tau, \cdots\},$$

where $\tau > 0$ is the so-called sampling-time, and approximate the time derivative with the first difference, namely

$$\dot{x}(k\tau) \approx \frac{x(k\tau + \tau) - x(k\tau)}{\tau}.$$

The differential equation can thus be approximated by

$$\frac{x(k\tau + \tau) - x(k\tau)}{\tau} = f(x(k\tau))$$

yielding the *integration algorithm*

$$x_{k+1} = x_k + \tau f(x_k), \tag{1.11}$$

with $k \geq 0$, and with x_0 given. The equation (1.11) is known as the Euler discrete-time approximation of the differential equation (1.10). The sequence $\{x_k\}$ obtained from the Euler approximate model is such that

$$\|x_k - x(k\tau)\| \leq \tau\psi(k)$$

where $\psi(k)$ is a function of k which is not, in general, bounded. This implies that the use of the Euler approximation yields an error that can be reduced (under certain technical conditions) reducing the sampling interval τ but may become unbounded as $k \to \infty$, i.e. the Euler approximate model cannot be used for long term prediction of the solution of differential equations.

1.2.10 Google page rank

The speed, and success, of Google can be attributed in large part to the efficiency of the search algorithm which, linked with a good hardware architecture, creates an excellent search engine.

The main part of the search engine is *PageRank*^{TM}, a system for ranking web pages developed by Google's founders Larry Page and Sergey Brin at Stanford University.

The main idea of the algorithm is the following. The web can be represented as an oriented (and sparse) graph in which the nodes are the web pages and the oriented paths between nodes are the hyperlinks. The basic idea of PageRank is to *walk* randomly on the graph assigning to each node a *vote* proportional to the frequency of return to the node. If $x_i(k)$ denotes the vote of the node i at time k, one has

$$x_i(k+1) = \sum_{j:j\to i} \frac{x_j(k)}{n_j},$$

where n_j is the number of nodes connected to the node i, for $i = 1, \ldots, N$, where $N \approx$ 3.000.000.000. The graph, representing the web, is not strongly connected, therefore to improve the algorithm for computing the vote, one considers a random jump, with probability p (typically 0.15) to another node (i.e. another page). As a result

$$x_i(k+1) = (1-p) \sum_{j:j\to i} \frac{x_j(k)}{n_j} + p \sum_{j=1}^{N} \frac{x_j(k)}{N}. \tag{1.12}$$

Collecting the variables x_i in a vector x we can rewrite the equations for the votes in the form

$$x(k+1) = Ax(k),$$

for some matrix $A \in I\!R^{N\times N}$. It is not difficult to prove that the matrix A has one eigenvalue equal to one and all other eigenvalues λ_i are such that $|\lambda_i| < 1$. This implies that (we will discuss this issue in detail, after introducing the notion of stability in Chapter 2)

$$\lim_{k\to\infty} x(k) = \bar{x},$$

where all elements $\bar{x}_i$ of $\bar{x}$ are non-negative and bounded. The vector $\bar{x}$ (after a certain normalization) is Google Page Rank. The computation of $\bar{x}$ is numerically very difficult, and it is performed once a month.

1.3 The notion of system

The discussion in Section 1.2 highlights the fact that it is possible to describe the behaviour of several *objects*, natural or artificial, by means of mathematical expressions (differential or difference equations) of diverse forms, and with diverse properties.

The notion of system is thus introduced to provide tools to study such a wide variety of objects on the basis of their mathematical, hence abstract, description. Therefore, by definition, an abstract system is an entity which does not depend upon the physical properties of the associated object. This implies that it is possible to associate the same system to several different objects, and at the same time several systems can be associated to the same object (depending upon the properties that have to be investigated).

We stress that the definition of an abstract notion of system has the advantage that it allows to interprete and study, within a unified framework, diverse phenomena and processes, and provides a unique *language* for several different areas of applications. However, because of its generality, it raises several difficult issues, which can be solved or addressed from several perspectives.

In this notes we give a definition of system which is based on the consideration of the input and output signals. With this in mind, note that the simplest way of *associating* a system to an object is to consider all possible behaviours (as a function of time) of the input signals and of the corresponding output signals. This approach does not depend upon the physical properties of the signals and upon the mechanisms which determine such signals.

Remark. Throughout these notes we assume that the objects under study are deterministic. Similar considerations can be performed in a probabilistic setting. These however require somewhat more sophisticated mathematical tools. ◇

The process of *association* of a system to an object can be regarded as the collection of data from experiments performed on the object thought of as a *black box*. The experiments can be carried out as follows: fix an initial time instant t_0, consider a possible input signal for all $t \geq t_0$ and the corresponding output signals. In this way we collect one or more pairs of functions, denoted as input-output pairs, which are defined for all $t \geq t_0$. Collecting together all such pairs we have a set of input-output pairs, which is used to obtain the definition of system.

In particular, if we consider the set $\mathcal{U}$ of all input signals and the set $\mathcal{Y}$ of all output signals, we have that all input-output pairs determine a relation $\mathcal{S}$ which is such that

$$\mathcal{S} \subset \mathcal{U} \times \mathcal{Y}.$$

This implies that the natural way of giving a formal definition of system is to define an abstract system as a set of relations, where each relation describes all input-output pairs obtained from experiments performed starting from a given time instant.

In particular, consider an ordered subset T of the set $\mathbb{R}$, which is the set of time instants of interest for the system, and define the subset of future time instants[2]

$$F(t_0) = \{t \in T \mid t \geq t_0\};$$

[2] F stands for "future".

the set $U^{F(t_0)}$ of all input functions defined for $t \geq t_0$, and the set $Y^{F(t_0)}$ of all output functions defined for $t \geq t_0$. Then, a relation

$$S_{t_0} \subset U^{F(t_0)} \times Y^{F(t_0)}$$

can be used to describe all experiments, hence all input-ouput pairs, starting at t_0.

From the above discussion we conclude that an abstract system can be defined as the set of all relations S_{t_0} for all $t_0 \in T$. Note however that the sets S_{t_0} and S_{t_1}, for $t_1 > t_0$ are not independent, because we can consider some of the pairs in S_{t_1} as obtained from experiments started at t_0 and disregarding all data for $t < t_1$. A formal definition of system must therefore take this issue into consideration.

Example 1.1 *Consider a synchronous D-type flip-flop with input u and output y. The set T is the set of all time instants in which the clock* goes high, *the set of all input values is $U = [0, 1]$, the set of all output values is $Y = [0, 1]$, and the set of all input and output functions is such that $U^{F(t_0)} = Y^{F(t_0)}$ and this is the set of all sequences of '1's and '0's. To understand the structure of the relation S_{t_0} we have to consider the behaviour of the flip-flop. For example, for some given t_0, the pair of input and output sequences*

$$\{'0011', '0010'\}$$

does not belong to S_{t_0}, whereas the pairs

$$\{'0011', '0001'\} \qquad \{'0011', '1001'\}$$

do. This is consistent with the fact that the output lags the input of one clock cycle, that the initial value of y cannot be determined from the current input and that the value of u at some time instant does not affect y at the same time instant. This implies that all pairs of input and output sequence of the form

$$\{'001\tilde{x}', '\hat{x}001'\}$$

with $\tilde{x}$ and $\hat{x}$ either '1' or '0' belong to S_{t_0}. Note finally that, for a given input sequence we can generate several (two in this example) output sequences.

Definition 1.1 *Consider an ordered subset T of $\mathbb{R}$ and two (non-empty) sets U and Y. An abstract system is a set of relations*

$$\mathcal{S} = \{S_{t_0} \subset U^{F(t_0)} \times Y^{F(t_0)} \mid t_0 \in T\}$$

such that[3] *for all $t_0 \in T$ and for all $t_1 \in F(t_0)$*

$$(u_0, y_0) \in S_{t_0} \Rightarrow (u_0|_{F(t_1)}, y_0|_{F(t_1)}) \in S_{t_1}. \tag{1.13}$$

For this system T is the set of time instants, U the set of values of the input signal, and Y the set of values of the output signal.

[3] $u|_{F(t_1)}$ denotes the restriction of u to $t \in F(t_1)$.

Condition (1.13) implies that the relation S_{t_1} contains all input-output pairs which are obtained *truncating* any other input-output pair originated at a time instant $t_0 \leq t_1$. Note however, that the relation S_{t_1} may contain input-output pairs which cannot be obtained by truncation of other pairs. There is, however, a class of systems, of special interest in applications, for which every relation contains only pairs obtained by truncation. This class is characterized as follows.

Definition 1.2 *A system is uniform if there exists a relation*

$$S \subset U^T \times Y^T \tag{1.14}$$

such that for all $t_0 \in T$

$$(u, y) \in S \Rightarrow (u|_{F(t_0)}, y|_{F(t_0)}) \in S_{t_0}$$

and

$$(u_0, y_0) \in S_{t_0} \Rightarrow \exists (u, y) \in S \ : \ (u|_{F(t_0)}, y|_{F(t_0)}) = (u_0, y_0).$$

A uniform system can therefore be assigned by means of the relation S which, roughly speaking, allows to generate all relations S_{t_0} for $t_0 \in T$.

Example 1.2 *Consider an ideal and instantaneous quantizer, i.e. a device which receives at its input a signal u and delivers at its output a signal y, which is obtained via quantization, with a certain quantization interval q. For such a system $T = I\!R$, $U = I\!R$ and*

$$Y = \{\cdots, -2q, -q, 0, q, 2q, \cdots\}.$$

The relation S_{t_0} is given in Figure 1.4, and it is not hard to argue that the system is uniform.

1.3.1 Parametric representations

The use of relations to represent a system provides a very powerful point of view which is applicable to a very general set of objects. It is however interesting to study if it is possible to determine all input-output pairs by means of functions. To this end, we recall the following basic result.

Lemma 1.1 *Consider two non-empty sets A and B and a relation*

$$R \subset A \times B.$$

Then it is always possible to define a set C and a function[4]

$$f: \ C \times \mathcal{D}R \to \mathcal{R}R$$

[4] $\mathcal{D}R$ denotes the domain of R, and $\mathcal{R}R$ denotes the range of R.

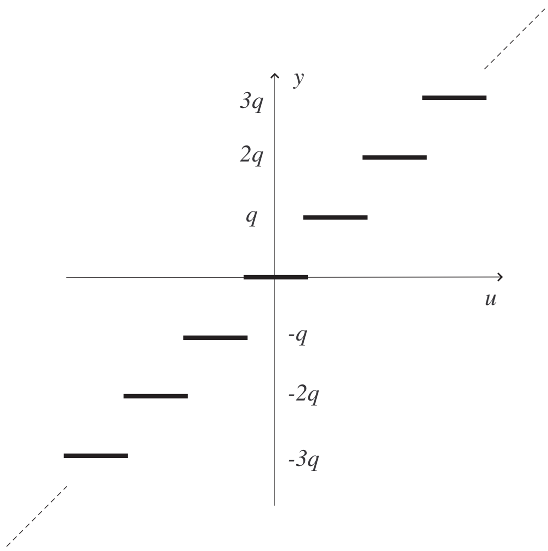


Figure 1.4: The relation S_{t_0} for an ideal quantizer.

such that

$$(a, b) \in R \Rightarrow \exists c \in C \ : \ b = f(c, a)$$

and

$$c \in C, \ a \in \mathcal{D}R \Rightarrow (a, f(c, a)) \in R.$$

This lemma shows that the function f can be used to specify all, and only, the pairs in the relation R. The set C is the set of parameters and the function f is a parametric representation of R. Finally, the association of f and C to R is said *parameterization.*

The result expressed in Lemma 1.1 can be used to obtain parametric representations for a system $\mathcal{S}$. To this end, for any relation S_{t_0} it is possible to perform a parameterization, i.e. it is possible to define a set of parameters X_{t_0} and a function

$$f_{t_0} : \ X_{t_0} \times \mathcal{D}S_{t_0} \to \mathcal{R}S_{t_0}.$$

It is then possible to define a parametric representation of the system $\mathcal{S}$ by means of a set of functions

$$\mathcal{F} = \{f_{t_0} : \ X_{t_0} \times \mathcal{D}S_{t_0} \to \mathcal{R}S_{t_0} \mid t_0 \in T\}.$$

Note that, for uniform systems, described by the relation (1.14), it is enough to consider a single function

$$f : \ X \times \mathcal{D}S \to \mathcal{R}S.$$

Remark. For given $x_0 \in X_{t_0}$ and given $u_0 \in \mathcal{D}S_{t_0}$, the function f_{t_0} can be used to compute the output of the system as $y = f_{t_0}(x_0, u_0)$. ◇

1.3.2 Causal systems

The class of systems considered so far are so-called *oriented*, i.e. there is a natural flow of *information* from the input to the output. This implies that we regard the input as a cause and the output as a consequence. However, these quantities are functions of time. It is therefore natural to study the relationship between the observed effect at time $\bar{t}$ and the time evolution of the causes. Because abstract systems are often used to describe the behaviour of physical objects or processes, it is natural to consider that the above relation be causal, i.e. the output at time $\bar{t}$ should depend only upon the input at times $t < \bar{t}$, or possibly upon the input at times $t \leq \bar{t}$.

It is not easy to formalize this idea to input-output relations. The simplest way is to resort to the parametric representation of the system and to consider the following definition.

Definition 1.3 *A system $\mathcal{S}$ is causal if it possesses at least one parametric representation $\mathcal{F}$ which is such that for all $t_0 \in T$, for all $x_0 \in X_{t_0}$ and for all $\bar{t} \in F(t_0)$*

$$u|_{[t_0,\bar{t}]} = u'|_{[t_0,\bar{t}]} \Rightarrow f_{t_0}(x_0, u)(\bar{t}) = f_{t_0}(x_0, u')(\bar{t}).$$

A system $\mathcal{S}$ is strictly causal if it possesses at least one parametric representation $\mathcal{F}$ which is such that for all $t_0 \in T$, for all $x_0 \in X_{t_0}$ and for all $\bar{t} \in F(t_0)$

$$u|_{[t_0,\bar{t})} = u'|_{[t_0,\bar{t})} \Rightarrow f_{t_0}(x_0, u)(\bar{t}) = f_{t_0}(x_0, u')(\bar{t}).$$

We stress that the difference between the notions of causality and strict causality is only on the *constraint* on u. In the former case u and u' have to be identical for all $t \in [t_0, \bar{t}]$, in the latter for all $t \in [t_0, \bar{t})$.

1.3.3 The notion of state

The crucial point in the definition of abstract system by means of a relation is the fact that, in the relation S_{t_0}, to any given input signal we can associate several output signals. To single out one output signal it is thus necessary to specify, besides the input function, further information. This infomation is associated to the notion of state.

To understand this notion, recall that we have associated, via the process of parameterization, to each relation S_{t_0} a function which associates to an input signal and a parameter $x_0 \in X_{t_0}$ a single output signal. The parameter x_0 may be therefore regarded as the additional piece of information needed to specify the output signal in a unique way. However, to give a formal and precise, hence useful, definition, we have to make sure that the parameterizations performed at each time instant be related in some way, i.e. they cannot be independent but must satisfy a so-called consistency condition.

This consistency property will be discussed in the framework of causal systems.

To begin with, note that the set X_{t_0}, associated with the parameterization S_{t_0}, may be different from the set X_{t_1}, associated with the parameterization S_{t_1}, and so on for all

$t \in T$. It is therefore convenient to define a unique set X such that all sets $X_{\bar{t}}$, with $\bar{t} \in T$, are subsets of X. Consider now an element $x_0 \in X$, thought of as an element of X_{t_0}, which is the set of parameters associated to the parameterization of S_{t_0}, and a second element $x_1 \in X$, thought of as an element of X_{t_1}, which is the set of parameters associated to the parameterization of S_{t_1}, with $t_1 > t_0$. If the system is causal there should be a relation between x_1 and x_0.

In particular, if we assume that x_0 depends only upon the input values for $t < t_0$ and x_1 depends only upon the input values for $t < t_1$, then, recalling that $t_1 > t_0$, it is natural to assume that x_1 depends upon the input values for $t \in [t_0, t_1)$ and upon x_0, and this can be written as

$$x_1 = \phi(t_1, t_0, x_0, u_{[t_0,t_1)})$$

for some function ϕ[5].

The discussion above can be extended to any pair of elements in X, hence it is possible to define the function ϕ for all pairs of elements of X, for all pairs $\{t_0, t_1\}$ such that $t_0 < t_1$, and once t_0 is fixed to all $u \in \mathcal{D}S_{t_0}$.

Finally, for any $t_1 \geq t_0$ the output is computed as

$$y(t_1) = f_{t_0}(x_0, u_{[t_0,t_1]})(t_1).$$

Setting $t_0 = t_1$ in the above relation allows to obtain the relation

$$y(t_1) = \eta(t_1, x_1, u(t_1)),$$

which shows that the output at time t_1 depends (only) upon t_1, the parameter x_1 (which represents the memory of the system) and the value of the input signal at time t_1.

We conclude this discussion noting that, to a causal system, we have associated a space of parameters and two functions which provide an alternative representation of input-output pairs. In fact, for all $t_1 \geq t_0$, we have

$$y(t_1) = \eta(t_1, \phi(t_1, t_0, x_0, u_{[t_0,t_1)}), u(t_1))$$

and this yields, varying x_0 in X and u in $\mathcal{D}S_{t_0}$, all input-output pairs in S_{t_0}.

This representation is characterized by the appearance, a-side the input and output signals, of an auxiliary quantity which takes values in the space X. The role of this quantity is to summarize the effect of past input values, hence to render unique the determination of the present and future output. This quantity is denominated state, or state variable, and the space X is denominated state space.

1.3.4 Definition of state

In the previous section we have informally introduced the notion of state, and we have highlighted its main properties. We now provide a formal definition of state.

[5] With an alternative *convention* we could have obtained $x_1 = \phi(t_1, t_0, x_0, u_{[t_0,t_1]})$.

Definition 1.4 *Given a system $\mathcal{S}$ and a space of input functions $\mathcal{U}$. A set X is a state space for the system $\mathcal{S}$ if there exist two functions*

$$\phi : T \times T \times X \times \mathcal{U} \to X$$
$$\eta : T \times X \times U \to Y$$

such that the following conditions hold.

- *For all $t_0 \in T$, $u \in \mathcal{U}$ and $x_0 \in X$*
$$\{(u_0, y_0) \in U^{F(t_0)} \times Y^{F(t_0)} \; : \; u_0(t) = u(t), y_0(t) = \eta(t, \phi(t, t_0, x_0, u_{[t_0,t)}, u(t))\} = S_{t_0}.$$
- *(Causality) For all $t_0 \in T$, for all $t \geq t_0$ and for all $x_0 \in X$*
$$u|_{[t_0,t)} = u'|_{[t_0,t)} \Rightarrow \phi(t, t_0, x_0, u|_{[t_0,t)}) = \phi(t, t_0, x_0, u'|_{[t_0,t)}).$$
- *(Consistency) For all $t \in T$, and for all $u \in \mathcal{U}$*
$$\phi(t, t, x_0, u) = x_0.$$
- *(Separation)*[6] *For all $t_0 \in T$, for all $t \geq t_0$, for all $x_0 \in X$ and for all $u \in \mathcal{U}$*
$$t > t_1 > t_0 \Rightarrow \phi(t, t_0, x_0, u_{[t_0,t)}) = \phi(t, t_1, \phi(t_1, t_0, x_0, u_{[t_0,t_1)}), u_{[t_1,t)}).$$

The function ϕ is called state transition function, and the function η output transformation. The triple $\{X, \phi, \eta\}$ is called state space representation, or input-state-output representation, of $\mathcal{S}$.

Remark. In what follows, and to simplify the equations, we use the notation $\phi(t, t_0, x_0, u)$ in place of $\phi(t, t_0, x_0, u_{[t_0,t)})$. ◇

Remark. For the special class of systems in which U, Y and X are composed of a finite number of elements, for examples all digital electronics systems, the above definition is equivalent to the definition of a Mealy-type finite state machine. To obtain the definition of a Moore-type finite state machine it is necessary to alter the definition of state, requiring that the state represent the effect of all past and current input values (see Footnote 5). ◇

At this stage, one may wonder if the given definition of state space representation is sufficiently general to develop a systematic theory. To this end, we need to address the issues of existence, and then of unicity of state space representations. While a complete treatment of such topics is outside the scope of these notes, we give a few important results and facts.

[6] This property is sometimes called semigroup property.

Theorem 1.1 *Given a system $\mathcal{S}$ and a space of input functions*[7] $\mathcal{U}$. *The system has a state space representation if and only if it is causal.*

The above statement implies that, under mild technical assumptions, a causal system admits at least one state space representation. However, such representation need not be unique. In fact, given a system $\mathcal{S}$ and a state space representation $\{X, \phi, \eta\}$ it is possible to obtain other state space representations, for example by means of one of the following procedures.

- (State space transformation) Consider a system $\mathcal{S}$, with state space representation $\{X, \phi, \eta\}$. Let $\psi : X \to Z$ be an invertible map, i.e. there exists $\psi^{-1} : Z \to X$ such that $z = \psi(\psi^{-1}(z))$ and $x = \psi^{-1}(\psi(x))$, for all $z \in Z$ and $x \in X$. Define the function
$$\phi_z : T \times T \times Z \times \mathcal{U} \to \tilde{Z}$$
such that
$$\phi_z(t, t_0, z, u_{[t_0,t)}) = \psi(\phi(t, t_0, \psi^{-1}(z), u_{[t_0,t)})$$
and the function
$$\eta_z : T \times Z \times U \to \tilde{Y}$$
such that
$$\eta_z(t, z, u) = \eta(t, \psi^{-1}(z), u).$$
Then $\{Z, \phi_z, \eta_z\}$ is a state space representation of $\mathcal{S}$.

- (State augmentation) Consider a system $\mathcal{S}$, with state space representation $\{X, \phi, \eta\}$. Let $X_a = X \times \tilde{X}$, where $\tilde{X}$ is a non-empty set,
$$\phi_a = \begin{bmatrix} \phi \\ \tilde{\phi} \end{bmatrix},$$
for some function
$$\tilde{\phi} : T \times T \times X_a \times \mathcal{U} \to \tilde{X},$$
and $\eta_a : T \times T \times X_a \times \mathcal{U} \to Y$ is such that $\eta_a - \eta = 0$. Then $\{X_a, \phi_a, \eta_a\}$ is a state space representation of $\mathcal{S}$.

We conclude that if a system $\mathcal{S}$ admits a state space representation, it admits an infinite number of representations. Thus, it makes sense to distinguish between the various representations of a system, and to determine (if possible) a representation which is more convenient, or adequate, for a certain goal. To this end, we conclude this section introducing a few new concepts, which will be useful in the study a state space representations of particular classes of systems.

[7] To be precise, we should assume that the space $\mathcal{U}$ be complete, i.e. that it is closed with respect to concatenation and that, for all $t \in T$, $\{u(t) \in U \ : \ u \in \mathcal{U}\} = U$.

Definition 1.5 (Equivalent representations) *Two state space representations* $\{X, \phi, \eta\}$ *and* $\{X', \phi', \eta'\}$ *of a system* $\mathcal{S}$ *are equivalent if*

- *for all* $t_0 \in T$ *and for all* $x_0 \in X$ *there exists* $x_0' \in X'$ *such that for all* $u \in \mathcal{U}$ *and* $t \geq t_0$
$$\eta(t, \phi(t, t_0, x_0, u), u(t)) = \eta'(t, \phi'(t, t_0, x_0', u), u(t));$$
- *for all* $t_0 \in T$ *and for all* $x_0' \in X'$ *there exists* $x_0 \in X$ *such that for all* $u \in \mathcal{U}$ *and* $t \geq t_0$
$$\eta'(t, \phi'(t, t_0, x_0', u), u(t)) = \eta(t, \phi(t, t_0, x_0, u), u(t)).$$

Definition 1.6 (Equivalent states) *Consider a system* $\mathcal{S}$ *and a state space representation* $\{X, \phi, \eta\}$*. Two elements* x_a *and* x_b *of* X *are equivalent at* t_0 *if for all* $u \in \mathcal{U}$ *and for all* $t \geq t_0$
$$\eta(t, \phi(t, t_0, x_a, u), u(t)) = \eta(t, \phi(t, t_0, x_b, u), u(t)).$$

Remark. Equivalent states are sometimes referred to as non-distinguishable states, because it is not possible to distingh between them by measurements of the output. $\diamond$

Definition 1.7 (Reduced state space) *Consider a system* $\mathcal{S}$ *and a state space representation* $\{X, \phi, \eta\}$*. The state space* X *is said to be reduced at* t_0 *if there are no pairs of states equivalent at* t_0*. If* X *is reduced at* t_0 *the representation* $\{X, \phi, \eta\}$ *is said reduced at* t_0*.*

Remark. A state space can be reduced at some time t_0 but not reduced at some other time t_1. It is possible to give a notion of reduction independent of time requiring that X be reduced at least at one time instant. $\diamond$

1.3.5 Classification

The notion of system introduced is very general. In applications, it is often possible to consider special classes of systems, i.e. to restrict our interest to systems with special properties. To clarify this issue we introduce a classification of systems on the basis of some of the key ingredients discussed.

Definition 1.8 *A system* $\mathcal{S}$ *is a continuous-time system if* $T = \mathbb{R}$*. A system* $\mathcal{S}$ *is a discrete-time system if* $T = Z$*.*

Remark. There is a class of system, increasingly studied and used in applications, in which for some state variables $T = \mathbb{R}$ and for some other state variables $T = Z$, i.e.

some variables vary continuously with time, and other variables vary only at discrete time instants. This is the case in systems where a physical component, for example a robot, is connected with a supervisor, for example a machine that decides which operation the robot has to perform. These systems are denominated hybrid systems. $\diamond$

Definition 1.9 *A system $\mathcal{S}$ is said time-invariant if for all $t_0 \in T$ and for all δ such that $t_0 + \delta \in T$*

$$(u_0(t), y_0(t)) \in S_{t_0} \Rightarrow (u_0(t-\delta), y_0(t-\delta)) \in S_{t_0+\delta}.$$

Definition 1.10 *A state space representation is said time-invariant if for all $t_0 \in T$, all $x_0 \in X$, all $u \in \mathcal{U}$ and all $\bar{t} \in T$*

$$\phi(t, t_0, x_0, u) = \phi(t - t_0, 0, x_0, u) \qquad \eta(t, x, u(t)) = \eta(\bar{t}, x, u(t)).$$

Remark. For a time-invariant representation we can always select $t_0 = 0$. $\diamond$

Definition 1.11 *A system $\mathcal{S}$ is said linear if X, U and Y are linear spaces and if, for all $t_0 \in T$, S_{t_0} is a linear subspace of $U^{F(t_0)} \times Y^{F(t_0)}$. A system S which is not linear is called nonlinear.*

Definition 1.12 *A state space representation is linear if*

- *the sets U, Y and X are linear spaces;*
- *the set $\mathcal{U}$ is a linear subspace of U^T;*
- *for all $t_0 \in T$ and $t \in T$ such that $t \geq t_0$ the function ϕ is linear on $X \times \mathcal{U}$;*
- *for all $t \in T$ the function η is linear on $X \times U$.*

Definition 1.13 *A state space representation is a finite state representation if the sets U, Y and X have a finite number of elements.*

Definition 1.14 *A state space representation is a finite-dimensional representation if the sets U, Y and X are linear, finite-dimensional, spaces.*

1.3.6 Generating functions

In this section we show that, under certain regularity assumptions, it is possible to obtain an alternative description of a system. To begin with, consider discrete-time systems and rewrite the state transition function for $t - t_0 = 1$, i.e.

$$x(t+1) = \phi(t+1, t, x(t), u_{[t,t+1)}) = \phi(t+1, t, x(t), u(t)).$$

This equation shows that for a discrete-time system the value of the state at time $t+1$ depends upon t, $x(t)$ and $u(t)$. We can therefore write

$$x(t+1) = f(t, x(t), u(t)),$$

where the function f is called generating function, or one-step state update function. Note that, from the function f it is possible to reconstruct (uniquely) the state transition function ϕ. Thus, the triple $\{X, f, \eta\}$ can be regarded as the state space representation of a discrete-time system.

It is now natural to wonder if a similar representation can be obtained for continuous-time systems. To this end, consider the generating function of a discrete-time system and note that (if X is a linear space)

$$x(t+1) - x(t) = f(t, x(t), u(t)) - x(t),$$

which shows that the variation of the state in a time unit is a function of t, $x(t)$ and $u(t)$.

This means that, for continuous-time systems, we are looking at the class of systems for which the rate of change of the state $x(t)$ can be written as a function of t, $x(t)$ and $u(t)$. These systems have special importance in applications, where they arise naturally whenever first principles are used to derive their representation (see some of the examples in Section 1.2).

Motivated by these considerations we say that the state space representation $\{X, \phi, \eta\}$ is regular, or differentiable, if there exists a function $f : I\!R \times X \times U \to X$ such that, for any t_0, for any $x_0 \in X$ and for any $u \in \mathcal{U}$ the function $\phi(t, t_0, x_0, u)$ is, for all $t \in F(t_0)$ the (unique) solution of the differential equation

$$\frac{d\phi(t, t_0, x_0, u)}{dt} = f(t, \phi(t, t_0, x_0, u), u(t)), \tag{1.15}$$

with the initial condition

$$\phi(t_0, t_0, x_0, u) = x_0.$$

Note that equation (1.15) can be rewritten as

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

We conclude noting that a regular representation $\{X, \phi, \eta\}$ can be alternatively described by the triple $\{X, f, \eta\}$.

Remark. A state space representation $\{X, f, \eta\}$ is time-invariant in f does not depend explicitely on t and η is as in Definition 1.10. ◇

1.4 Examples revisited

We conclude this chapter by revisiting the examples discussed in Section 1.2 in terms of the concepts, and notions introduced.

- The system (1.1) is a discrete-time, time-invariant, linear, finite-dimensional system without input. To obtain a state space representation $\{X, f, \eta\}$ consider the state variables $x_1 = F_{n-2}$ and $x_2 = F_{n-1}$ and note that

  $$X = \{(x_1, x_2) \in I\!R^2\}$$

  $$f(x) = \begin{bmatrix} x_2 \\ x_1 + x_2 \end{bmatrix}$$

  and $\eta(x) = x_1 + x_2$.

- The system (1.2) is a continuous-time, nonlinear, time-invariant, finite dimensional system, with input $I \in I\!R^+$, and state $(x, y) \in I\!R^+ \times I\!R^+$, described by a generating function. For such a system we have not defined an output transformation.

- The system (1.3) is a discrete-time, nonlinear, finite-dimensional system, with input $u \in I\!R^+$, state $(x_1, x_2, x_3) \in I\!R^+ \times I\!R^+ \times I\!R^+$, and output $y \in I\!R^+$, described by means of a generating function. The system is nonlinear because the input, state and output spaces are not linear spaces.

- The system (1.4) is a continuous-time, nonlinear, time-invariant, finite-dimensional system with input $u \in [-\tan\delta_M, \tan\delta_M]$ and state $(l_{os}, \vartheta_{os}, \varphi_{os}) \in I\!R \times (-\pi, \pi] \times (-\pi, \pi]$, and described by means of a generating function. For such a system we have not defined an output transformation.

- The system (1.4) is a continuous-time, nonlinear, time-invariant, finite-dimensional system, without input, and with state $(x, y, z) \in I\!R^3$, described by means of a generating function. For such a system we have not defined an output transformation.

- The system (1.5)-(1.6) is a continuous-time, nonlinear, finite-dimensional system, without input, with state $(x_1, x_2, x_3) \in I\!R^3$, and output $y \in I\!R^2$, described by a generating function. The system is nonlinear because the output transformation is not linear.

- The system (1.7) is a continuous-time, nonlinear, time-invariant, finite-dimensional system, with input $T_b \in I\!R^+$, and state $(\omega, v) \in I\!R^2$, described by a generating function. The output variable can be selected as the longitudinal slip $\lambda \in [-1, 1]$.

- The system (1.8) is a continuous-time, linear, time-invariant, finite-dimensional system, without input, and with state $(x_1, \cdots, x_n) \in I\!R^n$, described by a generating function. For such a system we have not defined an output transformation.

  The system (1.9) is a continuous-time, time-invariant, infinite-dimensional, linear system without input of with state the set of all functions $x(y,t)$ defined and twice differentiable in $[0,L] \times I\!R$. For such a system we have not defined an output transformation.

- The system (1.10) (resp. (1.11)) is a continuous- (resp. discrete-) time, time-invariant, nonlinear (in general), finite-dimensional system, with state $x \in X$ and without input. For such a system we have not defined an output transformation.

- The system (1.12) is a discrete-time, linear, time-invariant, finite-dimensional system, without input, and with state $x \in I\!R^N$. For such a system the output transformation can be regarded as the identity map.

Chapter 2

Stability

2.1 Introduction

In Chapter 1 we have seen that, under some regularity conditions, continuous- and discrete-time causal systems, with state space X, can be described by means of a generating function and an output transformation, namely

$$\dot{x} = f(t, x, u) \qquad y = \eta(t, x, u) \tag{2.1}$$

and[1]

$$x^+ = f(t, x, u) \qquad y = \eta(t, x, u), \tag{2.2}$$

where all signals have to be understood as evaluated at time t, and $t \in I\!R$ if the system is continuous-time, whereas $t \in Z$ if the system is discrete-time. In what follows, whenever convenient and for compactness, we also use the notation

$$\sigma x = f(t, x, u) \qquad y = \eta(t, x, u), \tag{2.3}$$

where σx stands for $\dot{x}$ if the system is continuous-time, and σx stands for x^+ if the system is discrete-time.

2.2 Existence and unicity of solutions

The simplest question that can be posed in the study of the equations (2.1) and (2.2) is the following.

Given an initial time t_0, an initial value of the state $x(t_0) = x_0$ and an input signal $u \in U^{F(t_0)}$, is it possible to obtain a *solution* of the equation (2.3)? By a solution we mean a function $x(t)$, defined for all $t \geq t_0$, and such that

$$\sigma x(t) = f(t, x(t), u(t))$$

for all[2] $t \in F(t_0)$, or for all $t \in [t_0, \bar{t})$, for some $\bar{t} > t_0$.

[1] To simplify notation we replace $x(t+1)$ with x^+ and $x(t)$ with x.

[2] It is enough to require that the equality holds for almost all t, i.e. the condition may be violated for some $t \in T_s \subset T$, provided that T_s has zero Lebesgue measure. To illustrate this point consider the differential equation

$$\dot{x} = -\text{sign}(x), \tag{2.4}$$

where the signum function is defined as

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x > 0. \end{cases}$$

For a given $x(0) > 0$ we have

$$x(t) = \begin{cases} x(0) - t & \text{for } t \leq x(0) \\ 0 & \text{for } t \geq x(0), \end{cases}$$

which shows that equation (2.4) does not hold for all t, in fact $x(t)$ is not differentiable at $t = x(0)$.

The answer to this question is trivial in the case of discrete-time systems, provided the state space X coincides with $I\!R^n$ and the generating function is continuous. In fact, if the function f is continuous and the input signal $u(t)$ is bounded for all finite $t \in T$, then equation (2.2) describes a continuous mapping from $T \times I\!R^n \times U$ to $I\!R^n$, hence the solution $x(t)$ is unique and it is such that $x(t) \subset X = I\!R^n$ for all finite $t \geq t_0$.

Note however that, if the function f is not continuous, or if the state space X is not $I\!R^n$, then solutions of the equation (2.2) may not be defined for all $t \geq t_0$.

In the case of continuous-time systems the situation is much more involved, and continuity of f is not enough to guarantee existence and uniqueness of solutions of equation (2.1). To understand the issues involved in this problem we start considering two examples.

Example 2.1 *Consider the nonlinear system, without input, described by*

$$\dot{x} = x^3, \tag{2.5}$$

with $x \in I\!R$, and the initial condition $x(0) = x_0$. A simple integration by parts yields

$$x(t) = \frac{x_0}{\sqrt{1 - 2x_0^2 t}},$$

which shows that if $x_0 \neq 0$ then $x(t)$ is defined only for

$$t \in [0, \frac{1}{2x_0^2}).$$

This situation is often referred to as the existence of a finite escape time for the solutions of the differential equation. Note that the time of escape depends upon the initial condition.

Example 2.2 *Consider the nonlinear system, without input, described by*

$$\dot{x} = x^{1/3}, \tag{2.6}$$

with $x \in I\!R$, and the initial condition $x(0) = 0$. Clearly $x(t) = 0$ is a solution of the differential equation for the given initial condition. However

$$x(t) = \left(\frac{2}{3}t\right)^{\frac{3}{2}}$$

is also a solution of the differential equation for the given initial condition.

Remark. The finite escape time phenomenon is not only a mathematical curiosity, but it may naturally arise in applications. For example, consider a mass m, sliding without friction along a line, and acted upon by an external force F. Let x be the position of the mass, and assume that the mass is constrained to stay in the set $I = (-1, 1)$, to avoid contacts with hard physical constraints (this may be the case in a car suspension). By Newton's law, the differential equation of the mass is

$$m\ddot{x} = F,$$

which holds provided $x \in I$. The differential equation can be rewritten in state space representation (see Example 2.8 for a general discussion on high order differential/difference equations) as

$$\dot{x}_1 = x_2 \qquad \dot{x}_2 = \frac{F}{m},$$

where $x_1 = x \in [-1, 1]$ and $x_2 = \dot{x}$. To take into account the presence of the constraints we can define a new variable

$$z = \psi(x) = -\log(1 - x) + \log(1 + x)$$

and note that the function ψ maps the interval $[-1, 1]$ into $I\!R$. Note now that the system can be rewritten as

$$\dot{z} = \frac{(e^z + 1)^2}{2e^z} x_2 \qquad \dot{x}_2 = \frac{F}{m}.$$

Suppose that $F = 0$ for all t, that $x_1(0) = 0$ and that $x_2(0) \neq 0$. Then $x_2(t) = x_2(0)$ for all t and

$$z(t) = \log \frac{1 + x_2(0)t}{1 - x_2(0)t}$$

which shows that $z(t)$ has finite escape time at $t = \frac{1}{|x_2(0)|}$. This is the mathematical description of the obvious fact that the mass, with a nonzero initial velocity, and without any external action, hits one of the constraints in finite time, i.e. at the time of escape of $z(t)$. Note that, even in the presence of a nonzero force, the problem of determining F to avoid hitting the constraints, i.e. to avoid the occurence of finite escape time, is non-trivial. ◇

Example 2.3 *Consider the Euler approximate models of systems (2.5) and (2.6), with sampling time τ, namely*

$$x^+ = x + \tau x^3$$

and

$$x^+ = x + \tau x^{1/3}.$$

Note that these systems do not present the finite escape time phenomenon, neither the existence of multiple solutions. This shows that the numerical integration of nonlinear continuous-time systems may be very difficult and requires, in general, the use of dedicated algorithms.

From the above examples we infer that the issues of existence and unicity of solutions for continuous-time systems may be very delicate. It is possible to show that if the function

f is continuous then, for any x_0, the differential equation (2.1) has at least one solution, defined for all $t \geq t_0$ and such that $t - t_0$ is sufficiently small (local existence). The critical issue is therefore the existence of trajectories for $t \to \infty$ (global existence). It is very hard to characterize such a property, and we therefore discuss only sufficient conditions.

Theorem 2.1 (Existence of solutions for Lipschitz differential equations) *Consider the differential equation*

$$\dot{x} = f(t, x), \tag{2.7}$$

with $x \in I\!R^n$, and the initial condition $x(t_0) = x_0$. Suppose that f is piecewise continuous in t and it is such that the (global) Lipschitz condition

$$\|f(t, x) - f(t, y)\| \leq L\|x - y\|$$

holds for all $x \in I\!R^n$ and $y \in I\!R^n$, and for some constant $L > 0$. Then for any x_0 the differential equation (2.7) has a unique solution defined for all $t \geq t_0$.

As a direct consequence of the above statement we have the following result.

Corollary 2.1 (Existence of solutions for linear differential equations) *Consider the differential equation*

$$\dot{x} = A(t)x + g(t),$$

with $x \in I\!R^n$, and the initial condition $x(t_0) = x_0$. Assume that $A(t)$ and $g(t)$ are piecewise continuous. Then for any $x_0 \in I\!R^n$ the differential equation has a unique solution defined for all $t \geq t_0$.

The Lipschitz condition of Theorem 2.1 is very restrictive, and there are several differential equations for which it does not hold, but which have unique solutions defined for all $t \geq t_0$. Note that existence conditions for non-Lipschitz differential equations may be very involved. We thus restrict our discussion to a simple sufficient condition which is however very important in applications.

Theorem 2.2 *Consider the differential equation (2.7), with $x \in I\!R^n$, and the initial condition $x(t_0) = x_0$. Suppose that f is piecewise continuous in t and it is differentiable*[3] *in x. Suppose, in addition, that there exists a compact set W such that, for all $x_0 \in W$, the solutions of the differential equation remain in W for all $t \geq t_0$. Then for any $x_0 \in W$ the differential equation (2.7) has a unique solution defined for all $t \geq t_0$.*

Example 2.4 *Consider the differential equation*

$$\dot{x} = a(t)x - x^3 + u,$$

[3] Locally Lipschitz is enough.

with $x \in I\!R$, $a(t)$ continuous and such that $|a(t)| \leq 1$, and $u \in [-1, 1]$. We now show that for any $x_0 \in I\!R$ the differential equation admits a unique solutions for all $t \geq t_0$. Note first that, for any fixed $\bar{x} > 2$,

$$x > \bar{x} \Rightarrow \dot{x} < 0 \qquad\qquad x < -\bar{x} \Rightarrow \dot{x} > 0.$$

This implies that, for all $x_0 \in [-\bar{x}, \bar{x}]$, the corresponding solution $x(t)$ is such that $x(t) \in [-\bar{x}, \bar{x}]$, for all $t \geq t_0$. Hence, by Theorem 2.2 we infer that, for any $x_0 \in I\!R$ the differential equation has a unique solution for all $t \geq t_0$.

2.3 Trajectory, motion and equilibrium

We now consider again a system described by means of an input-state-output representation $\{X, \phi, \eta\}$ and define a few typical dynamic behaviours of the system.

Definition 2.1 (Trajectory) *Consider a system $\{X, \phi, \eta\}$. A trajectory is the set*

$$\mathcal{T} = \{x \in X \ : \ x = \phi(t, t_0, x_0, u)\} \subset X,$$

i.e. is the set of points in X reached by the state $x(t)$, for $t \geq t_0$, and for a specific initial state x_0 and input signal u.

Definition 2.2 (Motion) *Consider a system $\{X, \phi, \eta\}$. A motion is the set*

$$\mathcal{M} = \{(t, x(t)) \in T \times X \ : \ t \in F(t_0),\ x(t) = \phi(t, t_0, x_0, u)\} \subset T \times X,$$

i.e. is the set of points in $T \times X$ taken by the pairs $(t, x(t))$, for $t \geq t_0$, and for a specific initial state x_0 and input signal u.

The main differences between a trajectory and a motion are that they leave in different spaces, and the motion is parameterized by t, whereas the trajectory does not contain any information on t. This means that the trajectory provides solely information on the points of the state space X visited by the system during his evolution, whereas the motion specifies in addition when each point has been visited. Note, however, that the (natural) projection of a motion along T yields a trajectory.

Figure 2.1 gives an example of a motion, and of the corresponding trajectory, for a continuous-time system, with $X = I\!R^2$ and $t_0 = 0$.

Definition 2.3 (Equilibrium) *Consider a system $\{X, \phi, \eta\}$. Assume the input u is constant, i.e. $u(t) = u_0$ for all t and for some constant u_0. A state x_e is an equilibrium of the system associated to the input u_0 if*

$$x_e = \phi(t, t_0, x_e, u_0),$$

for all $t \geq t_0$, i.e. an equilibrium is a trajectory composed of a single point.

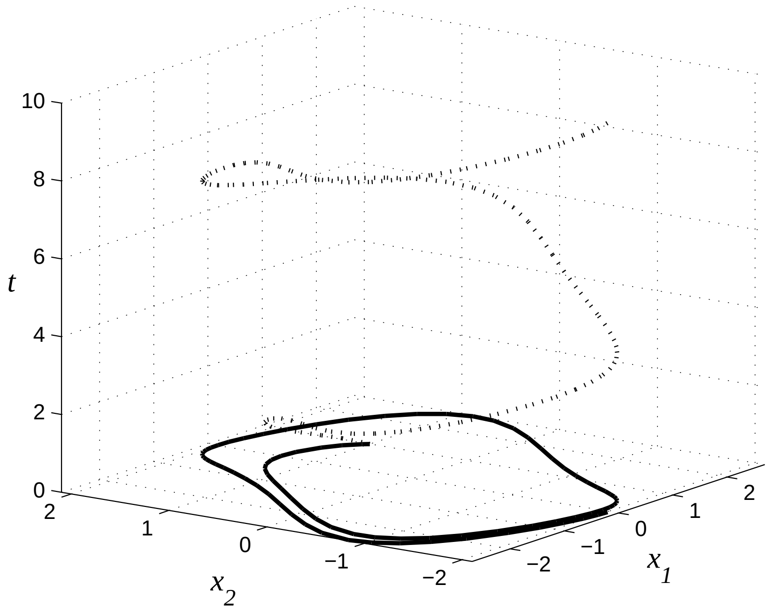


Figure 2.1: A motion (dashed line) and the corresponding trajectory (solid line).

If the system $\{X, \phi, \eta\}$ possesses a generating function, hence can be described by means of the triple $\{X, f, \eta\}$, then the computation of equilibria requires the solution of the system of equations

$$0 = f(t, x_e, u_0),$$

for continuous-time systems, and of the systems of equations

$$x_e = f(t, x_e, u_0),$$

for discrete-time systems.

2.3.1 Linear systems

In this section we discuss the notions introduced for the special class of linear, time-invariant, finite-dimensional systems, i.e. systems described by equations of the form

$$\sigma x = Ax + Bu \qquad y = Cx + Du, \tag{2.8}$$

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$, $y(t) \in I\!R^p$ and $A \in I\!R^{n\times n}$, $B \in I\!R^{n\times m}$, $C \in I\!R^{p\times n}$, and $D \in I\!R^{p\times m}$.

Proposition 2.1 (Equilibria of linear systems) *Consider a linear, time-invariant, system*

$$\sigma x = Ax + Bu,$$

with $x \in I\!R^n$ and $u(t) \in I\!R^m$. The set of equilibria is a linear subspace[4]*. Moreover, the following hold.*

- *For $u(t) = u_0 = 0$, the origin is always an equilibrium.*
- *For continuous-time systems, if A is invertible, for any u_0 there is a unique equilibrium $x_e = -A^{-1}Bu_0$. If A is not invertible the system has either infinitely many equilibria (spanning a linear subspace) or it has no equilibria.*
- *For discrete-time systems, if $I - A$ is invertible, for any u_0 there is a unique equilibrium $x_e = (I - A)^{-1}Bu_0$. If $I - A$ is not invertible the system has either infinitely many equilibria (spanning a linear subspace) or it has no equilibria.*

Proposition 2.2 (Trajectories of linear, continuous-time, systems) *Consider the continuous-time, time-invariant, linear system*

$$\dot{x} = Ax + Bu \qquad\qquad y = Cx + Du,$$

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$, $y(t) \in I\!R^p$ and the initial condition[5] *$x(0) = x_0$. Then*[6]

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \tag{2.9}$$

[4]This property holds also for general linear systems, i.e. not necessarily time-invariant.

[5]Without loss of generality we set $t_0 = 0$.

[6] Given a (square) matrix F, the matrix exponential e^{Ft} is formally defined as

$$e^{Ft} = I + Ft + \frac{(Ft)^2}{2!} + \frac{(Ft)^3}{3!} + \cdots .$$

The matrix exponential has the following properties, which can be derived from its definition.

- For every t_1 and t_2, $e^{Ft_1}e^{Ft_2} = e^{F(t_1+t_2)}$.
- $e^{Ft}e^{\tilde{F}t} = e^{\tilde{F}t}e^{Ft} = e^{(F+\tilde{F})t}$ if and only $F\tilde{F} = \tilde{F}F$, i.e. if and only if F and $\tilde{F}$ commute.
- $(e^{Ft})^{-1} = e^{-Ft}$ and $(e^{Ft})' = e^{F't}$.
- If v is an eigenvector of F with eigenvalue λ then v is also an eigenvector of e^{Ft} with eigenvalue $e^{\lambda t}$.
- $\frac{d}{dt}e^{Ft} = Fe^{Ft} = e^{Ft}F$.
- $T^{-1}e^{Ft}T = e^{(T^{-1}FT)t}$.
- If F is diagonal, i.e. $F = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ then e^{Ft} is also diagonal and $e^{Ft} = \text{diag}(e^{\lambda_1 t}, e^{\lambda_2 t}, \cdots, e^{\lambda_n t})$.

Finally,

$$e^{Ft} = \sum_{i=1}^{r}\sum_{k=1}^{m_i} R_{ik}\frac{t^{k-1}}{(k-1)!}e^{\lambda_i t},$$

for some $m_i \geq 1$ (see Footnote 9), where r is the number of distinct eigenvalues of F.

and

$$y(t) = Ce^{At}x_0 + \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau + Du(t). \tag{2.10}$$

Proof. To begin with consider the differential equation $\dot{x}(t) = Bu(t)$, with $x(0) = x_0$ and note that, by a simple integration,

$$x(t) = x_0 + \int_0^t Bu(\tau)d\tau.$$

Consider now the variable

$$z(t) = e^{-At}x(t)$$

and note that $z(0) = x(0)$, $x(t) = e^{At}z(t)$ and

$$\dot{z} = e^{-At}Bu.$$

Hence

$$z(t) = z_0 + \int_0^t e^{-A\tau}Bu(\tau)d\tau,$$

from which we obtain directly equation (2.9). Equation (2.10) is trivially obtained replacing $x(t)$ in the output transformation. ◁

Example 2.5 *Let $F_\lambda \in \mathbb{R}^{n\times n}$ be defined as*

$$F_\lambda = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix}$$

then

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

To establish the claim note that $F_\lambda = \lambda I + F_0$ and that I and F_0 commutes, hence

$$e^{F_\lambda t} = e^{(\lambda I + F_0)t} = e^{\lambda t}e^{F_0 t},$$

where, by definition of the matrix exponential,

$$e^{F_0 t} = \begin{bmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \cdots & \frac{t^{n-2}}{(n-2)!} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & t \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}.$$

Example 2.6 *Let*

$$F_\lambda = \begin{bmatrix} \lambda & \omega \\ -\omega & \lambda \end{bmatrix}$$

then

$$e^{F_\lambda t} = e^{\lambda t} \begin{bmatrix} \cos \omega t & \sin \omega t \\ -\sin \omega t & \cos \omega t \end{bmatrix}.$$

To establish the claim note that $F_\lambda = \lambda I + F_0$, *and that* I *and* F_0 *commutes, hence*

$$e^{F_\lambda t} = e^{(\lambda I + F_0)t} = e^{\lambda t} e^{F_0 t},$$

where, by definition of the matrix exponential, and recalling the series expansions of $\sin \omega t$ *and* $\cos \omega t$,

$$e^{F_0 t} = \begin{bmatrix} \cos \omega t & \sin \omega t \\ -\sin \omega t & \cos \omega t \end{bmatrix}.$$

Example 2.7 *Consider a continuous-time, time-invariant, linear system with* $x \in I\!R^2$, $u(t) \in I\!R$ *and* $y(t) \in I\!R$. *Suppose that* $x(0) = [1, 1]'$, *that* $u(t) = 0$ *for all* t, *and that*

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Note that

$$A = L^{-1} \tilde{A} L = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}^{-1}$$

hence

$$e^{At} = e^{(L^{-1}\tilde{A}L)t} = L^{-1} \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-2t} \end{bmatrix} L = e^{-t} \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} + e^{-2t} \begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix}.$$

Therefore

$$x(t) = \begin{bmatrix} -2e^{-2t} + 3e^{-t} \\ 4e^{-2t} - 3e^{-t} \end{bmatrix}$$

and $y(t) = -2e^{-2t} + 3e^{-t}$. *Note that, the state and the output are linear combinations of exponential functions with exponents given by* t *times the eigenvalues of* A.

Remark. Continuous-time systems are *reversible*, i.e. the knowledge of $x(t)$ and of the input in the interval $[0, t)$ allows to compute x_0. In fact, from equation (2.9) we obtain

$$x_0 = e^{-At}x(t) - \int_0^t e^{-A\tau}Bu(\tau)d\tau.$$

◇

Proposition 2.3 (Trajectories of linear, discrete-time, systems) *Consider the discrete-time, time-invariant, linear system*

$$x(k+1) = Ax(k) + Bu(k) \qquad\qquad y(k) = Cx(k) + Du(k),$$

with $x \in X = I\!R^n$, $u(k) \in I\!R^m$, $y(k) \in I\!R^p$ *and the initial condition* $x(0) = x_0$. *Then*

$$x(k) = A^k x_0 + \sum_{i=0}^{k-1} A^{k-1-i}Bu(i) \tag{2.11}$$

and

$$y(k) = CA^k x_0 + \sum_{i=0}^{k-1} CA^{k-1-i}Bu(i) + Du(k). \tag{2.12}$$

Proof. Using the state space representation of the system we have that

$$\begin{array}{rcl} x(1) & = & Ax_0 + Bu(0) \\ x(2) & = & Ax(1) + Bu(1) = A^2x_0 + ABu(0) + Bu(1) \\ x(3) & = & Ax(2) + Bu(2) = A^3x_0 + A^2Bu(0) + ABu(1) + Bu(2) \\ & \vdots & \end{array}$$

from which we obtain the expression of $x(k)$. Finally, $y(k)$ is obtained replacing $x(k)$ in the output transformation. ◁

Remark. The expression of $x(k)$ can be rewritten as

$$x(k) = A^k x_0 + \left[\begin{array}{cccc} B & AB & \cdots & A^{k-1}B \end{array}\right] \left[\begin{array}{c} u(k-1) \\ u(k-2) \\ \vdots \\ u(0) \end{array}\right].$$

This expression highlights the role of the matrix $\left[B,\ AB, \cdots,\ A^{k-1}B\right]$ in the computation of $x(k)$. ◇

Remark. If the matrix A is invertible then the system is *reversible*, i.e. the knowledge of $x(k)$ and of the input sequence in the interval $[0, k)$ allows to compute x_0. In fact, from equation (2.11) we obtain

$$x_0 = A^{-k}x(k) - \sum_{i=0}^{k-1} A^{-i-1}Bu(i).$$

◇

Remark. Reversibility of continuous-time systems does not require any assumption on the matrix A. This is because, for any A, the matrix e^{At} is invertible for any t. ◇

Remark. The matrices CA^hB, known as Markov parameters, appearing in equation (2.12) have a simple and interesting interpretation for single-input single-output discrete-time systems. Suppose that $x(0) = 0$, that $u(0) = 1$ and that $u(i) = 0$ for $i \geq 1$. Then

$$y(0) = D \qquad y(1) = CB \qquad y(2) = CAB \qquad \cdots \qquad y(h) = CA^{h-1}B,$$

i.e. the output yields directly information on the matrices A, B, C and D. The problem of determining such matrices, hence a state space representation for the system, from the above output sequence is the so-called realization problem. ◇

It is interesting to interprete the results of the Propositions 2.2 and 2.3 in the light of the general discussion in Chapter 1. Equations (2.9) and (2.11) show that the state of the system at time t is the linear combination of two contributions, the former depends only upon the initial condition x_0, and is denoted free response of the state of the system, the latter depends only upon the input signal u and is denoted forced response of the state of the system. Note that the initial condition and the input can be regarded as two independent causes acting on the system, hence equations (2.9) and (2.11) show that the principle of superposition holds for such systems.

Analogously, equations (2.10) and (2.12) show that the output of the system at time t is the linear combination of two contributions, the former depends only upon the initial condition x_0, and is denoted free response of the output of the system, the latter depends only upon the input signal u and is denoted forced response of the output of the system.

Finally, note that equations (2.9) and (2.10) ((2.11) and (2.12), resp.) yield directly the functions ϕ and η of a state space representation of the system.

Remark. Consider the continuous-time, time-invariant, linear system

$$\dot{x} = Ax + Bu \qquad\qquad y = Cx$$

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$ and $y(t) \in I\!R^p$. Suppose the system is *between* a zero-order hold and a sampler with sampling period T, i.e. the input is constant over each time interval $[kT,(k+1)T)$, for $k \geq 0$, and the output is measured at $t = kT$, for $k \geq 0$. The system viewed *from outside* the zero-order hold and the sampler is a discrete-time, time-invariant, linear system. To obtain a state space representation of this discrete-time system let $x_k = x(kT)$, $u_k = u(kT)$ and $y_k = y(kT)$. Integrating the above differential equation for $t \in [kT,(k+1)T)$ yields (recall equation (2.9))

$$x_{k+1} = e^{AT}x_k + \int_0^T e^{A(T-\tau)}Bd\tau\ u_k = A_d x_k + B_d u_k,$$

whereas the output transformation is given by $y_k = Cx_k$. These equations provide a state space representation of the discrete-time system seen from outside the zero-order hold and the sampler. Note that, unlike the approximate discrete-time models described in Section 1.2.9, the obtained discrete-time model is exact, i.e. under the stated operating conditions $x_k = x(kT)$, for all $k \geq 0$. ◇

Example 2.8 (Input/output models) *Linear, time-invariant, systems can be also described by means of high-order differential or difference equations involving only the* external *signals, i.e. the input and the output. Consider for simplicity single-input single output-systems and the equation*[7]

$$\sigma^n y + a_{n-1}\sigma^{n-1}y + \cdots + a_1\sigma y + a_0 y = b_{n-1}\sigma^{n-1}u + \cdots + b_1\sigma u + b_0 u, \qquad (2.15)$$

where the a_i and b_i are constant coefficients. We now show that this equation defines a linear, time-invariant, system as discussed in Chapter 1. To this end we derive a state space representation with a generating function by means of the following procedure (known as realization).

Assume, for simplicity, that $b_{n-1} = b_{n-2} = \cdots = b_1 = 0$ and that $b_0 \neq 0$. Let $x \in X = I\!R^n$, and define

$$x_1 = y \qquad x_2 = \sigma y \qquad \cdots \qquad x_{n-1} = \sigma^{n-2}y \qquad x_n = \sigma^{n-1}y.$$

Note that

$$\sigma x = \begin{bmatrix} \sigma x_1 \\ \sigma x_2 \\ \vdots \\ \sigma x_{n-1} \\ \sigma x_n \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_n \\ -a_0x_1 - a_1x_2 - \cdots - a_{n-1}x_n + b_0u \end{bmatrix}$$

[7] The *operator* σ^i is such that

$$\sigma^i\alpha(t) = \begin{cases} \alpha(t) & i = 0 \\ \frac{d^i}{dt^i}\alpha(t) & i > 1 \end{cases} \qquad (2.13)$$

for continuous-time systems, and such that

$$\sigma^i\alpha(t) = \begin{cases} \alpha(t) & i = 0 \\ \alpha(t+i) & i > 1 \end{cases} \qquad (2.14)$$

for discrete-time systems. To simplify notation we write α instead of $\sigma^0\alpha$ and $\sigma\alpha$ instead of $\sigma^1\alpha$.

hence

$$\sigma x = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-1} & -a_n \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b_0 \end{bmatrix} u,$$

and

$$y = [1, 0, \cdots, 0]x.$$

The above equations provide the state space representation of the system defined by equation (2.15), under the considered simplifying assumptions. Note that, a similar procedure can be used in the general case, i.e. in the case of nonzero b_i's.

2.4 Linearization

Linear systems can be often used to approximate the behaviour of nonlinear systems around given operating conditions. The procedure that allows to associate a linear system to a nonlinear system together with one of its operating conditions is called linearization.

The advantage in using a linear system to *approximate* a nonlinear system resides clearly in the fact that linear systems are simpler to study, and it is possible to assess several of their properties with simple tests. In addition, it is also often possible to determine local properties of the original nonlinear system, i.e. properties around a certain operating condition, by properties of the linearization.

The process of linearization is performed on a nonlinear system and for a specific initial condition and input signal, i.e. it is assumed that a system, described by equations of the form (2.3), an initial (nominal) state $x(t_0) = x_0$ and a (nominal) input function u_N are assigned.

Suppose, in addition, that the motion of the system, for the given initial state and input signal, is well-defined for all $t \geq t_0$. Let $x_N(t)$ be such that $x_N(t_0) = x_0$ and, for $t \geq t_0$,

$$\sigma x_N = f(t, x_N, u_N),$$

and let

$$y_N(t) = \eta(t, x_N, u_N).$$

$x_N(t)$ and $y_N(t)$ are referred to as the nominal state and the nominal output trajectories.

Consider now a perturbed initial condition and a perturbed input signal, i.e. assume that the system (2.3) has been initialized at time t_0 with $x(t_0) = x_N(0) + \delta_x(t_0)$ and that the input signal is

$$u_P = u_N + \delta_u.$$

Let $x_P(t)$ be such that $x_P(t_0) = x_N(0) + \delta_x(t_0)$ and, for $t \geq t_0$,

$$\sigma x_P = f(t, x_P, u_P),$$

and let

$$y_P(t) = \eta(t, x_P, u_P).$$

$x_P(t)$ and $y_P(t)$ are referred to as the perturbed state and the perturbed output trajectories.

Note now that

$$\delta_x(t) = x_P(t) - x_N(t)$$

is such that

$$\sigma\delta_x = \sigma x_P - \sigma x_N = f(t, x_P, u_P) - f(t, x_N, u_N) = f(t, x_N + \delta_x, u_N + \delta_u) - f(t, x_N, u_N),$$

and define $\delta_y = y_P(t) - y_N(t)$. If the generating function is differentiable, using Taylor series expansion around x_N and u_N, we obtain[8]

$$\begin{aligned}\sigma\delta_x &= f(t, x_N + \delta_x, u_N + \delta_u) - f(t, x_N, u_N) \\ &= f(t, x_N, u_N) + \frac{\partial f(t,x,u)}{\partial x}(t, x_N, u_N)\delta_x + \frac{\partial f(t,x,u)}{\partial u}(t, x_N, u_N)\delta_u + \\ &\quad O(\|\delta_x\|^2 + \|\delta_u\|^2) - f(t, x_N, u_N).\end{aligned}$$

If the perturbations $\delta_x(0)$ and δ_u are sufficiently small, and as long as $\delta_x(t)$ remains small, it is possible to approximate $\sigma\delta_x$ with the *linear* terms of the Taylor series expansion, i.e.

$$\sigma\delta_x \approx \frac{\partial f(t,x,u)}{\partial x}(t, x_N, u_N)\delta_x + \frac{\partial f(t,x,u)}{\partial u}(t, x_N, u_N)\delta_u = A(t)\delta_x + B(t)\delta_u. \quad (2.16)$$

Similarly, if η is differentiable,

$$\begin{aligned}\delta_y &= \eta(t, x_P, u_P) - \eta(t, x_N, u_N) \\ &= \eta(t, x_N, u_N) + \tfrac{\partial \eta(t,x,u)}{\partial x}(t, x_N, u_N)\delta_x + \tfrac{\partial \eta(t,x,u)}{\partial u}(t, x_N, u_N)\delta_u + \\ &\quad O(\|\delta_x\|^2 + \|\delta_u\|^2) - \eta(t, x_N, u_N),\end{aligned}$$

hence δ_y can be approximated, provided δ_u and δ_x are small, with the *linear* terms of the Taylor series expansion, i.e.

$$\delta_y \approx \frac{\partial \eta(t,x,u)}{\partial x}(t, x_N, u_N)\delta_x + \frac{\partial \eta(t,x,u)}{\partial u}(t, x_N, u_N)\delta_u = C(t)\delta_x + D(t)\delta_u. \quad (2.17)$$

In summary, we have shown that, under suitable differentiability assumptions, the behaviour of a nonlinear system around a nominal operating condition can be approximated

[8] The Jacobian of a funcion $g : I\!R^n \to I\!R^m$ at a point $\bar{x}$ is defined as

$$\frac{\partial g(x)}{\partial x}(\bar{x}) = \begin{bmatrix} \frac{\partial g_1(x)}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial g_1(x)}{\partial x_n}(\bar{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(x)}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial g_m(x)}{\partial x_n}(\bar{x}) \end{bmatrix}.$$

considering input, state and output perturbations around the nominal behaviour and such perturbations are the input, state and output variables of a linear, time-varying system.

We stress, once more, that the approximation makes sense only if δ_x and δ_u are small. While it is possible to select δ_u small by selecting u_P *close* to u_N, selecting $\delta_x(t_0)$ small does not guarantee that $\delta_x(t)$ be small for all $t \geq t_0$. We will see in Section 2.7 that there is an important practical situation in which a small $\delta_x(t_0)$ implies that $\delta_x(t)$ remains small for all $t \geq t_0$.

Remark. If we consider, as nominal operating condition, a constant input and an equilibrium point, and if in addition the functions f and η do not depend explicitely on time, then the linearized system is time-invariant, i.e. the matrices $A(t)$, $B(t)$, $C(t)$ and $D(t)$ have constant entries. ◇

Example 2.9 *Consider the nonlinear, time-invariant, system*

$$\dot{x} = \sin x + u \qquad y = \sin 2x,$$

with $x \in I\!R$, $u(t) \in I\!R$ and $y(t) \in I\!R$, and the nominal operating condition $u_N = 0$ and $x_N(0) = 0$, yielding $x_N(t) = 0$ and $y_N(t) = 0$ for all $t \geq 0$. The linearization of the system around the given nominal behaviour is given by

$$\dot{\delta}_x = \frac{\partial(\sin x + u)}{\partial x}(0,0)\delta_x + \frac{\partial(\sin x + u)}{\partial u}(0,0)\delta_u = \delta_x + \delta_u$$

and

$$\delta_y = \frac{\partial \sin 2x}{\partial x}(0,0)\delta_x + \frac{\partial \sin 2x}{\partial u}(0,0)\delta_u = 2\delta_x.$$

Note that, as discussed above, the linearized system is time-invariant.

Consider again the same system but with operating condition $u_N = -1$ and $x_N(0) = \pi/2$, yielding $x_N(t) = \pi/2$ and $y_N(t) = 0$ for all $t \geq 0$. The linearized system is now

$$\dot{\delta}_x = \delta_u \qquad \delta_y = -2\delta_x.$$

Note the (obvious) fact that linearizing a nonlinear system around different operating conditions yields different linearized systems!

Example 2.10 *The problem of minimizing a function $f : I\!R^n \to I\!R$ can be dealt with, if the function is twice differentiable, considering the sequence of points generated, starting from some initial guess x_0 of the minimum, by the so-called Newton's iteration, namely*

$$x^+ = x - \left[\frac{\partial^2 f(x)}{\partial x^2}(x)\right]^{-1} \frac{\partial f(x)}{\partial x}(x). \tag{2.18}$$

This is in general a nonlinear, discrete-time system, and any stationary point of the function f is an equilibrium of this system. Let $\bar{x}$ be a stationary point of f such that the matrix

$$\frac{\partial^2 f(x)}{\partial x^2}(\bar{x}) \tag{2.19}$$

is nonsingular. The linearization of system (2.18) around such stationary point is given by

$$\delta_x^+ = \left(I - \frac{\partial \left[\frac{\partial^2 f(x)}{\partial x^2}(x)\right]^{-1} \frac{\partial f(x)}{\partial x}(x)}{\partial x}(\bar{x}) \right) \delta_x = 0.$$

Example 2.11 *Consider the system*

$$\begin{array}{rcl} \dot{x}_1 & = & g(x_1 x_2) x_1 \\ \dot{x}_2 & = & -2x_2 + x_3 \\ \dot{x}_3 & = & -x_3, \end{array} \tag{2.20}$$

with $x = (x_1, x_2, x_3) \in I\!R^3$, where $g(x_1 x_2)$ is a differentiable function such that $g(1) = 1$. Consider an initial condition $(x_1(0), x_2(0), x_3(0))$ such that $x_1(0) \neq 0$, $x_2(0) = 1/x_1(0)$ and $x_3(0) = x_2(0)$ and note that the corresponding trajectory is

$$x_1(t) = e^t x_1(0) \qquad x_2(t) = \frac{e^{-t}}{x_1(0)} \qquad x_3(t) = \frac{e^{-t}}{x_1(0)}. \tag{2.21}$$

The system linearized around this nominal trajectory is described by

$$\begin{array}{rcl} \dot{\delta}_{x_1} & = & \left(\frac{\partial g(x_1 x_2)}{\partial x_1}(e^t x_1(0), \frac{e^{-t}}{x_1(0)}) + 1 \right) \delta_{x_1} + \frac{\partial g(x_1 x_2)}{\partial x_2} \left(e^t x_1(0), \frac{e^{-t}}{x_1(0)} \right) \delta_{x_2} \\ \dot{\delta}_{x_2} & = & -2\delta_{x_2} + \delta_{x_3} \\ \dot{\delta}_{x_3} & = & -\delta_{x_3}, \end{array} \tag{2.22}$$

and this is not, in general, time-invariant.

2.5 Lyapunov stability

To study the qualitative behaviour of a system, hence to describe the properties of its trajectories for all $t \in T$ and for $t \to \infty$, we introduce the notion of stability. In addition, this notion allows to study the behaviour of trajectories close to an equilibrium point or to a certain motion.

Remark. The notion of stability that we discuss has been introduced in 1882 by the Russian mathematician A.M. Lyapunov, in its Doctoral Thesis (hence it is often referred to as Lyapunov stability). There are other notions of stability due for example to Lagrange (Lagrange stability) or introduced in the past 20 years. Nevertheless, the concept of Lyapunov stability is the most commonly used in applications. ◇

2.5.1 Definition

Consider a system described via a state space representation $\{X, \phi, \eta\}$ and denote with $x(t)$ the value of the function $\phi(t, t_0, x_0, 0)$, i.e. the value of the state at time t when the input is identically zero and $x(t_0) = x_0$. Recall that $x(t)$ describes the so-called free evolution of the system. Suppose that it is possible to define a norm on the space X. (If $X \subset I\!R^n$ the Euclidean norm can be defined.)

Definition 2.4 *[Lyapunov stability] Consider a system $\{X, \phi, \eta\}$ and an equilibrium point x_e. The equilibrium is stable (in the sense of Lyapunov) if for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, t_0) > 0$ such that*

$$\|x(t_0) - x_e\| < \delta$$

implies

$$\|x(t) - x_e\| < \epsilon,$$

for all $t \geq t_0$.

In stability theory, the quantity $x(t_0) - x_e$ is called initial perturbation, and $x(t) = \phi(t, t_0, x_0, 0)$ is called perturbed evolution.

Therefore, an equilibrium x_e is stable if for any neighborhood of x_e (even very small) the perturbed evolution stays within this neighborhood for all initial perturbations belonging to a sufficiently small neighborhood of x_e.

The definition of stability can be interpreted as follows. An equilibrium point x_e is stable if however we select a *tolerable* deviation ϵ, there exists a (sufficiently small) region with the equilibrium x_e in its interior, such that all initial perturbations in this region give rise to trajectories which are within the *tolerable* deviation.

Remark. The constant δ is in general a function of t_0. If it is possible to define a δ which does not depend upon t_0, we say that the equilibrium is uniformly stable. Note that if an equilibrium of a time-invariant system is stable, it is also uniformly stable. $\diamond$

The property of stability dictates a condition on the free evolution of the system for all $t \geq t_0$. Note, however, that in the definition of stability we have not requested that the perturbed evolution converges, for $t \to \infty$, to x_e. This convergence property is very important in applications, as it allows to characterize the situation in which not only the perturbed evolution remains close to the unperturbed evolution, but it also converges to the initial (unperturbed) evolution. To capture this property we introduce a new definition.

Definition 2.5 *[Asymptotic stability] Consider a system $\{X, \phi, \eta\}$ and an equilibrium point x_e. The equilibrium is asymptotically stable if it is stable and if there exists a constant $\delta_a = \delta_a(\epsilon, t_0)$ such that*

$$\|x(t_0) - x_e\| < \delta_a$$

implies

$$\lim_{t\to\infty} \|x(t) - x_e\| = 0. \tag{2.23}$$

In summary, an equilibrium point is asymptotically stable if it is stable and whenever the initial perturbation is inside a certain neighborhood of x_e the perturbed evolution converges, as $t \to \infty$, to the equilibrium point, which is said to be attractive. From a physical point of view this means that all sufficiently small initial perturbations give rise to effects which can be a-priori bounded (stability) and these vanish as $t \to \infty$ (convergence).

It is important to realize that convergence does not imply stability: it is possible to have an equilibrium of a system which is not stable (i.e. it is unstable), yet for all initial perturbations the perturbed evolution converges to the equilibrium.

Remark. To define the notion of uniform asymptotic stability, it is important to understand the role of t_0 in the convergence property. The existence of the limit in equation (2.23) implies that, for any $\kappa > 0$ there is a time t_a such that

$$\|x(t) - x_e\| \le \kappa,$$

for all $t \ge t_a$. The value t_a is a function of κ (consistently with the definition of limit), but may be also function of t_0. Therefore, the convergence property depends upon t_0 in two ways: firstly through the constant $\delta_a(\epsilon, t_0)$ and secondly on the fact that the speed of convergence of $\|x(t) - x_e\|$, which can be measured by $t_a - t_0$, depends upon t_0. If δ_a and $t_a - t_0$ are not function of t_0 then the convergence is uniform. Finally, if the equilibrium is uniformly stable, and if the convergence property is uniform, the equilibrium is uniformly asymptotically stable. ◇

Example 2.12 *Consider the discrete-time system*

$$x_{k+1} = -x_k,$$

with $x \in I\!R$. *This system has the unique equilibrium* $x_e = 0$. *Note that, for any initial condition* $x_0 \in I\!R$ *one has*

$$x_{2i-1} = -x_0 \qquad\qquad x_{2i} = x_0,$$

for all $i \ge 1$. *This implies that the equilibrium is uniformly stable, but not attractive.*

Example 2.13 *Consider the discrete-time system*

$$x^+ = \begin{cases} 2x & \text{if } |x| < 1 \\ 0 & \text{if } |x| \ge 1, \end{cases}$$

with $x \in I\!R$. *This system has the unique equilibrium* $x_e = 0$. *This equilibrium is attractive, i.e. for any initial condition the trajectories converge to* x_e, *but it is not stable.*

Example 2.14 *Consider the continuous-time system*

$$\dot{x}_1 = \psi(t, x_1, x_2)x_2 \qquad \dot{x}_2 = -\psi(t, x_1, x_2)x_1,$$

with $\psi(t, x_1, x_2) > 0$ *for all* (t, x_1, x_2). *The system has the unique equilibrium* $x_e = 0$. *This equilibrium is stable, but not attractive. To see this note that, along the trajectories of the system,*

$$x_1\dot{x}_1 + x_2\dot{x}_2 = 0,$$

and this implies that, along the trajectories of the system, $x_1^2(t) + x_2^2(t)$ *is constant for every* $t \in T$, *i.e.*

$$x_1^2(t) + x_2^2(t) = x_1^2(t_0) + x_2^2(t_0).$$

Therefore the trajectory of the system with initial state $(x_1(t_0), x_2(t_0))$ *remains on a circle of radius* $x_1^2(t_0) + x_2^2(t_0)$ *for all* t, *hence the condition for stability holds with*

$$\delta(\epsilon, t_0) = \epsilon.$$

Finally, because δ *does not depend upon* t_0 *the equilibrium* $x_e = 0$ *is uniformly stable.*

Definition 2.6 *[Global asymptotic stability] Consider a system* $\{X, \phi, \eta\}$ *and an equilibrium point* x_e. *The equilibrium is globally asymptotically stable if it is stable and if, for all* $x(t_0) \in X$,

$$\lim_{t\to\infty} \|x(t) - x_e\| = 0.$$

Remark. The property of global asymptotic stability is very strong, and it has important implications on the structure of the underlying state space realization. For example, in the case of finite-dimensional, time-invariant systems, it implies that $X = I\!R^n$. This fact, which is the consequence of a very delicate theorem of J.W. Milnor, is sometimes informally explained with the sentence "it is not possible to comb the hair on a sphere without leaving a crown somewhere". ◇

Obviously, the property of global asymptotic stability is much stronger than the property of asymptotic stability (which is often referred to as local asymptotic stability), as it requires that the effect of all initial perturbations vanishes as $t \to \infty$.

If an equilibrium x_e is not globally asymptotically stable, it is possible to determine a region of X, containing x_e, such that for all initial conditions in this region the free evolution converges to x_e. This region is known as region of attraction of the equilibrium x_e. Note that if x_e is globally asymptotically stable then its region of attraction coincides with X.

The property of asymptotic stability can be strengthened imposing conditions on the convergence speed of $\|x(t) - x_e\|$.

Definition 2.7 *[Exponential stability] Consider a system* $\{X, \phi, \eta\}$ *and an equilibrium point* x_e*. The equilibrium is exponentially stable if there exists* $\lambda > 0$ *such that for all* $\epsilon > 0$ *there exists a* $\delta = \delta(\epsilon) > 0$ *such that*

$$\|x(t_0) - x_e\| < \delta$$

implies

$$\|x(t) - x_e\| < \epsilon e^{-\lambda(t-t_0)}, \tag{2.24}$$

for all $t \geq t_0$*.*

Remark. The property of exponential stability implies the property of stability and the property of uniform asymptotic stability. ◇

Definition 2.8 *[Stability of motion] Consider a system* $\{X, \phi, \eta\}$ *and a motion*

$$\mathcal{M} = \{(t, x(t)) \in T \times X \ : \ t \in F(t_0), \ x(t) = \phi(t, t_0, x_0, u)\}.$$

The motion is stable if for every $\epsilon > 0$ *there exists a* $\delta = \delta(\epsilon, t_0) > 0$ *such that*

$$\|x(t_0) - x_0\| < \delta$$

implies

$$\|\phi(t, t_0, x(t_0), u) - \phi(t, t_0, x_0, u)\| < \epsilon, \tag{2.25}$$

for all $t \geq t_0$*.*

The notion of stability of motion is substantially similar to the notion of stability of an equilibrium. The important issue is that the time-parameterization is important, i.e. a motion is stable if, for small initial perturbations, for any $t \geq t_0$ the perturbed evolution is close to the non-perturbed evolution. This does not mean that if the perturbed and un-perturbed trajectories are close then the motion is stable: in fact the trajectories may be close, but may be followed with different timing, which means that for some $t \geq t_0$ condition (2.25) may be violated.

Example 2.15 *To illustrate, informally, the fact that an unperturbed and a perturbed trajectory can be close, yet the nominal unperturbed motion is not stable, consider two identical chemical reactions taking place one in the presence of a catalyst and the other without catalyst. The two reactions follow the same trajectory, but with a different time parameterization, i.e one is faster than the other. Then, the distance between the two motions may be larger than any pre-assigned bound.*

2.5.2 Stability of linear systems

The notion of stability relies on the knowledge of the trajectories of the system. As a result, even if this notion is very elegant, and useful in applications, it is in general very hard to assess stability of an equilibrium or of a motion. There are, however, classes of systems for which it is possible to give stability conditions without relying upon the knowledge of the trajectories.

Linear systems belong to one such class. Therefore, in this section, we study the stability of linear systems ad we show that, because of the linear structure, it is possible to assess the properties of stability and attractivity in a simple way.

To begin with, we recall some properties of linear representations.

Proposition 2.4 *Consider a system with a linear state space representation. Then (asymptotic) stability of one motion implies (asymptotic) stability of all motions. In particular, (asymptotic) stability of any motion implies and is implied by (asymptotic) stability of the equilibrium* $x_e = 0$.

Proof. It is enough to prove the second claim. Consider a motion $\phi(t, t_0, x_0, u)$ and note that, by definition, the motion is stable if for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, t_0) > 0$ such that

$$\|x(t_0) - x_0\| < \delta$$

implies

$$\|\phi(t, t_0, x(t_0), u) - \phi(t, t_0, x_0, u)\| < \epsilon,$$

for all $t \geq t_0$. However, by linearity of the state space representation

$$\phi(t, t_0, x(t_0), u) - \phi(t, t_0, x_0, u) = \phi(t, t_0, x(t_0) - x_0, 0).$$

Hence, the motion is stable if and only if the equilibrium $x_e = 0$ is stable. An analogous argument can be used to prove the asymptotic stability claim. ◁

The above statement, together with the result in Proposition 2.1, implies the following important properties.

Proposition 2.5 *If the origin of the linear representation of a system is asymptotically stable, then necessarily, the origin is the only equilibrium of the system for* $u = 0$*. Moreover, asymptotic stability of the zero equilibrium is always global. Finally, uniform asymptotic stability implies exponential stability.*

The above discussion shows that the stability properties of a motion (e.g. an equilibrium) of a linear representation are inherited by all motions of the system. Moreover, for linear representations local properties are always global properties. This means that, with some abuse of terminology, we can refer the stability properties to the linear representation, for example we say that a linear representation is stable to mean that all its motions are stable. Note that it does not make sense to say that a nonlinear representation or a nonlinear system is stable, despite the fact that this terminology is often used!

The stability results discussed in this section apply to general linear representations, in particular it is not necessary to assume that $X = \mathbb{R}^n$, for some $n \geq 1$. However, in this case, i.e. in the case of a finite-dimensional linear representation, it is possible to obtain very simple stability tests.

To derive such tests, note that, by linearity of ϕ and by the finite dimensionality of $X = \mathbb{R}^n$, we have that

$$\phi(t, t_0, x_0, 0) = \Phi(t, t_0)x_0,$$

where $\Phi(t, t_0)$, defined for all $t \geq t_0$, is a square matrix of dimension $n \times n$, known as the state transition matrix. For finite-dimensional linear representations this matrix plays a central role in the study of stability properties.

Proposition 2.6 *A linear, finite-dimensional, representation is stable if and only if*

$$\|\Phi(t, t_0)\| \leq k, \tag{2.26}$$

for all $t \geq t_0$ and for some $k > 0$ possibly dependent on t_0.

Proof. We prove only the sufficient part. For, suppose condition (2.26) holds. Then

$$\|x(t)\| = \|\Phi(t, t_0)x_0\| \leq k\|x_0\|.$$

Therefore, for any $\epsilon > 0$ the selection

$$\delta = \frac{\epsilon}{k}$$

is such that

$$\|x_0\| < \delta$$

implies

$$\|x(t)\| < \epsilon,$$

for all $t \geq t_0$, which is the stability property for the equilibrium $x_e = 0$. ◁

A similar result, with conceptually similar proof, holds with respect to the property of asymptotic stability.

Proposition 2.7 *A linear, finite-dimensional, representation is asymptotically stable if and only if*

$$\|\Phi(t, t_0)\| \leq k,$$

for all $t \geq t_0$, and for some $k > 0$ possibly dependent on t_0, and

$$\lim_{t \to \infty} \Phi(t, t_0) = 0.$$

We conclude this section noting that the above statements can be further simplified, if we assume in addition that the linear representation is time-invariant. To develop these tests, we need the following fact.

Proposition 2.8 *The equilibrium x_0 of a linear, finite-dimensional, time-invariant representation is asymptotically stable if and only if it is attractive.*

This result implies that for linear, finite-dimensional, time-invariant representations attractivity of the zero equilibrium implies stability of the zero equilibrium. This implies that stability is a property of the matrix A (see equation (2.8)). Moreover, as the properties are uniform, attractivity implies uniform asymptotic stability, and hence exponential stability.

Proposition 2.9 *The equilibrium $x_e = 0$ of a linear, finite-dimensional, time-invariant representation is stable if and only if the following conditions hold.*

- *In the case of continuous-time systems, the eigenvalues of A with geometric multiplicity[9] equal to one have non-positive real part, and the eigenvalues of A with geometric multiplicity larger than one have negative real part.*
- *In the case of discrete-time systems, the eigenvalues of A with geometric multiplicity equal to one have modulo not larger than one, and the eigenvalues of A with geometric multiplicity larger than one have modulo smaller than one.*

Proof. Recall that, for the considered class of representations, stability implies and is implied by boundedness of the state transition matrix.

For continuous-time systems the state transition matrix, with $t_0 = 0$, is (see Footnote 6)

$$e^{At} = \sum_{i=1}^{r} \sum_{k=1}^{m_i} R_{ik} \frac{t^{k-1}}{(k-1)!} e^{\lambda_i t},$$

where m_i is the geometric multiplicity of the eigenvalue λ_i. This matrix is bounded if and only if the conditions in the statement hold.

[9] To define the geometric multiplicity of an eigenvalue we need to recall a few facts. Consider a matrix $A \in I\!R^{n\times n}$ and a polynomial $p(\lambda)$. The polynomial $p(\lambda)$ is a zeroing polynomial for A if $p(A) = 0$. Note that, by Cayley-Hamilton Theorem, the characteristic polynomial of A is a zeroing polynomial for A. Among all zeroing polynomials there is a unique monic polynomial $p_M(\lambda)$ with smallest degree. This polynomial is called the minimal polynomial of A. Note that the minimal polynomial of A is a divisor of the characteristic polynomial of A. If A has r distinct eigenvalues λ_1, ..., λ_r then

$$p_M(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_r)^{m_r},$$

where the numbers m_i denote by definition the geometric multiplicity of λ_i. This means that the geometric multiplicity of λ_i equals the multiplicity of λ_i as a root of $p_M(\lambda)$. Recall, finally, that the multiplicity of λ_i as a root of the characteristic polynomial is called algebraic multiplicity.

Similarly, for discrete-time systems, the state transition matrix, for $t_0 = 0$ and $t \geq 1$, is

$$A^t = \sum_{i=1}^{r} \sum_{k=1}^{m_i} R_{ik} \frac{t^{k-1}}{(k-1)!} \lambda_i^{t-k+1},$$

and this is bounded if and only if the conditions in the statement hold. ◁

Proposition 2.10 *The equilibrium $x_e = 0$ of a linear, finite-dimensional, time-invariant representation is asymptotically stable if and only if the following conditions hold.*

- *In the case of continuous-time systems, the eigenvalues of A have all negative real part.*
- *In the case of discrete-time systems, the eigenvalues of A have all modulo smaller than one.*

Proof. The proof is similar to the one of the previous proposition, once it is noted that, for the considered class of representations, asymptotic stability implies and is implied by boundedness and convergence of the state transition matrix. ◁

We conclude this discussion with an alternative characterization of asymptotic stability, the proof of which is outside the scope of these lecture notes.

Proposition 2.11 *The equilibrium $x_e = 0$ of a linear, finite-dimensional, time-invariant representation is asymptotically stable if and only if the following conditions hold.*

- *In the case of continuous-time systems, there exists a positive definite matrix*[10] $P = P'$ *such that*

$$A'P + PA < 0.$$

[10] A square and symmetric matrix $P \in I\!R^{n \times n}$ is positive definite, denoted $P > 0$, if

$$v'Pv > 0,$$

for all nonzero vectors $v \in I\!R^n$. Note that the symmetry condition is without loss of generality. In fact, a nonsymmetric matrix M is the sum of a symmetric matrix P and an anti-symmetric matrix Q. Hence

$$v'Mv = v'(P+Q)v = v'Pv.$$

To test positivity of a symmetric matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{12} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{13} & p_{23} & p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \vdots \ddots & \vdots & \\ p_{1n} & p_{2n} & p_{3n} & \cdots & p_{nn} \end{bmatrix}$$

- *In the case of discrete-time systems, there exists a positive definite matrix $P = P'$ such that*
$$A'PA - P < 0.$$

To complete our discussion we stress that stability properties are associated to the specific state space representation of the system that we consider. Thus, another state space representation of the same system may have different stability properties. Nevertheless, representations related by a change of coordinates with specific properties have the same stability properties. We discuss this issue with reference to linear representations and linear change of coordinates. To this end, consider a change of coordinates described by

$$x(t) = L(t)\hat{x}(t), \tag{2.27}$$

with $L(t)$ invertible for all t, and note that the state transition matrix of the representation with state $\hat{x}$ is given by
$$\hat{\Phi}(t, t_0) = L^{-1}(t)\Phi(t, t_0)L(t_0).$$
As a consequence, the following results hold.

Proposition 2.12 *Consider a state space representation of a linear, finite-dimensional system, and assume it is (asymptotically) stable. Then any representation obtained by means of a change of variable of the form (2.27) is (asymptotically) stable if and only if*

$$\|L(t)\| \leq k_1 \qquad \|L^{-1}(t)\| \leq k_2, \tag{2.28}$$

for some constants k_1 and k_2, and for all t.

Corollary 2.2 *Consider a state space representation of a linear, finite-dimensional, time-invariant system, and assume it is (asymptotically) stable. Then any representation obtained by means of a change of variable of the form (2.27) with $L(t)$ constant and invertible is (asymptotically) stable.*

Proof. It is enough to note that if $L(t)$ is constant and invertible then condition (2.28) holds. ◁

Example 2.16 *To show the importance of the bounds (2.28) consider the system described by*
$$\dot{x} = -x,$$

we could use the Sylvester test, which states that $P = P' > 0$ if and only if

$$p_{11} > 0 \qquad \begin{vmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{vmatrix} > 0 \qquad \begin{vmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{vmatrix} > 0 \quad \cdots \quad \det P > 0.$$

Note finally, that a matrix P is negative definite, denoted $P < 0$ is $-P > 0$.

with $x \in I\!R$, which is asymptotically stable, and the change of coordinates

$$\hat{x}(t) = e^{\alpha t} x(t).$$

Note that, for $\alpha \neq 0$, this change of coordinates is such that conditions (2.28) do not hold. The state space representation with the state variable $\hat{x}$ is given by

$$\dot{\hat{x}} = (\alpha - 1)\hat{x},$$

and its stability depends upon the value of α!

2.6 Coordinates transformations

Because of the importance that it has in the forthcoming chapters, we elaborate on the operation of coordinates transformation. This operation is often used to simplify the state space representation of a given system or to highlight some specific property.

Consider a continuous-time, finite-dimensional, linear system described by the equations

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

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$, $y(t) \in I\!R^p$, and the change of coordinates

$$x(t) = L(t)\hat{x}(t), \tag{2.29}$$

with $L(t)$ invertible for all t. The state space representation in the new coordinates is given by

$$\begin{aligned} \dot{\hat{x}} = & \left[L^{-1}(t)(A(t)x + B(t)u) + \dot{L}^{-1}(t)x\right]_{x=L(t)\hat{x}} \\ = & (L(t)^{-1}A(t)L(t) + \dot{L}^{-1}(t)L(t))\hat{x} + L^{-1}(t)B(t)u, \end{aligned}$$

and

$$y = CL(t)\hat{x} + D(t)u.$$

Similarly, for a a discrete-time, finite-dimensional, linear system described by the equations

$$x^+ = A(t)x + B(t)u \qquad\qquad y = C(t)x + D(t)u,$$

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$, $y(t) \in I\!R^p$, and the change of coordinates

$$x(t) = L(t)\hat{x}(t),$$

with $L(t)$ invertible for all t, the state space representation in the new coordinates is given by

$$\hat{x}^+ = L^{-1}(t+1)A(t)L(t)\hat{x} + L^{-1}(t+1)B(t)u,$$

and

$$y = CL(t)\hat{x} + D(t)u.$$

In the case of time-invariant systems, we focus on time-invariant coordinates transformations, i.e. $x = L\hat{x}$, with L invertible. The transformed system is described by

$$\sigma\hat{x} = L^{-1}AL\hat{x} + L^{-1}Bu \qquad\qquad y = CL\hat{x} + Du.$$

The matrices A and $L^{-1}AL$ have the same eigenvalues, and the same characteristic and minimal polynomials, and they are said to be similar.

Motivated by this discussion we introduce the following definition.

Definition 2.9 (Algebraic equivalent systems) *The state space representations*

$$\sigma x = Ax + Bu \qquad\qquad y = Cx + Du$$

with $x \in I\!R^n$, $u(t) \in I\!R^m$, $y(t) \in I\!R^p$, *and*

$$\sigma\hat{x} = \hat{A}\hat{x} + \hat{B}\hat{u} \qquad\qquad \hat{y} = \hat{C}\hat{x} + \hat{D}\hat{u}$$

with $\hat{x} \in I\!R^n$, $\hat{u}(t) \in I\!R^m$, $\hat{y}(t) \in I\!R^p$, *are algebraically equivalent if there exists a nonsingular matrix* L *such that*

$$\hat{A} = L^{-1}AL \qquad \hat{B} = L^{-1}B \qquad \hat{C} = CL \qquad \hat{D} = D.$$

Note that, for algebraically equivalent representations, one has

$$CA^iB = \hat{C}\hat{A}^i\hat{B}$$

and

$$Ce^{At}B = \hat{C}e^{\hat{A}t}\hat{B}.$$

Example 2.17 (Lyapunov transformation) *Consider a system, without input, described by*

$$\dot{x} = A(t)x, \tag{2.30}$$

with $x \in I\!R^n$, *and* $A(t)$ *periodic of period* T, *i.e.* $A(t+T) = A(t)$, *for all* $t \in I\!R$. *A coordinates transformation* $x(t) = L(t)\hat{x}(t)$, *with* $L(t)$ *periodic of period* T, *differentiable, bounded and nonsingular for every* t, *and with bounded inverse, is called a Lyapunov transformation if the matrix*

$$\tilde{A} = L^{-1}(t)A(t)L(t) + \dot{L}^{-1}(t)L(t)$$

has constant entries. Note that in general, for a given periodic matrix $A(t)$ *it is not possible to decide on the existence of a Lyapunov transformation. However, if such a transformation does exist, then it is possible to decide stability of the system (2.30) by computing the eigenvalues of* $\tilde{A}$.

2.7 Stability in the first approximation

In Section 2.4 we have seen that it is possible to approximatly describe a nonlinear system around a given operating condition by means of a linear system. This approximation makes sense only if the *deviation variables* δ_u and δ_x are small for all t, i.e. only if δ_u is selected sufficiently small and if a sort of stability property holds for δ_x.

In this section we clarify this issue. In particular we show that, in several cases of practical interest, from properties of the linearized model it is possible to infer properties of the nominal motion of the nonlinear system.

Proposition 2.13 *Consider a nonlinear, time-invariant, system*

$$\sigma x = f(x),$$

with $x \in X$*. Let* x_e *be an equilibrum of the system and let*

$$\sigma \delta_x = A\delta_x$$

be the corresponding linearized system.

- *In the case of continuous-time systems, if* A *is non-singular then the equilibrium* x_e *is isolated.*
- *In the case of discrete-time systems, if* $A - I$ *is non-singular then the equilibrium* x_e *is isolated.*

Remark. An equilibrum of a nonlinear system may be isolated even if the matrix A, or $A - I$, which describes the linearized system around the equilibrium is singular. For example, consider the discrete-time system

$$x_1^+ = x_1 + x_2^2 \qquad\qquad x_2^+ = x_2 + x_1^3,$$

and the equilibrium $(x_1, x_2) = (0, 0)$. Note that this is the unique equilibrium of the system. Nevertheles, the matrix A of the linearized system around the equilibrium is equal to the identity, hence $A - I$ is identically equal to zero, hence it is singular. $\diamond$

The result in Proposition 2.13 is important because only isolated equilibria may be asymptotically stable, hence it provides a necessary condition for asymptotic stability of the equilibrium of the nonlinear system. Sufficient conditions can be also developed, as given in the following statement.

Proposition 2.14 *Consider a nonlinear, time-invariant, system*

$$\sigma x = f(x),$$

with $x \in X$. Let x_e be an equilibrum of the system and let

$$\sigma\delta_x = A\delta_x$$

be the corresponding linearized system.

- *Asymptotic stability of the linearized system implies (local) asymptotic stability of the equilibrium $x = x_e$ of the nonlinear system.*
- *In the case of continuous-time systems, if the linearized system is unstable because the matrix A has eigenvalues with positive real part then the equilibrium $x = x_e$ of the nonlinear system is unstable.*
- *In the case of discrete-time systems, if the linearized system is unstable because the matrix A has eigenvalues with modulo larger than one then the equilibrium $x = x_e$ of the nonlinear system is unstable.*

The second and third claims cannot be relaxed replacing the fact that one eigenvalue has positive real part, in the case of continuous-time systems, or modulo larger than one, in the case of discrete-time systems, with the property that the linearized system is unstable, as illustrated in the following example.

Example 2.18 *Consider the system*

$$\dot{x}_1 = x_2 \qquad \dot{x}_2 = -x_1^3,$$

with $x = (x_1, x_2) \in I\!R^2$. The origin is an equilibrium of the system, and the linearized system around such an equilibrium is

$$\dot{\delta}_x = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \delta_x,$$

hence it is unstable. However, the zero equilibrium of the nonlinear system is stable. To prove this note that

$$x_1^3\dot{x}_1 + x_2\dot{x}_2 = 0,$$

hence

$$\frac{x_1^4(t)}{4} + \frac{x_2^2(t)}{2}$$

is constant along all trajectories of the system. As a results, using similar arguments as those in Example 2.14, we conclude stability of the equilibrium.

Example 2.19 *Consider the Newton iteration discussed in Example 2.10. Therein, it has been shown that the system linearized around any stationary point such that condition (2.19) holds is*

$$\delta_x^+ = 0,$$

hence the stationary point, which is an equilibrium of the nonlinear system, is locally asymptotically stable. This property justifies the fast convergence rate of Newton's method whenever initialized close to a stationary point, provided the Hessian matrix of the function be nonsingular at the point.

Example 2.20 *Proposition 2.14 holds also for non-time-invariant systems. For example, consider the system in Example 2.11. Suppose that, for all $t \in I\!R$,*

$$\frac{\partial g(x_1 x_2)}{\partial x_1}\left(e^t x_1(0), \frac{e^{-t}}{x_1(0)}\right) + 1 \qquad\qquad \frac{\partial g(x_1 x_2)}{\partial x_2}\left(e^t x_1(0), \frac{e^{-t}}{x_1(0)}\right)$$

are bounded and

$$\frac{\partial g(x_1 x_2)}{\partial x_1}\left(e^t x_1(0), \frac{e^{-t}}{x_1(0)}\right) + 1 < -\epsilon < 0,$$

for some constant $\epsilon > 0$. Then, by direct integration of the linear differential equations, it is possible to conclude that the system (2.22) is asymptotically stable. This, in turn, implies that the motion associated to the trajectory (2.21) is locally asymptotically stable.

Example 2.21 *Newton's iteration*

$$x^+ = x - \left[\frac{\partial^2 f(x)}{\partial x^2}(x)\right]^{-1} \frac{\partial f(x)}{\partial x}(x).$$

is not well-defined, hence not applicable, for all x such that the matrix

$$\left[\frac{\partial^2 f(x)}{\partial x^2}(x)\right]$$

is singular. To avoid this difficulty it is sometimes modified into

$$x^+ = x - \det\left(\frac{\partial^2 f(x)}{\partial x^2}(x)\right)\left[\frac{\partial^2 f(x)}{\partial x^2}(x)\right]^{-1} \frac{\partial f(x)}{\partial x}(x). \tag{2.31}$$

This modification, known as discrete Branin method, is well-defined for all x. However, it has the disadvantage that there are equilibria of the system (2.31) which are not stationary points of the function f. (These points are called extraneous singularities). Moreover, the system linearized around a stationary point x_e of f is given by

$$\delta_x^+ = \left(1 - \det\left(\frac{\partial^2 f(x)}{\partial x^2}(x_e)\right)\right)\delta_x,$$

hence it is not possible to claim anything on the local stability of the equilibrium x_e.

Chapter 3

The structural properties

3.1 Introduction

In this chapter we focus on linear, time-invariant, systems, namely systems described by equations of the form

$$\sigma x = Ax + Bu \qquad y = Cx + Du, \tag{3.1}$$

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$, $y(t) \in I\!R^p$ and A, B, C, and D matrices of appropriate dimensions and with constant entries.

For such a class of systems we study the input-to-state and state-to-output interactions, which are characterized by means of the so-called structural properties: reachability and controllability, observability and reconstructability, respectively.

These properties allow to quantify and describe in a precise way the effect of input signals on the state of the system and the ability to reconstruct the state of the system by means of measurements of the output variable. These properties are naturally defined in terms of properties of the trajectories of the system. However, because of linearity and time-invariance, it is possible to characterize such properties in terms of properties of the matrices arising in the state space representation.

3.2 Reachability and controllability

When studying the input-to-state interaction we can take two different points of view. In the former, we assume that the initial state of the system, i.e. the state of the system at time $t = 0$, is fixed and we consider the problem of determining the states of the system that can be reached applying a certain input signal over a given period of time. In this case we study the so-called reachability property. In the latter, we assume that the final state of the system, at some time T, is fixed and we aim at determining all initial states that can be steered, by means of a certain input signal, to the selected final state. In this case we study the so-called controllability property.

In the study of reachability and controllability, whenever the input signal that drives a certain initial state to a certain final state is not unique, we could impose constraints on such an input signal, e.g. we could consider the input signal with minimum energy, or with minimum amplitude, or the input signal which achieves the *transfer* in minimum time. If no constraint is imposed all input signals achieving the considered trasfer are equivalent.

For linear systems the properties of reachability and controllability are referred to the state $x = 0$, hence we say that a state is reachable to mean that it is reachable from $x = 0$ and that a state is controllable to mean that it is controllable to $x = 0$. Note moreover that, because these properties are used to describe the input-to-state interaction they, trivially, depend only upon properties of the matrices A and B.

3.2.1 Reachability of discrete-time systems

Consider a linear, time-invariant, discrete-time, system. Let $x(0) = 0$ and consider an input sequence $u(0)$, $u(1)$, $u(2)$, $\cdots$, $u(k-1)$. The state reached at $t = k$ is given by

$$x(k) = \begin{bmatrix} B & AB & \cdots & A^{k-1}B \end{bmatrix} \begin{bmatrix} u(k-1) \\ u(k-2) \\ \vdots \\ u(0) \end{bmatrix}.$$

This implies that the set of states that can be reached at $t = k$ is a linear space, i.e. it is the subspace $\mathcal{R}_k$ spanned by all linear combinations of the columns of the matrix

$$R_k = \begin{bmatrix} B & AB & \cdots & A^{k-1}B \end{bmatrix},$$

i.e.

$$\mathcal{R}_k = \mathrm{Im} R_k.$$

The set $\mathcal{R}_k$ is a vector space, denoted as the reachable subspace in k steps. If $\mathcal{R}_k = X$, i.e. $\mathrm{rank} R_k = n$, then all states of the system are reachable in (at most) k steps and the system is said to be reachable in k steps.

As k varies, we have a sequence of subspaces, namely

$$\mathcal{R}_1, \ \mathcal{R}_2, \ \cdots, \ \mathcal{R}_k, \ \cdots. \tag{3.2}$$

This sequence of subspaces is such that the following properties hold.

Proposition 3.1 *The sequence of subspaces (3.2) is such that*

$$\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \cdots \subseteq \mathcal{R}_k \subseteq \cdots.$$

Moreover, if for some $\bar{k}$, $\mathcal{R}_{\bar{k}} = \mathcal{R}_{\bar{k}+1}$, then, for all $k \geq \bar{k}$, $\mathcal{R}_k = \mathcal{R}_{\bar{k}}$. Finally,

$$\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \cdots \subseteq \mathcal{R}_n = \mathcal{R}_{n+1}.$$

Proof. To prove the first claim note that if a state $\bar{x}$ is reached from zero in k steps, using the input sequence $u(0)$, $u(1)$, $\cdots$, $u(k-1)$, then the same state is also reached from zero in $k+1$ steps, using the input sequence 0, $u(0)$, $u(1)$, $\cdots$, $u(k-1)$, hence, for all $k \geq 1$, $\mathcal{R}_k \subseteq \mathcal{R}_{k+1}$.

To prove the second claim it is enough to show that

$$\mathcal{R}_{\bar{k}} = \mathcal{R}_{\bar{k}+1} \Rightarrow \mathcal{R}_{\bar{k}+1} = \mathcal{R}_{\bar{k}+2}, \tag{3.3}$$

or, equivalently that if $\mathcal{R}_{\bar{k}} = \mathcal{R}_{\bar{k}+1}$, then any $\bar{x} \in \mathcal{R}_{\bar{k}+2}$ belongs also to $\mathcal{R}_{\bar{k}+1}$. For, let $\bar{x}$ be an element of $\mathcal{R}_{\bar{k}+2}$. This means that there is an input sequence which steers the state of the system from $x(0) = 0$ to $\bar{x}$ in $\bar{k}+2$ steps. Consider now the state reached after $\bar{k}+1$ steps, using the same input sequence, which we denote with $\tilde{x}$. By assumption, $\tilde{x} \in \mathcal{R}_{\bar{k}+1} = \mathcal{R}_{\bar{k}}$, hence there is an input sequence which steers the state of the system from $x(0) = 0$ to $\tilde{x}$, in $\bar{k}$ steps. However, by definition of $\tilde{x}$ it is possible to steer $\tilde{x}$ to $\bar{x}$ in one step, hence there is an input sequence which steers $x(0) = 0$ to $\bar{x}$, in $\bar{k}+1$ steps, which proves the claim.

To prove the third claim note that if, for some $k < n$, $\mathcal{R}_k = \mathcal{R}_{k+1}$ then the claim follows from equation (3.3). Suppose now that, for all k, the dimension of $\mathcal{R}_{k+1}$ is strictly larger than the dimension of $\mathcal{R}_k$. This implies that the sequence

$$\dim \mathcal{R}_k$$

is strickly increasing at each step. However, this sequence is bounded (from above) by n, and this proves the claim. ◁

Definition 3.1 *Consider the discrete-time system (3.1). The subspace $\mathcal{R} = \mathcal{R}_n$ is the reachability subspace of the system.*

The matrix $R = R_n$ is the reachability matrix of the system.

The system is said to be reachable if $\mathcal{R} = X = I\!R^n$.

Remark. By definition,

$$\mathcal{R} = \mathrm{Im} R,$$

hence the discrete-time system (3.1) is reachable if and only if

$$\mathrm{rank} R = n. \tag{3.4}$$

Equation (3.4) is known as Kalman reachability rank condition, and was derived by R.E. Kalman in the 60's. ◇

Remark. From the above discussion it is obvious that, in a n-dimensional, linear, discrete-time system, if a state $\bar{x}$ is reachable, then it is reachable in at most n steps. This does not mean that n steps are necessarily required, i.e. the state $\bar{x}$ could be reached in less than n steps. In a reachable system, the smallest integer $k^\star$ such that

$$\mathrm{rank} R_{k^\star} = n$$

is called the reachability index of the system. Note that, for single-input reachable systems, necessarily, $k^\star = n$. ◇

Example 3.1 *Consider a discrete-time system with* $x \in \mathbb{R}^3$,

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Then

$$\mathcal{R}_1 = \mathrm{span}B \qquad \mathcal{R}_2 = \mathcal{R}_3 = \mathcal{R} = \mathrm{span}\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Hence the system is not reachable, and its reachable subspace has dimension two.

Example 3.2 *Consider a discrete-time system with* $x \in \mathbb{R}^3$,

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}.$$

Then

$$\mathcal{R}_1 = \mathrm{span}B \qquad \mathcal{R}_2 = \mathrm{span}\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \\ \gamma & 0 \end{bmatrix} \qquad \mathcal{R}_3 = \mathrm{span}\begin{bmatrix} \alpha & \beta & \gamma \\ \beta & \gamma & 0 \\ \gamma & 0 & 0 \end{bmatrix}.$$

As a result, the system is reachable if and only if $\gamma \neq 0$. *Moreover, if* $\gamma = 0$ *and* $\beta \neq 0$, *the system is not reachable and the reachable subspace has dimension two. Finally, if* $\gamma = \beta = 0$ *and* $\alpha \neq 0$, *the system is not reachable and the reachable subspace has dimension one.*

The reachability subspace $\mathcal{R}$ has the following important property, the proof of which is a simple consequence of the definition of the subspace.

Proposition 3.2 *The reachability subspace contains the subspace* span$B$, *i.e.*

$$\mathrm{span}B \subseteq \mathcal{R},$$

and it is A*-invariant, i.e.*

$$A\mathcal{R} \subseteq \mathcal{R}.$$

We conclude this section noting that algebraically equivalent systems have the same reachability properties. In particular, consider two algebraically equivalent systems, with state x and $\hat{x}$, respectively. Let L be the coordinates transformation matrix, as given in equation (2.29), $\mathcal{R}_k$ and $\hat{\mathcal{R}}_k$ the reachability subspaces, and R and $\hat{R}$ the reachability matrices, respectively. Then

$$\hat{\mathcal{R}}_k = L^{-1}\mathcal{R}_k,$$

hence

$$\hat{R} = L^{-1}R,$$

and one of the two systems is reachable if and only if the other is.

3.2.2 Controllability of discrete-time systems

The results established for the reachability property can be easily exploited to characterize the controllability property. In fact, for a linear, time-invariant, discrete-time system, a state $x^\star$ is controllable (to zero) in k steps if there exists an input sequence $u(0)$, $u(1)$, $\cdots$, $u(k-1)$ that drives the state from $x(0) = x^\star$ to $x(k) = 0$, i.e.

$$0 = A^k x^\star + \begin{bmatrix} B & AB & \cdots & A^{k-1}B \end{bmatrix} \begin{bmatrix} u(k-1) \\ u(k-2) \\ \vdots \\ u(0) \end{bmatrix},$$

or equivalently

$$-A^k x^\star = \begin{bmatrix} B & AB & \cdots & A^{k-1}B \end{bmatrix} \begin{bmatrix} u(k-1) \\ u(k-2) \\ \vdots \\ u(0) \end{bmatrix}.$$

This last equation implies that $x^\star$ is controllable if the state $-A^k x^\star$ is reachable in k steps, hence if

$$-A^k x^\star \in \mathcal{R}_k. \tag{3.5}$$

It is easy to see that the set of all $x^\star$ such that equation (3.5) holds is a vector space, denoted by $\mathcal{C}_k$, and called the controllability subspace in k steps.

A linear, discrete-time, system is controllable in k steps if

$$\mathrm{Im} A^k \subseteq \mathcal{R}_k.$$

Example 3.3 *Consider the system in Example 3.1. The system is controllable in two steps. In fact*

$$A^2 = A \in \mathcal{R}_2 = \mathcal{R}.$$

Example 3.4 *Consider the system in Example 3.2. The system is controllable in three steps no matter the values of α, β, and γ. Note in fact that $A^3 = 0$. The system is controllable in two steps if $\gamma = 0$ and $\alpha \neq 0$ or $\beta \neq 0$. Finally, it is controllable in one step if $\gamma = 0$, and $\alpha\beta \neq 0$.*

Similarly to the reachability subspaces, as k varies we have a sequence of controllability subspaces, namely

$$\mathcal{C}_1,\ \mathcal{C}_2,\ \cdots,\ \mathcal{C}_k,\ \cdots. \tag{3.6}$$

This sequence of subspaces is such that the following properties hold.

Proposition 3.3 *The sequence of subspaces (3.6) is such that*

$$\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \cdots \subseteq \mathcal{C}_k \subseteq \cdots .$$

Moreover, if for some $\bar{k}$, $\mathcal{C}_{\bar{k}} = \mathcal{C}_{\bar{k}+1}$, *then, for all* $k \geq \bar{k}$, $\mathcal{C}_k = \mathcal{C}_{\bar{k}}$. *Finally,*

$$\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \cdots \subseteq \mathcal{C}_n = \mathcal{C}_{n+1}.$$

Proof. The proof of this statement is similar to the one of Proposition 3.1. We simply remark that, if a state is controllable in k steps, using the input sequence $u(0)$, $u(1)$, $\cdots$, $u(k-1)$, then the same state is also controllable in $k+1$ steps, using the input sequence $u(0)$, $u(1)$, $\cdots$, $u(k-1)$, 0. ◁

Definition 3.2 *Consider the discrete-time system (3.1). The subspace* $\mathcal{C} = \mathcal{C}_n$ *is the controllability subspace of the system.*

The system is said to be controllable if $\mathcal{C} = X = I\!R^n$.

The discrete-time system (3.1) is controllable if and only if

$$\mathrm{Im} A^n \subseteq \mathcal{R}. \tag{3.7}$$

In particular, if A is nilpotent, i.e. $A^q = 0$, for some $q \leq n$, then for any B (even $B = 0$) the system is controllable. Note that a reachable system is controllable, but the converse statement does not hold. In particular

$$\mathcal{R} \subseteq \mathcal{C} \subseteq X = I\!R^n.$$

3.2.3 Construction of input signals

The study of the properties of reachability and controllability leads to the following question. Is it possible to explicitly construct an input sequence which steers the state of the system from an initial condition x_0, i.e. $x(0) = x_0$, to a final condition x_f in k steps, i.e. $x(k) = x_f$?

To answer this question, consider the problem of determining an input sequence $u(0)$, $u(1)$, $\cdots$, $u(k-1)$ such that

$$x_f - A^k x_0 = R_k U_{k-1}, \tag{3.8}$$

where

$$U_{k-1} = \begin{bmatrix} u(k-1) \\ u(k-2) \\ \vdots \\ u(1) \\ u(0) \end{bmatrix} \in I\!R^{km}.$$

To solve the problem we have to solve the linear system (3.8) in the unknown U_{k-1}. This system has a solution if, and only if,

$$x_f - A^k x_0 \in \mathrm{Im} R_k, \tag{3.9}$$

which clearly shows the role of the matrices R_k in solving the considered problem. Note that the input sequence achieving the desired goal may not be unique. In particular, several solutions can be obtained as the linear combination of a particular solution of equation (3.9) and a solution of the homogeneous equation

$$R_k U_{k-1} = 0.$$

In the special case of a reachable system, it is possible to obtain an explicit expression for one input sequence solving the considered problem in n steps. To this end, note that by reachability, $\mathrm{rank} R_n = n$, hence the condition expressed in equation (3.9), with $k = n$, holds.

Consider now an input signal defined as

$$U_{n-1} = R_n' v,$$

where v has to be determined. Using this definition, and setting $k = n$, equation (3.8) becomes

$$x_f - A^n x_0 = R_n R_n' v,$$

where the matrix $R_n R_n'$ is square and invertible. Hence, a control sequence solving the considered problem in n steps is given by

$$U_{n-1} = R_n' (R_n R_n')^{-1} (x_f - A^n x_0).$$

It is possible to show that, among all input sequences steering the state of the system from x_0 to x_f in n steps the one constructed has minimal norm (energy).

3.2.4 Reachability and controllability of continuous-time systems

The properties of reachability and controllability for linear, time-invariant, continuous-time systems can be assessed using the same ideas exploited in the case of discrete-time systems. However, the tools are more involved as the input-state relation is expressed by means of an integral (see equation (2.9)).

Consider the reachability problem, i.e. the initial state of the system is $x(0) = 0$ and we want to characterize all states $\bar{x}$ that can be reached in some interval of time t, i.e. all states such that, for some input function $u(t)$,

$$\bar{x} = \int_0^t e^{A(t-\tau)} B u(\tau)\, d\tau.$$

Note now that, by Cayley-Hamilton Theorem

$$e^{At} = \alpha_0(t) I + \alpha_1(t) A + \cdots + \alpha_{n-1}(t) A^{n-1},$$

for some scalar functions $\alpha_i(t)$. Hence

$$\bar{x} = \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \begin{bmatrix} \int_0^t \alpha_0(t-\tau)u(\tau)\, d\tau \\ \int_0^t \alpha_1(t-\tau)u(\tau)\, d\tau \\ \vdots \\ \int_0^t \alpha_{n-1}(t-\tau)u(\tau)\, d\tau \end{bmatrix}.$$

This implies that a state $\bar{x}$ is reachable only if

$$\bar{x} \in \mathrm{Im} \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} = \mathrm{Im}R.$$

We now prove the converse fact, i.e. that if a state is in the image of R then it is reachable. To this end, define the controllability Gramian

$$W_t = \int_0^t e^{A(t-\tau)}BB'e^{A'(t-\tau)}\, d\tau,$$

with $t > 0$, and note that[1]

$$\mathrm{Im}R = \mathrm{Im}W_t. \tag{3.10}$$

Selecting

$$u(\tau) = B'e^{A'(t-\tau)}\beta,$$

where β is a constant vector, yields

$$\bar{x} = W_t\beta. \tag{3.11}$$

Hence, to assess reachability of the state $\bar{x} \in \mathrm{Im}R$ it is sufficient to show that equation (3.11) has (at least) one solution β. However, this fact holds trivially by condition (3.10).

Remark. Unlike the case of discrete-time systems, where the set of reachable states depends upon the length of the input sequence, for continuous-time systems if a state is reachable, then it is reachable in any (possibly small) interval of time. ◇

Definition 3.3 *Consider the continuous-time system (3.1). The subspace $\mathcal{R}$ is the reachability subspace of the system.*

The matrix R is the reachability matrix of the system.

The system is said to be reachable if $\mathcal{R} = X = \mathbb{R}^n$.

We summarize the above discussion with a formal statement.

[1]The proof of this property is not trivial.

Proposition 3.4 *Consider the continuous-time system (3.1). The following statements are equivalent.*

- *The system is reachable.*
- $\text{rank}R = n$.
- *For all $t > 0$ the controllability Gramian W_t is positive definite.*

Remark. If a system is reachable then it is possible to explicitly determine one input signal which steers the state of the system from $x(0) = 0$ to any $\bar{x}$ in a give time $t > 0$. In fact, to determine one such input signal it is sufficient to solve the equation (3.11) which, if the system is reachable, has the unique solution

$$\beta = W_t^{-1}\bar{x},$$

i.e. the input signal

$$u(\tau) = B'e^{A'(t-\tau)}W_t^{-1}\bar{x} \tag{3.12}$$

steers the state of the system from $x(0) = 0$ to $x(t) = \bar{x}$. Similarly to what discussed in Section 3.2.3, it is possible to prove that, among all input signals steering the state from 0 to $\bar{x}$ in time t, the input signal (3.12) is the one with minimum energy. Similar considerations can be done to determine an input signal steering a nonzero initial state to a given final state. ◇

To discuss the property of controllability note that a state $\bar{x}$ is controllable (to zero) in time $t > 0$ if there exists an input signal such that

$$0 = e^{At}\bar{x} + \int_0^t e^{A(t-\tau}Bu(\tau)\,d\tau.$$

This, however, implies that

$$e^{At}\bar{x} \in \mathcal{R}$$

hence

$$\bar{x} \in e^{-At}\mathcal{R}.$$

This implies that the set of controllable states in time $t > 0$ is the set

$$\mathcal{C}_t = e^{-At}\mathcal{R},$$

which has the same dimension as $\mathcal{R}$, by invertibility of e^{-At} for all t, and it is contained in $\mathcal{R}$, by the fact that $\mathcal{R}$ is A-invariant, hence it is trivially e^{-At}-invariant. As a consequence, for all $t > 0$,

$$\mathcal{C}_t = \mathcal{R},$$

which shows that the set $\mathcal{C}$ of controllable states does not depend upon $t > 0$ and that a continuous-time system is controllable if and only if it is reachable (unlike what happens for discrete-time systems, for which reachability implies, but it is not implied by, controllability).

Example 3.5 *Consider the linear electric network in Figure 3.1. Assume $R_1 > 0$, $R_2 > 0$, $L > 0$ and $C > 0$. The input u is the driving voltage and the output y is the current supplied.*

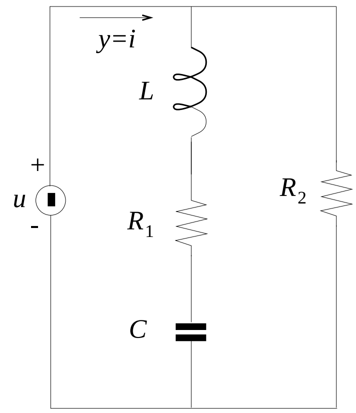


Figure 3.1: A linear electric network.

Let x_1 be the current through L and x_2 the voltage across C. By Kirchhoff's laws we have

$$\dot{x}_1 = -\frac{R_1}{L}x_1 - \frac{1}{L}x_2 + \frac{1}{L}u \qquad \dot{x}_2 = \frac{1}{C}x_1$$

and

$$y = x_1 + \frac{1}{R_2}u.$$

Therefore

$$A = \begin{bmatrix} -\frac{R_1}{L} & -\frac{1}{L} \\ \frac{1}{C} & 0 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix}.$$

The reachability matrix is

$$R = \begin{bmatrix} \frac{1}{L} & -\frac{R_1}{L^2} \\ 0 & \frac{1}{LC} \end{bmatrix}$$

hence the system is reachable (and controllable) for any L and C.

Example 3.6 *Consider the linear electric network in Figure 3.2. Assume $R_1 > 0$, $R_2 > 0$, $L > 0$ and $C > 0$. The input u is the driving voltage and the output y is the current supplied.*

Let x_1 be the voltage across C and x_2 the current through L. By Kirchhoff's laws we have

$$\dot{x}_1 = -\frac{1}{R_1C}x_1 + \frac{1}{R_1C}u \qquad \dot{x}_2 = -\frac{R_2}{L}x_2 + \frac{1}{L}u$$

and

$$y = -\frac{1}{R_1}x_1 + x_2 + \frac{1}{R_1}u.$$

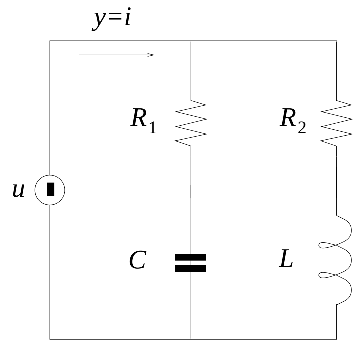


Figure 3.2: A linear electric network.

The reachability matrix is

$$R = \begin{bmatrix} \frac{1}{R_1 C} & -\frac{1}{R_1^2 C^2} \\ \frac{1}{L} & -\frac{R_2}{L^2} \end{bmatrix}$$

and

$$\det R = \frac{1}{R_1 C L}\left(\frac{1}{R_1 C} - \frac{R_2}{L}\right),$$

hence the system is reachable (and controllable) provided $R_1 R_2 C \neq L$.

3.2.5 A canonical form for reachable systems

In this section we focus on single-input systems and we show that the property of reachability allows to write the system in a special form, known as reachability canonical form.

Consider the system (3.1), with $m = 1$, and suppose the system is reachable, i.e. the rank of the reachability matrix is equal to n.

By reachability, there is a (row) vector l such that

$$lB = 0 \qquad lAB = 0 \qquad lA^{n-2}B = 0 \qquad lA^{n-1}B = 1. \tag{3.13}$$

In fact, conditions (3.13) can be rewritten as

$$lR = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{3.14}$$

hence

$$l = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} R^{-1}$$

is well defined.

The vector l has the following important property.

Lemma 3.1 *Let l be as in equation (3.14). Then the square matrix*

$$T = \begin{bmatrix} l \\ lA \\ \vdots \\ lA^{n-2} \\ lA^{n-1} \end{bmatrix}$$

is invertible.

Proof. Consider the matrix

$$TR = \begin{bmatrix} l \\ lA \\ \vdots \\ lA^{n-2} \\ lA^{n-1} \end{bmatrix} \begin{bmatrix} B & AB & \cdots & A^{n-2}B & A^{n-1}B \end{bmatrix}$$

$$= \begin{bmatrix} lB & lAB & \cdots & lA^{n-2}B & lA^{n-1}B \\ lAB & lA^2B & \cdots & lA^{n-1}B & lA^nB \\ \vdots & \vdots & \cdot^{\cdot^{\cdot}} & \vdots & \vdots \\ lA^{n-2}B & lA^{n-1}B & \cdots & \cdots & \cdots \\ lA^{n-1}B & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

and note that, by conditions (3.13),

$$TR = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & lA^nB \\ \vdots & \vdots & \cdot^{\cdot^{\cdot}} & \vdots & \vdots \\ 0 & 1 & \cdots & \cdots & \cdots \\ 1 & lA^nB & \cdots & \cdots & \cdots \end{bmatrix}$$

which shows that $|\det(TR)| = 1$, hence T is invertible. ◁

The matrix T can be used to define a new set of coordinates $\hat{x}$ such that

$$\hat{x} = Tx.$$

To derive the state space representation of the system in the $\hat{x}$ coordinates we could use the general discussion in Section 2.6. However, it is easier to proceeed in an alternative way. For, consider the auxiliary signal

$$\hat{x}_1 = lx$$

and note that

$$\sigma \hat{x}_1 = lAx = \hat{x}_2,$$
$$\sigma \hat{x}_i = lA^i x = \hat{x}_{i+1},$$

for $i = 1, \cdots, n-1$ and

$$\sigma \hat{x}_n = lA^n x + u = lA^n T^{-1} \hat{x} + u.$$

As a result, in the new coordinates $\hat{x}$ one has

$$\sigma \hat{x} = A_r \hat{x} + B_r \hat{x}, \tag{3.15}$$

where

$$A_r = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \cdots & -\alpha_{n-1} \end{bmatrix} \qquad B_r = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

and

$$\begin{bmatrix} -\alpha_0 & -\alpha_1 & -\alpha_2 & \cdots & -\alpha_{n-1} \end{bmatrix} = lA^n T^{-1}.$$

Note that for any α_i, the system (3.15) is reachable, hence a system described by equation (3.15) is said to be in reachability canonical form.

The matrix A_r is in companion form. It is worth noting that its characteristic polynomial is

$$p(s) = s^n + \alpha_{n-1} s^{n-1} + \alpha_{n-2} s^{n-2} + \cdots + \alpha_1 s + \alpha_0,$$

i.e. it depends only upon the elements of the last row.

Example 3.7 *Consider the system in Example 3.5. This system is reachable (and controllable) for any L and C. To write the system in reachability canonical form we have to find a (row) vector l such that conditions (3.13) hold, namely*

$$l \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} = 0 \qquad l \begin{bmatrix} -\frac{R_1}{L^2} \\ \frac{1}{LC} \end{bmatrix} = 1,$$

yielding

$$l = \begin{bmatrix} 0 & LC \end{bmatrix}.$$

Finally

$$T = \begin{bmatrix} l \\ lA \end{bmatrix} = \begin{bmatrix} 0 & LC \\ L & 0 \end{bmatrix}$$

and the system in the transformed coordinates is described by

$$\dot{\hat{x}} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{LC} & -\frac{R_1}{L} \end{bmatrix} \hat{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u,$$

i.e. it is in reachability canonical form.

3.2.6 Description of non-reachable systems

In this section we study systems which are not reachable, i.e. systems described by the equation (3.1) and such that

$$\text{rank}R = \rho < n.$$

Under this assumption, consider a set of coordinates $\hat{x}$ such that

$$x = L\hat{x},$$

and the matrix L is constructed as follows. The first ρ columns of L are ρ linearly independent columns of the matrix R, and the last $n - \rho$ columns are selected in such a way that the matrix L is invertible[2].

The system in the $\hat{x}$ coordinates, which is algebraically equivalent to the system in the x coordinates, is described by the equations

$$\sigma\hat{x} = \hat{A}\hat{x} + \hat{B}u = L^{-1}AL\hat{x} + L^{-1}Bu.$$

We now show that, because of the way in which L has been constructed, the matrices $\hat{A}$ and $\hat{B}$ have a special structure. To this end note that

$$L\hat{A} = AL \qquad L\hat{B} = B$$

and partition the matrices L, A, $\hat{A}$, B and $\hat{B}$ as

$$L = \left[\begin{array}{c|c} L_{11} & L_{12} \\ \hline L_{12} & L_{22} \end{array}\right], \qquad A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{12} & A_{22} \end{array}\right], \qquad \hat{A} = \left[\begin{array}{c|c} \hat{A}_{11} & \hat{A}_{12} \\ \hline \hat{A}_{21} & \hat{A}_{22} \end{array}\right],$$

$$B = \left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right], \qquad \hat{B} = \left[\begin{array}{c} \hat{B}_1 \\ \hline \hat{B}_2 \end{array}\right],$$

where L_{11}, A_{11} and $\hat{A}_{11}$ have dimensions $\rho \times \rho$; L_{12}, A_{12} and $\hat{A}_{12}$ have dimensions $\rho \times n - \rho$; L_{21}, A_{21} and $\hat{A}_{21}$ have dimensions $n - \rho \times \rho$; L_{22}, A_{22} and $\hat{A}_{22}$ have dimensions $n - \rho \times n - \rho$; B_1 and $\hat{B}_1$ have dimensions $\rho \times m$ and B_2 and $\hat{B}_2$ have dimensions $n - \rho \times m$.

Recall that, by construction, $\mathcal{R}$ is spanned by the first ρ columns of the matrix L, and that $\mathcal{R}$ is A-invariant and contains B. This implies that the submatrices $\hat{A}_{21}$ and $\hat{B}_2$ have to be identically zero, i.e. in the $\hat{x}$ coordinates the system is described by

$$\sigma\hat{x} = \left[\begin{array}{c} \sigma\hat{x}_1 \\ \sigma\hat{x}_2 \end{array}\right] = \left[\begin{array}{cc} \hat{A}_{11} & \hat{A}_{12} \\ 0 & \hat{A}_{22} \end{array}\right]\hat{x} + \left[\begin{array}{c} \hat{B}_1 \\ 0 \end{array}\right]u, \tag{3.16}$$

where $\hat{x}_1 \in I\!R^{\rho}$ and $\hat{x}_2 \in I\!R^{n-\rho}$. The reachability matrix of the system in the $\hat{x}$ coordinates is

$$\hat{R} = \left[\begin{array}{cccc} \hat{B}_1 & \hat{A}_{11}\hat{B}_1 & \cdots & \hat{A}_{11}^{n}\hat{B}_1 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

[2]It is always possible to determine such $n - \rho$ vectors. Moreover it is possible to select them among the vectors e_i of the canonical basis.

and, because $R = L\hat{R}$, it has rank ρ. This implies that the subsystem

$$\sigma\hat{x}_1 = \hat{A}_{11}\hat{x}_1 + \hat{B}_1 u \tag{3.17}$$

is reachable. The subsystem (3.17) is called the reachable subsystem of the system (3.1), whereas the subsystem

$$\sigma\hat{x}_2 = \hat{A}_{22}\hat{x}_2, \tag{3.18}$$

which is clearly not affected by the input, is called the unreachable subsystem of the system (3.1). The eigenvalues of the matrix $\hat{A}_{22}$, which are a subset of the eigenvalues of the matrix A, are called unreachable modes.

Example 3.8 *Consider the system in Example 3.6, and assume $R_1R_2C = L$, which implies that the system is not reachable (and not controllable), as the rank of the reachability matrix is equal to one. To obtain a decomposition into reachable and unreachable subsystems, let*

$$L = \begin{bmatrix} \frac{1}{R_1C} & 1 \\ \frac{1}{L} & 0 \end{bmatrix}$$

yielding

$$\hat{A} = L^{-1}AL = \begin{bmatrix} -\frac{R_2}{L} & 0 \\ 0 & -\frac{1}{R_1C} \end{bmatrix} \qquad \hat{B} = L^{-1}B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

3.2.7 PBH reachability test

The decomposition of a system in reachable and unreachable parts allows to derive an alternative test for reachability.

Proposition 3.5 (Popov-Belevich-Hautus (PBH) reachability test) *Consider the system (3.1). The system is reachable if and only if*

$$\text{rank} \left[\begin{array}{c|c} sI - A & B \end{array} \right] = n$$

for all $s \in \mathbb{C}$.

Remark. The matrix

$$\left[\begin{array}{c|c} sI - A & B \end{array} \right]$$

is called reachability pencil. Note that the rank condition in Proposition 3.5 holds trivially for all s which are not eigenvalues of A, hence the condition has to be checked only for the n (complex) numbers which are eigenvalues of A. ◇

Proof. (Necessity) We prove the necessity by contradiction. Suppose the system is reachable and that, for some $s^\star \in \mathbb{C}$,

$$\text{rank} \left[\begin{array}{c|c} s^\star I - A & B \end{array} \right] < n.$$

Then there is a vector w such that

$$w' \left[\begin{array}{c|c} s^\star I - A & B \end{array} \right] = 0,$$

hence

$$w'B = 0 \qquad\qquad w'A = s^\star w'.$$

As a result

$$w'AB = 0 \qquad w'A^2B = 0 \qquad \cdots \qquad w'A^{n-1}B = 0$$

or, equivalently,

$$w'R = 0,$$

which implies that the system is not reachable, hence the contradiction.
(Sufficiency) Again, we prove the statement by contradiction. Suppose

$$\text{rank} \left[\begin{array}{c|c} sI - A & B \end{array} \right] = n$$

for all $s \in \mathbb{C}$ and the system is not reachable. Consider the change of coordinates which transforms the system into the form (3.16) and note that, for the transformed system one has[3]

$$\text{rank} \left[\begin{array}{cc|c} sI - \hat{A}_{11} & -\hat{A}_{12} & \hat{B}_1 \\ 0 & sI - \hat{A}_{22} & 0 \end{array} \right] = n.$$

However, this is not true as this matrix loses rank for all s which are eigenvalues of $\hat{A}_{22}$, hence the contradiction. $\lhd$

Remark. The PBH test allows to compute the unreachable modes of a system without performing the decomposition into reachable and unreachable parts. In fact, the unreachable modes are all the complex numbers for which the reachability pencil loses rank. $\diamond$

Example 3.9 *Consider again the system in Example 3.6. The reachability pencil is*

$$\left[\begin{array}{cc|c} s + \frac{1}{R_1C} & 0 & \frac{1}{R_1C} \\ 0 & s + \frac{R_2}{L} & \frac{1}{L} \end{array} \right].$$

This matrix has rank two for all $s \neq -\frac{1}{R_1C}$ *and* $s \neq -\frac{R_2}{L}$. *For* $s = -\frac{1}{R_1C}$ *or* $s = -\frac{R_2}{L}$ *the reachability pencil loses rank, i.e. the system is not reachable, if*

$$-\frac{1}{R_1C} + \frac{R_2}{L} = 0,$$

and this condition is the same as the one derived in Example 3.6.

---

[3] Note that the rank of the reachability pencils of algebraically equivalent systems is the same for all $s \in \mathbb{C}$. This is a consequence of the identity

$$\left[\begin{array}{c|c} sI - \hat{A} & \hat{B} \end{array} \right] = \left[\begin{array}{c|c} sI - L^{-1}AL & L^{-1}B \end{array} \right] = L^{-1} \left[\begin{array}{c|c} sI - A & B \end{array} \right] \left[\begin{array}{cc} L & 0 \\ 0 & I \end{array} \right].$$

Example 3.10 *Consider a continuous-time system described by the equation*

$$\dot{x} = Ax + Bu \tag{3.19}$$

and its Euler approximate discrete-time model

$$x^+ = x + TAx + TBu, \tag{3.20}$$

where T denotes the sampling-time. Suppose the continuous-time system is controllable, i.e.

$$\text{rank} \left[\ sI - A \ \middle| \ B \ \right] = n,$$

for all $s \in \mathbb{C}$. Consider now the reachability pencil of the discrete-time model (3.20), i.e.

$$\left[\ sI - (I + TA) \ \middle| \ TB \ \right]$$

and note that, by reachability of the continuous-time system,

$$\text{rank} \left[\ sI - (I + TA) \ \middle| \ TB \ \right] = \text{rank} \left[\ \frac{s-1}{T} I - A \ \middle| \ B \ \right] = \text{rank} \left[\ zI - A \ \middle| \ B \ \right] = n,$$

where $z = \frac{s-1}{T}$. Hence, reachability of system (3.19) implies, and it is implied by, reachability of system (3.20).

Suppose now that the continuous-time system is not reachable, and that all non-reachable modes are in the open left part of the complex plane, i.e. they have negative real part hence they are asymtotically stable. This implies that the discrete-time Euler approximate model is not reachable, however the non-reachable modes belong to the set described by

$$\left\{ z \in \mathbb{C} \mid \text{Real}(z) < -\frac{1}{T} \right\},$$

hence it is not possible to decide a-priori if they are stable or otherwise.

3.3 Observability and reconstructability

In this section we study the state-to-output relation, and we focus on the problem of determining the state of a system, at a given time instant, from measurements of the input and output signals. This problem, of great importance in applications, can be addressed from two different perspectives.

In the former, we assume that the state at time t has to be determined on the basis of current and future measurements. In this case we deal with a so-called observability problem. In the latter, we assume that the state at time t has to be determined from past and current measurements. In this case we have a so-called reconstructability problem. Observability problems typically arise in real-time problems, where one has to determine the state of a system from the actual measurements, whereas a classical reconstructability problem arises in weather forecast, where one wishes to determine the current weather from past measurements.

3.3.1 Observability of discrete-time systems

Consider a linear, time-invariant, discrete-time system. Two states x_a and x_b are indistinguishable in the future in k steps if for any input sequence $u(0)$, $u(1)$, ..., $u(k-1)$ the corresponding output sequences y_a and y_b, coincide for the first k steps, i.e.

$$y_a(t) = y_b(t) \tag{3.21}$$

for all $t \in [0, k]$. By equation (2.12), condition (3.21) is equivalent to

$$CA^t x_a = CA^t x_b \tag{3.22}$$

for all $t \in [0, k]$. This implies that the property of indistinguishability in the future does not depend upon the input sequence, i.e. it is a property of the free response of the output of the system, hence of the matrices A and C. Note, moreover, that condition (3.22) can be rewritten as

$$x_a - x_b \in \ker O_k$$

where

$$O_k = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^k \end{bmatrix}.$$

We say that two states are indistinguishable in the future if they are indistinguishable in the future in k steps for all $k \geq 0$. Note that, by Cayley-Hamilton Theorem, two states x_a and x_b are indistinguishable in the future if

$$x_a - x_b \in \ker O_{n-1}$$

Definition 3.4 *A state x is not observable in k steps if it is not distinguishable in the future in k steps from the zero state. It is not observable if it is not observable in k steps for all k.*

As a consequence of the above discussion, we conclude that a state x is not observable in k steps if

$$x \in \ker O_k$$

and it is not observable if

$$x \in \ker O_{n-1}.$$

Note that the set of non-observable (in k steps) states is a subspace.

Definition 3.5 *Consider the discrete-time system (3.1). The subspace $\ker O_{n-1}$ is the unobservable subspace of the system.*

The matrix $O = O_{n-1}$ is the observability matrix of the system.

The system is said to be observable if $\ker O = \{0\}$.

Remark. The discrete-time system (3.1) is observable if and only if

$$\text{rank}O = n. \tag{3.23}$$

Equation (3.23) is known as Kalman observability rank condition, and was derived by R.E. Kalman in the 60's. ⋄

Example 3.11 *Consider a discrete-time system with $x \in \mathbb{R}^3$, $y \in \mathbb{R}$,*

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} \alpha & \beta & \gamma \end{bmatrix}.$$

Then

$$O_0 = C \qquad O_1 = \begin{bmatrix} \alpha & \beta & \gamma \\ 0 & 0 & \gamma \end{bmatrix}. \qquad O_2 = \begin{bmatrix} \alpha & \beta & \gamma \\ 0 & 0 & \gamma \\ 0 & 0 & \gamma \end{bmatrix}.$$

Hence the system is not observable. The unobservable subspace has dimension one if $\alpha\gamma \neq 0$ or $\beta\gamma \neq 0$, it has dimension two either if $\gamma = 0$ and $\alpha \neq 0$ or $\beta \neq 0$, or if $\gamma \neq 0$ and $\alpha = 0$ and $\beta = 0$.

Remark. The existence of a non-empty unobservable subspace, implies that from current and future output and input measurements it is not possible to determine the current state. In fact this can be determined modulo an element in ker$O$. ⋄

Example 3.12 *Consider the system in Example 3.11 and assume $\alpha \neq 0$, $\beta = 0$ and $\gamma \neq 0$. Thus the unobservable subspace has dimension one and*

$$\ker O = \text{span} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Therefore it is not possible to distinguish, using current and future measurements, between states which have the same x_1 and x_3 components, but different x_2 component. This means that, for example, from current and past measurements, it is possible to conclude that the initial state belongs to the straight line in Figure 3.3, but it is not possible to determine at which point on the line the initial state is.

The unobservable subspace $\ker O$ has the following important property, the proof of which is a simple consequence of the definition of the subspace.

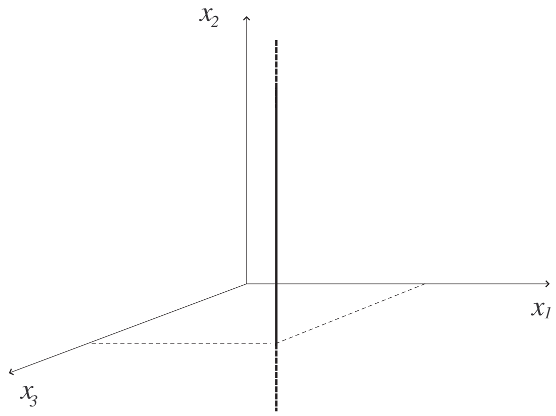


Figure 3.3: Admissible initial states for an unobservable system.

Proposition 3.6 *The unobservable subspace contains the subspace* $\ker C$*, i.e.*

$$\ker C \subseteq \ker O,$$

and it is A-invariant, i.e.

$$A \ker O \subseteq \ker O.$$

We conclude this section noting that algebraically equivalent systems have the same observability properties. In particular, consider two algebraically equivalent systems, then one of the two systems is observable if and only if the other is.

3.3.2 Reconstructability of discrete-time systems

In this section we study the property of reconstructability. Consider a linear, time-invariant, discrete-time system, assume that the input sequence $u(0)$, $u(1)$, …, $u(k-1)$ and the output sequence $y(0)$, $y(1)$, …, $y(k)$ are known and consider the problem of determining the state of the system at time k, i.e. $x(k)$.

Of course, if the system is observable in k steps, then the considered problem is solvable. In fact, the input and output sequences determine a unique state $x(0)$, from which it is possible to compute $x(k)$.

If the system is not observable in k steps, then the initial state cannot be uniquely determined, but it can determined modulo an element in the unobservable subspace. This means that, if $x(0)$ is an initial state which is consistent with the input and output sequences, then all states described by

$$\tilde{x}(0) = x(0) + \ker O_k$$

are also consistent with the same input and output sequences.

Consider now the initial state $\tilde{x}(0)$ and the resulting state at time k, namely

$$\tilde{x}(k) = A^k x(0) + \sum_{t=0}^{k-1} A^{k-t-1} B u(t) + A^k \mathrm{ker} O_k = x(k) + A^k \mathrm{ker} O_k.$$

As a result, for given input and output sequences, the state of the system at time k is uniquely defined if

$$A^k \mathrm{ker} O_k = \{0\},$$

or, equivalently, if

$$\mathrm{ker} O_k \subseteq \ker A^k \tag{3.24}$$

whereas it is not uniquely defined if the above condition does not hold.

A system is said reconstructable in k steps if condition (3.24) holds. It is said reconstructable if it is reconstructable in k steps for all k. Recalling that

$$\ker A^k = \ker A^n$$

for all $k \geq n$, a system is reconstructable if and only if

$$\mathrm{ker} O \subseteq \mathrm{ker} A^n. \tag{3.25}$$

Note that, the observability rank condition (3.23) implies, but it is not implied, by condition (3.25).

Example 3.13 *Consider a discrete-time system with $x \in I\!R^3$, $y \in I\!R$,*

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix},$$

i.e.

$$x_1^+ = x_1 + x_2 \qquad x_2^+ = x_1 \qquad x_3^+ = x_2,$$

and $y = x_1$. The observability matrix is

$$O = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 0 \end{bmatrix},$$

hence the system is not observable. The unobservable subspace has dimension one and is given by

$$\mathrm{ker} O = \mathrm{span} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Note now that

$$A^3 = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix},$$

hence

$$\ker A^3 = \ker O,$$

which implies that the system is reconstructable.

This fact can be shown directly. Suppose we want to determine $x(2)$ *from* $y(0)$, $y(1)$ *and* $y(2)$. *To this end, note that*

$$x_1(2) = y(2),$$

$$x_2(2) = x_1(1) = y(1)$$

and

$$x_3(2) = x_2(1) = x_1(0) = y(0),$$

which proves the claim.

Example 3.14 *Consider again the system in Example 3.13. Note that*

$$\ker A^2 = \ker O,$$

hence the system is reconstructable in two steps. This property can be shown directly. In fact, suppose we want to determine $x(1)$ *from* $y(0)$ *and* $y(1)$. *This can be achieved noting that*

$$x_1(1) = y(1),$$

$$x_2(1) = x_1(0) = y(0)$$

and

$$x_3(1) = x_2(0) = x_1(1) - x_1(0) = y(1) - y(0).$$

3.3.3 Computation of the state

The property of observability highlights the ability to determine the state of a system from present and future measurements. From a practical point of view it is however important not only to characterize the observability property, but also to have a procedure that allows to effectively compute the unknown state.

Consider a linear, time-invariant, discrete-time system, and note that, by equation (2.12), the output response of the system is a linear combination of the free response and of the forced response. The latter is known, once the input signal is known, whereas the former depends upon the initial state, which has to be determined. This means that the problem of determining the initial state of a system from current and future (input and output) measurements is equivalent to the problem of determining the initial state of the system from the knowledge of its free output response, i.e. it is possible to assume, without loss of generality, that the input sequence is identically equal to zero.

To solve this problem note that

$$Y_k = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(k) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^k \end{bmatrix} x(0) = O_k x(0),$$

with $Y_k \in I\!R^{p\,(k+1)}$. If the system is observable then $O_{n-1} = O$ is full rank, hence the equation

$$Y_{n-1} = Ox(0) \qquad (3.26)$$

has the (unique) solution

$$x(0) = (O'O)^{-1}O'Y_{n-1}, \qquad (3.27)$$

which provides a simple way to compute the state $x(0)$ of the system.

Remark. Equation (3.27) shows how to compute the initial state of a system from current and future measurements. This information, in turn, can be used to determine the state at time k using equation (2.11). ◇

Remark. Measurements are naturally affected by noise, i.e. in practice equation (3.26) has to be replaced by

$$Y_{n-1} + \nu = Ox(0),$$

where ν represents a vector of additive noise affecting the output measurements. (For simplicity we assume that the input sequence is identically zero.) It is possible to prove that the initial state given by equation (3.27) yields a free output response which minimizes the norm (energy) of the error between the actual free output response and the calculated one. ◇

3.3.4 Observability and reconstructability for continuous-time systems

The properties of observability and reconstructability for continuous-time systems can be assessed using the same ideas exploited in the case of discrete-time systems.

Consider a linear, time-invariant, continuous-time system. Two states x_a and x_b are indistinguishable in the future over the interval $[0, t]$ if for any input signal u the corresponding output responses coincide in the interval $[0, t]$. This property, recalling equation (2.10), and noting that the forced responses do not depend upon the initial states, is equivalent to the condition

$$Ce^{A\tau}x_a = Ce^{A\tau}x_b,$$

for all $\tau \in [0, t]$, or to the condition

$$Ce^{A\tau}(x_a - x_b) = 0, \tag{3.28}$$

for all $\tau \in [0, t]$.

The function $Ce^{A\tau}$ is analytic, hence condition (3.28) is equivalent to

$$Ce^{A\tau}(x_a - x_b)|_{\tau=0} = 0 \quad \frac{d}{d\tau}Ce^{A\tau}(x_a - x_b)|_{\tau=0} = 0 \quad \cdots \quad \frac{d^i}{d\tau^i}Ce^{A\tau}(x_a - x_b)|_{\tau=0} = 0 \quad \cdots,$$

yielding, by a property of the matrix exponential,

$$C(x_a - x_b) = 0 \qquad CA(x_a - x_b) = 0 \qquad \cdots \qquad CA^i(x_a - x_b) = 0 \qquad \cdots. \tag{3.29}$$

Note finally that, by Cayley-Hamilton Theorem, (3.29) is equivalent to (recall the definition of the observability matrix O)

$$O(x_a - x_b) = 0. \tag{3.30}$$

In summary, if two states x_a and x_b are indistinguishable in the future over the interval $[0, t]$ then they have to be such that condition (3.30) holds.

Remark. Unlike discrete-time systems, in the continuous-time case if two states are indistinguishabe in the future over an interval $[0, t]$ then they are indistinguishabe in the future over any interval $[0, \bar{t}]$, with $\bar{t} > 0$. ⋄

From the above discussion, we conclude that, for a continuous-time system, all states which are indistinguishable in the future from the zero state, i.e. the unobservable states, are those, and only those, belonging to $\ker O$, as expressed in the following statement.

Proposition 3.7 *Consider the continuous-time system (3.1). The following statements are equivalent.*

- *The system is observable.*
- $\text{rank} O = n$.
- *For all $t > 0$ the observability Gramian*

$$V_t = \int_0^t e^{A'\tau} C' C e^{A\tau} d\tau$$

  is positive definite.

Proof. We have only to prove the claim on the observability Gramian. Note first that $V_t = V_t' \geq 0$, hence we only need to show that $\text{rank} V_t = n$, for all $t > 0$, if and only if the system is observable.

Suppose the system is observable and rank$V_t < n$, for some $t > 0$. This implies that there exists a vector w such that

$$w'V_t w = 0,$$

which implies

$$\int_0^t w' e^{A'\tau} C' C e^{A\tau} w d\tau = 0$$

hence

$$\int_0^t \|Ce^{A\tau} w\| d\tau = 0.$$

This last equality implies that the function

$$Ce^{A\tau} w$$

is identically equal to zero on the interval $[0, t]$. As a result, by analyticity of the function and a property of the matrix exponential,

$$Cw = 0 \qquad CAw = 0 \qquad \cdots \qquad CA^{n-1}w = 0,$$

or equivalently $Ow = 0$, which contradicts the observability assumption.

Suppose now that $V_t > 0$ for all $t > 0$. To show that the system is observable consider the function

$$\delta(t) = \int_0^t e^{A'\tau} C' y(\tau) d\tau$$

and note that

$$\delta(t) = V_t x(0).$$

Hence, by positivity of V_t we can uniquely determine $x(0)$ processing future measurements, i.e. the system is observable. ◁

We conclude this section studying the property of reconstructability for linear, continuous-time systems. For, note that two states x_a and x_b are indistinguishable in the past over the interval $[-t, 0]$ if for any input signal u the corresponding output responses coincide in the interval $[-t, 0]$.

Using arguments similar to the ones used in the case of discrete-time systems, we conclude that all states are distinguishable in the past, over the interval $[-t, 0]$, from the zero state if

$$e^{At} \ker O = \{0\}.$$

However, because the matrix e^{At} is invertible for all t, the above condition is equivalent to

$$\ker O = \{0\},$$

i.e. a continuous-time system is reconstructable if and only if it is observable.

Example 3.15 *Consider the system in Example 3.5. The observability matrix is*

$$O = \begin{bmatrix} 1 & 0 \\ -\frac{R_1}{L} & -\frac{1}{L} \end{bmatrix}$$

hence the system is observable for any zero L *(recall that* $L > 0$*).*

Example 3.16 *Consider the system in Example 3.6. The observability matrix is*

$$O = \begin{bmatrix} -\frac{1}{R_1} & 1 \\ \frac{1}{R_1^2 C} & -\frac{R_2}{L} \end{bmatrix}$$

and

$$\det O = \frac{R_2}{R_1 L} - \frac{1}{R_1^2 C},$$

hence the system is observable if, and only if, $R_1 R_2 C \neq L$*.*

3.3.5 Duality

In the study of the structural properties there is a strong relation between the results on reachability and observability and the results on controllability and reconstructability. This relation can be defined formally by introducing the notion of dual system.

Definition 3.6 *Consider the system, with state* $x \in X = I\!R^n$*, input* $u \in I\!R^m$*, and output* $y \in I\!R^p$*, described by the equations*

$$\sigma x = Ax + Bu \qquad\qquad y = Cx + Du. \tag{3.31}$$

The system, with state $\xi \in \Xi = I\!R^n$*, input* $v \in I\!R^p$*, and output* $\eta \in I\!R^m$*, described by the equations*

$$\sigma\xi = A'\xi + C'v \qquad\qquad \eta = B'\xi + D'v \tag{3.32}$$

is called the dual system of system (3.31), which is called the primal system of system (3.32).

To understand the importance and the usefulness of this notion of duality, let R and O be the reachability and observability matrix, respectively, of the system (3.31) and let $R^\star$ and $O^\star$ be the reachability and observability matrix, respectively, of the dual system (3.32).

Then, trivially,

$$R^\star = O' \qquad\qquad O^\star = R'.$$

Moreover, for discrete-time systems, the following implications hold

$$\text{Im}A^n \subseteq R \Leftrightarrow (\text{Im}A^n)^\perp \supseteq R^\perp \Leftrightarrow \ker A^n \supseteq \ker R' = \ker O^\star.$$

As a result, the following statement holds.

Proposition 3.8 *Consider the system (3.31) and its dual (3.32).*

- *System (3.31) is reachable if and only if system (3.32) is observable.*
- *System (3.31) is observable if and only if system (3.32) is reachable.*
- *System (3.31) is controllable if and only if system (3.32) is reconstructable.*
- *System (3.31) is reconstructable if and only if system (3.32) is controllable.*

The notion of duality and the above statement allow to derive results similar to those presented in Sections 3.2.5, 3.2.6 and 3.2.7, but with respect to the observability property.

Proposition 3.9 *Consider the system (3.1) with $p = 1$. Suppose the system is observable. Then the system is algebraically equivalent to a system of the form*

$$\sigma\hat{x} = A_o\hat{x} + B_o u \qquad\qquad y = C_o\hat{x} + D_o u \tag{3.33}$$

with

$$A_0 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & 0 & -\alpha_1 \\ 0 & 1 & \cdots & 0 & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & -\alpha_{n-2} \\ 0 & 0 & \cdots & 0 & 1 & -\alpha_{n-1} \end{bmatrix} \qquad C_0 = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Note that for any α_i, the system (3.33) is observable, hence a system described by the equations (3.33) is said to be in observability canonical form.

Proposition 3.10 *Consider the system (3.1). Suppose the system is not observable. Then the system is algebraically equivalent to a system of the form*

$$\sigma\hat{x} = \begin{bmatrix} \sigma\hat{x}_1 \\ \sigma\hat{x}_2 \end{bmatrix} = \begin{bmatrix} \hat{A}_{11} & 0 \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}\hat{x} + \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} u,$$

$$y = \begin{bmatrix} \hat{C}_1 & 0 \end{bmatrix}\hat{x} + \hat{D}u.$$

Moreover, the system

$$\sigma\hat{x}_1 = \hat{A}_{11}\hat{x}_1 + \hat{B}_1 u \qquad\qquad y = \hat{C}_1\hat{x}_1 + \hat{D}u \tag{3.34}$$

is observable.

The subsystem (3.34) is called the observable subsystem of the system (3.1), whereas the subsystem

$$\sigma\hat{x}_2 = \hat{A}_{22}\hat{x}_2 + \hat{B}_2 u \qquad\qquad y = 0, \tag{3.35}$$

which clearly does not contribute to the output, is called the unobservable subsystem of the system (3.1). The eigenvalues of the matrix $\hat{A}_{22}$, which are a subset of the eigenvalues of the matrix A, are called unobservable modes.

Proposition 3.11 (PBH observability test) *Consider the system (3.1). The system is observable if and only if*

$$\text{rank}\begin{bmatrix} sI - A \\ \hline C \end{bmatrix} = n$$

for all $s \in \mathbb{C}$.

Example 3.17 (Kalman decomposition) *Consider the system (3.1), the associated reachable subspace* $\mathcal{R}$ *and unobservable subspace* $\ker O$. *Let*

$$\mathcal{X}_1 = \mathcal{R} \cap \ker O,$$

with $\dim \mathcal{X}_1 = n_1$, *and note that, by A-invariance of* $\mathcal{R}$ *and* $\ker O$, *also* $\mathcal{X}_1$ *is A-invariant. Let* $\mathcal{X}_2$, $\mathcal{X}_3$ *and* $\mathcal{X}_4$ *be subspaces, of dimension* n_2, n_3, *and* n_4, *respectively, such that*[4]

$$\mathcal{X}_1 \oplus \mathcal{X}_2 = \mathcal{R} \qquad\qquad \mathcal{X}_1 \oplus \mathcal{X}_3 = \ker O \qquad\qquad \mathcal{X}_1 \oplus \mathcal{X}_2 \oplus \mathcal{X}_3 \oplus \mathcal{X}_4 = X = \mathbb{R}^n.$$

Note that, by construction, $n_1 + n_2 + n_3 + n_4 = n$. *Consider now a change of coordinates described by*

$$x = L\hat{x}$$

where the matrix L *is constructed as follows. The first* n_1 *columns of* L *are a basis for* $\mathcal{X}_1$. *The subsequent* n_2 *columns are a basis for* $\mathcal{X}_2$, *the subsequent* n_3 *columns are a basis for* $\mathcal{X}_3$, *and finally, the last* n_4 *columns are a basis for* $\mathcal{X}_4$.

In the coordinates $\hat{x}$ *the system is described by*

$$\begin{aligned} \sigma\hat{x} &= \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} & \hat{A}_{13} & \hat{A}_{14} \\ 0 & \hat{A}_{22} & 0 & \hat{A}_{24} \\ 0 & 0 & \hat{A}_{33} & \hat{A}_{34} \\ 0 & 0 & 0 & \hat{A}_{44} \end{bmatrix}\hat{x} + \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \\ 0 \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} 0 & \hat{C}_2 & 0 & \hat{C}_4 \end{bmatrix}\hat{x} + \hat{D}u. \end{aligned} \tag{3.36}$$

The state space representation (3.36) is known as Kalman's canonical form. From this representation it is possible to single out the following subsystems.

[4]The symbol $\oplus$ denotes the direct sum between subspaces. The direct sum of two subspaces is defined as follows. Let $\mathcal{X}$ and $\mathcal{Y}$ be two subspaces of a vector space $\mathcal{W}$. The sum of these subspaces, denoted $\mathcal{X} + \mathcal{Y}$, is the set of all vectors $x + y$, with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. If $\mathcal{X} \cap \mathcal{Y} = \{0\}$ the sum is a direct sum and it is denoted $\mathcal{X} \oplus \mathcal{Y}$.

- *The reachable and unobservable subsystem*

$$\sigma \hat{x}_1 = \hat{A}_{11}\hat{x}_1 + \hat{B}_1 u \qquad y = 0.$$

- *The reachable and observable subsystem*

$$\sigma \hat{x}_2 = \hat{A}_{22}\hat{x}_2 + \hat{B}_2 u \qquad y = \hat{C}_2\hat{x}_2 + \hat{D}u.$$

- *The unreachable and unobservable subsystem*

$$\sigma \hat{x}_3 = \hat{A}_{33}\hat{x}_3 \qquad y = 0.$$

- *The unreachable and observable subsystem*

$$\sigma \hat{x}_4 = \hat{A}_{44}\hat{x}_4 \qquad y = \hat{C}_4\hat{x}_4.$$

These subsystems, and their role in the overall system, are illustrated in Figure 3.4.

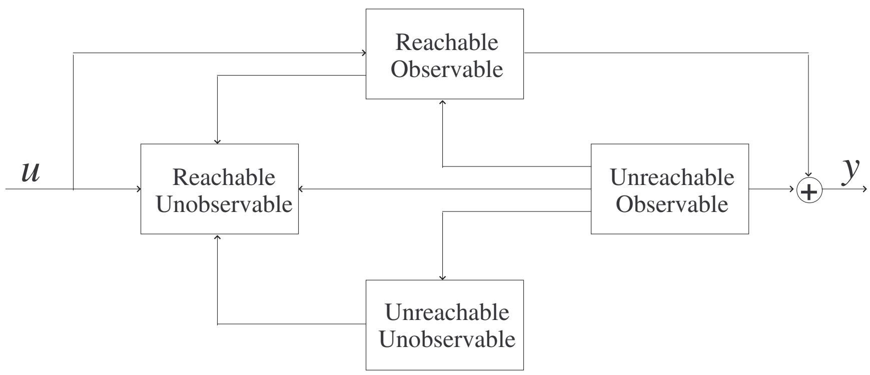


Figure 3.4: Illustrative representation of a system in Kalman's canonical form.

We conclude this example noting that a state space representation which is reachable and observable is called minimal. From a nonminimal state space representation it is always possible to obtain a (reduced) minimal representation. This is constructed computing Kalman's canonical form and then disregarding the unreachable and/or unobservable parts.

Note that only the reachable and observable subsystem contribute to the input-output behaviour of the system. This can be proved, for example, noting that the transfer function of the system (3.36) is

$$G(s) = \hat{C}_2(sI - \hat{A}_{22})^{-1}\hat{B}_2 + \hat{D}.$$

Example 3.18 *Consider the system*

$$\begin{array}{rcl} \dot{x}_1 & = & x_2 + u \\ \dot{x}_2 & = & \alpha x_1 + u \\ y & = & x_1 - x_2 \end{array}$$

with $x_1 \in I\!R$, $x_2 \in I\!R$, $u \in I\!R$, $y \in I\!R$ *and* α *a constant parameter.*

The reachability matrix is

$$R = \begin{bmatrix} 1 & 1 \\ 1 & \alpha \end{bmatrix},$$

hence the system is reachable (and controllable) if $\alpha \neq 1$*. If* $\alpha = 1$ *the reachability pencil is*

$$\left[\begin{array}{c|c} sI - A & B \end{array}\right] = \left[\begin{array}{cc|c} s & -1 & 1 \\ -1 & s & 1 \end{array}\right]$$

and this has rank one for $s = -1$ *and rank two for any other* s*. Therefore, for* $\alpha = 1$*, the unreachable mode is* $s = -1$*.*

The observability matrix is

$$O = \begin{bmatrix} 1 & -1 \\ -\alpha & 1 \end{bmatrix},$$

hence the system is observable if $\alpha \neq 1$*. If* $\alpha = 1$ *the observability pencil is*

$$\left[\begin{array}{c} sI - A \\ \hline C \end{array}\right] = \left[\begin{array}{cc} s & -1 \\ -1 & s \\ \hline 1 & -1 \end{array}\right]$$

and this has rank one for $s = 1$ *and rank two for any other* s*. Therefore, for* $\alpha = 1$*, the unobservable mode is* $s = 1$*.*

Note that, if $\alpha = 1$ *the system does not have a subsystem which is reachable and observable, hence the transfer function is identically equal to zero, i.e. there is no direct input-output connection. To see this, note that*

$$CB = 0 \qquad CAB = 1 - \alpha$$

and, by Cayley-Hamiton Theorem, CA^iB*, with* $i > 1$*, can be written in terms of* CB *and* CAB*.*

Chapter 4

Design tools

4.1 Introduction

In Chapter 3 we have considered the problem of determining an input signal driving the state of the system to a given final condition and the problem of determining the state of the system from measurements of the input and output signals. The proposed solutions rely upon information on the system over a finite and predetermined time interval and generate the required input signal or the required state estimate off-line. Typically, the input signal is computed *a priori* and the state estimate *a posteriori*.

This approach is unsatisfactory for various reasons: the effect of disturbances, model errors and uncertainties is not taken into consideration and the actual evolution of the system is not considered in the solution of the problems, hence these solutions are *open-loop*.

This implies that it makes sense to seek methods, to control or estimate the state of the system, which are based on current information. Such methods provide *closed-loop* solutions to the considered problems.

Closed-loop solutions can be constructed in several ways. In the case of linear, finite-dimensional, time-invariant systems it is natural to solve the above problems considering the interconnection, of the system to be controlled or of the system the state of which has to be estimated, with another linear, finite-dimensional, time-invariant system. This system, which has to be designed, processes the current information and generates the current input to be applied to the system to achieve a specific goal, or the current estimate of the state of the system.

4.2 The notion of feedback

Consider a linear, finite-dimensional, time-invariant system and the problem of determining an input signal such that a certain objective is achieved.

Typically, the input signal has to be such that the state of the system has to be driven to zero with a given speed of convergence (state regulation) or the output of the system has to follow a pre-assigned reference value (output tracking).

We are interested in determining the input signal in closed-loop form, i.e. the input signal at time t has to be a function of the available information (state or output) at the same time instant. This implies that the input signal is generated by means of a feedback mechanism. This mechanism can be instantaneous, i.e. the input signal is generated instantaneously by processing the available information. In this case we have a static feedback. Alternatively, the input signal can be generated processing the available information through a dynamical device. In this case we have a dynamic feedback.

Assume that the system to be controlled is described by equations of the form

$$\sigma x = Ax + Bu \qquad\qquad y = Cx + Du, \tag{4.1}$$

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$, $y(t) \in I\!R^p$ and A, B, C, and D matrices of appropriate dimensions and with constant entries. Assume, in addition, that the system which generates the input signal is linear, finite-dimensional and time-invariant. Then we may have the following four configurations.

- *Static output feedback.* The input signal is generated via the relation

$$u = Ky + v, \tag{4.2}$$

with K a constant matrix of appropriate dimensions and v a new external signal. The resulting closed-loop system is described by the equations

$$\begin{aligned} \sigma x &= (A + B(I + KD)^{-1}KC)x + Bv \\ y &= (I + D(I + KD)^{-1}K)Cx + Dv, \end{aligned} \tag{4.3}$$

which are well-defined provided the matrix $I + KD$ is invertible. Note that this is always the case if $D = 0$.

- *Static state feedback.* The input signal is generated via the relation

$$u = Kx + v, \tag{4.4}$$

with K a constant matrix of appropriate dimensions and v a new external signal. The resulting closed-loop system is described by the equations

$$\begin{aligned} \sigma x &= (A + BK)x + Bv \\ y &= (C + DK)x + Dv. \end{aligned} \tag{4.5}$$

- *Dynamic output feedback.* The input signal is generated by the system[1]

$$\sigma\xi = F\xi + Gy \qquad u = K\xi + v \tag{4.6}$$

with F, G and K constant matrices of appropriate dimensions and v a new external signal. The resulting closed-loop system is described by the equations

$$\begin{aligned} \sigma x &= Ax + BK\xi + Bv \\ \sigma\xi &= (F + GDK)\xi + GCx + GDv \\ y &= Cx + DK\xi + Dv. \end{aligned} \tag{4.7}$$

- *Dynamic state feedback.* The input signal is generated by the system[2]

$$\sigma\xi = F\xi + Gx \qquad u = K\xi + v \tag{4.8}$$

[1] We consider the simplest version of dynamic output feedback. A more general form is given by $\sigma\xi = F\xi + Gy + Hv$, $u = K\xi + Jy + Lv$.

[2] We consider the simplest version of dynamic state feedback. A more general form is given by $\sigma\xi = F\xi + Gx + Hv$, $u = K\xi + Jx + Lv$.

with F, G and K constant matrices of appropriate dimensions and v a new external signal. The resulting closed-loop system is described by the equations

$$\begin{aligned} \sigma x &= Ax + BK\xi + Bv \\ \sigma\xi &= F\xi + Gx \\ y &= Cx + DK\xi + Dv. \end{aligned} \tag{4.9}$$

In what follows we study in detail the static state feedback (Section 4.3) and the dynamic output feedback (Section 4.6) configurations. This is mainly due to the fact that these two configurations allow to solve most control problems for linear systems. Moreover, while the use of static output feedback is very appealing in practice because it results in a simple to implement control strategy, the study of the properties of system (4.3) as a function of K is very difficult[3]. Finally, dynamic state feedback is useful only in very specific problems, such as the so-called noninteracting control problem with stability[4], which are not the subject of these notes.

4.3 State feedback

Consider a system described by the equations (4.1), the state feedback control law (4.4) and the resulting closed-loop system (4.5).

The use of state feedback modifies the input-to-state interaction, i.e. the system $\sigma x = Ax + Bu$ is replaced by the system $\sigma x = (A + BK)x + Bv$. As a result, it makes sense to study which properties of the system are left unchanged by the application of state feedback (i.e. are feedback invariant) and which properties can be modified as a function of K.

Proposition 4.1 *The system (4.1) is reachable if and only if the system (4.5) is reachable.*

Proof. Note that

$$\operatorname{rank}\left[\begin{array}{c|c} sI - A & B \end{array}\right] = \operatorname{rank}\left[\begin{array}{c|c} sI - A & B \end{array}\right]\left[\begin{array}{cc} I & 0 \\ -K & I \end{array}\right] = \operatorname{rank}\left[\begin{array}{c|c} sI - (A + BK) & B \end{array}\right].$$

Hence, by Hautus test, the claim holds. ◁

[3]In the case $m = p = 1$ and $D = 0$ the root locus method can be used to study the eigenvalues of the matrix $A + BKC$.

[4]The noninteracting control problem with stability, in the case $m = p > 1$, can be informally described as the problem of designing a system described by equations of the form (4.8) such that the closed-loop system (4.9) is asymptotically stable and composed of m *decoupled* systems, e.g. the Markov parameters are diagonal matrices.

Proposition 4.2 *Consider the system (4.1). Suppose the system is not reachable and it is described by equations of the form (3.16). Then the system (4.5) is described by equations of the same form. Moreover, the systems (4.1) and (4.5) have the same unreachable modes.*

Proof. By assumption, system (4.1) is described by equations of the form

$$\sigma x = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} x + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u.$$

As a result, system (4.5) is described by

$$\sigma x = \begin{bmatrix} A_{11} + B_1 K_1 & A_{12} + B_1 K_2 \\ 0 & A_{22} \end{bmatrix} x + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} v,$$

where

$$K = \begin{bmatrix} K_1 & K_2 \end{bmatrix},$$

which proves the claim. ◁

This result implies that state feedback does not modify unreachable modes, hence to evaluate the effect of the feedback on the dynamics of the system it is sufficient to consider only the reachable subsystem.

We focus initially on single-input reachable systems.

Proposition 4.3 *Consider system (4.1). Assume $m = 1$ (i.e. the system has only one input) and suppose the system is reachable. Let $p(s)$ be a monic polynomial of degree n. Then there is a (unique) K such that the characteristic polynomial of $A + BK$ is equal to $p(s)$.*

Proof. By reachability of the system it is possible to write the system in reachability canonical form (see Section 3.2.5). Let T be the transformation matrix defined in Section 3.2.5, i.e. let

$$A_r = TAT^{-1} \qquad B_r = TB.$$

Let

$$p(s) = s^n + \tilde{\alpha}_{n-1} s^{n-1} + \tilde{\alpha}_{n-2} s^{n-2} + \cdots + \tilde{\alpha}_1 s + \tilde{\alpha}_0$$

and define

$$K_r = \begin{bmatrix} \alpha_0 - \tilde{\alpha}_0 & \alpha_1 - \tilde{\alpha}_1 & \cdots & \alpha_{n-1} - \tilde{\alpha}_{n-1} \end{bmatrix}.$$

Note that

$$A_r + B_r K_r = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\tilde{\alpha}_0 & -\tilde{\alpha}_1 & -\tilde{\alpha}_2 & \cdots & -\tilde{\alpha}_{n-1} \end{bmatrix},$$

hence $p(s)$ is the characteristic polynomial of $A_r + B_r K_r$. Finally, let

$$K = K_r T$$

and note that

$$A_r + B_r K_r = T(A + BK)T^{-1},$$

which shows that $p(s)$ is the characteristic polynomial of $A + BK$. To prove unicity, let $\hat{K} \neq K$ and define

$$\hat{K}_r = \hat{K} T^{-1},$$

yielding

$$A_r + B_r \hat{K}_r \neq A_r + B_r K_r,$$

hence the characteristic polynomial of $A_r + B_r \hat{K}_r$ is not $p(s)$. $\triangleleft$

The main disadvantage of the above result is that, to compute the feedback gain K which assigns the eigenvalues of the closed-loop system, it is necessary to transform the system in reachability canonical form. This transformation is however not needed, as shown in the following statement.

Proposition 4.4 (Ackermann's formula) *Consider system (4.1). Assume* $m = 1$ *(i.e. the system has only one input) and suppose the system is reachable. Let* $p(s)$ *be a monic polynomial of degree* n*. Then*

$$K = -\begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} R^{-1} p(A)$$

is such that the characteristic polynomial of $A + BK$ *is equal to* $p(s)$*.*

Remark. Propositions 4.3 and 4.4 provide a constructive way to assign the characteristic polynomial, hence the eigenvalues, of system (4.5). Note that, for low order systems, i.e. if $n = 2$ or $n = 3$, it may be convenient to compute directly the characteristic polynomial of $A + BK$ and then compute K using the principle of identity of polynomials, i.e. K should be such that the coefficients of the polynomials $\det(sI - (A + BK))$ and $p(s)$ coincide. $\diamond$

The result summarized in Proposition 4.3 can be extended to multi-input systems.

Proposition 4.5 *Consider system (4.1) and suppose the system is reachable. Let* $p(s)$ *be a monic polynomial of degree* n*. Then there is a* K *such that the characteristic polynomial of* $A + BK$ *is equal to* $p(s)$*.*

Note that in the case $m > 1$ the feedback gain K assigning the characterisitc polynomial of the matrix $A + BK$ is not unique. Finally, to compute such feedback gain we may either use the direct approach discussed in the above Remark or exploit the following fact.

Lemma 4.1 (Heymann) *Consider system (4.1) and suppose the system is reachable. Let b_i be a nonzero column of the matrix B. Then there is a matrix G such that the single-input system*

$$\sigma x = (A + BG)x + b_i v \tag{4.10}$$

is reachable.

Exploiting Lemma 4.1 it is possible to design a matrix K such that the characteristic polynomial of $A + BK$ equals some monic polynomial $p(s)$ of degree n in two steps. First we compute a matrix G such that the system (4.10) is reachable, and then we use Ackermann's formula to compute a matrix k such that the characteristic polynomial of

$$A + BG + b_i k$$

is $p(s)$.

We conclude this section noting that the property of reachability implies the existence of a state feedback gain assigning the eigenvalues of system (4.5). Conversely, it is possible to conclude reachability of system (4.1) by a property of system (4.5).

Proposition 4.6 *System (4.1) is reachable if and only if it is possible to arbitrarily assign the eigenvalues of $A + BK$.*

4.3.1 Stabilizability

The main goal of a state feedback control law is to render the closed-loop system asymptotically stable. This goal may be achieved, as discussed in the previous section, if the system is reachable. However, reachability is not necessary to achieve this goal. In fact, as highlighted in Proposition 4.2, the unreachable modes are not modified by the application of state feedback. This implies that there exists a matrix K such that system (4.5) is asymptotically stable if and only if the unreachable modes of system (4.1) have negative real part, in the case of continuous-time systems, or have modulo smaller than one, in the case of discrete-time systems.

To capture this situation we introduce a new definition.

Definition 4.1 (Stabilizability) *System (4.1) is stabilizable if its unreachable modes have negative real part, in the case of continuous-time systems, or have modulo smaller than one, in the case of discrete-time systems.*

Example 4.1 (Dead-beat control) *Consider a discrete-time system described by equations of the form*

$$x^+ = Ax + Bu,$$

and the problem of designing a static, state feedback control law, described by the equation (4.4), yielding a closed-loop system (4.5) such that, for any initial condition $x(0)$ and for $v = 0$, one has

$$x(k) = 0,$$

for all $k \geq N$, and for some $N > 0$. A control law achieving this goal is called a dead-beat control law. To achieve this goal it is necessary to select K such that

$$(A + BK)^N = 0$$

or, equivalently, such that the matrix $A + BK$ has all eigenvalues equal to 0. Therefore, there exists a dead-beat control law for the considered system if and only if the system is controllable. Note that $N \leq n$.

4.4 The notion of filtering

Consider a linear, finite-dimensional, time-invariant system and the problem of estimating its state from measurements of the input and output signals.

We are interested in determining an on-line estimate, i.e. the estimate at time t has to be a function of the available information (input and output) at the same time instant. This implies that the estimate is generated by means of a device (known as filter) processing the current input and output of the system and generating a state estimate. The filter may be instantaneous, i.e. the estimate is generated instantaneously by processing the available information. In this case we have a static filter. Alternatively, the state estimate can be generated processing the available information through a dynamical device. In this case we have a dynamic filter.

Assume that the system to be controlled is described by equations of the form (4.1) and assume[5] that $D = 0$. Assume, in addition, that the filter which generates the on-line estimate is linear, finite-dimensional and time-invariant. Then we may have the following two configurations.

- *Static filter.* The state estimate is generated via the relation

  $$x_e = My + Nu, \tag{4.11}$$

  with M and N constant matrices of appropriate dimensions. The resulting interconnected system is described by the equations

  $$\begin{aligned} \sigma x &= Ax + Bu \\ x_e &= MCx + Nu. \end{aligned} \tag{4.12}$$

[5] This assumption is without loss of generality. In fact, if $y = Cx + Du$ and u are measurable then also $\tilde{y} = Cx$ is measurable.

- *Dynamic filter.* The state estimate is generated by the system

$$\sigma\xi = F\xi + Ly + Hu \qquad x_e = M\xi + Ny + Pu \tag{4.13}$$

with F, L, H, M, N and P constant matrices of appropriate dimensions. The resulting interconnected system is described by the equations

$$\begin{aligned} \sigma x &= Ax + Bu \\ \sigma\xi &= F\xi + LCx + Hu \\ x_e &= M\xi + NCx + Pu. \end{aligned} \tag{4.14}$$

In what follows we study in detail the dynamic filter configuration. This is mainly due to the fact that this configuration allows to solve most estimation problems for linear systems. Moreover, while the use of a static filter is very appealing, it provides a useful alternative only in very specific situations.

4.5 State observer

A state observer is a filter that allows to estimate, asymptotically or in finite time, the state of a system from measurements of the input and output signals.

The simplest possible observer can be constructed considering a copy of the system, the state of which has to be estimated. This means that a candidate observer for system (4.1) is given by

$$\sigma\xi = A\xi + Bu \qquad x_e = \xi. \tag{4.15}$$

To assess the properties of this candidate state observer let

$$e = x - x_e$$

be the estimation error and note that

$$\sigma e = Ae.$$

As a result, if $e(0) = 0$ then $e(t) = 0$ for all t and for any input signal u. However, if $e(0) \neq 0$ then, for any input signal u, $e(t)$ will be bounded only if the system (4.1) is stable, and will converge to zero only if the system (4.1) is asymptotically stable. If these conditions do not hold, the estimation error will not be bounded and system (4.15) does not qualify as a state observer for system (4.1).

The intrinsic limitation of the observer (4.15) is that it does not use all the available information, i.e. it does not use the knowledge of the output signal y. This observer is therefore an open-loop observer.

To exploit the knowledge of y we modify the observer (4.15) adding a term which depends upon the available information on the estimation error, which is given by

$$y_e = Cx_e - y.$$

This modification yields a candidate state observer described by

$$\sigma\xi = A\xi + Bu + Ly_e \qquad\qquad x_e = \xi. \tag{4.16}$$

To assess the properties of this candidate state observer note that $e = x - x_e$ is such that

$$\sigma e = (A + LC)e. \tag{4.17}$$

The matrix L (known as output injection gain) can be used to shape the dynamics of the estimation error. In particular, we may select L to assign the characteristic polynomical $p(s)$ of $A + LC$. To this end, note that

$$p(s) = \det(sI - (A + LC)) = \det(sI - (A' + C'L')).$$

Hence, there is a matrix L which arbitrarily assigns the characteristic polynomial of $A+LC$ if and only if the system

$$\sigma\xi = A'\xi + C'v$$

is reachable, or equivalently, if and only if the system (4.1) is observable.

We summarize the above discussion with two formal statements.

Proposition 4.7 *Consider system (4.1) and suppose the system is observable. Let $p(s)$ be a monic polynomial of degree n. Then there is*[6] *a matrix L such that the characteristic polynomial of $A + LC$ is equal to $p(s)$.*

Proposition 4.8 *System (4.1) is observable if and only if it is possible to arbitrarily assign the eigenvalues of $A + LC$.*

4.5.1 Detectability

The main goal of a state observer is to provide an on-line estimate of the state of a system. This goal may be achieved, as discussed in the previous section, if the system is observable. However, observability is not necessary to achieve this goal. In fact, similarly to what discussed in Proposition 4.2, the unobservable modes are not modified by the output injection gain. This implies that there exists a matrix L such that system (4.17) is asymptotically stable if and only if the unobservable modes of system (4.1) have negative real part, in the case of continuous-time systems, or have modulo smaller than one, in the case of discrete-time systems.

To capture this situation we introduce a new definition.

Definition 4.2 (Detectability) *System (4.1) is detectable if its unobservable modes have negative real part, in the case of continuous-time systems, or have modulo smaller than one, in the case of discrete-time systems.*

[6] For single-output systems the matrix L assigning the characteristic polynomial of $A + LC$ is unique.

Example 4.2 (Dead-beat observer) *Consider a discrete-time system described by equations of the form*

$$x^+ = Ax + Bu \qquad\qquad y = Cx,$$

and the problem of designing a state observer, described by the equation (4.16), such that, for any initial condition $x(0)$ and for any u,

$$e(k) = 0,$$

for all $k \geq N$, and for some $N > 0$. A state observer achieving this goal is called a dead-beat state observer. To achieve this goal it is necessary to select L such that

$$(A + LC)^N = 0$$

or, equivalently, such that the matrix $A + LC$ has all eigenvalues equal to 0. Therefore, there exists a dead-beat state observer for the considered system if and only if the system is reconstructable. Note that $N \leq n$.

4.5.2 Reduced order observer

We have shown that, under the hypotheses of observability or detectability, it is possible to design an asymptotic observer of order n for the system (4.1). However, this observer is somewhat over-sized, i.e. it gives an estimate for the n components of the state vector, without making use of the fact that some of these components can be directly determined from the output function, *e.g.* if $y = x_1$ there is no need to reconstruct x_1.

Therefore, it makes sense to design a *reduced order observer*, i.e. a device that estimates only the part of the state vector which is not directly attainable from the output. To this end, consider the system (4.1) with $D = 0$ and assume that the matrix C has p independent columns[7]. Then there exists a matrix Q such that

$$QC = [I \;\; C_2].$$

Let

$$v = Qy = QCx = x_1 + C_2 x_2,$$

in which $x_1 \in I\!R^p$ and $x_2 \in I\!R^{n-p}$ denote the first p and the last $n - p$ components of the state x. Observe that the vector v is measurable.

From the definition of v we conclude that if v and x_2 are known then x_1 can be easily computed, i.e. there is no need to construct a dynamic observer for x_1.

Define now the new coordinates

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = Tx = \begin{bmatrix} I & C_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

[7]This is the case if rank$C = p$, whereas if rank$C < p$ it is always possible to eliminate redundant lines.

and note that, by construction,

$$v = Qy = \hat{x}_1.$$

In the new coordinates the system, with output v, is described by equations of the form

$$\begin{aligned} \sigma\hat{x}_1 &= \tilde{A}_{11}\hat{x}_1 + \tilde{A}_{12}\hat{x}_2 + \tilde{B}_1 u \\ \sigma\hat{x}_2 &= \tilde{A}_{21}\hat{x}_1 + \tilde{A}_{22}\hat{x}_2 + \tilde{B}_2 u \\ v &= \hat{x}_1. \end{aligned}$$

In order to construct an observer for $\hat{x}_2$ consider the system

$$\sigma\xi = F\xi + Hv + Gu,$$

with state ξ, driven by u and v, and with output

$$w = \xi + Lv.$$

The idea is to select the matrices F, H, G and L in such a way that w be an estimate for $\hat{x}_2$. Let $w - \hat{x}_2$ be the observation error. Then

$$\begin{aligned} \sigma w - \sigma\hat{x}_2 &= F\xi + Hv + Gu + L\left[\tilde{A}_{11}\hat{x}_1 + \tilde{A}_{12}\hat{x}_2 + \tilde{B}_1 u\right] - \left[\tilde{A}_{21}\hat{x}_1 + \tilde{A}_{22}\hat{x}_2 + \tilde{B}_2 u\right] \\ &= F\xi + \left(H + L\tilde{A}_{11} - \tilde{A}_{12}\right)\hat{x}_1 + \left[L\tilde{A}_{12} - \tilde{A}_{22}\right]\hat{x}_2 + \left[G + L\tilde{B}_1 - \tilde{B}_2\right]u. \end{aligned} \tag{4.18}$$

To have convergence of the estimation error to zero, regardless of the initial conditions and of the input signal, we must have

$$\sigma(w - \hat{x}_2) = F(w - \hat{x}_2) \tag{4.19}$$

and F must have all eigenvalues with negative real part, in the case of continuous-time systems, or with modulo smaller than one, in the case of discrete-time systems.

Comparing equations (4.18) and (4.19), we obtain that the matrices F, H, G, and L must be such that

$$\begin{aligned} L\tilde{A}_{12} - \tilde{A}_{22} &= -F \\ H + L\tilde{A}_{11} - \tilde{A}_{21} &= FL \\ G + L\tilde{B}_1 - \tilde{B}_2 &= 0. \end{aligned}$$

We now show how the previous equations can be solved and how the stability condition of F can be enforced. Detectability of the system implies that the (reduced system)

$$\sigma\tilde{\xi} = \tilde{A}_{22}\tilde{\xi} \qquad \tilde{y} = \tilde{A}_{12}\xi$$

is detectable. As a result, there exists a matrix L such that the matrix

$$F = \tilde{A}_{22} - L\tilde{A}_{12}$$

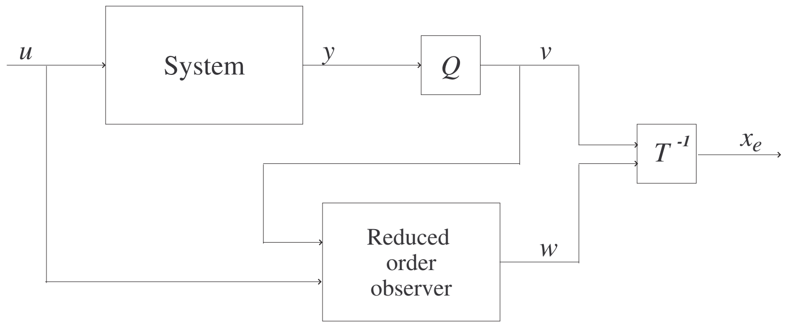


Figure 4.1: A reduced order observer.

has all all eigenvalues with negative real part, in the case of continuous-time systems, or with modulo smaller than one, in the case of discrete-time systems. Then the remaining equations are solved by

$$H = FL - L\tilde{A}_{11} + \tilde{A}_{21} \qquad\qquad G = -L\tilde{B}_1 + \tilde{B}_2.$$

Finally, from $\hat{x}_1 = v$ and the estimate w of $\hat{x}_2$ we build an estimate x_e of the state x inverting the transformation T, *i.e.*

$$\begin{bmatrix} x_{1e} \\ x_{2e} \end{bmatrix} = \begin{bmatrix} I & -C_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}.$$

The (conceptual) structure of the reduced order observer is shown in Figure 4.1.

4.6 The separation principle

In Section 4.3 it has been shown that system (4.1) can be stabilized by means of a state feedback control law, provided the system is stabilizable. Moreover, in Section 4.5 it has been shown that the state of system (4.1) can be (asymptotically) estimated provided the system is detectable.

It is therefore natural to discuss the properties resulting from the use of a state feedback control law in which the state is replaced by an estimate generated by a state observer.

To this end, consider system (4.1) with $D = 0$, the state feedback control law

$$u = Kx + v,$$

the state observer

$$\sigma\xi = (A + LC)\xi + Bu - Ly \qquad\qquad x_e = \xi,$$

and the control law obtained replacing the state x with its estimate x_e, namely

$$u = Kx_e + v.$$

The overall system is described by equations of the form

$$\begin{array}{rcl} \begin{bmatrix} \sigma x \\ \sigma \xi \end{bmatrix} & = & \begin{bmatrix} A & BK \\ -LC & A+LC+BK \end{bmatrix} + \begin{bmatrix} B \\ B \end{bmatrix} v \\ y & = & Cx. \end{array} \tag{4.20}$$

To study this system consider the coordinate

$$e = x - x_e$$

and note that the system can be rewritten in the form

$$\begin{array}{rcl} \begin{bmatrix} \sigma x \\ \sigma e \end{bmatrix} & = & \begin{bmatrix} A+BK & -BK \\ 0 & A+LC \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} v \\ y & = & Cx. \end{array} \tag{4.21}$$

From this representation it is possible to draw the following conclusions.

The characteristic polynomial of the matrix

$$\begin{bmatrix} A+BK & -BK \\ 0 & A+LC \end{bmatrix}$$

is given by the product of the characeristic polynomials of the matrices $A + BK$ and $A + LC$. This result, known as the separation principle, implies that the designs of the state feedback and of the state observer can be carried out independently. Therefore, the problem of asymptotic stabilization of the system (4.1) by means of a dynamic output feedback control law can be solved provided system (4.1) is stabilizable and detectable. The control law stabilizing the system is described by the equations

$$\sigma\xi = (A + BK + LC)\xi - Ly \qquad\qquad u = K\xi + v,$$

i.e. it is a dynamic output feedback control law.

System (4.21), hence system (4.20), is not reachable. In fact, by Hautus test, we note that the unreachable modes are all the eigenvalues of $A+LC$. This implies that the state observer does not contribute to the input-output behaviour of the closed-loop system, i.e.

$$\begin{bmatrix} C & 0 \end{bmatrix} \begin{bmatrix} A+BK & -BK \\ 0 & A+LC \end{bmatrix}^t \begin{bmatrix} B \\ 0 \end{bmatrix} = C(A+BK)^t B,$$

and, similarly,

$$\begin{bmatrix} C & 0 \end{bmatrix} e^{\left(\begin{bmatrix} A+BK & -BK \\ 0 & A+LC \end{bmatrix} t \right)} \begin{bmatrix} B \\ 0 \end{bmatrix} = Ce^{(A+BK)t}B.$$

Therefore the input-output behaviour of the closed-loop system resulting from the use of an output feedback controller designed on the basis of the separation principle coincides with the input-output behaviour of the closed-loop system resulting from the use of the underlying state feedback controller.

4.7 Tracking and regulation

In the previous sections we have considered the simplest possible design problems, namely the stabilization and observation problems. Practical problems, however, present themselves in a more complex form. In particular, the system to be controlled may be affected by disturbances, and the output of the system does not have to be regulated to zero, but should asymptotically track a certain, prespecified, reference signal.

In this section we discuss this control problem and present possible solutions. To begin with, consider a system to be controlled described by equations of the form

$$\sigma x = Ax + Bu + Pd \qquad e = Cx + Qd, \tag{4.22}$$

with $x \in X = I\!R^n$, $u(t) \in I\!R^m$, $e(t) \in I\!R^p$, $d(t) \in I\!R^r$, and A, B, P, C and Q matrices of appropriate dimensions and with constant entries.

The signal $d(t)$, denoted exogeneous signal, is in general composed of two components: the former models a set of disturbances acting on the system to be controlled, the latter a set of reference signals. In what follows we assume that the exogeneous signal is generated by a linear system, denoted exosystem, described by the equation

$$\sigma d = Sd, \tag{4.23}$$

with S a matrix with constant entries. Note that, under this assumption, it is possible to generate, for example, constant or polynomial references/disturbances and sinusoidal references/disturbances with any given frequency.

The variable $e(t)$, denoted tracking error, is a measure of the error between the ideal behaviour of the system and the actual behaviour. Ideally, the variable $e(t)$ should be regulated to zero, i.e. should converge asymptotically to zero, despite the presence of the disturbances. If this happens we say that the tracking error is regulated to zero, i.e. converges asymptotically to zero, hence the disturbances are not affecting the asymptotic behaviour of the system and the output $Cx(t)$ is asymptotically tracking the reference signal $-Qd(t)$.

In general, the tracking error does not naturally converge to zero, hence it is necessary to determine an input signal $u(t)$ which *drives* it to zero. The simplest possible way to construct such an input signal is to assume that it is generated via static feedback of the state $x(t)$ of the system to be controlled and of the state $d(t)$ of the exosystem, i.e.

$$u = Kx + Ld. \tag{4.24}$$

In practice, it is unrealistic to assume that both $x(t)$ and $d(t)$ are measurable, hence it may be more natural to assume that the input signal $u(t)$ is generated via dynamic feedback of the error signal only, i.e. it is generated by the system

$$\sigma \chi = F\chi + Ge \qquad u = H\chi, \tag{4.25}$$

with $\chi(t) \in I\!R^\nu$, for some $\nu > 0$, and F, G and H matrices with constant entries.

In summary, it is possible to formally pose the regulator problem as follows.

Definition 4.3 (Full information regulator problem) *Consider the system (4.22), driven by the exosystem (4.23) and interconnected with the controller (4.24). The full information regulator problem is the problem of determining the matrices K and L of the controller such that*[8]

(S) the system

$$\sigma x = (A + BK)x$$

is asymptotically stable;

(R) all trajectories of the system

$$\sigma d = Sd \qquad \sigma x = (A + BK)x + (BL + P)d \qquad e = Cx + Qd \tag{4.26}$$

are such that

$$\lim_{t\to\infty} e(t) = 0.$$

Definition 4.4 (Error feedback regulator problem) *Consider the system (4.22), driven by the exosystem (4.23) and interconnected with the controller (4.25). The error feedback regulator problem is the problem of determining the matrices F, G and H of the controller such that*

(S) the system

$$\sigma x = Ax + BH\chi \qquad \sigma\chi = F\chi + GCx$$

is asymptotically stable;

(R) all trajectories of the system

$$\sigma d = Sd \quad \sigma x = Ax + BH\chi + Pd \quad \sigma\chi = F\chi + G(Cx + Qd) \quad e = Cx + Qd \tag{4.27}$$

are such that

$$\lim_{t\to\infty} e(t) = 0.$$

4.7.1 The full information regulator problem

Consider the full information regulator problem and assume the following.

Assumption 4.1 *The matrix S of the exosystem has all eigenvalues with non-negative real part, in the case of continuous-time systems, or with modulo not smaller than one, in the case of discrete-time systems.*

Assumption 4.2 *The system (4.22) with $d = 0$ is reachable.*

[8] (S) stands for stability and (R) for regulation.

Assumption 4.1 implies that there are no initial conditions $d(0)$ such that the signal $d(t)$ converges (asymptotically) to zero. This assumption is not restrictive. In fact, disturbances converging to zero do not have any effect on the asymptotic behaviour of the system, and references which converge to zero can be tracked simply by driving the state of the system to zero, i.e. by stabilizing the system.

Assumption 4.2 implies that it is possible to arbitrarily assign the eigenvalues of the matrix $A+BK$ by a proper selection of K. Note that, in practice, this assumption can be replaced by the weaker assumption that the system (4.22) with $d = 0$ is stabilizable.

We now present a preliminary result which is instrumental to derive a solution to the full information regulator problem.

Lemma 4.2 *Consider the full information regulator problem. Suppose Assumption 4.1 holds. Suppose, in addition, that there exists matrices K and L such that condition (S) holds.*

Then condition (R) holds if and only if there exists a matrix $\Pi \in I\!R^{n\times r}$ such that the equations

$$\Pi S = (A + BK)\Pi + (P + BL) \qquad\qquad 0 = C\Pi + Q \tag{4.28}$$

hold.

Proof. Consider the system (4.26) and the coordinates transformation

$$\hat{d} = d \qquad\qquad \hat{x} = x - \Pi d,$$

where Π is the solution of the equation[9]

$$\Pi S = (A + BK)\Pi + (P + BL).$$

Note that, by condition (S) and Assumption 4.1, there is a unique matrix Π which solves this equation. In the new coordinates $\hat{x}$ and $\hat{d}$ the system is described by the equations

$$\sigma\hat{d} = S\hat{d} \qquad \sigma\hat{x} = (A + BK)\hat{x} \qquad e = C\hat{x} + (C\Pi + Q)\hat{d}.$$

Note now that, by condition (S) $\lim_{t\to\infty} \hat{x} = 0$, hence condition (R) holds, by Assumption 4.1, if and only if

$$C\Pi + Q = 0.$$

In summary, under the state assumptions, condition (R) holds if and only if there exists a matrix Π such that equations (4.28) hold. $\triangleleft$

We are now ready to state and prove the result which provides conditions for the solution of the full information regulator problem.

[9] This equation is a so-called Sylvester equation. The Sylvester equation is a (matrix) equation of the form

$$A_1 X = X A_2 + A_3,$$

in the unknown X. This equation has a unique solution, for any A_3, if and only if the matrices A_1 and A_2 do not have common eigenvalues.

Theorem 4.1 *Consider the full information regulator problem. Suppose Assumptions 4.1 and 4.2 hold.*

There exists a full information control law described by the equation (4.24) which solves the full information regulator problem if and only if there exist two matrices Π *and* Γ *such that the equations*

$$\Pi S = A\Pi + B\Gamma + P \qquad\qquad 0 = C\Pi + Q \tag{4.29}$$

hold.

Proof. (Necessity) Suppose there exist two matrices K and L such that conditions (S) and (R) of the full information regulator problem hold. Then, by Lemma 4.2, there exists a matrix Π such that equations (4.28) hold. As a result, the matrices Π and $\Gamma = K\Pi + L$ are such that equations (4.29) hold.
(Sufficiency) The proof of the sufficiency is constructive. Suppose there are two matrices Π and Γ such that equations (4.29) hold. The full information regulator problem is solved selecting K and L as follows.

The matrix K is any matrix such that the system

$$\sigma x = (A + BK)x$$

is asymptotically stable. By Assumption 4.2 such a matrix K does exist.

The matrix L is selected as

$$L = \Gamma - K\Pi.$$

This selection is such that condition (S) of the full information regulator problem holds, hence to complete the proof we have only to show that, with K and L as selected above, the equations (4.28) hold. This is trivially the case. In fact, replacing L in (4.28) yields the equations (4.29), which hold by assumption. As a result, also condition (R) of the full information regulator problem holds, and this completes the proof. ◁

The proof of Theorem 4.1 implies that a controller (it is not the only one) which solves the full information regulator problem is described by the equation

$$u = Kx + (\Gamma - K\Pi)d,$$

with K such that a stability condition holds, and Π and Γ such that equations (4.29) hold. By Assumption 4.2 the stability condition can be always satisfied. As a result, the solution of the full information regulator problem relies upon the existence of a solution of equations (4.29).

4.7.2 The FBI equations

Equations (4.29), known as the Francis-Byrnes-Isidori (FBI) equations, are linear equations in the unknown Π and Γ, for which the following statement holds.

Lemma 4.3 (Hautus) *The equations (4.29), in the unknown* Π *and* Γ*, are solvable for any* P *and* Q *if and only if*

$$\operatorname{rank}\begin{bmatrix} sI - A & B \\ C & 0 \end{bmatrix} = n + p, \tag{4.30}$$

for all s *which are eigenvalues of the matrix* S.

Remark. The equations (4.29) can be rewritten in compact form as

$$\begin{bmatrix} A & B \\ C & 0 \end{bmatrix}\begin{bmatrix} \Pi \\ \Gamma \end{bmatrix} - \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Pi \\ \Gamma \end{bmatrix} S = \begin{bmatrix} -P \\ -Q \end{bmatrix},$$

which is a so-called generalized Sylvester equation. ◇

For single-input, single-output systems (i.e. $m = p = 1$) the condition expressed by Lemma 4.3 has a very simple interpretation. In fact, the complex number s such that

$$\operatorname{rank}\begin{bmatrix} sI - A & B \\ C & 0 \end{bmatrix} < n + 1$$

are the zeros of the system

$$\sigma x = Ax + Bu \qquad\qquad y = Cx,$$

which coincides with the roots of the numerator polynomial of the transfer function

$$W(s) = C(sI - A)^{-1}B,$$

i.e. the zeros of $W(s)$. This implies that, for single-input, single-output systems the full information regulator problem is solvable if and only if the eigenvalues of the exosystem are not zeros of the transfer function of the system (4.22), with input u, output e and $d = 0$.

4.7.3 The error feedback regulator problem

To provide a solution to the error feedback regulator problem we need to introduce a new assumption.

Assumption 4.3 *The system*

$$\begin{bmatrix} \sigma x \\ \sigma d \end{bmatrix} = \begin{bmatrix} A & P \\ 0 & S \end{bmatrix}\begin{bmatrix} x \\ d \end{bmatrix} \qquad\qquad e = \begin{bmatrix} C & Q \end{bmatrix}\begin{bmatrix} x \\ d \end{bmatrix} \tag{4.31}$$

is observable.

Note that Assumption 4.3 implies observability of the system

$$\sigma x = Ax \qquad y = Cx. \tag{4.32}$$

To prove this property note that observability of the system (4.31) implies that

$$\operatorname{rank}\begin{bmatrix} C & Q \\ CA & \star \\ \vdots & \vdots \\ CA^{n+r-1} & \star \end{bmatrix} = n + r.$$

This, in turn, implies

$$\operatorname{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n+r-1} \end{bmatrix} = n$$

and, by Cayley-Hamilton Theorem,

$$\operatorname{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = n,$$

which implies observability of system (4.32). Similarly to what discussed in the case of Assumption 4.2, Assumption 4.3 can be replaced by the weaker assumption that the system (4.31) is detectable.

We are now ready to state and prove the result which provides conditions for the solution of the error feedback regulator problem.

Theorem 4.2 *Consider the error feedback regulator problem. Suppose Assumptions 4.1, 4.2 and 4.3 hold.*

There exists an error feedback control law described by the equation (4.25) which solves the full information regulator problem if and only if there exist two matrices Π *and* Γ *such that the equations*

$$\Pi S = A\Pi + B\Gamma + P \qquad 0 = C\Pi + Q \tag{4.33}$$

hold.

Remark. Theorem 4.2 can be alternatively stated as follows.

Consider the error feedback regulator problem. Suppose Assumptions 4.1, 4.2 and 4.3 hold. Then the error feedback regulator problem is solvable if and only if the full information regulator problem is solvable. ◇

Proof. (Necessity) The proof of the necessity is similar to the proof of the necessity of Theorem 4.1, hence omitted.
(Sufficiency) The proof of the sufficiency is constructive. Suppose there are two matrices Π and Γ such that equations (4.33) hold. Then, by Theorem 4.1 the full information control law

$$u = Kx + (\Gamma - K\Pi)d,$$

with K such that the system $\sigma x = (A + BK)x$ is asymptoticaly stable, solves the full information regulator problem. This control law is not implementable, because we only measure e. However, by Assumption 4.3, it is possible to build asymptotic estimates ξ and δ of x and d, hence implement the control law

$$u = K\xi + (\Gamma - K\Pi)\delta. \tag{4.34}$$

To this end, consider an observer described by the equation

$$\begin{bmatrix} \sigma\xi \\ \sigma\delta \end{bmatrix} = \begin{bmatrix} A & P \\ 0 & S \end{bmatrix} \begin{bmatrix} \xi \\ \delta \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \left(\begin{bmatrix} C & Q \end{bmatrix} \begin{bmatrix} \xi \\ \delta \end{bmatrix} - e \right) + \begin{bmatrix} B \\ 0 \end{bmatrix} \begin{bmatrix} K & \Gamma - K\Pi \end{bmatrix} \begin{bmatrix} \xi \\ \delta \end{bmatrix}.$$

Note that the estimation errors $e_x = x - \xi$ and $e_d = d - \delta$ are such that

$$\begin{bmatrix} \sigma e_x \\ \sigma e_d \end{bmatrix} = \left(\begin{bmatrix} A & P \\ 0 & S \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \begin{bmatrix} C & Q \end{bmatrix} \right) \begin{bmatrix} e_x \\ e_d \end{bmatrix}, \tag{4.35}$$

hence, by Assumption 4.3, there exist G_1 and G_2 that assign the eigenvalues of this error system.

Note now that the control law (4.34) can be rewritten as

$$u = Kx + (\Gamma - K\Pi)d - (Ke_x + (\Gamma - K\Pi)e_d),$$

hence the control law is composed of the full information control law, which solves the considered regulator problem, and of an additive disturbance which decays exponentially to zero. Such a disturbance does not affect the regulation requirement, provided the closed-loop system is asymptotically stable. Therefore, to complete the proof we need to show that condition (S) holds. For, note that, in the coordinates x, e_x and e_d the closed-loop system, with $d = 0$, is described by the equations

$$\begin{bmatrix} \sigma x \\ \sigma e_x \\ \sigma e_d \end{bmatrix} = \begin{bmatrix} A + BK & -BK & -B(\Gamma - K\Pi) \\ 0 & A + G_1C & P + G_1Q \\ 0 & G_2C & S + G_2Q \end{bmatrix} \begin{bmatrix} x \\ e_x \\ e_d \end{bmatrix}. \tag{4.36}$$

Recall that the matrices G_1 and G_2 have been selected to render system (4.35) asymptotically stable, and that K is such that the system $\sigma x = (A + BK)x$ is asymptotically stable. As a result, system (4.36) is asymptotically stable. $\triangleleft$

4.7.4 The internal model principle

The proof of Theorem 4.2 implies that a controller (it is not the only one) which solves the error feedback regulator problem is described by equations of the form (4.25) with $\chi = \begin{bmatrix} \xi & \delta \end{bmatrix}'$,

$$F = \begin{bmatrix} A + G_1C + BK & P + G_1Q + B(\Gamma - K\Pi) \\ G_2C & S + G_2Q \end{bmatrix},$$
$$G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}, \qquad H = \begin{bmatrix} K & \Gamma - K\Pi \end{bmatrix}, \tag{4.37}$$

K, G_1 and G_2 such that a stability condition holds, and Π and Γ such that equations (4.33) hold. This controller, and in particular the matrix F, possesses a very interesting property.

Proposition 4.9 (Internal model property) *The matrix F in equation (4.37) is such that*

$$F\Sigma = \Sigma S,$$

for some matrix Σ of rank r. In particular, any eigenvalue of S is also an eigevalue of F.

Proof. Let

$$\Sigma = \begin{bmatrix} \Pi \\ I \end{bmatrix}$$

and note that rank$\Sigma = r$, by construction, and that

$$\begin{aligned} F\Sigma &= \begin{bmatrix} A\Pi + G_1C\Pi + BK\Pi + P + G_1Q + B(\Gamma - K\Pi) \\ -G_2C\Pi + S - G_2Q \end{bmatrix} \\ &= \begin{bmatrix} (A\Pi + B\Gamma + P) + G_1(C\Pi + Q) \\ S - G_2(C\Pi + Q) \end{bmatrix} = \begin{bmatrix} \Pi S \\ S \end{bmatrix} = \Sigma S, \end{aligned}$$

hence the first claim. To prove the second claim, let λ be an eigenvalue of S and v the corresponding eigenvector. Then $Sv = \lambda v$, hence

$$F\Sigma v = \Sigma S v = \lambda \Sigma v,$$

which shows that λ is an eigenvalue of F with eigenvector Σv, and this proves the second claim. ◁

It is possible to prove that the property highlighted in Proposition 4.9 is shared by all error feedback control laws which solve the considered regulation problem, and not only the proposed controller. This property, which is often referred to as the internal model principle, can be interpreted as follows. The control law solving the regulator problem has to *contain* a copy of the exosystem, i.e. it has to be able to generate, when $e = 0$, a copy of the exogeneous signal.

Exercises

Exercise 1 *Consider the matrix*

$$A = \begin{bmatrix} 7 & -6 & 2 \\ 8.8 & -7.6 & 2.8 \\ 9.6 & -7.2 & 2.6 \end{bmatrix}.$$

1. *Compute the characteristic polynomial of A and hence show that the eigenvalues of A are 1, -1 and 2.*

2. *Compute three linearly independent eigenvectors.*

3. *Find a similarity transformation L such that $\hat{A} = L^{-1}AL$ is a diagonal matrix.*

4. *Determine $\exp(At)$ as a function of t.*

5. *Determine $\sin(At)$ as a function of t.*

Solution 1

1. The characteristic polynomal is

$$\det(sI - A) = s^3 - 2s^2 - s + 2 = (s^2 - 1)(s - 2),$$

hence the claim.

2. The matrix A has three distinct eigenvalues, hence has three linearly independent eigenvectors. These are computed solving the equation

$$Av = \lambda v$$

with $\lambda = 1, -1, 2$. The eigenvector associated to $\lambda = 1$ is $v_1 = [1/2, 1, 3/2]'$; the eigenvector associated to $\lambda = -1$ is $v_2 = [1, 4/3, 0]'$; the eigenvector associated to $\lambda = 2$ is $v_3 = [1, 1/2, 2]'$. Note that, as suggested by the theory,

$$\det([v_1, v_2, v_3]) \neq 0,$$

hence the three vectors are linearly independent.

3. Let $V = [v_1, v_2, v_3]$ and

$$\hat{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Note that

$$AV = V\hat{A},$$

hence, $L = V$ is such that $L^{-1}AL$ is a diagonal matrix.

4. By a property of the matrix exponential

$$e^{At} = e^{(V\hat{A}V^{-1})t} = Ve^{\hat{A}t}V^{-1},$$

where

$$e^{\hat{A}t} = \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{2t} \end{bmatrix}.$$

5. Similarly to the definition of the matrix exponential we have

$$\sin(At) = At - \frac{(At)^3}{3!} + \frac{(At)^5}{5!} - \cdots,$$

hence

$$\sin(At) = V\sin(\hat{A}t)V^{-1},$$

with

$$\sin(\hat{A}t) = \begin{bmatrix} \sin t & 0 & 0 \\ 0 & -\sin t & 0 \\ 0 & 0 & \sin 2t \end{bmatrix}.$$

Exercise 2 *Let A be a square matrix of dimension n and let $\lambda_1, \cdots \lambda_n$ denote its eigenvalues. Assume for simplicity that A has n distinct eigenvalues.*

1. *Show that the determinant of A is equal to the product of the eigenvalues of A:*

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n.$$

2. *Show that the trace of A is equal to the sum of the eigenvalues of A:*

$$\text{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

($\text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$.)

3. *Show that if $\text{tr}(A) > 0$, the system $\dot{x} = Ax$ is unstable. Is the converse true?*

4. *Show that if the system $\dot{x} = Ax$ is asymptotically stable then $\text{tr}(A) < 0$.*

5. *Assume $n = 3$, $\mathrm{tr}(A) = T$ and $\det(A) = \omega^2 T$. Assume moreover that A has two purely imaginary eigenvalues. Discuss the stability of the system $\dot{x} = Ax$ as a function of T and ω and explain how to compute exp(At) as a function of t if the eigenvectors of A are known.*

Solution 2 If A has n distinct eigenvalues then there exists a nonsingular matrix L such that

$$L^{-1}AL = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \tilde{A}$$

for some $\lambda_i \in \mathbb{C}$.

1. Note that

$$\det(A) = \det(L)\,\det(L^{-1})\,\det(\tilde{A}) = \lambda_1\lambda_2\cdots\lambda_n.$$

2. The characteristic polynomial of A is given by

$$\det(sI - A) = (s - a_{11})(s - a_{22})\cdots(s - a_n) + \cdots = s^n + \alpha_{n-1}s^{n-1} + \cdots$$

with

$$\alpha_{n-1} = -a_{11} - a_{22} - \cdots - a_{nn} = -\mathrm{tr}(A).$$

Moreover

$$\det(sI - A) = (s - \lambda_1)(s - \lambda_2)\cdots(s - \lambda_n) = s^n - (\lambda_1 + \lambda_2 + \cdots + \lambda_n)s^{n-1} + \cdots.$$

Hence

$$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

3. If $\mathrm{tr}(A) > 0$ then there exists a λ_i with positive real part. Hence the system $\dot{x} = Ax$ is unstable. Instability of the system $\dot{x} = Ax$ does not imply $\mathrm{tr}(A) > 0$, as the sum of the real parts of the eigenvalues may be negative even if some of the real parts are positive.

4. If the system $\dot{x} = Ax$ is asymptotically stable then the real parts of all eigenvalues are negative, hence $\mathrm{tr}(A) < 0$.

5. Let $n = 3$ and, as stated,

$$\lambda_1 = -i\omega \qquad \lambda_2 = i\omega \qquad \lambda_3 = \alpha,$$

then

$$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \lambda_3 = T \qquad \det(A) = \lambda_1\lambda_2\lambda_3 = \omega^2 T.$$

As a result, $\alpha = T$. Hence the system $\dot{x} = Ax$ is stable, not asymptotically if $T \leq 0$ and it is unstable if $T > 0$. Finally let v_1, v_2 and v_3 be the eigenvectors of A corresponding to the eigenvalues λ_1, λ_2 and λ_3. Hence

$$AV = A \left[\begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \right] = \left[\begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \right] \left[\begin{array}{ccc} -i\omega & 0 & 0 \\ 0 & i\omega & 0 \\ 0 & 0 & \alpha \end{array} \right] = V\Lambda.$$

As a result $A = V\Lambda V^{-1}$, hence

$$e^{At} = Ve^{\Lambda t}V^{-1}$$

with

$$e^{\Lambda t} = \left[\begin{array}{ccc} e^{-i\omega t} & 0 & 0 \\ 0 & e^{i\omega t} & 0 \\ 0 & 0 & e^{\alpha t} \end{array} \right].$$

Exercise 3 *Consider the discrete-time system* $x_{k+1} = Ax_k$.

1. *Let*

$$A = \left[\begin{array}{cc} -1 & 1 \\ 0 & -1 \end{array} \right].$$

Consider the initial state

$$x(0) = \left[\begin{array}{c} 0 \\ 1 \end{array} \right]$$

and plot $x(k)$ *on the state space for* $k = 1, 2, 3, 4$. *Exploiting the obtained result discuss the stability of the equilibrium* $x = 0$.

2. *Let*

$$A = \left[\begin{array}{cc} -1 & 1 \\ 0 & 1 \end{array} \right].$$

Consider the initial state

$$x(0) = \left[\begin{array}{c} 0 \\ 1 \end{array} \right]$$

and plot $x(k)$ *on the state space for* $k = 1, 2, 3, 4$. *Exploiting the obtained result discuss the stability of the equilibrium* $x = 0$.

Solution 3

1. Note that

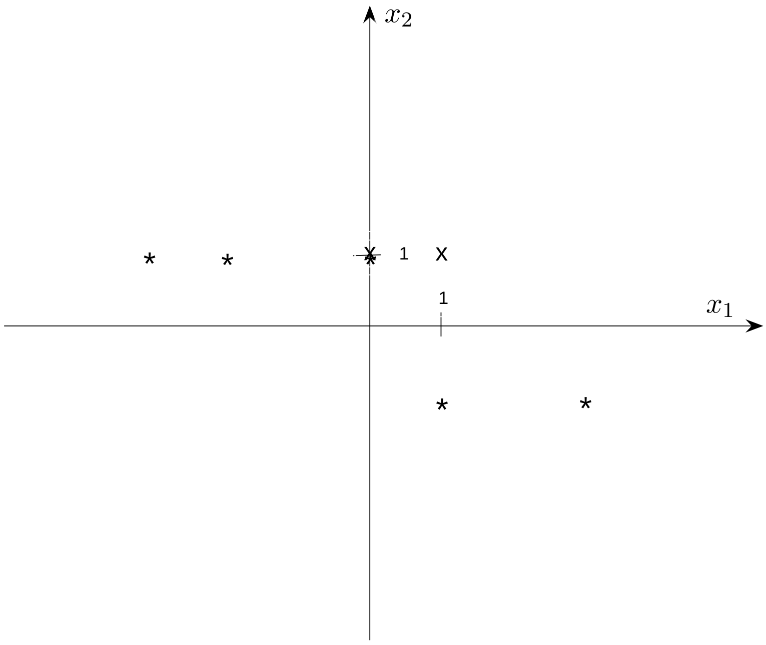
$$x(1) = \left[\begin{array}{c} 1 \\ -1 \end{array} \right] \quad x(2) = \left[\begin{array}{c} -2 \\ 1 \end{array} \right] \quad x(3) = \left[\begin{array}{c} 3 \\ -1 \end{array} \right] \quad x(4) = \left[\begin{array}{c} -4 \\ 1 \end{array} \right],$$

and these are indicated in the figure with $\star$ signs. This implies that the equilibrium $x = 0$ is unstable. (Note that to decide instability of an equilibrium it is enough that one trajectory does not satisfy the "$\epsilon - \delta$" argument in the definition of stability.)

2. Note that

$$x(1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad x(2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad x(3) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad x(4) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and these are indicated in the figure with X signs. This trajectory is such that the "$\epsilon - \delta$" argument holds, however we cannot conclude stability of the equilibrium $x = 0$ only from properties of one trajectory.



Exercise 4 *Consider the discrete-time system*

$$x_{k+1} = Ax_k = \begin{bmatrix} 1 & 1 \\ a & -1 \end{bmatrix} x_k$$

with $a \in I\!R$.

1. *Show that the system is asymptotically stable for all* $a \in (-2, 0)$ *and it is unstable for* $a < -2$ *and* $a > 0$.

2. *Let* $a = 0$*. Discuss the stability properties of the system.*

3. *Let* $a = -2$*. Discuss the stability properties of the system.*

Solution 4 The characteristic polynomial of the matrix A is

$$\det(sI - A) = (s-1)(s+1) - a = s^2 - 1 - a.$$

Hence the eigenvalues of A are

$$\lambda_1 = +\sqrt{1+a} \qquad \lambda_2 = -\sqrt{1+a}.$$

1. The system is asymptotically stable if (and only if)

   $$|\lambda_1| < 1 \qquad |\lambda_2| < 1.$$

   Observe that λ_1 and λ_2 are real if $a \geq -1$ and are imaginary if $a < -1$. Moreover, $|\lambda_1| = |\lambda_2|$ and this is smaller than one if (and only if) $a \in (-2, 0)$.

2. If $a = 0$ the eigenvalues are real and have modulo equal to one. Moreover they are distinct, hence they have geometric and algebraic multiplicity equal to one. Therefore the system is stable, not asymptotically.

3. If $a = -2$ the eigenvalues are imaginary and have modulo equal to one. Moreover they are distinct, hence they have geometric and algebraic multiplicity equal to one. Therefore the system is stable, not asymptotically.

Exercise 5 *Consider two systems* $\sigma x_i = A_i x_i + B_i u$*,* $y_i = C_i x_i$*, with* $i = 1, 2$ *and*

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \qquad B_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad C_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix} \qquad B_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad C_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Note that the systems have the same input.

1. *Show that if* $x_1(0) = x_2(0) = 0$ *then* $y_1(t) = y_2(t)$ *for all* t*. (This is equivalent to* $C_1 A_1^k B_1 = C_2 A_2^k B_2$ *for all* $k \geq 0$*, in discrete-time and to* $C_1 e^{A_1 t} B_1 = C_2 e^{A_2 t} B_2$ *for all* $t \geq 0$*, in continuous-time.)*

2. *The two systems are not algebraically equivalent.*

Solution 5

1. If $x_1(0) = x_2(0) = 0$ then (suppose the systems are discrete-time)

$$y_1(k) = \sum_{i=0}^{k-1} C_1 A_1^{k-1-i} B_1 u(i) \qquad\qquad y_2(k) = \sum_{i=0}^{k-1} C_2 A_2^{k-1-i} B_2 u(i)$$

   Note that $A_1 = A_2'$, $B_1 = C_2'$ and $C_1 = B_2'$. Hence

$$C_1 A_1^k B_1 = (C_1 A_1^k B_1)' = B_1'(A_1')^k C_1' = C_2 A_2^k B_2$$

   which proves that the output responses of the two systems coincide.

2. The systems are algebraically equivalent if and only if there exists a nonsingular L such that

$$A_1 = L^{-1}A_2L \quad (\text{or } LA_1 = A_2L) \qquad B_1 = L^{-1}B_2 \quad (\text{or } LB_1 = B_2) \qquad C_1 = C_2L.$$

   Setting

$$L = \begin{bmatrix} L_1 & L_2 \\ L_3 & L_4 \end{bmatrix}$$

   yields

$$LB_1 = \begin{bmatrix} L_1 \\ L_3 \end{bmatrix} = B_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 1 & 1 \end{bmatrix} = C_2L = \begin{bmatrix} L_1 & L_2 \end{bmatrix},$$

   hence $L_1 = L_2 = L_3 = 1$. Finally, for any L_4,

$$LA_1 = \begin{bmatrix} 1 & 3 \\ 1 & 1+2L_4 \end{bmatrix} \neq A_2L = \begin{bmatrix} 1 & 1 \\ 1 & 3+2L_4 \end{bmatrix},$$

   which shows that the two systems are not algebraically equivalent.
   It is worth noting that the system with state x_1 is not reachable but observable, whereas the system with state x_2 is reachable but not observable (it is the dual of the system with state x_1). Finally, if two systems are reachable and observable and the forced responses of the output coincide then the systems are algebraically equivalent.

Exercise 6 *The function $y = x - \log x$ has a minimum for $x = 1$. This minimum can be computed using Newton's algorithm which yields the discrete-time system*

$$x_{k+1} = 2x_k - x_k^2.$$

1. *Compute the equilibrium points of this system.*

2. *Study the stability properties of the equilibrium points.*

3. *Show that*

   (a) *if* $x_0 = 2$ *or* $x_0 = 0$ *then* $x_k = 0$*, for all* $k \geq 1$*;*
   (b) *if* $x_0 > 2$ *then* $x_{k+1} < x_k < 0$*, for all* $k \geq 1$*;*
   (c) *if* $x_0 \in (0, 2)$ *then* $x_k \in (0, 1)$*, for all* $k \geq 1$*, and* $\lim_{k\to\infty} x_k = 1$*.*

Solution 6

1. At an equilibrium point of a discrete-time system we have $x_{k+1} = x_k$. Hence, the equilibrium points of the system are the solutions of the equation

$$x = 2x - x^2.$$

   We have therefore two equilibria

$$x = 0 \qquad\qquad x = 1.$$

2. The linearization of the system around an equilibrium point x_{eq} is described by

$$(\delta_x)_{k+1} = (2 - 2x_{eq})(\delta_x)_k.$$

   Hence, for $x_{eq} = 0$ we have $(\delta_x)_{k+1} = 2(\delta_x)_k$, which is an unstable system, while for $x_{eq} = 1$ we have $(\delta_x)_{k+1} = 0$, which is a stable system. As a result, by the principle of stability in the first approximation, the equilibrium point $x = 0$ is unstable and the equilibrium point $x = 1$ is locally asymptotically stable.

3. If $x_0 = 2$, then $x_1 = 0$, and by definition of equilibrium $x_k = 0$ for all $k > 1$. If $x_0 = 0$ then, again by definition of equilibrium, $x_k = 0$, for all $k > 0$.
   Note now that the relation

$$x_{k+1} = 2x_k - x_k^2,$$

   implies $x_{k+1} < x_k$ if and only if $x_k < 0$ or $x_k > 1$.
   If $x_0 > 2$ then $x_1 < 0$, hence $x_2 < x_1 < 0$ and, in general,

$$x_{k+1} < x_k < \cdots < x_2 < x_1 < 0.$$

   This shows that, for any $x_0 > 2$ the sequence $\{x_k\}$ diverges, i.e. $\lim_{k\to\infty} x_k = -\infty$.
   If $x_0 \in (0, 2)$ then $x_1 \in (0, 1)$. Moreover, $x_1 \in (0, 1)$ implies that $x_2 \in (0, 1)$, and so on. Therefore, if $x_0 \in (0, 2)$ then $x_k \in (0, 1)$ for all $k > 0$. Note now that if $x_k \in (0, 1)$ then $x_{k+1} > x_k$. As a result, for any $x_0 \in (0, 2)$ we have

$$0 < x_1 < x_2 < \cdots < x_k < x_{k+1} < 1,$$

   which states that the sequence $\{x_k\}$ is monotonically increasing and bounded from above, hence should have a limit. However, any limit of the sequence has to be an equilibrium point of the system, which implies that, for any $x_0 \in (0, 2)$ we have that $\lim_{k\to\infty} x_k = 1$.

Exercise 7 *Consider the discrete-time system*

$$\begin{aligned} x_{1,k+1} &= \frac{1}{2}x_{1,k} + \alpha(x_{2,k})x_{2,k} \\ x_{2,k+1} &= x_{2,k} - \alpha(x_{2,k})x_{1,k} - r(x_{2,k})x_{2,k} \end{aligned}$$

with

$$r(x_{2,k}) = \frac{2}{3}(\alpha(x_{2,k}))^2,$$

and $\alpha(\cdot)$ *a differentiable function.*

1. *Show that if*

$$0 < |\alpha(x_{2,k})| < \frac{\sqrt{3}}{2}$$

the origin is a (locally) asymptotically stable equilibrium.

Solution 7 To study the stability of the origin we use the principle of stability in the first approximation.

1. The linearization of the system around the zero equilibrium is described by

$$\delta_x(k+1) = A\delta_x(k) = \begin{bmatrix} \frac{1}{2} & \alpha \\ -\alpha & 1 - \frac{2}{3}\alpha^2 \end{bmatrix} \delta_x(k).$$

The characteristic polynomial of the matrix A is

$$\det(sI - A) = s^2 + s\left(\frac{2}{3}\alpha^2 - \frac{3}{2}\right) + \left(\frac{2}{3}\alpha^2 + \frac{1}{2}\right) = s^2 + s\eta + (\eta + 2)$$

with $\eta = \left(\frac{2}{3}\alpha^2 - \frac{3}{2}\right)$. Note that we are interested in values of α^2 in the set $(0, 3/4)$ and hence of η in the set $(-3/2, -1)$. We need to show that, for $\eta \in (-3/2, -1)$, the roots of the characteristic polynomial above have modulo not larger than one. The simplest way to perform this test, without computing the roots, is to consider the change of variable

$$s \to \frac{1+\xi}{1-\xi},$$

and then apply Routh test to the resulting polynomial. Note in fact that is $|s| < 1$ then the real part of ξ is negative. The polynomial to study is thus

$$p(\xi) = (1+\xi)^2 + (1+\xi)(1-\xi)\eta + (1-\xi)^2(\eta+2) = 3\xi^2 + 2(-1-\eta)\xi + (3+2\eta).$$

By Routh test, this polynomial has all roots with negative real part if and only if $\eta \in (-3/2, -1)$. Hence, the matrix A has all roots with modulo smaller than one if and only if $\alpha^2 \in (0, 3/4)$. By the principle of stability in the first approximation, for all $\alpha^2 \in (0, 3/4)$ the zero equilibrium of the nonlinear system is locally asymptotically stable. Note that, if $\alpha^2 = 0$ or $\alpha^2 = 3/4$, the linearized system is stable, not asymptotically, but we cannot conclude anything on the stability properties of the zero equilibrium of the nonlinear system.

Exercise 8 *Consider the nonlinear model of a bioreactor described by the equations*

$$\begin{aligned}\dot{X} &= \mu(S)X - Xu \\ \dot{S} &= -\mu(S)X + (S_{in} - S)u \\ y &= S + kS^3,\end{aligned}$$

in which $S \geq 0$*,* $X \geq 0$*,* $S_{in} > 1$ *is a constant,* k *is a constant,* u *is an external signal and*

$$\mu(S) = 2\frac{S}{1 + 2S + S^2}.$$

1. *Suppose* u *is constant and determine all equilibrium points of the system in the following cases*

   (a) $u \in (0, 1/2)$*;*

   (b) $u = 1/2$*;*

   (c) $u > 1/2$.

   Sketch the position of the equilibrium points on the (X, S)*-plane as a function of* u*.*

2. *Write the linearized model of the system around the equilibrium point with both nonzero components determined in part 1.(b). Study the stability of such linearized system.*

Solution 8

1. The equilibrium points are the solutions of the equations $\dot{X} = 0$ and $\dot{S} = 0$. From the first equation we obtain $X = 0$ or $u = \mu(S)$. The first alternative yields, exploiting the second equation, $S = S_{in}$. The second alternative yields, exploiting again the second equation, $X = S_{in} - S$, where S is the solution of $u = \mu(S)$.
   Note now that the function $\mu(S)$ is always non-negative, it is zero for $S = 0$, it tends to zero as $S \to \infty$, and it has a maximum for $S = 1$ equal to $1/2$. As a result we obtain the following equilibria.

   (a) $u \in (0, 1/2)$. There are three equilibria: $(0, S_{in})$, $(S_{in} - S_1, S_1)$ and $(S_{in} - S_2, S_2)$, where S_1 and S_2 are the two solutions of $u = \mu(S)$ with $u \in (0, 1/2)$.

   (b) $u = 1/2$. There are two equilibria: $(0, S_{in})$ and $(S_{in} - 1, 1)$.

   (c) $u > 1/2$. There is one equilibrium: $(0, S_{in})$.

2. The system linearized around the equilibrium point $(S_{in} - 1, 1)$ is described by (note that $\frac{d\mu}{dS}(1) = 0$)

$$\dot{\delta}_x = A\delta_x + Bu = \begin{bmatrix} 0 & 0 \\ -1/2 & -1/2 \end{bmatrix} \delta_x + \begin{bmatrix} 1 - S_{in} \\ S_{in} - 1 \end{bmatrix} \delta_u$$

$$\delta_y = \begin{bmatrix} 0 & 1 + 3k \end{bmatrix} \delta_x.$$

   The linearized system is stable, not asymptotically, but nothing can be concluded on the stability of the equilibrium of the nonlinear system.

Exercise 9 *The (simplified and normalized) model of a patient in the presence of an infectious disease is described by the equations*

$$\dot{x} = 1 - x - Txy \qquad \dot{y} = Txy - y - ISy,$$

in which x represents the number of non-infected cells, y represents the number of infected cells, T represents the effect of the therapy and $IS \in (0,1)$ represents the action of the immune system.

1. *Determine the two equilibrium points of the system. Show that one equilibrium corresponds to a healthy patient, i.e. the number of infected cells is zero, and one equilibrium corresponds to an ill patient, i.e. the number of infected cells is non-zero.*
2. *Write the linearized models of the system around the two equilibrium points.*
3. *Without therapy $T > 1+IS$. Show that the equilibrium point associated to a healthy patient is unstable and the one associated to an ill patient is locally asymptotically stable.*
4. *With therapy $T < 1+IS$. Show that the equilibrium point associated to a healthy patient is locally asymptotically stable and the one associated to an ill patient is unstable.*

Solution 9

1. The equilibrium points of the system are the solutions of the equations $\dot{x} = 0$ and $\dot{y} = 0$. From the second equation we have $y = 0$ or $x = \frac{1+IS}{T}$. From the equation $\dot{x} = 0$, the former yields $x = 1$, and the latter yields $y = \frac{T-1-IS}{T(1+IS)}$. Hence, the equilibrium $(1,0)$ corresponds to a healty patient, and the equilibrium $(\frac{1+IS}{T}, \frac{T-1-IS}{T(1+IS)})$ corresponds to an ill patient.
2. The linearized model of the system around the first equilibrium point is

$$\begin{bmatrix} \dot{\delta}_x \\ \dot{\delta}_y \end{bmatrix} = \begin{bmatrix} -1 & -T \\ 0 & T-1-IS \end{bmatrix} \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix}.$$

The linearized model of the system around the second equilibrium point is

$$\begin{bmatrix} \dot{\delta}_x \\ \dot{\delta}_y \end{bmatrix} = \begin{bmatrix} -\frac{T}{1+IS} & -1-IS \\ \frac{T-1-IS}{1+IS} & 0 \end{bmatrix} \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix}.$$

3. Suppose $T > 1+IS$. The system linearized around the equilibrium associated to a healthy patient has an eigenvalue with positive real part, hence the equilibrium is unstable. On the contrary, the system linearized around the equilibrium associated to a ill patient has both eigenvalues with negative real part, hence the equilibrium is (locally) asymptotically stable.

4. Suppose $T < 1 + IS$. The system linearized around the equilibrium associated to a healthy patient has both eigenvalues with negative real part, hence the equilibrium is (locally) asymptotically stable. On the contrary, the system linearized around the equilibrium associated to a ill patient has one eigenvalue with positive real part, hence the equilibrium is unstable.

Exercise 10 *Consider the model of a simple robot arm given by*

$$\ddot{\phi} + m \sin \phi = T,$$

in which ϕ is the angle of the arm with respect to a vertical line directed upward, T is the applied torque, and m is a positive parameter.

1. *Let $T = 0$. Determine the equilibrium points of the system.*
2. *Let $T = 0$. Consider the energy of the system $E = \frac{1}{2}\dot{\phi}^2 - m \cos \phi$. Compute $\dot{E}$ and show that E remains constant for all t.*
3. *Let $T = -k\dot{\phi}$ with $k > 0$. Show that the equilibrium $(\phi, \dot{\phi}) = (0, 0)$ is locally asymptotically stable.*

Solution 10

1. The equilibrium points of the system are such that $\phi(t)$ is constant for all t, hence are such that $\dot{\phi} = \ddot{\phi} = 0$. As a result, the equilibrium points are the solutions of the equation $\sin \phi = 0$. This equation has infinitely many solutions, however, from a physical point of view, we have only two solutions $\phi = 0$ (the arm is directed upward) and $\phi = \pi$ (the arm is directed downward).
2. Note that
$$\dot{E} = \dot{\phi}\ddot{\phi} + m \sin \phi \dot{\phi}$$
and replacing the equation describing the motion of the arm yields
$$\dot{E} = T\dot{\phi}.$$
Hence, if $T = 0$ we have that $\dot{E} = 0$, hence $E(t)$ is constant for all t.
3. Setting $T = -k\dot{\phi}$ yields
$$\ddot{\phi} + k\dot{\phi} + m \sin \phi = 0.$$
To study the stability of the equilibrium $(0, 0)$ we linearize this equation around this point and we obtain
$$\ddot{\delta}_\phi + k\dot{\delta}_\phi + m\delta_\phi = 0.$$
The characteristic polynomial associated with this linear system is (Laplace transform the equation, and factor $\mathcal{L}(\phi(s))$)
$$s^2 + ks + m,$$
which has all roots with negative real part, by Routh test. As a result, the equilibrium $(0, 0)$ of the nonlinear system is locally asymptotically stable.

Exercise 11 *Consider the nonlinear system modelling the dynamics of the angular velocities of a rigid body in space (for example a satellite), described by the equations (known as Euler's equations)*

$$\begin{aligned}\dot{x}_1 &= \frac{I_2 - I_3}{I_1} x_2 x_3 \\ \dot{x}_2 &= \frac{I_3 - I_1}{I_2} x_3 x_1 \\ \dot{x}_3 &= \frac{I_1 - I_2}{I_3} x_1 x_2,\end{aligned}$$

with $I_1 \neq 0$, $I_2 \neq 0$ and $I_3 \neq 0$.

1. *Determine the equilibrium points of the system.*
2. *Consider the equilibrium $\tilde{x} = (\alpha, 0, 0)$ with $\alpha \neq 0$. Compute the linearized model around the equilibrium $\tilde{x}$, and study its stability as a function of I_1, I_2 and I_3. Discuss the stability properties of the equilibrium $\tilde{x}$ of the nonlinear system.*

Solution 11

1. The equilibrium points of the system are the solutions of the equations $\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0$. From the equation $\dot{x}_1 = 0$ we have $x_2 = 0$ or $x_3 = 0$. Replacing $x_2 = 0$ in the second and third equations yields the constraint $x_3 x_1 = 0$. As a result, we have the two families of equilibrium points described by

   $$(0, 0, \star) \qquad\qquad (\star, 0, 0),$$

   where $\star$ is any real number. Finally, from $x_3 = 0$ we have another family of equilibrium points described by

   $$(0, \star, 0).$$

   All the above equilibria describe steady rotations around an axis of symmetry (these are sometimes called, in mechanics, relative equilibria).

2. The system linearized around the given equilibrium is described by

   $$\dot{\delta}_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{I_3 - I_1}{I_2}\alpha \\ 0 & \frac{I_1 - I_2}{I_3}\alpha & 0 \end{bmatrix} \delta_x.$$

   As a result, if

   $$\frac{I_3 - I_1}{I_2} \frac{I_1 - I_2}{I_3} < 0$$

   the linearized system is stable, but no conclusions can be drawn on the stability properties of the given equilibrium of the nonlinear system. If

   $$\frac{I_3 - I_1}{I_2} \frac{I_1 - I_2}{I_3} > 0$$

   the linearized system is unstable, hence the given equilibrium of the nonlinear systems is unstable.

Exercise 12 *A continuous-time system described by the equations*

$$\dot{x} = Ax + Bu \qquad y = Cx$$

is passive if there exists a matrix $P = P' > 0$ *such that*

$$A'P + PA \leq 0 \qquad PB = C'.$$

Consider the continuous-time system

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -2 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \qquad y = \begin{bmatrix} 0 & 1 \end{bmatrix} x.$$

1. *Show that the system is passive.*
2. *Let* $u = -Ky$. *Write the equations of the closed-loop system and show that the system is asymptotically stable for all* $K > 0$.
3. *The zeros of a system with* $x \in I\!R^n$, $u \in I\!R$ *and* $y \in I\!R$, *are the complex numbers* $\bar{s}$ *such that*
$$\text{rank} \begin{bmatrix} \bar{s}I - A & B \\ C & 0 \end{bmatrix} < n + 1.$$
Show that the considered system has a zero at $\bar{s} = 0$.

Finally, show that a passive system does not have zeros $\bar{s}$ *with positive real part.*

Solution 12

1. Let
$$P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$
and note that
$$PB = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = C'$$
and
$$A'P + PA = 0,$$
hence the system is passive.
2. The closed-loop system is described by the equations
$$\dot{x} = (A - KBC)x = \begin{bmatrix} 0 & 1 \\ -2 & -K \end{bmatrix} x = A_{cl}x.$$
The characteristic polynomial of the matrix A_{cl} is given by
$$\det(sI - A_{cl}) = s^2 + Ks + 2,$$
hence, by Routh test, its roots have negative real part if and only if $K > 0$, hence the closed-loop system is asymptotically stable for all $K > 0$.

3. Note that

$$\begin{bmatrix} \bar{s}I - A & B \\ C & 0 \end{bmatrix} = \begin{bmatrix} \bar{s} & -1 & 0 \\ -2 & \bar{s} & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Setting $\bar{s} = 0$ yields the matrix

$$\begin{bmatrix} 0 & -1 & 0 \\ -2 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The determinant of this matrix is equal to zero, hence the matrix has rank smaller than three, which proves that the system has a zero for $\bar{s} = 0$.
(This definition of zeros for single-input, single-output systems is consistent with the definition based on the transfer function of the system. For example, for the considered system, the transfer function is

$$W(s) = C(sI - A)^{-1}B = \frac{s}{s^2 + 2},$$

hence the system has a zero at zero.)

To prove that a passive, single-input, single-output system does not have zeros with positive real part we proceed as follows. Suppose that the system has a zero at some $\bar{s}$, hence

$$\text{rank} \begin{bmatrix} \bar{s}I - A & B \\ C & 0 \end{bmatrix} < n + 1.$$

This implies that there exist a nonzero (complex) vector v and a scalar w such that

$$(\bar{s}I - A)v + Bw = 0 \qquad\qquad Cv = 0 \;\; (\text{or} \;\; v^\star C' = 0),$$

where $v^\star$ is the adjoint of v. Then, left multiplying by P,

$$0 = (\bar{s}P - PA)v + PBw = (\bar{s}P - PA)v + C'w,$$

hence, left multiplying by $v^\star$,

$$0 = v^\star \left((\bar{s}P - PA)v + C'w \right) = v^\star(\bar{s}P - PA)v.$$

As a result

$$\bar{s}v^\star Pv = v^\star PAv$$

and

$$\bar{s}^\star v^\star Pv = v^\star A'Pv.$$

Adding these two last equations yields

$$(\bar{s} + \bar{s}^\star)v^\star Pv = v^\star(A'P + PA)v.$$

Finally, note that $v^\star Pv > 0$, $v^\star(A'P + PA)v \leq 0$ and $(\bar{s} + \bar{s}^\star)$ is equal to twice the real part of $\bar{s}$, which has to be non-positive, as claimed.

Exercise 13 *Consider the discrete-time system $x(k+1) = Ax(k) + Bu(k)$. Let*

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

1. *Compute the reachability matrix R.*
2. *Determine if the system is reachable and compute the set of reachable states.*
3. *Determine all states x_I such that $x(0) = x_I$ and $x(1) = 0$.*

Solution 13

1. The reachability matrix is

$$R = \begin{bmatrix} B & AB & A^2B \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & -1 & 1 \\ 2 & -2 & -2 \end{bmatrix}.$$

2. The first two columns of the reachability matrix are linearly independent, and $\det(R) = 0$, hence the system is not reachable. The set of reachable states is two dimensional and it is described by the linear combination of the first two columns of the reachable matrix.

3. We have to determine all states which are controllable in one step. Instead of using the definition of controllable states in one step, we perform a direct calculation. Let

$$x(0) = x_I = \begin{bmatrix} x_{I,1} \\ x_{I,2} \\ x_{I,3} \end{bmatrix}$$

and note that

$$x(1) = Ax(0) + Bu(0) = \begin{bmatrix} x_{I,2} + u(0) \\ -x_{I,1} + u(0) \\ 2(x_{I,2} + u(0)) \end{bmatrix}.$$

The condition $x(1) = 0$ implies $x_{I,1} = u(0)$, $x_{I,2} = -u(0)$, hence all states that can be controlled to zero in one step are given by

$$x_I = \begin{bmatrix} u(0) \\ -u(0) \\ x_{I,3} \end{bmatrix},$$

and this is a two dimensional set. Note that this implies that the considered system has an eigenvalue at zero.

Exercise 14 *Consider the discrete-time system $x(k+1) = Ax(k) + Bu(k)$. Let*

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

1. *Compute the reachable subspaces in one step, two steps and three steps.*

2. *Using Hautus test determine the unreachable modes.*

3. *Show that the system is controllable in two steps.*

Solution 14 The reachability matrix is

$$R = \begin{bmatrix} B & AB & A^2B \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and it has rank equal to two.

1. The set of reachable states in one step is

$$\mathcal{R}_1 = \text{span} B = \text{span} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

   The set of reachable states in two steps is

$$\mathcal{R}_2 = \text{span}[B, AB] = \text{span} \begin{bmatrix} 1 & -1 \\ -1 & -1 \\ 0 & 0 \end{bmatrix} = \text{span} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

   The set of reachable states in three steps is

$$\mathcal{R}_3 = \text{span}[B, AB, A^2B] = \mathcal{R}_2.$$

2. The reachability pencil is

$$\left[\, sI - A \mid B \,\right] = \left[\begin{array}{ccc|c} s & -1 & 0 & 1 \\ 1 & s & 0 & -1 \\ 0 & 0 & s & 0 \end{array}\right].$$

   This matrix has rank three for all $s \neq 0$, hence the unreachable mode is $s = 0$.

3. The system is controllable, since the unreachable modes are at $s = 0$. To show that it is controllable in two steps note that

$$A^2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

hence

$$\text{Im}A^2 \subseteq \mathcal{R}_2,$$

which proves the claim.

Exercise 15 *Consider the continuous-time system $\dot{x} = Ax + Bu$. Let*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Compute the set of states that can be reached from the state

$$x_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Solution 15 Note that

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau = \begin{bmatrix} 0 \\ e^t \end{bmatrix} + \begin{bmatrix} \int_0^t e^{t-\tau}u(\tau)d\tau \\ 0 \end{bmatrix}.$$

Note that, by a proper selection of $u(\tau)$ in the interval $[0, t)$ it is possible to *assign* $\int_0^t e^{t-\tau}u(\tau)d\tau$. Therefore, the states that can be reached at time t from x_0 are described by

$$x(t) = \begin{bmatrix} 0 \\ e^t \end{bmatrix} + \lambda B,$$

with $\lambda \in I\!R$.

Exercise 16 *Consider the discrete-time system $x(k+1) = Ax(k)$, $y(k) = Cx(k)$. Let*

$$A = \begin{bmatrix} 0 & -4 & 0 \\ 1 & 4 & 0 \\ 0 & -4 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}.$$

Determine if the system is observable and compute the unobservable subspace.

Solution 16 The observability matrix is

$$O = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 0 \\ 4 & 12 & 0 \end{bmatrix}.$$

This matrix has rank two, hence the system is not observable. The unobservable subspace $\ker O$ is spanned by the vector

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

this means that it is not possible to obtain information on the third component of the state from measurements of the output.

Exercise 17 *Consider the continuous-time system*

$$\dot{x} = Ax + Bu \qquad\qquad y = Cx$$

and its Euler discrete-time approximate model

$$x_{k+1} = x_k + T(Ax_k + Bu_k) \qquad\qquad y_k = Cx_k$$

where $T > 0$ is the sampling time.

Show that the continuous-time system is observable if and only if the Euler discrete-time approximate model is observable.

Solution 17 By observability of the continuous-time system we have that

$$\text{rank} \left[\frac{sI - A}{C} \right] = n,$$

for all $s \in \mathbb{C}$. Consider now the observability pencil of the Euler model, namely

$$\left[\frac{sI - (I + TA)}{C} \right].$$

Note now that, for all $s \in \mathbb{C}$,

$$n = \text{rank} \left[\frac{sI - A}{C} \right] = \text{rank} \left[\frac{\frac{s-1}{T} I - A}{C} \right] = \text{rank} \left[\frac{zI - A}{C} \right],$$

where $z = \frac{s-1}{T}$. Hence, observability of the continuous-time system implies, and it is implied by, observability of the Euler discrete-time approximate model.

Exercise 18 *Consider the system $\sigma x = Ax$, $y = Cx$, with*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad\qquad C = \begin{bmatrix} I & 0 \end{bmatrix},$$

and A_{ij} matrices of appropriate dimensions. Show that the system is observable if and only if the (simpler) system $\sigma\xi = A_{22}\xi$ with output $\eta = A_{12}\xi$ is observable.

Solution 18 *Observability of the system $\sigma x = Ax$, $y = Cx$ implies, and is implied by,*

$$\text{rank}\left[\begin{array}{cc} sI - A_{11} & -A_{12} \\ A_{21} & sI - A_{22} \\ \hline I & 0 \end{array}\right] = n$$

for all $s \in \mathbb{C}$. Suppose that $y(t) \in \mathbb{R}^p$, then

$$\text{rank}\left[\begin{array}{cc} sI - A_{11} & -A_{12} \\ A_{21} & sI - A_{22} \\ \hline I & 0 \end{array}\right] = p + \text{rank}\left[\begin{array}{c} -A_{12} \\ sI - A_{22} \end{array}\right].$$

Note now that the matrix

$$\left[\begin{array}{c} -A_{12} \\ sI - A_{22} \end{array}\right]$$

has full rank for all $s \in \mathbb{C}$ if and only if the system $\sigma\xi = A_{22}\xi$ with output $\eta = A_{12}\xi$ is observable, and this proves the claim.

Exercise 19 *Consider the continuous-time system*

$$\begin{aligned} \dot{x} &= \left[\begin{array}{cc} 3 & -1+\epsilon \\ 1 & 2-\epsilon \end{array}\right] x + \left[\begin{array}{c} 0 \\ 1 \end{array}\right] u \\ y &= \left[\begin{array}{cc} -1 & 1 \end{array}\right] x. \end{aligned}$$

1. *Show that the system is reachable for all $\epsilon \neq 1$.*
2. *Let $\epsilon = 1$. Write the system in the canonical form for non-reachable systems, i.e. determine coordinates $\hat{x}$, defined by*

   $$x = L\hat{x}$$

   for some matrix L, such that, in the $\hat{x}$ coordinates, the system is described by equations of the form

   $$\dot{\hat{x}} = \left[\begin{array}{cc} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{array}\right] \hat{x} + \left[\begin{array}{c} \tilde{B}_1 \\ 0 \end{array}\right] u,$$

   with $\tilde{A}_{11}$ and $\tilde{B}_1$ such that $\dot{\hat{x}}_1 = \tilde{A}_{11}\hat{x}_1 + \tilde{B}_1 u$ is reachable. (Compute explicitly L, $\tilde{A}_{11}$, $\tilde{A}_{12}$, $\tilde{A}_{22}$ and $\tilde{B}_1$.)
3. *Show that the system is observable for all $\epsilon \neq 1/2$.*
4. *Let $\epsilon = 1/2$. Determine, using Hautus test, the unobservable modes.*

Solution 19

1. The reachability matrix is
$$R = \begin{bmatrix} 0 & \epsilon - 1 \\ 1 & 2 - \epsilon \end{bmatrix}.$$
If $\epsilon \neq 1$ then $\det(R) \neq 0$, hence the system is reachable.

2. If $\epsilon = 1$ the system is not reachable. Let L be constructed setting the first column equal to the first column of the reachable matrix (which is nonzero) and selecting the second column to render L non-singular, for example
$$L = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$
(Any matrix L of the form
$$L = \begin{bmatrix} 0 & L_{12} \\ 1 & L_{21} \end{bmatrix},$$
with $L_{12} \neq 0$, can be used.)
Note now that (recall that $\epsilon = 1$ and note that $L^{-1} = L$)
$$\begin{aligned} \dot{\hat{x}} &= L^{-1} \begin{bmatrix} 3 & 0 \\ 1 & 1 \end{bmatrix} L\hat{x} + L^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} u = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \hat{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} -1 & 1 \end{bmatrix} L\hat{x} = \begin{bmatrix} 1 & -1 \end{bmatrix} \hat{x}. \end{aligned}$$
Hence $\tilde{A}_{11} = 1$, $\tilde{A}_{12} = 1$, $\tilde{A}_{22} = 3$, $\tilde{B}_1 = 1$ and the system $\dot{\hat{x}}_1 = \tilde{A}_{11}\hat{x}_1 + \tilde{B}_1 u = \hat{x}_1 + u$ is reachable. (Note that the system is not stabilizable, and the unreachable mode is $s = 3$.)

3. The observability matrix is
$$O = \begin{bmatrix} -1 & 1 \\ -2 & 3 - 2\epsilon \end{bmatrix}.$$
Note that $\det(O) = 2\epsilon - 1$. Therefore the system is observable if $\epsilon \neq 1/2$.

4. The observability pencil, for $\epsilon = 1/2$, is
$$\left[\begin{array}{cc} s-3 & 1/2 \\ -1 & s - 3/2 \\ \hline 1 & -1 \end{array}\right].$$
As the system is not observable we know that the observability pencil loses rank, i.e. it has rank equal to one, for some s. To compute such an s consider the submatrix
$$\left[\begin{array}{cc} -1 & s - 3/2 \\ \hline 1 & -1 \end{array}\right].$$
This has rank equal to one for $s = 5/2$, which is therefore the unobservable mode.

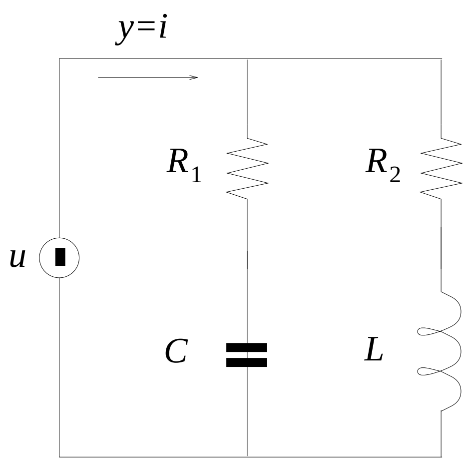


Figure A: The electrical network for Exercise 20.

Exercise 20 *Consider the linear electrical network in Figure A. Let u be the driving voltage.*

1. *Using Kirchhoff's laws, or otherwise, express the dynamics of the circuit in the standard state-space form*

$$\dot{x} = Ax + Bu \qquad y = Cx + Du$$

   Take x_1 to be the voltage across the capacitor, x_2 to be the current through the inductor and the output to be the current supplied by the generator.

2. *Derive a condition on the parameters R_1, R_2, C and L under which the pair (A, B) is controllable.*

3. *Derive a condition on the parameters R_1, R_2, C and L under which the pair (A, C) is observable.*

4. *Assume $R_1R_2C = L$ and $R_1 \neq R_2$. Derive the Kalman canonical form for the system.*

5. *Assume $R_1R_2C = L$ and $R_1 \neq R_2$. Define the controllable subspace and the unobservable subspace. Illustrate these subspaces as lines in R^2.*

Solution 20 Let x_1 denote the voltage across C and x_2 the current through L.

1. Kirchhoff's laws yield

$$u = x_1 + R_1C_1\dot{x}_1 \qquad u = R_2x_2 + l\dot{x}_2$$

and

$$y = i = x_2 + \frac{u - x_1}{R_1}.$$

As a result,

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = Ax + Bu = \begin{bmatrix} -\frac{1}{R_1 C} & 0 \\ 0 & -\frac{R_2}{L} \end{bmatrix} x + \begin{bmatrix} \frac{1}{R_1 C} \\ \frac{1}{L} \end{bmatrix} u$$

and

$$y = Cx + Du = \begin{bmatrix} -\frac{1}{R_1} & 1 \end{bmatrix} x + \frac{1}{R_1} u.$$

2. The reachability matrix is

$$R = \begin{bmatrix} \frac{1}{R_1 C} & -\frac{1}{R_1^2 C^2} \\ \frac{1}{L} & -\frac{R_2}{L^2} \end{bmatrix}$$

and

$$\det(R) = \frac{1}{R_1 C L} \left(\frac{1}{R_1 C} - \frac{R_2}{L} \right).$$

Hence the system is reachable and controllable if, and only if,

$$R_1 R_2 C \neq L.$$

3. The observability matrix is

$$O = \begin{bmatrix} -\frac{1}{R_1} & 1 \\ \frac{1}{R_1^2 C} & -\frac{R_2}{L} \end{bmatrix}$$

and

$$\det(O) = \frac{1}{R_1} \left(\frac{R_2}{L} - \frac{1}{R_1 C} \right).$$

Hence the system is observable if, and only if,

$$R_1 R_2 C \neq L.$$

4. If $R_1 R_2 C = L$ then the reachable subspace is

$$\mathcal{R} = \text{span} \begin{bmatrix} R_2 \\ 1 \end{bmatrix}$$

and the unobservable subspace is

$$\ker \mathcal{O} = \text{span} \begin{bmatrix} R_1 \\ 1 \end{bmatrix}.$$

Note that, as $R_1 \neq R_2$ these two subspaces are independent. Let

$$L = \begin{bmatrix} R_2 & R_1 \\ 1 & 1 \end{bmatrix}$$

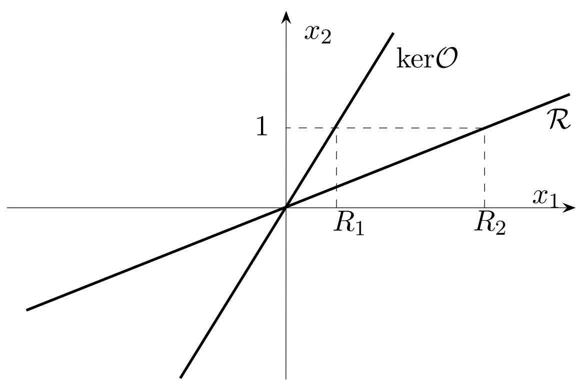
and note that the matrices of the system in Kalman canonical form are

$$\tilde{A} = L^{-1}AL = \begin{bmatrix} -\frac{R_2}{L} & 0 \\ 0 & -\frac{R_2}{L} \end{bmatrix} \qquad \tilde{B} = L^{-1}B = \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix}$$

and

$$\tilde{C} = CL = \begin{bmatrix} 1 - \frac{R_2}{R_1} & 0 \end{bmatrix}.$$

5. The subspaces are indicated in the figure.



Exercise 21 *Consider the electrical network depicted in Figure B.*

1. *Using Kirchoff's laws, or otherwise, write a state space description of the system.*
2. *Let $R_1 = R_2 = R$ and $C_1 = C_2 = C$. Compute the controllability and observability matrices and their ranks.*
3. *Let $R_1 = R_2 = R$ and $C_1 = C_2 = C$. Compute the Kalman canonical form of the system.*

Solution 21 Let x_j be the voltage across the capacitor C_j, considered positive "from left to right", and i_j the current through the capacitor C_j, considered positive "from left to right". The input current i is positive in the upward direction, and the output voltage is the voltage, positive "upward" between the two open terminals.

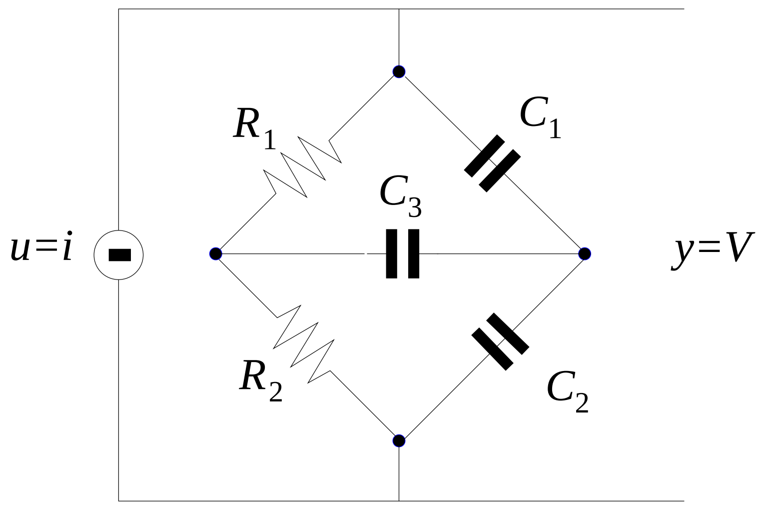


Figure B: The electrical network for Exercise 21.

1. Using the above conventions we have

$$i_1 = C_1\dot{v}_1 \qquad i_2 = C_2\dot{v}_2 \qquad i_3 = C_3\dot{v}_3$$

$$i_1 + i_2 + i_3 = 0 \qquad v_3 - v_2 = R_2(i - i_1 - i_3) \qquad v_1 - v_3 = R_1(i - i_1).$$

As a result

$$\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} C_1\dot{v}_1 \\ C_2\dot{v}_2 \\ C_3\dot{v}_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{1}{R_2} & \frac{1}{R_2} \\ \frac{1}{R_1} & 0 & -\frac{1}{R_1} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

yielding

$$\begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \dot{v}_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{C_1R_1} & 0 & \frac{1}{C_1R_1} \\ 0 & -\frac{1}{C_2R_2} & \frac{1}{C_2R_2} \\ \frac{1}{C_3R_2} & \frac{1}{C_3R_2} & -\frac{1}{C_3R_1} - \frac{1}{C_3R_2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} -\frac{1}{C_1} \\ \frac{1}{C_2} \\ 0 \end{bmatrix} i.$$

Finally

$$y = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

2. Setting $R_1 = R_2 = R$, $C_1 = C_2 = C$, $RC = 1/\alpha$, $RC_3 = 1/\beta$, and $u = i/C$ yields

$$\begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \dot{v}_3 \end{bmatrix} = \begin{bmatrix} -\alpha & 0 & \alpha \\ 0 & -\alpha & \alpha \\ \beta & \beta & -2\beta \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} u.$$

The reachability matrix is

$$R = \begin{bmatrix} -1 & \alpha & -\alpha^2 \\ 1 & -\alpha & \alpha^2 \\ 0 & 0 & 0 \end{bmatrix}.$$

R has rank one, hence the system is not reachable and not controllable. The observability matrix is

$$O = \begin{bmatrix} 1 & -1 & 0 \\ -\alpha & \alpha & 0 \\ \alpha^2 & -\alpha^2 & 0 \end{bmatrix}.$$

O has rank one, hence the system is not observable.

3. The reachability subspace is

$$\mathcal{R} = \text{span} B = \text{span} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

The unobservable subspace is

$$\ker \mathcal{O} = \text{span} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that

$$\mathcal{X}_1 = \mathcal{R} \cap \ker \mathcal{O} = \emptyset \qquad \mathcal{X}_2 = \mathcal{R} \qquad \mathcal{X}_3 = \ker \mathcal{O} \qquad \mathcal{X}_4 = \emptyset,$$

hence $x = L\hat{x}$, with

$$L = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Finally, Kalman canonical form is given by the equations

$$\dot{\hat{x}} = \begin{bmatrix} -\alpha & 0 & 0 \\ 0 & -\alpha & \alpha \\ 0 & 2\beta & -2\beta \end{bmatrix} \hat{x} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u$$

$$y = \begin{bmatrix} -2 & 0 & 0 \end{bmatrix} \hat{x}.$$

Exercise 22 *The linearized model of an orbiting satellite about a circular orbit of radius* $r_0 > 0$ *and angular velocity* $\omega_0 \neq 0$ *is described by the equations*

$$\dot{x} = Ax + Bu = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 3\omega_0^2 & 0 & 0 & 2r_0\omega_0^2 \\ 0 & 0 & 0 & 1 \\ 0 & -2\omega_0/r_0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1/r_0 \end{bmatrix} u$$

$$y = Cx = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x.$$

The output components are variations in radius and angle of the orbit and the input components are radial and tangential forces.

1. *Show that the system is controllable.*
2. *Design a state feedback control law*

$$u = Kx + Gv = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \end{bmatrix} x + \begin{bmatrix} g_{11} & 0 \\ 0 & g_{22} \end{bmatrix} v$$

such that

- *the matrix $A+BK$ has all eigenvalues equal to -1 and it is block diagonal, i.e.*

$$A + BK = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}$$

with $F_i \in I\!R^{2\times 2}$;

- *the closed-loop system has unity DC gain, i.e.*

$$-C(A+BK)^{-1}BG = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Solution 22

1. Consider the following submatrix of the reachability matrix

$$\tilde{R} = [B\ AB] = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2\omega_0^2 \\ 0 & 0 & 0 & 1/r_0 \\ 0 & 1/r_0 & -2\omega_0/r_0 & 0 \end{bmatrix}$$

and note that its determinant is $-1/r_0^2 \neq 0$. Hence the system is reachable and controllable.

2. Consider the closed-loop system

$$\dot{x} = (A+BK)x + BGv$$

and note that

$$A + BK = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 3\omega_0^2 + k_{11} & k_{12} & k_{13} & 2r_0\omega_0^2 + k_{14} \\ 0 & 0 & 0 & 1 \\ k_{21}/r_0 & -2\omega_0/r_0 + k_{22}/r_0 & k_{23}/r_0 & k_{24}/r_0 \end{bmatrix}.$$

Hence, selecting

$$k_{11} = -3\omega_0^2 - 1 \quad k_{12} = -2 \quad k_{13} = 0 \quad k_{14} = -2r_0\omega_0^2$$

$$k_{21} = 0 \quad k_{22} = 2\omega_0 \quad k_{23} = -r_0 \quad k_{24} = -2r_0$$

yields

$$A + BK = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & -2 \end{bmatrix}$$

which shows that the first condition has been achieved.

To achieve the second condition note that

$$-C(A + BK)^{-1}BG = \begin{bmatrix} g_{11} & 0 \\ 0 & g_{22}/r_0 \end{bmatrix}.$$

Hence, it suffices to select

$$g_{11} = 1 \qquad g_{22} = r_0.$$

Exercise 23 *Consider the continuous-time system $\dot{x} = Ax + Bu$. Let*

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Find a matrix K such that $\sigma(A + BK) = \{-1, -2\}$. Solve the problem in two ways:

1. *using the general theory discussed in the lectures;*
2. *using a direct computation, i.e. without computing the reachability matrix of the system.*

Solution 23

1. The general theory states that the state feedback is given by

$$K = -\begin{bmatrix} 0 & 1 \end{bmatrix} R^{-1}p(A),$$

where R is the reachability matrix and $p(s)$ is the desired closed-loop characteristic polynomial, in this case $p(s) = (s + 1)(s + 2) = s^2 + 3s + 2$. As a result,

$$K = -\begin{bmatrix} 0 & 1 \end{bmatrix} \left(\frac{1}{4} \begin{bmatrix} 7 & -3 \\ -1 & 1 \end{bmatrix} \right) \begin{bmatrix} 12 & 16 \\ 24 & 36 \end{bmatrix} = -\begin{bmatrix} 3 & 5 \end{bmatrix},$$

yielding

$$A + BK = \begin{bmatrix} -2 & -3 \\ 0 & -1 \end{bmatrix},$$

which has eigenvalues equal to -1 and -2, as requested.

2. Let

$$K = \begin{bmatrix} K_1 & K_2 \end{bmatrix}$$

and note that

$$A + BK = \begin{bmatrix} 1 + K_1 & 2 + K_2 \\ 3 + K_1 & 4 + K_2 \end{bmatrix}.$$

The characteristic polynomial of $A + BK$ is

$$\det(sI - (A + BK)) = s^2 + s(-5 - K_1 - K_2) + (-2K_2 + 2K_1 - 2),$$

and this should be equal to $p(s) = s^2 + 3s + 2$. As a result, K_1 and K_2 should be such that

$$-5 - K_1 - K_2 = 3 \qquad -2K_2 + 2K_1 - 2 = 2,$$

which yields $K_1 = -3$ and $K_2 = -5$. Note that, because the system has only one input and it is reachable, the state feedback assigning the eigenvalues is unique!

Exercise 24 *Consider the continuous-time system*

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 1 & -2 \\ 3 & -1 \end{bmatrix} x + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u \\ y &= \begin{bmatrix} 3 & -1 \end{bmatrix} x. \end{aligned}$$

1. *Show that the system is controllable and observable.*

2. *Assume zero initial state. Compute the response of the system when u is a unity step applied at $t = 0$.*

3. *Design a state feedback control law*

$$u = Kx + Gr$$

such that the closed-loop system has two eigenvalues at -3.

Solution 24

1. The reachability matrix is

$$R = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix},$$

which is full rank, hence the system is reachable and controllable.
The observability matrix is

$$O = \begin{bmatrix} 3 & -1 \\ 0 & -5 \end{bmatrix},$$

which is full rank, hence the system is observable.

2. We have to compute the forced response of the output of the system. Note that $\sigma(A) = \{i\sqrt{5}, -i\sqrt{5}\}$, hence the forced response of the output of the system has the form
$$y(t) = y_0 + y_1 \sin(\sqrt{5}t + \phi).$$
Note now that (recall that $u = 1$, for $t \geq 0$, and that $x(0) = 0$)
$$y(0) = \begin{bmatrix} 3 & -1 \end{bmatrix} x(0) = 0$$
$$\dot{y}(0) = \begin{bmatrix} 0 & -5 \end{bmatrix} x(0) + 4u = 4$$
$$\ddot{y}(0) = \begin{bmatrix} -15 & 5 \end{bmatrix} x(0) + 5u + 4\dot{u} = 5.$$
Therefore, we have to determine y_0, y_1 and ϕ from the equations
$$y_0 + y_1 \sin\phi = 0 \qquad \sqrt{5}y_1 \cos\phi = 4 \qquad -5y_1 \sin\phi = 5.$$
This yields
$$y_1 = \sqrt{\frac{21}{5}} \qquad \phi = \arctan\left(-\frac{\sqrt{5}}{4}\right) \qquad y_0 = \sqrt{\frac{21}{5}}.$$

3. Let
$$K = \begin{bmatrix} K_1 & K_2 \end{bmatrix}$$
and note that
$$A + BK = \begin{bmatrix} 1 + K_1 & -2 + K_2 \\ 3 + K_1 & -1 - K_2 \end{bmatrix}.$$
The characteristic polynomial of $A + BK$ is
$$\det(sI - (A + BK)) = s^2 + s(-K_1 + K_2) + (-3K_1 + 5 - 4K_2),$$
and this should be equal to $p(s) = (s + 3)^2 = s^2 + 6s + 9$. As a result, K_1 and K_2 should be such that
$$-K_1 + K_2 = 6 \qquad -3K_1 + 5 - 4K_2 = 9,$$
which yields $K_1 = -4$ and $K_2 = 2$.

Exercise 25 *Consider the continuous-time system*
$$\begin{aligned} \dot{x} &= \begin{bmatrix} 3 & -1 + \epsilon \\ 1 & 2 - \epsilon \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \\ y &= \begin{bmatrix} -1 & 1 \end{bmatrix} x. \end{aligned}$$

1. *Show that the system is controllable for any $\epsilon \neq 1$. Study the stabilisability of the system for $\epsilon = 1$.*

2. *Show that the system is observable for any $\epsilon \neq 1/2$. Study the detectability of the system for $\epsilon = 1/2$.*
3. *Assume $\epsilon = 0$. Design a state feedback control law*

$$u = Kx + Gr$$

such that the closed-loop system has two eigenvalues equal to -2.

4. *Show that the state feedback control law designed above stabilizes the system for any $\epsilon \in (-4, 1/7)$.*

Solution 25

1. The reachability matrix is

$$R = \begin{bmatrix} 0 & \epsilon - 1 \\ 1 & 2 - \epsilon \end{bmatrix},$$

and $\det(R) = 1 - \epsilon$. As a result the system is reachable and controllable for $\epsilon \neq 1$. Let $\epsilon = 1$ and consider the reachability pencil

$$\left[\begin{array}{c|c} sI - A & B \end{array}\right] = \left[\begin{array}{cc|c} s-3 & 0 & 0 \\ -1 & s-1 & 1 \end{array}\right],$$

which has rank equal to one for $s = 3$. The system is therefore not stabilizable. Note that it is possible to obtain this conclusion without computing the reachability pencil. In fact, for $\epsilon = 1$ the eigenvalues of A are $\{3, 1\}$, hence if there is an unreachable mode this is associated to a value of s with positive real part.

2. The observability matrix is

$$O = \begin{bmatrix} -1 & 1 \\ -2 & -2\epsilon + 3 \end{bmatrix},$$

and $\det(O) = 2\epsilon - 1$. As a result the system is observable for $\epsilon \neq 1/2$. Let $\epsilon = 1/2$ and consider the observability pencil

$$\left[\begin{array}{c} sI - A \\ \hline C \end{array}\right] = \left[\begin{array}{cc} s-3 & 1/2 \\ -1 & s-3/2 \\ \hline -1 & 1 \end{array}\right].$$

Because the system is not observable, this matrix has to have rank equal to one for some s. To find such an s consider the submatrix

$$\left[\begin{array}{cc} -1 & s-3/2 \\ \hline -1 & 1 \end{array}\right].$$

Its determinant is $s - 5/2$, hence the unobservable mode is $s = 5/2$ and the system is not detectable. Note that it is possible to obtain this conclusion without computing the observability pencil. In fact, for $\epsilon = 1/2$ the eigenvalues of A are $\{5/2, 2\}$, hence if there is an unobservable mode this is associated to a value of s with positive real part.

3. If $\epsilon = 0$ we have
$$A = \begin{bmatrix} 3 & -1 \\ 1 & 2 \end{bmatrix}.$$
Setting $K = [K_1,\ K_2]$ yields
$$\det(sI - (A + BK)) = s^2 + s(-5 - K_2) + (K_1 + 3K_2 + 7),$$
which should be equal to $(s + 2)^2$. This is achieved setting
$$K_1 = 24 \qquad K_2 = -9.$$

4. Consider now the matrix
$$A + BK = \begin{bmatrix} 3 & \epsilon - 1 \\ 25 & -7 - \epsilon \end{bmatrix}.$$
Its characteristic polynomial is
$$\det(sI - (A + BK)) = s^2 + s(4 + \epsilon) + (4 - 28\epsilon),$$
which has both roots with negative real part (by Routh test) if and only if $\epsilon \in (-4, 1/7)$.

Exercise 26 *Consider the continuous-time system $\dot{x} = Ax + Bu$, $y = Cx$, with*
$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -5 & -6 \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}.$$
Design an asymptotic observer having three eigenvalues at -10.

Solution 26 *An asymptotic observer is described by*
$$\dot{\xi} = A\xi + L(C\xi - y) + Bu = (A + LC)\xi - Ly + Bu$$
for some $L = [L_1\ L_2\ L_3]'$, where ξ is the asymptotic estimate of x provided the matrix $A + LC$ has all eigenvalues with negative real part. Note that
$$A + LC = \begin{bmatrix} L_1 & 1 & 0 \\ L_2 & 0 & 1 \\ L_3 & -5 & -6 \end{bmatrix},$$
and its characteristic polynomial is
$$s^3 + s^2(6 - L_1) + s(5 - 6L_1 - L_2) + (-L_3 - 5L_1 - 6L_2).$$
This should be equal to
$$(s + 10)^3 = s^3 + 30s^2 + 300s + 1000.$$
As a result,
$$L_1 = -24 \qquad L_2 = -151 \qquad L_3 = 26.$$

Exercise 27 *Consider the continuous-time system $\dot{x} = Ax$, $y = Cx$. Let*

$$A = \begin{bmatrix} 0 & 1 \\ 0 & -\alpha \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

1. *Show, using Hautus test, that the system is observable for all α.*

2. *Design an asymptotic observer for the system. Select the output injection gain L such that the matrix $A - LC$ has two eigenvalues equal to -3.*

3. *Suppose that one can measure $y(t)$ and a delayed copy of $y(t)$ given by $y(t-\tau)$, with $\tau > 0$. Assume (for simplicity) that $\alpha \neq 0$.*
   For $t \geq \tau$, express the vector

   $$Y(t) = \begin{bmatrix} y(t) \\ y(t-\tau) \end{bmatrix}$$

   from $x(0)$.
   Show that the relation determined above can be used, for any $\tau > 0$, to compute $x(0)$ as a function of $Y(t)$, where $t \geq \tau$. Argue that the above result can be used to determine $x(t)$ from $Y(t)$, for $t \geq \tau$, exactly.

Solution 27

1. The observability pencil is

   $$\left[\begin{array}{cc} s & -1 \\ 0 & s+\alpha \\ \hline 1 & 0 \end{array}\right],$$

   which has rank two for any s and any α. Hence the system is observable.

2. An asymptotic observer is described by

   $$\dot{\xi} = A\xi + L(C\xi - y) = (A + LC)\xi - Ly$$

   for some $L = [L_1 \ L_2]'$, where ξ is the asymptotic estimate of x provided the matrix $A + LC$ has all eigenvalues with negative real part. Note that

   $$A + LC = \begin{bmatrix} L_1 & 1 \\ L_2 & -\alpha \end{bmatrix}$$

   and its characteristic polynomial is

   $$s^2 + s(\alpha - L_1) - \alpha L_1 - L_2.$$

   This should be equal to $(s+3)^2 = s^2 + 6s + 9$, yielding

   $$L_1 = \alpha - 6 \qquad L_2 = -9 + (6 - \alpha)\alpha.$$

3. Note that
$$y(t) = Ce^{At}x(0)$$
and replacing t with $t-\tau$ one has
$$y(t-\tau) = Ce^{A(t-\tau)}x(0).$$
Then, for $t \geq \tau$,
$$Y(t) = \begin{bmatrix} y(t) \\ y(t-\tau) \end{bmatrix} = \begin{bmatrix} C \\ Ce^{-A\tau} \end{bmatrix} e^{At}x(0).$$
For the given A and C (using $\alpha \neq 0$) we have
$$\begin{bmatrix} C \\ Ce^{-A\tau} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -\frac{e^{\alpha\tau}-1}{\alpha} \end{bmatrix}$$
which is invertible for all $\alpha \neq 0$ and all $\tau > 0$. Hence
$$x(t) = e^{At}x(0) = \begin{bmatrix} 1 & 0 \\ 1 & -\frac{e^{\alpha\tau}-1}{\alpha} \end{bmatrix}^{-1} Y(t).$$
The above relation implies that, for all $t \geq \tau$, it is possible to obtain exactly $x(t)$.

Exercise 28 *Consider the discrete-time system*
$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 1 & -2 \\ 3 & -1 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u_k, \\ y_k &= \begin{bmatrix} 3 & -1 \end{bmatrix} x_k. \end{aligned}$$

1. *Show that the system is observable.*
2. *Design an asymptotic observer, with state $\hat{x}_k$, such that $e_k = x_k - \hat{x}_k = 0$ for all $k \geq N$. Determine the smallest value of N for which the above condition can be satisfied.*
3. *Let*
$$u_k = K\hat{x}_k + v_k$$
with $K = [3/4, 3/4]$. Write the equations of the closed-loop system, with state $[x_k, \hat{x}_k]$, input v_k and output y_k, and determine the eigenvalues of this system.

Solution 28

1. The observability matrix is
$$O = \begin{bmatrix} 3 & -1 \\ 0 & -5 \end{bmatrix},$$
which has rank equal to two. The system is therefore observable.

2. An asymptotic observer is described by

$$\xi_{k+1} = A\xi_k + L(C\xi_k - y_k) + Bu_k = (A + LC)\xi_k - Ly_k + Bu_k$$

for some $L = [L_1\ L_2]'$, where ξ_k is the asymptotic estimate of x provided the matrix $A + LC$ has all eigenvalues with negative real part. To obtain a dead-beat observer L should be such that both eigenvalues of $A + LC$ are zero. Note that

$$A + LC = \begin{bmatrix} 1 + 3L_1 & -2 - L_1 \\ 3 + 3L_2 & -1 - L_2 \end{bmatrix},$$

and

$$\det(sI - (A + LC)) = s^2 + s(L_2 - 3L_1) + 5(1 + L_2).$$

Hence,

$$L_1 = -1/3 \qquad\qquad L_2 = -1.$$

With this selection of L we have $(A + LC)^2 = 0$, hence $N = 2$. To prove that the smallest N for which the considered condition holds is $N = 2$ it is enough to observe that there is no selection of L such that $(A + LC)^1 = 0$.

3. By the separation principle the eigenvalues of the closed-loop system are the eigenvalues of the observer and of the matrix

$$A + BK = \begin{bmatrix} 7/4 & -5/4 \\ 9/4 & -7/4 \end{bmatrix}.$$

Hence, the eigenvalues of the closed-loop system are $\{1/2, -1/2, 0, 0\}$.

Exercise 29 *Consider the simplified model of a ship described by the equation*

$$\begin{aligned} M\ddot{\theta} + d\dot{\theta} + c\alpha &= w \\ \dot{\alpha} + \alpha &= u \end{aligned}$$

where θ denotes the heading angle error (the angle between the ship's heading and the desired heading), α denotes the rudder angle, w denotes a disturbance due to wind, and u is the control input. M and c are positive parameters, and d is a non-negative parameter.

1. Write the equation of the system, with state $(\theta, \dot{\theta}, \alpha)$, input (w, u) and output θ in standard state space form.

2. Let $w = 0$. Show that the system is controllable.

3. Show that the system is observable.

4. Let $w = 0$, $M = 1$, $c = 1$ and $d = 0$. Design an output feedback controller applying the separation principle. In particular, select the state feedback gain K such that the matrix $A - BK$ has three eigenvalues equal to -1 and the output injection gain L such that the matrix $A - LC$ has three eigenvalues equal to -3.

Solution 29

1. The description of the system in standard state space form is (set $x = (\theta, \dot{\theta}, \alpha)'$)

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -d/M & -c/M \\ 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1/M & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix}$$

$$y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} x.$$

2. The reachability matrix is

$$R = \begin{bmatrix} 0 & 0 & -c/M \\ 0 & -c/M & c/M(d/M+1) \\ 1 & -1 & 1 \end{bmatrix}$$

and this has full rank for all positive c and M. The system is reachable and controllable.

3. The observability matrix is

$$O = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -d/M & -c/M \end{bmatrix}$$

and this has full rank for all positive c and M. The system is observable.

4. Let $K = [K_1 \ K_2 \ K_3]$ and note that

$$A + BK = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ K_1 & K_2 & -1 + K_3 \end{bmatrix},$$

and that the characteristic polynomial of this matrix is $s^3 + (1 - K_3)s^2 + (K_2)s + (K_1)$. Hence the selection

$$K_1 = 1 \qquad K_2 = 3 \qquad K_3 = -2$$

is such that all eigenvalues of $A + BK$ are equal to -1. Let $L = [L_1 \ L_2 \ L_3]'$ and note that

$$A + LC = \begin{bmatrix} L_1 & 1 & 0 \\ L_2 & 0 & -1 \\ L_3 & 0 & -1 \end{bmatrix},$$

and that the characteristic polynomial of this matrix is $s^3 + (1 - L_1)s^2 + (-L_1 - L_2)s + (-L_2 + L_3)$. Hence the selection

$$L_1 = -8 \qquad L_2 = -19 \qquad L_3 = 8$$

is such that all eigenvalues of $A + LC$ are equal to -3. Finally, the controller is $\dot{\xi} = (A + BK + LC)\xi - Ly$, $u = K\xi$.

Exercise 30 *Consider the continuous-time system*

$$\dot{x} = \begin{bmatrix} 1 & -2 \\ 3 & -1 \end{bmatrix} x + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u,$$

$$y = \begin{bmatrix} 3 & -1 \end{bmatrix} x$$

1. *Design an asymptotic observer with a double pole at* -6.

2. *Suppose* x_0 *is the observer state evaluated in part 1. Let*

$$u = Kx_0 + Gr$$

with $K = [-4, 2]$. *Compute the eigenvalues of the closed-loop system.*

Solution 30

1. An asymptotic observer is described by

$$\dot{\xi} = A\xi + L(C\xi - y) + Bu = (A + LC)\xi - Ly + Bu$$

for some $L = [L_1 \; L_2]'$, where ξ is the asymptotic estimate of x provided the matrix $A + LC$ has all eigenvalues with negative real part. Note that

$$A + LC = \begin{bmatrix} 1 + 3L_1 & -2 - L_1 \\ 3 + 3L_2 & -1 - L_2 \end{bmatrix},$$

and its characteristic polynomial is

$$s^2 + s(L_2 - 3L_1) + 5(1 + L_2).$$

This should be equal to

$$(s + 6)^2 = s^2 + 12s + 36.$$

As a result,

$$L_1 = -29/15 \qquad L_2 = 31/5.$$

2. By the separation principle the eigenvalues of the closed-loop system are the eigenvalues of the observer and of the matrix

$$A + BK = \begin{bmatrix} -3 & 0 \\ 7 & -3 \end{bmatrix}.$$

Hence, the eigenvalues of the closed-loop system are $\{-3, -3, -6, -6\}$.