Module Notes

Advanced Control Systems

ECS778P/ECS654U

Dr. Aidan O. T. Hogg {a.hogg@qmul.ac.uk}

Queen Mary University of London (last updated: January 6, 2025)

Contents

These lecture notes were originally created by Prof. Alessandro Astolfi and have been adapted for this module by Dr. Aidan O. T. Hogg. I am deeply grateful to Prof. Astolfi for his exceptional work and generosity in allowing these notes to be adapted. For suggestions or comments, please email a.hogg@qmul.ac.uk.

Preface

These notes were originally created by Prof. Alessandro Astolfi as a reference for a third-year 'Control Engineering' module in the Electrical and Electronic Engineering (EEE) Department at Imperial College London and as supplementary material for the 'Automatic Control' module at the University of Rome Tor Vergata. Subsequently, these notes have now been revised by Dr Aidan O. T. Hogg for the 3rd/4th-year 'Advanced Control Systems' module in the Electrical Engineering and Computer Science (EECS) Department at Queen Mary University of London.

The primary aim of these notes is to provide a self-contained, rigorous foundation in systems theory, along with an introduction to state-space analysis and design methods for linear systems. The content draws heavily from the Teoria dei sistemi by A. Ruberti and A. Isidori (Boringheri, 1985) and Prof. Astolfi's research in nonlinear control theory. While various approaches to the subject exist, the so-called behavioural approach, developed by J. Willems, is probably the most compelling alternative.

It is hoped that these notes will serve as a valuable starting point for further exploration into more advanced topics in systems and control theory. Although there are numerous excellent textbooks—such as A Linear Systems Primer by Panos J. Antsaklis and Anthony N. Michel—that cover similar material in depth, these notes intend to offer a concise yet precise introduction rather than to replace a more comprehensive study.

AIDAN O. T. HOGG

1 Introduction

In the (abstract) definition of the 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 emphasis 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 its properties and the conditions under which it can be introduced in detail.

1.1 Examples

Let's start by considering a few motivating examples that are instrumental in introducing the notion of an (abstract) system.

1.1.1 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 preceding resistors, the net resistance is a rational number having the 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 it represents.

1.1.2 Model of an infectious disease

There are several mathematical models which describe the interactions between HIV and immunocytes in the 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 a 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.

1.1.3 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

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

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 graduate after three years. Finally, in the *extreme* situation in which, $\alpha_i(k) = 0$ for all k and for some i, we have

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

1.1.4 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 the 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.

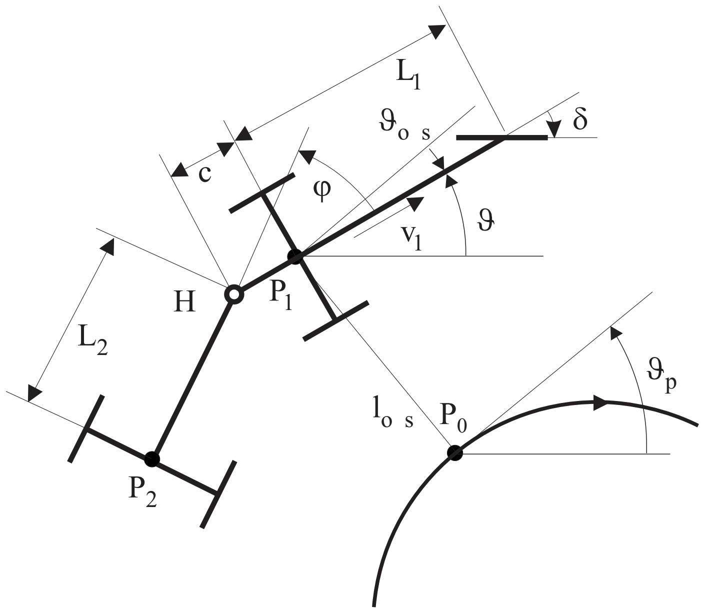

Figure 1.1: Vehicle's geometry and path-tracking offsets l_{os} and ϑ_{os}.

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

$$\begin{aligned}
\dot{l}_{os} &= -\sigma\,|v_1|\sin\vartheta_{os}\ , \\
\dot{\vartheta}_{os} &= v_1\frac{u}{L_1} - \sigma\,|v_1|\,\frac{\cos\vartheta_{os}}{R_1 + l_{os}}\ , \\
\dot{\varphi}_{os} &= -\frac{v_1}{L_2}\sin\left(\varphi_{os}+\varphi_p\right) - \frac{v_1}{L_1L_2}\left(c\cos\left(\varphi_{os}+\varphi_p\right)+L_2\right)u\ ,
\end{aligned}$$

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

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

1.1.5 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 void of outer space cools the top. 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* atmosphere behaviour and to justify why weather forecasts can be erroneous and have been recently shown to play an important role in models of lasers and electrical generators. Note that the Lorenz equations are still at the basis of modern weather forecast algorithms.

1.1.6 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

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

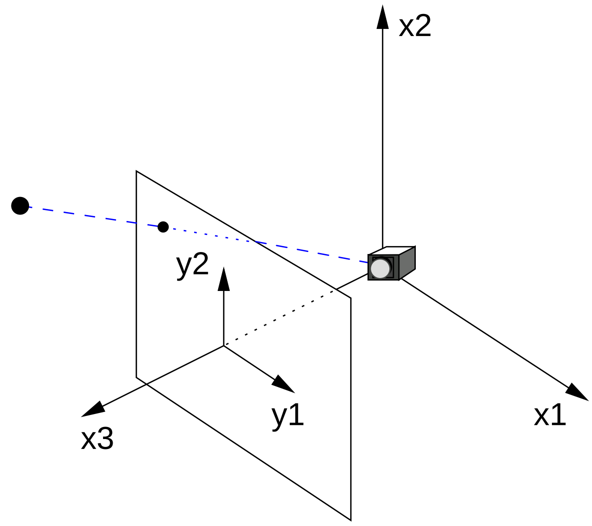

Figure 1.2: Diagram of the perspective vision system.

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.

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

$$y = \begin{bmatrix} y_1, & y_2 \end{bmatrix}' = \epsilon \begin{bmatrix} \frac{x_1}{x_3}, & \frac{x_2}{x_3} \end{bmatrix}' , \tag{1.6}$$

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 of reconstructing the coordinates x_1, x_2, x_3 from measurements of the image-space coordinates y_1, y_2.

1.1.7 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

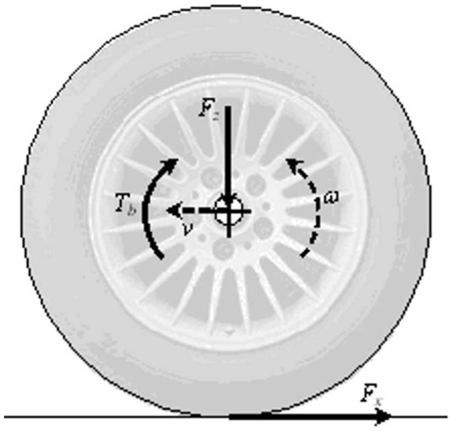

Figure 1.3: Quarter car vehicle model.

- 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 behaviour 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

1.1.8 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.

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

The string can be described by

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

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 is given by

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

where

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

◇

1.1.9 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. Except for 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.

1.1.10 Google page rank

The speed and success of Google can be largely attributed to the efficiency of the search algorithm, which, when 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 j, 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.

1.2 The notion of system

The discussion in Section 1.1 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 a system is thus introduced to provide tools to study such a wide variety of objects based on 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 with several different objects, and at the same time, several systems can be associated with the same object (depending upon the properties that have to be investigated).

We stress that the definition of an abstract notion of a system has the advantage that it allows for interpretation 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 these notes, we define a system by considering 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 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. This way, we collect one or more pairs of functions, denoted as input-output pairs, defined for all $t \geq t_0$. When we collect all such pairs together, we have a set of input-output pairs, which is used to define a 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 a 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 $I\!R$, which is the set of time instants of interest for the system, and define the subset of future time instants[2]

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

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-output 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 a system must, therefore, consider this issue.

Definition 1.1. *Consider an ordered subset T of $I\!R$ and two (non-empty) sets U and Y. An abstract*

[2] F stands for "future".

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 *is the set of values of the input signal, and* Y *is the set of values of the output signal.*

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 of all relations S_{t_0} for $t_0 \in T$.

1.2.1 Parametric representations

The use of relations to represent a system provides a very powerful point of view which applies 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]

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

such that

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

[3] $u|_{F(t_1)}$ denotes the restriction of u to $t \in F(t_1)$.
[4] $\mathcal{D}R$ denotes the domain of R, and $\mathcal{R}R$ denotes the range of R.

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 $y = f_{t_0}(x_0, u_0)$. $\diamond$

1.2.2 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 the above relation to 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 $t \le \bar{t}$.

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 $t \in [t_0, \bar{t})$.

1.2.3 The notion of state

The crucial point in defining an abstract system by means of a relation is 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 further information besides the input function. This information is associated with 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 are 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}, maybe 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 with 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 relationship 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

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

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

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

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

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 by 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

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

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

The appearance characterizes this representation, 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.

1.2.4 Definition of state

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

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*

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

such that the following conditions hold.

- *For all $t_0 \in T$, $u \in \mathcal{U}$ and $x_0 \in X$*

$$\{(u_0, y_0) \in U^{F(t_0)} \times Y^{F(t_0)} \ : \ u_0(t) = u(t), y_0(t) = g(t, \phi(t, t_0, x_0, u_{[t_0,t)}), u(t))\} = S_{t_0} \ .$$

- *(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 the state transition function, and the function g output transformation. The triple $\{X, \phi, g\}$ 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 example, 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 the 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 must address the issues of existence and the 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.

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 $\{X, \phi, g\}$, 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 $\{X, \phi, g\}$. 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

  $$g_z : T \times Z \times U \to \tilde{Y} \ ,$$

  such that

  $$g_z(t, z, u) = g(t, \psi^{-1}(z), u) \ .$$

[6] This property is sometimes called semigroup property.

[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 $t \in T$, $\{u(t) \in U \ : \ u \in \mathcal{U}\} = U$.

Then $\{Z, \phi_z, g_z\}$ is a state space representation of $\mathcal{S}$.

- (State augmentation) Consider a system $\mathcal{S}$, with state space representation $\{X, \phi, g\}$. 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 $g_a : T \times T \times X_a \times \mathcal{U} \to Y$ is such that $g_a - g = 0$. Then $\{X_a, \phi_a, g_a\}$ 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 by introducing a few new concepts that will be useful in studying state space representations of particular classes of systems.

Definition 1.5 (Equivalent representations). *Two-state space representations $\{X, \phi, g\}$ and $\{X', \phi', g'\}$ 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 $t \geq t_0$*

$$g(t, \phi(t, t_0, x_0, u), u(t)) = g'(t, \phi'(t, t_0, x_0', u), u(t)) .$$

- *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 $t \geq t_0$*

$$g'(t, \phi'(t, t_0, x_0', u), u(t)) = g(t, \phi(t, t_0, x_0, u), u(t)) .$$

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

$$g(t, \phi(t, t_0, x_a, u), u(t)) = g(t, \phi(t, t_0, x_b, u), u(t)) .$$

Remark. Equivalent states are sometimes referred to as non-distinguishable states because it is not possible to distinguish between them by measurements of the output. ◇

Definition 1.7 (Reduced state space). *Consider a system $\mathcal{S}$ and a state space representation $\{X, \phi, g\}$. 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 $\{X, \phi, g\}$ is said to be reduced at t_0.*

Remark. A state space can be reduced at some time t_0 but not at another 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. ◇

1.2.5 Classification

The notion of a 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 based on some of the key ingredients discussed.

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

Remark. There is a class of systems, increasingly studied and used in applications, in which for some state variables $T = I\!R$ 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 $\bar{t} \in T$*

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

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 g 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.*

1.2.6 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 the 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 $\{X, f, g\}$ 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 1.1).

Motivated by these considerations, we say that the state space representation $\{X, \phi, g\}$ 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 by noting that a regular representation $\{X, \phi, g\}$ can be alternatively described by the triple $\{X, f, g\}$.

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

1.3 Examples revisited

We conclude this chapter by revisiting the examples discussed in Section 1.1 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 $\{X, f, g\}$ 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 $g(x) = x_1 + x_2$.
- 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.

2 State-Space Descriptions of Systems

2.1 Introduction

State-space representations offer a comprehensive view of a system's internal dynamics, while input-output representations focus on the system's external behaviour and its interaction with external signals.

In this section, we explore the state-space and input-output descriptions of systems. The state-space approach provides an internal view of system dynamics, while the input-output approach, also known as the external description, captures the system's response to inputs. We will discuss continuous-time systems represented by ordinary differential equations and discrete-time systems represented by ordinary difference equations, focusing on linear systems.

Let us first consider once more systems described by equations of the form

$$\dot{x} = f(t, x, u) \ , \tag{2.16a}$$

$$y = g(t, x, u) \ , \tag{2.16b}$$

where $x \in R_n$, $y \in R_p$, $u \in R_m$, $f : R \times R_n \times R_m \to R_n$, and $g : R \times R_n \times R_m \to R_p$. Here t denotes continuous-time, u denotes system input, y denotes system output, and x denotes the state of the system. To simplify notation we denote $u(t)$, $x(t)$, $y(t)$ as u, x and y, respectively.

The equation (2.16a) is called the *state equation*, (2.16b) is called the *output equation*, and (2.16a) and (2.16b) constitute the *state-space description* of continuous-time finite-dimensional systems.

If the system is in discrete-time, $t \in Z$, we use the notation

$$x^+ = f(t, x, u) \ , \quad y = g(t, x, u) \ , \tag{2.17}$$

where we simplify notation by replacing $x(t+1)$ with x^+. For convenience and compactness, we also use the notation

$$\sigma x = f(t, x, u) \ , \quad y = g(t, x, u) \ , \tag{2.18}$$

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 State-space representation

An important special case of (2.16) is systems described by the most general state-space representation of a linear time-invariant system, which is written in the following form

$$\sigma x = Ax(t) + Bu(t) \ , \tag{2.19a}$$

$$y = Cx(t) + Du(t) \ , \tag{2.19b}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{q\times n}$, and $D \in \mathbb{R}^{q\times p}$ for p inputs, q outputs and n state variables. Such equations may arise in the modelling process of a physical system, or they may be a consequence of a linearization process. It should be noted that for brevity, t is often omitted.

2.3 Trajectory and motion

We now consider again a system described using an input-state-output representation $\{X, \phi, g\}$ where ϕ is the state transition function. We now define a few typical dynamic behaviours of the system.

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

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

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

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

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

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

The main difference between a trajectory and a motion is that they live in different spaces. A motion is parameterized by t, whereas a 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 its evolution. In contrast, 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 $X = \mathbb{R}^2$ and $t_0 = 0$.

2.4 Equilibrium

Definition 2.3 (Equilibrium)**.** *Consider a system $\{X, \phi, g\}$. 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 with 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.

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

$$0 = f(t, x_e, u_0) \ , \tag{2.20}$$

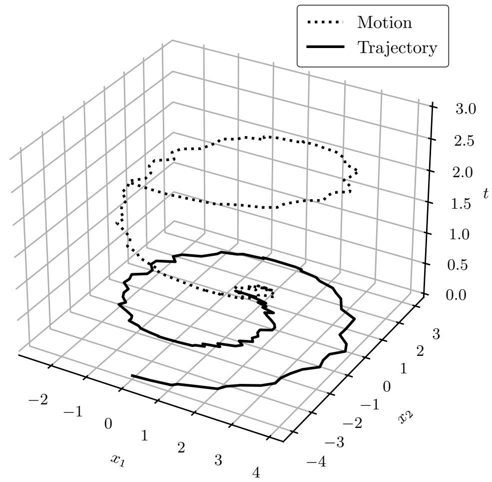

Figure 2.1: A motion (dashed line) and the corresponding trajectory (solid line).

for continuous-time systems and of the systems of equations

$$x_e = f(t, x_e, u_0) \;, \tag{2.21}$$

for discrete-time systems.

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

$$\sigma x = Ax + Bu \;, \tag{2.22}$$

with $x \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$. The set of equilibria is a linear subspace. 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.*

2.5 Trajectories of linear continuous-time systems

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

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

with $x \in X = \mathbb{R}_n$, $u(t) \in \mathbb{R}_m$, $y(t) \in \mathbb{R}_p$, initial condition $x(0) = x_0$ and without loss of generality t_0 can be set to $t_0 = 0$. Then

$$x(t) = \underbrace{e^{At}x_0}_{\textit{free response}} + \underbrace{\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau}_{\textit{forced response}} , \tag{2.23}$$

and

$$y(t) = \underbrace{Ce^{At}x_0}_{\textit{free response}} + \underbrace{\int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau + Du(t)}_{\textit{forced response}} , \tag{2.24}$$

where the free response is the component of the solution that arises solely from the system's initial conditions in the absence of any input, while the forced response is the portion of the solution that results directly from the applied input.

Proof. First we need to formally define the matrix exponential e^{Ft} for a given(square) matrix F as

$$e^{Ft} = I + Ft + \frac{(Ft)^2}{2!} + \frac{(Ft)^3}{3!} + \cdots , \tag{2.25}$$

where 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 $e^{Ft} = \text{diag}(e^{\lambda_1 t}, e^{\lambda_2 t}, \cdots, e^{\lambda_n t})$.

Now consider the simple case where $A = 0$ and the differential equation is $\dot{x}(t) = Bu(t)$, 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

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

Hence

$$z(t) = z_0 + \int_0^t Bu(\tau)d\tau \ ,$$

from which we obtain directly equation (2.23) by substituting $z(t) = z_0 + \int_0^t Bu(\tau)d\tau$ into $x(t) = e^{At}z(t)$. Then (2.24) can be trivially obtained by replacing $x(t)$ in the output transformation. ◁

2.6 Transfer function

The transfer function only captures the input/output properties of a system. That is to say, it only captures the system's forced response. To calculate the transfer function, we have to set the initial conditions to $x(0) = 0$ then

$$Y(s) = \mathscr{L}\Big\{ \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau + Du(t) \Big\} \ , \tag{2.26}$$

$$Y(s) = [C(sI - A)^{-1}B + D]U(s) \ , \tag{2.27}$$

where the transfer function, $G(s)$, is

$$G(s) = C(sI - A)^{-1}B + D \ . \tag{2.28}$$

2.7 Trajectories of linear discrete-time systems

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

$$x(k+1) = Ax(k) + Bu(k) \ , \quad y(k) = Cx(k) + Du(k) \ , \tag{2.29}$$

with $x(k) \in X = \mathbb{R}^n$, $u(k) \in \mathbb{R}^p$, $y(k) \in \mathbb{R}^q$ and the initial condition $x(0) = x_0$. Then

$$x(k+1) = \underbrace{A^k x_0}_{\textit{free response}} + \underbrace{\sum_{i=0}^{k-1} A^{k-1-i}Bu(i)}_{\textit{forced response}} \ , \tag{2.30}$$

and

$$y(k) = \underbrace{CA^k x_0}_{\textit{free response}} + \underbrace{\sum_{i=0}^{k-1} CA^{k-1-i}Bu(i) + Du(k)}_{\textit{forced response}} \ , \tag{2.31}$$

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

$$\begin{aligned} x(1) &= Ax_0 + Bu(0) \ , \\ x(2) &= Ax(1) + Bu(1) = A^2x_0 + AB(0) + Bu(1) \ , \\ x(3) &= Ax(2) + Bu(2) = A^3x_0 + A^2B(0) + AB(1) + Bu(2) \ , \\ &\vdots \end{aligned}$$

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

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

$$x(k) = A^k x_0 + \begin{bmatrix} B & AB & \cdots & A^{k-1}B \end{bmatrix} \begin{bmatrix} u(k-1) \\ u(k-2) \\ \vdots \\ u(0) \end{bmatrix} . \tag{2.32}$$

This expression highlights the role of the matrix $\begin{bmatrix} B & AB & \cdots & A^{k-1}B \end{bmatrix}$ 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 u(k) in the interval $[0, k)$ allows to compute x_0. In fact, from (2.30) we obtain

$$x_0 = A^{-k}x(k) - \sum_{i=0}^{k-1} A^{-i-1}Bu(i) , \tag{2.33}$$

◇

Remark. The 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 C, A, and B, known as Markov parameters, appearing in (2.31) 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

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

i.e. the output yields direct 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 interpret the results of Propositions 2.2 and 2.3. Equations (2.23) and (2.30) 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 (2.23) and (2.30) show that the principle of superposition holds for such systems.

Analogously, (2.24) and (2.31) 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 (2.23) and (2.30) ((2.24) and (2.31), respectively) yield directly the functions ϕ and g of a state-space representation of the system.

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

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

with $x \in X = \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$ and $y(t) \in \mathbb{R}^q$. 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 (2.23))

$$x(t) = e^{AT} x_k + \int_0^T e^{A(T-\tau)} B u(\tau) d\tau \; u_k = A_d x_k + B_d u_k \,, \tag{2.34}$$

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 the obtained discrete-time model is exact under the stated operating conditions $x_k = x(kT)$, for all $k \geq 0$. ◇

2.8 Coordinate transformations

Because of the importance of this concept in the forthcoming sections, 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, time-invariant, finite-dimensional, linear system described by the equations

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

with $x \in X = \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^q$, and the change of coordinates

$$x = L\hat{x} \,, \tag{2.35}$$

where L is invertible. The state-space representation in the new coordinates is given by

$$\begin{aligned} \dot{\hat{x}} &= L^{-1}\dot{x} \,, \\ &= L^{-1}(Ax + Bu) = L^{-1}(AL\hat{x} + Bu) \,, \\ &= \underbrace{L^{-1}AL}_{\hat{A}}\hat{x} + \underbrace{L^{-1}B}_{\hat{B}}u \,, \end{aligned}$$

and

$$y = \underbrace{CL}_{\hat{C}}\hat{x} + \underbrace{D}_{\hat{D}}u \; .$$

The matrices A and $\hat{A}$ 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 2.4 (Algebraic equivalent systems)**.** *The state-space representations*

$$\sigma x = Ax + Bu \; , \quad y = Cx + Du \; ,$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^q$, and

$$\sigma \hat{x} = \hat{A}\hat{x} + \hat{B}\hat{u} \; , \quad \hat{y} = \hat{C}\hat{x} + \hat{D}\hat{u} \; ,$$

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

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

Note that, for algebraically equivalent representations, one has

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

and

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

3 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 us to study the behaviour of trajectories close to an equilibrium point or to a certain motion.

Remark. The notion of stability that we discuss was introduced in 1882 by the Russian mathematician A.M. Lyapunov in his Doctoral Thesis (hence, it is often referred to as Lyapunov stability). There are other notions of stability, such as Lagrange stability. Nevertheless, the concept of Lyapunov stability is the most commonly used in applications. $\diamond$

3.1 Definition

Consider a system described via a state-space representation $\{X, \phi, g\}$ 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 $X \subset \mathbb{R}^n$, the Euclidean norm can be defined.)

Definition 3.1 (Lyapunov stability). *Consider a system $\{X, \phi, g\}$ 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*

$$\|x(t_0) - x_e\| < \delta \; , \tag{3.36}$$

implies

$$\|x(t) - x_e\| < \epsilon \; , \tag{3.37}$$

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 neighbourhood of x_e (even very small), the perturbed evolution stays within this neighbourhood for all initial perturbations belonging to a sufficiently small neighbourhood 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 us 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 3.2 (Asymptotic stability). *Consider a system $\{X, \phi, g\}$ 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

$$\lim_{t\to\infty} \|x(t) - x_e\| = 0 \ . \tag{3.38}$$

In summary, an equilibrium point is asymptotically stable if it is stable and whenever the initial perturbation is inside a certain neighbourhood 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 (3.38) implies that for any $\kappa > 0$ there is a time ta such that

$$\|x(t) - x_e\| \leq \kappa \ , \tag{3.39}$$

for all $t \geq t_a$. The value t_a is a function of κ (consistently with the definition of limit) but may also be a 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 functions 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. ⋄

Definition 3.3 (Global asymptotic stability). *Consider a system $\{X, \phi, g\}$ 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 for the structure of the underlying state-space realization. For example, in the case of finite-dimensional, time-invariant systems, it implies that $X = \mathbb{R}^n$. 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 $t \to \infty$.

Suppose an equilibrium x_e is not globally asymptotically stable. In that case, 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 the region of attraction of the equilibrium x_e. Note that if x_e is globally asymptotically stable, then its region of attraction will coincide with X.

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

Definition 3.4 (Exponential stability). *Consider a system $\{X, \phi, g\}$ 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

$$\|x(t) - x_e\| < \epsilon e^{-\lambda(t-t_0)} , \tag{3.40}$$

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 3.5 (Stability of motion). *Consider a system $\{X, \phi, g\}$ and a motion*

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

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

$$\|\phi(t, t_0, x(t_0), u) - \phi(t, t_0, x_0, u)\| < \epsilon , \tag{3.41}$$

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 unperturbed 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 (3.41) may be violated.

3.2 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 the 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, and 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 3.1. *Consider a system with a linear state-space representation. Then (asymptotic) stability of one motion implies (asymptotic) stability of all motions. In particular, the (asymptotic) stability of any motion implies and is implied by the (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

$$\|x(t_0) - x_0\| < \delta \ , \tag{3.42}$$

implies

$$\|\phi(t, t_0, x(t_0), u) - \phi(t, t_0, x_0, u)\| < \epsilon \ , \tag{3.43}$$

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

$$\phi(t, t_0, x(t_0), u) - \phi(t, t_0, x_0, u) = \phi(t, t_0, x(t_0) - x_0, 0) \ . \tag{3.44}$$

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 3.2. *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, the 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 to the stability properties of the linear representation. For example, we say that a linear representation is stable, meaning 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 the 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 3.3. *A linear, finite-dimensional representation is stable if and only if*

$$\|\Phi(t, t_0)\| \leq k \ , \tag{3.45}$$

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 (3.45) 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 $x_e = 0$. $\triangleleft$

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

Proposition 3.4. *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 by 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 to assume time-invariance.

Proposition 3.5. *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, the attractivity of the zero equilibrium implies the stability of the zero equilibrium. This implies that stability is a property of the matrix A (see equation (2.19)). Moreover, as the properties are uniform, attractivity implies uniform asymptotic stability and hence exponential stability.

Proposition 3.6. *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[8] equal to one have a non-positive real part, and the eigenvalues of A with geometric multiplicity larger than one have a 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 the boundedness of the state transition matrix.

For continuous-time systems the state transition matrix, with $t_0 = 0$, is (see equation (2.25))

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

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.

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

[8] To define the geometric multiplicity of an eigenvalue, we need to recall a few facts. Consider a matrix $A \in \mathbb{R}^{n\times n}$ 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.

- *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 that 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. $\triangleleft$

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

Proposition 3.8. *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*[9] $P = P'$ *such that*

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

- *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 with 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 that are 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

$$x(t) = L(t)\hat{x}(t) \ , \tag{3.46}$$

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

$$v'Pv > 0 \ ,$$

for all nonzero vectors $v \in \mathbb{R}^n$. 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 the positivity of a symmetric matrix

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

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 $L(t)$ 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 3.9. *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 (3.46) is (asymptotically) stable if and only if*

$$\|L(t)\| \le k_1 \ , \qquad\qquad \|L^{-1}(t)\| \le k_2 \ , \tag{3.47}$$

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

Corollary 3.1. *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 (3.46) 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 (3.47) holds. ◁

4 Structural Properties - Reachability and Controllability

4.1 Introduction

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

$$\sigma x = Ax + Bu \ , \quad y = Cx + Du \ , \tag{4.48}$$

with $x \in X = \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^q$ 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 the so-called structural properties: reachability and controllability, observability and reconstructability, respectively.

These properties allow us 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.

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 by 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 to determine 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 transfer are equivalent.

For linear systems, the properties of reachability and controllability are referred to as the state $x = 0$; hence, we say that a state is reachable means that it is reachable from $x = 0$ and that a state is controllable means 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 the properties of the matrices A and B.

4.2 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

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

This 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

$$R_k = \left[\begin{array}{cccc} B & AB & \cdots & A^{k-1}B \end{array} \right],$$

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

$$\mathcal{R}_1, \ \mathcal{R}_2, \ \cdots, \ \mathcal{R}_k, \ \cdots. \tag{4.49}$$

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

Proposition 4.1. *The sequence of subspaces (4.49) 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

$$\mathcal{R}_{\bar{k}} = \mathcal{R}_{\bar{k}+1} \Rightarrow \mathcal{R}_{\bar{k}+1} = \mathcal{R}_{\bar{k}+2}, \tag{4.50}$$

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

Definition 4.1. *Consider the discrete-time system (4.48). 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 $\mathcal{R} = X = \mathbb{R}^n$.

Remark. By definition,

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

hence the discrete-time system (4.48) is reachable if and only if

$$\mathrm{rank} R = n. \tag{4.51}$$

Equation (4.51) 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, $k^\star = n$. ◇

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 4.2. *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 by 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 (2.35), $\mathcal{R}_k$ 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.

4.3 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

$$-A^k x^\star \in \mathcal{R}_k. \tag{4.52}$$

It is easy to see that the set of all $x^\star$ such that equation (4.52) 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

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

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

$$\mathcal{C}_1,\ \mathcal{C}_2,\ \cdots,\ \mathcal{C}_k,\ \cdots. \tag{4.53}$$

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

Proposition 4.3. *The sequence of subspaces (4.53) 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 4.1. 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 4.2. *Consider the discrete-time system (4.48). The subspace $\mathcal{C} = \mathcal{C}_n$ is the controllability subspace of the system.*

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

The discrete-time system (4.48) is controllable if and only if

$$\mathrm{Im} A^n \subseteq \mathcal{R}. \tag{4.54}$$

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

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

4.4 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

$$x_f - A^k x_0 = R_k U_{k-1}, \tag{4.55}$$

where

$$U_{k-1} = \begin{bmatrix} u(k-1) \\ u(k-2) \\ \vdots \\ u(1) \\ u(0) \end{bmatrix} \in \mathbb{R}^{k\,m}.$$

To solve the problem, we have to solve the linear system (4.55) in the unknown U_{k-1}. This system has a solution if, and only if,

$$x_f - A^k x_0 \in \mathrm{Im} R_k, \tag{4.56}$$

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 (4.56) 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, $\text{rank} R_n = n$, hence the condition expressed in equation (4.56), 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 (4.55) 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).

4.5 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 (2.23)).

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[10]

$$\mathrm{Im} R = \mathrm{Im} W_t. \tag{4.57}$$

Selecting

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

where β is a constant vector, yields

$$\bar{x} = W_t \beta. \tag{4.58}$$

Hence, to assess the reachability of the state $\bar{x} \in \mathrm{Im}R$, it is sufficient to show that equation (4.58) has (at least) one solution β. However, this fact holds trivially by condition (4.57).

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 4.3. *Consider the continuous-time system (4.48). 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.

Proposition 4.4. *Consider the continuous-time system (4.48). The following statements are equivalent.*

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

- *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 given time $t > 0$. In fact, to determine one such input signal, it is sufficient to solve the equation (4.58), which, if the system is reachable, has a unique solution

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

i.e. the input signal

$$u(\tau) = B'e^{A'(t-\tau)}W_t^{-1}\bar{x} \tag{4.59}$$

steers the state of the system from $x(0) = 0$ to $x(t) = \bar{x}$. Similarly to what was discussed in Section 4.4, it is possible to prove that, among all input signals steering the state from 0 to $\bar{x}$ in time t, the input signal (4.59) 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).

4.6 A canonical form for reachable systems

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

Consider the system (4.48), 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

$$lB = 0 \qquad lAB = 0 \qquad lA^{n-2}B = 0 \qquad lA^{n-1}B = 1. \tag{4.60}$$

In fact, conditions (4.60) can be rewritten as

$$lR = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{4.61}$$

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 4.1. *Let l be as in equation (4.61). Then, the square matrix*

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

is invertible.

Proof. Consider the matrix

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

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

and note that, by conditions (4.60),

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

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

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 2.8. However, it is easier to proceed 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

$$\sigma\hat{x} = A_r\hat{x} + B_r\hat{x}, \tag{4.62}$$

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 (4.62) is reachable, hence a system described by equation (4.62) 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.

4.7 Description of non-reachable systems

In this section, we study systems which are not reachable, i.e. systems described by the equation (4.48) 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[11].

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

$$L\hat{A} = AL \qquad\qquad L\hat{B} = B$$

and partition the matrices L, A, $\hat{A}$, B and $\hat{B}$ as

$$L = \left[\begin{array}{c|c} L_{11} & L_{12} \\ \hline L_{21} & L_{22} \end{array}\right], \qquad A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right], \qquad \hat{A} = \left[\begin{array}{c|c} \hat{A}_{11} & \hat{A}_{12} \\ \hline \hat{A}_{21} & \hat{A}_{22} \end{array}\right],$$

$$B = \left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right], \qquad \hat{B} = \left[\begin{array}{c} \hat{B}_1 \\ \hline \hat{B}_2 \end{array}\right],$$

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

$$\sigma\hat{x} = \begin{bmatrix} \sigma\hat{x}_1 \\ \sigma\hat{x}_2 \end{bmatrix} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ 0 & \hat{A}_{22} \end{bmatrix} \hat{x} + \begin{bmatrix} \hat{B}_1 \\ 0 \end{bmatrix} u, \tag{4.63}$$

where $\hat{x}_1 \in \mathbb{R}^{\rho}$ and $\hat{x}_2 \in \mathbb{R}^{n-\rho}$. The reachability matrix of the system in the $\hat{x}$ coordinates is

$$\hat{R} = \begin{bmatrix} \hat{B}_1 & \hat{A}_{11}\hat{B}_1 & \cdots & \hat{A}_{11}^{n}\hat{B}_1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and, because $R = L\hat{R}$, it has rank ρ. This implies that the subsystem

$$\sigma\hat{x}_1 = \hat{A}_{11}\hat{x}_1 + \hat{B}_1 u \tag{4.64}$$

[11] 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.

is reachable. The subsystem (4.64) is called the reachable subsystem of the system (4.48), whereas the subsystem

$$\sigma\hat{x}_2 = \hat{A}_{22}\hat{x}_2, \tag{4.65}$$

which is clearly not affected by the input, is called the unreachable subsystem of the system (4.48). The eigenvalues of the matrix $\hat{A}_{22}$, which are a subset of the eigenvalues of the matrix A, are called unreachable modes.

4.8 PBH reachability test

The decomposition of a system in reachable and unreachable parts allows us to derive an alternative test for reachability.

Proposition 4.5 (Popov-Belevich-Hautus (PBH) reachability test)**.** *Consider the system (4.48). 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 the reachability pencil. Note that the rank condition in Proposition 4.5 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 (4.63) and note that, for the transformed system, one has[12]

$$\text{rank} \left[\begin{array}{cc|c} sI - \hat{A}_{11} & -\hat{A}_{12} & \hat{B}_1 \\ 0 & sI - \hat{A}_{22} & 0 \end{array} \right] = n.$$

However, this is not true, as this matrix loses rank for all s which are eigenvalues of $\hat{A}_{22}$, hence the contradiction. $\triangleleft$

Remark. The PBH test allows one 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. $\diamond$

[12]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

$$\left[\begin{array}{c|c} sI - \hat{A} & \hat{B} \end{array} \right] = \left[\begin{array}{c|c} sI - L^{-1}AL & L^{-1}B \end{array} \right] = L^{-1} \left[\begin{array}{c|c} sI - A & B \end{array} \right] \left[\begin{array}{cc} L & 0 \\ 0 & I \end{array} \right].$$

5 Structural Properties - Observability and Reconstructability

5.1 Introduction

In this Chapter 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.

5.2 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 $u(0)$, $u(1)$, $\dots$, $u(k-1)$ the corresponding output sequences y_a and y_b, coincide for the first k steps, i.e.

$$y_a(t) = y_b(t) \tag{5.66}$$

for all $t \in [0, k]$. By equation (2.31), condition (5.66) is equivalent to

$$CA^t x_a = CA^t x_b \tag{5.67}$$

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 (5.67) 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 5.1. *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 5.2. *Consider the discrete-time system (4.48). 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 (4.48) is observable if and only if

$$\mathrm{rank} O = n. \tag{5.68}$$

Equation (5.68) is known as Kalman observability rank condition, and was derived by R.E. Kalman in the 60's. ◇

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 $\ker O$. ◇

The unobservable subspace $\ker O$ has the following important property, the proof of which is a simple consequence of the definition of the subspace.

Proposition 5.1. *The unobservable subspace is contained in the subspace* $\ker C$*, i.e.*

$$\ker C \supseteq \ker O,$$

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.

5.3 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 $u(0)$, $u(1)$, $\ldots$, $u(k-1)$ and the output sequence $y(0)$, $y(1)$, $\ldots$, $y(k)$ 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

$$\tilde{x}(k) = A^k x(0) + \sum_{t=0}^{k-1} A^{k-t-1} Bu(t) + A^k \ker O_k = x(k) + A^k \ker O_k.$$

As a result, for given input and output sequences, the state of the system at time k is uniquely defined if

$$A^k \ker O_k = \{0\},$$

or, equivalently, if

$$\ker O_k \subseteq \ker A^k \tag{5.69}$$

whereas it is not uniquely defined if the above condition does not hold.

A system is said reconstructable in k steps if condition (5.69) 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

$$\ker O \subseteq \ker A^n. \tag{5.70}$$

Note that, the observability rank condition (5.68) implies, but it is not implied, by condition (5.70).

5.4 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 (2.31), 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

$$Y_{n-1} = Ox(0) \tag{5.71}$$

has the (unique) solution

$$x(0) = (O'O)^{-1}O'Y_{n-1}, \tag{5.72}$$

which provides a simple way to compute the state $x(0)$ of the system.

Remark. Equation (5.72) 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 (2.30). ⋄

Remark. Measurements are naturally affected by noise, i.e. in practice equation (5.71) 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 (5.72) yields a free output response which minimizes the norm (energy) of the error between the actual free output response and the calculated one. ⋄

5.5 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 (2.24), 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

$$Ce^{A\tau}(x_a - x_b) = 0, \tag{5.73}$$

for all $\tau \in [0, t]$.

The function $Ce^{A\tau}$ is analytic, hence condition (5.73) 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,

$$C(x_a - x_b) = 0 \qquad CA(x_a - x_b) = 0 \qquad \cdots \qquad CA^i(x_a - x_b) = 0 \qquad \cdots. \tag{5.74}$$

Note finally that, by Cayley-Hamilton Theorem, (5.74) is equivalent to (recall the definition of the observability matrix O)

$$O(x_a - x_b) = 0. \tag{5.75}$$

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 (5.75) 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 5.2. *Consider the continuous-time system (4.48). The following statements are equivalent.*

- *The system is observable.*
- $\text{rank}O = n$.
- *For all $t > 0$ the observability Gramian*

$$V_t = \int_0^t e^{A'\tau}C'Ce^{A\tau}d\tau$$

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

$$\int_0^t w'e^{A'\tau}C'Ce^{A\tau}wd\tau = 0$$

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. $\triangleleft$

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.

5.6 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 5.3. *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*

$$\sigma x = Ax + Bu \qquad y = Cx + Du. \tag{5.76}$$

The system, with state $\xi \in \Xi = I\!R^n$, input $v \in I\!R^p$, and output $g \in I\!R^m$, described by the equations

$$\sigma \xi = A'\xi + C'v \qquad g = B'\xi + D'v \tag{5.77}$$

is called the dual system of system (5.76), which is called the primal system of system (5.77).

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 (5.76) and let $R^\star$ and $O^\star$ be the reachability and observability matrix, respectively, of the dual system (5.77).

Then, trivially,

$$R^\star = O' \qquad O^\star = R'.$$

Moreover, for discrete-time systems, the following implications hold

$$\mathrm{Im}A^n \subseteq R \Leftrightarrow (\mathrm{Im}A^n)^\perp \supseteq R^\perp \Leftrightarrow \ker A^n \supseteq \ker R' = \ker O^\star.$$

As a result, the following statement holds.

Proposition 5.3. *Consider the system (5.76) and its dual (5.77).*

- *System (5.76) is reachable if and only if system (5.77) is observable.*
- *System (5.76) is observable if and only if system (5.77) is reachable.*
- *System (5.76) is controllable if and only if system (5.77) is reconstructable.*
- *System (5.76) is reconstructable if and only if system (5.77) is controllable.*

The notion of duality and the above statement allow to derive results similar to those presented in Sections 4.6, 4.7 and 4.8, but with respect to the observability property.

Proposition 5.4. *Consider the system (4.48) with $p = 1$. Suppose the system is observable. Then the system is algebraically equivalent to a system of the form*

$$\sigma \hat{x} = A_o\hat{x} + B_o u \qquad y = C_o\hat{x} + D_o u \tag{5.78}$$

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 (5.78) is observable, hence a system described by the equations (5.78) is said to be in observability canonical form.

Proposition 5.5. *Consider the system (4.48). 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

$$\sigma\hat{x}_1 = \hat{A}_{11}\hat{x}_1 + \hat{B}_1 u \qquad y = \hat{C}_1\hat{x}_1 + \hat{D}u \tag{5.79}$$

is observable.

The subsystem (5.79) is called the observable subsystem of the system (4.48), whereas the subsystem

$$\sigma\hat{x}_2 = \hat{A}_{22}\hat{x}_2 + \hat{B}_2 u \qquad y = 0, \tag{5.80}$$

which clearly does not contribute to the output, is called the unobservable subsystem of the system (4.48). The eigenvalues of the matrix $\hat{A}_{22}$, which are a subset of the eigenvalues of the matrix A, are called unobservable modes.

Proposition 5.6 (PBH observability test)**.** *Consider the system (4.48). The system is observable if and only if*

$$\text{rank} \begin{bmatrix} sI - A \\ \hline C \end{bmatrix} = n$$

for all $s \in \mathbb{C}$.

6 Design Tools

6.1 Introduction

In Chapters 4 and 5 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.

6.2 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

$$\sigma x = Ax + Bu \qquad y = Cx + Du, \tag{6.81}$$

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

$$u = Ky + v, \tag{6.82}$$

with K a constant matrix of appropriate dimensions and v a new external signal. The resulting closed-loop system is described by the equations

$$\begin{aligned} \sigma x &= (A + B(I - KD)^{-1}KC)x + (I - KD)^{-1}Bv \\ y &= (I + D(I - KD)^{-1}K)Cx + (I - KD)^{-1}Dv, \end{aligned} \tag{6.83}$$

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

$$u = Kx + v, \tag{6.84}$$

with K a constant matrix of appropriate dimensions and v a new external signal. The resulting closed-loop system is described by the equations

$$\begin{aligned} \sigma x &= (A + BK)x + Bv \\ y &= (C + DK)x + Dv. \end{aligned} \tag{6.85}$$

- *Dynamic output feedback.* The input signal is generated by the system[13]

$$\sigma\xi = F\xi + Gy \qquad u = K\xi + v \tag{6.86}$$

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

$$\begin{aligned} \sigma x &= Ax + BK\xi + Bv \\ \sigma\xi &= (F + GDK)\xi + GCx + GDv \\ y &= Cx + DK\xi + Dv. \end{aligned} \tag{6.87}$$

- *Dynamic state feedback.* The input signal is generated by the system[14]

$$\sigma\xi = F\xi + Gx \qquad u = K\xi + v \tag{6.88}$$

with F, G and K constant matrices of appropriate dimensions and v a new external signal. The

[13] We consider the simplest version of dynamic output feedback. A more general form is given by $\sigma\xi = F\xi + Gy + Hv$, $u = K\xi + Jy + Lv$.

[14] We consider the simplest version of dynamic state feedback. A more general form is given by $\sigma\xi = F\xi + Gx + Hv$, $u = K\xi + Jx + Lv$.

resulting closed-loop system is described by the equations

$$\begin{aligned} \sigma x &= Ax + BK\xi + Bv \\ \sigma\xi &= F\xi + Gx \\ y &= Cx + DK\xi + Dv. \end{aligned} \tag{6.89}$$

In what follows we study in detail the static state feedback (Section 6.3) and the dynamic output feedback (Section 6.6) 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 (6.83) as a function of K is very difficult[15]. Finally, dynamic state feedback is useful only in very specific problems, such as the so-called noninteracting control problem with stability[16], which are not the subject of these notes.

6.3 State feedback

Consider a system described by the equations (6.81), the state feedback control law (6.84) and the resulting closed-loop system (6.85).

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 6.1. *The system (6.81) is reachable if and only if the system (6.85) 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. ◁

Proposition 6.2. *Consider the system (6.81). Suppose the system is not reachable and it is described by equations of the form (4.63). Then the system (6.85) is described by equations of the same form. Moreover, the systems (6.81) and (6.85) have the same unreachable modes.*

Proof. By assumption, system (6.81) is described by equations of the form

$$\sigma x = \left[\begin{array}{cc} A_{11} & A_{12} \\ 0 & A_{22} \end{array}\right] x + \left[\begin{array}{c} B_1 \\ 0 \end{array}\right] u.$$

[15] In the case $m = p = 1$ and $D = 0$ the root locus method can be used to study the eigenvalues of the matrix $A + BKC$.

[16] 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 (6.88) such that the closed-loop system (6.89) is asymptotically stable and composed of m *decoupled* systems, e.g. the Markov parameters are diagonal matrices.

As a result, system (6.85) 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 6.3. *Consider system (6.81). 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 4.6). Let T be the transformation matrix defined in Section 4.6, 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

$$A_r + B_r\hat{K}_r \neq A_r + B_rK_r,$$

hence the characteristic polynomial of $A_r + B_r\hat{K}_r$ is not $p(s)$. ◁

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 6.4 (Ackermann's formula)**.** *Consider system (6.81). 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 6.3 and 6.4 provide a constructive way to assign the characteristic polynomial, hence the eigenvalues, of system (6.85). 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. ◇

The result summarized in Proposition 6.3 can be extended to multi-input systems.

Proposition 6.5. *Consider system (6.81) 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 6.1 (Heymann)**.** *Consider system (6.81) 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*

$$\sigma x = (A + BG)x + b_iv \tag{6.90}$$

is reachable.

Exploiting Lemma 6.1 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 (6.90) is reachable, and then we use Ackermann's formula to compute a matrix k such that the characteristic polynomial of

$$A + BG + b_ik$$

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 (6.85). Conversely, it is possible to conclude reachability of system (6.81) by a property of system (6.85).

Proposition 6.6. *System (6.81) is reachable if and only if it is possible to arbitrarily assign the eigenvalues of $A + BK$.*

6.3.1 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 6.2, the unreachable modes are not modified by the application of state feedback. This implies that there exists a matrix K such that system (6.85) is asymptotically stable if and only if the unreachable modes of system (6.81) 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 6.1 (Stabilizability)**.** *System (6.81) 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.*

6.4 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 (6.81) and assume[17] 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

$$x_e = My + Nu, \tag{6.91}$$

[17] This assumption is without loss of generality. In fact, if $y = Cx + Du$ and u are measurable then also $\tilde{y} = Cx$ is measurable.

with M and N constant matrices of appropriate dimensions. The resulting interconnected system is described by the equations

$$\begin{aligned} \sigma x &= Ax + Bu \\ x_e &= MCx + Nu. \end{aligned} \tag{6.92}$$

- *Dynamic filter.* The state estimate is generated by the system

$$\sigma\xi = F\xi + Ly + Hu \qquad\qquad x_e = M\xi + Ny + Pu \tag{6.93}$$

with F, L, H, M, N and P constant matrices of appropriate dimensions. The resulting interconnected system is described by the equations

$$\begin{aligned} \sigma x &= Ax + Bu \\ \sigma\xi &= F\xi + LCx + Hu \\ x_e &= M\xi + NCx + Pu. \end{aligned} \tag{6.94}$$

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.

6.5 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 (6.81) is given by

$$\sigma\xi = A\xi + Bu \qquad\qquad x_e = \xi. \tag{6.95}$$

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 (6.81) is stable, and will converge to zero only if the system (6.81) is asymptotically stable. If these conditions do not hold, the estimation error will not be bounded and system (6.95) does not qualify as a state observer for system (6.81).

The intrinsic limitation of the observer (6.95) 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 (6.95) 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

$$\sigma\xi = A\xi + Bu + Ly_e \qquad\qquad x_e = \xi. \tag{6.96}$$

To assess the properties of this candidate state observer note that $e = x - x_e$ is such that

$$\sigma e = (A + LC)e. \tag{6.97}$$

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 (6.81) is observable.

We summarize the above discussion with two formal statements.

Proposition 6.7. *Consider system (6.81) and suppose the system is observable. Let $p(s)$ be a monic polynomial of degree n. Then there is*[18] *a matrix L such that the characteristic polynomial of $A+LC$ is equal to $p(s)$.*

Proposition 6.8. *System (6.81) is observable if and only if it is possible to arbitrarily assign the eigenvalues of $A + LC$.*

6.5.1 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 6.2, the unobservable modes are not modified by the output injection gain. This implies that there exists a matrix L such that system (6.97) is asymptotically stable if and only if the unobservable modes of system (6.81) 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 6.2 (Detectability)**.** *System (6.81) 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.*

[18]For single-output systems the matrix L assigning the characteristic polynomial of $A + LC$ is unique.

6.5.2 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 (6.81). 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 (6.81) with $D = 0$ and assume that the matrix C has p independent columns[19]. 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

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = Tx = \begin{bmatrix} I & C_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

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

[19]This is the case if rank$C = p$, whereas if rank$C < p$ it is always possible to eliminate redundant lines.

$w - \hat{x}_2$ be the observation error. Then

$$\begin{aligned}\sigma w - \sigma \hat{x}_2 &= F\xi + Hv + Gu + L\left[\tilde{A}_{11}\hat{x}_1 + \tilde{A}_{12}\hat{x}_2 + \tilde{B}_1 u\right] - \left[\tilde{A}_{21}\hat{x}_1 + \tilde{A}_{22}\hat{x}_2 + \tilde{B}_2 u\right] \\ &= F\xi + \left(H + L\tilde{A}_{11} - \tilde{A}_{21}\right)\hat{x}_1 + \left[L\tilde{A}_{12} - \tilde{A}_{22}\right]\hat{x}_2 + \left[G + L\tilde{B}_1 - \tilde{B}_2\right]u.\end{aligned} \tag{6.98}$$

To have convergence of the estimation error to zero, regardless of the initial conditions and of the input signal, we must have

$$\sigma(w - \hat{x}_2) = F(w - \hat{x}_2) \tag{6.99}$$

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 (6.98) and (6.99), 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

$$F = \tilde{A}_{22} - L\tilde{A}_{12}$$

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

$$H = FL - L\tilde{A}_{11} + \tilde{A}_{21} \qquad G = -L\tilde{B}_1 + \tilde{B}_2.$$

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 6.1.

6.6 The separation principle

In Section 6.3 it has been shown that system (6.81) can be stabilized by means of a state feedback control law, provided the system is stabilizable. Moreover, in Section 6.5 it has been shown that the state of system (6.81) 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

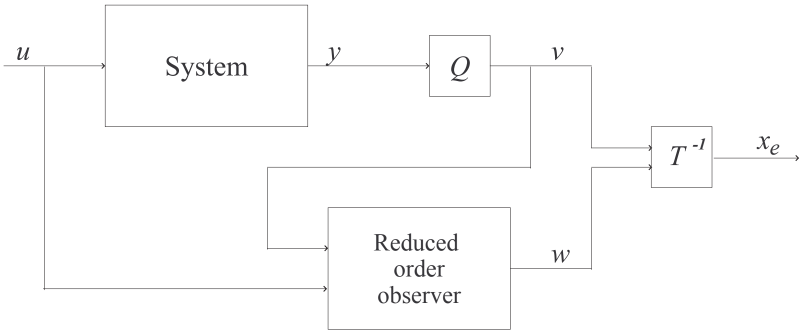

Figure 6.1: A reduced order observer.

law in which the state is replaced by an estimate generated by a state observer.

To this end, consider system (6.81) with $D = 0$, the state feedback control law

$$u = Kx + v,$$

the state observer

$$\sigma\xi = (A + LC)\xi + Bu - Ly \qquad x_e = \xi,$$

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

$$\begin{aligned} \begin{bmatrix} \sigma x \\ \sigma\xi \end{bmatrix} &= \begin{bmatrix} A & BK \\ -LC & A + LC + BK \end{bmatrix} \begin{bmatrix} x \\ \xi \end{bmatrix} + \begin{bmatrix} B \\ B \end{bmatrix} v \\ y &= Cx. \end{aligned} \tag{6.100}$$

To study this system, consider the coordinate

$$e = x - x_e$$

and note that the system can be rewritten in the form

$$\begin{aligned} \begin{bmatrix} \sigma x \\ \sigma e \end{bmatrix} &= \begin{bmatrix} A + BK & -BK \\ 0 & A + LC \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} v \\ y &= Cx. \end{aligned} \tag{6.101}$$

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 characteristic 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 (6.81) by means of a dynamic output feedback control law can be solved provided system (6.81) 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 (6.101), hence system (6.100), 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.

6.7 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

$$\sigma x = Ax + Bu + Pd \qquad\qquad e = Cx + Qd, \tag{6.102}$$

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

$$\sigma d = Sd, \tag{6.103}$$

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.

$$u = Kx + Ld. \tag{6.104}$$

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

$$\sigma\chi = F\chi + Ge \qquad\qquad u = H\chi, \tag{6.105}$$

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 6.3 (Full information regulator problem)**.** *Consider the system (6.102), driven by the exosystem (6.103) and interconnected with the controller (6.104). The full information regulator problem is the problem of determining the matrices K and L of the controller such that*[20]

(S) the system

$$\sigma x = (A + BK)x$$

is asymptotically stable;

(R) all trajectories of the system

$$\sigma d = Sd \qquad \sigma x = (A + BK)x + (BL + P)d \qquad e = Cx + Qd \tag{6.106}$$

[20] (S) stands for stability and (R) for regulation.

are such that

$$\lim_{t\to\infty} e(t) = 0.$$

Definition 6.4 (Error feedback regulator problem)**.** *Consider the system (6.102), dri-ven by the exosystem (6.103) and interconnected with the controller (6.105). 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

$$\sigma d = Sd \quad \sigma x = Ax + BH\chi + Pd \quad \sigma\chi = F\chi + G(Cx + Qd) \quad e = Cx + Qd \tag{6.107}$$

are such that

$$\lim_{t\to\infty} e(t) = 0.$$

6.7.1 The full information regulator problem

Consider the full information regulator problem and assume the following.

Assumption 6.1. *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 6.2. *The system (6.102) with $d = 0$ is reachable.*

Assumption 6.1 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 6.2 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 (6.102) 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 6.2. *Consider the full information regulator problem. Suppose Assumption 6.1 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 $\Pi \in \mathbb{R}^{n\times r}$ such that the equations

$$\Pi S = (A + BK)\Pi + (P + BL) \qquad 0 = C\Pi + Q \tag{6.108}$$

hold.

Proof. Consider the system (6.106) and the coordinates transformation

$$\hat{d} = d \qquad \hat{x} = x - \Pi d,$$

where Π is the solution of the equation[21]

$$\Pi S = (A + BK)\Pi + (P + BL).$$

Note that, by condition (S) and Assumption 6.1, 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 6.1, 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 (6.108) hold. ◁

We are now ready to state and prove the result which provides conditions for the solution of the full information regulator problem.

Theorem 6.1. *Consider the full information regulator problem. Suppose Assumptions 6.1 and 6.2 hold.*

There exists a full information control law described by the equation (6.104) which solves the full information regulator problem if and only if there exist two matrices Π and Γ such that the equations

$$\Pi S = A\Pi + B\Gamma + P \qquad 0 = C\Pi + Q \tag{6.109}$$

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 6.2, there exists a matrix Π such that equations (6.108) hold. As a result, the matrices Π and $\Gamma = K\Pi + L$ are such that equations (6.109) hold.

(Sufficiency) The proof of the sufficiency is constructive. Suppose there are two matrices Π and Γ such that equations (6.109) hold. The full information regulator problem is solved selecting K and L as follows.

[21] 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.

The matrix K is any matrix such that the system

$$\sigma x = (A + BK)x$$

is asymptotically stable. By Assumption 6.2 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 (6.108) hold. This is trivially the case. In fact, replacing L in (6.108) yields the equations (6.109), 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 6.1 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 (6.109) hold. By Assumption 6.2 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 (6.109).

6.7.2 The FBI equations

Equations (6.109), known as the Francis-Byrnes-Isidori (FBI) equations, are linear equations in the unknown Π and Γ, for which the following statement holds.

Lemma 6.3 (Hautus)**.** *The equations (6.109), in the unknown Π and Γ, are solvable for any P and Q if and only if*

$$\operatorname{rank}\begin{bmatrix} sI - A & B \\ C & 0 \end{bmatrix} = n + p, \tag{6.110}$$

for all s which are eigenvalues of the matrix S.

Remark. The equations (6.109) 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 6.3 has a very simple interpretation. In fact, the complex number s such that

$$\text{rank}\begin{bmatrix} sI - A & B \\ C & 0 \end{bmatrix} < n+1$$

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 (6.102), with input u, output e and $d = 0$.

6.7.3 The error feedback regulator problem

To provide a solution to the error feedback regulator problem we need to introduce a new assumption.

Assumption 6.3. *The system*

$$\begin{bmatrix} \sigma x \\ \sigma d \end{bmatrix} = \begin{bmatrix} A & P \\ 0 & S \end{bmatrix}\begin{bmatrix} x \\ d \end{bmatrix} \qquad\qquad e = \begin{bmatrix} C & Q \end{bmatrix}\begin{bmatrix} x \\ d \end{bmatrix} \tag{6.111}$$

is observable.

Note that Assumption 6.3 implies observability of the system

$$\sigma x = Ax \qquad\qquad y = Cx. \tag{6.112}$$

To prove this property note that observability of the system (6.111) implies that

$$\text{rank}\begin{bmatrix} C & Q \\ CA & \star \\ \vdots & \vdots \\ CA^{n+r-1} & \star \end{bmatrix} = n + r.$$

This, in turn, implies

$$\text{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n+r-1} \end{bmatrix} = n$$

and, by Cayley-Hamilton Theorem,

$$\text{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = n,$$

which implies observability of system (6.112). Similarly to what discussed in the case of Assumption 6.2, Assumption 6.3 can be replaced by the weaker assumption that the system (6.111) 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 6.2. *Consider the error feedback regulator problem. Suppose Assumptions 6.1, 6.2 and 6.3 hold.*

There exists an error feedback control law described by the equation (6.105) which solves the full information regulator problem if and only if there exist two matrices Π *and* Γ *such that the equations*

$$\Pi S = A\Pi + B\Gamma + P \qquad 0 = C\Pi + Q \tag{6.113}$$

hold.

Remark. Theorem 6.2 can be alternatively stated as follows.

Consider the error feedback regulator problem. Suppose Assumptions 6.1, 6.2 and 6.3 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 6.1, hence omitted.
(Sufficiency) The proof of the sufficiency is constructive. Suppose there are two matrices Π and Γ such that equations (6.113) hold. Then, by Theorem 6.1 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 6.3, it is possible to build asymptotic estimates ξ and δ of x and d, hence implement the control law

$$u = K\xi + (\Gamma - K\Pi)\delta. \tag{6.114}$$

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

$$\begin{bmatrix} \sigma e_x \\ \sigma e_d \end{bmatrix} = \left(\begin{bmatrix} A & P \\ 0 & S \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \begin{bmatrix} C & Q \end{bmatrix} \right) \begin{bmatrix} e_x \\ e_d \end{bmatrix}, \tag{6.115}$$

hence, by Assumption 6.3, there exist G_1 and G_2 that assign the eigenvalues of this error system.

Note now that the control law (6.114) 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

$$\begin{bmatrix} \sigma x \\ \sigma e_x \\ \sigma e_d \end{bmatrix} = \begin{bmatrix} A + BK & -BK & -B(\Gamma - K\Pi) \\ 0 & A + G_1 C & P + G_1 Q \\ 0 & G_2 C & S + G_2 Q \end{bmatrix} \begin{bmatrix} x \\ e_x \\ e_d \end{bmatrix}. \tag{6.116}$$

Recall that the matrices G_1 and G_2 have been selected to render system (6.115) asymptotically stable, and that K is such that the system $\sigma x = (A + BK)x$ is asymptotically stable. As a result, system (6.116) is asymptotically stable. ◁

6.7.4 The internal model principle

The proof of Theorem 6.2 implies that a controller (it is not the only one) which solves the error feedback regulator problem is described by equations of the form (6.105) with $\chi = \begin{bmatrix} \xi & \delta \end{bmatrix}'$,

$$F = \begin{bmatrix} A + G_1 C + BK & P + G_1 Q + B(\Gamma - K\Pi) \\ G_2 C & S + G_2 Q \end{bmatrix},$$

$$G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}, \qquad H = \begin{bmatrix} K & \Gamma - K\Pi \end{bmatrix}, \tag{6.117}$$

K, G_1 and G_2 such that a stability condition holds, and Π and Γ such that equations (6.113) hold. This controller, and in particular the matrix F, possesses a very interesting property.

Proposition 6.9 (Internal model property)**.** *The matrix F in equation (6.117) 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 $\text{rank}\Sigma = r$, 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. $\triangleleft$

It is possible to prove that the property highlighted in Proposition 6.9 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.