Mathematical Systems Theory

November 2020

Gjerrit Meinsma
Arjan van der Schaft

Department of Applied Mathematics
University of Twente

Preface These are the lecture notes for the course "Systems Theory" which is part of the module "Dynamical Systems" (201500103). The material is based on the notes from 2002 "Inleiding Wiskundige Systeemtheorie voor Informatici" written by Arjan van der Schaft. Over the years the material has been re-ordered and expanded by Gjerrit Meinsma. Till 2017 the notes were in Dutch. Reinie Erné took care of the translation into English (not this preface!).

Contents

Chapter 1

Dynamical Systems

In this opening chapter, we introduce terminology from systems theory and classify different types of dynamical systems. We illustrate these briefly using a series of examples from different fields, highlighting the interdisciplinary nature of the subject.

1.1 Introduction

Words such as *system*, *systems theory*, and *systems engineering* have become ingrained in various fields such as data processing, electrical engineering, economics, management, biology, theoretical computer science, and mathematics; consequently, the exact meaning of these terms is no longer clear-cut. We will therefore first attempt to describe more precisely what we will be studying in this course.

The word *system* refers to an object, device, phenomenon, or part of the environment that causes certain measurable quantities in that environment to be interrelated. We call the measurable quantities variables. In this course, we will mainly concern ourselves with *dynamical* systems. These are systems in which the variables evolve over time. In this case, the variables are often called *signals*. The variables are mostly real valued—the position of a mass in a mechanical system, the current through a wire in an electrical circuit, the height of the interest rate in a model of a national economy, etc.—or discrete—the position of a switch, a symbol equal to 0 or 1, corresponding to “on” or “off”.

In order to reproduce and analyze the dynamical behavior of a system, we will consider a *mathematical model* of the system, which, to some extent, shows how the different variables in the system evolve in relation to one another. In many cases, one and the same system can be connected to different mathematical models, corresponding to different compromises between the precision or descriptive quality of the model and its simplicity. The choice of the mathematical model can also depend on which problem pertaining to the system we wish to study.

Since mathematical models are themselves also systems (of a more abstract nature), it is common to use the word "system" for both the (physical) object of study and its mathematical model. Although systems theory also deals with formulating mathematical models for specific systems, a system in this course will always be an (idealized) mathematical system.

It is essential in systems theory that some of the variables describe the relation between the system and the *environment* of the system, or between the system and other systems. These variables are called the interconnection or *external variables*, and any other variables are called *internal variables*. Think, for example, of a watch, where the external signals could include the position of the watch hands, and the internal variables could include the state of the cogs and whatever else is inside the watch. A useful representation of a system is a box with lines to the environment, where the internal variables are associated with the box and the external variables are associated with the lines; see Figure 1.1.

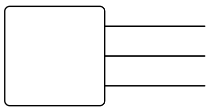


FIGURE 1.1: System.

In a so-called *black box* approach to the system, we cannot or do not wish to describe what happens in the box (i.e., the black box), and the description of the system will only concern the evolution of the external variables (think of the watch). A more detailed description of the box may also be available, for example that it is made up of a number of *subsystems* that are linked through their external variables; see Figure 1.2.

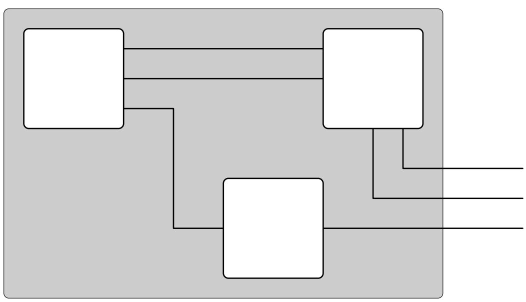


FIGURE 1.2: Complex system.

In many cases, it is useful to separate the external variables into *inputs* and *outputs*. Input variables can be set arbitrarily by the environment of the system—like the voltage across a voltage source in an electrical circuit or whether or not keys are pressed—while, on the other hand, the output variables are set by the input variables and internal variables—like the current supplied by a voltage source or the symbols that appear on the monitor, respectively. Input/output systems are depicted as in Figure 1.3; the arrows toward the box indicate inputs, while the arrows leaving the box indicate outputs.

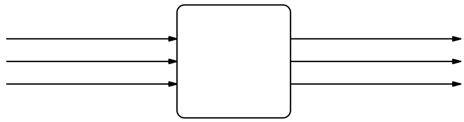


FIGURE 1.3: System with inputs and outputs.

In this course we will mainly deal with input/output systems because of their great importance for applications, and because they are easier to analyze than more general systems.

1.1.1 Examples

Using a number of examples, we will now briefly present typical problems and questions from systems theory. These examples serve as motivation, and we will encounter many of them again further on in the syllabus. Some problems are too ambitious to solve in an introductory course such as this one, some first need to be idealized because they are otherwise too complex, while others can be solved completely using the techniques that will be introduced in this course.

- *Steering and observing.* What is necessary to steer a car? On the one hand, the driver must be able to *control* the behavior of the car to some extent (which is why there is a steering wheel!), and on the other hand, the driver must *observe* the surroundings well enough to control successfully. What does the driver need to observe, and how aggressively may we pull on the steering wheel to keep the car on the road?

  The panacea here is *feedback*. Feedback is omnipresent, for example in biomedical systems such as the regulation of your blood sugar levels. We will discuss feedback extensively.

- *Tracking.* How can we, while walking, follow a moving object perfectly with our eyes? If we can describe this with a mathematical model, then we might also be able to let cameras follow moving objects. Other similar problems: How can a control tower track the trajectories of planes? How can we ensure that the laser beam in the CD player follows the track on the CD?

- *Filtering & signal processing.* How do equalizers and noise suppressors work? How can we remove undesirable properties of signals? The “device” that does this is often called a filter. What are convolution filters (popular in, for example, video-image processing)? Why are jpeg files so small, and why are jpeg images often more fuzzy than the original picture?

- *Robust control.* How can we design a cruise control system that ensures that the car maintains a constant speed regardless of road slope or wind conditions? Similarly, how can we ensure that the laser beam in the CD player continues to follow the track despite perturbations?

- *Uncertain models.* Adjusting the temperature in an unfamiliar shower can lead to painful situations. This is because we do not know the shower, in other words, because we have a defective model of the operation of the shower. In general, it is difficult to draw conclusions using defective models. How can we bypass the defects? For example, we could take more time to adjust the temperature, or is there a more advance solution? A well-known example of successfully controlling uncertain models is *Black's negative feedback amplifier*.

- *Divide and conquer.* To model complex systems, it is helpful to view a complex system as an interconnection of subsystems. When are the interconnections set correctly? How can we simulate complex systems by combining simulations of the subsystems?

- *System identification.* How can we improve a mathematical model of, say, a car by experimenting with the car? Which experiments are necessary to obtain as much information as possible? We will not discuss this in this course.

- *Stochastic models.* The influence of noise, wind gusts, measurement errors, etc. are difficult to model. In such cases, it may be helpful to see these as realizations of stochastic processes. Systems with stochastic components form an important subfield of systems theory. We will not discuss it here.

1.2 Axioms of Dynamical Systems

As already indicated, we deal mainly with *dynamical* systems, that is, systems in which the variables can change over "time". In this section, we will formalize what we mean by a "dynamical system".

Suppose that we are interested in the motion of "the falling apple". Under the influence of gravity, the vertical distance w from the apple to a certain point of reference will change as a function of time according to the equation[1]

$$\ddot{w}(t) = -g, \tag{1.1}$$

where g is the gravitational acceleration. Now, in the literature, it is not unusual to identify systems with equations, so that one, for example, speaks of "the system $\ddot{w}(t) = -g$". This is rather unsatisfactory, because someone else could just as well

[1] By definition, $\dot{w}(t) = \frac{\mathrm{d}}{\mathrm{d}t} w(t)$; likewise, $\ddot{w}(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} w(t)$, etc.

describe the motion of the falling apple using

$$w(t) = -\frac{g}{2}t^2 + at + b \qquad \text{for some } a, b \in \mathbb{R}. \tag{1.2}$$

Instinctively, this is the same system, but the representation is different. If we want (1.1) and (1.2) to denote the same system, then we must define the notion of "system" using what these representations have in common: the set $\mathfrak{B}$ of possible signals,

$$\begin{aligned} \mathfrak{B} &:= \{w : \mathbb{R} \to \mathbb{R} \mid \ddot{w}(t) = -g\} \\ &= \{w : \mathbb{R} \to \mathbb{R} \mid \exists a, b \in \mathbb{R} : w(t) = -\tfrac{g}{2}t^2 + at + b \ \forall t\}. \end{aligned} \tag{1.3}$$

This idea is the basis of what is known as the *behavioral* approach to systems theory, where the *set of possible signals* is taken as the starting point. In the behavioral approach, a system is formally defined as follows.

Definition 1.2.1 (Dynamical system). A *dynamical system* Σ is a triple $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ with $\mathbb{T}$ and $\mathbb{W}$ sets and $\mathfrak{B} \subseteq \{w : \mathbb{T} \to \mathbb{W}\}$. □

We call $\mathbb{T}$ the *time axis*; it is the set of times we consider. The set $\mathbb{W}$ is called the *signal space*, which is the space of values the signals can take on, and $\mathfrak{B}$ is called the *behavior* of the system. The behavior is nothing more than the set of signals that can occur in the system. In the example of the falling apple, the time variable t is real valued, so for the time axis we can for example choose $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = [0, \infty)$. The height $w(t)$ is also real, so $\mathbb{W} = \mathbb{R}$ is a natural choice, and we have already described the behavior $\mathfrak{B}$ in (1.3).

Here are three other examples.

Example 1.2.2 (Dynamical systems).

- *The falling apple 2.* Suppose that until time $t = 0$, the falling apple hangs at a height of 5 meters, and that it is then released. The behavior for nonnegative time is then

  $$\mathfrak{B} = \{w : \mathbb{T} \to \mathbb{W} \mid \ddot{w}(t) = -g, \ w(0) = 5, \ \dot{w}(0) = 0\},$$

  and we choose $\mathbb{T} = [0, \infty)$ and $\mathbb{W} = \mathbb{R}$.

- *Bank balance 1.* Suppose that on January 2 of the year 0, we put w_0 euros into a savings account, at a fixed interest rate of 5% a year. If interest is paid only once a year (say on January 2), then we can view the bank balance as a dynamical system with discrete time axis the years:

  $$\mathbb{T} = (0, 1, 2, 3, \ldots).$$

  As signal space, we can use $\mathbb{W} = \mathbb{R}$, and we define the behavior as

  $$\mathfrak{B} = \{w : \mathbb{T} \to \mathbb{W} \mid w[t+1] = (1 + \tfrac{5}{100})w[t], \ w(0) = w_0\}.$$

Note that the representation of the behavior is not unique. We could also have written

$$\mathfrak{B} = \{w : \mathbb{T} \to \mathbb{W} \mid w[t] = w_0(1 + \tfrac{5}{100})^t\}.$$

Of course, the bank balance does not depend on the choice of the representation. In other words, it is the behavior that matters, not one of the many representations of it. This is exactly what is formalized by Definition 1.2.1.

- *Bank balance 2.* We can also define the "bank balance" for an arbitrary input $w_0 \in \mathbb{R}$ (that may be negative). In that case, we could choose the behavior to be

$$\mathfrak{B} = \{w : \mathbb{T} \to \mathbb{W} \mid w[t+1] = (1 + \tfrac{5}{100})\, w[t]\}$$

with $\mathbb{T}$ and $\mathbb{W}$ as defined earlier.

□

The behavioral approach is axiomatic and precise, but often overly formal and laborious. It is not necessary to give the triple $(\mathbb{T}, \mathbb{W}, \mathfrak{B})$ for every system. We will, for example, still say "the system $\ddot{w}(t) = -g$" because it is usually clear from the context which triple $(\mathbb{T}, \mathbb{W}, \mathfrak{B})$ is meant. However, if we want to put on the finishing touches, then it is important that we can rely on this formal definition of a dynamical system.

We can distinguish between many different types of systems and signals. We continue this chapter with an overview of a number of types and we illustrate them using a series of examples.

1.3 Two Signals

In systems theory and signal processing, in addition to the standard functions from mathematics, we also use the (unit) *step function* $\mathbb{1} : \mathbb{R} \to \mathbb{R}$, defined as

$$\mathbb{1}(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$

(see Figure 1.4), and the *delta function* $\delta(t)$. The delta function will be studied in detail in another course. It is, simply put, the function $\delta : \mathbb{R} \to \mathbb{R}$ that is zero everywhere except on a tiny interval around zero, where it has a spike that is so high that the integral over the spike is 1,

$$\int_{0^-}^{0^+} \delta(t)\, \mathrm{d}t = 1;$$

see Figure 1.4. The delta function is the derivative of the step function.

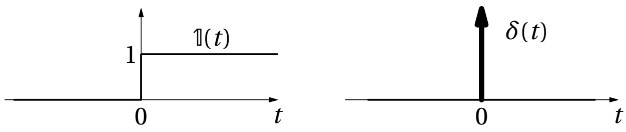


FIGURE 1.4: Step function and delta function.

1.4 Continuous and Discrete Time

We can be brief about this. In the examples in this course, the time will often change continuously, and we will usually take $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = \mathbb{R}_+ := [0, \infty)$. We call systems with such time axes *continuous-time systems*. A second important class of systems is that where time can take on a countable number of values, for example $\mathbb{T} = \mathbb{Z}$ or a subset thereof. We call such systems *discrete-time systems*. So, in Example 1.2.2 the "falling apple 2" is a continuous-time system and the "bank balance" is a discrete-time system.

We generally denote the time variable by the letter t. The value of a *discrete-time* signal w is sometimes denoted by $w[t]$ instead of the more common $w(t)$.

1.5 Linearity and Time Invariance

Definition 1.5.1 (Linearity). A system $(\mathbb{T}, \mathbb{W}, \mathfrak{B})$ is *linear* if $\mathbb{W}$ is a vector space and $\mathfrak{B}$ is a subspace of[2] $\{w : \mathbb{T} \to \mathbb{W}\}$, that is,

1. $0 \in \mathfrak{B}$;

2. if $w_1, w_2 \in \mathfrak{B}$, then $w_1 + w_2 \in \mathfrak{B}$;

3. if $w \in \mathfrak{B}$, then $\lambda w \in \mathfrak{B}$ for all scalars λ.

□

In a linear system, the behavior is by definition a subspace of a vector space, and therefore itself a vector space. Linearity is a strong property and simplifies the analysis of systems.

Example 1.5.2 (Electrical circuit). Figure 1.5 shows an RC circuit, where u is the voltage supplied by the voltage source and y is the voltage across the capacitor.

[2] If we want to be precise: because $\mathbb{W}$ is a vector space, the set of signals $\{w : \mathbb{T} \to \mathbb{W}\}$ naturally is a vector space under addition and scalar multiplication defined pointwise by $(w + v)(t) := w(t) + v(t)$ and $(\alpha v)(t) := \alpha\, v(t)$.

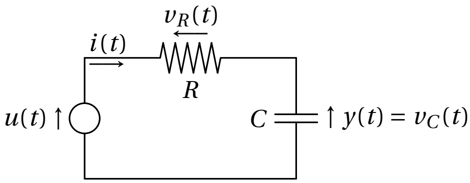


FIGURE 1.5: An RC circuit (Example 1.5.2).

The standard model for the relation between u and y is the differential equation[3]

$$\dot{y}(t) + \alpha y(t) = \alpha u(t) \tag{1.5}$$

with $\alpha = \frac{1}{RC} > 0$. We define the behavior of this system as all possible u, y:

$$\mathfrak{B} = \{(u, y) : \mathbb{R} \to (\mathbb{R}, \mathbb{R}) \mid \dot{y} + \alpha y = \alpha u\}.$$

We do not worry about the differentiability of the signals.

The system is linear because the behavior is a subspace. Indeed,

1. $(y, u) = (0, 0)$ satisfies $\dot{y} + \alpha y = u$, so $(0, 0) \in \mathfrak{B}$;

2. if $(u_1, y_1), (u_2, y_2) \in \mathfrak{B}$, then the sum $(\tilde{u}, \tilde{y}) = (u_1, y_1) + (u_2, y_2)$ is also an element of $\mathfrak{B}$ because it satisfies the system equation:

$$\begin{aligned} \dot{\tilde{y}} + \alpha \tilde{y} &= (\dot{y}_1 + \dot{y}_2) + \alpha(y_1 + y_2) \\ &= (\dot{y}_1 + \alpha y_1) + (\dot{y}_2 + \alpha y_2) \\ &= \alpha u_1 + \alpha u_2 \\ &= \alpha(u_1 + u_2) \\ &= \alpha \tilde{u}; \end{aligned}$$

3. if $(u, y) \in \mathfrak{B}$, then $(\tilde{u}, \tilde{y}) := \lambda(u, y)$ is also an element of $\mathfrak{B}$ because it also satisfies the system equation:

$$\begin{aligned} \dot{\tilde{y}} + \alpha \tilde{y} &= (\lambda \dot{y}) + \alpha(\lambda y) \\ &= \lambda(\dot{y} + \alpha y) = \lambda \alpha u = \alpha \tilde{u}. \end{aligned}$$

We apparently do not need to solve the differential equation explicitly to verify the linearity. □

[3] Deduction: By Kirchhoff's laws, we have

$$u(t) = v_R(t) + y(t) = Ri(t) + y(t). \tag{1.4}$$

The standard model for the capacitor is that the voltage across the capacitor is proportional to the electric charge on it, $y(t) = q(t)/C$. The derivative of the electric charge is current, so $C\dot{y}(t) = i(t)$. Substituting this expression for $i(t)$ in (1.4) gives (1.5).

In practice, linearity rarely holds exactly. For example, the *RC* circuit will no longer be linear if the current is so high that the resistor threatens to burn out.

A property that often holds is *time invariance*. Roughly speaking, this means that the system behaves the same way today as it does tomorrow. To define this properly, we introduce the *delay operator* or *backshift operator* σ. This works on signals and is defined by

$$(\sigma^\tau w)(t) = w(t-\tau),$$

Figure 1.6 gives some insight into this operator.

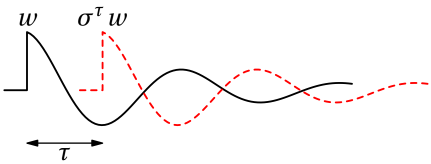


FIGURE 1.6: Graph of a signal $w : \mathbb{R} \to \mathbb{R}$ and of the delayed signal $\sigma^\tau w : \mathbb{R} \to \mathbb{R}$.

Definition 1.5.3 (Time invariance). Let $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = \mathbb{Z}$. A system $(\mathbb{T}, \mathbb{W}, \mathfrak{B})$ is *time invariant* if $w \in \mathfrak{B}$ implies $\sigma^\tau w \in \mathfrak{B}$ for all $\tau \in \mathbb{T}$. We call a system *time varying* if it is not time invariant. □

One could say that time-invariant systems have no built-in clock.

Example 1.5.4 (Electrical circuit—continued). The *RC* circuit with behavior

$$\mathfrak{B} = \{(u, y) : \mathbb{R} \to (\mathbb{R}, \mathbb{R}) \mid \dot{y} + \alpha y = \alpha u\}$$

is time invariant. Intuitively, this is because the resistance R and capacity C, and therefore $\alpha = 1/(RC)$, do not depend on time. Formally, we can show it as follows: let $(u, y) \in \mathfrak{B}$ and for arbitrary $t_0 \in \mathbb{R}$, define the delayed signals

$$\tilde{u}(t) := u(t - t_0), \quad \tilde{y}(t) := y(t - t_0).$$

Then we also have $(\tilde{u}, \tilde{y}) \in \mathfrak{B}$ because it satisfies the system equation:

$$\begin{aligned} \dot{\tilde{y}}(t) + \alpha \tilde{y}(t) &= \dot{y}(t - t_0) + \alpha y(t - t_0) \\ &= \alpha u(t - t_0) \\ &= \alpha \tilde{u}(t). \end{aligned}$$

Because this holds for all t_0, the system is time invariant. □

In the next section, we will also encounter time-varying systems.

1.6 Input and Output

In some cases, we can divide the entries of $w \in \mathfrak{B}$ into "cause" and "effect". We often encounter these types of systems in signal processing and in communication theory.

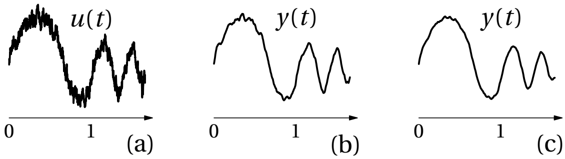


FIGURE 1.7: (a): A signal with noise u; (b) moving average y for $T = 0.03$; (c) for $T = 0.09$.

Example 1.6.1 (Moving averages & noise reduction). Strong fluctuations (noise) in a signal $u : \mathbb{R} \to \mathbb{R}$ can be suppressed by averaging the signal over a certain period T:

$$y(t) := \frac{1}{T} \int_{t-T}^{t} u(\tau)\, \mathrm{d}\tau. \tag{1.6}$$

We can expect y to behave somewhat less erratically than u, but as long as T is not too great, the graph of y will follow the course of u reasonably well; see Figure 1.7. The period T in part (c) of this figure is greater than that in part (b); consequently, y is smoother and flatter in (c) than it is in (b).

The system with behavior $\{\left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] : \mathbb{R} \to \mathbb{R}^2 \mid (1.6) \text{ holds}\}$ is called a *moving-average system.* □

Characteristic of this system is that

- the signal u is not constrained by the system equations.

Every signal u is permitted (subject to a technical integrability condition). Moreover,

- the signal y is completely determined by u.

In other words, we can write

$$y = \mathscr{H}(u)$$

for some map $\mathscr{H}$. Many systems have these two properties. We call them *input/output systems*, where the signal u is the *input* and the signal y is the *output* or *response* of the system. Figure 1.8 depicts such a system. In this course, we will always denote the *input* by u, and draw the corresponding arrow in the diagram as pointing *toward* the box. The interpretation is that u is not restricted by

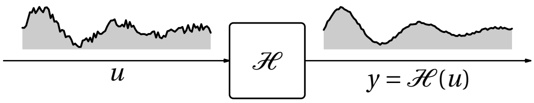


FIGURE 1.8: A system with input u and output y.

the system, and may be chosen freely. We denote the *output* by y, and the corresponding arrow points away from the box, as if y is produced by the box and its input.

The linearity of an input/output system $y = \mathscr{H}(u)$ corresponds exactly to the linearity of the map $\mathscr{H}$.

Lemma 1.6.2 (Linearity and time-invariant maps). *A system with behavior* $\mathfrak{B} = \{(u, y) : \mathbb{T} \to (\mathbb{U}, \mathbb{Y}) \mid y = \mathscr{H}(u)\}$ *for some map* $\mathscr{H}$ *is linear if and only if* $\mathscr{H}$ *is a linear map:*

1. $\mathscr{H}(u_1 + u_2) = \mathscr{H}(u_1) + \mathscr{H}(u_2)$*;*
2. $\mathscr{H}(\lambda u) = \lambda \mathscr{H}(u)$ *for all scalars* λ*.*

The system is time invariant if and only if the output of the delay is equal to the delay of the output (see Figure 1.9), that is,

$$\mathscr{H}(\sigma^\tau u) = \sigma^\tau \mathscr{H}(u) \qquad \forall \tau \in \mathbb{T},\ u : \mathbb{T} \to \mathbb{U}.$$

Proof. Exercise 1.10. ∎

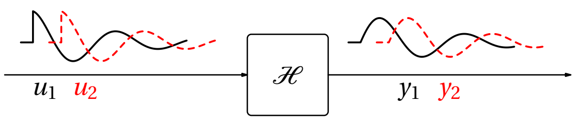


FIGURE 1.9: A time-invariant system $y = \mathscr{H}(u)$.

More succinctly, a map $\mathscr{H}$ is time invariant exactly when $\mathscr{H}$ commutes with the delay operator: $\mathscr{H}\sigma^\tau = \sigma^\tau \mathscr{H}$ for all $\tau \in \mathbb{T}$.

Example 1.6.3 (Moving-average system). The system (1.6) is linear because the map $\mathscr{H}$ defined by

$$y = \mathscr{H}(u) : \qquad y(t) = \frac{1}{T}\int_{t-T}^{t} u(\tau)\,\mathrm{d}\tau$$

is a linear map. Indeed,

$$\begin{aligned}
\mathscr{H}(u_1 + u_2)(t) &= \frac{1}{T}\int_{t-T}^{t} (u_1 + u_2)(\tau)\,\mathrm{d}\tau \\
&= \frac{1}{T}\int_{t-T}^{t} u_1(\tau)\,\mathrm{d}\tau + \frac{1}{T}\int_{t-T}^{t} u_2(\tau)\,\mathrm{d}\tau \\
&= \mathscr{H}(u_1)(t) + \mathscr{H}(u_2)(t)
\end{aligned}$$

and for every scalar λ,

$$\begin{aligned}
\mathscr{H}(\lambda u)(t) &= \frac{1}{T}\int_{t-T}^{t} (\lambda u)(\tau)\,\mathrm{d}\tau \\
&= \lambda \frac{1}{T}\int_{t-T}^{t} u(\tau)\,\mathrm{d}\tau \\
&= \lambda \mathscr{H}(u)(t).
\end{aligned}$$

The system is also time invariant. To show this, let $y_0 = \mathscr{H}(u_0)$ and define the delayed variables by $\tilde{u}(t) := u_0(t-t_0), \tilde{y}(t) := y_0(t-t_0)$. We must show that $\mathscr{H}(\tilde{u}) = \tilde{y}$. This is true because

$$\begin{aligned}
\mathscr{H}(\tilde{u})(t) &= \frac{1}{T}\int_{t-T}^{t} \tilde{u}(\tau)\,\mathrm{d}\tau \\
&= \frac{1}{T}\int_{t-T}^{t} u(\tau - t_0)\,\mathrm{d}\tau \\
&\quad \{\text{substitute } s = \tau - t_0\} \\
&= \frac{1}{T}\int_{(t-t_0)-T}^{t-t_0} u(s)\,\mathrm{d}s \\
&= y_0(t-t_0) \\
&= \tilde{y}(t),
\end{aligned}$$

so the system is time invariant. □

Example 1.6.4 (Linear time-varying system). Suppose that $u, y : \mathbb{R} \to \mathbb{R}$ are related through

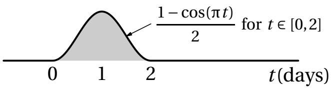
$$y(t) = t u(t), \qquad t \in \mathbb{R}.$$

This system is linear, because the map $\mathscr{H}(u) = \alpha u$ with $\alpha(t) = t$ is a linear map (verify), but the system is not time invariant. To show this, it suffices to give one counterexample to the time invariance. Choose, for example, $u_0(t) = 1$ (the constant function 1). The output is then $y_0(t) = t$. But the output of the delay $\tilde{u}(t) = u_0(t-1) = 1$ is then t, which is not equal to the delay of the output $y_0(t-1) = t-1$. So the system is not time invariant. □

Example 1.6.5 (Drainage system). A nice illustration of a time-invariant system is a drainage system. As input $u(t)$, we take the number of liters of water that fall onto a soccer field in Dinkel per unit of time, and the output $y(t)$ is the number of liters of water that flow from the field into the river the Dinkel per unit of time. If it rains one day, for example

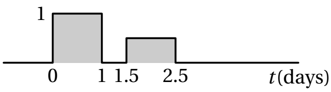
$u_0(t) =$ 

then, if nothing evaporates, all of the water will eventually flow into the Dinkel, but this takes time. The outflow for our $u_0(t)$ could be as follows:

$y_0(t) =$ 

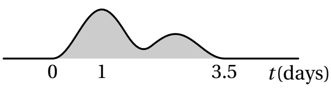
It takes a while before the outflow $y(t)$ begins, and after the last drop of rain $u(t)$, it takes another day before the outflow is over. This system is time invariant, because if it starts raining a day later, the water will also flow into the Dinkel a day later. If the system is linear, then for an influx

$$u_1(t) = u_0(t) + \frac{1}{2}u_0(t-1.5)$$

$=$ 

we have the outflow

$$\begin{aligned} y_1(t) &= \mathscr{H}(u_0 + \tfrac{1}{2}\sigma^{1.5}u_0)(t) \\ &= \mathscr{H}(u_0)(t) + \tfrac{1}{2}\mathscr{H}(\sigma^{1.5}u_0)(t) && \text{(linearity)} \\ &= \mathscr{H}(u_0)(t) + \tfrac{1}{2}\sigma^{1.5}\mathscr{H}(u_0)(t) && \text{(time invari-} \\ &= y_0(t) + \tfrac{1}{2}y_0(t-1.5) && \text{-ance)} \end{aligned}$$

$=$ 

Incidentally, it is nice to realize that because of the conservation of mass, we always have "total influx = total outflow": $\int_{-\infty}^{\infty} u(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} y(t)\,\mathrm{d}t$. □

1.6.1 Causality

Your car only starts moving *after* you start it. This is what we call "causality". Most systems are causal, and when we design systems, we must take into account that the implementation must be causal. Though if we may edit our signals "off-line"—think of the restoration of music on old LPs—then "noncausal" systems are allowed, but this means that you first record the entire LP onto your computer before you produce any sound.

The causality of input/output systems $y = \mathscr{H}(u)$ is often defined as follows.

Definition 1.6.6 (Causal map). Let $\mathbb{T} = \mathbb{Z}$ or $\mathbb{T} = \mathbb{R}$. A system $y = \mathscr{H}(u)$ is *causal* if a future change in the input $u(t)$ does not have an effect on the past of the output, that is, if for every $t_0 \in \mathbb{T}$ and for every two inputs u_1, u_2 that satisfy

$$u_1(t) = u_2(t) \quad \forall t < t_0,$$

we have

$$y_1(t) = y_2(t) \quad \forall t < t_0,$$

where $y_1 = \mathscr{H}(u_1)$ and $y_2 = \mathscr{H}(u_2)$. □

Causality is illustrated by Figure 1.10. An example of a causal system is $y(t) = u(t-1)$. By contrast, $y(t) = u(t+1)$ is a noncausal (or acausal) system.

Example 1.6.7 (Moving-average system). The moving-average system

$$y(t) := \frac{1}{T}\int_{t-T}^{t} u(\tau)\,\mathrm{d}\tau$$

is causal (assuming $T > 0$). This is because the integral is taken over only $\tau \in [t-T, t]$, hence does not use the future $> t$. Formally, we can prove it as follows: Suppose

$$u_1(t) = u_2(t) \quad \forall t < t_0.$$

Then for all $t < t_0$, we have

$$\begin{aligned} y_1(t) &= \frac{1}{T}\int_{t-T}^{t} u_1(\tau)\,\mathrm{d}\tau \\ &= \frac{1}{T}\int_{t-T}^{t} u_2(\tau)\,\mathrm{d}\tau \\ &= y_2(t). \end{aligned}$$

(We are allowed to replace $u_1(\tau)$ by $u_2(\tau)$ because $\tau \in [t-T, t]$, so $\tau \le t < t_0$.) □

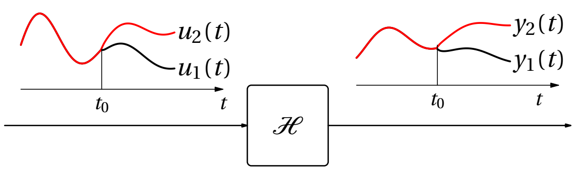


FIGURE 1.10: In a causal system, a change in the input does not affect the past of the output.

1.7 States

Note that the moving-average system depends not only on u, but also on the period T over which the average is taken. When we wish to make this dependence explicit, we write

$$y = \mathscr{H}(u, T).$$

An example similar to the previous ones is that of “the car” with input u the position of the accelerator pedal and output y the position of the car. The position of the accelerator pedal, u, is not restricted (within certain limits) by the dynamics of the car, while the position y is, of course, restricted by the dynamics of the car, and is influenced by u. But unlike in systems represented by maps, the output y is not *completely* determined by the input u. For example, the position y also depends on the initial position of the car,

$$y = \mathscr{H}(u, x_1, x_2, \ldots, x_n),$$

where the x_j are the coordinates of the initial position (and possibly other factors that we want to make explicit).

As in the previous section, we call u the *input* and y the *output* or *response* of the system, and we use Figure 1.8 to illustrate it.

Example 1.7.1 (Electrical circuit). Consider, once more, the RC circuit of Figure 1.5. The relation between the source voltage u and the voltage y across the capacitor is given by

$$\dot{y} + \alpha y = \alpha u \tag{1.7}$$

with $\alpha = \frac{1}{RC} > 0$. The solution of this differential equation can be determined using variation of constants (we will elaborate on this later; see, for example, page 38), which gives

$$y(t) = \mathrm{e}^{-\alpha t}\frac{q(0)}{C} + \int_0^t \mathrm{e}^{-\alpha(t-\tau)}\alpha u(\tau)\,\mathrm{d}\tau \quad \forall t \in \mathbb{R}.$$

We see that at any time t, the voltage $y(t)$ depends on the voltage u supplied over the time interval $[0, t]$, but also on the charge $q(0)$ present on the capacitor at time 0. In abstract terms,

$$y(t) = \mathscr{H}(q(0), u(\tau)\big|_{\tau\in[0,t]})$$

for some map $\mathscr{H}$. □

In this course, a prominent role is played by causal systems where the relation between u and y is of the form

$$y(t) = \mathscr{H}(x(t_0), u(\tau)\big|_{\tau\in[t_0,t]}, t) \quad \forall t, t_0 \in \mathbb{T},\ t \geq t_0 \tag{1.8}$$

for some map $\mathscr{H}$. The electrical circuit of Example 1.7.1 is of this form, with $x = q$. If we view t_0 as the “present” and $t_1 > t_0$ as the “future”, then (1.8) says that the value of $y(t_1)$ in the future depends only on the input $u(t)$ over the future interval $[t_0, t_1]$ and the value of x at the present time t_0. That is, the present value, $x(t_0)$, of x is all that is needed from the past to determine the future. You could say that the variable $x(t_0)$ contains all past information necessary to determine the future. For this reason, x is called a *state* of the system. This property will now be used as the definition of a state. The following is an informal definition.

Definition 1.7.2 (State). If a system with input u and output y is of the form (1.8) for some map $\mathscr{H}$ and time axis $\mathbb{T} \subseteq \mathbb{R}$, then we call x a *state* of the system. □

In addition to the "information-theoretical" interpretation of the notion of state, the state of a system often has a physical meaning. In many physical systems, the distribution of energy over the system forms the state of the system. This is closely related to the fact that physical systems can often be interpreted as an interconnection of energy-storing subsystems that interact with one another by means of energy flows. In Example 1.7.1, the total energy of the system is given by

$$E = \frac{1}{2C} q^2,$$

the energy due to the charge q "carried by" the capacitor.

In the electrical circuit, the state $x(t_0)$ at any time t_0 is an element of $\mathbb{R}$. We then say that $\mathbb{X} = \mathbb{R}$, the *state space* $\mathbb{X}$ of the system is $\mathbb{R}$, and because $\mathbb{R}$ is finite dimensional, we say that the model of the electrical circuit is *finite dimensional.*

Definition 1.7.3 (Finite and infinite dimensional). A system with input u and output y and time axis $\mathbb{T} \subseteq \mathbb{R}$ is *finite dimensional* if there exists a finite-dimensional space $\mathbb{X}$ such that (1.8) holds for some map $\mathscr{H}$ with $x(t_0) \in \mathbb{X}$ for all $t_0 \in \mathbb{T}$. □

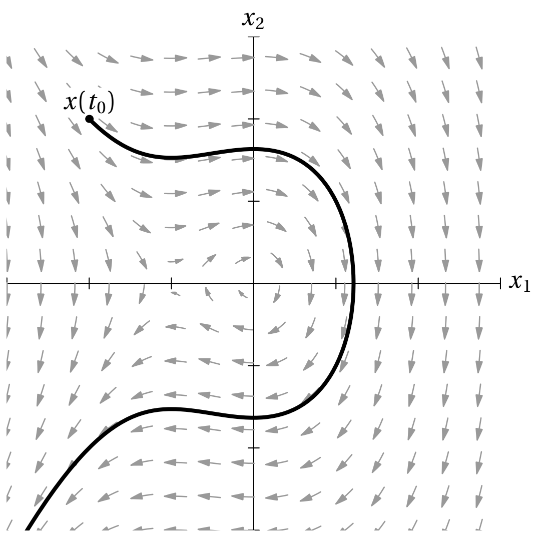


FIGURE 1.11: A vector field (Example 1.7.4).

Example 1.7.4 (Vector fields). In a first course on ordinary differential equations, many systems of the form

$$\dot{x}(t) = f(x(t)), \qquad x : \mathbb{R} \to \mathbb{R}^n, \ f : \mathbb{R}^n \to \mathbb{R}^n$$

are studied. In almost all cases, the evolution of $x(t)$ is fully determined by an initial condition $x(t_0)$. That is, if $x(t_0)$ is given for some t_0, then in general there is a unique extension of $x(t)$ with that initial condition. In other words, we can write

$$x(t) = \mathscr{H}(x(t_0), t), \qquad t \geq t_0. \tag{1.9}$$

That this holds, is intuitively clear when we look at a plot of the vector field f: Figure 1.11 shows an example of a vector field (the set of arrows) for $n = 2$. More precisely, it shows the vector field of

$$f(x_1, x_2) = \begin{bmatrix} x_2 \\ -x_1(1+x_1) \end{bmatrix}.$$

For a given $x(t_0) = \left[\begin{smallmatrix} x_1(t_0) \\ x_2(t_0) \end{smallmatrix}\right] \in \mathbb{R}^2$, the $x(t) = \left[\begin{smallmatrix} x_1(t) \\ x_2(t) \end{smallmatrix}\right]$ can do nothing else than "flow" with the vector field. Because of (1.9), the variable x can be viewed as a state and therefore the initial condition $x(t_0)$ is also called the *initial state*, and because $x(t) \in \mathbb{X} := \mathbb{R}^n$, the system is finite dimensional. □

In physical examples such as 1.7.1, finite dimensionality of the state space is often a consequence of idealization. For example, the charge q in Example 1.7.1 is assumed to be concentrated in one point. For some physical systems, or for a greater accuracy of the models, it is necessary to take into account that some quantities are distributed spatially. We call the resulting models *distributed parameter systems*, as opposed to the finite-dimensional *lumped parameter systems*.

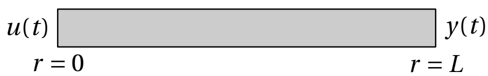


FIGURE 1.12: Heated beam.

Example 1.7.5 (Heated beam–infinite-dimensional system). Suppose that a thin beam of length L is thermally isolated from the environment. We denote by $T(t, r)$ the temperature at time t and place r (see Figure 1.12). At the left extremity of the beam, a heat source $u(t)$ is added, and we are interested in the temperature at the other extremity, $y(t) := T(t, L)$.

Over time, the temperature at every point of the beam influences the temperature at the extremity, so we must take the entire *function* $T(t, \cdot)$ as state:

$$x(t_0) : [0, L] \to \mathbb{R}, \qquad x(t_0)(r) = T(t_0, r). \tag{1.10}$$

From physical considerations, we may assume that the temperature varies smoothly over the beam, and therefore that the function in (1.10) is infinitely differentiable. Therefore, for the state space we take $\mathbb{X} = C^\infty[0, L]$ (the infinite-dimensional space of infinitely differentiable functions from $[0, L]$ to $\mathbb{R}$). The system is infinite-dimensional. In Appendix A.1, the model is described in greater detail. □

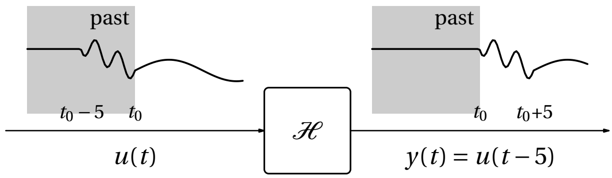


FIGURE 1.13: System with delay.

The seemingly innocent continuous-time system

$$y(t)=u(t-5), \qquad (t\in\mathbb{R})$$

is also infinite dimensional. This becomes clear in Figure 1.13. As the state must contain that information from the past that may influence the output in the future, the entire input u over the previous five time periods— , see Figure 1.13— must be included in the state. The "smallest" state $x(t_0)$ for every t_0 is therefore the *function*

$$x(t_0):[0,5]\to\mathbb{R}, \qquad x(t_0)(\tau)=u(t_0-\tau).$$

For the space of continuous outputs u, this gives the infinite-dimensional state space

$$\mathbb{X}=\{x:\mathbb{R}\to C[0,5]\}$$

($C[0,5]$ is the space of continuous functions from $[0,5]$ to $\mathbb{R}$.) Note that the state is not unique, we could also have chosen a state with a longer memory

$$x(t_0):[0,9]\to\mathbb{R}, \qquad x(t_0)(\tau)=u(t_0-\tau).$$

This is not practical, but it is permitted.

The notion of state also exists for discrete-time systems.

Example 1.7.6 (Discrete-time system). Consider the discrete-time system

$$y[t]=\max_{j\leq t}u[j], \qquad (t\in\mathbb{Z},\ u[t]\in\mathbb{R}).$$

In other words, the output $y[t]$ remembers the hitherto maximum value of the input. The only thing we need to remember from the past $u[j]$ of the input for $j\leq t_0$ to determine $y[t]$ for $t\geq t_0$ is the maximum value of $u[j]$ during that past:

$$x[t_0]=\max_{j\leq t_0}u[j].$$

This $x[t_0]$ indeed satisfies the definition of state, because

$$y[t]=\max\bigl(x[t_0],u[t_0+1],u[t_0+2],\dots,u[t]\bigr) \quad \forall t\geq t_0,$$

and this is of the form (1.8). The system is finite dimensional because $x[t]\in\mathbb{R}$. $\square$

1.8 State Representations

An effective modeling method consists in formulating states and indicating how one state passes to the next.

1.8.1 Discrete-Time Systems

Example 1.8.1 (Parking meter). A 60-minute parking meter can be modeled as follows. The meter has 61 possible states, one for every minute,

$$\mathbb{X} := \{0, 1, 2, \dots, 60\}.$$

The only way to influence the remaining time is by inserting money into the meter. As possible inputs, we set

$$\mathbb{U} := \{0.2€, 0.50€, 1€, \varnothing\}.$$

(The $\varnothing$ means that we do not insert any money.) With each insertion of money, the remaining time increases by respectively 10, 25, or 60 minutes, In order not to get bogged down in more complex models, we then propose that the parking meter accepts only one coin per minute! The operation of the parking meter is now fully determined by indicating how one state passes on to the next:

$$x[t+1] = \begin{cases} \max(0, x[t]-1) & \text{if } u[t] = \varnothing \\ \min(60, x[t]-1+10) & \text{if } u[t] = 0.2€ \\ \min(60, x[t]-1+25) & \text{if } u[t] = 0.5€ \\ \min(60, x[t]-1+60) & \text{if } u[t] = 1€ \end{cases}$$

for all $t \in \mathbb{N}$ and $x[0] \in \mathbb{X}$. As possible outputs, we can choose

$$\mathbb{Y} := \{\text{empty}, \text{running}\}$$

with $y[t] = \text{empty}$ if $x[t] = 0$ and $y[t] = \text{running}$ otherwise. □

The model has deliberately been kept simple (there are innumerable variants, just try to make such a model for a coffee machine). The example illustrates that there are cases where the signals do not need to be real valued. The input spaces $\mathbb{U}$ and output spaces $\mathbb{Y}$ can also consist of symbols.

The model of the parking meter is an example of a discrete-time *state representation*. These are representations of the form

$$\begin{aligned} x[t+1] &= f(x[t], u[t], t), \qquad & u[t] \in \mathbb{U},\ x[t] \in \mathbb{X},\ t \in \mathbb{Z}, \\ y[t] &= h(x[t], u[t], t), \qquad & y[t] \in \mathbb{Y}. \end{aligned} \tag{1.11}$$

State representations form a very rich class of representations, and are ideal for simulations. The function f is called the *next-state function*, and h is called the *output function.* Note that the next-state function $f : \mathbb{X} \times \mathbb{U} \times \mathbb{Z} \to \mathbb{X}$ indicates how the state $x[t]$ at any time t passes on to the state $x[t+1]$ at time $t+1$ under the influence of the input $u[t]$ at time t. So this function fixes the *dynamics* of the system. The output function does nothing more than determine the output $y[t]$ at time t as a function of the state and input at that same time.

Lemma 1.8.2. *The x in* (1.11) *is a state.*

Proof. See Exercise 1.18. ■

Example 1.8.3 (Parity checker). Consider an input space $\mathbb{U}$ consisting of two letters a and b, so $\mathbb{U} = \{a, b\}$, and consider the space $\mathbb{U}^*$ of all "words" that can be formed using the letters a and b. We want to construct a system that tells us whether the number of times the letter a occurs in a word is even or odd. Our output space $\mathbb{Y}$ consists of the two symbols E and O, for even and odd, respectively.

We introduce a state space with two elements x_o, x_e, and define the next-state function as follows:

$$\begin{aligned} f(x_o, b) &= x_o \\ f(x_o, a) &= x_e \\ f(x_e, b) &= x_e \\ f(x_e, a) &= x_o, \end{aligned}$$

while the output function depends only on x, and is given by

$$\begin{aligned} h(x_o) &= O \\ h(x_e) &= E. \end{aligned}$$

Note that neither f nor h depends explicitly on t. It is now easy to verify that if at time $t = 0$ we begin in the state $x[0] = x_e$, and starting at $t = 0$ we input a word consisting of the letters a and b, then the output at time $t \geq 0$ will indicate whether, at this time, the letter a has occurred an even or odd number of times. A convenient way to represent the system we have just defined is the description as a *finite automaton*; see Figure 1.14. □

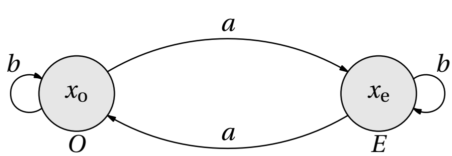


FIGURE 1.14: Parity checker.

Example 1.8.4 (String detector). The input set $\mathbb{U}$ in this example consists of all symbols on a keyboard. We want to design a system that detects the string "IWS".

The output set is $\mathbb{Y} = \{0,1\}$, and the output should have value 1 every time the string IWS occurs in the input string.

As state space, we take the set $\mathbb{X}$ consisting of three elements α, β, γ. The next-state function f is determined by the table below (the columns give the output, and the rows give the input, where R denotes any input value that is not I, W, or S.)

	I	W	S	R
α	β	α	α	α
β	β	γ	α	α
γ	β	α	α	α

Furthermore, we define the output function by setting $h(\gamma, \mathrm{S}) = 1$ and $h = 0$ elsewhere. We can summarize this using the finite automaton in Figure 1.15, where the three states are represented by circles, and the input/output corresponding to each transition arrow is written next to it. If we initialize the system in the state α (or equivalently, bring the system into the state α using the input R), then the system carries out the desired task and produces the output 1 every time it encounters the input string IWS. □

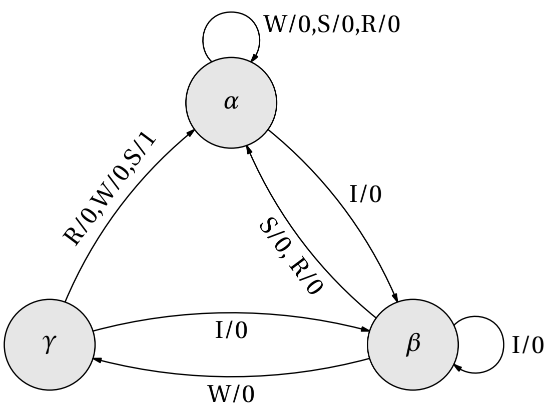


FIGURE 1.15: String detector.

Note that the time axis $\mathbb{T} = \mathbb{Z}$ in Examples 1.8.3 and 1.8.4 plays a somewhat different role than it does in Example 1.8.1. Indeed, in Examples 1.8.3 and 1.8.4, the times 1,2,3, etc. only indicate an *order*—time 1 comes before time 2, and time 2 comes before time 3—without referring to a physical time as in Example 1.8.1. Systems such as those in Examples 1.8.3 and 1.8.4 are therefore also called *discrete-event* systems (a series of events takes place at consecutive, but not previously specified, times). It is good to note that $\mathbb{U}, \mathbb{X}$, and $\mathbb{Y}$ in Examples 1.8.1, 1.8.3, and 1.8.4 are *finite* sets. They are therefore also known as *finite* (input/output) *automatons.*

Discrete-time systems of type (1.11) are also widely used in signal processing. Numerical processing of continuous-time processes in general requires a discretization at some point in time. Discrete-time systems also occur naturally in economical systems, if only because economical systems describe "laws of nature" and those are often discrete, such as, for example, only paying interest at regular intervals.

Example 1.8.5 (Economical model). A very simple model of a national economy is a follows. For every year t, we define the following quantities:

$$\begin{aligned} y[t] &= \text{national product (in the year } t\text{)},\\ c[t] &= \text{consumption},\\ i[t] &= \text{investments},\\ u[t] &= \text{government spending}. \end{aligned}$$

These quantities satisfy the balance equation

$$y[t] = c[t] + i[t] + u[t].$$

We assume that the consumption is a fixed part of the national product, so that $c[t] = my[t]$ for some m with $0 \leq m \leq 1$. This gives

$$y[t] = my[t] + i[t] + u[t]$$

or

$$i[t] = (1-m)y[t] - u[t].$$

We furthermore assume that the *growth* of the national product is proportional (with factor $r[t]$) to the investments. We then have

$$y[t+1] - y[t] = r[t]i[t] = r[t](1-m)y[t] - r[t]u[t]$$

or

$$y[t+1] = [1 + r[t](1-m)]\, y[t] - r[t]u[t]. \tag{1.12}$$

□

1.8.2 Nonlinear Continuous-Time Systems

For continuous-time systems too, an effective modeling method is to formulate states and indicate how they evolve over time The continuous-time analog of (1.11) is the continuous-time *state representation*

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), t), \quad u(t) \in \mathbb{R}^{n_u}, \quad x(t) \in \mathbb{R}^n, \quad t \in \mathbb{R}\\ y(t) &= h(x(t), u(t), t), \quad y(t) \in \mathbb{R}^{n_y}. \end{aligned} \tag{1.13}$$

(We usually denote the number of components of a vector u by n_u. For example, $w(t) \in \mathbb{R}^{n_w}$.) Solvability now becomes a problem, because for some functions f, the existence and uniqueness of solutions of the differential equation (1.13) is not easy, or simply does not hold. A sufficient condition for the existence and uniqueness of solutions of (1.13) for given $x(0) = x_0$ and u is that $f(x)$ is continuously differentiable in x and has only a finite number of discontinuities. We will not discuss the solvability.

Example 1.8.6 (Predator-prey model). This is a classic example from population dynamics. Let x_1 be the number of anchovies (=prey) and x_2 the number of salmon (=predator). Let, further, u_1 be the fraction of anchovies that is caught per unit of time, and let u_2 be the fraction of salmon that is caught per unit of time. The Lotka–Volterra differential equations for x_1 and x_2 are

$$\begin{aligned} \dot{x}_1 &= ax_1 - bx_1x_2 - u_1x_1 \\ \dot{x}_2 &= cx_1x_2 - dx_2 - u_2x_2, \end{aligned} \tag{1.14}$$

where $a, b, c,$ and d are positive constants. The term ax_1 comes from the natural increase in anchovies when there is no salmon or fishing. Conversely, $-dx_2$ is the decrease in the salmon population when there is no food (=anchovies). The terms $-bx_1x_2$ and cx_1x_2 are the result of salmon eating anchovies. It is not unreasonable that these terms are proportional to the product x_1x_2, because the number of times a salmon comes across a school of anchovies is proportional to the number of anchovies, x_1, so for the total population of salmon, the number of salmon-anchovy encounters is proportional to x_1x_2. As output, we can, for example, take the number of salmon,

$$y = x_2. \tag{1.15}$$

(Note that, in fact, the variables $u_1, u_2, x_1, x_2,$ and y are all ≥ 0.) □

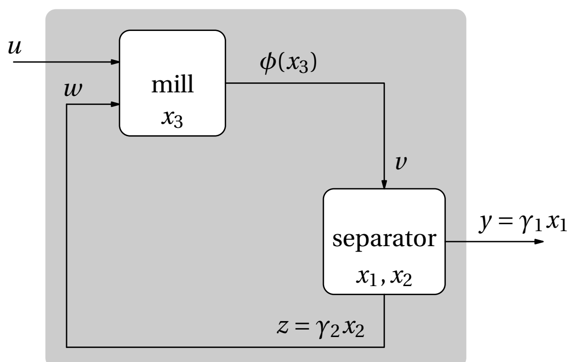


FIGURE 1.16: Configuration of the milling process.

Example 1.8.7 (Milling process—interconnection of subsystems). Consider an industrial milling process. Such a process can often be described as the interconnection of two subsystems, namely the "mill" that grinds the material and the "separator" that separates the ground material into fine and coarse particles. The mill is fed raw material, and produces a certain outflow of ground material. A simple model for the mill is

$$\textit{mill:} \quad \dot{x}_3 = u + w - \phi(x_3), \tag{1.16}$$

where x_3 is the amount of material in the mill, u is the influx of raw material, $\phi(x_3)$ is the outflow of ground material modeled using some function ϕ, and w is the recycling of coarse material from the separator; see Figure 1.16. The separator can be modeled by

$$\textit{separator:} \quad \begin{cases} \dot{x}_1 = -\gamma_1 x_1 + (1-\alpha)v \\ \dot{x}_2 = -\gamma_2 x_2 + \alpha v \\ \;\, y = \gamma_1 x_1 \\ \;\, z = \gamma_2 x_2, \end{cases} \tag{1.17}$$

where x_1 is the amount of fine particles, x_2 is the amount of coarse particles, $y = \gamma_1 x_1$ is the outflow of fine particles (the end product of the process), $z = \gamma_2 x_2$ is the outflow of coarse particles, v is the influx of material, and α is a separation constant ($0 < \alpha < 1$).

The two subsystems (1.16) and (1.17) (mill and separator) are linked through the interconnection

$$\textit{interconnection:} \quad \begin{cases} w = z, \\ v = \phi(x_3); \end{cases} \tag{1.18}$$

see Figure 1.16. The total system with input u and output y is now described by the state representation

$$\textit{total system:} \quad \begin{cases} \dot{x}_1 = -\gamma_1 x_1 + (1-\alpha)\phi(x_3) \\ \dot{x}_2 = -\gamma_2 x_2 + \alpha\phi(x_3) \\ \dot{x}_3 = \gamma_2 x_2 - \phi(x_3) + u \\ \;\, y = \gamma_1 x_1. \end{cases} \tag{1.19}$$

The structure of the system is hard to find in this description. □

1.9 Implicit Representations

From a modeling point of view, implicit representations are often the most natural ones. By implicit representations, we mean that the system equations only give *relations* between the variables, rather than maps from some of the variables to others. Furthermore, such a mathematical model often consists of a *collection* of relations between variables, rather than one relation (equation). For example, several conservation laws (of mass and energy) and laws of Newton, Hooke, Kirchoff, etc. can hold, and each of them leads to an equation. Together, the equations then form the mathematical model of the system.

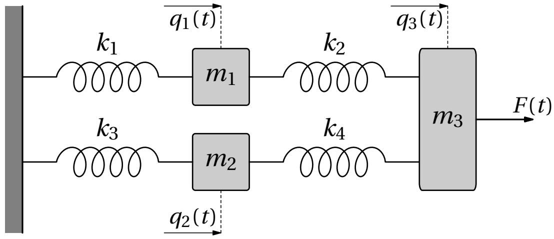


FIGURE 1.17: A spring-mass system.

Example 1.9.1 (Spring-mass system). Consider the spring-mass system of Figure 1.17. Two masses m_1 and m_2 are attached to a wall (on the left) by springs with spring constants k_1 and k_3 and to a mass m_3 (on the right) by springs with spring constants k_2 and k_4. We can exert a force F on the mass m_3. The positions of the three masses m_i with respect to their equilibrium points are denoted by q_i.

For each of the three masses, we can apply Newton's second law in combination with Hooke's law. This gives three motion equations that together describe the system

$$\begin{cases} m_1\ddot{q}_1 &= -k_1q_1 + k_2(q_3 - q_1), \\ m_2\ddot{q}_2 &= -k_3q_2 + k_4(q_3 - q_2), \\ m_3\ddot{q}_3 &= -k_2(q_3 - q_1) - k_4(q_3 - q_2) \ + \ F. \end{cases}$$

There are many variables. This system can be rewritten as a state representation (but we will not do this systematically). □

1.10 Exercises

1.1 Determine whether the systems with the following behavior are linear.

(a) $\{y : \mathbb{R} \to \mathbb{R} \mid \dot{y} = y\}$

(b) $\{(u, y) : \mathbb{R} \to (\mathbb{R}, \mathbb{R}) \mid \dot{y} = u\}$

(c) $\{(u, y) : \mathbb{R} \to (\mathbb{R}, \mathbb{R}) \mid y(t) = tu(t-1)\}$

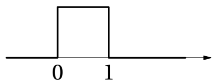
(d) $\{(u, y) : \mathbb{R} \to (\mathbb{R}, \mathbb{R}) \mid y^2(t) = u^2(t)\}$

(e) $\{(u, y) : \mathbb{Z} \to (\mathbb{R}, \mathbb{R}) \mid y[t] = u[-t]\}$

(f) $\{(u, y) : \mathbb{Z} \to (\mathbb{R}, \mathbb{R}) \mid y[t] = t\sin(u[t])\}$

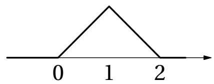
1.2 Let $u, y : \mathbb{R} \to \mathbb{R}$. Suppose that we have a linear time-invariant system $y = \mathscr{H}(u)$ and that we know that the response $y_0(t)$ to

$$u_0(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & \text{elsewhere} \end{cases}$$



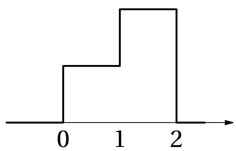
is equal to

$$y_0(t) = \begin{cases} 1 - |t-1| & 0 < t < 2 \\ 0 & \text{elsewhere} \end{cases}$$



Determine the response $y(t)$ to

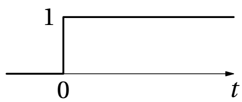
$u(t) =$



and make a sketch of this output.

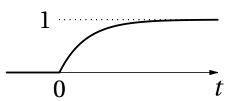
1.3 Let $u, y : \mathbb{R} \to \mathbb{R}$. Suppose that we have a linear time-invariant system $y = \mathscr{H}(u)$ and that we know that the output for the step function

$u_0(t) = \mathbb{1}(t)$



is equal to

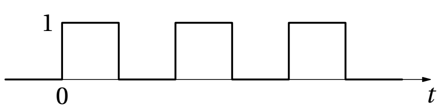
$$y_0(t) = \begin{cases} 0 & t < 0 \\ 1 - \mathrm{e}^{-2t} & t \geq 0 \end{cases}$$



Determine the output $y(t)$ for the square wave

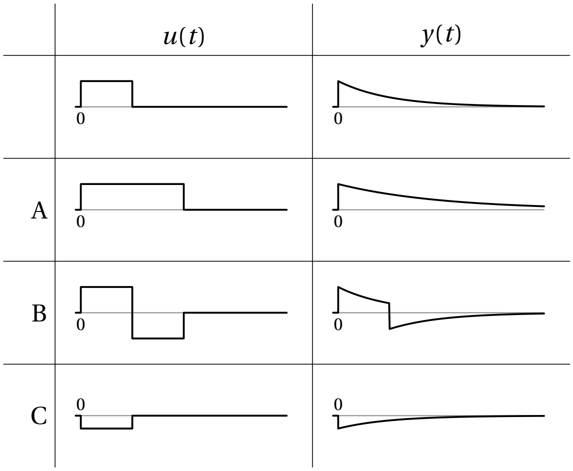
$$u(t) = u_0(t) - u_0(t-1) + u_0(t-2) - u_0(t-3) + \cdots$$

$=$



and make a sketch of this output.

1.4 Let $u, y : \mathbb{R} \to \mathbb{R}$, and assume that the system $y = \mathscr{H}(u)$ is linear and time invariant. The first row of the table below gives the input/output pair (u, y) of this system.



Which pairs (u, y) in the rows A, B, C correspond to the given pair? Indicate for each row A, B, C what you use: linearity, time invariance, or both.

1.5 Let $u, y : \mathbb{R} \to \mathbb{R}$. Set

$$y(t) = 2u(t) - 3u(t-1).$$

(a) Is this system linear?

(b) Is this system time invariant?

1.6 *Linearity, time invariance, and causality.* In this exercise, we have $t \in \mathbb{R}$ and $u, y : \mathbb{R} \to \mathbb{R}$.

(a) Is the system $y(t) = u^2(t)$ linear?

(b) Is the system $y(t) = u^2(t)$ time invariant?

(c) Is the system $y(t) = u^2(t)$ causal?

(d) Is the system $y(t) = t^2 u(t)$ linear?

(e) Is the system $y(t) = t^2 u(t)$ time invariant?

(f) Is the system $y(t) = t^2 u(t)$ causal?

(g) Is the system $y(t) = u(t^2)$ linear?

(h) Is the system $y(t) = u(t^2)$ time invariant?

(i) Is the system $y(t) = u(t^2)$ causal?

1.7 Consider the system $y(t) = u(\alpha t)$ with $u, y : \mathbb{R} \to \mathbb{R}$ and with α a given real number.

(a) Is the system linear?

(b) For which $\alpha \in \mathbb{R}$ is the system time invariant?

(c) For which α is the system causal?

1.8 Consider the continuous-time system

$$y(t) = u(t)u(t-2), \qquad t \in \mathbb{R}.$$

(a) Is the system linear?

(b) Is the system time invariant?

(c) Is the system causal?

1.9 *Moving average.* Let $u, y : \mathbb{R} \to \mathbb{R}$. Consider the moving average

$$y(t) = \frac{1}{2}\int_{t-1}^{t+1} u(\tau)\,\mathrm{d}\tau.$$

(a) Is the system linear?

(b) Is the system time invariant?

(c) Is the system causal?

1.10 *Linearity and time invariance of maps $y = \mathscr{H}(u)$.* Prove Lemma 1.6.2.

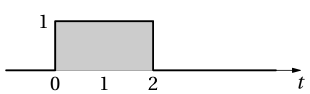
1.11 Let $u, y : \mathbb{R} \to \mathbb{R}$. Assume that the system $y = \mathscr{H}(u)$ is linear and time invariant. Show that the output for the derivative is the derivative of the output, that is: if $\mathscr{H}(u) = y$, then

$$\mathscr{H}(\dot{u}) = \dot{y}.$$

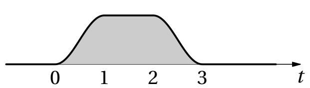
[Hint: First consider the output for $u(t+\epsilon) - u(t)$.]

1.12 Consider the drainage system of Example 1.6.5 and determine whether the system is linear and time invariant.

(a) Use a sketch to show that the output for



is equal to



(b) Sketch the output for $u(t) = \mathbb{1}(t)$.

(c) Determine the impulse response $h(t)$ and argue why this $h(t)$ is a realistic impulse response for the drainage system. [Hint: Consider Exercise 1.11.]

(d) Verify that $\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t = 1$ and explain with words why this is logical in this drainage system [Hint: conservation of mass].

(e) Determine the output $y(t)$ for the constant signal $u(t) = 5$ and explain why the answer makes sense.

1.13 Consider the linear system

$$y(t) = \int_{-\infty}^{\infty} K(t,s)\,u(s)\,\mathrm{d}s.$$

(a) Show that the system is time invariant if and only if $K(t,s) = K(t+\tau, s+\tau)$ for all t, s, τ.

(b) Next, show that in a time-invariant system, the function $K(t,s)$ depends only on $t-s$, so it can be written as $K(t,s) = h(t-s)$ for some function h.

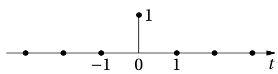
[Remark: It is therefore time invariant if and only if it is a *convolution*: $y(t) = \int_{-\infty}^{\infty} h(t-s)\,u(s)\,\mathrm{d}s$.]

(c) Show that the system is causal if and only if $K(t,s) = 0$ for all $t < s$

(d) Show that if $K(t,s) = h(t-s)$, then the system is causal if and only if $h(t) = 0$ for all $t < 0$.

1.14 *Linear time-invariant map.* Let $u, y : \mathbb{Z} \to \mathbb{R}$ (so a discrete-time system), and assume that the system $y = \mathscr{H}(u)$ is linear and time invariant. Define the impulse response $h : \mathbb{Z} \to \mathbb{R}$ as $h = \mathscr{H}(\Delta)$ with Δ the discrete pulse,

$$\Delta[t] = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}$$



(a) Let $u_0 \in \mathbb{R}$. Show that $\mathscr{H}(u_0\Delta) = u_0 h$.

(b) Sketch the graph of $\sigma^1\Delta$.

(c) Sketch the graph of $u_0\Delta + u_1\sigma^1\Delta$ for $u_0 = 1, u_1 = 1/2$.

(d) Show that $u = \sum_{k=-\infty}^{\infty} u[k]\sigma^k\Delta$.

(e) Show that $y = \mathscr{H}(u)$ can be written as $y = \sum_{k=-\infty}^{\infty} u[k]\sigma^k h$.

(f) Show that $y[t] = \sum_{k=-\infty}^{\infty} u[k]h[t-k]$.

(g) Show that $y[t] = \sum_{k=-\infty}^{\infty} u[t-k]h[k]$.

The last two forms are called (discrete-time) *convolutions*.

1.15 *Toeplitz matrix.* Let $u, y : \mathbb{Z} \to \mathbb{R}$. In the previous exercise, we showed that every linear time-invariant system $y = \mathscr{H}(u)$ is a convolution,

$$y[t] = \sum_{k=-\infty}^{\infty} u[k]h[t-k] = \sum_{k=-\infty}^{\infty} u[t-k]h[k].$$

Because the time axis $\mathbb{T} = \mathbb{Z}$ is countable, we can also represent the signals u, y as (infinitely long) column vectors. We denote $y[t]$ by y_t (the tth element of the column vector). Show that the convolution looks as follows:

$$\begin{bmatrix} \vdots \\ y_{-1} \\ y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & h_0 & h_{-1} & h_{-2} & h_{-3} & \ddots \\ \ddots & h_1 & h_0 & h_{-1} & h_{-2} & \ddots \\ \ddots & h_2 & h_1 & h_0 & h_{-1} & \ddots \\ \ddots & h_3 & h_2 & h_1 & h_0 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ u_{-1} \\ u_0 \\ u_1 \\ u_2 \\ \vdots \end{bmatrix}.$$

All descending diagonals of this (infinite) matrix are constant. Such matrices are called *Toeplitz* matrices.

Under which conditions on h_t is the system causal?

1.16 *State for a discrete-time system.* Determine a state for $y[t] = u[t] + u[t-1]$ by writing it in the form (1.11).

1.17 *State for a discrete-time system.* Determine a state $x : \mathbb{Z} \to \mathbb{X}$ for the discrete-time system $y[t] = u[t-100],\ t \in \mathbb{Z}$, and write the system in the state representation (1.11).

1.18 *State for a discrete-time system.* Show that in the system (1.11), the variable $x[t]$ has the properties of a state.

1.19 *Finite-dimensional system in continuous time.* Consider the system

$$y(t) = \int_{-\infty}^{t} u^2(\tau)\,\mathrm{d}\tau.$$

(a) Determine a state for the system.

(b) Write the system in the form (1.13).

1.20 *Finite-dimensional system.* Show that the discrete-time system

$$y[t] = \sum_{k=0}^{\infty} \frac{1}{2^k} u[t-k] \qquad (t \in \mathbb{Z},\ u[t] \in \mathbb{R})$$

is finite dimensional.

1.21 Let the input $u[t]$ be the average temperature on day $t \in \mathbb{N}$ and the output $y[t]$ be the number of times up to and including day t that the average temperature has risen ($u[t] > u[t-1]$). Construct a finite-dimensional state for this system.

1.22 Determine a state for the system in Exercise 1.8.

1.23 Let $u, y : \mathbb{R} \to \mathbb{R}$. Determine a state for the system $y(t) = \frac{1}{2}\int_{t-2}^{t} u(\tau)\,\mathrm{d}\tau$.

1.24 *Predator-prey model as interconnection.* In Example 1.8.7 about milling, the system was made up of subsystems that were later linked. This is common for complex systems. It may also be more transparent for the predator-prey model of Example 1.8.6. In analogy to Figure 1.16, represent the predator-prey model as an interconnection of two subsystems, with one system for the predator dynamics and one subsystem for the prey dynamics.

[Remark: The advantages of such a modular representation are, among others, that it clarifies the structural connections and that it makes it easier to adapt the model of, say, the predator without having to rewrite the whole system.]

Tougher Exercises

1.25 *State representations.* We are given matrices $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times n_u}, C \in \mathbb{R}^{n_y\times n}, D \in \mathbb{R}^{n_y\times n_u}$ for certain indices $n, n_u, n_y \in \mathbb{N}$. Consider the system with behavior

$$\mathfrak{B} = \left\{ \begin{bmatrix} u \\ y \end{bmatrix} : \mathbb{R} \to \begin{bmatrix} \mathbb{R}^{n_u} \\ \mathbb{R}^{n_y} \end{bmatrix} \;\middle|\; \exists x : \mathbb{R} \to \mathbb{R}^n \text{ such that } \begin{bmatrix} \dot{x}(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \right\}.$$

(a) Is the system linear?

(b) Is the system time invariant?

1.26 Is the discrete-time system

$$y[t] = \sum_{m=1}^{\infty} \frac{1}{m} u[t-m] \quad (t \in \mathbb{Z},\ u[t] \in \mathbb{R})$$

infinite dimensional? (This exercise is very complicated!)

Chapter 2

State Representations

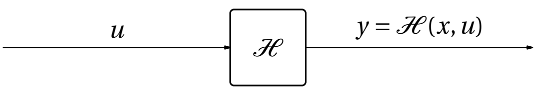


FIGURE 2.1: A system with a state.

In this chapter, we consider state representations. These describe systems where the output y depends not only on the input u, but also on other variables x_j that together form a a state of the system. Figure 2.1 illustrates this. This chapter focuses on the dynamical properties of the state x. For the moment, we will not pay much attention to the roles of the input and output, but this will change in later chapters.

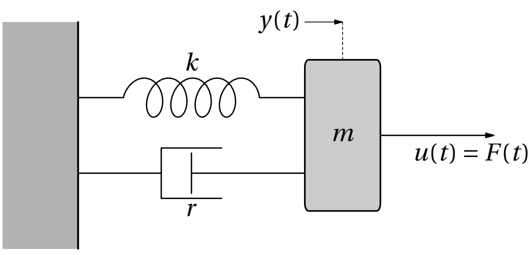


FIGURE 2.2: Mass-spring-damper system.

2.1 State Representations

Example 2.1.1 (Mass-spring-damper system). Suppose that we have a mass m that can move along a horizontal straight line; see Figure 2.2. The mass is attached to a wall by a spring with spring constant k, and undergoes friction that is proportional to the velocity of the mass. The mass also has an outward force $F(t)$ acting on it.

Let $q(t)$ be the distance from the mass to its equilibrium point (where the spring does not exert any force). We moreover assume that the external force $F(t)$ can change arbitrarily, so as input, we take $u(t) := F(t)$. As output, we take the position of the mass, $y(t) = q(t)$. The equation of motion is

$$m\ddot{q}(t) + r\dot{q}(t) + kq(t) = u(t). \tag{2.1}$$

Suppose that, from a certain time $t = t_0$ on, we experiment with $u(t) = F(t)$. It may be intuitively clear that if at $t = t_0$, we know both the position $q(t_0)$ and the velocity $\dot{q}(t_0)$ of the mass, then the position and velocity are fully determined for given $u(t),\ t \geq t_0$. For this reason, we define the two signals $x_1(t)$ and $x_2(t)$ as

$$x_1(t) = q(t), \qquad x_2(t) = \dot{q}(t).$$

The second-order differential equation (2.1) can now be rewritten as two first-order differential equations,

$$\begin{aligned}\dot{x}_1(t) &= x_2(t),\\ \dot{x}_2(t) &= -\frac{k}{m}x_1(t) - \frac{r}{m}x_2(t) + \frac{1}{m}u(t).\end{aligned}$$

In matrix notation, this is

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix} = \begin{bmatrix}0 & 1\\ -\frac{k}{m} & -\frac{r}{m}\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} + \begin{bmatrix}0\\ \frac{1}{m}\end{bmatrix}u(t). \tag{2.2}$$

As output y, we take the position, $y = q$. In matrix notation, this is

$$y(t) = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}.$$

In summary: the mass-spring-damper system has a state representation of the form

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t),\\ y(t) &= Cx(t),\end{aligned} \tag{2.3}$$

where $x = \left[\begin{smallmatrix}x_1\\ x_2\end{smallmatrix}\right]$ is the state vector and $A = \left[\begin{smallmatrix}0 & 1\\ -k/m & -r/m\end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix}0\\ 1/m\end{smallmatrix}\right]$, and $C = \left[\begin{smallmatrix}1 & 0\end{smallmatrix}\right]$. □

In this chapter, we consider systems with several inputs,

$$u(t) = \begin{bmatrix}u_1(t)\\ \vdots\\ u_{n_u}(t)\end{bmatrix},$$

and several outputs,

$$y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_{n_y}(t) \end{bmatrix}$$

that are related through a differential equation of the form (2.3). In the example above, the state consists of two components $x_1(t)$ and $x_2(t)$. In general, x will consist of several components; we denote their number by n,

$$x(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

This n-vector is called the *state* (it has the characteristics of a state, as we will show later on). The first equation in (2.3),

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

is called the *state equation*. Now, A is an $n \times n$ matrix and B is an $n \times n_u$ matrix. The second equation in (2.3) is $y(t) = Cx(t)$. This is called the *output equation* of the system. We will also consider an output equation that is a bit more general,

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

where C is an $n_y \times n$ matrix and D is an $n_y \times n_u$ matrix. Note that the output equation does not contain any derivatives. In short, the type of representation that we will consider in this chapter is

$$\begin{cases} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t). \end{cases} \tag{2.4}$$

To simplify the notation, we often leave out the time and write

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

Example 2.1.2 (Simulation using state representations). Why would people want to consider state representations? One of the reasons is that they are well suited to simulation. To simulate (2.4), we discretize the state equation $\dot{x} = Ax + Bu$. Let $h > 0$ be a (small) step size; then

$$\begin{aligned} x(t+h) &\approx x(t) + h\dot{x}(t) \\ &= x(t) + h(Ax(t) + Bu(t)). \end{aligned}$$

So the simulated signals $x(kh)$ and $u(kh)$, $k \in \mathbb{N}$, satisfy

$$x((k+1)h) \approx x(kh) + h[Ax(kh) + Bu(kh)].$$

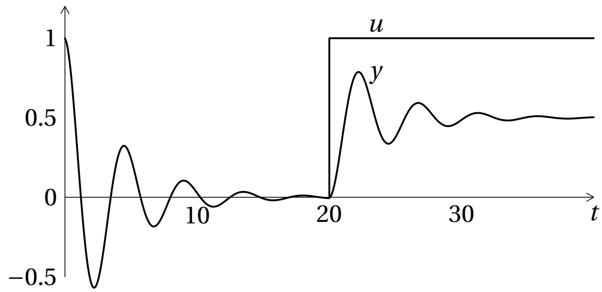


FIGURE 2.3: Simulation of the mechanical system of Example 2.1.1, with $m = 1, k = 2, r = 1/2$ and $q(0) = 1,\ \dot{q}(0) = 0$.

This is an explicit recurrence relation in $x(kh)$, $k \in \mathbb{N}$, in a form that can be programmed directly provided $u(kh)$ and an initial state $x(0)$ are known. This method is called the *Euler method*. Simulation software packages use more advanced methods.

Figure 2.3 shows the simulated output $y(t)$ of the system of Example (2.1.1) produced by the MATLAB command `lsim` (see Appendix A.11). □

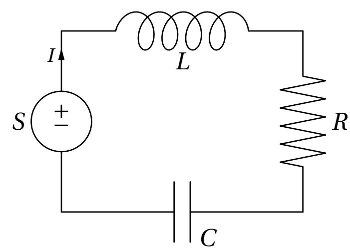


FIGURE 2.4: *RLC* network.

Example 2.1.3 (*RLC* electrical circuit). Consider the electrical circuit of Figure 2.4. This consists of a capacitor C, an inductor (coil) L, a resistor R, and a voltage source S. Let V_C, V_L, V_R, and V be the voltages across C, L, R, and S, respectively, and let I_C, I_L, I_R, and I be the currents through these elements. Kirchoff's voltage and current laws lead to the following balance equations:

$$V = V_L + V_C + V_R, \qquad I_L = I_R = I_C = I. \tag{2.5}$$

The constitutive equations of a linear capacitor, inductor, and resistor are given by

$$\begin{cases} V_C = \frac{1}{C} q \\ \dot{q} = I_C \end{cases}, \qquad \begin{cases} I_L = \frac{1}{L} \phi \\ \dot{\phi} = V_L \end{cases}, \qquad V_R = R I_R \tag{2.6}$$

for constants C, L, and R. Here q is the charge on the capacitor and ϕ is the magnetic flux of the inductor. By choosing $x(t) = \left[\begin{smallmatrix} q(t) \\ \phi(t) \end{smallmatrix}\right]$ for the state vector $x(t)$, we can rewrite (2.5) and (2.6) as

$$\begin{bmatrix} \dot{q}(t) \\ \dot{\phi}(t) \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{L} \\ -\frac{1}{C} & -\frac{R}{L} \end{bmatrix} \begin{bmatrix} q(t) \\ \phi(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} V(t). \tag{2.7}$$

This is a differential equation with input $u(t) := V(t)$, the voltage across the voltage source. As output $y(t)$, we can for example take the current I through the voltage source, in which case

$$y(t) = \begin{bmatrix} 0 & \frac{1}{L} \end{bmatrix} \begin{bmatrix} q(t) \\ \phi(t) \end{bmatrix}, \tag{2.8}$$

or the charge on the capacitor, in which case

$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} q(t) \\ \phi(t) \end{bmatrix}. \tag{2.9}$$

□

Note that the state representation (2.2) for the mass-spring-damper system in Example 2.1.1 closely resembles (2.7) for the *RLC* circuit. This becomes even clearer by taking for the state vector $x(t)$ of the mass-spring-damper system the vector $\left[\begin{smallmatrix} q(t) \\ p(t) \end{smallmatrix}\right]$, with $p(t) := m\dot{q}(t)$ the *impulse* of the mass. In that case, instead of (2.2), we obtain

$$\begin{bmatrix} \dot{q}(t) \\ \dot{p}(t) \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{m} \\ -k & -\frac{r}{m} \end{bmatrix} \begin{bmatrix} q(t) \\ p(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t). \tag{2.10}$$

These equations in q, p, and u are *exactly* equal to equation (2.7) in q, ϕ, and V for $m = L$, $k = \frac{1}{C}$, and $r = R$. We see that two physically very different systems can be equivalent *mathematically*. This mathematical equivalence of physically distinct systems is one of the reasons for the existence of mathematical systems theory, which, after all, studies *general* mathematical systems, which can be applied to systems occurring in *a variety of* fields of science, in this case to both mechanical networks and electrical circuits.

2.2 Solutions of State Equations

In this section, we determine the general solution of the state equations

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

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n_u}$, $C \in \mathbb{R}^{n_y\times n}$, $D \in \mathbb{R}^{n_y\times n_u}$, and $x(t)$ an n-dimensional signal. We assume that $u(t)$ is given, and that we are looking for the state $x(t)$ and output $y(t)$. As soon as we have $x(t)$, we also have the output $y(t)$, because $y(t) = Cx(t) + Du(t)$. In other words, the only difficulty lies in the state equation $\dot{x}(t) = Ax(t) + Bu(t)$.

Example 2.2.1 (Variation of constants). In this example, we determine $x(t)$ in the case $n = n_u = 1$. That is, A and B are scalars, $A = a$, $B = b$, and the state has only one component, $x(t) = x_1(t)$. Consider

$$\dot{x}_1(t) = ax_1(t) + bu(t). \tag{2.11}$$

For $u(t) \equiv 0$, this equation reduces to the homogeneous equation $\dot{x}_1(t) = ax_1(t)$. The solution of this homogeneous equation is known to be

$$x_1(t) = z\mathrm{e}^{at}, \quad z \in \mathbb{C}$$

for an arbitrary constant z. The method of *variation of constants* consists in writing a candidate solution $x_1(t)$ of (2.11) as

$$x_1(t) = z(t)\mathrm{e}^{at},$$

where $z(t)$ is now a function of t. Every $x_1(t)$ can be written as $x_1(t) = z(t)\mathrm{e}^{at}$ because e^{at} is invertible. We have

$$\begin{aligned} \dot{x}_1(t) &= ax_1(t) + bu(t) \\ &\iff \dot{z}(t)\mathrm{e}^{at} + az(t)\mathrm{e}^{at} = az(t)\mathrm{e}^{at} + bu(t) \\ &\iff \dot{z}(t)\mathrm{e}^{at} = bu(t) \\ &\iff \dot{z}(t) = \mathrm{e}^{-at}bu(t) \\ &\iff z(t) = z_0 + \int_0^t \mathrm{e}^{-a\tau}bu(\tau)\,\mathrm{d}\tau \qquad (z_0 \in \mathbb{C}). \end{aligned}$$

The general solution $x_1(t) = z(t)\mathrm{e}^{at}$ is therefore

$$\begin{aligned} x_1(t) &= \mathrm{e}^{at}\Big(z_0 + \int_0^t \mathrm{e}^{-a\tau}bu(\tau)\,\mathrm{d}\tau\Big) \\ &= \mathrm{e}^{at}z_0 + \int_0^t \mathrm{e}^{a(t-\tau)}bu(\tau)\,\mathrm{d}\tau. \end{aligned} \tag{2.12}$$

□

We will see that the method of variation of constants also works for state with more than one component ($n > 1$). In the example above, we used the exponential function e^{at}. In the general n-dimensional case, its role is taken over by the *matrix exponential* e^{At}, with $A \in \mathbb{R}^{n\times n}$. In analogy with the Taylor series expansion of e^a,

$$\mathrm{e}^a = \sum_{k=0}^{\infty} \frac{1}{k!}a^k = 1 + a + \frac{1}{2!}a^2 + \frac{1}{3!}a^3 + \cdots$$

we define the following.

Definition 2.2.2 (Matrix exponential). The *matrix exponential* e^A of a matrix $A \in \mathbb{R}^{n\times n}$ is defined as

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = I + A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \cdots . \tag{2.13}$$

□

For every square matrix A, this series is convergent.

Lemma 2.2.3 (Matrix exponential). *Let $A, F \in \mathbb{R}^{n\times n}$. Four characteristic properties of the matrix exponential are:*

1. $e^0 = I$ *for the zero matrix* $0 \in \mathbb{R}^{n\times n}$.
2. *If* $AF = FA$, *then* $e^A e^F = e^{A+F}$.
3. e^A *is invertible and* $(e^A)^{-1} = e^{-A}$.
4. *Let* $t \in \mathbb{R}$; *then* $\frac{d}{dt} e^{At} = A e^{At} = e^{At} A$.

Proof.

1. Follows immediately from the definition (formula (2.13)).
2. Written out, the product $e^A e^F$ is

$$\begin{aligned} e^A e^F &= \Big(\sum_{k=0}^{\infty} \frac{1}{k!} A^k\Big)\Big(\sum_{m=0}^{\infty} \frac{1}{m!} F^m\Big) \\ &= \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \frac{1}{k!\,m!} A^k F^m. \end{aligned}$$

   We see that the coefficient of $A^k F^m$ is equal to $\frac{1}{k!m!}$. Written out, e^{A+F} gives

$$\begin{aligned} e^{A+F} &= \sum_{n=0}^{\infty} \frac{1}{n!} (A+F)^n \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} A^k F^{n-k} \right) \\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{1}{k!(n-k)!} A^k F^{n-k}. \end{aligned}$$

   Here too, the coefficient of the factor $A^k F^m$ is equal to $\frac{1}{k!m!}$. So $e^A e^F = e^{A+F}$. (So, where did we use that $AF = FA$? See Exercise 2.6.)

3. Apply part 2 with $F = -A$.

4. On every time interval $[a, b]$, the series of the derivative $\sum_k \frac{\mathrm{d}}{\mathrm{d}t}(\frac{1}{k!}A^k t^k)$ can be shown to converge uniformly. Hence the derivative of the series is the series of the derivatives (that is, summation and differentiation may be interchanged):

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At} &= \frac{\mathrm{d}}{\mathrm{d}t}\Big(\sum_{k=0}^{\infty}\frac{1}{k!}A^k t^k\Big) \\
&= \sum_{k=0}^{\infty}\frac{1}{k!}A^k\frac{\mathrm{d}}{\mathrm{d}t}t^k \\
&= \sum_{k=1}^{\infty}\frac{1}{(k-1)!}A^k t^{k-1} \\
&= \{\text{take } m := k-1\} = \sum_{m=0}^{\infty}\frac{1}{m!}A^{m+1}t^m \\
&= \mathrm{e}^{At}A.
\end{aligned}$$

Since A commutes with itself, we also have $\mathrm{e}^{At}A = \sum_{n=0}^{\infty}\frac{1}{n!}t^n A^{n+1} = A\mathrm{e}^{At}$.

■

With these properties of the matrix exponential, we can redo the method of variation of constants of Example 2.2.1. We write the candidtate solution $x(t)$ of $\dot{x}(t) = Ax(t) + Bu(t)$ as

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

with $z(t) \in \mathbb{R}^n$ a yet to be determined function of time. Every $x(t)$ can be written this way, because e^{At} is invertible. We have

$$\begin{aligned}
&\dot{x}(t) = Ax(t) + Bu(t) \\
\iff\ & A\mathrm{e}^{At}z(t) + \mathrm{e}^{At}\dot{z}(t) = A\mathrm{e}^{At}z(t) + Bu(t) \\
\iff\ & \mathrm{e}^{At}\dot{z}(t) = Bu(t) \\
\iff\ & \dot{z}(t) = \mathrm{e}^{-At}Bu(t) \\
\iff\ & z(t) = z(t_0) + \int_{t_0}^{t}\mathrm{e}^{-A\tau}Bu(\tau)\,\mathrm{d}\tau \qquad (z(t_0) \in \mathbb{R}^n) \\
\iff\ & x(t) = \mathrm{e}^{At}\Big(\mathrm{e}^{-At_0}x(t_0) + \int_{t_0}^{t}\mathrm{e}^{-A\tau}Bu(\tau)\,\mathrm{d}\tau\Big) \\
\iff\ & x(t) = \mathrm{e}^{A(t-t_0)}x(t_0) + \int_{t_0}^{t}\mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau.
\end{aligned}$$

In summary, we have the following result.

Theorem 2.2.4 (Solution of the state equations). *Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n_u}$. The state equation $\dot{x}(t) = Ax(t) + Bu(t)$ with given initial condition $x(t_0) = x_0 \in \mathbb{R}^n$*

and input $u(t)$ has a unique solution $x(t)$, given by

$$x(t) = \mathrm{e}^{A(t-t_0)} x_0 + \int_{t_0}^{t} \mathrm{e}^{A(t-\tau)} B u(\tau)\,\mathrm{d}\tau. \tag{2.14}$$

The output $y(t) = Cx(t) + Du(t)$ follows uniquely, and is

$$y(t) = C\mathrm{e}^{A(t-t_0)} x_0 + \int_{t_0}^{t} C\mathrm{e}^{A(t-\tau)} B u(\tau)\,\mathrm{d}\tau + Du(t). \tag{2.15}$$

□

The integrals exist for every locally bounded input.[1] Note that (2.15) is of the form

$$y(t) = \mathscr{H}(x(t_0), u(\tau)\big|_{\tau\in[t_0,t]}) \quad \forall\, t_0, t \geq t_0.$$

The variable x is therefore a state, and the system is finite dimensional (see page 16).

If the input is the zero function, $u(t) = 0$, then the dynamics are fully determined by the initial state. The state then satisfies $\dot{x}(t) = Ax(t)$, and according to (2.14), its solution is

$$x(t) = \mathrm{e}^{A(t-t_0)} x(t_0).$$

2.2.1 The Entries of e^{At}

To better understand the behavior of the system, it is useful to write the matrix exponential e^{At} in a more explicit form. Only in a few cases can this be done directly using the power series (2.13). An example where the power series is handy, is a diagonal matrix. If Λ is diagonal,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}, \tag{2.16}$$

then the corresponding matrix exponential $\mathrm{e}^{\Lambda t}$ is nothing else than the diagonal matrix of the exponentials,

$$\mathrm{e}^{\Lambda t} = \begin{bmatrix} \mathrm{e}^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathrm{e}^{\lambda_n t} \end{bmatrix}.$$

[1] Locally bounded means bounded on every finite time interval. There is a subtlety: we have until now tacitly assumed that $\dot{x}$ exists, but if u is bounded but discontinuous, then the x defined in (2.14) is well defined and continuous, but not differentiable in the classic sense. What does $\dot{x} = Ax + Bu$ mean in this case? Appendix A.2 gives more details on this.

This follows from (2.13),

$$
\begin{aligned}
\mathrm{e}^{\Lambda t} &= \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} + \begin{bmatrix} \lambda_1 t & & 0 \\ & \ddots & \\ 0 & & \lambda_n t \end{bmatrix} + \frac{1}{2!} \begin{bmatrix} \lambda_1^2 t^2 & & 0 \\ & \ddots & \\ 0 & & \lambda_n^2 t^2 \end{bmatrix} + \cdots \\
&= \begin{bmatrix} 1 + \lambda_1 t + \frac{1}{2!}\lambda_1^2 t^2 + \cdots & & 0 \\ & \ddots & \\ 0 & & 1 + \lambda_n t + \frac{1}{2!}\lambda_n^2 t^2 + \cdots \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{e}^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & \mathrm{e}^{\lambda_n t} \end{bmatrix}.
\end{aligned}
$$

This result serves as a basis for the determination of the elements of e^{At} for general A. If, for example, A has an *eigendecomposition*, that is, if there exists an invertible matrix T such that

$$A = T\Lambda T^{-1}$$

for some diagonal matrix Λ, then

$$\mathrm{e}^{At} = T\mathrm{e}^{\Lambda t}T^{-1}$$

and because the elements of $\mathrm{e}^{\Lambda t}$ are easy to compute, the elements of e^{At} follow. We state this result for general matrices $\Lambda = M$.

Lemma 2.2.5. *Let M and T be square matrices of the same dimension, and assume that T is invertible. Then*

$$\mathrm{e}^{TMT^{-1}} = T\mathrm{e}^{M}T^{-1}.$$

Proof.

$$
\begin{aligned}
\mathrm{e}^{TMT^{-1}} &= I + TMT^{-1} + \frac{1}{2!}(TMT^{-1})(TMT^{-1}) + \cdots \\
&= I + TMT^{-1} + \frac{1}{2!}TM^2T^{-1} + \cdots \\
&= T\big(I + M + \frac{1}{2!}M^2 + \cdots\big)T^{-1} \\
&= T\mathrm{e}^{M}T^{-1}.
\end{aligned}
$$

■

The decomposition $A = T\Lambda T^{-1}$, with Λ diagonal, is called an eigendecomposition because the columns of T are then eigenvectors, and the corresponding diagonal elements of Λ are the eigenvalues. Indeed, if we multiply $A = T\Lambda T^{-1}$ on the right by T, we get

$$AT = T \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix},$$

hence the kth column T_k of T satisfies $AT_k = \lambda_k T_k$. This relation allows us to determine T and Λ (provided that they exist; see further on), and we conclude that the computation of e^{At} can be reduced to the computation of the eigenvalues $\lambda_1, \dots, \lambda_n$ of A and the corresponding linearly independent eigenvectors $v_1, \dots, v_n$. The eigenvalues λ_i of A are the zeros of the *characteristic polynomial* $\chi_A(\lambda)$ of A, defined by

$$\chi_A(\lambda) = \det(\lambda I - A).$$

By the fundamental theorem of algebra, the characteristic polynomial has exactly n zeros $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ (of which some can coincide, in which case those zeros are called *multiple*).

Example 2.2.6 (Matrix exponential for a mass-spring-damper system). Consider the mass-spring-damper system from Example 2.1.1 and first take $k = 0$ (no spring). Then

$$A = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{r}{m} \end{bmatrix}. \tag{2.17}$$

The characteristic polynomial is

$$\chi_A(\lambda) = \det(\begin{bmatrix} \lambda & -1 \\ 0 & \lambda + \frac{r}{m} \end{bmatrix}) = \lambda(\lambda + \tfrac{r}{m}).$$

This has zeros $\lambda_1 = 0$, $\lambda_2 = -\frac{r}{m}$. The corresponding eigenvectors v_1 and v_2 can be determined using the equations

$$\begin{aligned} 0 &= (\lambda_1 I - A) v_1 = \begin{bmatrix} 0 & -1 \\ 0 & \frac{r}{m} \end{bmatrix} v_1, \\ 0 &= (\lambda_2 I - A) v_2 = \begin{bmatrix} -\frac{r}{m} & -1 \\ 0 & 0 \end{bmatrix} v_2. \end{aligned}$$

This gives, for example,

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} 1 \\ -\frac{r}{m} \end{bmatrix}.$$

We therefore take T and Λ to be

$$\begin{aligned} T &= \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -\frac{r}{m} \end{bmatrix}, \\ \Lambda &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{r}{m} \end{bmatrix}, \end{aligned}$$

and therefore

$$\begin{aligned}
e^{At} &= T e^{\Lambda t} T^{-1} \\
&= \begin{bmatrix} 1 & 1 \\ 0 & -\frac{r}{m} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{-\frac{r}{m}t} \end{bmatrix} \begin{bmatrix} 1 & \frac{m}{r} \\ 0 & -\frac{m}{r} \end{bmatrix} \\
&= \begin{bmatrix} 1 & \frac{m}{r}(1 - e^{-\frac{r}{m}t}) \\ 0 & e^{-\frac{r}{m}t} \end{bmatrix}.
\end{aligned}$$

If we do not exert any force u on the mass, then the state $x = \left[\begin{smallmatrix} q \\ \dot{q} \end{smallmatrix}\right]$ is given by $x(t) = e^{At}x(0)$, that is,

$$\begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} = \begin{bmatrix} 1 & \frac{m}{r}(1 - e^{-\frac{r}{m}t}) \\ 0 & e^{-\frac{r}{m}t} \end{bmatrix} \begin{bmatrix} q(0) \\ \dot{q}(0) \end{bmatrix}.$$

This seems quite reasonable, because if we write it out, we get

$$\begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} = \begin{bmatrix} q(0) + \frac{m}{r}(1 - e^{-\frac{r}{m}t})\dot{q}(0) \\ e^{-\frac{r}{m}t}\dot{q}(0) \end{bmatrix},$$

and we see that the initial velocity $\dot{q}(0)$ decreases exponentially. In the end (in the limit $t = \infty$), the velocity becomes zero, and the mass comes to rest at $q(\infty) = q(0) + \frac{m}{r}\dot{q}(0)$. We now also see that heavier masses take longer to come to a standstill, and that for them, the final position lies further away from the initial position.

Next, take $k \neq 0$ and $r = 0$ (no damper). The A-matrix is then equal to

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{bmatrix}.$$

This has characteristic polynomial $\chi_A(\lambda) = \lambda^2 + \frac{k}{m}$, which has imaginary zeros, $\lambda_1 = i\omega,\ \lambda_2 = -i\omega$ with $\omega = \sqrt{k/m}$. The corresponding eigenvectors have complex components (verify this yourself),

$$v_1 = \begin{bmatrix} -i \\ \omega \end{bmatrix}, \quad v_2 = \begin{bmatrix} i \\ \omega \end{bmatrix},$$

but the matrix exponential does not:

$$\begin{aligned}
e^{At} &= \begin{bmatrix} -i & i \\ \omega & \omega \end{bmatrix} \begin{bmatrix} e^{i\omega t} & 0 \\ 0 & e^{-i\omega t} \end{bmatrix} \begin{bmatrix} -i & i \\ \omega & \omega \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \dfrac{e^{i\omega t} + e^{-i\omega t}}{2} & \dfrac{e^{i\omega t} - e^{-i\omega t}}{2i\omega} \\ -\omega\dfrac{e^{i\omega t} - e^{-i\omega t}}{2i} & \dfrac{e^{i\omega t} + e^{-i\omega t}}{2} \end{bmatrix} \\
&= \begin{bmatrix} \cos(\omega t) & \frac{\sin(\omega t)}{\omega} \\ -\omega\sin(\omega t) & \cos(\omega t) \end{bmatrix}.
\end{aligned}$$

We have used Euler's formula, $\mathrm{e}^{\pm \mathrm{i}\omega t} = \cos(\omega t) \pm \mathrm{i}\sin(\omega t)$. If we do not exert any force u on the mass, the state $x = \left[\begin{smallmatrix} q \\ \dot{q} \end{smallmatrix}\right]$ is given by $x(t) = \mathrm{e}^{At}x(0)$, which in this case means that

$$\begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega t) & \frac{\sin(\omega t)}{\omega} \\ -\omega\sin(\omega t) & \cos(\omega t) \end{bmatrix} \begin{bmatrix} q(0) \\ \dot{q}(0) \end{bmatrix}.$$

Starting with $q(0) = 1, \dot{q}(0) = 0$, we for example obtain the motion

$$\begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega t) \\ -\omega\sin(\omega t) \end{bmatrix}.$$

The mass keeps swinging back and forth about zero. The period of the motion is $2\pi/\omega = 2\pi\sqrt{m/k}$. Large masses m (with respect to k) need more time to complete a period. □

If $\det(\lambda I - A)$ has multiple zeros, then it may happen that A does not have n linearly independent eigenvectors. The matrix A is then not diagonalizable, and the procedure described above breaks down. If the matrix A is not diagonalizable, we can use the Jordan normal form of A. Every square matrix has a *Jordan normal form*. This means that every square matrix A can be written as

$$A = TJT^{-1}$$

with T an invertible matrix and J the Jordan normal form of A, which is a matrix of a block-diagonal structure

$$J = \begin{bmatrix} J_1 & 0 & \cdots \\ 0 & J_2 & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix}$$

with J_k $(k = 1, 2, \ldots)$ the so-called *Jordan blocks* of the form

$$J_k = \begin{bmatrix} \lambda_k & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix}.$$

A special case is that where all J_k are of dimension 1×1, that is, $J_k = \lambda_k \in \mathbb{C}$. In this case, the decomposition is an eigendecomposition. In general, however, the Jordan blocks J_k have a higher dimension, and then ones appear on the super-diagonal. For every Jordan block J_k, the matrix exponential can be determined using the definition (2.13). This gives (see Exercise 2.8) the upper triangular matrix

$$\mathrm{e}^{J_k t} = \mathrm{e}^{\lambda_k t} \begin{bmatrix} 1 & t & \frac{1}{2!}t^2 & \cdots & \frac{1}{(m-1)!}t^{m-1} \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \frac{1}{2!}t^2 \\ \vdots & \ddots & \ddots & \ddots & t \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}. \tag{2.18}$$

It now follows that e^{At} equals

$$\mathrm{e}^{At} = T \begin{bmatrix} \mathrm{e}^{J_1 t} & 0 & \cdots \\ 0 & \mathrm{e}^{J_2 t} & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix} T^{-1}.$$

Note that the elements of $\mathrm{e}^{J_k t}$ are of the form $\alpha t^m \mathrm{e}^{\lambda t}$ for some $m \in \mathbb{N}$ and $\lambda, \alpha \in \mathbb{C}$, and because *every* square matrix A has a Jordan normal form, all elements of *every* matrix exponential are linear combinations of functions of the form $t^m \mathrm{e}^{\lambda t}$. For complex λ, the function $\mathrm{e}^{\lambda t}$ is usually split into a real part and an imaginary part,

$$\lambda = \mu + \mathrm{i}\omega \implies \mathrm{e}^{\lambda t} = \mathrm{e}^{\mu t}\mathrm{e}^{\mathrm{i}\omega t} = \mathrm{e}^{\mu t}\big(\cos(\omega t) + \mathrm{i}\sin(\omega t)\big).$$

Hence, as we already saw in Example 2.2.6, there can also be terms with cosines and sines in the elements of e^{At}.

2.2.2 Coordinate Transformations

Another way to look at the computation of e^{At} in Lemma 2.2.5 is as follows. The choice of a state is not unique. In Example 2.1.3, we already proposed to replace the state $x = \left[\begin{smallmatrix} q \\ \dot q \end{smallmatrix}\right]$ of the mass-spring-damping system by $x = \left[\begin{smallmatrix} q \\ m\dot q \end{smallmatrix}\right]$ in order to make a comparison with the *RLC* network. More generally, if we have a state x, then we can choose to change to the transformed state z defined by

$$z = T^{-1}x \quad \text{for some invertible } T \in \mathbb{R}^{n\times n}. \tag{2.19}$$

Such a transformation is called a *state transformation*. Because $x = Tz$, the z_k are the coordinates with respect to the basis formed by the columns $T_k \in \mathbb{R}^n$ of T, while the x_k are the coordinates with respect to the *standard basis* of $\mathbb{R}^n$.

Substituting $z = T^{-1}x$ and $x = Tz$ in the equations

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

gives

$$\begin{aligned} \dot z &= T^{-1}\dot x \\ &= T^{-1}(Ax + Bu) \\ &= T^{-1}ATz + T^{-1}Bu \\ y &= CTz + Du. \end{aligned}$$

In other words, a coordinate transformation $x \to z = T^{-1}x$ corresponds to the following transformation of system matrices:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \to \begin{bmatrix} T^{-1}AT & T^{-1}B \\ CT & D \end{bmatrix}. \tag{2.20}$$

We say that two state representations (A, B, C, D) and $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ are *isomorphic* if they can be changed into each other through a state transformation. Note that a state transformation does not change the relation between u and y (see Exercise 2.23).

This transformation changes the homogeneous equation (that is, the equation for $u \equiv 0$)

$$\dot{x} = Ax$$

into

$$\dot{z} = \Lambda z, \qquad \Lambda := T^{-1}AT.$$

This holds for every $\Lambda = T^{-1}AT$, but is particularly interesting when Λ, as before, is diagonal,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

In this case, $\dot{z} = \Lambda z$ is nothing else than n uncoupled equations in the components of z,

$$\dot{z}_k = \lambda_k z_k \qquad \forall k \in \{1, \ldots, n\}.$$

As we know, the solution of this is $z_k(t) = \mathrm{e}^{\lambda_k t} z_k(0)$, so that we obtain

$$x(t) = Tz(t) = T\mathrm{e}^{\Lambda t} z(0) = T\mathrm{e}^{\Lambda t} T^{-1} x(0).$$

2.2.3 Geometric Interpretation of the Solutions of $\dot{x} = Ax$

Because the transformation matrix T does not depend on time, $x(t)$ and $z(t) = T^{-1}x(t)$ exhibit a quantitatively similar dynamical behavior. Because $\dot{z} = \Lambda z$ is uncoupled (provided that Λ is diagonal), it is easier to first analyze the behavior in the z-domain and later transfer the analysis to the x-domain. This is particularly helpful in providing insight into second-order systems. These are systems whose state has two components ($n = 2$).

Example 2.2.7 (Phase portraits of second-order systems). Consider $\dot{x} = Ax$ and assume that $A \in \mathbb{R}^{2\times 2}$ and that the matrix is diagonalizable.

1. *Stable node.* If A has two different negative eigenvalues, then after the transformation $z = T^{-1}x$, the system equation is given by

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \qquad (\lambda_1 < \lambda_2 < 0).$$

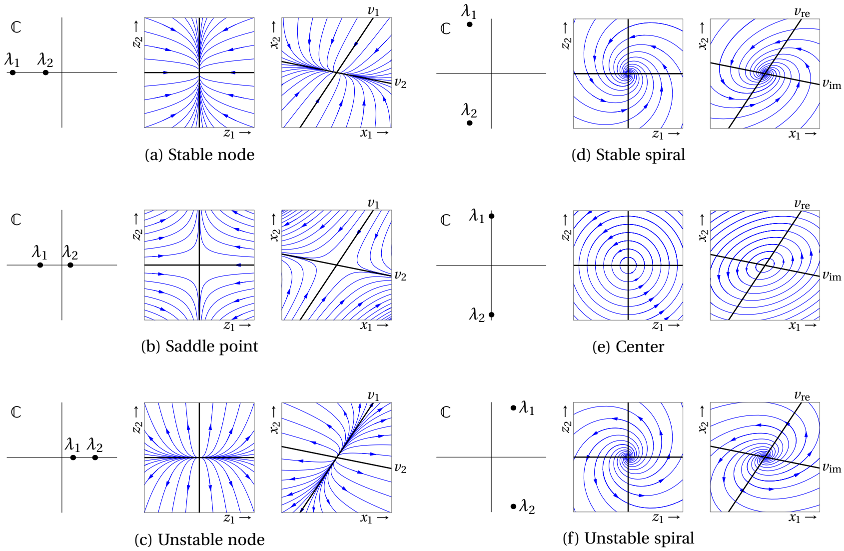


FIGURE 2.5: Phase portraits for diagonalizable second-order systems $\dot{x} = Ax$ for different types of eigenvalues $\lambda_{1,2} \in \mathbb{C}$. In the phase portraits in the (x_1, x_2)-plane, the eigenvectors v_1, v_2 are also shown.

Both states converge to zero, but z_1 does so more quickly because $\lambda_1 < \lambda_2 < 0$. The trajectories in the (z_1, z_2)-plane therefore converge more quickly to the z_2-axis than to the z_1-axis. See Figure 2.5(a). Such plots, where z_2 is set out against z_1, are called *phase portraits*. Figure 2.5(a) also shows the phase portrait of $\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = T\left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]$ (on the right). With each column v_1, v_2 of T corresponds a characteristic trajectory,

$$x(t) = v_1 z_1(t), \qquad x(t) = v_2 z_2(t).$$

These are the trajectories along the eigenvectors v_1 and v_2; we can recognize them as the straight lines in the phase portrait.

2. *Saddle point.* Figure 2.5(b) shows phase portraits in the case where A has one real negative eigenvalue and one real positive eigenvalue. To the negative eigenvalue corresponds a characteristic trajectory $x(t) = v_1 z_1(t)$ that converges to the origin. To the positive eigenvalue corresponds a characteristic trajectory $x(t) = v_2 z_2(t)$ that moves away from the origin. All other trajectories also diverge.

3. *Unstable node.* If both eigenvalues of A are positive (and real), both components of z (and therefore of $x = Tz$) diverge; see Figure 2.5(c).

4. *Stable spiral.* If the eigenvalues of A are not real, they are necessarily each other's complex conjugates, $\lambda_1 = \mu + \mathrm{i}\omega$, $\lambda_2 = \mu - \mathrm{i}\omega$. The corresponding

eigenvectors v_1 and v_2 can then also be chosen as complex conjugates (verify): $v_1 = v_{\mathrm{re}} + \mathrm{i}v_{\mathrm{im}}$ en $v_2 = v_{\mathrm{re}} - \mathrm{i}v_{\mathrm{im}}$. After the transformation $z = \begin{bmatrix} v_1 & v_2 \end{bmatrix}^{-1} x$, we get

$$\begin{bmatrix} \dot z_1 \\ \dot z_2 \end{bmatrix} = \begin{bmatrix} \mu + \mathrm{i}\omega & 0 \\ 0 & \mu - \mathrm{i}\omega \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

but this does not give much insight because $z(t)$ is then complex valued. Instead, we use the transformation $z = \begin{bmatrix} v_{\mathrm{re}} & -v_{\mathrm{im}} \end{bmatrix}^{-1} x$. This gives

$$\begin{bmatrix} \dot z_1 \\ \dot z_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix}}_{\Lambda} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

from which we can deduce (see Exercise 2.9) that

$$\mathrm{e}^{\Lambda t} = \mathrm{e}^{\mu t} \begin{bmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{bmatrix}.$$

Figure 2.5(d) shows the situation when $\mu < 0$. The solutions spiral around the origin and converge to it.

5. *Center.* Consider the stable focus, but this time with $\mu = 0$. In this case, the phase portrait consists of concentric circles around the origin (in the (z_1, z_2)-plane) and ellipses around the origin (in the (x_1, x_2)-plane). See Figure 2.5(e).

6. *Unstable spiral.* If the eigenvalues are complex, namely $\lambda_1 = \mu + \mathrm{i}\omega$ and $\lambda_2 = \mu - \mathrm{i}\omega$ with $\mu > 0$, then x_1 and x_2 diverge. This looks the same as the stable spiral, except that it spirals away from the origin; see Figure 2.5(f).

If the eigenvalues coincide, $\lambda_1 = \lambda_2$, then the A-matrix may not be diagonalizable. We do not consider these cases. They will be studied in the course *Ordinary Differential Equations.* □

Example 2.2.8 (Classification of the damping for second-order systems). Consider, once more, the mass-spring-damper system from Example 2.1.1 with $k > 0$ and $r > 0$, and assume that we are not exerting any external force $u = F$, so $F \equiv 0$. We then have

$$\dot x = Ax \quad \text{with} \quad A = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{r}{m} \end{bmatrix}.$$

The eigenvalues of A are

$$\lambda_{1,2} = -\frac{r}{2m} \pm \sqrt{\frac{r^2}{4m^2} - \frac{k}{m}}.$$

Depending on the values of k and r, we can distinguish three cases:

Underdamped. If $\frac{r^2}{4m^2} - \frac{k}{m} < 0$, then λ_1 and λ_2 are complex eigenvalues, with $\operatorname{Re}\lambda_1 = \operatorname{Re}\lambda_2 = -\frac{r}{2m} < 0$. This is called the *underdamped* situation and occurs when the damping r is weak. The phase portrait is as in Figure 2.5(d). Both x_1 and x_2 converge to zero, but they do so oscillating about zero. The mass m therefore continuously oscillates about the equilibrium, and only comes to a halt in the limit.

Overdamped. If $\frac{r^2}{4m} - \frac{k}{m} > 0$, then λ_1 and λ_2 are both real, and we have $\lambda_{1,2} < 0$. This is called the *overdamped* situation and occurs when r is sufficiently large. The phase portrait is as in Figure 2.5(a). In this case, x_1 and x_2 can at most have a local extremum, after which they converge monotonically to zero. The mass m therefore shoots through the origin at most once, after which it converges monotonically to it.

Critically damped. The boundary case between underdamped and overdamped is when $\frac{r^2}{4m^2} - \frac{k}{m} = 0$. In this case, λ_1 and λ_2 are real and equal, $\lambda_1 = \lambda_2 = -\frac{r}{2m} < 0$. This is called the *critically damped* situation. The mass just barely refrains from oscillating, and the system has in common with the overdamped situation that the mass passes through the equilibrium at most once.

In the critically damped situation, the damper is just strong enough to prevent oscillations, but not strong enough to make the system excessively sluggish; see Figure 2.6. □

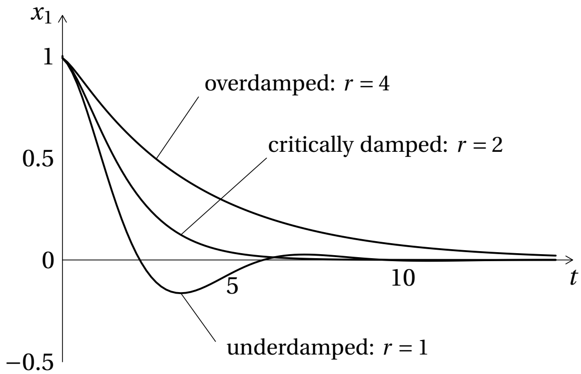


FIGURE 2.6: Position $x_1(t)$ in an overdamped, critically damped, and underdamped mass-spring-damper system (Example 2.2.8 with $m = 1$ and $k = 1$ and $x_1(0) = 1, \dot{x}_1(0) = 0$.)

2.3 Stability of Equilibrium Points

In the previous example, we saw that in certain cases of $\dot{x} = Ax$, the solution $x(t)$ converges to the origin for $t \to \infty$, and that in other cases, $x(t)$ diverges away from the origin. In both cases, if $x(t_0)$ is at the origin—$x(t_0) = 0$—then in theory, $x(t)$ always remains at the origin, $x(t) \equiv 0$. We therefore call the origin an *equilibrium point*. However, a small disturbance can move $x(t)$ away from this equilibrium point, and then the situation depends on whether $x(t)$ returns to the equilibrium point or not. We will analyze this phenomenon in this section, for both linear and nonlinear systems.

Definition 2.3.1 (Equilibrium point). An $x^* \in \mathbb{R}^n$ is an *equilibrium point* of a differential equation if it is a constant solution of the differential equation. □

The equilibrium points x^* of

$$\dot{x}(t) = f(x(t)), \quad x(t) \in \mathbb{R}^n, \tag{2.21}$$

are therefore the zeros of f,

$$0 = f(x^*). \tag{2.22}$$

Because $x(t) = x^*$ satisfies the differential equation, equilibrium points are also called equilibrium *solutions*.

It should be clear that for every matrix A, the origin is an equilibrium point of $\dot{x} = Ax$. If A is invertible, then the origin is the only equilibrium point of $\dot{x} = Ax$ (see also Exercise 2.1).

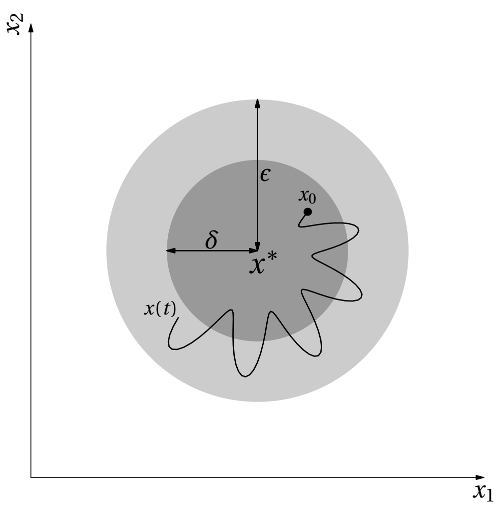


FIGURE 2.7: Stability for 2-dimensional $x = (x_1, x_2)$.

Definition 2.3.2 (Stability). Consider (2.21) with initial condition $x(0) = x_0$. An equilibrium point x^* is called

1. *stable* if

$$\forall \epsilon > 0 \exists \delta > 0 : \|x_0 - x^*\| < \delta \Rightarrow \|x(t) - x^*\| < \epsilon \ \forall t \geq 0$$

(see Figure 2.7);

2. *asymptotically stable* if it is stable and we can choose $\delta > 0$ such that

$$\|x_0 - x^*\| < \delta \Rightarrow \lim_{t\to\infty} x(t) = x^*. \tag{2.23}$$

□

Stability means that in the future, the state stays in any chosen (small) neighborhood of the equilibrium point if we begin sufficiently close to the equilibrium point; see Figure 2.7. This definition also applies to nonlinear systems such as $\dot{x}(t) = f(x(t))$. For linear systems $\dot{x}(t) = Ax(t)$ it is somewhat simpler, and δ is only there for show. The asymptotic stability of $\dot{x} = Ax$ is all about the position of the eigenvalues of A.

Lemma 2.3.3 (Asymptotic stability). *Consider the system $\dot{x}(t) = Ax(t)$ with equilibrium point $x^* = 0$. The following four statements are equivalent:*

1. *$x^* = 0$ is asymptotically stable.*
2. *$\lim_{t\to\infty} x(t) = 0$ for all $x(0) = x_0 \in \mathbb{R}^n$.*
3. *All eigenvalues of A have negative real part.*
4. *There exist $\sigma > 0$, $M > 0$ such that $\|e^{At}\| \leq M e^{-\sigma t}$ for $t \geq 0$.*

Proof.

1 $\Longrightarrow$ 2. Because of the asymptotic stability, $\tilde{x}(t)$ converges to zero when $\|\tilde{x}_0\| < \delta$ for some fixed $\delta > 0$. However, *every* $x_0 \in \mathbb{R}^n$ is a scalar multiple $x_0 = c\tilde{x}_0$ of an $\tilde{x}_0$ that is so small that $\|\tilde{x}_0\| < \delta$. By the linearity, we have $x(t) = c\tilde{x}(t)$; therefore $x(t)$ also converges to zero.

2 $\Longrightarrow$ 3. By contradiction: For every eigenvalue λ of A with eigenvector v, $x(t) := ve^{\lambda t}$ is a solution of $\dot{x} = Ax$. If $\operatorname{Re}\lambda \geq 0$, then $x(t) := ve^{\lambda t}$ does not converge to zero for $t \to \infty$. (If λ is not real, take $x(t) = \operatorname{Re} ve^{\lambda t}$, and verify that this signal is not the zero function.)

3 $\Longrightarrow$ 4. Take $\sigma := -\frac{1}{2}\max_i \operatorname{Re}\lambda_i(A)$; then $\sigma > 0$. Take the norm $\|B\| := \max_{ij} |B_{ij}|$. The eigenvalues of $\sigma I + A$ have a negative real part, so the elements of $e^{(\sigma I + A)t}$ are linear combinations of terms of the form $t^k e^{\lambda t}$ with $k \in \mathbb{N}$, $\lambda \in \mathbb{C}$, $\operatorname{Re}(\lambda) < 0$. The elements are therefore bounded for positive time, and consequently $\|e^{(\sigma I+A)t}\|$ is bounded for positive time. That is, there exists an $M > 0$ such that

$$\|e^{(\sigma I+A)t}\| \le M \text{ for all } t \ge 0.$$

Statement 4 now follows from $\|e^{(\sigma I+A)t}\| = \|e^{\sigma t}e^{At}\| = e^{\sigma t}\|e^{At}\|$.

4 $\Longrightarrow$ 1. Because $\|e^{At}\| \le Me^{-\sigma t}$ for $t \ge 0$, we have $\|x(t)\| = \|e^{At}x(0)\| \le ce^{-\sigma t}\|x(0)\|$ for some c. Hence $\lim_{t\to\infty} x(t) = 0$. For a given $\epsilon > 0$, we have $\|x(t)\| < \epsilon$ for all $t > 0$ if $\|x(0\| < \delta := \epsilon/c$.

The choice of norms on $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$ is not relevant because they are all equivalent. ■

If $\operatorname{Re}\lambda_i(A) > 0$ for some i, then the system $\dot{x} = Ax$ is unstable because regardless of how small (in norm) the eigenvector v for the eigenvalue $\lambda_i(A)$ is taken, the possible solution $x(t) = ve^{\lambda_i(A)t}$ always diverges.

Example 2.3.4 (Mass-spring-damper). The mass-spring-damper system has A-matrix given by

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{r}{m} \end{bmatrix}.$$

Suppose that all parameters are positive, $m, k, r > 0$. The eigenvalues of A are the zeros of

$$\begin{aligned} \chi_A(\lambda) &= \det(\lambda I - A) \\ &= \lambda(\lambda + \tfrac{r}{m}) + \tfrac{k}{m} \\ &= \lambda^2 + \tfrac{r}{m}\lambda + \tfrac{k}{m} \\ &= \frac{m\lambda^2 + r\lambda + k}{m}. \end{aligned}$$

The zeros $\lambda_{1,2}$ of this polynomial are

$$\lambda_{1,2} = \frac{-r \pm \sqrt{r^2 - 4km}}{2m}.$$

If the discriminant $r^2 - 4km$ is negative, then both zeros have real part $-r/(2m) < 0$. So the system is then asymptotically stable. If the discriminant is ≥ 0, then the following argument holds: since $km > 0$, we have $r^2 - 4km < r^2$, or $\sqrt{r^2 - 4km} < \sqrt{r^2} = r$. In other words, if $-r$ is negative, then $-r \pm \sqrt{r^2 - 4km}$ are both also negative. The system is therefore asymptotically stable for all positive m, k, r. □

2.3.1 Asymptotically stable Polynomials and the Routh–Hurwitz Test

We have seen that the asymptotic stability of

$$\dot{x} = Ax$$

can be determined using the eigenvalues of the matrix A, that is, the zeros of the characteristic polynomial,

$$\chi_A(\lambda).$$

Now, for general polynomials of degree greater than four, there is no finite expression for the zeros[2]. However, to test stability we do not really need to know the zeros of a polynomial. We only need to figure out if all zeros have negative real part, and for that finite tests do exist! This is a well-known result, developed independently by Edward Routh (1831–1907) and Adolf Hurwitz (1859–1919). To honor the latter, we have the following notion.

Definition 2.3.5 (Strictly Hurwitz). A polynomial is said to be *asymptotically stable* or *strictly Hurwitz* if all its zeros have negative real part. □

Here is the famous result.

Theorem 2.3.6 (Routh–Hurwitz test). *A polynomial*

$$a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n, \qquad a_0 \neq 0$$

is asymptotically stable if and only if all $n+1$ numbers in the first column of the Routh table *have the same sign. The Routh table is of the form*

$$\begin{array}{lllll} a_0 & a_2 & a_4 & a_6 & \cdots \\ a_1 & a_3 & a_5 & a_7 & \cdots \\ b_0 & b_2 & b_4 & \cdots & \\ b_1 & b_3 & b_5 & \cdots & \\ \cdots & & & & \end{array}$$

It has $n+1$ rows. The first two rows follow from the polynomial. The third row is constructed from the two rows above it (rows 1 and 2) by

$$\begin{bmatrix} b_0 & b_2 \cdots \end{bmatrix} = \begin{bmatrix} a_2 & a_4 & \cdots \end{bmatrix} - \frac{a_0}{a_1}\begin{bmatrix} a_3 & a_5 & \cdots \end{bmatrix}.$$

Every following row k is constructed in the same manner from the rows directly above it (rows $k-2$ and $k-1$.) □

A proof of this theorem is given in Appendix A.3.

[2] Numerically, however, finding the zeros is in general no problem.

Example 2.3.7 (Routh–Hurwitz). Consider the third-degree polynomial

$$\lambda^3 + \lambda^2 + \lambda + c$$

depending on $c \in \mathbb{R}$. The Routh table then consists of $n+1=4$ rows:

$$\begin{array}{cc} 1 & 1 \\ 1 & c \\ 1-c & \\ c & \end{array}$$

All four numbers in the first column have the same sign exactly when $0 < c < 1$, hence the polynomial is asymptotically stable if and only $0 < c < 1$. (If $c = 0$ it has a zero at $\lambda = 0$, if $c = 1$ it has two imaginary zeros $s_{1,2} = \pm\mathrm{i}$.) □

Example 2.3.8 (Degree five). Consider the fifth-degree polynomial

$$2\lambda^5 + 1\lambda^4 + 4\lambda^3 + 3\lambda^2 + 6\lambda + 5.$$

The Routh table is now

$$\begin{array}{ccc} 2 & 4 & 6 \\ 1 & 3 & 5 \\ -2 & -4 & \\ 1 & 5 & \\ 6 & & \\ 5 & & \end{array}$$

The six elements of the first column do not all have the same sign, so the polynomial is not asymptotically stable.

(An extension of the Routh-Hurwitz test says that there are as many unstable zeros as there are sign changes in the ordered sequence of the first column $(2,1,-2,1,6,5)$. So here that is 2. The zeros, obtained numerically, are $0.65763 \pm 1.21259\mathrm{i}, -0.51573 \pm 1.18754\mathrm{i}$ and -0.78379.) □

Example 2.3.9 (Dividing by zero). The Routh table of $\lambda^2 + 4$ should have 3 elements, but it breaks down because we cannot divide by zero,

$$\begin{array}{cc} 1 & 4 \\ 0 & \\ ? & \end{array}$$

The polynomial is therefore not asymptotically stable. □

With the Routh–Hurwitz test, we can for instance show that a second-degree polynomial

$$a\lambda^2 + b\lambda + c, \qquad a \neq 0$$

is asymptotically stable if and only if all three coefficients a, b, c have the same sign; see Exercise 2.12. It immediately gives us that the characteristic polynomial $(m\lambda^2 + r\lambda + k)/m$ of the mass-spring-damper system is asymptotically stable if and only if m, k, r have the same sign.

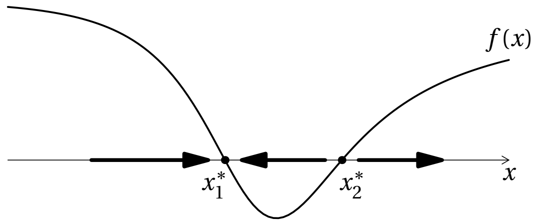


FIGURE 2.8: A nonlinear system with two equilibrium points.

2.4 Linearization

Asymptotic stability of an equilibrium point in a *non*linear system can often be determined by approximating the system by a linear system. Suppose, to get an idea, that the state has one component and satisfies

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

with f as sketched in Figure 2.8. This figure shows the two equilibrium points, x_1^* and x_2^* (the zeros of f). Now, if at some time $x(t)$ is just to the left of the left equilibrium point x_1^*, then we see in Figure 2.8 that $\dot{x} = f(x) > 0$, and therefore that $x(t)$ increases, that is, moves toward x_1^*. If $x(t)$ is just to the *right* of x_1^*, then according to the figure, $\dot{x} = f(x) < 0$, that is, $x(t)$ *de*creases, and therefore once again moves toward x_1^*. It looks like x_1^* is asymptotically stable. The second equilibrium point x_2^* is not stable: if $x(t)$ lies to the right of x_2^*, then $\dot{x} = f(x) > 0$, and therefore x moves away from x_2^*.

It looks like x_1^* is asymptotically stable, while x_2^* is not. To be able to draw this conclusion, we do not need to know f exactly. It suffices to know whether f is decreasing in the neighborhood of the equilibrium (then it is asymptotically stable) or increasing (then it is unstable). Whether it is decreasing of increasing follows from the derivative $\frac{\partial f}{\partial x}(x^*)$ at the equilibrium point. This leads to the following lemma.

Lemma 2.4.1 (Linear stability, scalar case). *Consider the differential equation*

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

with equilibrium point $x^* \in \mathbb{R}$ and assume that $f : \mathbb{R} \to \mathbb{R}$ is differentiable. Then the system is

1. *asymptotically stable in x^* if $\frac{\partial f}{\partial x}(x^*) < 0$;*

2. *not stable in x^* if $\frac{\partial f}{\partial x}(x^*) > 0$.*

Proof. Without loss of generality, we assume that $x^* = 0$ (otherwise, we first shift x by the constant x^*). Because f is differentiable and $f(x^*) = f(0) = 0$, we have $f(x) = Ax + \mathrm{o}(x)$ with $A = \frac{\partial f}{\partial x}(0)$ and o a "little o" function, that is,

$$\dot{x} = Ax + \mathrm{o}(x). \tag{2.24}$$

Suppose that $A > 0$. Because of the $o(x)$, there exists an $\epsilon > 0$ such that

$$\dot{x} \geq \frac{A}{2}x \qquad \forall 0 < x < \epsilon.$$

So if $0 < x(t) < \epsilon$, then $x(t)$ increases at least as quickly as the solution of $\dot{x} = \frac{A}{2}x$, which is $e^{A/2t}x(0)$. This $e^{A/2t}x(0)$ goes to infinity. So for every $\delta \in (0,\epsilon)$ and every $0 < x(0) < \delta$, there will be a time t such that $x(t) \geq \epsilon$. The system is therefore unstable.

Conversely, if $A < 0$, then it follows from (2.24) that for every $\epsilon > 0$ there exists a $\delta \in (0,\epsilon)$ such that

$$\dot{x}(t) \leq \frac{A}{2}x(t) \qquad \forall 0 < x(t) < \delta.$$

For such an $x(t)$, the function $x(t)$ goes to zero almost as quickly as the solution of $\dot{x} = A/2x$, which is $e^{A/2t}x(0)$, so that we still have $x(t) \leq \delta < \epsilon$ for all $t > 0$ and $\lim_{t\to\infty} x(t) \leq \lim_{t\to\infty} e^{A/2t}x(0) = 0$. The same argument works for negative $x(t) \in (-\delta, 0)$. ■

In the multidimensional case, the role of $\frac{\partial f}{\partial x}(x^*)$ is taken over by a matrix of derivatives.

Theorem 2.4.2 (Linear stability). *Consider the system of differential equations (2.21) with equilibrium point $x^* \in \mathbb{R}^n$, and assume that $f : \mathbb{R} \to \mathbb{R}^n$ is continuously differentiable. Define the Jacobian matrix*

$$A = \frac{\partial f}{\partial x}(x^*) := \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x^*) & \frac{\partial f_1}{\partial x_2}(x^*) & \cdots & \frac{\partial f_1}{\partial x_n}(x^*) \\ \frac{\partial f_2}{\partial x_1}(x^*) & \frac{\partial f_2}{\partial x_2}(x^*) & \cdots & \frac{\partial f_2}{\partial x_n}(x^*) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1}(x^*) & \frac{\partial f_n}{\partial x_2}(x^*) & \cdots & \frac{\partial f_n}{\partial x_n}(x^*) \end{bmatrix}. \tag{2.25}$$

Then the equilibrium point x^* of $\dot{x} = f(x)$ is

- *asymptotically stable if* $\operatorname{Re}\lambda < 0$ *for all eigenvalues λ of A;*
- *unstable if A has at least one eigenvalue λ with* $\operatorname{Re}\lambda > 0$.

Proof (sketch). Let $\delta_x(t)$ be the difference between $x(t)$ and x^*, that is, $\delta_x(t) = x(t) - x^*$. This difference satisfies

$$\begin{aligned} \dot{\delta}_x(t) &= \dot{x}(t) - \dot{x}^* \\ &= \dot{x}(t) \\ &= f(x(t)) \\ &= f(x^* + \delta_x(t)) \\ &= A\delta_x(t) + o(\delta_x(t)). \end{aligned}$$

The difference vector $\delta_x = x - x^*$ therefore asymptotically satisfies the linear equation $\dot{\delta}_x = A\delta_x$. From Lemma 2.3.3, we therefore expect that $\lim_{t\to\infty} \|x(t) - x^*\| = 0$ for sufficiently small $\|x(0) - x^*\|$, provided that Re $\lambda < 0$ for all eigenvalues λ of A. For a true proof, we need Lyapunov functions (Appendix A.4.) ■

As argued in the proof, the difference vector $\delta_x = x - x^*$ asymptotically satisfies

$$\dot{\delta}_x = A\delta_x. \tag{2.26}$$

This linear system (2.26) is called the *linearization* or the *linearized system* of $\dot{x}(t) = f(x(t))$ around the equilibrium point x^*. The matrix of derivatives in (2.25) is called the *Jacobian matrix* or *Jacobian* of f at $x = x^*$.

Example 2.4.3 (First-order differential equation). Consider the differential equation

$$\dot{x} = -\sin(2x). \tag{2.27}$$

The function $f(x) = -\sin(2x)$ has many zeros, including

$$x^* = 0.$$

The idea of linearization is that in the neighborhood of $x^* = 0$, the sine function is barely distinguishable from the tangent line with slope

$$A = f'(x^*) = -2\cos(0) = -2$$

(see Figure 2.9). The solution $x(t)$ of (2.27) is likely to be very similar to $x^* + \delta_x(t) = \delta_x(t)$, where $\delta_x(t)$ is the solution of the linearization

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

□

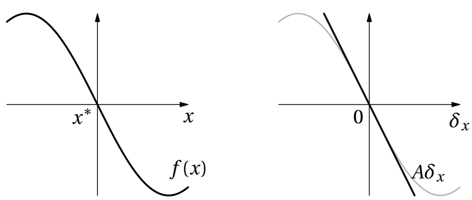


FIGURE 2.9: Nonlinear $f(x)$ (left) and its linear approximation $A\delta_x$ (right).

Example 2.4.4 (Pendulum). The standard model for the pendulum (Figure 2.10) is

$$m\ell\ddot{\phi} + mg\sin(\phi) = 0. \tag{2.28}$$

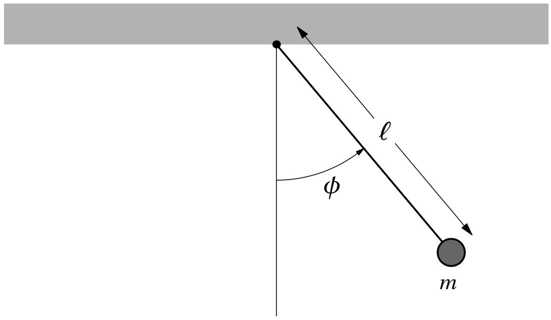


FIGURE 2.10: The pendulum.

Here ϕ denotes the angle the pendulum makes with the vertical axis, m is the mass of the pendulum, ℓ the length of the cable (that we assume to be weightless), and g is the gravitational acceleration. This model is nonlinear because of the term $\sin(\phi)$. To determine equilibrium points, we first write (2.28) as a state representation. Choose

$$x := \begin{bmatrix} \phi \\ \dot{\phi} \end{bmatrix}.$$

This gives

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} x_2 \\ -\frac{g}{\ell}\sin(x_1) \end{bmatrix}}_{f(x)}. \tag{2.29}$$

The corresponding Jacobian $\frac{\partial f}{\partial x}(x)$ for general x is

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{\ell}\cos(x_1) & 0 \end{bmatrix}.$$

Now, we have $f(x^*) = 0$ exactly when $x_1^* = k\pi$ ($k \in \mathbb{Z}$) and $x_2^* = 0$. These are the two vertical positions: hanging (downward) and standing (upward).

The linearization around the hanging position $x^* = (0,0)$ is

$$\begin{aligned} \dot{\delta}_x &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(0,0) & \frac{\partial f_1}{\partial x_2}(0,0) \\ \frac{\partial f_2}{\partial x_1}(0,0) & \frac{\partial f_2}{\partial x_2}(0,0) \end{bmatrix} \delta_x \\ &= \underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{g}{\ell} & 0 \end{bmatrix}}_{A} \delta_x. \end{aligned}$$

The Jacobian A has only imaginary eigenvalues, so Thm. 2.4.2 does not say anything about stability properties of the nonlinear system.

The linearization around the standing position $x^* = (\pi, 0)$ is

$$\begin{aligned}\dot{\delta}_x &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\pi,0) & \frac{\partial f_1}{\partial x_2}(\pi,0) \\ \frac{\partial f_2}{\partial x_1}(\pi,0) & \frac{\partial f_2}{\partial x_2}(\pi,0) \end{bmatrix} \delta_x \\ &= \underbrace{\begin{bmatrix} 0 & 1 \\ +\frac{g}{\ell} & 0 \end{bmatrix}}_{A} \delta_x.\end{aligned}$$

The Jacobian A now has eigenvalues $\lambda = \pm\sqrt{g/\ell}$. Since one is positive, the nonlinear system is unstable, by Thm. 2.4.2. □

It is also possible to linearize along a given trajectory (which may be time varying).

Definition 2.4.5 (Linearization along a trajectory). Consider the system

$$\begin{cases} \dot{x}(t) &= f(x(t), u(t)), \\ y(t) &= h(x(t), u(t)), \end{cases}$$

and a given solution $(u^*(t), x^*(t), y^*(t))$ of this system. The *linearization* along this solution is the linear system

$$\begin{cases} \dot{\delta}_x(t) &= A\delta_x(t) + B\delta_u(t) \\ \delta_y(t) &= C\delta_x(t) + D\delta_u(t) \end{cases}$$

with A, B, C, D the Jacobians

$$A = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x^*, u^*) & \cdots & \frac{\partial f_1}{\partial x_n}(x^*, u^*) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1}(x^*, u^*) & \cdots & \frac{\partial f_n}{\partial x_n}(x^*, u^*) \end{bmatrix},$$

$$B = \begin{bmatrix} \frac{\partial f_1}{\partial u_1}(x^*, u^*) & \cdots & \frac{\partial f_1}{\partial u_{n_u}}(x^*, u^*) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial u_1}(x^*, u^*) & \cdots & \frac{\partial f_n}{\partial u_{n_u}}(x^*, u^*) \end{bmatrix},$$

$$C = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x^*, u^*) & \cdots & \frac{\partial h_1}{\partial x_n}(x^*, u^*) \\ \vdots & \vdots & \vdots \\ \frac{\partial h_n}{\partial x_1}(x^*, u^*) & \cdots & \frac{\partial h_n}{\partial x_n}(x^*, u^*) \end{bmatrix},$$

$$D = \begin{bmatrix} \frac{\partial h_1}{\partial u_1}(x^*, u^*) & \cdots & \frac{\partial h_1}{\partial u_{n_u}}(x^*, u^*) \\ \vdots & \vdots & \vdots \\ \frac{\partial h_n}{\partial u_1}(x^*, u^*) & \cdots & \frac{\partial h_n}{\partial u_{n_u}}(x^*, u^*) \end{bmatrix}.$$

□

In this setting, the solution $(u^*(t), x^*(t), y^*(t))$ is usually called an (equilibrium) solution or (equilibrium) trajectory of the system. Do note that x^* and u^* may depend on time, and that the Jacobians A, B, C, D above may be time varying.

Example 2.4.6 (Predator-prey model). Consider the predator-prey model of Example 1.8.6:

$$\begin{aligned}\dot{x}_1 &= ax_1 - bx_1x_2 - u_1x_1,\\ \dot{x}_2 &= cx_1x_2 - dx_2 - u_2x_2,\\ y &= x_2,\end{aligned} \tag{2.30}$$

and consider the solution

$$\begin{aligned}u_1^*(t) &= 0,\\ u_2^*(t) &= 0,\\ x_1^*(t) &= \frac{d}{c},\\ x_2^*(t) &= \frac{a}{b},\\ y^*(t) &= \frac{a}{b}.\end{aligned}$$

This is a (constant) equilibrium solution. Linearization of (2.30) around this equilibrium solution gives the linear system

$$\begin{aligned}\begin{bmatrix}\dot{\delta}_{x_1}\\ \dot{\delta}_{x_2}\end{bmatrix} &= \begin{bmatrix}0 & -\frac{bd}{c}\\ \frac{ac}{b} & 0\end{bmatrix}\begin{bmatrix}\delta_{x_1}\\ \delta_{x_2}\end{bmatrix} + \begin{bmatrix}-\frac{d}{c} & 0\\ 0 & -\frac{a}{b}\end{bmatrix}\begin{bmatrix}\delta_{u_1}\\ \delta_{u_2}\end{bmatrix}\\ \delta_y &= \begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}\delta_{x_1}\\ \delta_{x_2}\end{bmatrix}.\end{aligned} \tag{2.31}$$

The solutions of (2.30) in the neighborhood of the equilibrium solution given above can be approximated by

$$\begin{aligned}u_1 &= u_1^* + \delta_{u_1},\\ u_2 &= u_2^* + \delta_{u_1},\\ x_1 &\approx \frac{d}{c} + \delta_{x_1},\\ x_2 &\approx \frac{a}{b} + \delta_{x_2},\\ y &\approx \frac{a}{b} + \delta_y.\end{aligned}$$

Note that here the linearization (2.31) is *time invariant*; this is because the original nonlinear system (2.30) is time invariant *and* the linearization takes place around a solution (x^*, u^*, y^*) that does not depend on time. □

2.5 Higher-Order Differential Equations

We conclude this chapter with a remark about higher-order differential equations. An nth order differential equation

$$y^{(n)}(t) + p_{n-1}y^{(n-1)}(t) + \cdots + p_1 y^{(1)}(t) + p_0 y(t) = q_0 u(t)$$

can easily be changed into a state representation using the substitution

$$x := \begin{bmatrix} y \\ y^{(1)} \\ \vdots \\ y^{(n-2)} \\ y^{(n-1)} \end{bmatrix}.$$

This gives the nth order state representation

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -p_0 & -p_1 & \cdots & \cdots & -p_{n-1} \end{bmatrix} x + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ q_0 \end{bmatrix} u \\ y &= \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \end{bmatrix} x. \end{aligned} \tag{2.32}$$

Using this, the theory of the previous sections also applies to these types of differential equations. However, this method does not automatically generalize to the case where the differential equation also contains a derivative of u, as in the following general nth order differential equation:

$$\begin{aligned} & y^{(n)}(t) + p_{n-1}y^{(n-1)}(t) + \cdots + p_1 y^{(1)}(t) + p_0 y(t) \\ = \quad & q_n u^{(n)}(t) + q_{n-1}u^{(n-1)}(t) + \cdots + q_1 u^{(1)}(t) + q_0 u(t). \end{aligned} \tag{2.33}$$

This type of differential equation also admits a state representation, but its construction is different. The following example shows how.

Example 2.5.1 (Simulation diagram). Consider the differential equation

$$\ddot{y} + 5\dot{y} + 6y = 7\dot{u} + 8u. \tag{2.34}$$

To derive the state representation, we bring all terms except $\ddot{y}$ to the right-hand side of the equation,

$$\ddot{y} = -5\dot{y} - 6y + 7\dot{u} + 8u.$$

Next, we integrate the equation as often as necessary to get rid of the derivatives

$$\begin{aligned} y &= \iint [-5\dot{y} - 6y + 7\dot{u} + 8u] \\ &= \int \Big[-5y + 7u + \int [-6y + 8u]\Big]. \end{aligned}$$

As a last step, we assign a state component x_k, with $k = 1,2$, to each of the antiderivatives,

$$y = \underbrace{\int \left[-5y + 7u + \underbrace{\int [-6y + 8u]}_{x_1}\right]}_{x_2}.$$

The so defined state components satisfy

$$\begin{cases} \dot{x}_1 = -6y + 8u, \\ \dot{x}_2 = -5y + 7u + x_1, \\ \;\; y = x_2. \end{cases}$$

We can now read off the state representation directly:

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 0 & -6 \\ 1 & -5 \end{bmatrix} x + \begin{bmatrix} 8 \\ 7 \end{bmatrix} u, \\ y &= \begin{bmatrix} 0 & \quad 1 \end{bmatrix} x. \end{aligned}$$

□

The general process to construct a state representation from an ordinary nth order differential equation

$$\begin{aligned} &y^{(n)} + p_{n-1} y^{(n-1)} + \cdots + p_0 y \\ &\qquad = q_n u^{(n)} + q_{n-1} u^{(n-1)} + \cdots + q_0 u \end{aligned} \tag{2.35}$$

is not more complicated than the example we just gave. For typographical reasons, we only present the method for third-order differential equations. We first bring all terms except $y^{(n=3)}$ to the right of the equal sign,

$$y^{(3)} = q_3 u^{(3)} + [q_2 u^{(2)} - p_2 y^{(2)}] + [q_1 u^{(1)} - p_1 y^{(1)}] + [q_0 u - p_0 y].$$

Then we integrate $n = 3$ times,

$$y = q_3 u + \int \left[q_2 u - p_2 y + \int \left[q_1 u - p_1 y + \int [q_0 u - p_0 y]\right]\right]$$

and assign a state component $x_1, \ldots, x_{n=3}$ to each of the antiderivatives,

$$y = q_3 u + \underbrace{\int \left[q_2 u - p_2 y + \underbrace{\int \left[q_1 u - p_1 y + \underbrace{\int [q_0 u - p_0 y]}_{x_1}\right]}_{x_2}\right]}_{x_3}. \tag{2.36}$$

This way, the state components satisfy

$$\begin{aligned}
\dot{x}_1 &= q_0 u - p_0 y, \\
\dot{x}_2 &= q_1 u - p_1 y + x_1, \\
\dot{x}_3 &= q_2 u - p_2 y + x_2, \\
y &= q_3 u + x_3,
\end{aligned}$$

or, in matrix form,

$$\dot{x} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x + \begin{bmatrix} q_0 & -p_0 \\ q_1 & -p_1 \\ q_2 & -p_2 \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix}, \tag{2.37}$$

$$y = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} x + q_3 u. \tag{2.38}$$

In particular, we see that $y = x_3 + q_3 u$. This allows us to eliminate y in (2.37), and the result is a state representation that is called the *observer canonical form* (the reasons for this name will become clear in § 3.5):

$$\begin{aligned}
\dot{x} &= \begin{bmatrix} 0 & 0 & -p_0 \\ 1 & 0 & -p_1 \\ 0 & 1 & -p_2 \end{bmatrix} x + \begin{bmatrix} q_0 - p_0 q_3 \\ q_1 - p_1 q_3 \\ q_2 - p_2 q_3 \end{bmatrix} u, \\
y &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} x + q_3 u.
\end{aligned}$$

For general n we have the following result.

Lemma 2.5.2 (Observer canonical form). *A pair $u, y : \mathbb{R} \to \mathbb{R}$ is a solution of*

$$y^{(n)} + p_{n-1} y^{(n-1)} + \cdots + p_0 y = q_n u^{(n)} + \cdots + q_0 u$$

if and only if there exists an $x : \mathbb{R} \to \mathbb{R}^n$ such that

$$\dot{x} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix} x + \begin{bmatrix} q_0 - p_0 q_n \\ q_1 - p_1 q_n \\ \vdots \\ \vdots \\ q_{n-1} - p_{n-1} q_n \end{bmatrix} u, \tag{2.39}$$

$$y = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} x + q_n u.$$

Proof (idea). For simplicity, we again take $n = 3$. It follows from the above that if (u, y) satisfies (2.35), then we can take x_i as constructed in (2.36). These x_i satisfy (2.37,2.38) by construction. Conversely, if (2.37,2.38) hold, then substitution shows that x satisfies (2.36). Differentiate (2.36) n times, and we obtain (2.35). ∎

Note that the A -matrix of the observer canonical form is the transpose of the A-matrix of (2.32).

Example 2.5.3 (Observer canonical form). As we know, a pair (u, y) satisfies $\dot{y} = \dot{u}$ if and only if $y = u + c$ for some constant $c \in \mathbb{R}$. The observer canonical form of $\dot{y} = \dot{u}$ is (verify this yourself)

$$\dot{x} = 0,$$
$$y = x + u.$$

Since $\dot{x}$ is zero, we have that x is constant. The state x takes over the role of c in $y = u + c$. □

The state x in the observer canonical form usually does not have a physical interpretation such as the current or voltage in Example 2.1.3.

2.5.1 Polynomial Representation

With the differential equation (2.35), we associate the polynomials $P(s)$ and $Q(s)$ defined by

$$P(s) = s^n + p_{n-1}s^{n-1} + \cdots + sp_1 + p_0$$

and

$$Q(s) = q_n s^n + q_{n-1}x^{n-1} + \cdots + sq_1 + q_0.$$

We can then write the differential equation (2.35) compactly as

$$P(\tfrac{\mathrm{d}}{\mathrm{d}t})y = Q(\tfrac{\mathrm{d}}{\mathrm{d}t})u.$$

Example 2.5.4. We can write the differential equation

$$m\ddot{y}(t) + r\dot{y}(t) + ky(t) = u(t)$$

as $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ with

$$P(s) = ms^2 + rs + k, \qquad Q(s) = 1.$$

□

This notation is going to be useful.

2.5.2 Asymptotic Stability of Differential Equations

Earlier in this chapter, we defined asymptotic stability for systems of the form $\dot{x} = Ax$. The following definition is a slight generalization of this. It is a generalization that also makes it applicable for ordinary differential equations.

Definition 2.5.5 (Asymptotic stability). A linear system $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ with input u and output y is *asymptotically stable* if $\lim_{t\to\infty} y(t) = 0$ for all solutions of the homogeneous part $P(\frac{\mathrm{d}}{\mathrm{d}t})y = 0$. □

For the system $\dot{y} = Ay + Bu$, this is equivalent to the asymptotic stability defined earlier. The homogeneous system $P(\frac{\mathrm{d}}{\mathrm{d}t}y) = 0$ is the system we get by taking $u(t) = 0$. Now, we know from calculus that to every zero $\lambda \in \mathbb{C}$ of the polynomial $P(\lambda)$ there corresponds a solution $y(t) = \mathrm{e}^{\lambda t}$. So for asymptotic stability, all these zeros λ must have negative real parts. It turns out that this is also sufficient, as we will now show. To see this, we represent the differential equation

$$P(\tfrac{\mathrm{d}}{\mathrm{d}t})y = 0$$

in the equivalent observer canonical form

$$\dot{x} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix} x, \qquad y = x_n. \tag{2.40}$$

(This follows from Lemma 2.5.2 with $u(t) = 0$.) The matrix in (2.40) is known as a *(right) companion* matrix.

Lemma 2.5.6 (Companion matrix). *The right companion matrix*

$$A = \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix}$$

has characteristic polynomial

$$\det(sI - A) = s^n + p_{n-1}s^{n-1} + \cdots + sp_1 + p_0. \tag{2.41}$$

Proof. See Exercise 2.7. ■

Lemma 2.5.7 (Asymptotic stability). *The system $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ is asymptotically stable if and only if $P(\lambda)$ is an asymptotically stable polynomial.* □

Proof. If $P(\lambda)$ is asymptotically stable, then $\dot{x} = Ax$ is asymptotically stable, and, hence also $y(t) = x_n(t)$ converges to zero for $t \to \infty$.

If $P(\lambda)$ is not asymptotically stable, then there is a zero λ_0 with a nonnegative real part, and then $y(t) = \mathrm{e}^{\lambda_0 t}$ satisfies $P(\frac{\mathrm{d}}{\mathrm{d}t})y = 0$ yet does not converge to zero for $t \to \infty$. So then $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ is not asymptotically stable. ■

Remarkably a very similar result holds for *systems* of differential equations. For example, if $y : \mathbb{R} \to \mathbb{R}^{n_y}$ and $u : \mathbb{R} \to \mathbb{R}^{n_u}$ are vectors of signals, and satisfy a differential equation, say,

$$\begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t} + 1 & -1 \\ 2 & \frac{\mathrm{d}}{\mathrm{d}t} + 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t} + 1 \\ -2 \end{bmatrix} u,$$

then the claim is that all possible solutions y_1, y_2 of the homogeneous equation

$$\begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t}+1 & -1 \\ 2 & \frac{\mathrm{d}}{\mathrm{d}t}+3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0$$

converge to zero if and only if the *determinant* of the corresponding polynomial matrix

$$P(s) := \begin{bmatrix} s+1 & -1 \\ 2 & s+3 \end{bmatrix}$$

is asymptotically stable. In this example, we have

$$\det(P(s)) = (s+1)(s+3)+2 = s^2+4s+5,$$

and this is indeed asymptotically stable. This is the result:

Lemma 2.5.8 (Asymptotic stability). *Let P and Q be real polynomial matrices with P square, and suppose P and Q have the same number of rows. Then the system of differential equations*

$$P(\tfrac{\mathrm{d}}{\mathrm{d}t})y = Q(\tfrac{\mathrm{d}}{\mathrm{d}t})u$$

is asymptotically stable if and only if $\det(P)$ *is asymptotically stable.*

Proof. Let m denote the number of rows of P.

Suppose $\det(P)$ is not asymptotically stable. Then $\det(P(s_0)) = 0$ for some $s_0 \in \mathbb{C}$ with $\mathrm{Re}(s_0) \geq 0$. Let $v \in \mathbb{C}^m$ be a nonzero vector such that $P(s_0)v = 0$. It is easy to check that then $y(t) := v\mathrm{e}^{s_0 t}$ satisfies $P(\frac{\mathrm{d}}{\mathrm{d}t})y = 0$. Since $\mathrm{Re}(s_0) \geq 0$ it follows that $y(t)$ does not converge to zero, hence the system is not asymptotically stable.

Next suppose $\det(P)$ is asymptotically stable. We are going to cheat a little, in that we assume sufficient smoothness of solutions y of $P(\frac{\mathrm{d}}{\mathrm{d}t})y = 0$. Let R be the *adjugate* of P. This adjugate is also a polynomial matrix, and it has the property that

$$RP = \det(P)I.$$

If $P(\frac{\mathrm{d}}{\mathrm{d}t})y = 0$ then also $\det(P)Iy = R(\frac{\mathrm{d}}{\mathrm{d}t})P(\frac{\mathrm{d}}{\mathrm{d}t})y$ is zero. Written out this says that

$$\begin{bmatrix} \det(P(\frac{\mathrm{d}}{\mathrm{d}t})) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \det(P(\frac{\mathrm{d}}{\mathrm{d}t})) \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = 0.$$

Therefore every component y_i satisfies the differential equation $\det(P(\frac{\mathrm{d}}{\mathrm{d}t}))y_i = 0$. Since $\det(P)$ is asymptotically stable this implies, via Lemma 2.5.7, that $\lim_{t\to\infty} y_i(t) = 0$ for every component y_i of y. The system hence is asymptotically stable. ■

Example 2.5.9 (Spring-mass system). The output $y := \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}^{\mathrm{T}}$ of the spring-mass system of Example 1.9.1 for $F = 0$ satisfies

$$P(\tfrac{\mathrm{d}}{\mathrm{d}t})y = 0$$

with

$$P(s) = \begin{bmatrix} m_1 s^2 + k_1 + k_2 & 0 & -k_2 \\ 0 & m_2 s^2 + k_3 + k_4 & -k_4 \\ -k_2 & -k_4 & m_3 s^2 + k_2 + k_4 \end{bmatrix}.$$

Suppose, for simplicity, that all masses are equal, $m := m_1 = m_2 = m_3$, and that all spring constants are also equal, $k := k_1 = k_2 = k_3 = k_4$. Then $P(s)$ becomes

$$P(s) = \begin{bmatrix} \Omega(s) & 0 & -k \\ 0 & \Omega(s) & -k \\ -k & -k & \Omega(s) \end{bmatrix}, \quad \Omega(s) := ms^2 + 2k.$$

The determinant of $P(s)$ is

$$\det P(s) = \Omega(s)(\Omega^2(s) - 2k^2).$$

Now, $\Omega(\lambda) = m\lambda^2 + 2k$ has two imaginary zeros, $\lambda = \pm\mathrm{i}\sqrt{2k/m}$, so these are also the zeros of $\det P(s)$. The system is therefore not asymptotically stable (which was to be expected). □

2.6 Exercises

2.1 Comprehension questions (on the whole chapter). Prove or give a counterexample.

(a) There exists an $A \in \mathbb{R}^{n\times n}$ such that

$$\mathrm{e}^{At} = \begin{bmatrix} \mathrm{e}^t & \mathrm{e}^t \\ 0 & \mathrm{e}^{-t} \end{bmatrix}.$$

(b) There exists an $A \in \mathbb{R}^{n\times n}$ such that

$$\mathrm{e}^{At} = \begin{bmatrix} \mathrm{e}^t & t\mathrm{e}^t \\ 0 & \mathrm{e}^{-t} \end{bmatrix}.$$

(c) If $A \in \mathbb{R}^{n\times n}$ is invertible, then 0 is the only equilibrium point of $\dot{x} = Ax$.

(d) If $A \in \mathbb{R}^{n\times n}$ is singular, then $\dot{x} = Ax$ has infinitely many equilibrium points.

(e) If $A \in \mathbb{R}^{n\times n}$ is singular, then no equilibrium point of $\dot{x} = Ax$ is stable.

(f) Let $A \in \mathbb{R}^{n\times n}$, and let $v \in \mathbb{R}^n$ be a constant signal. If x^* is an equilibrium point of $\dot{x} = Ax + v$, then the equilibrium point is asymptotically stable if and only if $\dot{z} = Az$ is asymptotically stable.

(g) If $\dot{x}(t) = f(x(t))$ is asymptotically stable, then its linearization is also asymptotically stable.

2.2 Let

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

(a) Determine e^{At} using the definition.

(b) Determine the general solution $(x_1(t), x_2(t), x_3(t))$ of

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = x_3, \quad \dot{x}_3 = 0$$

using part (a) and verify your answer.

2.3 Determine e^{At} for the following matrices:

(a) $A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

(c) $A = \begin{bmatrix} -8 & 8 \\ -15 & 14 \end{bmatrix}$

(d) $A = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$

(e) $A = \begin{bmatrix} 0 & 2 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

(f) $A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$

(g) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

2.4 Determine e^{At} for $A = \begin{bmatrix} 4 & -2 \\ 3 & -1 \end{bmatrix}$ and verify that $\frac{d}{dt} e^{At} = A e^{At}$.

2.5 Determine e^{At} for the antidiagonal $n \times n$ matrix with ones on the antidiagonal,

$$A = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & .\cdot^{\cdot} & .\cdot^{\cdot} & 0 \\ 0 & .\cdot^{\cdot} & .\cdot^{\cdot} & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

[Hint: Use the definition of the matrix exponential, and realize that $\cosh(t) := \frac{1}{2}(e^t + e^{-t})$ and $\sinh(t) := \frac{1}{2}(e^t - e^{-t})$.]

2.6 *Commuting matrices.* In which step of the proof of Lemma 2.2.32 do we use the assumption that $AF = FA$?

2.7 *Companion matrix.* Consider the *(right) companion matrix*

$$A = \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

This is a square matrix with a one under each diagonal element and zeros elsewhere, except in the last column.

(a) Show that the characteristic polynomial $\chi_A(s)$ is equal to

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

(b) Let λ be an eigenvalue of A. Show that $v \in \mathbb{C}^{1 \times n}$ is a corresponding *left* eigenvector if-and-only-if

$$v = c \begin{bmatrix} 1 & \lambda & \lambda^2 & \cdots & \lambda^{n-1} \end{bmatrix}$$

for some nonzero constant c.

(c) Let A be the companion matrix

$$A = \begin{bmatrix} 0 & 0 & -6 \\ 1 & 0 & -11 \\ 0 & 1 & -6 \end{bmatrix}.$$

Determine e^{At} using parts (a) and (b).

2.8 *Matrix exponential of a Jordan block.* Prove equation (2.18).

2.9 *Stable spiral.* Consider part 4 of Example 2.2.7.

(a) Show that after the transformation $z = \begin{bmatrix} v_{\mathrm{re}} & -v_{\mathrm{im}} \end{bmatrix}^{-1} x$, the system $\dot{x} = Ax$ is given by $\dot{z} = \Lambda z$ with

$$\Lambda = \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix}.$$

(b) Determine $\mathrm{e}^{\Lambda t}$ for Λ as above. (Hint: Use that $\left[\begin{smallmatrix} \mu & 0 \\ 0 & \mu \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & -\omega \\ \omega & 0 \end{smallmatrix}\right]$ commute.)

(c) Show that $\|\mathrm{e}^{\Lambda t} z_0\| = \mathrm{e}^{\mu t}\|z_0\|$, where $\|\cdot\|$ is the usual Euclidean norm. (This shows that for $\mu < 0$, the solutions z of $\dot{z} = \Lambda z$ are all strictly decreasing in the sense that $\|z(t)\|$ is strictly monotonically decreasing.)

2.10 *Stability.* Study the asymptotic stability of the system $\dot{x} = Ax$ for the following matrices:

(a) $A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$

(c) $A = \begin{bmatrix} -2 & 1 \\ 2 & -2 \end{bmatrix}$

(d) $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

(e) $A = \begin{bmatrix} 0 & A_{12} \\ A_{21} & 0 \end{bmatrix}$ with $A_{ij} \in \mathbb{R}^{n_i \times n_j}$

(f) $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ with $A_{ij} \in \mathbb{R}^{n_i \times n_j}$

2.11 Is $P(\lambda) = \lambda^5 + 2\lambda^4 + 3\lambda^2 + 4\lambda + 5$ asymptotically stable?

An extension of Thm. 2.3.6 is as follows: *if* the first column of the Routh table does not contain zeros, then the polynomial does not have imaginary zeros, and the number of sign changes in the first column of the table – as you go from the first entry down to the last entry – is equal to the number of unstable zeros. (A zero $\lambda \in \mathbb{C}$ is *unstable* if $\operatorname{Re}(\lambda) \geq 0$).

How many unstable zeros does $\lambda^5 + 2\lambda^4 + 3\lambda^2 + 4\lambda + 5$ have?

2.12 *Routh–Hurwitz for second- and third-degree polynomials.*

(a) Prove that a general degree two polynomial $a\lambda^2 + b\lambda + c$ (with $a \neq 0$) is asymptotically stable if and only if a, b, c all have the same sign.

(b) Consider a general degree three polynomial

$$a\lambda^3 + b\lambda^2 + c\lambda + d \qquad \text{with } a \neq 0.$$

Under what conditions on a, b, c, d is this polynomial asymptotically stable?

2.13 *Stationary solution.* Consider (2.4) and assume that it is asymptotically stable. If $u(t)$ is a constant signal, $u(t) = u_\infty$, then $y(t)$ also converges to a constant signal $y_\infty = \lim_{t\to\infty} y(t)$. Express y_∞ in terms of A, B, C, D, u_∞ (without using integrals).

2.14 Consider the nonlinear system

$$\dot{x}_1 = x_1^2 - x_1^4.$$

(a) Determine all equilibrium points.

(b) Determine the linearization around each equilibrium point.

(c) For each *linearization*, indicate whether it is asymptotically stable.

2.15 Consider the nonlinear system

$$\begin{aligned} \dot{x}_1 &= (x_1^2 + x_2^2 - 1)(x_1 + x_2), \\ \dot{x}_2 &= (x_1^2 + x_2^2 - 1)(-x_1 + x_2). \end{aligned} \tag{2.42}$$

(a) Determine all equilibrium points.

(b) Linearize the system around the equilibrium point $x^* = (0,0)$.

(c) Is this linearization asymptotically stable?

(d) Is the nonlinear system (2.42) asymptotically stable around the equilibrium point $x^* = (0,0)$?

(e) Is the nonlinear system (2.42) asymptotically stable around the equilibrium point $x^* = (1,0)$?

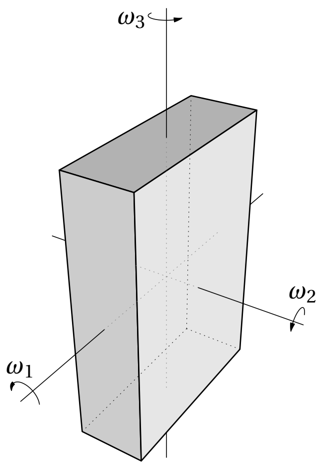


FIGURE 2.11: Tumbling domino tiles.

2.16 Consider the nonlinear system of the tumbling domino tile; see Figure 2.11. The equations of motion are

$$\dot{\omega}_1 = \tfrac{I_2 - I_3}{I_1}\omega_2\omega_3,$$
$$\dot{\omega}_2 = \tfrac{I_3 - I_1}{I_2}\omega_3\omega_1,$$
$$\dot{\omega}_3 = \tfrac{I_1 - I_2}{I_3}\omega_1\omega_2.$$

Here ω_k is the angular velocity about the kth axis, and I_k is the moment of inertia of the mass with respect to the kth axis.

(a) Determine all equilibrium points.

(b) Linearize the system around the equilibrium point $(\omega_1^*, \omega_2^*, \omega_3^*) = (1, 0, 0)$.

(c) Assume $0 < I_3 < I_1 < I_2$. Is the linearization in part (b) asymptotically stable?

(d) Assume $0 < I_3 < I_1 < I_2$. Is the nonlinear system asymptotically stable around the equilibrium point $(\omega_1^*, \omega_2^*, \omega_3^*) = (1, 0, 0)$?

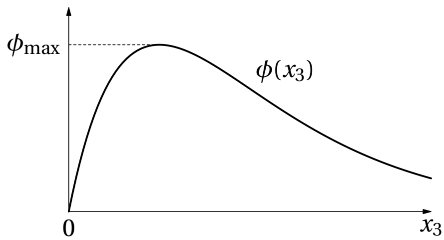


FIGURE 2.12: A potential milling function ϕ.

2.17 *Milling.* Consider the milling process of Example 1.8.7.

(a) Deduce that the function ϕ (the "milling function") is qualitatively of the form indicated in Figure 2.12.

(b) At a constant influx u^* of material, the equilibrium points (x_1^*, x_2^*, x_3^*) of the milling process is given by

$$\begin{aligned} x_1^* &= \frac{u^*}{\gamma_1}, \\ x_2^* &= \frac{\alpha u^*}{\gamma_2(1-\alpha)}, \\ \phi(x_3^*) &= \frac{u^*}{(1-\alpha)}. \end{aligned}$$

So there are two equilibrium points if

$$u^* < (1-\alpha)\phi_{\max}.$$

Study the stability of both equilibrium points. (It turns out that the equilibrium point on the right is unstable. In industrial terminology, this is called *mill plugging*.)

2.18 *Coupled pendulum.* Two pendula, each with a mass m, are connected by a spring with spring constant k; see Figure 2.13. The motion is describe approximately by

$$\begin{aligned} m\ell^2\ddot{\phi}_1 &= -mg\ell\sin(\phi_1) - k\ell^2(\sin(\phi_1) - \sin(\phi_2)), \\ m\ell^2\ddot{\phi}_2 &= -mg\ell\sin(\phi_2) + k\ell^2(\sin(\phi_1) - \sin(\phi_2)). \end{aligned}$$

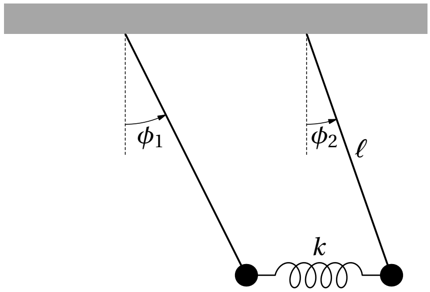


FIGURE 2.13: Coupled pendulum.

(a) Write the system in the form $\dot{x} = f(x)$ with $x = (\phi_1, \dot{\phi}_1, \phi_2, \dot{\phi}_2)$.

(b) Linearize these two equations around $(\phi_1, \dot{\phi}_1, \phi_2, \dot{\phi}_2) = (0,0,0,0)$ and write the linearize state equation with $\delta_x = (\phi_1, \dot{\phi}_1, \phi_2, \dot{\phi}_2)$.

2.19 Consider the nonlinear system

$$\dot{x}_1(t) = x_2(t),$$
$$\dot{x}_2(t) = -x_1(t) - x_2^2(t) + u(t).$$

(a) Show that $u(t) = \cos^2(t)$, $x_1(t) = \sin(t)$, $x_2(t) = \cos(t)$ is a solution of this system.

(b) Determine the linearization of the system along this solution.

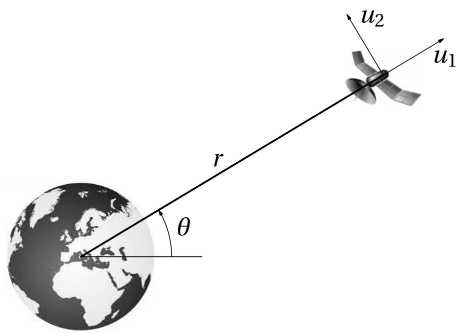


FIGURE 2.14: Satellite orbiting the earth.

2.20 *Satellite and linearization.* Consider a satellite orbiting the earth. We take earth to be a perfect ball, and assume that we have a satellite with mass $m = 1$. We decompose the force of the jets propelling the satellite into a component u_1 in the radial direction and a component u_2 in the tangential direction. Newton's law gives

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

(a) Give a nonlinear state representation $\dot{x} = f(x)$ of the system with state $x = \left[r, \dot{r}, \theta, \dot{\theta}\right]^{\mathrm{T}}$.

(b) This system has an equilibrium *trajectory*

$$\begin{aligned} x_1(t) &= R, \\ x_2(t) &= 0, \\ x_3(t) &= \Omega t, \\ x_4(t) &= \Omega, \\ u_1 = u_2 &\equiv 0. \end{aligned}$$

Determine the relation between the radius R, the angular velocity Ω, and the gravitational acceleration g.

(c) Linearize this system around the equilibrium trajectory.

(d) Determine the eigenvalues of the linearized subsystem (expressed in terms of Ω).

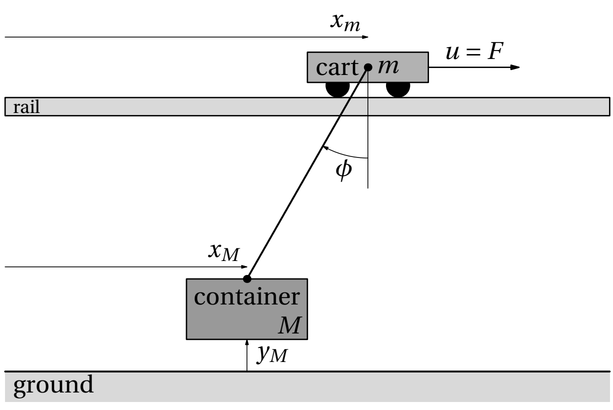


FIGURE 2.15: Container transfer.

2.21 *Container transfer.* In ports such as the one in Rotterdam, hundreds of containers are transferred to and from ships daily. Figure 2.15 shows the situation schematically. A cart with mass m can be pulled without friction over a fixed rail. To this cart is fixed a cable of length L, from which a container hangs. The mass M of the container is in general much greater than that of the cart. By exerting a suitable force F on the cart, we can try to move the container M from the leftmost position (the ship) to the rightmost position (the quay).

As input, we take $u := F$. After the necessary modeling (which we will spare you), it turns out that we can describe the system by the system of equations

$$\begin{cases} x_M(t) = x_m(t) - L\sin(\phi(t)), \\ y_M(t) = L - L\cos(\phi(t)), \\ m\ddot{x}_m(t) = F(t) - \sin(\phi(t))b(t), \\ M\ddot{x}_M(t) = \sin(\phi(t))b(t), \\ M\ddot{y}_M(t) = \cos(\phi(t))b(t) - Mg. \end{cases} \tag{2.43}$$

Here g is the gravitational acceleration, $g \approx 9.81\frac{\mathrm{m}}{\mathrm{s}^2}$. The term $b(t)$ is the tension in the cable, which is assumed to be the same everywhere in the cable. The tension can be eliminated from these equations, and by choosing

$$x(t) := \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} := \begin{bmatrix} x_m(t) \\ \dot{x}_m(t) \\ \phi(t) \\ \dot{\phi}(t) \end{bmatrix},$$

the implicit representation (2.43) can be changed into a nonlinear state representation

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ \frac{F - MLx_4^2\sin(x_3) - Mg\sin(x_3)\cos(x_3)}{m + M\sin^2(x_3)} \\ x_4 \\ \frac{F\cos(x_3) - MLx_4^2\sin(x_3)\cos(x_3) - (m+M)g\sin(x_3)}{L(m + M\sin^2(x_3))} \end{bmatrix}. \tag{2.44}$$

(a) Write (2.44) in the form

$$\dot{x} = \mathscr{L}(x, F) + R(x, F)$$

for some linear map $\mathscr{L}$, where R is a "little o" function. [Hint: Use Taylor series expansions of cos and sin, and use that $\frac{1}{1+R(s)} = 1 + R(s)$ in the sense of little o functions.]

(b) Linearize (2.44) around $x^* = 0$, $F^* = 0$. [Hint: Do this without differentiating.]

2.22 *Differentiated and integrated outputs.* Consider (2.4) with $D = 0$.

(a) Determine a state representation for the system with input u and output $z := \dot{y}$.

(b) Determine a state representation for the system with input u and output z, where z is an arbitrary signal satisfying $\dot{z} = y$.

2.23 *Equivalent representation.* Let $x:\mathbb{R}^n \to \mathbb{R}^n$ and $\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] \in \mathbb{R}^{(n+n_y)\times(n+n_u)}$. Show that

$$\mathfrak{B} = \{\left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] : \mathbb{R} \to \mathbb{R}^{n_u+n_y} \mid \exists x, \left[\begin{smallmatrix} \dot{x} \\ y \end{smallmatrix}\right] = \left[\begin{smallmatrix} T^{-1}AT & T^{-1}B \\ CT & D \end{smallmatrix}\right]\left[\begin{smallmatrix} x \\ u \end{smallmatrix}\right]\}$$

is independent of $T \in \mathbb{R}^{n\times n}$ (provided that it is invertible).

2.24 *Equivalent representation.* Consider the external behavior

$$\mathfrak{B}_j = \{\left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] : \mathbb{R} \to \left[\begin{smallmatrix} \mathbb{R} \\ \mathbb{R} \end{smallmatrix}\right] \mid \exists x, \left[\begin{smallmatrix} \dot{x} \\ y \end{smallmatrix}\right] = \left[\begin{smallmatrix} A_j & B_j \\ C_j & D_j \end{smallmatrix}\right]\left[\begin{smallmatrix} x \\ u \end{smallmatrix}\right]\}$$

for

- $A_1 = \left[\begin{smallmatrix} -1 & 0 \\ 0 & -2 \end{smallmatrix}\right]$, $B_1 = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, and $C_1 = \left[\begin{smallmatrix} 1 & 0 \end{smallmatrix}\right]$
- $A_2 = \left[\begin{smallmatrix} -2 & 0 \\ 0 & -1 \end{smallmatrix}\right]$, $B_2 = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$, and $C_2 = \left[\begin{smallmatrix} 0 & 1 \end{smallmatrix}\right]$
- $A_3 = \left[\begin{smallmatrix} -1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $B_3 = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, and $C_3 = \left[\begin{smallmatrix} 0 & 1 \end{smallmatrix}\right]$
- $A_4 = \left[\begin{smallmatrix} -1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $B_4 = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, and $C_4 = \left[\begin{smallmatrix} 1 & 0 \end{smallmatrix}\right]$
- $A_5 = -1$, $B_5 = 1$, and $C_5 = 1$

(a) Which state representations are isomorphic?

(b) Which among $\mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{B}_3, \mathfrak{B}_4, \mathfrak{B}_5$ are equal?

2.25 *State representations.* Give state representations of $\dot{y}+y=z$, $\dot{z}+z=v$, $\dot{v}+v=u$ with input u and output y.

2.26 Determine state representations of the following systems:

(a) $y^{(1)} + 4y = 2u$

(b) $y^{(1)} + 4y = u^{(1)} + 2u$

(c) $\dot{y} + 2y = \dot{u} - 3u$

(d) $3y^{(3)} + 2y^{(2)} + y^{(1)} = u^{(2)}$

(e) $y^{(4)} = u$

(f) $y^{(3)} = u^{(3)}$

2.27 Consider

$$\begin{cases} \dot{x} &= Ax + Bu, \\ y &= Cx + u. \end{cases}$$

Show that y can also be chosen as input (and u as output) and determine the corresponding state representation

$$\begin{cases} \dot{x} &= \cdots x + \cdots y, \\ u &= \cdots x + \cdots y. \end{cases}.$$

2.28 *Asymptotic stability.* Check the asymptotic stability of $P(\frac{\mathrm{d}}{\mathrm{d}t})y = 0$ for the following P:

(a) $P(s) = s^2 + 2s + 1$

(b) $P(s) = s^2 + 2s - 3$

(c) $P(s) = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$

(d) $P(s) = \begin{bmatrix} 1 & s \\ s & 1 \end{bmatrix}$

(e) $P(s) = \begin{bmatrix} 1 & s \\ s+1 & s^2+s \end{bmatrix}$

(f) $P(s) = \begin{bmatrix} s^2 - 2s + 2 & -s+1 \\ 3s-2 & -2 \end{bmatrix}$

Tougher Exercises

2.29 *Differential equations that do not admit a state representation.*

(a) Explain that $y = \dot{u}$ does not admit a state representation. [Hint: Find a continuous u for which y is not continuous, and explain that this is not possible in state representations.]

(b) Explain that $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ does not admit a state representation if the degree of Q is greater than that of P.

(c) Prove that $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ admits a state representation if and only if the degree of Q is less than or equal to that of P.

2.30 *Monotonically decreasing state*

(a) Prove that $\frac{d}{dt}\|x(t)\|^2 < 0$ for all solutions $x(t) \neq 0$ of $\dot{x} = Ax$ if and only if the matrix $A + A^{\mathrm{T}}$ has only negative eigenvalues.

(b) Prove that all eigenvalues of $A \in \mathbb{R}^{n\times n}$ have negative real part if $A + A^{\mathrm{T}}$ has only negative eigenvalues.

2.31 *Smoothness of the output.* Consider (2.4).

(a) Under what conditions on A, B, C, D is y continuous for every bounded u?

(b) Under what conditions on A, B, C, D is y continuously differentiable for every bounded u?

(c) Let $k \in \mathbb{N}$. Under what conditions on A, B, C, D is y at least k times continuously differentiable for every bounded u?

2.32 *Alternating system.* Assume that we have two matrices $A_1, A_2 \in \mathbb{R}^{n\times n}$, and consider the alternating system $\dot{x}(t) = A(t)x(t)$ with $A(t) \in \mathbb{R}^{n\times n}$ a time-varying matrix that switches between A_1 and A_2, that is,

$$\text{for all } t, \text{ either } A(t) = A_1 \text{ or } A(t) = A_2. \tag{2.45}$$

(a) Give an example of two asymptotically stable systems $\dot{x} = A_1 x$ and $\dot{x} = A_2 x$ for which $\dot{x}(t) = A(t)x(t)$ is not stable. [Hint: Take $n = 2$ and let the points where the system switches depend on $x(t)$.]

(b) Assume that A_1 and A_2 commute. Show that $\dot{x} = A(t)x$ is asymptotically stable for all $A(t)$ of the form (2.45) if and only if both $\dot{x} = A_1 x$ and $\dot{x} = A_2 x$ are asymptotically stable.

2.33 Let A be a 2×2 matrix with distinct eigenvalues $\lambda_1 \neq \lambda_2$. We can write A as

$$A = \lambda_2 P_1 + \lambda_1 P_2,$$

for

$$P_1 := \frac{1}{\lambda_2 - \lambda_1}(A - \lambda_1 I),$$
$$P_2 := \frac{1}{\lambda_1 - \lambda_2}(A - \lambda_2 I).$$

(a) Show that $P_1 P_2 = P_2 P_1$.

(b) Show that $P_1^2 = P_1$ en $P_2^2 = P_2$.

(c) Show that

$$\mathrm{e}^{At} = \mathrm{e}^{\lambda_2 t} P_1 + \mathrm{e}^{\lambda_1 t} P_2.$$

(d) Find a similar method to determine e^{At} when the eigenvalues coincide.

2.34 *Schur–Cohn–Jury criterion.* In this exercise, $p(s)$ is a real polynomial. We know that a system $p(\frac{\mathrm{d}}{\mathrm{d}t})y = 0$ is asymptotically stable if and only if all zeros of $p(\lambda) = 0$ have negative real part. This seems to imply that asymptotic stability can only be determined by computing the zeros. That is not the case. A simple test suffices:

(a) Show that the coefficients of an asymptotically stable polynomial $p(s)$ all have the same sign.

(b) Show that a nonconstant polynomial $p(s)$ is asymptotically stable only when $|p(-1)/p(1)| < 1$. [*Hint:* Use part (a).]

(c) Prove that a nonconstant polynomial $p(s)$ is asymptotically stable if and only if $|p(-1)/p(1)| < 1$ and

$$q(s) := p(s) - \frac{p(-1)}{p(1)} p(-s)$$

is asymptotically stable.

[*Hint:* Study the convex combinations defined as $r_\eta := (1-\eta)p+\eta q$ and show that all these convex combinations have the same degree and the same imaginary zeros if $|p(-1)/p(1)| < 1$.]

The good news is that $q(-1) = 0$, so the problem can be reduced to that of the polynomial $q(s)/(s+1)$ (of lower degree). Continuing this way, the problem is solved in n steps (with $n = \deg(p)$).

2.35 *Discretized systems.* We are given the continuous-time system (2.4). Assume that the input u is piecewise constant of the form

$$u(t) = u(kT) \quad \text{for } kT \le t < (k+1)T, \quad k \in \mathbb{Z}.$$

When we are only interested in the values of the output y at times kT, that is,

$$\tilde{y}[k] := y(kT), \qquad k \in \mathbb{Z},$$

it suffices to study the discrete-time system

$$\begin{aligned} \tilde{x}[k+1] &= F\tilde{x}[k] + G\tilde{u}[k], \\ \tilde{y}[k] &= C\tilde{x}[k] + D\tilde{u}[k]. \end{aligned}$$

(a) Show this and express F and G in terms of A and B.

(b) Determine F and G for $A = \left[\begin{smallmatrix} -2 & 0 \\ 0 & -3 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$.

Chapter 3

Controllability and Observability

In Chapter 2, we spent much time analyzing the autonomous system $\dot{x} = Ax$, in other words, the system $\dot{x} = Ax + Bu$ with input equal to zero, $u \equiv 0$. However, in applications we can often choose u any way we want (within reasonable limits), and thereby direct the behavior of x to some degree. This is called *controllability*. For example, it is because of the driver that a car does not go about its business autonomously but rather follows a trajectory determined by the driver.

Another fundamental notion from systems theory, which is closely connected to controllability, is *observability*. In order to steer the car successfully, we must of course keep our eyes and ears open. The question is then what we must watch, and what we must listen to. Is what we see and hear even sufficient to keep the car on the road? The systems theory abstraction of this idea begins with the question whether we can reconstruct, or *observe*, the internal variables (the state x) based on only the external variables (u, y).

3.1 Reachability

Consider the initially-at-rest system

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = 0 \tag{3.1}$$

and recall that the state $x(t)$ can be expressed explicitly as

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

By reachability, we mean the possibility to reach, from the resting position $x(0) = 0$, any arbitrary state using a well-chosen input signal.

Definition 3.1.1 (Reachability). A system $\dot{x} = Ax + Bu$ is *reachable* if for every $x_1 \in \mathbb{R}^n$ and $x(0) = 0$, there is a $t_1 > 0$ and an input $u : [0, t_1] \to \mathbb{R}^{n_u}$ such that $x(t_1) = x_1$. □

We also say that "the pair (A, B)" is reachable. Reachability says that *eventually*, we can reach any desired state from $x(0) = 0$. It does not say that we should be able to do so during a previously imposed time horizon t_1, just that such a finite horizon t_1 exists (possibly depending on x_1).

Example 3.1.2 (Reachability). A simple example of an unreachable system is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u. \tag{3.2}$$

Because the A-matrix is diagonal and $u(t)$ can only influence the second state component x_2, it will be clear that x_1 cannot be influenced (controlled) by u. This system is therefore not reachable. Reachability is more difficult to analyze for the system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u. \tag{3.3}$$

Here too, u cannot influence the first state component x_1 *directly*, but the difference with the previous system is that x_2 can influence the first component x_1, and because x_2 in turn can be influenced by u, it might still be possible to sent x_1 to a desired value. But if that is possible, can it also be done in such a way that, at the same time, x_2 is sent to a (different) desired value? □

We first study which states $x(t_1)$ can be reached for a given time horizon $t_1 > 0$. We denote the set of possible states $x(t_1)$ by $\mathbb{X}(t_1)$:

$$\mathbb{X}(t_1) = \left\{ \int_0^{t_1} \mathrm{e}^{A(t_1 - \tau)} B u(\tau) \, \mathrm{d}\tau \;\middle|\; u : [0, t_1] \to \mathbb{R}^{n_u} \right\}. \tag{3.4}$$

It is a subspace of $\mathbb{R}^n$ (see Exercise 3.2).

Lemma 3.1.3 (Unreachable states). *Let $t_1 > 0$ and $\eta \in \mathbb{R}^n$. The following four statements are equivalent:*

1. $\eta \perp \mathbb{X}(t_1)$*; that is,* $\eta^{\mathrm{T}} x(t_1) = 0$ *for all possible* $x(t_1)$.
2. $\eta^{\mathrm{T}} \mathrm{e}^{At} B = 0$ *for all* $t \in [0, t_1]$.
3. $\eta^{\mathrm{T}} A^k B = 0$ *for all* $k = 0, 1, \ldots$.
4. $\eta^{\mathrm{T}} \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} = 0$.

Proof. We prove (1) $\implies$ (2) $\implies$ (3) $\implies$ (4) $\implies$ (1).

(1) $\implies$ (2): We have $\eta^{\mathrm{T}} x(t_1) = 0$ for all u. This holds, in particular, for $u(\tau) = (\eta^T \mathrm{e}^{A(t_1-\tau)} B)^{\mathrm{T}}$. For this input, we have

$$\begin{aligned}
0 &= \eta^{\mathrm{T}} x(t_1) \\
&= \int_0^{t_1} \eta^{\mathrm{T}} \mathrm{e}^{A(t_1-\tau)} B u(\tau)\, \mathrm{d}\tau \\
&= \int_0^{t_1} \|\eta^{\mathrm{T}} \mathrm{e}^{A(t_1-\tau)} B\|^2\, \mathrm{d}\tau.
\end{aligned}$$

This implies that $\eta^{\mathrm{T}} \mathrm{e}^{At} B = 0$ for all $t = t_1 - \tau \in [0, t_1]$.

(2) $\implies$ (3). Differentiating the equality $\eta^{\mathrm{T}} \mathrm{e}^{At} B = 0$ a number of times gives $\eta^{\mathrm{T}} A^k \mathrm{e}^{At} B = 0 \forall k = 0, 1, 2, \ldots$. For $t = 0$, this says that $\eta^{\mathrm{T}} A^k B = 0$.

(3) $\implies$ (4). Trivial.

(4) $\implies$ (1). We use the Cayley–Hamilton theorem. This theorem says that a matrix A satisfies its own characteristic equation. That is, if

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

then $\chi_A(A) = 0$, that is,

$$A^n = -(p_{n-1}A^{n-1} + p_{n-2}A^{n-2} + \cdots + p_1 A + p_0 I).$$

So A^n is a linear combination of lower powers of A. But then A^{n+1} is also a linear combination of $I, A, \ldots, A^{n-1}$ because

$$\begin{aligned}
&A^{n+1} \\
&= AA^n \\
&= -A\big(p_{n-1}A^{n-1} + p_{n-2}A^{n-2} + \cdots + p_1 A + p_0 I\big) \\
&= -(p_{n-1}A^n + p_{n-2}A^{n-1} + \cdots + p_1 A^2 + p_0 A) \\
&= -p_{n-1}A^n + \text{linear combination of } A^{n-1}, \ldots, A \\
&= \text{linear combination of } A^{n-1}, \ldots, A, I.
\end{aligned}$$

Continuing this way, we see that every A^k is a linear combination of $I, A, A^2, \ldots, A^{n-1}$. Consequently, every $A^k B$ is a linear combination of $B, AB, A^2B, \ldots, A^{n-1}B$.

Now, if $\eta^{\mathrm{T}} \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} = 0$, then the Cayley–Hamilton theorem implies that $\eta^{\mathrm{T}} A^k B = 0$ for all $k \geq 0$, and consequently that for every input, we have

$$\begin{aligned}
\eta^{\mathrm{T}} x(t_1) &= \int_0^{t_1} \eta^{\mathrm{T}} \mathrm{e}^{A(t_1-\tau)} B u(\tau)\, \mathrm{d}\tau \\
&= \int_0^{t_1} \sum_{k=0}^{\infty} \eta^{\mathrm{T}} A^k B \frac{(t_1-\tau)^k}{k!} u(\tau)\, \mathrm{d}\tau \\
&= 0.
\end{aligned}$$

In other words, η is orthogonal to every element of $\mathbb{X}(t_1)$. ■

The set $\mathbb{X}(t_1)$ and the column space of the matrix $\begin{bmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{bmatrix}$ therefore have the same orthogonal complement. But as $\mathbb{X}(t_1)$ is a subspace of $\mathbb{R}^n$, this means that $\mathbb{X}(t_1)$ is *equal* to this column space:

$$\begin{aligned}\mathbb{X}(t_1) &= \operatorname{im}(\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}) \\ &:= \{Bu_0 + ABu_1 + \cdots + A^{n-1}Bu_{n-1} \mid u_i \in \mathbb{R}^{n_u}\}.\end{aligned}$$

Also, because this column space does not depend on t_1, the space $\mathbb{X}(t_1)$ is also independent of t_1 (provided $t_1 > 0$). Apparently, the states that are at all reachable, are reachable for every positive horizon $t_1 > 0$, regardless of how small it is. The matrix

$$\mathscr{C} := \begin{bmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{bmatrix} \tag{3.5}$$

is called the *controllability matrix*. It follows from the above that the set of reachable states $x(t)$ for $t > 0$ is equal to the column space

$$\operatorname{im}(\mathscr{C}).$$

This is a subspace of $\mathbb{R}^n$ and is called the *reachable subspace*. The system is thus reachable if and only if $\operatorname{im}(\mathscr{C}) = \mathbb{R}^n$. This is the case if and only if $\mathscr{C}$ has full row rank (i.e. rank n). The input u often consists of one element. In this case, B is a matrix with one column, and the controllability matrix $\mathscr{C}$ is therefore square. For square matrices, full row rank is equivalent to invertibility. Reachability can now be tested easily.

Example 3.1.4. The system (3.2) has state dimension $n = 2$ and input dimension $n_u = 1$. The controllability matrix $\mathscr{C}$ is then a 2×2 matrix, namely

$$\mathscr{C} = \begin{bmatrix} B & AB \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}.$$

This matrix is not invertible, hence does not have full row rank. The system (3.2) is therefore not reachable (as we had already deduced in Example 3.1.2). The controllability matrix of the system (3.3) is equal to

$$\mathscr{C} = \begin{bmatrix} B & AB \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

This matrix is invertible (and therefore has full row rank), and this system is therefore reachable. □

We summarize this reachability test in the following theorem. In this theorem, we also explicitly give an input that realizes the control objective.

Theorem 3.1.5 (Reachability). *Let $\mathscr{C}$ be the controllability matrix as in* (3.5)*. The following five statements are equivalent:*

1. (A, B) *is reachable.*

2. $\operatorname{im}(\mathscr{C}) = \mathbb{R}^n$.

3. *$\mathscr{C}$ has full row rank.*

4. *The* controllability Gramian $P(t)$*, defined as*

$$P(t) = \int_0^t e^{A\tau} B B^{\mathrm{T}} e^{A^{\mathrm{T}}\tau} \, d\tau,$$

is invertible for all $t > 0$.

5. *The* controllability Gramian $P(t)$ *is invertible for some* $t > 0$.

If these hold, then every state x_1 *is reachable for every positive horizon* $t_1 > 0$ *and*

$$u_*(t) := B^{\mathrm{T}} e^{A^{\mathrm{T}}(t_1 - t)} P^{-1}(t_1) x_1 \tag{3.6}$$

is one of the many inputs that realize this: $x(t_1) = x_1$. *Moreover, the (squared) norm of this* u_* *is*

$$\|u_*\|^2 := \int_0^{t_1} u_*^{\mathrm{T}}(t) u_*(t) \, dt = x_1^{\mathrm{T}} P^{-1}(t_1) x_1,$$

and no other input that realizes $x(t_1) = x_1$ *has a smaller norm.*

Proof. We prove (1) $\implies$ (2) $\implies$ (3) $\implies$ (4) $\implies$ (5) $\implies$ (1).

(1) $\implies$ (2) is trivial.

(2) $\implies$ (3): Proof by contradiction: If $\mathscr{C}$ has less than full row rank, then there exists an $\eta \in \mathbb{R}^n$ with $\eta \neq 0$ such that $\eta^{\mathrm{T}} \mathscr{C} = 0$. But then by Lemma 3.1.3, the nonzero η is orthogonal to the reachable subspace $\mathbb{R}^n$. This is impossible, so $\mathscr{C}$ has full row rank.

(3) $\implies$ (4): Let $t > 0$. We must prove that $P(t)\eta = 0$ implies $\eta = 0$. Suppose $P(t)\eta = 0$. We then also have $\eta^{\mathrm{T}} P(t) \eta = 0$ and therefore

$$\begin{aligned}
0 &= \eta^{\mathrm{T}} P(t) \eta \\
&= \int_0^t \eta^{\mathrm{T}} e^{A\tau} B B^{\mathrm{T}} e^{A^{\mathrm{T}}\tau} \eta \, d\tau \\
&= \int_0^t \|\eta^{\mathrm{T}} e^{A\tau} B\|^2 \, d\tau.
\end{aligned}$$

This is only possible if $\eta^{\mathrm{T}} e^{A\tau} B$ is identical to the zero function (on $[0, t]$). As in Lemma 3.1.3, it follows that $\eta^{\mathrm{T}} A^k B = 0$ for all $k \geq 0$ and therefore, by the same lemma, that $\eta^{\mathrm{T}} \mathscr{C} = 0$. Since $\mathscr{C}$ has full row rank, this can only happen if $\eta = 0$.

(4) $\implies$ (5) is trivial.

(5) $\implies$ (1) follows by verifying that (3.6) holds (see Exercise 3.6).

In Exercise 3.6, you must show that $\|u_*\|^2 = x_1^{\mathrm{T}} P^{-1}(t_1) x_1$. It remains to prove that u_* has optimal norm. Suppose that u_1 is one of the inputs that achieve $x(t_1) = x_1$. It then follows by linearity that

$$\int_0^{t_1} e^{A(t_1 - \tau)} B [u_1(\tau) - u_*(\tau)] \, d\tau = x_1 - x_1 = 0.$$

Consequently,

$$\begin{aligned}
&\int_0^{t_1} u_*^{\mathrm{T}}(\tau)[u_1(\tau)-u_*(\tau)]\,\mathrm{d}\tau \\
&= \int_0^{t_1} x_1^{\mathrm{T}} P^{-1}(t_1)\mathrm{e}^{A(t_1-\tau)} B[u_1(\tau)-u_*(\tau)]\,\mathrm{d}\tau \\
&= x_1^{\mathrm{T}} P(t_1)^{-1} \int_0^{t_1} \mathrm{e}^{A(t_1-\tau)} B[u_1(\tau)-u_*(\tau)]\,\mathrm{d}\tau \\
&= 0.
\end{aligned}$$

(in the course *Linear Structures*, we would say that u_* and u_1-u_* are orthogonal to each other in a suitable inner product.) It thus follows that the norm of u_1 is at least that of u_*:

$$\begin{aligned}
&\|u_1\|^2 \\
&\quad = \|u_* + (u_1-u_*)\|^2 \\
&\quad = \int_0^{t_1} (u_* + (u_1-u_*))^{\mathrm{T}} (u_* + (u_1-u_*))\,\mathrm{d}\tau \\
&\quad = \int_0^{t_1} u_*^{\mathrm{T}} u_* + 2\underbrace{u_*^{\mathrm{T}}(u_1-u_*)}_{\text{integrates to 0}} + (u_1-u_*)^{\mathrm{T}}(u_1-u_*)\,\mathrm{d}\tau \\
&\quad = \|u_*\|^2 + \|u_1-u_*\|^2 \\
&\quad \geq \|u_*\|^2.
\end{aligned}$$

■

3.2 Controllability

Reachability is defined for systems with $x(0)=0$. For systems with an arbitrary initial state $x(0)=x_0$ we have the following analogous definition.

Definition 3.2.1 (Controllability). A system $\dot{x}=Ax+Bu$ is *controllable* if for every pair of states $x_0, x_1 \in \mathbb{R}^n$, and $x(0)=x_0$, there is a $t_1 \geq 0$ and an input $u:[0,t_1]\to\mathbb{R}^{n_u}$ such that $x(t_1)=x_1$. □

Controllability obviously implies reachability, but for our type of system, $\dot{x}=Ax+Bu$, reachability also implies controllability. Indeed, if the system is reachable, then there also exists an input u_* that sends $x(0)=0$ to $x(t_1)=x_1-\mathrm{e}^{At_1}x_0$. This u_* sends $x(0)=x_0$ to $x(t_1)=x_1$ because

$$x(t_1) = \mathrm{e}^{At_1}x_0 + \underbrace{\int_0^{t_1} \mathrm{e}^{A(t_1-\tau)} B u_*(\tau)\,\mathrm{d}\tau}_{x_1-\mathrm{e}^{At_1}x_0} = x_1.$$

In short, controllability and reachability are equivalent. From here on, we will usually call it controllability.

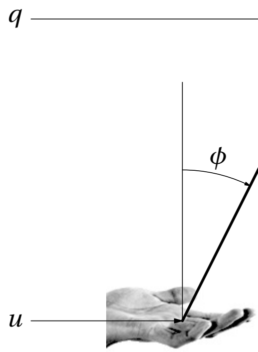


FIGURE 3.1: Inverted pendulum.

Example 3.2.2 (Juggler). Consider an ideal pendulum consisting of a mass m attached to a massless rigid stick of length ℓ_1, which can rotate freely in its pivot point (the palm of the hand); see Figure 3.1. Assume that we can set the position of the hand freely in one horizontal direction; this is the input variable u. We indicate the angle of the pendulum with respect to the vertical position by ϕ. Newton's second law gives the differential equation

$$\cos(\phi)\,\ddot{u} + \ell\ddot{\phi} = g\sin(\phi). \tag{3.7}$$

(The mass m does not play a role in the model[1].) Linearizing around $\phi = \dot{\phi} = 0$, $u = 0$ boils down to replacing $\sin(\phi)$ by ϕ and $\cos(\phi)$ by 1. This gives

$$\ddot{u} + \ell\ddot{\phi} = g\phi. \tag{3.8}$$

As state variables, we choose $q = u+\ell\phi$ and $v := \dot{q} = \dot{u}+\ell\dot{\phi}$. This q is the horizontal displacement of the top of the pendulum, and v is its velocity. With these choices, we can write (3.8) as the state representation

$$\begin{bmatrix}\dot{q}\\ \dot{v}\end{bmatrix} = \begin{bmatrix}0 & 1\\ \frac{g}{\ell} & 0\end{bmatrix}\begin{bmatrix}q\\ v\end{bmatrix} + \begin{bmatrix}0\\ -\frac{g}{\ell}\end{bmatrix}u. \tag{3.9}$$

The controllability matrix $\mathscr{C}$ is now given by

$$\mathscr{C} = \begin{bmatrix}0 & -\frac{g}{\ell}\\ -\frac{g}{\ell} & 0\end{bmatrix}. \tag{3.10}$$

Since $\det(\mathscr{C}) = g^2/\ell^2 \neq 0$, we have that (3.9) is controllable. In the neighborhood of the vertical position, we can therefore control both the *position* and the *velocity* of the pendulum by moving the pivot point. Not bad!

[1]This differential equation is derived in Appendix A.5

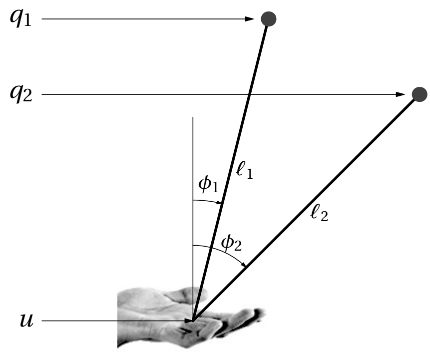


FIGURE 3.2: Two inverted pendula.

Now, assume that our juggler has *two* ideal pendula on his hand; see Figure 3.2. As above, we obtain linearized equations

$$\begin{aligned} \ddot{u} + \ell_1\ddot{\phi}_1 &= g\phi_1, \quad q_1 := u + \ell_1\phi_1 \\ \ddot{u} + \ell_2\ddot{\phi}_2 &= g\phi_2, \quad q_2 := u + \ell_2\phi_2 \end{aligned} \tag{3.11}$$

and the state representation

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{v}_1 \\ \dot{v}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{g}{\ell_1} & 0 & 0 & 0 \\ 0 & \frac{g}{\ell_2} & 0 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -\frac{g}{\ell_1} \\ -\frac{g}{\ell_2} \end{bmatrix} u. \tag{3.12}$$

Let $\alpha := -\frac{g}{\ell_1}, \beta := -\frac{g}{\ell_2}$. Then the controllability matrix is

$$\mathscr{C} = \begin{bmatrix} 0 & \alpha & 0 & -\alpha^2 \\ 0 & \beta & 0 & -\beta^2 \\ \alpha & 0 & -\alpha^2 & 0 \\ \beta & 0 & -\beta^2 & 0 \end{bmatrix}. \tag{3.13}$$

(You may want to verify this.) This controllability matrix has rank 4 if and only if the matrix $\left[\begin{smallmatrix} \alpha & -\alpha^2 \\ \beta & -\beta^2 \end{smallmatrix}\right]$ has rank 2, or, equivalently, $\alpha\beta^2 - \alpha^2\beta \neq 0$, that is, $\alpha \neq \beta$. Hence the system is controllable if $\ell_1 \neq \ell_2$ and uncontrollable if $\ell_1 = \ell_2$. It may not be surprising that the system is uncontrollable if the pendula have the same length. That the system *is* controllable if the lengths differ is less intuitive, in fact it is quite spectacular! □

3.3 Kalman Controllability Decomposition & the Hautus Test

Consider a system in state z (not x) and suppose it has the following structure,

$$\begin{bmatrix} \dot{z}_{\mathrm{c}} \\ \dot{z}_{\mathrm{uc}} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} z_{\mathrm{c}} \\ x_{\mathrm{uc}} \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u \tag{3.14}$$

with $z_c \in \mathbb{R}^q$ and $z_{uc} \in \mathbb{R}^{n-q}$, for some $q \leq n$. Because of the zero blocks in the lower left corner of the A-matrix and in the lower part of the B-matrix, it is intuitively clear that we cannot reach z_{uc} using u. Indeed, if $z_{uc}(0) = 0$, then it follows from $\dot{z}_{uc} = A_{22} z_{uc}$ that $z_{uc}(t) = 0$ for all t, regardless of the choice of u. This also follows from the reachable subspace, $\text{im}(\mathscr{C}_z)$, because its controllability matrix equals

$$\mathscr{C}_z = \begin{bmatrix} B_1 & A_{11}B_1 & \cdots & A_{11}^{n-1}B_1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{3.15}$$

(verify this yourself), and the reachable subspace $\text{im}(\mathscr{C}_z)$ therefore satisfies

$$\text{im}(\mathscr{C}_z) \subseteq \begin{bmatrix} \mathbb{R}^q \\ 0 \end{bmatrix}.$$

If z_{uc} begins in the origin, $z_{uc}(0) = 0$, then it stays in the origin, $z_{uc}(t) = 0 \forall t$, and then the system reduces to

$$\dot{z}_c = A_{11} z_c + B_1 u.$$

It is clear that the system (3.14) is uncontrollable if z_{uc} has at least one component (if $n - q > 0$). We will now show that using a state transformation $z = T^{-1}x$, *every* system can be written in the form above, and such that the subsystem $\dot{z}_c = A_{11} z_c + B_1 u$ *is* controllable.

Lemma 3.3.1 (Kalman controllability decomposition). *For every system $\dot{x} = Ax + Bu$, there exists a state transformation $z = T^{-1}x$ such that in the new coordinates, we have*

$$\begin{bmatrix} \dot{z}_c \\ \dot{z}_{uc} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} z_c \\ z_{uc} \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u \tag{3.16}$$

with (A_{11}, B_1) controllable.

Proof. If $\mathscr{C}$ has full row rank, then we do not need a transformation: $z = x$, $A_{11} = A$, and A_{22} is the "empty" matrix.

Now, suppose that $\mathscr{C}$ does not have full row rank, $q := \text{rank}(\mathscr{C}) < n$. In this case, the reachable subspace is a q-dimensional subspace of $\mathbb{R}^n$. Take a basis $\{v_1, \ldots, v_q\}$ of this subspace, and extend it to a basis of $\mathbb{R}^n$,

$$(v_1, \ldots, v_q, \tilde{v}_{q+1}, \ldots, \tilde{v}_n).$$

Let z be the coordinate vector of x with respect to this new basis, that is,

$$x = Tz, \qquad T := \begin{bmatrix} v_1 & v_2 & \cdots & \tilde{v}_n \end{bmatrix}.$$

So $z = T^{-1}x$. By construction, x is in the reachable subspace iff z is of the form $z = (z_1, \ldots, z_q, 0, \ldots, 0)$. Hence the reachable subspace in terms of z is $\left[\begin{smallmatrix} \mathbb{R}^q \\ 0 \end{smallmatrix}\right]$. The state transformation transforms the system $\dot{x} = Ax + Bu$ into $\dot{z} = T^{-1}ATz + T^{-1}Bu$, and the controllability matrix into

$$\begin{aligned}\mathscr{C}_z &:= \begin{bmatrix} T^{-1}B & T^{-1}ATT^{-1}B & \cdots & T^{-1}A^{n-1}B \end{bmatrix} \\ &= T^{-1}\mathscr{C}.\end{aligned} \tag{3.17}$$

Since the reachable subspace for z is $\left[\begin{smallmatrix}\mathbb{R}^q\\0\end{smallmatrix}\right]$ and all columns of the new matrix $T^{-1}B$ are part of the reachable subspace, the new matrix $T^{-1}B$ must be of the form $\left[\begin{smallmatrix}B_1\\0\end{smallmatrix}\right]$. Consequently, the transformed system is of the form

$$\begin{bmatrix}\dot{z}_c\\ \dot{z}_{uc}\end{bmatrix} = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}z_c\\ z_{uc}\end{bmatrix} + \begin{bmatrix}B_1\\0\end{bmatrix}u.$$

If we would now have $A_{21} \neq 0$, then, because of the reachability of z_c, we could send $z_c(t)$ (for some t) to a vector satisfying $A_{21}z_c(t) \neq 0$. But then, we would have $\dot{z}_{uc}(t) \neq 0$, which contradicts the fact that the reachable subspace is equal to $\left[\begin{smallmatrix}\mathbb{R}^q\\0\end{smallmatrix}\right]$. A contradiction, so $A_{21} = 0$. ■

Example 3.3.2. Consider the system

$$\dot{x} = \begin{bmatrix}0 & 1 & 1\\ 2 & 0 & 3\\ 1 & 0 & 1\end{bmatrix}x + \begin{bmatrix}1\\1\\0\end{bmatrix}u.$$

The controllability matrix is then

$$\mathscr{C} = \begin{bmatrix}1 & 1 & 3\\ 1 & 2 & 5\\ 0 & 1 & 2\end{bmatrix}.$$

This has rank 2 and is therefore not invertible. The first two columns of $\mathscr{C}$ span the reachable subspace. As new basis $\{v_1, v_2, v_3\}$, we choose the first two columns of $\mathscr{C}$ and an arbitrary column vector v_3 that is independent of the first two, for example $v_3 = (0,0,1)^{\mathrm{T}}$. So

$$T = \begin{bmatrix}v_1 & v_2 & v_3\end{bmatrix} = \begin{bmatrix}1 & 1 & 0\\ 1 & 2 & 0\\ 0 & 1 & 1\end{bmatrix}.$$

The transformed system $\dot{z} = T^{-1}ATz + T^{-1}Bu$ then becomes (verify this yourself)

$$\dot{z} = \left[\begin{array}{cc|c}0 & 1 & -1\\ 1 & 2 & 2\\ \hline 0 & 0 & 1\end{array}\right]z + \left[\begin{array}{c}1\\0\\ \hline 0\end{array}\right]u.$$

And the transformed controllability matrix becomes

$$\mathscr{C}_z = T^{-1}\mathscr{C} = \left[\begin{array}{ccc}1 & 0 & 1\\ 0 & 1 & 2\\ \hline 0 & 0 & 0\end{array}\right].$$

It confirms that the reachable subspace for z is $\left[\begin{smallmatrix}\mathbb{R}^2\\0\end{smallmatrix}\right]$. □

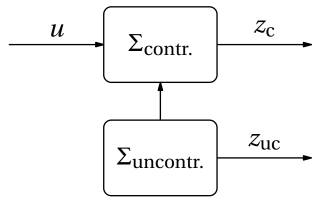


FIGURE 3.3: Kalman controllability decomposition.

Figure 3.3 illustrates this decomposition. The controllable part z_{c} is influenced by the input u and the uncontrollable part z_{uc}, while the uncontrollable part is influenced by nothing (apart from an initial condition $z_{\text{uc}}(0)$).

It should be clear that a state transformation $z = T^{-1}x$ does not change controllability: a controllable system with state x is still controllable in state $z = T^{-1}x$, and an *un*controllable system is still uncontrollable after a transformation. The importance of the Kalman controllability decomposition is, among other things, that it translates controllability into matrix properties, which are often easier to handle. A good example is the proof of the Hautus test[2].

Theorem 3.3.3 (Hautus test). *The system $\dot{x} = Ax + Bu$ is controllable if and only if the $n \times (n + n_u)$ matrix depending on s,*

$$\begin{bmatrix} sI - A & B \end{bmatrix},$$

has full row rank for all $s \in \mathbb{C}$.

Proof. Suppose that $\begin{bmatrix} sI - A & B \end{bmatrix}$ does not have full row rank for some s. Then there exists an $s_0 \in \mathbb{C}$ and a nonzero vector $\eta \in \mathbb{C}^n$ such that $\eta^{\mathrm{T}} \begin{bmatrix} s_0 I - A & B \end{bmatrix} = 0$. In particular, η^{T} is a left eigenvector of the A-matrix: $\eta^{\mathrm{T}} A = s_0 \eta^{\mathrm{T}}$. This η is orthogonal to the reachable subspace, because

$$\begin{aligned} \eta^{\mathrm{T}}\mathscr{C} &= \eta^{\mathrm{T}} \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \\ &= \begin{bmatrix} \eta^{\mathrm{T}}B & s_0\eta^{\mathrm{T}}B & \cdots & s_0^{n-1}\eta^{\mathrm{T}}B \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}. \end{aligned}$$

The controllability matrix therefore does not have full row rank, and the system is uncontrollable.

Conversely, suppose that the system is uncontrollable. Since a state transformation does not change the rank of $\begin{bmatrix} sI - A & B \end{bmatrix}$ (see Exercise 3.7), we may assume without loss of generality that the system is in the form of the Kalman controllability decomposition (3.16). The matrix A_{22} of (3.16) is nonempty because the system

[2]Malo Hautus was a professor at the University of Eindhoven. The test is also called the *PBH test*, after Popov, Belevitch, and Hautus. They discovered the test shortly after one another in 1966, 1968, and 1969, respectively.

is uncontrollable. Let $\eta_{\mathrm{uc}}^{\mathrm{T}}$ be a left eigenvector of A_{22} and let s_0 be its eigenvalue. (That is, $\eta_{\mathrm{uc}}^{\mathrm{T}} A_{22} = s_0 \eta_{\mathrm{uc}}^{\mathrm{T}}$ and $\eta_{\mathrm{uc}}^{\mathrm{T}} \neq 0$.) We then have

$$\begin{bmatrix} 0 & \eta_{\mathrm{uc}}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} s_0 I - A_{11} & -A_{12} & B_1 \\ 0 & s_0 I - A_{22} & 0 \end{bmatrix} = 0.$$

Hence $\begin{bmatrix} s_0 I - A & B \end{bmatrix}$ does not have full row rank. ■

Example 3.3.4. The system

$$\dot{x} = \begin{bmatrix} -2 & 1 \\ 0 & -1 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u$$

is uncontrollable because

$$\mathscr{C} = \begin{bmatrix} B & AB \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix},$$

and this is not invertible. This also follows from the Hautus test: consider the matrix

$$\begin{bmatrix} sI - A & B \end{bmatrix} = \begin{bmatrix} s+2 & -1 & 1 \\ 0 & s+1 & 1 \end{bmatrix}.$$

Since we are only interested in the rank, we may carry out elementary operations on the rows and columns. By subtracting the first row from the second, we obtain

$$\begin{bmatrix} s+2 & -1 & 1 \\ -s-2 & s+2 & 0 \end{bmatrix}. \tag{3.18}$$

The last row is zero if $s = -2$, so the rank of the matrix decreases for $s = -2$. Consequently, the system is uncontrollable. □

In the above example we could easily spot the rank loss of (3.18). For other more complicated systems that may be harder. However, what always should work is this: since $sI - A$ loses rank at precisely the eigenvalues s of A, the matrix $\begin{bmatrix} sI - A & B \end{bmatrix}$ can only lose rank if s is an eigenvalue of A. In the above example the eigenvalues of A are $s_1 = -1$ and $s_2 = -2$ and so to test for controllability we need only verify the rank of *two* matrices

$$\begin{bmatrix} (-1)I - A & B \end{bmatrix} = \begin{bmatrix} 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} (-2)I - A & B \end{bmatrix} = \begin{bmatrix} 0 & -1 & 1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Clearly the rank of the second is less than 2 (and it happens at $s = -2$).

3.4 Observability

The second fundamental property of systems is *observability*. Loosely speaking, observability means that we can determine the state by looking at only the external behavior (u, y). The output now also plays a role. We consider systems of the form

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t). \end{aligned} \tag{3.19}$$

Definition 3.4.1 (Observability). A system (3.19) is *observable* if there exists a $t_1 > 0$ such that for every triple of solutions (u_1, x_1, y_1), (u_2, x_2, y_2) of (3.19) with the same external behavior,

$$u_1(t) = u_2(t), \quad y_1(t) = y_2(t) \qquad \forall t \in [0, t_1],$$

also the state is the same,

$$x_1(t) = x_2(t) \qquad \forall t \in [0, t_1].$$

□

The system (3.19) is therefore observable if, from the knowledge of the input and output signals over a sufficiently long time interval $[0, t_1]$, we can uniquely determine the state signal on the interval $[0, t_1]$. The following $n_y n \times n$ matrix is crucial in the characterization of observability:

$$\mathscr{O} := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}. \tag{3.20}$$

This matrix $\mathscr{O}$ is called the *observability matrix* (compare this with the definition of the controllability matrix $\mathscr{C}$).

We first study the case where the external signals are zero at all times, $u(t) = 0, y(t) = 0$. This reduces the system (3.19) to

$$\begin{aligned} \dot{x}(t) &= Ax(t), \\ 0 &= Cx(t) \qquad \forall t > 0; \end{aligned}$$

that is,

$$Ce^{At}x_0 = 0 \qquad \forall t > 0.$$

For observability, this last equality must hold only for $x_0 = 0$. In general, x_0 does not need to be zero, and we define the t_1-*unobservable subspace* $\mathbb{X}^{\mathrm{uo}}(t_1)$ as those initial states for which the output is zero over $[0, t_1]$,

$$\mathbb{X}^{\mathrm{uo}}(t_1) := \{x_0 \in \mathbb{R}^n \mid Ce^{At}x_0 = 0 \forall t \in [0, t_1]\}. \tag{3.21}$$

Lemma 3.4.2 (Unobservable subspace). *Let $t_1 > 0$ and $\eta \in \mathbb{R}^n$. The following four statements are equivalent:*

1. $\eta \in \mathbb{X}^{uo}(t_1)$.

2. $Ce^{At}\eta = 0$ *for all* $t \in [0, t_1]$.

3. $CA^k\eta = 0$ *for all* $k = 0, 1, 2, \dots$

4. $\eta \in \ker(\mathscr{O})$.

Proof. We prove (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (1). The proofs closely resemble those of Lemma 3.1.3 on reachability.

(1) $\Longrightarrow$ (2) is by definition of $\mathbb{X}^{\text{uo}}(t_1)$. (2) $\Longrightarrow$ (3) and (3) $\Longrightarrow$ (4) are exercises (Exercise 3.22). This leaves (4) $\Longrightarrow$ (1): By the Cayley–Hamilton theorem, for every $k \geq 0$, the matrix A^k is a linear combination of $I, A, \dots, A^{n-1}$. If $\eta \in \ker(\mathscr{O})$, then by definition $C\eta = 0, CA\eta = 0, \dots, CA^{n-1}\eta = 0$, so by the Cayley–Hamilton theorem, $CA^k\eta = 0$ for all $k \geq 0$. It follows that

$$\begin{aligned}
Ce^{At}\eta &= C\big(I + tA + \tfrac{t^2}{2!}A^2 + \tfrac{t^3}{3!}A^3 + \cdots\big)\eta \\
&= C\eta + tCA\eta + \tfrac{t^2}{2!}CA^2\eta + \tfrac{t^3}{3!}CA^3\eta + \cdots \\
&= 0.
\end{aligned}$$

In other words, $\eta \in \mathbb{X}^{\text{uo}}(t_1)$. ∎

The t_1-unobservable subspace $\mathbb{X}^{\text{uo}}(t_1)$ is therefore equal to $\ker(\mathscr{O})$, and because the latter does not depend on t_1, the subspace $\mathbb{X}^{\text{uo}}(t_1)$ is apparently independent of t_1 as well (provided $t_1 > 0$). We therefore have

$$\ker(\mathscr{O}) = \{x_0 \in \mathbb{R}^n \mid Ce^{At}x_0 = 0 \,\forall t > 0\}.$$

This is called the *unobservable subspace*. It consists of all states for which $y(t) = 0$ for all t if $u(t) = 0$ for all t. The following theorem should not come as a surprise.

Theorem 3.4.3 (Observability). *Let $\mathscr{O}$ be the observability matrix defined in* (3.20). *The following three statements are equivalent:*

1. *The system is observable.*

2. $\ker(\mathscr{O}) = \{0\}$.

3. *The $\mathscr{O}$ has full column rank (rank n).*

Proof. We prove (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1).

(1) $\Longrightarrow$ (2): By contradiction: If $\ker(\mathscr{O}) \neq \{0\}$, then in addition to $\{0\}$, the unobservable subspace also contains a nonzero vector x_0. Then $x(t) = 0$ and $x(t) = e^{At}x_0$ both are consistent with zero external behavior for all time ($(u(t), y(t)) = (0,0)\,\forall t$). So the system is not observable.

(2) $\Longrightarrow$ (3). This is a standard result from linear algebra.

(3) $\Longrightarrow$ (1). Let $x_1, x_2: [0, t_1] \to \mathbb{R}^n$ be two states with the same external behavior $u: [0, t_1] \to \mathbb{R}^{n_u}, y: [0, t_1] \to \mathbb{R}^{n_y}$. In particular, we can express y two ways:

$$y(t) = Ce^{At}x_1(0) + \int_0^t Ce^{A(t-\tau)}Bu(\tau)\,dt + Du(t),$$
$$y(t) = Ce^{At}x_2(0) + \int_0^t Ce^{A(t-\tau)}Bu(\tau)\,dt + Du(t).$$

The difference between these two expressions of $y(t)$ shows that

$$Ce^{At}[x_1(0) - x_2(0)] = 0 \qquad \forall t \in [0, t_1].$$

So $x_1(0) - x_2(0)$ is an element of the unobservable subspace $\ker(\mathscr{O})$. Since $\mathscr{O}$ has full column rank, this subspace is $\{0\}$; that is, $x_1(0) = x_2(0)$. But then $x_1(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)\,dt = x_2(t)$ for all t. ■

Even though the definition of observability also considers the input u, observability apparently does not depend on the matrices B and D. We therefore often speak of the observability of the matrix *pair* (A, C).

Example 3.4.4 (Unobservable system). Consider the system

$$\dot{x} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x,$$
$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} x.$$

For every α, the state $x = \left[\begin{smallmatrix} 0 \\ \alpha \end{smallmatrix}\right]$ is a constant solution of this system (verify this yourself) and it gives $y(t) = 0$ for all t. So the state cannot be observed on the basis of the output y. This also follows from Theorem 3.4.3 because

$$\mathscr{O} = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

and this is not invertible. The nullspace, $\ker(\mathscr{O})$, is the unobservable subspace. In this case, this is $\left[\begin{smallmatrix} 0 \\ \mathbb{R} \end{smallmatrix}\right]$, so all vectors of the form $\left[\begin{smallmatrix} 0 \\ \alpha \end{smallmatrix}\right]$. □

The output y often consists of one element. In that case C is a matrix with one row and $\mathscr{O}$ is square. Then the system is observable if and only if $\mathscr{O}$ is invertible.

Example 3.4.5 (Observability in the inverted pendulum). Consider the system with two inverted pendula in Example 3.2.2. If we can only observe one of the two angles ϕ_1 and ϕ_2, then the system is unobservable. Take, for example, the output

$$y = \phi_1 = \begin{bmatrix} \frac{1}{\ell_1} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ v_1 \\ v_2 \end{bmatrix} - \frac{1}{\ell_1}u. \tag{3.22}$$

Then

$$\mathscr{O} = \begin{bmatrix} \frac{1}{\ell_1} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\ell_1} & 0 \\ \frac{g}{\ell_1^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{g}{\ell_1^2} & 0 \end{bmatrix} \tag{3.23}$$

and so (verify this yourself) $\text{rank}(\mathscr{O}) = 2 < 4$. If, on the other hand, we take for the output y the *difference* between the two angles, that is,

$$\begin{aligned} y &= \phi_1 - \phi_2 \qquad (3.24) \\ &= \begin{bmatrix} \frac{1}{\ell_1} & \frac{-1}{\ell_2} & 0 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ v_1 \\ v_2 \end{bmatrix} + \left(\frac{1}{\ell_2} - \frac{1}{\ell_1}\right) u, \end{aligned}$$

then the observability matrix equals

$$\mathscr{O} = \begin{bmatrix} \frac{1}{\ell_1} & \frac{-1}{\ell_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\ell_1} & \frac{-1}{\ell_2} \\ \frac{g}{\ell_1^2} & \frac{-g}{\ell_2^2} & 0 & 0 \\ 0 & 0 & \frac{g}{\ell_1^2} & \frac{-g}{\ell_2^2} \end{bmatrix}. \tag{3.25}$$

It is clear that the rank of this matrix decreases if $\ell_1 = \ell_2$. This is also the only way the rank can decrease (since the determinant of $\mathscr{O}$ is equal to $-g^2(\ell_1 - \ell_2)^2/(\ell_1\ell_2)^4$). Therefore, in the case $\ell_1 \neq \ell_2$, applying any input $u(t)$, $t \in [0, t_1]$, and recording the resulting output $y(t)$, $t \in [0, t_1]$ equal to the difference between the angles, allows us to uniquely determine the complete 4-dimensional state $x(t) = [q_1(t), q_2(t), v_1(t), v_2(t)]^{\mathrm{T}}$. Not bad! □

In analogy with he Kalman controllability decomposition, there is an observability decomposition.

Lemma 3.4.6 (Kalman observability decomposition). *Every system $\dot{x} = Ax + Bu$, $y = Cx + Du$ is isomorphic[3] to a system of the form*

$$\begin{aligned} \begin{bmatrix} \dot{z}_o \\ \dot{z}_{uo} \end{bmatrix} &= \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} z_o \\ z_{uo} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u, \qquad (3.26) \\ y &= \begin{bmatrix} C_1 & 0 \end{bmatrix} \begin{bmatrix} z_o \\ z_{uo} \end{bmatrix} + Du \end{aligned}$$

with (C_1, A_{11}) observable. This form is called the Kalman observability decomposition.

Proof. (A, C) is observable if and only if $(A^{\mathrm{T}}, C^{\mathrm{T}})$ is controllable. Apply the Kalman controllability decomposition to $(A^{\mathrm{T}}, C^{\mathrm{T}})$. ■

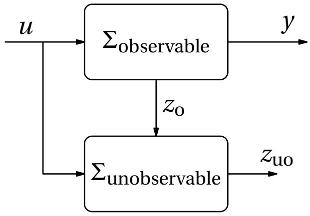


FIGURE 3.4: Kalman observability decomposition.

Figure 3.4 illustrates this decomposition. The input u and the observable state z_{o} may influence the unobservable state z_{uo}, but this unobservable z_{uo} does not influence the output y. The proof of Lemma 3.4.6 does not immediately yield an algorithm for finding this decomposition. The next example explains how to find it.

Example 3.4.7 (How to determine the Kalman observability decomposition). Consider the system without input

$$\dot{x} = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & 0 \\ 1 & 3 & 1 \end{bmatrix} x,$$
$$y = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} x.$$

The observability matrix of this system is

$$\mathscr{O}_x = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 3 & 5 & 2 \end{bmatrix},$$

and it is not invertible. (We added a subscript x to $\mathscr{O}$ to stress that this is with respect to state x.) It is not difficult to show that $\ker(\mathscr{O})$ is spanned by $(1,-1,1)^{\mathrm{T}}$. This is how we choose the *last* column of T. Next we simply choose the other columns of T in such a way that T is invertible. For example,

$$T = \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{array}\right].$$

Now, by construction, the unobservable subspace $\ker(\mathscr{O})$ is spanned by the last column of T. So in the new coordinates, $z := T^{-1}x$, it is spanned by $z = (0,0,1)^{\mathrm{T}}$. In these new coordinates, we have the desired decomposition (verify this yourself)

$$\dot{z} = \left[\begin{array}{cc|c} -1 & -1 & 0 \\ 2 & 3 & 0 \\ \hline 1 & 3 & -1 \end{array}\right] z,$$
$$y = \left[\begin{array}{cc|c} 1 & 1 & 0 \end{array}\right] z$$

[3] See § 2.2.2

and the observability matrix becomes

$$\mathcal{O}_z = \mathcal{O}_x T = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 3 & 5 & 0 \end{array}\right].$$

As expected, the unobservable subspace $\ker(\mathcal{O}_z)$ in the new state z are the vectors of the form $z = (0, 0, z_3)^{\mathrm{T}}$. □

The observability matrix of (3.26) is

$$\mathcal{O}_z = \begin{bmatrix} C_1 & 0 \\ C_1 A_{11} & 0 \\ \vdots & \vdots \\ C_1 A_{11}^{n-1} & 0 \end{bmatrix}$$

and the unobservable subspace, $\ker(\mathcal{O}_z)$, this is the set of states of the form

$$z = \begin{bmatrix} 0 \\ z_{\mathrm{uo}} \end{bmatrix}.$$

To conclude, we formulate the Hautus test for observability.

Theorem 3.4.8 (Hautus test—observability). *The system $\dot{x} = Ax + Bu$, $y = Cx + Du$ is observable if and only if the $(n + n_y) \times n$ matrix depending on s,*

$$\begin{bmatrix} sI - A \\ C \end{bmatrix},$$

has full column rank for all $s \in \mathbb{C}$.

Proof. See Exercise 3.23. ■

3.5 Canonical Representations

We present a number of canonical representations for systems that are either controllable or observable. We restrict ourselves to systems with a single input and a single output.

The first canonical form has a controllability matrix equal to the identity.

Lemma 3.5.1. *Suppose $n_u = 1$. For every controllable system $\dot{x} = Ax + Bu$ there is a unique state transformation $v = T^{-1}x$ for which the system takes the form*

$$\dot{v} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix} v + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} u. \tag{3.27}$$

This has controllability matrix $\mathcal{C}_v = I$, and the p_i are the coefficients of the characteristic polynomial of the A-matrix: $\det(sI - A) = s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0$.

Proof. A state transformation $v = T^{-1}x$ transforms the controllability matrix $\mathscr{C}_x$ into $\mathscr{C}_v = T^{-1}\mathscr{C}_x$ (see (3.17)). It is easy to see that the controllability matrix of (3.27) is $\mathscr{C}_v = I$ so *if* a tranformation to (3.27) exists then the T is unique: $T = \mathscr{C}_x$.

The $T = \mathscr{C}_x$ indeed does the job: then by construction $\mathscr{C}_v = I$. Since the first column of $\mathscr{C}_v = I$ is the transformed B, this transformed B must be the column vector $(1, 0, \dots, 0)$. Likewise, the second through nth columns of $\mathscr{C}_v = I$ are equal to the first through next-to-last columns of the transformed A (verify this yourself). Denote the last column of the transformed A-matrix as $(-p_0, \dots, -p_{n-1})$. We have seen in Exercise 2.7 that these p_i are the coefficients of the characteristic polynomial of the matrix. Since state transformations do not change characteristic polynomials this is also the characteristic polynomial of the matrix A. ■

In itself, this canonical form is of limited use, but it allows us to deduce the important *controller canonical form.* At first glance, this closely resembles the previous form, but the A-matrix is transposed, and the B-matrix is ordered differently.

Theorem 3.5.2 (Controller canonical form). *Suppose $n_u = 1$. Every controllable system $\dot{x} = Ax + Bu$ is isomorphic to a system of the form*

$$\dot{z} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -p_0 & -p_1 & \cdots & \cdots & -p_{n-1} \end{bmatrix} z + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} u. \tag{3.28}$$

This is called the controller canonical form. *This form is unique: the p_i are the coefficients of the characteristic polynomial* $\det(sI - A) = s^n + p_{n-1}s^{n-1} + \cdots + p_0$, *and $z = T^{-1}x$ where*

$$T = \mathscr{C}_x \mathscr{C}_z^{-1}.$$

Here $\mathscr{C}_x$ and $\mathscr{C}_z$ are the controllability matrices of $\dot{x} = Ax + Bu$ and the system (3.28)*, respectively.*

Proof. Define the p_i from the characteristic polynomial of the matrix A: $\det(sI - A) = s^n + p_{n-1}s^{n-1} + \cdots + p_0$. Since the matrix of (3.28) is a companion matrix its characteristic polynomial equals that of A.

Because the controller canonical form (3.28) is controllable (see Exercise 3.25), by Lemma 3.5.1 it is isomorphic to (3.27) through $v = \mathscr{C}_z^{-1}z$. The controllable system $\dot{x} = Ax + Bu$ is also isomorphic to (3.27) (through $v = \mathscr{C}_x^{-1}x$). So system $\dot{x} = Ax + Bu$ is isomorphic to (3.28), and $z = \mathscr{C}_z v = \mathscr{C}_z\mathscr{C}_x^{-1}x$ is the desired transformation from x to z. ■

The computation of $T = \mathscr{C}_x\mathscr{C}_z^{-1}$ can be laborious. It is just a bit easier to determine this T using that

$$T = \begin{bmatrix} \eta \\ \eta A \\ \vdots \\ \eta A^{n-1} \end{bmatrix}^{-1} \quad \text{with} \quad \eta := \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \mathscr{C}_x^{-1}. \tag{3.29}$$

This allows us to avoid having to determine $\mathscr{C}_z$. (Formula (3.29) is derived in Appendix A.6.)

Example 3.5.3 (Construction of a controller canonical form). Consider

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 1 & 3 \\ -2 & 2 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u, \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x. \end{aligned}$$

The controllability matrix and its inverse are

$$\mathscr{C}_x = \begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix}, \qquad \mathscr{C}_x^{-1} = \begin{bmatrix} 0 & 1 \\ 1/4 & -1/4 \end{bmatrix}.$$

The row vector η is defined as the last row of $\mathscr{C}_x^{-1}$, so $\eta = \begin{bmatrix} 1/4 & -1/4 \end{bmatrix}$. We can now compute the inverse of T from (3.29),

$$T^{-1} = \begin{bmatrix} 1/4 & -1/4 \\ 3/4 & 1/4 \end{bmatrix} \quad \text{and therefore} \quad T = \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix}.$$

The matrices of the controller canonical form (including output) now follow:

$$\begin{bmatrix} A_z & B_z \\ C_z & 0 \end{bmatrix} = \begin{bmatrix} T^{-1}AT & T^{-1}B \\ CT & 0 \end{bmatrix}.$$

This gives

$$\begin{aligned} \dot{z} &= \begin{bmatrix} 0 & 1 \\ -8 & 3 \end{bmatrix} z + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \\ y &= \begin{bmatrix} 1 & 1 \end{bmatrix} z. \end{aligned}$$

Actually only the C_z-matrix needs to be computed here because the B_z-matrix we know to be of the form $\begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}$ (as a column) and the A_z-matrix is a companion matrix whose coefficients in the bottom row are easily derived from the characteristic polynomial of the original matrix: $\det(\lambda I - A) = \det\left[\begin{smallmatrix} \lambda-1 & -3 \\ 2 & \lambda-2 \end{smallmatrix}\right] = \lambda^2 - 3\lambda + 8$. $\square$

In analogy with the controller canonical form, we have the observer canonical form.

Lemma 3.5.4 (Observer canonical form). *Suppose $n_u = n_y = 1$. Every observable system $\dot{x} = Ax + Bu, y = Cx$ is isomorphic to a system of the form*

$$\dot{z} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix} z + \begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ \vdots \\ q_{n-1} \end{bmatrix} u, \tag{3.30}$$
$$y = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} z.$$

This is called the observer canonical form. *This form is unique: the p_i are the coefficients of the characteristic polynomial of A, and $z = T^{-1}x$ for $T = \mathscr{O}_x^{-1}\mathscr{O}_z$, where $\mathscr{O}_x$ and $\mathscr{O}_z$ are, respectively, the observability matrices of $\dot{x} = Ax + Bu, y = Cx$ and the system (3.30). This T can also be determined using*

$$T = \begin{bmatrix} \eta & A\eta & \cdots & A^{n-1}\eta \end{bmatrix} \quad \textit{in which } \eta = \mathscr{O}_x^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Proof. The proof is analogous to that for the controllable canonical form: use that the system is observable if and only if the transposed system

$$\begin{cases} \dot{\tilde{x}} &= A^{\mathrm{T}}\tilde{x} + C^{\mathrm{T}}\tilde{u} \\ \tilde{y} &= B^{\mathrm{T}}\tilde{x} \end{cases}$$

is controllable. See also Appendix A.6. ■

Note that (3.30) is the state representation of the differential equation

$$y^{(n)} + p_{n-1}y^{(n-1)} + \cdots + p_0 y = q_{n-1}u^{(n-1)} + \cdots + q_0 u \tag{3.31}$$

(Lemma 2.5.2). So the theorem above says that in an observable system $\dot{x} = Ax + Bu, y = Cx$ the relation between input and output can also be represented simply using an ordinary differential equation (3.31). However, the main result of this theorem remains that observable systems can always be transformed into this special form. Due to its special structure, the observer canonical form is easier to analyze. We use this in the next chapter.

When is the observable canonical form controllable? The observer canonical form (3.30) of a differential equation (3.31) obviously is observable, but is it also controllable? Not always. To verify the controllability of the system in observer canonical form, we use the Hautus test. According to the Hautus test, the observer canonical form (3.30) is controllable if-and-only-if

$$\begin{bmatrix} sI - A & B \end{bmatrix} := \begin{bmatrix} s & 0 & \cdots & 0 & p_0 & q_0 \\ -1 & \ddots & \ddots & \vdots & p_1 & q_1 \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & s & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & s + p_{n-1} & q_{n-1} \end{bmatrix}$$

has full row rank for all $s \in \mathbb{C}$. Hence the system is uncontrollable if and only if for some $s \in \mathbb{C}$, we have $v^{\mathrm{T}} \begin{bmatrix} sI - A & B \end{bmatrix} = 0$ for some $v^{\mathrm{T}} \neq 0$. It follows from the first $n-1$ columns of $\begin{bmatrix} sI - A & B \end{bmatrix}$ that the only candidate for v^{T} is

$$v^{\mathrm{T}} = \begin{bmatrix} 1 & s & s^2 & \cdots & s^{n-1} \end{bmatrix}$$

(times a constant). This gives

$$\begin{aligned} \begin{bmatrix} 1 & s & s^2 & \cdots & s^{n-1} \end{bmatrix} & \begin{bmatrix} s & 0 & \cdots & 0 & p_0 & q_0 \\ -1 & \ddots & \ddots & \vdots & p_1 & q_1 \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & s & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & s + p_{n-1} & q_{n-1} \end{bmatrix} \\ = & \begin{bmatrix} 0 & \cdots & \cdots & 0 & P(s) & Q(s) \end{bmatrix}. \end{aligned}$$

It should be clear that this is the zero row if and only if $P(s) = Q(s) = 0$; that is, such s and v^{T} exist if and only if $P(s)$ and $Q(s)$ have at least one common factor. So we proved:

Lemma 3.5.5 (The observer canonical form is not always controllable.). *The observer canonical form of*

$$P(\tfrac{\mathrm{d}}{\mathrm{d}t})y = Q(\tfrac{\mathrm{d}}{\mathrm{d}t})u$$

is controllable if and only if $P(s)$ and $Q(s)$ do not have a common zero $s \in \mathbb{C}$. □

A simple example of $P(s)$ and $Q(s)$ with a common zero is $P(s) = Q(s) = s$. The corresponding differential equation $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ then becomes

$$\dot{y}(t) = \dot{u}(t).$$

Its solutions are $y(t) = c + u(t)$, with c a constant. The observer canonical form of this differential equation is

$$\begin{aligned} \dot{x}(t) &= 0, \\ y(t) &= x(t) + u(t) \end{aligned}$$

($x(t)$ takes over the role of the constant c). Clearly this system is uncontrollable. A natural question is now: is there another state representation for $\dot{y}(t) = \dot{u}(t)$ that *is* controllable? This is not the case! This suggests that we can say that the *system* $\dot{y}(t) = \dot{u}(t)$ is uncontrollable, without using the notion of state. This is indeed possible, but then we first need to define precisely what we mean by controllability for systems without states. This is explored in § 3.6.

Although the previous lemma shows that every system $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ has a controller canonical form provided that $P(s)$ and $Q(s)$ do not have a common zero, there does not seem to be a direct link between the coefficients of $P(s)$ and $Q(s)$ and the coefficients of the controller canonical form. But there is! The next result will probably come as a surprise.

Lemma 3.5.6 (Controller canonical form of differential equation $P(\frac{d}{dt})y = Q(\frac{d}{dt})u$). *Let $n_u = 1, n_y = 1$. If the observer canonical form*

$$\begin{aligned}
\dot{x} &= \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix} x + \begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ \vdots \\ q_{n-1} \end{bmatrix} u, \\
y &= \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} x.
\end{aligned}$$

is controllable, then the controller canonical form is the transpose of the observer canonical form,

$$\begin{aligned}
\dot{z} &= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -p_0 & -p_1 & \cdots & \cdots & -p_{n-1} \end{bmatrix} z + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} u, \\
y &= \begin{bmatrix} q_0 & q_1 & \cdots & \cdots & q_{n-1} \end{bmatrix} x.
\end{aligned}$$

Proof. In Chapter 5 we introduce transfer functions. We can use those to prove this theorem; see Exercise 5.19. ■

3.6 Behavioral Controllability

We have defined controllability as a property of state representations. This is the classic way to define controllability. For example, we could use it to show that the juggler with two sticks of equal length is an uncontrollable system. Our intuition concerning the juggler with two sticks is, however, such that we could already see this coming. Do we need the notion of state for this? The question that rises is as follows: is there a more intrinsic form of controllability (one without states)? The thought that it might exist, is unbearable for mathematicians and other theoretical scientists. It is time to return to the formal definition of a system. Recall that a dynamical system is a triple, $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$, with $\mathfrak{B} \subseteq \{w : \mathbb{T} \to \mathbb{W}\}$ the set of possible signals. This definition specifies that it is the set of signals that matter, not their representation. In this chapter, however, we have defined controllability for a specific representation of it (the state representation). If we want controllability to be a property of the system, then we must define it at the level of the system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$. This can easily be done: we call a system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ *controllable* if every signal w_1 in $\mathfrak{B}$ "from the past" can eventually be connected to any signal w_2 "from the future" in $\mathfrak{B}$. This is depicted in Figure 3.5.

Definition 3.6.1 (Behavioral controllability). A system $\Sigma = (\mathbb{R}, \mathbb{W}, \mathfrak{B})$ is *controllable* if for all $w_0, w_1 \in \mathfrak{B}$ and every $t_0 \in \mathbb{R}$, there exist a $t_1 \geq t_0$ and a $\tilde{w} \in \mathfrak{B}$ such

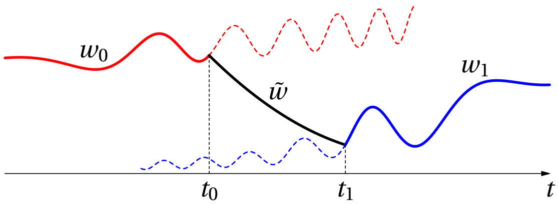


FIGURE 3.5: Controllability means that every possible signal w_0 "from the past" can eventually be connected to every possible signal w_1 "from the future".

that

$$\hat{w}(t) := \begin{cases} w_0(t) & \text{for } t < t_0 \\ \tilde{w}(t) & \text{for } t_0 \le t < t_1 \\ w_1(t) & \text{for } t_1 \le t \end{cases}$$

is also an element of $\mathfrak{B}$. □

Example 3.6.2 (Uncontrollable system). Consider the system with behavior $\mathfrak{B} = \{(u,y) \mid \dot{y} = \dot{u}\}$. The behavior consists of all signals (u,y) for which $y-u$ is constant.

This system is uncontrollable because both the zero function

$$\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and the constant function

$$\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

belong to the behavior, but they cannot be transformed into each other. □

Note that Definition 3.6.1 does not require states, but also no inputs or outputs, even no representations. Good.

Theorem 3.6.3 (Controllability for differential equations). *Let P, Q be two polynomials, not both zero. The system described by the differential equation*

$$P(\tfrac{\mathrm{d}}{\mathrm{d}t})w_1 = Q(\tfrac{\mathrm{d}}{\mathrm{d}t})w_2$$

is controllable (in the behavioral sense) if and only if P and Q do not have a common zero.

Proof (sketch). Without loss of generality, we assume that the degree of P is at least that of Q (otherwise we interchange the roles of P and Q).

Suppose that P and Q have no common zeros. Then the observer canonical form is controllable. Let (u_0, y_0, x_0) and (u_1, y_1, x_1) be possible solutions of the observer canonical form. Since this form is controllable, there exists an input $\tilde{u}: [t_0, t_t] \to \mathbb{R}$ that sends $x_1(t_0)$ to $x_2(t_1)$. Let $\tilde{y}: [t_0, t_1] \to \mathbb{R}^{n_y}$ be the resulting output. By the property of a state,

$$(u(t), y(t), x(t)) = \begin{cases} (u_0(t), y_0(t), x_0(t)), & t < t_0 \\ (\tilde{u}(t), \tilde{y}(t), \tilde{x}(t)), & t_0 \le t < t_1 \\ (u_1(t), y_1(t), x_1(t)), & t \ge t_1 \end{cases}$$

is then a solution of $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$. The system is controllable (in the behavioral sense).

Conversely, suppose that P and Q have a common zero. Then the observer canonical form is uncontrollable, and there exists a state $\hat{x}$ that is not in the reachable subspace $\mathrm{im}(\mathscr{C})$. Consider the two solutions

$$(u_0(t), y_0(t), x_0(t)) = (0, 0, 0)$$

and

$$(u_1(t), y_1(t), x_1(t)) = (0, C\mathrm{e}^{At}\hat{x}, \mathrm{e}^{At}\hat{x}).$$

The hypothesis that there exist a $t_1 > t_0$ and $\tilde{u}, \tilde{y}$ such that

$$(u(t), y(t)) = \begin{cases} (u_0(t), y_0(t)) & = (0, 0), & t < t_0 \\ (\tilde{u}(t), \tilde{y}(t)), & & t_0 \le t < t_1 \\ (u_1(t), y_1(t)) & = (0, C\mathrm{e}^{At}\hat{x}), & t \ge t_1 \end{cases}$$

satisfies the system equations leads to a contradiction. Indeed, because of the observability (over the time period $(-\infty, t_0)$), we have $x(t_0) = x_0(t) = 0$. But because of the observability over (t_1,∞), we also have $x(t) = \mathrm{e}^{At}\hat{x}$ for all $t > t_1$. The latter is in the unreachable subspace, but this is impossible because this subspace cannot be reached from $x(t_0) = 0$ for any input. Contradiction. So the system is uncontrollable in the behavioral sense. ■

3.7 Exercises

3.1 Comprehension questions (on the whole chapter). Prove or give a counterexample.

(a) If in a system with $x(t) \in \mathbb{R}^2$, there exist an input that sends $x(0) = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ to $x(1) = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ and an input that sends $x(0) = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ to $x(1) = \left[\begin{smallmatrix}1\\-2\end{smallmatrix}\right]$, then the system is controllable?

(b) If in a system with $x(t) \in \mathbb{R}^2$, there exists an input that sends $x(0) = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ to $x(1) = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ and an input that sends $x(0) = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ to $x(100) = \left[\begin{smallmatrix}1\\-2\end{smallmatrix}\right]$, then the system is controllable?

(c) If $C = 0 \in \mathbb{R}^{1\times n}$, then the system is not observable?

(d) Let $A \in \mathbb{R}^{n\times n}$. If the rank of $s_0 I - A$ is less than $n-1$, then (A,C) is unobservable?

(e) Let $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{1\times n}$. If the rank of $s_0 I - A$ is less than $n-1$, then (A,C) is unobservable?

(f) If (A,C_1) are (A,C_2) are observable, then (A,C_1+C_2) is also observable?

3.2 Let $t_1 > 0$. Show that $\mathbb{X}(t_1)$ as defined in (3.4) is a subspace of $\mathbb{R}^n$.

3.3 *Controllability.* Determine the controllability of the system $\dot{x} = Ax + Bu$ for the following pairs:

(a) $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

3.4 *Controllability.* Let M and N be matrices with the same number of rows, and assume that M is square. Show that the system

$$\dot{x} = \begin{bmatrix} 0 & M \\ M & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ N \end{bmatrix} u$$

is controllable if and only if M is invertible and

$$\dot{z} = M^2 z + Nu$$

is controllable.

3.5 *Controllability.* Determine the controllability of the systems $\dot{x} = Ax + Bu$ for the following pairs:

(a) $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(c) $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

3.6 Suppose $x_0 = 0$. Show that the u_* from (3.6) ensures that $x(t_1) = x_1$. Also show that $\|u_*\|^2 = x_1^{\mathrm{T}} P^{-1}(t_1) x_1$.

3.7 It is intuitively clear that the system $\dot{x} = Ax + Bu$ and transformed system $\dot{z} = T^{-1}ATz + T^{-1}Bu$ are either both controllable or both uncontrollable. In terms of the Hautus test this means that $\begin{bmatrix} sI - A & B \end{bmatrix}$ has full row rank for all $s \in \mathbb{C}$ iff $\begin{bmatrix} sI - T^{-1}AT & T^{-1}B \end{bmatrix}$ has full row rank for all $s \in \mathbb{C}$.

Specifically, show that for any given $s \in \mathbb{C}$, the matrix $\begin{bmatrix} sI - A & B \end{bmatrix}$ does not have full row rank iff $\begin{bmatrix} sI - T^{-1}AT & T^{-1}B \end{bmatrix}$ does not have full row rank. [Hint: use that a matrix W does not have full row rank iff $\eta^{\mathrm{T}} W = 0$ for some nonzero vector η.]

3.8 *Controllability.* Give a controllable system $\dot{x} = Ax + Bu$ such that for every column b of B, the system (A, b) is uncontrollable.

3.9 *Computation of the control signal u.* Compute the input signal $u : [0, 1] \to \mathbb{R}$ for the system

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

that sends $x(0) = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ to $x(1) = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ with minimal energy.

3.10 Determine the Kalman controllability decomposition of $\dot{x} = Ax + Bu$ for the following pairs:

(a) $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ -2 & 1 & 2 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

3.11 In analogy with the linear case, we call a nonlinear system

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

controllable if for every pair of states x_0, x_1, there exists a $t_1 \geq 0$ and an input u such that if $x(0) = x_0$ then $x(t_1) = x_1$.

Is the nonlinear system $\dot{x} = -x + u^2$ controllable?

3.12 *Diagonalizable matrices.* Suppose that (A, B) is controllable and $n_u = 1$, and suppose that A has an eigenvalue of multiplicity more than 1. Can A be diagonalizable?

3.13 *Controllability of a serial interconnection.* Consider the serial interconnection of Figure 3.6, with subsystems H_1 and H_2 given by

$$\begin{aligned} H_1 : &\quad \dot{x}_1 = A_1 x_1 + B_1 u, \quad z = C_1 x_1, \\ H_2 : &\quad \dot{x}_2 = A_2 x_2 + B_2 z, \quad y = C_2 x_2. \end{aligned}$$

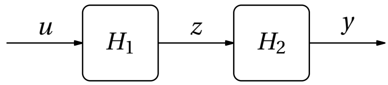


FIGURE 3.6: Serial interconnection.

If the interconnected system (with input u, output y, and state $\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$) is controllable is, are the subsystems H_1 and H_2 necessarily controllable?

3.14 Prove that (A, B) is uncontrollable if and only if there exists a $C \neq 0$ such that $C\mathrm{e}^{At}B = 0$ for all t.

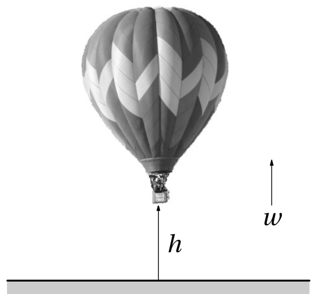


FIGURE 3.7: Hot air balloon.

3.15 *Hot air balloon.* The linearized equations for the motion of a hot air balloon are

$$
\begin{aligned}
\dot{\theta} &= -\frac{1}{\tau_1}\theta + u, \\
\dot{v} &= -\frac{1}{\tau_2}v + \sigma\theta + \frac{1}{\tau_2}w, \\
\dot{h} &= v,
\end{aligned}
$$

where θ is the temperature in the balloon, u is the added heat, v is the vertical velocity, and h is the elevation of the balloon; see Figure 3.7.

(a) Is the system controllable at a fixed wind speed w?

(b) Is the system controllable if we view u and w as control variables?

3.16 Let E be an invertible $n \times n$ matrix. Furthermore, let, as usual, A be an $n \times n$ matrix and B be a matrix with as many rows as A. Show that the system described by the implicit state representation

$$E\dot{x} = Ax + Bu$$

is controllable if and only if

$$\begin{bmatrix} sE - A & B \end{bmatrix}$$

has full row rank for all $s \in \mathbb{C}$.

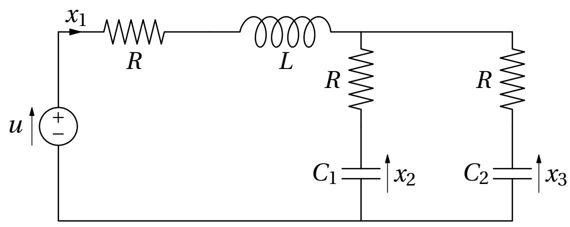


FIGURE 3.8: *RLC* circuit.

3.17 *Controllability of an RLC circuit.* Consider the *RLC* circuit of Figure 3.8. We take the voltage u over the voltage source as input. Straightforward modeling gives the model

$$\begin{bmatrix} L & RC_1 & 0 \\ 0 & RC_1 & -RC_2 \\ 0 & C_1 & C_2 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -R & -1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u$$

with u the voltage over the voltage source, $x_1 = i$ the current through the inductor, and x_2, x_3, respectively, the voltage over the capacitors C_1 and C_2. Assume that all constants R, C_1, C_2, L are greater than zero.

(a) Under which conditions on R, C_1, and C_2 is the system controllable? (Hint: Use the previous exercise.)

(b) Interpret your findings.

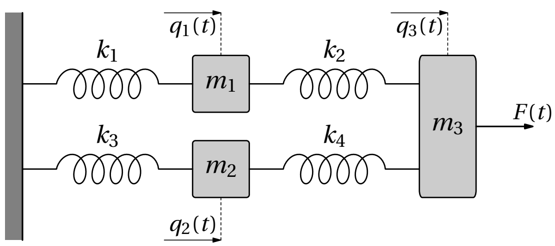


FIGURE 3.9: Spring-mass system.

3.18 *Controllability of a spring-mass system.* Consider the spring-mass system of Figure 3.9. Two masses m_1 and m_2 are attached to a wall (on the left) by springs with spring constants k_1 and k_3, and on the right to a mass m_3 by springs with spring constants k_2 and k_4. We can exert a force F on the mass m_3. The positions of the three masses with respect to their equilibrium points are denoted by q_i.

From Newton's second law and Hooke's law, we obtain the equations of motion

$$\begin{aligned} m_1\ddot{q}_1 &= -k_1q_1 + k_2(q_3-q_1),\\ m_2\ddot{q}_2 &= -k_3q_2 + k_4(q_3-q_2),\\ m_3\ddot{q}_3 &= -k_2(q_3-q_1) - k_4(q_3-q_2) \ + \ F. \end{aligned}$$

(a) Write this system in the form $\dot{x} = Ax + Bu$ with $u = F$ and state $x = [q_1, \dot{q}_1, q_2, \dot{q}_2, q_3, \dot{q}_3]^{\mathrm{T}}$.

(b) Suppose that $k_1 = k_2 = k_3 = k_4$. Under which conditions on the m_i is the system controllable? Interpret your findings.

(c) Suppose, once more, that $k_1 = k_2 = k_3 = k_4$. Choose q_3 as output. Under which conditions on the m_i is the system controllable? Interpret your findings.

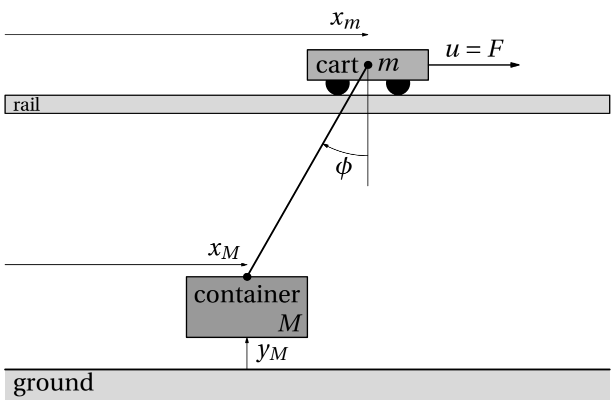


FIGURE 3.10: Container transfer.

3.19 *Container transfer.* Consider the problem of transferring containers (Figure 3.10; see also Exercise 2.21). As input, we take the force we can exert on the cart, $u = F$. The linearized system is

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0\\ 0 & 0 & -\frac{M}{m}g & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & -\frac{M+m}{Lm}g & 0 \end{bmatrix} x + \begin{bmatrix} 0\\ \frac{1}{m}\\ 0\\ \frac{1}{Lm} \end{bmatrix} u \tag{3.32}$$

with $x = [x_m, \dot{x}_m, \phi, \dot{\phi}]^{\mathrm{T}}$.

(a) Is the linearized system controllable?

(b) Is the linearized system observable if we take only the position $y = x_m$ of the cart as output?

(c) In order to control the container using u, is it necessary to observe not only the position of the cart, but also that of the container (for example with cameras)?

3.20 *Observability.* Determine the observability of the systems $\dot{x} = Ax,\ y = Cx$ for the following pairs:

(a) $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$

(c) $A = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$

(d) $A = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & ⋰ & ⋰ & 0 \\ 0 & ⋰ & ⋰ & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$, $C = \begin{bmatrix} 1 & 2 & \cdots & n \end{bmatrix}$

3.21 Determine the Kalman observability decomposition of $\dot{x} = Ax, y = Cx$ for the following pairs:

(a) $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 2 & 2 \end{bmatrix}$

(b) $A = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$

3.22 *Observability.* Prove the implications (2) $\Longrightarrow$ (3) and (3) $\Longrightarrow$ (4) of Lemma 3.4.2

3.23 *Hautus test for observability*

(a) Prove that the system (3.19) is observable if and only if $\operatorname{rank}\left[\begin{smallmatrix} sI-A \\ C \end{smallmatrix}\right] = n$ for all $s \in \mathbb{C}$.

(b) Under what conditions on $c_1, c_2, \ldots, c_n \in \mathbb{R}$ is the following system observable?

$$\dot{x} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & n \end{bmatrix} x$$
$$y = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} x.$$

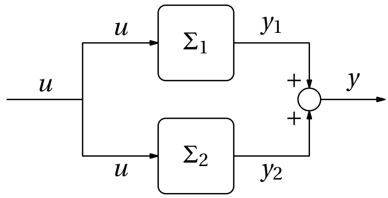


FIGURE 3.11: Parallel interconnection.

3.24 *Observability of a parallel interconnection.* Consider the configuration in Figure 3.11, with subsystems Σ_1 and Σ_2 that are both observable and controllable, with state representations

$$\begin{aligned}\Sigma_1: &\quad \dot{x}_1 = A_1 x_1 + B_1 u, \quad y_1 = C_1 x_1,\\ \Sigma_2: &\quad \dot{x}_2 = A_2 x_2 + B_2 u, \quad y_2 = C_2 x_2.\end{aligned}$$

(a) Explain in words that the system with input u, output $y = y_1 + y_2$, and state $\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ is unobservable if $A_1 = A_2$ and $C_1 = C_2$.

(b) Prove this using Theorem 3.4.3.

3.25 Show that the controller canonical form (3.28) is controllable.

3.26 Prove the observability of (3.30) using the Hautus test (Theorem 3.4.8).

3.27 Consider the system

$$\begin{aligned}\dot{x} &= \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u,\\ y &= \begin{bmatrix} -1 & 2 \end{bmatrix} x.\end{aligned} \tag{3.33}$$

(a) Determine the controller canonical form of this system:

$$\begin{aligned}\dot{z} &= \begin{bmatrix} 0 & 1 \\ ? & ? \end{bmatrix} z + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u,\\ y &= \begin{bmatrix} ? & ? \end{bmatrix} z.\end{aligned} \tag{3.34}$$

(b) Determine the observer canonical form of the system.

Tougher Exercises

3.28 Suppose $n_u = n_y = 1$. If

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

is controllable and observable, is it isomorphic to

$$\begin{aligned}\dot{z} &= A^{\mathrm{T}} z + C^{\mathrm{T}} u,\\ y &= B^{\mathrm{T}} z?\end{aligned}$$

3.29 *Reachability.* Consider Theorem 3.1.5.

(a) Use the fact that $\|u_*\|$ has minimal norm to prove that $x_1^{\mathrm{T}} P^{-1}(t) x_1$ is non-increasing as a function of t.

(b) Is $z^{\mathrm{T}} P(t) z$ non-increasing as a function of t?

(c) Suppose that A is stable. Does $\lim_{t\to\infty} P(t)$ exist?

(d) Show that $\dot{P}(t) = BB^{\mathrm{T}} + AP(t) + P(t)A^{\mathrm{T}}$.

(e) Show that if A is stable, then $P_\infty := P(\infty)$ satisfies $AP_\infty + P_\infty A^{\mathrm{T}} + BB^{\mathrm{T}} = 0$

3.30 *Dead beat control.* Dead beat control deals with constructing inputs u that bring the state to zero within a finite amount of time. We now consider discrete-time systems.

(a) Consider the *discrete*-time system

$$x[t+1] = Ax[t] + Bu[t], \qquad t \in \mathbb{Z}.$$

Show that "dead beat control" is possible for every initial state $x[0] \in \mathbb{R}^n$ if and only if

$$\begin{bmatrix} sI - A & B \end{bmatrix}$$

has full row rank for all $s \in \mathbb{C}, s \neq 0$.

(b) Does this same condition hold for continuous-time systems $\dot{x} = Ax + Bu$?

3.31 *Observing using outputs.* Consider the observable system with a single input and a single output

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

Prove that we can determine the state using *only* the knowledge of the output $y(t)$ for $t \geq 0$ if and only if $Ce^{At}B = 0$ for all t.

3.32 *Behavioral controllability.*

(a) Is the system in Exercise 3.18b with $w := (q_1, q_2, F)$ controllable in the sense of Theorem 3.6.3?

(b) The linearized model of the container transfer problem from which (3.32) is deduced, is (compare with (2.43)),

$$\begin{aligned} x_M(t) &= x_m(t) - L\phi(t), \\ y_M(t) &= 0, \\ m\ddot{x}_m(t) &= F(t) - \phi(t)b(t), \\ M\ddot{x}_M(t) &= \phi(t)b(t), \\ M\ddot{y}_M(t) &= b(t) - Mg. \end{aligned}$$

Show that $b(t)$ is constant. Is the system with $w = (x_m, x_M, \phi, F)$ controllable according to Theorem 3.6.3?

3.33 Theorem 3.1.5 on reachability introduced the *controllability Gramian* $P(t)$. The analogous theorem on observability (Theorem 3.4.3) lacks Gramians. We could have included them: prove that the following five statements are equivalent.

(a) The system is observable.

(b) $\ker(\mathscr{O}) = \{0\}$.

(c) The observability matrix $\mathscr{O}$ (3.20) has full column rank (rank n).

(d) The *observability Gramian* $Q(t)$ defined as

$$Q(t) = \int_0^t \mathrm{e}^{A^{\mathrm{T}}t} C^{\mathrm{T}} C \mathrm{e}^{At} \,\mathrm{d}t$$

is invertible for all $t > 0$

(e) The observability Gramian $Q(t)$ is invertible for some $t > 0$.

Now suppose that $u(t) = 0$ for all $t \in [0, t_1]$ and that the system is observable.

(f) Given a possible output $y : [0, t_1] \to \mathbb{R}^{n_y}$ show that the initial state $x(0)$ can be reconstructed from the output using

$$x(0) = Q^{-1}(t_1) \int_0^{t_1} \mathrm{e}^{A^{\mathrm{T}}t} C^{\mathrm{T}} y(t) \,\mathrm{d}t.$$

(g) In practice the measured output $\bar{y}(t)$ hardly ever equals the model $y(t) := C\mathrm{e}^{At}x(0)$. Show that

$$\bar{x}_0 := Q^{-1}(t_1) \int_0^{t_1} \mathrm{e}^{A^{\mathrm{T}}t} C^{\mathrm{T}} \bar{y}(t) \,\mathrm{d}t$$

solves the minimization problem

$$\min_{x_0} \|\bar{y} - C\mathrm{e}^{A\cdot} x_0\|.$$

Here the norm is the standard $\mathscr{L}_2$-norm of functions on the interval $[0, t_1]$:

$$\|z\| = \sqrt{\int_0^{t_1} z^{\mathrm{T}}(t) z(t) \,\mathrm{d}t}.$$

3.34 If a system is not reachable (i.e. not controllable) then, by definition, some states can not be reached. With this in mind it seems likely that "almost unreachable" systems require "large" inputs to control the state. This is confirmed by the following example. Let

$$\dot{x} = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} x + \begin{bmatrix} 1 & 0 \\ 0 & \beta \end{bmatrix} u.$$

Here both x and u have two components, and $\alpha, \beta \in \mathbb{R}$.

(a) Show that the system is uncontrollable iff $\beta = 0$.

(b) The smallest squared norm $\|u\|^2 = \int_0^{t_1} u_1^2(t) + u_2^2(t)\,\mathrm{d}t$ of all inputs that achieve $x(t_1) = x_1$ is given in Thm. 3.1.5. Compute this $\|u\|^2$.

(c) For which $x_1 \in \mathbb{R}^2$ do we have $\lim_{\beta\to 0}\|u\| = \infty$, and explain why you are not surprised by the answer.

(d) Suppose $\alpha \neq 0$. Show that $\|u\|$ decreases as t_1 increases. (In words this means: the longer the time, the smaller the required control action.)

(e) The minimal $\|u\|$ is also distinctively different for positive and negative α if t_1 is large: show that

$$\lim_{t_1\to\infty}\|u\|^2 = 0 \quad \text{if } \alpha > 0, \beta \neq 0 \tag{3.35}$$

and

$$\lim_{t_1\to\infty}\|u\|^2 = 2|\alpha|(x_{11}^2 + x_{12}^2/\beta^2) \quad \text{if } \alpha < 0, \beta \neq 0.$$

(Here x_{11}, x_{12} are the entries of $x_1 \in \mathbb{R}^2$.)

(f) Explain in words why the limit (3.35) makes sense.

Chapter 4

State Feedback and Observers

In the definition of controllability, we studied the existence of input signals $u : [0, t_1] \to \mathbb{R}^{n_u}$ that steer a given initial state $x(0) = x_0$ to a given desired state x_1 at some time t_1. This is a typical example of *open-loop control*: a time signal $u : [0, t_1] \to \mathbb{R}^{n_u}$ is programmed based on known system data and states x_0, x_1. This is depicted schematically in Figure 4.1.

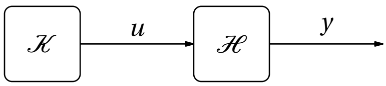


FIGURE 4.1: Open-loop control.

However, if such a computed input signal is used in an actual, real-life, system, then the end result will almost always be somewhat different from the computed and intended result. This is due to inevitable inaccuracies in the model, disturbances to the system, and an implementation of the computed input signal that is not 100% accurate. Although open-loop control is useful, in practice it will need to be supplemented with other methods.

What other control methods are there? Let us take another look at the example of the juggler (Example 3.2.2). We can imagine a juggler with a stick in his hand that he can move horizontally. Suppose that the juggler wants to balance the stick; in other words, he wants to keep the stick (approximately) upright. How should he proceed? In principle, he could, at any given time, determine the position and velocity of the stick and based on that (and on a precise knowledge of the equations of motion of the stick) compute how he needs to move his hand to put the stick upright and keep it there. Then he could—with his eyes closed—carry out this hand movement. This is an example of open-loop control.

Even if our juggler is capable of carrying out this computation, it is still clear that he cannot keep the stick balanced this way. After all, there will always be small errors or disturbances, and since the upright position of the stick is an unstable equilibrium the stick will fall anyway! The method he will use is completely different. He *observes* the position and velocity of the tip of the stick at every moment, and *based on this* adjusts his hand movement. This obvious method is the essence of *control theory*: the input $u(t)$ at every time t is determined as a function of a number of observed quantities $y(t)$ of the system up to this time t. This principle is called *feedback*, and the resulting control methodology is called *closed-loop control*. This is depicted schematically in Figure 4.2.

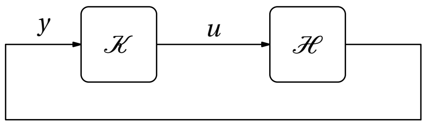


FIGURE 4.2: Feedback (closed-loop control).

In fact, more is going on in our juggler example, as well as in many other examples. There is a *learning* or *adaptive* element in the determination of the necessary feedback: based on the result of the feedback, we will adjust the feedback if necessary. For example, the function that shows how the movement of the hand depends on the position and velocity of the tip of the stick can be adjusted, in particular, if the properties of the system that is being controlled change (for example because an object is added onto the tip of the stick).

This feedback mechanism (whether adaptive or not) can be found in many biological, physical, and technical systems. In control engineering, the controller is usually itself an automatic mechanism (in contrast to our juggler). One of the classic examples of a mechanical controller comes from James Watt, and concerns the control of the steam engine, the "motor" behind the industrial revolution of the nineteenth century. The steam engine can be seen as a state system with as input u the steam supply and as output y the angular velocity of the governor shaft; see Figure 4.3. A typical problem is keeping the output y as close as possible to a previously determined constant value y_0 (constant number of revolutions per minute). The flyball governor developed by James Watt[1] achieves this by having two metal weights rotating about a shaft operate a lever that controls the steam supply; see Figure 4.3. If the angular velocity increases, the weights move outwards, automatically decreasing the steam supply (and indirectly decreasing the rpm). Conversely, at a lower angular velocity, the weights move inwards, increasing the steam supply. By adding this flyball governor to the steam engine, there is feedback from the output y to the input u, and we can show mathematically that the input y converges to a constant value y_0, regardless of natural variations in the steam input u.

In the case of Watt's flyball governor, we only need to feed back the output y. In other examples, however, it may be necessary to feed back the entire state vector.

[1] James Watt adapted Christiaan Huygens' centrifugal governor.

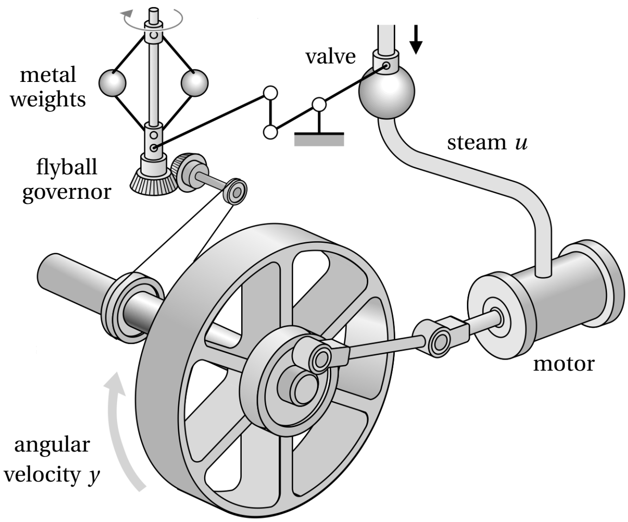


FIGURE 4.3: The steam engine with the flyball governor in the upper left corner. (This illustration is from [2, § 1.2].)

In this chapter, we analyze feedback control of state systems of the form[2]

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx \end{aligned} \tag{4.1}$$

using a static control law of the form

$$u = -Fx \qquad \text{for some } F \in \mathbb{R}^{n_u \times n}$$

or a dynamic control law

$$\begin{aligned} \dot{z} &= A_{\mathcal{K}} z + B_{\mathcal{K}} y, \\ u &= C_{\mathcal{K}} z \end{aligned}$$

for certain matrices $A_{\mathcal{K}}, B_{\mathcal{K}}, C_{\mathcal{K}}$. For state systems, control breaks up into two more or less independent problems: 1) How can we send the state x to zero by choosing the input u appropriately as a function of x? 2) This assumes that we have x at our disposal for feedback, but what if we can only measure part of the state, say y; under what conditions is the knowledge of (u, y) sufficient to reconstruct x? Next, we combine these problems and arrive at the celebrated result that says that state systems that are controllable and observable can be stabilized fully automatically in the sense that $\lim_{t\to\infty} x(t) = 0$, $\lim_{t\to\infty} u(t) = 0$, and $\lim_{t\to\infty} y(t) = 0$ regardless of the initial conditions.

[2]Note that the direct feedthrough term D in $y = Cx + Du$ is assumed to be zero. This chapter's theory also works well for $D \neq 0$, but the formulas are then more complicated.

4.1 Stabilizability

In Chapter 2, we saw that $\lim_{t\to\infty} x(t) = 0$ for all solutions of $\dot{x} = Ax$ if and only if the eigenvalues of A have negative real part. An extension of stability that includes the freedom of choice of u is called stabilizability.

Definition 4.1.1 (Stabilizability). A system $\dot{x} = Ax + Bu$ is *stabilizable* if for every $x(0) = x_0 \in \mathbb{R}^n$, there exists a $u : [0,\infty) \to \mathbb{R}^{n_u}$ such that $\lim_{t\to\infty} x(t) = 0$. □

The input u is then called a *stabilizing* input. The definition of stabilizability does not explain how to choose u. If we can measure the entire state $x(t)$ at every time t, then we can try, as control law, a function of the form

$$u(t) = \mathscr{F}\big(t, x(\tau)\big|_{\tau\le t}\big),$$

that is, a control law that uses the state measured up to time t. The simplest control law of this form is the linear *static state feedback* law,

$$u(t) = -Fx(t) \qquad \text{for some } F \in \mathbb{R}^{n_u\times n}.$$

(The inclusion of a minus sign is a convention.) This is called *static* feedback because at every time t, the input $u(t)$ depends only on the "current" state $x(t)$ (and not on its past). The choice of a static state feedback is not unnatural, because by the definition of a state, $x(t)$ contains "all necessary" information from the past to determine the future.

Example 4.1.2 (Open loop versus closed loop). This example illustrates the fundamental difference between open loop control and closed loop control, and it demonstrates that closed-loop control is superior.

Consider the system

$$\dot{x} = x + u.$$

This system is *un*stable. The stabilize it we need to choose the input appropriately. There are many inputs u that stabilize the system. Two of them are

$$\begin{aligned} \textit{open loop:} \quad & u(t) = -3\mathrm{e}^{-2t}x(0), \\ \textit{closed loop:} \quad & u(t) = -3x(t). \end{aligned} \tag{4.2}$$

The two inputs u give the same result (verify this yourself):

$$x(t) = \mathrm{e}^{-2t}x(0),$$

and therefore both inputs stabilize the system. However, there is a big difference: in the open-loop control case, the input, $u(t) = -3\mathrm{e}^{-2t}x(0)$, is determined by the initial state $x(0)$ and this initial state fixes the input for the rest of time. This is like looking at the system once in your life, and then controlling it with your eyes closed for the rest of time. In the closed-loop control case, the input,

$u(t) = -3x(t)$, at *every moment in time* is chosen depending on the state at that moment. This method requires the continuous observation of $x(t)$ and is, in that sense, more complicated. But it is also *much, much, much more robust*! Suppose, for example that the model $\dot{x} = x + u$ deviates a bit from the actual system, and that the actual system is

$$\dot{x} = 1.001x + u.$$

If we now apply the two control laws (4.2) to the actual system, we obtain (verify this yourself)

$$\begin{aligned} \textit{open loop:} \quad & x(t) = [\tfrac{3}{3.001}\mathrm{e}^{-2t} + \tfrac{0.001}{3.001}\mathrm{e}^{1.001t}]x(0), \\ \textit{closed loop:} \quad & x(t) = \mathrm{e}^{-1.999t}x(0). \end{aligned}$$

The open-loop law destabilizes it (because $\mathrm{e}^{1.001t}$ diverges), but the closed-loop feedback law still stabilizes the system, regardless of the small disturbance. So open-loop control can be extremely sensitive to modeling errors, while closed-loop control appears to be robust. This is typical, and in applications you always want to use closed loop control if the system itself is unstable. Open loop control is really not sufficient if the system is unstable. □

It is odd that the fundamental difference between open-loop and closed-loop control is not always well understood.

4.2 Static State Feedback

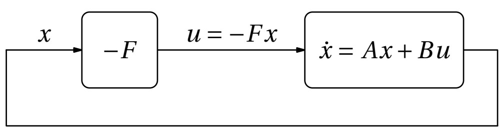


FIGURE 4.4: Static state feedback.

Probably the simplest closed loop control law is

$$u(t) = -Fx(t),$$

where F is some matrix. This is known as (linear, time-invariant) *static state feedback*. Application of this control to the system $\dot{x} = Ax + Bu$ modifies the dynamics into

$$\begin{aligned} \dot{x} &= Ax + Bu \\ &= Ax - BFx \\ &= (A - BF)x. \end{aligned}$$

If we can choose F in such a way that all eigenvalues of $A - BF$ have negative real part, then $u = -Fx$ is a stabilizing input.

Example 4.2.1 (Juggler). Consider, once more, the inverted pendulum of Example 3.2.2. For completeness, here are the system equations:

$$\begin{bmatrix} \dot{q}_1 \\ \dot{v}_1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{g}{\ell_1} & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ v_1 \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{g}{\ell_1} \end{bmatrix} u. \tag{4.3}$$

It should be clear that this system is not asymptotically stable (otherwise juggling would be a piece of cake). Since this is a second-order system, the characteristic polynomial has degree 2. We want a state feedback that puts the two eigenvalues of $A - BF$ in $s = -1$. So, we want

$$\chi_{A-BF}(s) = (s+1)^2 = s^2 + 2s + 1.$$

We write the candidate state feedback as $u = -Fx = -\begin{bmatrix} f_1 & f_2 \end{bmatrix} \begin{bmatrix} q_1 \\ v_1 \end{bmatrix}$; then

$$A - BF = \begin{bmatrix} 0 & 1 \\ \frac{g}{\ell_1}(1+f_1) & \frac{g}{\ell_1} f_2 \end{bmatrix}.$$

This has characteristic polynomial

$$\chi_{A-BF}(s) = s^2 - \frac{g}{\ell_1} f_2 s - \frac{g}{\ell_1}(1+f_1).$$

It equals $s^2 + 2s + 1$ if we choose

$$f_1 = -1 - \frac{\ell_1}{g}, \qquad f_2 = -2\frac{\ell_1}{g}.$$

Done. □

In the example above, we have put the closed-loop poles at $s = -1$ (twice), but we could just as well have chosen another pair of poles[3]. We will soon see that this has everything to do with controllability.

Example 4.2.2 (Controller canonical form & pole placement). Suppose our system is in controller canonical form,

$$\dot{z} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -p_0 & -p_1 & \cdots & \cdots & -p_{n-1} \end{bmatrix} z + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} u.$$

Under the state feedback law

$$u = -\begin{bmatrix} r_0 - p_0 & r_1 - p_1 & \cdots & r_{n-1} - p_{n-1} \end{bmatrix} z \tag{4.4}$$

[3] In the case of complex poles, a pair of complex conjugates.

the controlled system is again in controller canonical form,

$$\dot{z} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -r_0 & -r_1 & \cdots & \cdots & -r_{n-1} \end{bmatrix} z.$$

This is interesting because the characteristic polynomial of the controlled system is $R(s) := s^n + r_{n-1}s^{n-1} + \cdots + r_0$. It shows we have complete "control" over the characteristic polynomial of the controlled system: choose your favorite $s^n + r_{n-1}s^{n-1} + \cdots + r_0$ and then (4.4) does the job! □

In Theorem 3.5.2 we saw that every controllable system (with $n_u = 1$) is isomorphic to a controller canonical form, so the next result is probably not a surprise (but realize that this next result allows any $n_u \geq 1$):

Theorem 4.2.3 (Pole placement). *Consider the system $\dot{x} = Ax + Bu$ with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n_u}$. For every polynomial*

$$R(s) := s^n + r_{n-1}s^{n-1} + \cdots + r_0, \qquad r_k \in \mathbb{R},$$

there exists an $F \in \mathbb{R}^{n_u \times n}$ such that

$$\det(sI - (A - BF)) = R(s)$$

if and only if the system is controllable.

Proof. First of all, note that controllability does not change under a state transformation $z = T^{-1}x$, and that a state feedback $u = -Fx$ remains a state feedback after transformation: $u = -Fx = -F_z z$ with $F_z := FT$. The characteristic polynomial of $\dot{x} = (A - BF)x$ also does not change. Indeed, the transformation

$$A_z := T^{-1}AT, \qquad B_z := T^{-1}B, \qquad F_z := FT \tag{4.5}$$

gives

$$\begin{aligned} \det(sI - (A_z - B_z F_z)) &= \det(sI - T^{-1}(A - BF)T) \\ &= \det(T^{-1}(sI - (A - BF))T) \\ &= \det(sI - (A - BF)). \end{aligned}$$

If $\dot{x} = Ax + Bu$ is uncontrollable, then it follows from the Kalman controllability decomposition (see Eqn. (3.3.1)) that $\chi_{A-BF}(s)$ always has a factor $\chi_{A_{22}}(s)$. So the eigenvalues of A_{22} are eigenvalues of $A - BF$ for every F. Placing the poles arbitrarily is therefore not possible for uncontrollable systems.

Next, suppose that the system is controllable. We first contruct an F for the case $n_u = 1$, so $B \in \mathbb{R}^{n\times 1}$. In the previous example we found that $u = -F_z z$ with

$$F_z = \begin{bmatrix} r_0 - p_0 & r_1 - p_1 & \cdots & r_{n-1} - p_{n-1} \end{bmatrix} \tag{4.6}$$

does the job. Then $F := F_z T^{-1}$ does the job for $\dot{x} = Ax + Bu$.

Now, suppose $n_u > 1$. Unfortunately, it does not follow from the controllability of (A, B) that (A, B_k) is controllable for at least one column B_k of B (see Exercise 3.8). It is a bit more complicated than that. By Heymann's lemma, see Appendix A.7, for every $u_0 \in \mathbb{R}^{n_u}$ for which $b := Bu_0$ is nonzero, there exists an $\tilde{F}$ such that $(A - B\tilde{F}, b)$ is a controllable pair. This reduces the problem to the case $n_u = 1$. Indeed, the controllability of $(A - B\tilde{F}, b)$ implies that for every nth degree monic $R(s)$, there exists an F_1 such that $\chi_{A-B\tilde{F}-bF_1} = R$. Take $F = \tilde{F} + u_0 F_1$. ■

An immediate consequence of this theorem is that every controllable system is stabilizable. Choose, for example, $R(s) = (s+1)^n$; all zeros of this polynomial are in the left half-plane. This theorem also tells us that controllable systems are always stabilizable through static state feedback $u = -Fx$. The question that comes up is: are there systems that are stabilizable, but not through static state feedback $u = -Fx$? We will see that the answer is no, although there are stabilizable systems that are not controllable. Suppose, for example, that in the Kalman controllability decomposition, the system is given by

$$\begin{bmatrix} \dot{z}_{\text{c}} \\ \dot{z}_{\text{uc}} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} z_{\text{c}} \\ z_{\text{uc}} \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u. \tag{4.7}$$

This system is not controllable, but it can be stabilizable. A necessary condition for stabilizability is that $\dot{z}_{\text{uc}} = A_{22} z_{\text{uc}}$ is asymptotically stable, because that part cannot be influenced by u. This is also sufficient:

Theorem 4.2.4 (Stabilizability). *Consider the system $\dot{x} = Ax + Bu$. The following four statements are equivalent:*

1. *There exists an F such that $A - BF$ is asymptotically stable.*

   (So the system is stabilizable through static state feedback $u = -Fx$.)

2. *The system is stabilizable.*

3. *In the Kalman controllability decomposition (4.7) of $\dot{x} = Ax + Bu$ (with (A_{11}, B_1) controllable), the eigenvalues of A_{22} have negative real part.*

4. *$\begin{bmatrix} sI - A & B \end{bmatrix}$ has full row rank for all s with* $\operatorname{Re}(s) \geq 0$.

Proof. We prove the implications (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1). The equivalence of (3) and (4) is an exercise (Exercise 4.11). (1) $\Longrightarrow$ (2): Obvious. (2) $\Longrightarrow$ (3): That $\dot{z}_{\text{uc}} = A_{22} z_{\text{uc}}$ must be asymptotically stable is obvious. (3) $\Longrightarrow$ (1): See Exercise 4.10. (3) $\Longleftrightarrow$ (4): See Exercise 4.11. ■

We again note that this says that systems that are at all stabilizable — by open loop control or closed loop control, nonlinear or whatever method — are always stabilizable through static state feedback $u = -Fx$ as well. Nice.

4.2.1 Ackermann's Pole Placement Formula

The proof of the pole placement theorem (Theorem 4.2.3) is constructive. For $n_u > 1$, the construction of F is quite complicated, but for $n_u = 1$ it is simple and in essence does not use much more than a transformation to the controller canonical form. If we are only interested in F, then we do not even need to carry out the transformation, as we have the following result.

Lemma 4.2.5 (Ackermann's pole placement formula). *Suppose that (A, B) is controllable with $B \in \mathbb{R}^{n\times 1}$ (so $n_u = 1$). Write the characteristic polynomial of A as $\chi_A(s) = s^n + p_{n-1}s^{n-1} + \cdots + p_0$. Given a monic polynomial*

$$R(s) = s^n + r_{n-1}s^{n-1} + \cdots + r_0,$$

there is a unique F such that $\chi_{A-BF} = R$. This F equals

$$F = \begin{bmatrix} r_0 - p_0 & r_1 - p_1 & \cdots & r_{n-1} - p_{n-1} \end{bmatrix} \mathscr{C}_z \mathscr{C}_x^{-1} \tag{4.8}$$

with $\mathscr{C}_x$ and $\mathscr{C}_z$ the controllability matrices of, respectively, the pair (A, B) and (3.28). Equivalently, F can be determined using Ackermann's formula*:*

$$F = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \mathscr{C}_x^{-1} R(A). \tag{4.9}$$

Proof. By (4.5), the vector F is equal to $F = F_z T^{-1}$, with F_z as in (4.6). By Theorem 3.5.2, the matrix T is equal to $T = \mathscr{C}_x \mathscr{C}_z^{-1}$, and so (4.8) follows. The proof of (4.9) is more technical. Let $\eta := \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \mathscr{C}_x^{-1}$, that is,

$$\eta \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Then

$$\eta(A - BF)^k = \eta A^k \qquad \forall k < n \tag{4.10}$$

$$\eta(A - BF)^n = \eta A^n - F. \tag{4.11}$$

By the Cayley–Hamilton theorem, a matrix satisfies its own characteristic polynomial, $R(A - BF) = 0$. Hence, in particular, we have $\eta R(A - BF) = 0$. Using (4.10),(4.11), we can write this expression as

$$\begin{aligned} 0 &= \eta R(A - BF) \\ &= \eta(r_0 I + r_1(A - BF) + \cdots + (A - BF)^n) \\ &= r_0\eta I + r_1\eta A + \cdots + r_{n-1}\eta A^{n-1} + (\eta A^n - F) \\ &= \eta(r_0 I + r_1 A + \cdots + r_{n-1}A^{n-1} + A^n) - F \\ &= \eta R(A) - F. \end{aligned}$$

It follows that $F = \eta R(A)$. ∎

We should note that for large n, Ackermann's formula (4.9) is numerically ill-conditioned.

Example 4.2.6. Consider the system

$$\dot{x} = \begin{bmatrix} -1 & 0 \\ 4 & 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u. \tag{4.12}$$

We want a state feedback $u = -Fx$ that places the eigenvalues of the closed-loop system in -1 and -4. Ackermann's formula gives

$$\begin{aligned}
F &= \begin{bmatrix} 0 & 1 \end{bmatrix} \mathscr{C}_x^{-1} R(A) \qquad \text{use that } R(s) = (s+1)(s+4): \\
&= \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}^{-1} (A+I)(A+4I) \\
&= \underbrace{\underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}}_{\begin{bmatrix} 1/2 & 1/2 \end{bmatrix}} \begin{bmatrix} 0 & 0 \\ 4 & 2 \end{bmatrix}}_{\begin{bmatrix} 2 & 1 \end{bmatrix}} \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 10 & 5 \end{bmatrix}.
\end{aligned}$$

□

4.3 Observers

Many procedures for control of systems are based on the assumption that the entire state vector $x(t)$ can be measured. There is a good reason to use these types of control laws. Indeed, intuitively the state at any particular time contains all information necessary for the future behavior of the system. A control method that wants to influence the future behavior must therefore be based on the state. We have already seen an important example of such a control method in the previous section.

Often, however, we cannot measure the entire state vector. For physical systems, measuring certain quantities requires expensive measuring equipment, while economic systems, for example, require very extensive (statistical) measuring procedures. It may also happen that some state components with internal variables are not directly accessible for measurement. In all these cases, control must be based on the knowledge of *part* of the state vector. From here on, we will assume that this part consists of the output $y = Cx$ of the system (for convenience, we take $y = Cx$ and not the more general form $y = Cx + Du$; see Exercise 4.9).

Example 4.3.1 (Static output feedback). Consider the mass-spring-damper system of Example 2.1.1 and assume that the damping is zero:

$$\begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{bmatrix} \begin{bmatrix} q \\ \dot{q} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u.$$

As output, we take the position $y := q$. By closing the loop with static output feedback, $u = -Hy = -Hq$, for some $H \in \mathbb{R}$, we obtain

$$\begin{aligned}\begin{bmatrix}\dot q\\ \ddot q\end{bmatrix} &= \begin{bmatrix}0 & 1\\ -\frac{k}{m} & 0\end{bmatrix}\begin{bmatrix}q\\ \dot q\end{bmatrix} + \begin{bmatrix}0\\ \frac{1}{m}\end{bmatrix}(-Hq)\\ &= \begin{bmatrix}0 & 1\\ -\frac{k}{m}-\frac{H}{m} & 0\end{bmatrix}\begin{bmatrix}q\\ \dot q\end{bmatrix}.\end{aligned}$$

This has eigenvalues

$$\lambda_{1,2} = \pm\sqrt{-\frac{H+k}{m}}.$$

We see that the sum of the eigenvalues is always zero — regardless of how we choose H — and therefore that the system cannot be stabilized through static output feedback. □

This is a negative result. If only we could measure the entire state. How do we solve this problem? We know that if the system is observable, then based on the knowledge of the input and output during the time interval $[0, t_1]$, we can, *in principle*, uniquely determine the state at any time $t \in [0, t_1]$. The idea is now to let this determination of $x(t)$ be carried out automatically by a system we will call the *observer*. Figure 4.5 shows a block diagram of a system with observer. Both u and y are available to the observer, which uses them to construct an estimate $\hat x$ of x.

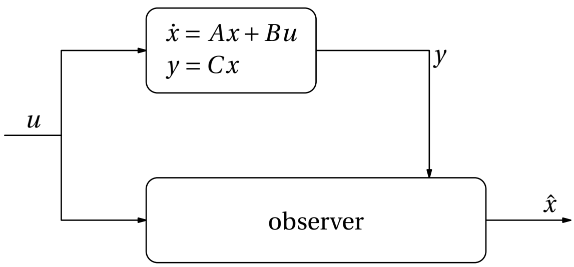


FIGURE 4.5: System with observer.

We do not require the observer to determine the state of the system $x(t)$ *exactly* at every point in time, but do require that it provides an estimate $\hat x(t)$ that improves with time. The idea is that with time, more and more information over the system is available to the observer, which should therefore be able to make better and better estimates. Systems that have such an observer are called detectable:

Definition 4.3.2 (Detectability). A system (4.1) is *detectable* if there exists an *observer* (a map from signals (u, y) to a signal $\hat x$) such that

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

for all initial conditions $x(0)$ and all inputs u. □

The definition of detectability does not say how to construct an observer. In principle we are allowed to use nonlinear and infinite-dimensional observers, but here we restrict ourselves to observers of the form

$$\textit{observer:}\quad \begin{cases} \dot{z} &= Pz + Qu + Ly, \\ \hat{x} &= Sz + Tu + Ry. \end{cases} \tag{4.13}$$

This is a dynamical system with input both u and y (because these are available), state z, and with output the estimate $\hat{x}$ of x. Actually, we restrict the search to an even simpler type of observer with $z = \hat{x}$, that is, $S = I$, $T = 0$, $R = 0$, so

$$\textit{observer:}\quad \dot{\hat{x}} = P\hat{x} + Qu + Ly, \tag{4.14}$$

and we also want it to satisfy the following condition:

Assumption 4.3.3. If $\hat{x}(t_0) = x(t_0)$ at some time t_0, then $\hat{x}(t) = x(t)$ for all $t \geq t_0$. □

In other words, “perfect once, perfect forever”. This assumption says that if the estimation error $e(t)$, defined as

$$e(t) := x(t) - \hat{x}(t),$$

ever becomes zero, that is, $e(t_0) = 0$, then we want $e(t) = 0$ for all $t > t_0$ for all possible inputs. This estimation error satisfies

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} e &= \frac{\mathrm{d}}{\mathrm{d}t}(x - \hat{x}) \\ &= (Ax + Bu) - (P\hat{x} + Qu + LCx) \\ &= (A - LC)x - P\hat{x} + (B - Q)u \\ &= (A - LC)e + (A - LC - P)\hat{x} + (B - Q)u. \end{aligned}$$

Based on this we choose P and Q equal to

$$P = A - LC, \qquad Q = B,$$

because then Assumption 4.3.3 is satisfied and we have

$$\dot{e} = (A - LC)e.$$

Note that the dynamics of the error $e := x - \hat{x}$ is now disconnected from the input! With this choice of P and Q, the observer (4.14) takes the form

$$\dot{\hat{x}} = (A - LC)\hat{x} + Bu + Ly. \tag{4.15}$$

By the way, this form (4.15) is equivalent to

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - \hat{y}) \quad \text{with } \hat{y} := C\hat{x}. \tag{4.16}$$

Both representations (4.15) and (4.16) are depicted in Figure 4.6. Though equivalent, the two representations give rise to different interpretations. The first representation, (4.15), gives the observer in the standard state form. In particular, we see that the observer is asymptotically stable if $A-LC$ is asymptotically stable[4]. The second representation, (4.16), shows that the observer can also be seen as a duplicate of the original system with as additional input a term that depends only on the difference $y(t)-\hat{y}(t)$. As long as $\hat{y}(t)$ is equal to $y(t)$, there is apparently no reason to adjust the dynamics of $\hat{x}$. If, however, $\hat{y}(t)$ differs from $y(t)$, then we cannot have $\hat{x}(t)=x(t)$, and the correction term $L(y(t)-\hat{y}(t))$ in (4.16) then comes into play.

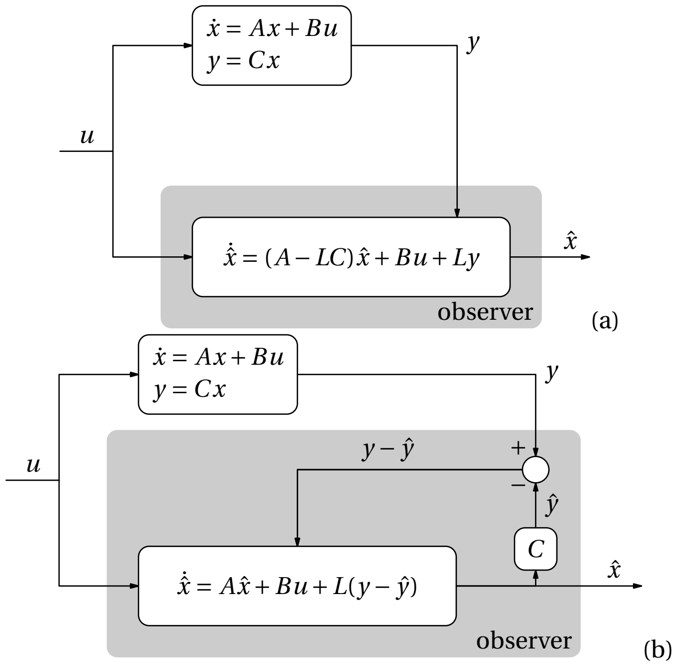


FIGURE 4.6: System with observer. The observers (a) and (b) are identical, but have been implemented differently.

The only freedom left in the observer is the matrix L. We still need to determine whether the estimation error $e(t)=x(t)-\hat{x}(t)$ goes to zero if $t\to\infty$. We have just seen that

$$\dot{e}=(A-LC)e.$$

Now, $\lim_{t\to\infty} e(t)=0$ for every initial condition $e(0)=x(0)-\hat{x}(0)$ if and only if all eigenvalues of $A-LC$ have negative real part. The question is whether we can find a matrix L that achieves this.

[4] A matrix $A\in\mathbb{R}^{n\times n}$ is called *asymptotically stable* if all eigenvalues of A have negative real part.

Theorem 4.3.4 (Observer pole placement). *The pair* (A,C) *is observable if and only if for every polynomial* $R(s) := s^n + r_{n-1}s^{n-1} + \cdots + r_0$*, there exists an* $L \in \mathbb{R}^{n\times n_y}$ *such that* $\det(sI-(A-LC)) = R(s)$.

If $n_y = 1$*, then the matrix L can be determined using Ackermann's formula,*

$$L = R(A)\mathcal{O}_x^{-1}\begin{bmatrix}0\\ \vdots\\ 0\\ 1\end{bmatrix} = \mathcal{O}_x^{-1}\mathcal{O}_z\begin{bmatrix}r_0-p_0\\ \vdots\\ r_{n-1}-p_{n-1}\end{bmatrix}$$

with $\mathcal{O}_x$ *the observability matrix of* (A,C) *and* $\mathcal{O}_z$ *the observability matrix of the transformed system* (3.30).

Proof. This result is the dual of Theorem 4.2.3 and Lemma 4.2.5. Recall that (A,C) is observable if and only if $(A^{\mathrm{T}},C^{\mathrm{T}})$ is controllable. By Theorem 4.2.3, the controllability of $(A^{\mathrm{T}},C^{\mathrm{T}})$ is equivalent to the existence of matrices $F \in \mathbb{R}^{n_y\times n}$ such that $\det(sI-(A^{\mathrm{T}}-C^{\mathrm{T}}F)) = R(s)$ for any choice of a monic nth degree polynomial $R(s)$. Take $L = F^{\mathrm{T}}$. ∎

Hence if (A,C) is observable, then we can certainly make $A-LC$ asymptotically stable. An immediate consequence of the observer pole placement theorem is therefore the following.

Theorem 4.3.5 (Observer). *If the system* (4.1) *is observable, then it is also detectable. In particular, any* $\dot{\hat{x}} = (A-LC)\hat{x} + Bu + Ly$ *is an observer provided we choose L such that* $A-LC$ *is asymptotically stable.* □

The zeros s of $\det(sI-(A-LC))$ are the eigenvalues of the observer and are also called the *observer poles*. As with state feedback, a logical question to ask is whether there are systems that are detectable—regardless of the form of observer—but for which there is no observer of the special form (4.15). The answer is no, which is good news:

Theorem 4.3.6 (Detectability). *Consider the system* $\dot{x} = Ax + Bu, y = Cx$*. The following four statements are equivalent:*

1. *There exists an L such that* $A-LC$ *is asymptotically stable.*

   (So the system is detectable with an observer of the form (4.15)*.)*

2. *The system is detectable.*

3. *In the Kalman observability decomposition, the eigenvalues of* A_{22} *have negative real part.*

4. *The matrix*

   $$\begin{bmatrix}sI-A\\ C\end{bmatrix}$$

   has full column rank for all s with $\mathrm{Re}\,(s) \geq 0$.

Proof. We prove (1) $\implies$ (2) $\implies$ (3) $\implies$ (4) $\implies$ (1).

(1) $\implies$ (2) is trivial.

(2) $\implies$ (3): Suppose $u(t) = 0$ and $x(t) = 0$. Then $y(t) = 0$. Since the system is detectable, there is an observer signal $\hat{x}(t)$ such that $\lim_{t\to\infty} \hat{x}(t) = \lim_{t\to\infty} \hat{x}(t) - x(t) = 0$.

Now, suppose $u(t) = 0$ and $x(t) = \left[\begin{smallmatrix} 0 \\ e^{A_{22}t} z_{uo} \end{smallmatrix}\right]$. This state is in the unobservable space, so we now also have $y(t) = 0$. But then the observer produces the same signal $\hat{x}(t)$ as above, so $\lim_{t\to\infty} \hat{x}(t) = 0$. Because it is an observer, we must have $\lim_{t\to\infty} \hat{x}(t) - x(t) = 0$. Hence $x(t) = \left[\begin{smallmatrix} 0 \\ e^{A_{22}t} z_{uo} \end{smallmatrix}\right]$ also converges to zero when $t \to \infty$. Since z_{uo} can be chosen arbitrarily, A_{22} must be asymptotically stable.

(3) $\implies$ (4): See Exercise 4.12.

(4) $\implies$ (1): It suffices to show that condition (4) ensures the existence of an L for which $A - LC$ is asymptotically stable. This is the dual of (4) $\implies$ (1) of Thm. 4.2.4. ■

In analogy with the stabilizability, we have that systems that are at all detectable (using whatever form of observer such as linear, nonlinear, finite or infinite dimensional, etc.), can also be detected through a dynamical linear observer of the form (4.15). Consequently, choosing this type of dynamic linear observer is not restrictive. Good.

Example 4.3.7 (Juggler). Consider, once again, the inverted pendulum of Example 4.2.1 and suppose that we can measure only the position q of the tip of the pendulum, and not its velocity. So

$$y = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C} \begin{bmatrix} q \\ v \end{bmatrix}.$$

The observer (4.16) for both the position and velocity is of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \hat{q} \\ \hat{v} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ \frac{g}{\ell_1} & 0 \end{bmatrix}}_{A} \begin{bmatrix} \hat{q} \\ \hat{v} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ -\frac{g}{\ell_1} \end{bmatrix}}_{B} u + \underbrace{\begin{bmatrix} l_1 \\ l_2 \end{bmatrix}}_{L} (y - \hat{y}).$$

We determine (l_1, l_2) such that $A - LC$ has characteristic polynomial $R(s) = (s+2)^2 = s^2 + 4s + 4$. (In particular, $A - LC$ is then asymptotically stable.) We have

$$\begin{aligned} \det(sI - (A - LC)) &= \begin{bmatrix} s + l_1 & -1 \\ -\frac{g}{\ell_1} + l_2 & s \end{bmatrix} \\ &= s^2 + l_1 s - \frac{g}{\ell_1} + l_2. \end{aligned}$$

We must therefore choose $l_1 = 4$ and $l_2 = 4 + \frac{g}{\ell_1}$. Done!

Because this system is observable, by Theorem 4.3.4 we can also use Ackermann's formula to determine L:

$$
\begin{aligned}
L &= (A^2 + 4A + 4I)\mathcal{O}^{-1}\begin{bmatrix}0\\1\end{bmatrix}\\
&= \begin{bmatrix}4+g/\ell_1 & 4\\ 4g/\ell_1 & 4+g/\ell_1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}^{-1}\begin{bmatrix}0\\1\end{bmatrix}\\
&= \begin{bmatrix}4\\ 4+g/\ell_1\end{bmatrix}.
\end{aligned}
$$

The result is the same. □

Example 4.3.8 (Observer canonical form). Consider a system in observer canonical form

$$
\begin{aligned}
\dot{x} &= \underbrace{\begin{bmatrix}0 & \cdots & \cdots & 0 & -p_0\\ 1 & \ddots & & \vdots & -p_1\\ 0 & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots\\ 0 & \cdots & 0 & 1 & -p_{n-1}\end{bmatrix}}_{A} x + \begin{bmatrix}q_0\\ q_1\\ \vdots\\ \vdots\\ q_{n-1}\end{bmatrix} u, \\
y &= \underbrace{\begin{bmatrix}0 & \cdots & \cdots & 0 & 1\end{bmatrix}}_{C} x.
\end{aligned}
\tag{4.17}
$$

In this case, it is easy to find a column vector L such that $A - LC$ is asymptotically stable: write L as

$$
L = \begin{bmatrix}l_0\\ l_1\\ \vdots\\ l_{n-1}\end{bmatrix}.
$$

Then we have

$$
A - LC = \begin{bmatrix}0 & \cdots & \cdots & 0 & -p_0 - l_0\\ 1 & \ddots & & \vdots & -p_1 - l_1\\ 0 & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots\\ 0 & \cdots & 0 & 1 & -p_{n-1} - l_{n-1}\end{bmatrix}.
$$

This is again a companion matrix, and the characteristic polynomial is therefore

$$
\chi_{A-LC}(s) = s^n + (p_{n-1} + l_{n-1})s^{n-1} + \cdots + (p_0 + l_0).
$$

It is clear that the coefficients of this polynomial can be chosen arbitrarily by choosing L accordingly. □

It looks like we can let observers act as quickly as we want. If we, for example, take $\det(sI-(A-LC)) = (s+100)^n$, then all eigenvalues of $A-LC$ are in $s = -100$ and consequently every solution $e(t)$ of $\dot{e} = (A-LC)e$ can be written as a linear combination of functions of the form $\mathrm{e}^{-100t} t^k$. This ensures that the estimation error $e(t)$ converges to zero very quickly, which implies that based on the output y, we will, in no time, have at our disposal an almost perfect estimate of x. This conclusion seems to contradict our discussion on page 126 that it is often impossible to measure entire state vectors. The problem is that fast (aggressive) observers are usually very sensitive to modeling errors and noise in the measurement of y, and although in theory they can be made arbitrarily fast, the presence of, for example, noise in the measurements limits the maximum speed. A proper consideration of this issue requires a stochastic setting—a subject in its own right. We will not discuss stochastics here, we just illustrate the problem:

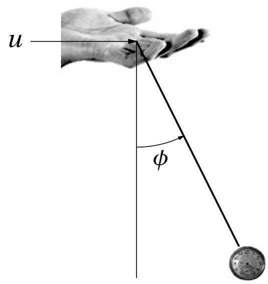


FIGURE 4.7: The hypnotist (Example 4.3.9).

Example 4.3.9 (Application: Hypnotist). A variation on the juggler is the hypnotist. Instead of stabilizing, the hypnotist wants to keep the pendulum in motion; see Figure 4.7. We denote the horizontal position of the hand by u and the angle the pendulum makes with the vertical axis by ϕ. The horizontal displacement of the pendulum (or watch) q is therefore $q = u + \ell \sin(\phi)$. Without derivation, we claim that the linearized equation of motion is

$$\ddot{q} + \frac{b}{m}\dot{q} + \frac{g}{\ell}q = \frac{g}{\ell}u.$$

Here ℓ is the length of the chord, b is a positive friction coefficient, g is the gravitational acceleration, and m is the mass of the watch. We take the values $m = 0.1[\mathrm{kg}]$, $\ell = 0.4[\mathrm{m}]$, $b = 0.05[\mathrm{kg/s}]$, and $g = 10[\mathrm{m/s^2}]$, so

$$\ddot{q} + 0.5\dot{q} + 2.5q = 2.5u.$$

With state variables q and $v := \dot{q}$, we get

$$\begin{bmatrix} \dot{q} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2.5 & -0.5 \end{bmatrix} \begin{bmatrix} q \\ v \end{bmatrix} + \begin{bmatrix} 0 \\ 2.5 \end{bmatrix} u$$
$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} q \\ v \end{bmatrix}.$$

We deliberately do not choose the observer canonical form, because that does not give the state variables we want.

We first take a nonaggressive observer, that is, one where the correction term $L(y-\hat{y})$ is not large. In this example, we could even take $L=0$, because the system itself is already asymptotically stable (with poles $-0.25\pm 1.56\mathrm{i}$). We choose L "small", $L=\left[\begin{smallmatrix}0.5\\0.5\end{smallmatrix}\right]$. Then the observer is

$$\begin{bmatrix}\dot{\hat{q}}\\ \dot{\hat{v}}\end{bmatrix}=\begin{bmatrix}-0.5 & 1.0\\ -3.0 & -0.5\end{bmatrix}\begin{bmatrix}\hat{q}\\ \hat{v}\end{bmatrix}+\begin{bmatrix}0\\ 2.5\end{bmatrix}u+\begin{bmatrix}0.5\\ 0.5\end{bmatrix}y.$$

The observer poles are $-0.5\pm 1.73\mathrm{i}$. Figure 4.8(top) shows the actual state (q,v) for $u(t)=\cos(\pi t)$ and $q(0)=2, v(0)=1$, and the state $(\hat{q},\hat{v})$ estimated by the observer (red, dashed). It seems that the observer constructs a very good estimate of (q,v) in less than 10 seconds. Because the input is a sinusoid, the states do not converge to zero. We will see later on (Chapter 5) that the states converge to sinusoids. This convergence is slower than the convergence of $(\hat{q},\hat{v})$ to (q,v) because the real part of the system poles

$$\operatorname{Re}(-0.25\pm 1.56\mathrm{i})=-0.25$$

is less negative than the real part of the observer poles

$$\operatorname{Re}(-0.5\pm 1.73\mathrm{i})=-0.5.$$

Next we take a more aggressive observer. We take $L=\left[\begin{smallmatrix}5\\5\end{smallmatrix}\right]$. The observer now becomes

$$\begin{bmatrix}\dot{\hat{q}}\\ \dot{\hat{v}}\end{bmatrix}=\begin{bmatrix}-5 & 1\\ -7.5 & -0.5\end{bmatrix}\begin{bmatrix}\hat{q}\\ \hat{v}\end{bmatrix}+\begin{bmatrix}0\\ 2.5\end{bmatrix}u+\begin{bmatrix}5\\ 5\end{bmatrix}y.$$

and its poles are $-2.75\pm 1.56\mathrm{i}$. With the same $u(t)=\cos(\pi t)$ and $q(0)=2, v(0)=1$ as before, the observer is much faster now; see Figure 4.8(bottom).

The more aggressive (faster) observer produces better results. But what happens if the measurement y of the position q is not perfect? We model the measurement error as an additional signal on q, and for simplicity assume that the measurement error is always $1/2$:

$$y=q+1/2.$$

With this perturbed y, it is the less aggressive (slower) observer that outperforms the more aggressive (faster) observer; see Figure 4.9.

For our system, there is also an observer that is completely independent of measurement errors in y; see Exercise 4.1g. □

4.4 Dynamical Output Feedback

We now have enough results to construct a control system $u=\mathscr{K}(y)$ that generates a signal u that stabilizes the given system, based only on the output y (and

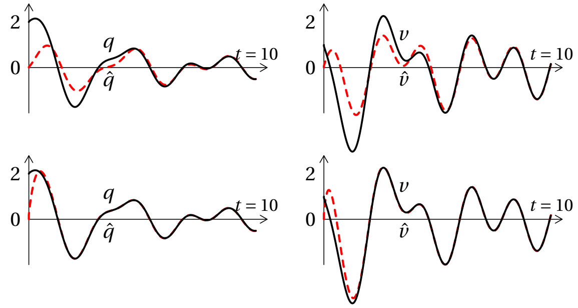


FIGURE 4.8: The actual (q, v) (black) and estimated $(\hat{q}, \hat{v})$ (red, dashed) for a slow observer (top) and for a fast observer (bottom). See Example 4.3.9.

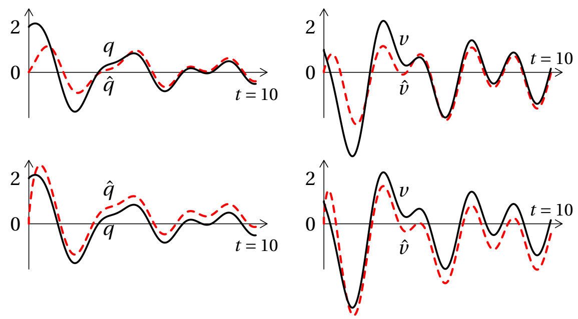


FIGURE 4.9: Actual (q, v) (black) and estimated $(\hat{q}, \hat{v})$ (red, dashed) for a slow observer (top) and for a faster observer (bottom). See Example 4.3.9.

not on the entire state). See Figure 4.2. As before the given system is assumed of the form

$$\textit{given system:} \quad \begin{cases} \dot{x} &= Ax + Bu, \\ y &= Cx \end{cases} \tag{4.18}$$

and as control system (aka *controller*), we propose a dynamical system that determines an estimate $\hat{x}$ of the state x of the given system from (u, y) using an observer, and that provides $u = -F\hat{x}$ based on this estimate. So,

$$\textit{controller:} \quad \begin{cases} \dot{\hat{x}} &= (A - LC)\hat{x} + Bu + Ly \quad \text{(observer)}, \\ u &= -F\hat{x} \qquad\qquad\qquad\quad\; \text{(feedback)}. \end{cases}$$

We can eliminate the term Bu in the observer by substituting $u = -F\hat{x}$, and this gives the controller in the standard form (i.e. with input y and output u),

$$\textit{controller:} \quad \begin{cases} \dot{\hat{x}} &= (A - LC - BF)\hat{x} + Ly, \\ u &= -F\hat{x}. \end{cases} . \tag{4.19}$$

The combination of the given system (4.18) and the controller (4.19) is called the closed-loop system, and it is described by

$$\textit{closed loop:} \quad \begin{bmatrix} \dot{x} \\ \dot{\hat{x}} \end{bmatrix} = \begin{bmatrix} A & -BF \\ LC & A - BF - LC \end{bmatrix} \begin{bmatrix} x \\ \hat{x} \end{bmatrix},$$

(verify this yourself). In the previous section, we saw that the dynamics of the estimation error $e := x - \hat{x}$ satisfy $\dot{e} = (A - LC)e$ and that these dynamics do not depend on u. So the same must hold here. Indeed,

$$\begin{aligned} \dot{e} &= \dot{x} - \dot{\hat{x}} \\ &= (Ax - BF\hat{x}) - (LCx + (A - BF - LC)\hat{x}) \\ &= (A - LC)(x - \hat{x}) \\ &= (A - LC)e. \end{aligned}$$

The dynamics of x in terms of x and e simplify to

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

The behavior of the closed loop can therefore equivalently be described by

$$\textit{closed loop:} \quad \begin{bmatrix} \dot{x} \\ \dot{e} \end{bmatrix} = \begin{bmatrix} A - BF & BF \\ 0 & A - LC \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix}.$$

Because of zero block here we can infer that the eigenvalues of the closed-loop system are equal to the eigenvalues of $A - BF$ together with the eigenvalues of $A - LC$! The conclusion is that the eigenvalues of the closed-loop system are equal to the eigenvalues we would get through the state feedback $u = -Fx$ together with the eigenvalues of the observer. This leads us to the central result of this chapter:

Theorem 4.4.1 (Stabilizing dynamical controller). *If the system (4.18) is stabilizable and detectable, then there exist matrices F and L such that $A-BF$ and $A-LC$ are asymptotically stable. In that case, the controller (4.19) stabilizes the system (4.18), in the sense that* $\lim_{t\to\infty} x(t) = 0$ *and* $\lim_{t\to\infty} \hat{x}(t) = 0$ *for all initial conditions* $x(0) = x_0$ *and* $\hat{x}(0) = \hat{x}_0$. □

This stabilization process is an example of a so-called *separation principle*: we can determine a state feedback law $u = -Fx$ and an estimate $\hat{x}$ for x independently of each other. Connecting the two using $u = -F\hat{x}$ gives the desired result.

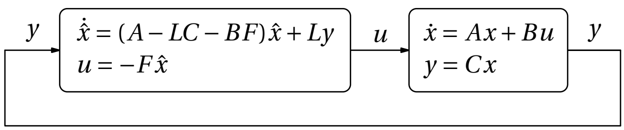


FIGURE 4.10: System with controller.

Figure 4.10 shows the closed-loop system schematically. The controller (the system at the top left) is also called a *compensator*. In contrast to observers, the controller has only y as input.

Example 4.4.2 (Juggler). In Example 4.2.1, we used $u = -Fx$ to place the eigenvalues of $A-BF$ in -1. For $\ell_1 = \frac{1}{2}g$, this gave $F = \begin{bmatrix} -3/2 & -1 \end{bmatrix}$. In Example 4.3.7, we placed the eigenvalues of the observer in -2. For $\ell_1 = \frac{1}{2}g$, this gave $L = \left[\begin{smallmatrix} 4 \\ 6 \end{smallmatrix}\right]$. Combining the observer and feedback $u = -F\hat{x}$ then gives (still for $\ell_1 = \frac{1}{2}g$)

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} q \\ v \\ \hat{q} \\ \hat{v} \end{bmatrix} = \underbrace{\left[\begin{array}{cc:cc} 0 & 1 & 0 & 0 \\ 2 & 0 & -3 & -2 \\ \hdashline 4 & 0 & -4 & 1 \\ 6 & 0 & -7 & -2 \end{array}\right]}_{=\left[\begin{smallmatrix} A & -BF \\ LC & A-BF-LC \end{smallmatrix}\right]} \begin{bmatrix} q \\ v \\ \hat{q} \\ \hat{v} \end{bmatrix}$$

with eigenvalues $-1, -1, -2, -2$.

The controller is

$$\begin{aligned} \dot{\hat{x}} &= \underbrace{\begin{bmatrix} -4 & +1 \\ -7 & -2 \end{bmatrix}}_{A-LC-BF} \hat{x} + \underbrace{\begin{bmatrix} 4 \\ 6 \end{bmatrix}}_{L} y, \\ u &= \underbrace{\begin{bmatrix} 3/2 & 1 \end{bmatrix}}_{-F} \hat{x}. \end{aligned}$$

One can check that here, the controller itself is also asymptotically stable. It is good to realize that this does not need to hold. In fact, it can be shown that certains systems can only be stabilized with an unstable controllers! □

In practice, controllers are also applied to *stable* systems. The aim is then to regulate the behavior in some other way. For instance, to speed up to convergence of the signals (see Exercise 4.15) or the steer the output to some *nonzero* value, possibly set by the user (think of the heating system where you set the desired room temperature). We will come back to this type of control in the next chapter.

4.5 Exercises

4.1 Comprehension questions (on the whole chapter). Prove or give a counterexample.

(a) If the system $\dot{x} = Ax + Bu$ is stabilizable, then for every $x(0)$ there exists a u such that $x(10) = 0$.

(b) If (A, B) is stabilizable, then so is $(A - BM, B)$.

(c) If $(A - MC, C)$ is detectable, then so is (A, C).

(d) If (A, B) is stabilizable, then so is $(A - LC, B)$.

(e) If $-A$ is asymptotically stable and (A, B) is stabilizable, then (A, B) is controllable. [This one is complicated!]

(f) If A is asymptotically stable, then (A, C) is detectable.

(g) If A is asymptotically stable, then $\dot{\hat{x}} = A\hat{x} + Bu$ is an observer.

4.2 Consider the system

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ \beta \end{bmatrix} u, \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x. \end{aligned}$$

(a) For which β is the system controllable?

(b) Show that $\frac{\mathrm{d}}{\mathrm{d}t}(-\beta x_1 + x_2) = (x_1 - \beta x_2)$, and explain why this implies that the system is not controllable if $\beta = \pm 1$.

(c) Is the system observable?

(d) Take $\beta = 2$. Give a state feedback $u = -Fx$ such that the closed loop has characteristic polynomial $s^2 + s + 1$.

(e) Take $\beta = 2$. Give an observer with double eigenvalue -1.

4.3 Consider the system

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u, \\ y &= \begin{bmatrix} 1 & 1 \end{bmatrix} x \end{aligned}$$

with $\alpha \in \mathbb{R}$.

(a) For which α is the system controllable?

(b) For which α is the system observable?

(c) Determine the characteristic polynomial of $\left[\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right]$.

(d) Determine the observer canonical form of the system (if it exists).

(e) Take $\alpha = 1$. Determine the controller canonical form of the system (if it exists).

(f) Take $\alpha = 1$. Determine a state feedback $u = -Fx$ that places the eigenvalues of $A - BF$ in $s = -2$ [twice].

(g) Show that if this system is not controllable, then the system is not stabilizable through a static state feedback $u = -Fx$.

4.4 *Third-order system.* We are given a system with

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}. \tag{4.20}$$

(a) Is the system observable? Is it controllable?

(b) Is $x = 0$ an asymptotically stable equilibrium point of $\dot{x} = Ax$?

(c) Is it possible to make the system asymptotically stable using a static output feedback $u(t) = -Hy(t)$? [Hint: you may want to know that for polynomials of the form $\lambda^n - c\lambda^{n-1} + \cdots$ the constant c equals the sum of all zeros of the polynomial[5].]

(d) Determine a state feedback $u = -Fx$ such that the eigenvalues of $A - BF$ are in $-1 \pm 2\mathrm{i}, -2$. [Hint: use properties of the companion matrix.]

4.5 *State feedback.* We are given the system $\dot{x} = Ax + Bu$ with

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

(a) Is $x = 0$ an asymptotically stable equilibrium point of $\dot{x} = Ax$?

(b) Is the system $\dot{x} = Ax + Bu$ controllable?

(c) Can we use the state feedback $u = -f_0 x_1 - f_1 x_2 - f_2 x_3 - f_3 x_4$ to place the eigenvalues of $\dot{x} = Ax + Bu$ in

i. $-2, -2, -1, -1$

ii. $-2, -2, -2, -1$

iii. $-2, -2, -2, -2$

[5] Just expand $\prod_{i=1}^{n} (\lambda - \lambda_i)$.

4.6 *Deadbeat control.* In some aspects, discrete-time systems are fundamentally different from continuous-time systems. Consider the n-dimensional discrete-time system

$$x[t+1] = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -p_0 & -p_1 & \cdots & \cdots & -p_{n-1} \end{bmatrix}}_{A} x[t] + \underbrace{\begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix}}_{B} u[t].$$

Suppose that the input quantity is chosen to be $u[t] = -Fx[t]$. Determine $F \in \mathbb{R}^{1\times n}$ such that all eigenvalues of the resulting feedback system lie in the origin. Verify that $x[t] := (A - BF)^t x[0]$ is then zero for all $t \geq n$.

4.7 *Feedback-2.* Consider (4.1) with A, B, C as in (4.20).

(a) Give an observer with eigenvalues -4, -5, -1.

(b) Give a state feedback for the system such that after applying state feedback, the system has eigenvalues $-1 \pm \mathrm{i}, -2$.

(c) Give a state representation of the stabilizing controller obtained using parts (a) and (b).

4.8 *Feedback-3.* Consider (4.1) with A, B, C as in (4.20). Give an observer with eigenvalues -2, -2, -3.

4.9 In Section 4.3, observers are constructed for systems of the form $\dot{x} = Ax + Bu,\ y = Cx + Du$ with $D = 0$. What adjustments of those observers are needed for the case $D \neq 0$?

4.10 *Stabilizability.* Prove the implication (3) $\Longrightarrow$ (1) of Thm. 4.2.4.

4.11 *Stabilizability.* Prove the equivalence of parts (3) and (4) of Thm. 4.2.4.

4.12 *Detectability.* Prove the implication (3) $\Longrightarrow$ (4) of Thm. 4.3.6.

4.13 *Mass-spring-damper system.* Consider the mass-spring-damper system of Example 2.1.1 and take $m = 1$.

(a) For which values of $k \geq 0, r \geq 0$ (and $m = 1$) is the system of Example 2.1.1

i. asymptotically stable

ii. controllable

iii. observable

(b) Determine a state feedback $u = -Fx$ that places the poles of the closed-loop system in -1 and -2. (Don't forget that F also depends on k and r.)

(c) Determine an observer with observer poles in -4 and -5. (Here too, the answer depends on k and r.)

(d) Using parts (b) and (c), determine a controller that stabilizes the system.

(e) Suppose that $u = -Fx$ stabilizes the system. For which constant v does $u = -Fx + v$ bring the mass to a rest 1 meter to the right of the equilibrium point?

4.14 Construct an observer for the nonlinear system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + g(u(t)), \\ y(t) &= Cx(t) \end{aligned}$$

with g an arbitrary function. (Assume that (A, C) is observable.)

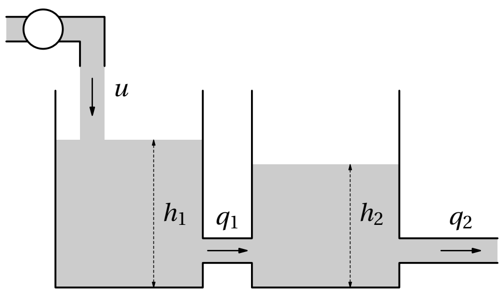


FIGURE 4.11: Two water tanks.

4.15 *Two-tank system.* Figure 4.11 shows a serial interconnection of two water tanks. The variables u, q_1, q_2 denote the water flow, and h_1 and h_2 denote the water heights in tanks 1 and 2, respectively. We linearize the system around a constant equilibrium solution (one of the many). That is, we write the variables as

$$\begin{aligned} u(t) &= u^* + \delta_u(t), \\ q_j(t) &= q_j^* + \delta_{q_j}(t), \\ h_j(t) &= h_j^* + \delta_{h_j}(t). \end{aligned}$$

Assuming that the tanks are identical and that q_1 depends only on the height *difference* $h_1 - h_2$, this gives the linearized model

$$\begin{bmatrix} \dot{\delta}_{h_1} \\ \dot{\delta}_{h_2} \end{bmatrix} = \frac{1}{S} \begin{bmatrix} -\frac{1}{R_1} & \frac{1}{R_1} \\ \frac{1}{R_1} & -\frac{1}{R_1} - \frac{1}{R_2} \end{bmatrix} \begin{bmatrix} \delta_{h_1} \\ \delta_{h_2} \end{bmatrix} + \frac{1}{S} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \delta_u, \tag{4.21}$$

where S is the area of the cross section of the tanks and the R_j are resistances.

For simplicity, take $S = R_1 = R_2 = 1$.

(a) Determine a state feedback that places the two poles in $-1 \pm \mathrm{i}$.

(b) Suppose that we can measure only the height of the first tank, $y := \delta_{h_1}$. Determine an observer with observer poles in -2 and -3.

(c) Using parts (a) and (b), determine a controller that stabilizes the system.

Remark. The given system (4.21) is itself already asymptotically stable, but the eigenvalues for $S = R_1 = R_2 = 1$ are -2.618 and -0.382, and as the latter is "close" to zero, fluctuation around the equilibrium point will die out only "slowly". For the controller you have constructed, the fluctuations die out more quickly because the eigenvalues of the closed-loop system are further away from the imaginary axis (are more negative).

4.16 Consider the system

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u, \\ y &= \begin{bmatrix} 1 & 1 \end{bmatrix} x \end{aligned}$$

with $\alpha \in \mathbb{R}$.

(a) For which α_{nondetec} is the system not detectable?

(b) Determine, for arbitrary α, an L_α for which $A - L_\alpha C$ has double eigenvalue -1.

(c) What happens to L_α when $\alpha \to \alpha_{\text{nondetec}}$? Why is this not surprising?

Tougher Exercises

4.17 It is not unusual to only be able to measure a state (or part of it) with some delay. Consider the system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= x(t-\eta) \end{aligned}$$

with $\eta > 0$. From $t = 0$ on, we connect this system to the "observer"

$$\hat{x}(t) = \mathrm{e}^{A\eta} y(t) + \int_{\max(0,t-\eta)}^{t} \mathrm{e}^{A(t-\tau)} Bu(\tau)\,\mathrm{d}\tau, \qquad t > 0$$

Show that $\hat{x}(t) = x(t)$ for all $t > \eta$.

4.18 *Delays.* In practice, measurements are often accompanied by delay. Determine an observer for the system (4.1) that has input $u(t)$ and delayed input $y(t-1)$, and for which $\lim_{t\to\infty} \|\hat{x}(t) - x(t)\| = 0$.

4.19 *Container transfer.* Consider once more the container transfer system. In particular, consider the linearization (3.32) on page 110. A reasonable mathematical model for how someone would direct the cart is

$$u(t) = c\big(r(t) - x_m(t)\big) - k\dot{x}_m(t), \qquad k, c > 0. \tag{4.22}$$

The term $c\big(r(t) - x_m(t)\big)$ is proportional to the distance to the point to which we want to send the cart. It is positive if the cart is to the left of $r(t)$, and negative it it is to the right of $r(t)$. So $c(d(t) - x_m(t))$ is a force in the direction of the target $r(t)$. To prevent too large accelerations, we have added the term $-k\dot{x}_m(t)$.

(a) Does this $u(t)$ stabilize the system? [Difficult?]

(b) Do you have any idea whether this $u(t)$ stabilizes the nonlinear system from every x_0? [Very difficult.]

Chapter 5

LTI Systems

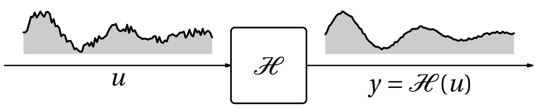


FIGURE 5.1: System with input u and output y.

In the previous three chapters, we concentrated on the state x of the system. The input u and output y played a less prominent role and were mainly tasked with observing and controlling the state.

In this chapter, we focus instead on the external behavior (u, y) of the system, often disregarding the state. We restrict ourselves to systems where the output y is completely determined by the input u, so

$$y = \mathscr{H}(u)$$

for some mapping $\mathscr{H}$. Again we call y the *response of the system* to u.

The analysis will mostly take place in what is called the *frequency domain.* The frequency domain is an alternative for the time domain. In many applications, it are the frequency properties that are important. Think, for example, of music, sonar, radar, and seismologic applications. In medical applications, too, signals with specific frequency properties are used. There are loads of other examples, including examples where one does not immediately think of frequencies. For instance, a cruise controller for a car is not designed to compensate for very fast fluctuations (high frequencies) such as those resulting from wind bursts, but rather to compensate for the influence of conditions that vary slowly (low frequencies), such as those resulting from a constant slope in the road.

5.1 LTI Systems

In this chapter, we restrict ourselves to systems $y = \mathscr{H}(u)$ with a one input and one output: $u, y : \mathbb{R} \to \mathbb{R}$. Usually we also assume linearity and time invariance:

Definition 5.1.1 (Linearity and time invariance—LTI). A system $y = \mathscr{H}(u)$ is *linear* if for all possible inputs u, u_1, u_2 and scalars λ we have

1. additivity: $\mathscr{H}(u_1 + u_2) = \mathscr{H}(u_1) + \mathscr{H}(u_2)$;
2. homogeneity: $\mathscr{H}(\lambda u) = \lambda \mathscr{H}(u)$.

A system is *time invariant* if the response to the delayed input is equal to the delayed response, that is,

$$\mathscr{H}(\sigma^\tau u) = \sigma^\tau \mathscr{H}(u) \qquad \forall \tau \in \mathbb{R}$$

for all possible inputs u and time delays $\tau \in \mathbb{R}$. We call a system *LTI* if it is both linear and time invariant. □

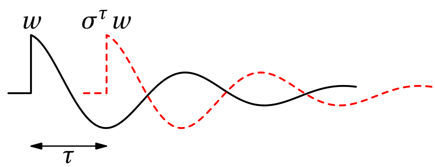
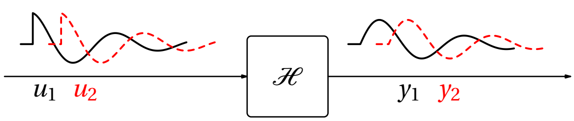


FIGURE 5.2: Graphs of a signal $w : \mathbb{R} \to \mathbb{R}$ and its delayed signal $\sigma^\tau w : \mathbb{R} \to \mathbb{R}$ for some $\tau > 0$.

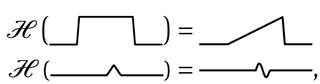
LTI is an initialism for “linear time-invariant”. The *delay operator* or *shift operator* σ^τ, used in the above definition, operates on signals, and is defined as

$$(\sigma^\tau w)(t) = w(t - \tau).$$

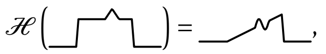
Figure 5.2 illustrates the delay operator. Roughly speaking, the time invariance property $\mathscr{H}(\sigma^\tau u) = \sigma^\tau \mathscr{H}(u)$ means that the system properties do not depend on time. In a time-invariant system, it does not matter whether we experiment with the system today or tomorrow. In other words, in a time-invariant system, the response to the delayed input is equal to the delayed response. Time invariance can be illustrated nicely:



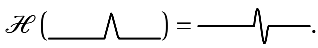
Here $y_1 = \mathscr{H}(u_1)$, $y_2 = \mathscr{H}(u_2)$, and $u_2(t) = u_1(t - \tau)$. Also linearity can be illustrated: suppose we know that



then additivity implies that



and homogeneity implies that



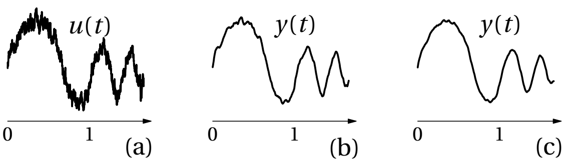


FIGURE 5.3: (a): A signal with noise; (b) averaged over $P = 0.05$; (c) averaged over $P = 0.1$.

Example 5.1.2 (Moving-average system). Figure 5.3(a) shows a signal $u(t)$ with noise. By averaging this signal for every time t over an interval of a certain length $P > 0$, we obtain the signal

$$y = \mathscr{H}(u): \qquad y(t) = \frac{1}{P}\int_{t-P}^{t} u(\tau)\,\mathrm{d}\tau.$$

This average $y(t)$ is in general smoother than $u(t)$, but as long as P is not too large, $y(t)$ will follow the course of $u(t)$ reasonably well. Figures 5.3(b) and (c) show the moving average $y(t)$ for $P = 0.05$ and $P = 0.1$, respectively. As expected, $y(t)$ is smoother in part (c) than in part (b).

We claim that this system is LTI. It is linear because

$$\begin{aligned}\mathscr{H}(u_1 + u_2)(t) &= \frac{1}{P}\int_{t-P}^{t} (u_1 + u_2)(\tau)\,\mathrm{d}\tau \\ &= \frac{1}{P}\int_{t-P}^{t} u_1(\tau)\,\mathrm{d}\tau + \frac{1}{P}\int_{t-P}^{t} u_2(\tau)\,\mathrm{d}\tau \\ &= \mathscr{H}(u_1)(t) + \mathscr{H}(u_2)(t)\end{aligned}$$

and for all scalars λ,

$$\begin{aligned}\mathscr{H}(\lambda u)(t) &= \frac{1}{P}\int_{t-P}^{t} (\lambda u)(\tau)\,\mathrm{d}\tau \\ &= \lambda\frac{1}{P}\int_{t-P}^{t} u(\tau)\,\mathrm{d}\tau \\ &= \lambda\mathscr{H}(u)(t).\end{aligned}$$

The system is also time invariant. To show this, let $y_0 = \mathscr{H}(u_0)$ and define the delayed signals $\tilde{u}(t) := u_0(t-t_0)$ and $\tilde{y}(t) := y_0(t-t_0)$. We must show that $\mathscr{H}(\tilde{u}) = \tilde{y}$. This is true because

$$\begin{aligned}
\mathscr{H}(\tilde{u})(t) &= \frac{1}{P}\int_{t-P}^{t} \tilde{u}(\tau)\,\mathrm{d}\tau \\
&= \frac{1}{P}\int_{t-P}^{t} u(\tau-t_0)\,\mathrm{d}\tau \quad \{\text{substitute } s=\tau-t_0\} \\
&= \frac{1}{P}\int_{(t-t_0)-P}^{(t-t_0)} u(s)\,\mathrm{d}s \\
&= y_0(t-t_0) \\
&= \tilde{y}(t),
\end{aligned}$$

so the system is also time invariant. □

One could say that time-invariant systems "have no built-in clock".

Example 5.1.3 (Linear time-varying system). Let

$$y(t) = t u(t), \qquad t \in \mathbb{R}.$$

This system is linear (verify this yourself), but it is not time invariant. To show this, it suffices to give one counterexample to the time invariance. Choose, for example, $u_0(t) = 1$ (the constant function 1) and its response $y_0(t) = t$, so:

$$\begin{aligned}
u_0(t) &= 1, \\
y_0(t) &= t.
\end{aligned}$$

Now consider their delayed copies

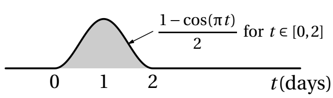
$$\begin{aligned}
\tilde{u}(t) &= u_0(t-1) = 1, \\
\tilde{y}(t) &= y_0(t-1) = t-1.
\end{aligned}$$

They do not satisfy the system equation because the response to $\tilde{u}(t) = 1$ is $t\tilde{u}(t) = t$ and not $\tilde{y}(t) = t-1$. So the system is not time invariant and therefore not LTI. □

Example 5.1.4 (Drainage system). A nice illustration of a time-invariant system is a drainage system. As input $u(t)$, we take the amount of liters of water that fall onto a soccer field per unit of time, and the output $y(t)$ is the amount of liters of water that flow into the river per unit of time. If it rains one day, for example

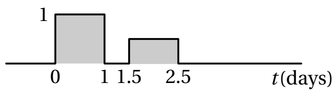
$u_0(t) =$ 

then, if nothing evaporates, all of the water will eventually flow into the river, but this takes time. The outflow for our $u_0(t)$ might be something like

$y_0(t) =$  .

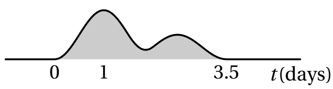
It takes a while before the outflow $y(t)$ begins, and after the last drop of rain $u(t)$, it takes another day before the outflow is over. It is natural to assume that the system is time invariant, because if it starts raining a day later, the water will also flow into the river a day later. If the system is also linear, then for an influx

$$u_1(t) = u_0(t) + \tfrac{1}{2} u_0(t-1.5)$$

$=$ 

we have the outflow

$$\begin{aligned} y_1(t) &= \mathscr{H}(u_0 + \tfrac{1}{2}\sigma^{1.5} u_0)(t) \\ &= \mathscr{H}(u_0)(t) + \tfrac{1}{2}\mathscr{H}(\sigma^{1.5} u_0)(t) \qquad \text{(linearity)} \\ &= \mathscr{H}(u_0)(t) + \tfrac{1}{2}\sigma^{1.5}\mathscr{H}(u_0)(t) \qquad \text{(time inv.)} \\ &= y_0(t) + \tfrac{1}{2} y_0(t-1.5) \end{aligned}$$

$=$ 

Incidentally, it is good to realize that because of the conservation of mass, we always have "total influx = total outflow": $\int_{-\infty}^{\infty} u(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} y(t)\,\mathrm{d}t$. □

We will shortly see that state models can also be viewed as LTI systems, but for that we first need a special result that says that LTI is essentially the same as convolution. We will rarely do computations with convolutions, but for the analysis of the system, this correspondence is important.

Definition 5.1.5 (Impulse response). The *impulse response* h of a system $y = \mathscr{H}(u)$ is the response to the delta function:

$$h = \mathscr{H}(\delta).$$

□

Figure 5.4 illustrates the impulse response. It is nothing more than the output y we obtain if we take the delta function for input u. Roughly speaking the impulse response $\mathscr{H}(\delta)$ is the output that we perceive if we strike the system with an enormous punch at time zero.

Theorem 5.1.6 (LTI = convolution). *Let $u, y : \mathbb{R} \to \mathbb{R}$ and assume that the impulse response $h := \mathscr{H}(\delta)$ exists. Then a system $y = \mathscr{H}(u)$ is LTI if and only if the output is the convolution of h and u,*

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

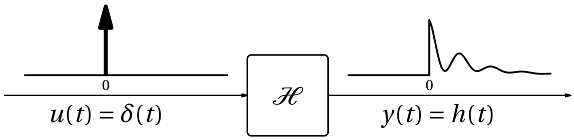


FIGURE 5.4: The impulse response h by definition is the response of the system to the impulse input $u = \delta$.

Proof (sketch). Suppose that the system is a convolution. The linearity follows from the fact that $h * (u_1 + u_2) = (h * u_1) + (h * u_2)$ and $h * (\lambda u_1) = \lambda(h * u_1)$ for all $\lambda \in \mathbb{R}$ and all $u_1, u_2 : \mathbb{R} \to \mathbb{R}$ (we assume that the convolution exists for all such inputs). The time invariance follows from the fact that the response to the delayed signal is the delayed response,

$$\begin{aligned}
h * \sigma^{t_0} u &= \int_{-\infty}^{\infty} h(\tau)(\sigma^{t_0} u)(\cdot - \tau)\,\mathrm{d}\tau \\
&= \int_{-\infty}^{\infty} h(\tau) u((\cdot - \tau) - t_0)\,\mathrm{d}\tau \\
&= \int_{-\infty}^{\infty} h(\tau) u((\cdot - t_0) - \tau)\,\mathrm{d}\tau \\
&= (h * u)(\cdot - t_0) \\
&= \sigma^{t_0}(h * u).
\end{aligned}$$

Convolution systems are therefore LTI.

Next, suppose that the system is LTI. Define $h = \mathscr{H}(\delta)$. Because of the time invariance, for every shift $\tau \in \mathbb{R}$ we have $h(\cdot - \tau) = \mathscr{H}(\delta(\cdot - \tau))$. Moreover, from the sifting property of delta functions, we have

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

Using this, we can write $y = \mathscr{H}(u)$ as a convolution:

$$\begin{aligned}
y &= \mathscr{H}(u) \\
&= \mathscr{H}\Big(\int_{-\infty}^{\infty} \delta(\cdot - \tau) u(\tau)\,\mathrm{d}\tau\Big) && \text{(sifting property)} \\
&= \int_{-\infty}^{\infty} \mathscr{H}\big(\delta(\cdot - \tau) u(\tau)\big)\,\mathrm{d}\tau && \text{(linearity)} \\
&= \int_{-\infty}^{\infty} \mathscr{H}\big(\delta(\cdot - \tau)\big) u(\tau)\,\mathrm{d}\tau && \text{(linearity)} \\
&= \int_{-\infty}^{\infty} h(\cdot - \tau) u(\tau)\,\mathrm{d}\tau && \text{(time invariance)} \\
&= \int_{-\infty}^{\infty} h(\tau) u(\cdot - \tau)\,\mathrm{d}\tau \\
&= h * u.
\end{aligned}$$

In other words, LTI systems are convolutions. ■

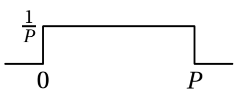
LTI systems hence are completely determined by the impulse response $h := \mathscr{H}(\delta)$. However, sometimes the impulse response does not exist (at least not as a normal function), for example in the case of the differentiator $y = \dot{u}$. This is an LTI system if we restrict ourselves to inputs that are differentiable, but the impulse response $h = \dot{\delta}$ is the derivative of the delta function, and that is not a regular function. However, the impulse response does exist for almost every physical LTI system, and conventional wisdom is that LTI and convolution are equivalent.

Example 5.1.7 (Moving-average system). Assume $P > 0$. In Example 5.1.2, we saw that the moving-average system

$$y(t) = \frac{1}{P}\int_{t-P}^{t} u(\tau)\,\mathrm{d}\tau \tag{5.1}$$

is LTI. We can therefore write it as a convolution. The impulse response $h = \mathscr{H}(\delta)$ is

$$\begin{aligned} h(t) &= \frac{1}{P}\int_{t-P}^{t} \delta(\tau)\,\mathrm{d}\tau \\ &= \frac{1}{P}\left[\mathbb{1}(\tau)\right]_{t-P}^{t} \\ &= \frac{1}{P}(\mathbb{1}(t) - \mathbb{1}(t-P)) \end{aligned}$$



and then the convolution $y = h * u$ becomes

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(t-\tau)u(\tau)\,\mathrm{d}\tau \\ &= \int_{t-P}^{t} \frac{1}{P}u(\tau)\,\mathrm{d}\tau. \end{aligned}$$

Indeed this equals (5.1). □

Example 5.1.8 (Integrator and the step function). The integrator is the system that sends u to the integral

$$y(t) = \int_{-\infty}^{t} u(\tau)\,\mathrm{d}\tau. \tag{5.2}$$

The integral is well defined for every initially-at-rest (and bounded) input u. This is an LTI system. An easy way to show this is by first determining the impulse response $h = \mathscr{H}(\delta)$:

$$h(t) = \int_{-\infty}^{t} \delta(\tau)\,\mathrm{d}\tau = [\mathbb{1}(\tau)]_{-\infty}^{t} = \mathbb{1}(t)$$

and then verifying that the system is the convolution $h * u$. This is the case, because

$$(h * u)(t) = \int_{-\infty}^{\infty} \mathbb{1}(t-\tau)u(\tau)\,\mathrm{d}\tau = \int_{-\infty}^{t} u(\tau)\,\mathrm{d}\tau = y(t).$$

Integrating is therefore convoluting with the step function. In particular, the integrator is LTI. □

Example 5.1.9 (Echo). Also the echo system

$$y(t) = u(t) + \tfrac{1}{2}u(t-1) + \tfrac{1}{4}u(t-2) + \frac{1}{8}u(t-3) + \cdots$$

is LTI. See Exercise 5.8. □

5.1.1 Initially-at-Rest State Representations

State representations

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{5.3}$$

can also be seen as LTI systems. Namely, there is often a moment at which we start experimenting with our system, and our system is usually *initially at rest*. By definition, this means that all signals involved are zero up to some t_0, in particular the input and the state,

$$u(t) = 0, \quad x(t) = 0 \qquad \forall\, t < t_0.$$

In that case, the general solution (2.14) of the state equation reduces to

$$y(t) = \int_{t_0}^{t} C\mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau + Du(t).$$

Since $u(t)$ is zero for $t < t_0$, we can also write this without making explicit use of t_0,

$$y(t) = \int_{-\infty}^{t} C\mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau + Du(t).$$

We recognize this expression as the convolution of u and the function h defined by

$$h(t) = C\mathrm{e}^{At}B\mathbb{1}(t) + D\delta(t), \tag{5.4}$$

with $\delta(t)$ the Dirac delta function, and $\mathbb{1}(t)$ the step function. Indeed, using the sifting property of delta functions, we have for h defined in (5.4) that

$$\begin{aligned}(h * u)(t) &= \int_{-\infty}^{\infty} h(t-\tau)u(\tau)\,\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} [C\mathrm{e}^{A(t-\tau)}B\mathbb{1}(t-\tau) + D\delta(t-\tau)]u(\tau)\,\mathrm{d}\tau \\ &= \int_{-\infty}^{t} C\mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau + Du(t) \\ &= y(t).\end{aligned}$$

Since the impulse response (5.4) is zero for negative time, we see that such initially-at-rest systems are causal (see Exercise 5.43), and because it is a convolution, it is also LTI. The integrator of Example 5.1.8 is such a system, described by

$$\begin{aligned}\dot{x} &= u, \\ y &= x.\end{aligned}$$

Another example is the following RC circuit.

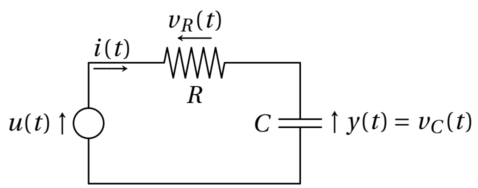


FIGURE 5.5: An RC circuit (Example 5.1.10).

Example 5.1.10 (*RC* circuit). Figure 5.5 shows an RC circuit, where u is the voltage supplied by the voltage source and y is the voltage across the capacitor. A state representation for this system is

$$\begin{aligned}\dot{q} &= -\alpha q + \tfrac{1}{R}u, \\ y &= \tfrac{1}{C}q,\end{aligned}$$

where q is the charge on the capacitor and $\alpha = 1/(RC)$. The solution of this equation is

$$y(t) = \frac{1}{C}\mathrm{e}^{-\alpha(t-t_0)}q(t_0) + \int_{t_0}^{t} \alpha\mathrm{e}^{-\alpha(t-\tau)}u(\tau)\,\mathrm{d}\tau.$$

Initially at rest now means that the initial charge $q(t_0)$, at some time t_0, is zero. If we also have $u(t) = 0$ for all $t < t_0$, then y is a convolution,

$$y(t) = \int_{-\infty}^{t} \mathrm{e}^{-\alpha(t-\tau)}\alpha u(\tau)\,\mathrm{d}\tau \quad \forall t \in \mathbb{R}.$$

The impulse response is

$$h(t) = \alpha e^{-\alpha t} \mathbb{1}(t).$$

□

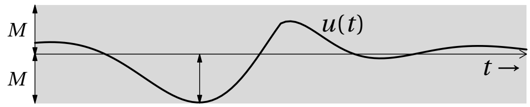


FIGURE 5.6: The narrowest interval $[-M, M]$ that contains $u(t)$ for all t.

5.2 BIBO Stability

For systems as mappings, $y = \mathscr{H}(u)$, stability is defined quite differently than for systems in state representation. Loosely speaking, a system $y = \mathscr{H}(u)$ is considered "stable" if the output is not big if the input is not big. There are many variants. Here, we discuss the popular "BIBO stability".

The ∞-norm $\|u\|_\infty$ of a signal u is defined as the supremum of $|u(t)|$ over all time,

$$\|u\|_\infty := \sup_{t\in\mathbb{R}} |u(t)|.$$

This norm is also called the *peak value* of the signal. It is the least $M \geq 0$ for which $u(t) \in [-M, M]$ for all t; see Fig. 5.6. Note that $\|u\|_\infty$ can be infinite, for example if $u(t) = t$. For a given input u, the *peak-to-peak gain* of $y = \mathscr{H}(u)$ is defined as

$$\frac{\|\mathscr{H}(u)\|_\infty}{\|u\|_\infty},$$

and we define $\|\mathscr{H}\|_1$ to be the *maximal* peak-to-peak gain of the system, that is, the largest possible peak-to-peak gain over all possible bounded inputs,

$$\|\mathscr{H}\|_1 := \sup_{\|u\|_\infty < \infty} \frac{\|\mathscr{H}(u)\|_\infty}{\|u\|_\infty}.$$

If for example $\|\mathscr{H}\|_1 = 3$, then the peak value of the output is at most 3 times as great as the peak value of the input, regardless of the chosen input. This $\|\mathscr{H}\|_1$ therefore is a possible measure of the "size" of the system. This is a system property.

Definition 5.2.1 (BIBO stability). A linear system $y = \mathscr{H}(u)$ is *(uniformly) BIBO stable* if $\mathscr{H}(u)$ exists for all bounded inputs (meaning $\|u\|_\infty < \infty$), and such that the maximal peak-to-peak gain is finite,

$$\|\mathscr{H}\|_1 < \infty.$$

□

BIBO is an acronym for “bounded input, bounded output”. In a BIBO-stable system, the output by definition is bounded if the input is. Using the impulse response, the maximal peak-to-peak gain can be determined explicitly.

Theorem 5.2.2 (BIBO stability). *The maximal peak-to-peak gain $\|\mathscr{H}\|_1$ of an LTI system with impulse response $h : \mathbb{R} \to \mathbb{R}$ is*

$$\|\mathscr{H}\|_1 = \int_{-\infty}^{\infty} |h(t)| \,\mathrm{d}t. \tag{5.5}$$

In particular, the system is BIBO stable if and only if the impulse response is absolutely integrable (that is, $\int_{-\infty}^{\infty} |h(t)| \,\mathrm{d}t < \infty$).

Proof. Let $y = \mathscr{H}(u)$. By definition of ∞-norm we have $|u(t)| \le \|u\|_\infty$ for all t. It follows that

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

Because this holds for all t, the peak value of y is at most $\|u\|_\infty \int_{-\infty}^{\infty} |h(\tau)| \,\mathrm{d}\tau$ and the maximal peak-to-peak gain is therefore at most $\int_{-\infty}^{\infty} |h(\tau)| \,\mathrm{d}\tau$. We claim that this upper bound is reached for

$$u(t) = \operatorname{sgn}(h(-t)).$$

Indeed, this u is bounded, $\|u\|_\infty = 1$, and it gives

$$\begin{aligned}
\frac{\|y\|_\infty}{\|u\|_\infty} &= \|y\|_\infty \\
&\ge y(0) \\
&= (h * u)(0) \\
&= \int_{-\infty}^{\infty} h(0-\tau) u(\tau) \,\mathrm{d}\tau \\
&= \int_{-\infty}^{\infty} |h(-\tau)| \,\mathrm{d}\tau \\
&= \int_{-\infty}^{\infty} |h(t)| \,\mathrm{d}t.
\end{aligned}$$

The integral $\int_{-\infty}^{\infty} |h(t)| \,\mathrm{d}t$ is therefore also a *lower bound* for the maximal peak-to-peak gain. ∎

This theorem also holds if h contains delta functions, in which case we must view $\int_{-\infty}^{\infty} |\delta(t)| \,\mathrm{d}t$ as being 1.

Example 5.2.3 (Delay). The maximal peak-to-peak gain of

$$y(t) = 3u(t-1)$$

is of course 3. This also follows from Theorem 5.2.2: the impulse response is $h(t) = 3\delta(t-1)$ and therefore $\|\mathscr{H}\|_1 = \int_{-\infty}^{\infty} |3\delta(t-1)|\,\mathrm{d}t = 3$. □

Example 5.2.4 (Single echo). Let

$$y(t) = u(t) + \tfrac{1}{2}u(t-1).$$

Then we have $|y(t)| = |u(t) + \frac{1}{2}u(t-1)| \leq \|u\|_\infty + \frac{1}{2}\|u\|_\infty = 1.5\|u\|_\infty$. The maximal peak-to-peak gain is at most 1.5, so the system is BIBO stable.

We can also apply Theorem 5.2.2: the impulse response is $h(t) = \delta(t) + \frac{1}{2}\delta(t-1)$, and so $\|\mathscr{H}\|_1 = \int_{-\infty}^{\infty} \delta(t) + \frac{1}{2}\delta(t-1)\,\mathrm{d}t = 1 + |\frac{1}{2}| = 1.5$. We see that the maximal peak-to-peak gain is equal to 1.5. □

Example 5.2.5 (Integrator). The integrator

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

is not BIBO stable because the impulse response is $h = \mathbb{1}$, and this is not absolutely integrable,

$$\int_{-\infty}^{\infty} |\mathbb{1}(t)|\,\mathrm{d}t = \int_{0}^{\infty} 1\,\mathrm{d}t = \infty.$$

That the integrator is not BIBO stable is also immediately clear, because the integral of the bounded $u(t) = \mathbb{1}(t)$ is the unbounded $y(t) = t\mathbb{1}(t)$. □

Every asymptotically stable system[1]

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

is (as an initially-at-rest system) also BIBO stable because in an asymptotically stable system, the elements of the matrix exponential e^{At} converge exponentially fast to zero when $t \to \infty$, and therefore $|h(t)| := |C\mathrm{e}^{At}B\mathbb{1}(t) + D\delta(t)|$ also converges exponentially fast to zero when $t \to \infty$, implying that $h(t)$ is absolutely integrable. Thus we have the following corollary. This is a simple result but useful:

Corollary 5.2.6 (Asymptotic stability implies BIBO). *Every asymptotically stable system $\dot{x} = Ax + Bu, y = Cx + Du$ is BIBO stable.* □

Example 5.2.7 (*RC* circuit). Consider once again the *RC* circuit of Example 5.1.10. Its state representation given in 5.1.10 is asymptotically stable since

[1]This means that all eigenvalues of A have negative real part.

$\alpha := 1/(RC) > 0$. Hence the system is BIBO-stable. Using $h(t) = \alpha e^{-\alpha t}\mathbb{1}(t)$, the maximal peak-to-peak gain is equal to

$$\begin{aligned}
\|\mathcal{H}\|_1 &= \int_0^\infty |\alpha e^{-\alpha t}| \,dt \\
&= \int_0^\infty \alpha e^{-\alpha t} \,dt \\
&= 1.
\end{aligned}$$

The voltage $y(t)$ over the capacitor can therefore never exceed (in amplitude) the peak voltage $\|u\|_\infty$ of the input. □

Example 5.2.8 (Maximal peak-to-peak gain when $h(t) \geq 0$). Let a_1, a_2 be positive numbers, and let $\beta \in [0, 1]$. The system

$$\begin{aligned}
\dot{x} &= \begin{bmatrix} -a_1 & 0 \\ a_1 & -a_2 \end{bmatrix} x + \begin{bmatrix} \beta \\ 1-\beta \end{bmatrix} u, \\
y &= \begin{bmatrix} 0 & a_2 \end{bmatrix} x
\end{aligned}$$

has eigenvalues $-a_1, -a_2$ and is therefore asymptotically stable and hence also BIBO stable. It can be shown that $h(t) \geq 0$ for all t (Appendix A.9). Assuming that this is so, we can compute the maximal peak-to-peak gain without determining matrix exponentials:

$$\begin{aligned}
\|\mathcal{H}\|_1 &= \int_{-\infty}^\infty |h(t)| \,dt \\
&= \int_{-\infty}^\infty h(t) \,dt \qquad \text{because } h(t) \geq 0 \\
&= \int_0^\infty h(t) \,dt \qquad \text{because } h(t) = 0 \,\forall t < 0 \\
&= \int_0^\infty C e^{At} B \,dt \\
&= [C A^{-1} e^{At} B]_0^\infty \\
&= -C A^{-1} B \\
&= \begin{bmatrix} 0 & a_2 \end{bmatrix} \frac{1}{a_1 a_2} \begin{bmatrix} a_2 & 0 \\ a_1 & a_1 \end{bmatrix} \begin{bmatrix} \beta \\ 1-\beta \end{bmatrix} = 1.
\end{aligned}$$

□

5.3 Step Response and DC Gain

In a BIBO-stable system, by definition the response to a constant signal exists (because constant signals are bounded). A special property of BIBO-stable LTI systems is that the response to a constant input is also constant. This follows easily from the convolution. For example, take $u = 1$ (the constant signal 1). Then the response is

$$\begin{aligned}\mathscr{H}(1)(t) &= (h * 1)(t) \\ &= \int_{-\infty}^{\infty} h(t-\tau)1\,\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} h(\tau)\,\mathrm{d}\tau\end{aligned}$$

and this final improper integral converges for BIBO-stable systems (see also Exercise 5.2) and the outcome is constant. It immediately follows from the linearity that the response to any constant input u_* is equal $(\int_{-\infty}^{\infty} h(\tau)\,\mathrm{d}\tau)u_*$. For this reason the number

$$\int_{-\infty}^{\infty} h(\tau)\,\mathrm{d}\tau$$

is called the *DC gain*[2].

For practical applications, we often consider the *step response*. The step response $\mathscr{H}(\mathbb{1})$ is the response to the unit step function $\mathbb{1}$. The step response models "switching on" the system at time zero. In a BIBO-stable LTI system, the step response exists because $\mathbb{1}$ is bounded. The step response is

$$\begin{aligned}\mathscr{H}(\mathbb{1})(t) &= (h * \mathbb{1})(t) \\ &= \int_{-\infty}^{\infty} h(\tau)\mathbb{1}(t-\tau)\,\mathrm{d}\tau \\ &= \int_{-\infty}^{t} h(\tau)\,\mathrm{d}\tau. \qquad (5.6)\end{aligned}$$

The step response is thus an antiderivative of the impulse response. In BIBO-stable systems, the step response converges to the DC gain, because

$$\begin{aligned}\lim_{t\to\infty} \mathscr{H}(\mathbb{1})(t) &= \lim_{t\to\infty} \int_{-\infty}^{t} h(\tau)\,\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} h(\tau)\,\mathrm{d}\tau.\end{aligned}$$

Example 5.3.1. Let $h(t) = \mathrm{e}^{-t/2}\mathbb{1}(t)$. The response to $u = 1$ is the constant $\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t = 2$ (verify this yourself). We denote the step response by $g(t)$. For

[2] DC is an initialism for *direct current*, a constant current (and AC means *alternating current*).

positive t, the step response is

$$\begin{aligned}
g(t) &= \mathscr{H}(\mathbb{1})(t) \\
&= \int_{-\infty}^{t} h(\tau)\,\mathrm{d}\tau \\
&= \int_{0}^{t} \mathrm{e}^{-\tau/2}\,\mathrm{d}\tau \\
&= \left[-2\mathrm{e}^{-\tau/2}\right]_0^t \\
&= 2(1-\mathrm{e}^{-t/2}).
\end{aligned}$$

This converges to 2 when $t \to \infty$. The graph of $g(t)$ in Figure 5.7 also includes that $g(t) = 0$ for $t < 0$. This type of step response is very common. The system must, as it were, get used to the new constant value of the input and grows exponentially quickly to a new stationary value. □

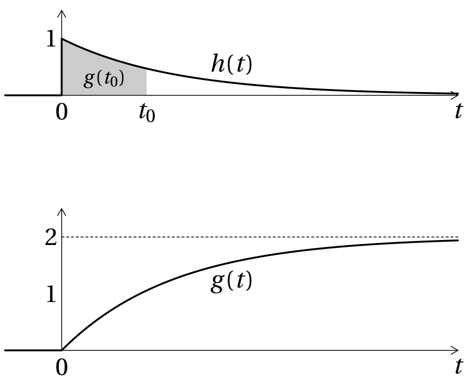


FIGURE 5.7: Graph of impulse response h and step response g of Example 5.3.1. The step response $g(t)$ is the integral of h from $-\infty$ to t.

Example 5.3.2 (Second-order system—step response). Consider a second-order system

$$p_2\ddot{y}(t) + p_1\dot{y}(t) + p_0 y(t) = q_0 u(t).$$

To analyze this system, it is advisable to first scale and rewrite the parameters. It follows from the Routh–Hurwitz test that the system is asymptotically stable iff the three parameters p_2, p_1, p_0 have the same sign. Such systems can always be written in the form

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = d\omega_0^2 u(t),$$

with $\zeta > 0, \omega_0 > 0$ and $d \in \mathbb{R}$ (see Exercise 5.40.) The advantage of this form is that ω_0 then has dimension "frequency" (1 divided by time) and that ζ is dimensionless. The parameter ζ is called the *relative damping*, and ω_0 is called the *natural* *resonance frequency*. Note that $\omega_0 t$ is dimensionless. In this form, we can depict the step response for all such second-order systems in a single figure; see Figure 5.8. (We do not derive the formulas in this course.) For small values of the relative damping,

$$0 < \zeta < 1,$$

the step response oscillates. For

$$\zeta \geq 1,$$

the step response is monotonic. The step response for the case $\zeta = 1$ is indicated in the figure by the dotted red line. It convergences to d "fast". The blue dashed line is the step response for

$$\zeta = \frac{1}{2}\sqrt{2}.$$

This step response also converges quite fast to d and only oscillates a tiny bit. (This case plays a role later.) □

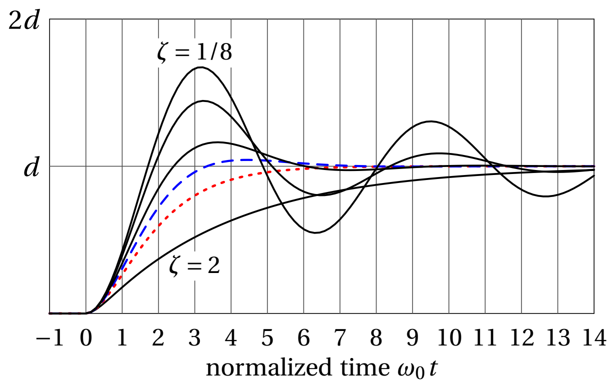


FIGURE 5.8: Graph of the step response of the system $\ddot{y} + 2\zeta\omega_0\dot{y} + \omega_0^2 y = d\omega_0^2 u$ for $\zeta = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2$ (the solid lines; $\zeta = 1$ is red dotted), and also for $\zeta = \frac{1}{2}\sqrt{2}$ (blue dashed line). All step responses converge to d (the DC gain). For $0 < \zeta < 1$ the step response oscillates. For $\zeta \geq 1$ the step response is monotonic.

5.4 Frequency Response

In the previous section, we saw that the response to a constant signal is constant. This holds more generally. A remarkable property of BIBO-stable LTI systems is that the response $y(t)$ to *any* harmonic input $u(t) = \mathrm{e}^{\mathrm{i}\omega t}$ exists and is again harmonic, and with the same frequency. An elegant derivation of this striking result is as follows. Denote the response to $u(t) = \mathrm{e}^{\mathrm{i}\omega t}$ by $y_\omega(t)$, so[3]

$$y_\omega(t) = \mathscr{H}(\mathrm{e}^{\mathrm{i}\omega t}).$$

Now, time-invariance gives

$$\mathscr{H}(\mathrm{e}^{\mathrm{i}\omega(t-t_0)}) = y_\omega(t - t_0),$$

but because of linearity, this is the same as

$$\begin{aligned}\mathscr{H}(\mathrm{e}^{\mathrm{i}\omega(t-t_0)}) &= \mathscr{H}(\mathrm{e}^{-\mathrm{i}\omega t_0}\mathrm{e}^{\mathrm{i}\omega t}) \\ &= \mathrm{e}^{-\mathrm{i}\omega t_0}\mathscr{H}(\mathrm{e}^{\mathrm{i}\omega t}) \\ &= \mathrm{e}^{-\mathrm{i}\omega t_0} y_\omega(t).\end{aligned}$$

As these two are the same, we have

$$\mathrm{e}^{-\mathrm{i}\omega t_0} y_\omega(t) = y_\omega(t - t_0).$$

This holds for all t_0 and all t, and in particular for $t_0 = t$. For $t_0 = t$, it says that

$$\mathrm{e}^{-\mathrm{i}\omega t} y_\omega(t) = y_\omega(0).$$

By multiplying left and right with $\mathrm{e}^{\mathrm{i}\omega t}$, we find the desired result (that the response is harmonic),

$$y_\omega(t) = H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t} \qquad \text{with } H(\mathrm{i}\omega) = y_\omega(0).$$

This result also says that *every* harmonic function $\mathrm{e}^{\mathrm{i}\omega t}$ is an eigenfunction of *every* BIBO-stable LTI system, and that $H(\mathrm{i}\omega)$ is the corresponding eigenvalue. Summary:

Theorem 5.4.1 (Frequency response). *Suppose that $y = \mathscr{H}(u)$ is BIBO and LTI, and that $\omega \in \mathbb{R}$. Then the response to the harmonic input*

$$u(t) = \mathrm{e}^{\mathrm{i}\omega t}$$

exists, and it is also harmonic, with the same frequency,

$$y(t) = H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}.$$

The scaling factor $H(\mathrm{i}\omega)$ — as a function of ω — is called the frequency response *of the system.* □

For $\omega = 0$ we recover the result of the previous section that constant inputs map to constant outputs, and the DC gain apparently is equal to $H(0)$.

We can often easily determine the frequency response $H(\mathrm{i}\omega)$ by simply substituting $u(t) = \mathrm{e}^{\mathrm{i}\omega t}$, $y(t) = H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}$ in the system equations.

[3] The notation $y_\omega = \mathscr{H}(\mathrm{e}^{\mathrm{i}\omega\cdot})$ is more correct mathematically, but is also not as nice to work with.

Example 5.4.2 (Delay system). Let $t_0 > 0$. The delay system

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

is LTI and BIBO (verify this yourself) and therefore the response to $u(t) = \mathrm{e}^{\mathrm{i}\omega t}$ is of the form $y(t) = H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}$. Using this, we can immediately determine the frequency response $H(\mathrm{i}\omega)$: let $u(t) = \mathrm{e}^{\mathrm{i}\omega t}$, then

$$y(t) = u(t - t_0) = \mathrm{e}^{\mathrm{i}\omega(t-t_0)} = \mathrm{e}^{-\mathrm{i}\omega t_0}\mathrm{e}^{\mathrm{i}\omega t}$$

and therefore we have

$$H(\mathrm{i}\omega) = \mathrm{e}^{-\mathrm{i}\omega t_0}.$$

The DC gain is $H(0) = \mathrm{e}^0 = 1$. This is not surprising because the response to a constant $u(t) = \bar{u}$ is of course the same constant, $y(t) = u(t - t_0) = \bar{u}$. □

Example 5.4.3 (Moving-average system). The moving average system (5.1) is BIBO stable (verify this yourself) and hence has a frequency response $H(\mathrm{i}\omega)$. Again we can determine $H(\mathrm{i}\omega)$ by using $u(t) = \mathrm{e}^{\mathrm{i}\omega t}$. Then

$$\begin{aligned} y(t) &= \frac{1}{P}\int_{t-P}^{t} \mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\tau \\ &= \frac{1}{P}\left[\frac{\mathrm{e}^{\mathrm{i}\omega\tau}}{\mathrm{i}\omega}\right]_{t-P}^{t} = \frac{\mathrm{e}^{\mathrm{i}\omega t} - \mathrm{e}^{\mathrm{i}\omega(t-P)}}{\mathrm{i}\omega P} \end{aligned}$$

This equals $H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}$ for

$$H(\mathrm{i}\omega) = \frac{1 - \mathrm{e}^{\mathrm{i}\omega(-P)}}{\mathrm{i}\omega P}.$$

□

We can also deduce the result of Theorem 5.4.1 using convolutions,

$$\begin{aligned} (h * \mathrm{e}^{\mathrm{i}\omega\cdot})(t) &= \int_{-\infty}^{\infty} h(\tau)\mathrm{e}^{\mathrm{i}\omega(t-\tau)}\,\mathrm{d}\tau \\ &= \underbrace{\left(\int_{-\infty}^{\infty} h(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau\right)}_{H(\mathrm{i}\omega):=}\mathrm{e}^{\mathrm{i}\omega t} \\ &= H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}. \end{aligned}$$

We recognize the frequency response $H(\mathrm{i}\omega)$ as the Fourier transform of the impulse response $h(t)$.

The nice thing about stable LTI systems is that they are completely determined by the frequency response. Namely, an LTI system is completely determined by the impulse response $h(t)$, and for BIBO-stable systems we have $\int_{-\infty}^{\infty}|h(t)|\,\mathrm{d}t < \infty$ and then Fourier theory guarantees that there is a bijection between $h(t)$ and $H(\mathrm{i}\omega)$.

For the zero frequency, $\omega = 0$, the above once again shows that the response to the constant signal $\mathrm{e}^{0t} = 1$ is equal to $H(0)$. This number is what we earlier called the *DC gain*. Indeed,

$$H(0) = \int_{-\infty}^{\infty} h(t)\mathrm{e}^{0t}\,\mathrm{d}t = \int_{-\infty}^{\infty} h(t)\,\mathrm{d}t$$

is the DC gain.

Because initially-at-rest systems $\dot{x} = Ax + Bu, y = Cx + Du$ are convolutions, also they have a frequency response, provided they are BIBO-stable. Again the frequency response follows easily from the system equations.

Lemma 5.4.4 (Frequency response of initially-at-rest state models). *If A is asymptotically stable, then the frequency response of the initially-at-rest system* (5.3) *exists, and is equal to*

$$H(\mathrm{i}\omega) = C(\mathrm{i}\omega I - A)^{-1}B + D.$$

Proof. From the asymptotic stability it follows that the system is BIBO, so $H(\mathrm{i}\omega)$ exists. We only prove it for $D = 0$. By (5.4), we have $h(t) = C\mathrm{e}^{At}B\mathbb{1}(t)$. Hence

$$\begin{aligned}
H(\mathrm{i}\omega) &= \int_{-\infty}^{\infty} h(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} C\mathrm{e}^{At}B\mathbb{1}(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t \\
&= \int_{0}^{\infty} C\mathrm{e}^{(A-\mathrm{i}\omega I)t}B\,\mathrm{d}t \\
&= \left[C(A - \mathrm{i}\omega I)^{-1}\mathrm{e}^{(A-\mathrm{i}\omega I)t}B\right]_0^{\infty} \\
&= C(\mathrm{i}\omega I - A)^{-1}B.
\end{aligned}$$

In the last step, we used that $\mathrm{e}^{(A-\mathrm{i}\omega I)\infty} = 0$. This is because A is asymptotically stable. Note that asymptotic stability implies that $\mathrm{i}\omega I - A$ is invertible for all $\omega \in \mathbb{R}$. ■

We already know that in BIBO-stable systems, the step response $\mathscr{H}(\mathbb{1})$ converges to $H(0)$. We can now generalize this result to arbitrary harmonic inputs:

Lemma 5.4.5 (Steady-state response). *Suppose the system is LTI and BIBO stable. Let $\omega \in \mathbb{R}$. Then the response $y(t)$ to the initially-at-rest harmonic input*

$$u(t) = \mathrm{e}^{\mathrm{i}\omega t}\mathbb{1}(t)$$

exists and is equal to

$$y(t) = \left(\int_{-\infty}^{t} h(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau\right)\mathrm{e}^{\mathrm{i}\omega t}.$$

Moreover, $y(t)$ converges to $H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}$ when $t \to \infty$. For this reason, $H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}$ is called the steady-state *response to* $\mathrm{e}^{\mathrm{i}\omega t}\mathbb{1}(t)$.

Proof. Left to the reader. It is the generalization of (5.6) for $u(t) = \mathbb{1}(t)$ to $u(t) = e^{i\omega t}\mathbb{1}(t)$. ∎

Example 5.4.6 (Complex harmonic signals). Let

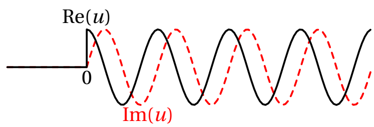
$$\begin{aligned} \dot{x} &= -0.1x + u, \\ y &= x. \end{aligned}$$

This is asymptotically stable and so has a frequency response. According to the above the frequency response equals

$$H(i\omega) = \frac{CB}{i\omega - A} = \frac{1}{i\omega + 0.1}.$$

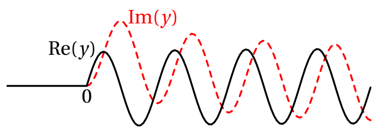
Lemma 5.4.5 now states that the response to the initially-at-rest complex harmonic signal

$$u(t) = e^{it}\mathbb{1}(t)$$



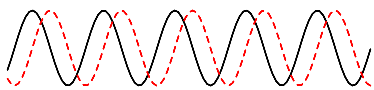
exists,

$$y(t)$$



and that it converges to the complex harmonic signal $H(i)e^{it}$, hence to

$$\frac{1}{i + 0.1}e^{it}.$$



The graphs confirm this convergence. □

We conclude this section by noting that asymptotically stable differential equations

$$y^{(n)} + p_{n-1}y^{(n-1)} + \cdots + p_0 y = q_n u^{(n)} + \cdots + q_0 u \tag{5.7}$$

also have a frequency response. Indeed, such equations are equivalent to the state equations

$$\begin{aligned} \dot{z} &= \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix} z + \begin{bmatrix} q_0 - p_0 q_n \\ q_1 - p_1 q_n \\ \vdots \\ \\ q_{n-1} - p_{n-1} q_n \end{bmatrix} u, \\ y &= \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} z + q_n u, \end{aligned}$$

and the eigenvalues of the A-matrix are the zeros of the characteristic polynomial $s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0$ of the differential equation, so the state model is then asymptotically stable as well, and, hence, the frequency response $H(\mathrm{i}\omega)$ therefore exists. To deduce the frequency response, we can avoid the state model: simply substitute $u(t) = \mathrm{e}^{\mathrm{i}\omega t}$ and $y(t) = H(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}$ in the system equations. Since the differentiation of harmonic functions $\mathrm{e}^{\mathrm{i}\omega t}$ is nothing more than multiplication by $\mathrm{i}\omega$, the differential equation (5.7) becomes the algebraic equation

$$\begin{aligned} &H(\mathrm{i}\omega)((\mathrm{i}\omega)^n + p_{n-1}(\mathrm{i}\omega)^{n-1} + \cdots + p_0) \\ &\qquad = q_n(\mathrm{i}\omega)^n + q_{n-1}(\mathrm{i}\omega)^{n-1} + \cdots + q_0, \end{aligned}$$

and, therefore,

$$H(\mathrm{i}\omega) = \frac{q_n(\mathrm{i}\omega)^n + q_{n-1}(\mathrm{i}\omega)^{n-1} + \cdots + q_0}{(\mathrm{i}\omega)^n + p_{n-1}(\mathrm{i}\omega)^{n-1} + \cdots + p_0}. \tag{5.8}$$

Done! We apply this in the next section.

5.5 Frequency Response – Real Form

For completeness we mention here the real (as in "non complex") form of the main result of the previous section. We assume that our systems are *real* meaning that they map real-valued inputs into real-valued outputs.

Lemma 5.5.1 (Frequency response). *In a real BIBO-stable LTI system $y = \mathscr{H}(u)$, the frequency response $H(\mathrm{i}\omega)$ exists for all $\omega \in \mathbb{R}$, and the response $y(t)$ to every sinusoid*

$$u(t) = \cos(\omega t)$$

exists and equals

$$y(t) = |H(\mathrm{i}\omega)|\cos(\omega t + \arg H(\mathrm{i}\omega)). \tag{5.9}$$

It has the same frequency as $u(t)$. Furthermore, the response to $u(t) = \cos(\omega t)\mathbb{1}(t)$ converges to (5.9) *as $t \to \infty$. In this setting,* (5.9) *is called the* steady-state response.

Proof. By linearity we have $\mathscr{H}(u_1 + \mathrm{i}u_2) = \mathscr{H}(u_1) + \mathrm{i}\mathscr{H}(u_2)$, and since it is a real system we also have that the real part of the response

$$\mathrm{Re}\,(\mathscr{H}(u_1 + \mathrm{i}u_2)) = \mathrm{Re}\,(\mathscr{H}(u_1) + \mathrm{i}\mathscr{H}(u_2)) = \mathscr{H}(u_1)$$

is equal to the response to the real part

$$\mathscr{H}(\mathrm{Re}\,(u_1 + \mathrm{i}u_2)) = \mathscr{H}(u_1).$$

In other words, $\operatorname{Re}\mathscr{H} = \mathscr{H}\operatorname{Re}$. Using this relation and Euler's formula — $e^{i\alpha} = \cos(\alpha) + i\sin(\alpha)$ — we see that

$$\begin{aligned}\mathscr{H}(\cos(\omega t)) &= \mathscr{H}(\operatorname{Re} e^{i\omega t}) \\ &= \operatorname{Re}\mathscr{H}(e^{i\omega t}) \\ &= \operatorname{Re}(H(i\omega)e^{i\omega t}) \\ &= \operatorname{Re}(|H(i\omega)|e^{i\Phi(\omega)}e^{i\omega t}) \\ &= |H(i\omega)|\cos(\omega t + \Phi(\omega)),\end{aligned}$$

where $\Phi(\omega) := \arg(H(i\omega))$.

Likewise, the response to the initially-at-rest $u_{\text{iar}}(t) := \cos(\omega t)\mathbb{1}(t)$ equals $y_{\text{iar}}(t) := \operatorname{Re}\mathscr{H}(e^{i\omega t}\mathbb{1}(t))$. Since the complex response $\mathscr{H}(e^{i\omega t}\mathbb{1}(t))$ converges to $\mathscr{H}(e^{i\omega t})$, its real part $y_{\text{iar}}(t)$ converges to $\operatorname{Re}\mathscr{H}(e^{i\omega t})$. ■

More specifically, we see that $|H(i\omega)|$ is the amplification factor of the amplitude of the sinusoid, and that $\arg(H(i\omega))$ is the additional phase the response has in comparison to the input. This is an important feature of $H(i\omega)$; see the next series of examples.

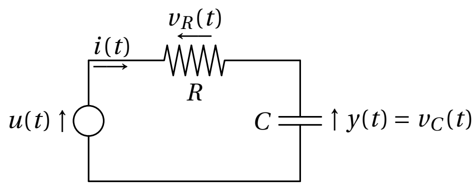


FIGURE 5.9: An RC circuit (Example 5.5.2).

Example 5.5.2 (*RC* circuit). By (5.8), the frequency response of the RC circuit (Figure 5.9) described by differential equation

$$\dot{y} + \alpha y = \alpha u, \qquad \alpha = \tfrac{1}{RC} > 0 \tag{5.10}$$

is equal to

$$H(i\omega) = \frac{\alpha}{i\omega + \alpha} = \frac{1}{i(\omega/\alpha) + 1}.$$

The amplitude of the frequency response goes to zero when $\omega \to \infty$. This means that rapidly fluctuating inputs u barely generate any response y. The DC gain is $H(0) = 1$. Therefore the zero-frequency signals (the constant signals $u(t) = u_0$) are transmitted unchanged. The response to

$$u(t) = \cos(\alpha t)$$

is

$$
\begin{aligned}
y(t) &= |H(\mathrm{i}\alpha)|\cos(\alpha t + \arg(H(\mathrm{i}\alpha))) \qquad (5.11)\\
&= |\tfrac{1}{\mathrm{i}+1}|\cos(\alpha t - \arg(\tfrac{1}{\mathrm{i}+1})).\\
&= \frac{1}{2}\sqrt{2}\cos(\alpha t - \pi/4).
\end{aligned}
$$

(See Figure 5.10.) The output y lags behind the input u. Delay (or time-lag effect) occurs in many stable systems. The system needs time, so to speak, to process changes in the input. □

Example 5.5.3 (Second-order system—resonance frequency). We continue with the system of Example 5.3.2,

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = d\omega_0^2 u(t).$$

This system is asymptotically stable for all $\zeta > 0$ and $\omega_0 > 0$. Hence for all such ζ, ω_0, the frequency response exists. It is

$$
\begin{aligned}
H(\mathrm{i}\omega) &= \frac{d\omega_0^2}{(\mathrm{i}\omega)^2 + 2\zeta\omega_0\mathrm{i}\omega + \omega_0^2}\\
&= \frac{d}{(\mathrm{i}\omega/\omega_0)^2 + 2\zeta(\mathrm{i}\omega/\omega_0) + 1}.
\end{aligned}
$$

Figure 5.11 shows the absolute value of the frequency response for a series of relative dampings $\zeta > 0$. Notice the large peak at $\omega = \omega_0$ if ζ is small. This is consistent with the oscillations in the step response as shown Figure 5.8. □

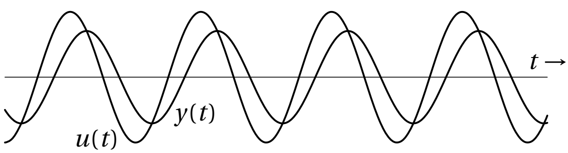


FIGURE 5.10: $u(t) = \cos(\alpha t)$ and $y(t) = |H(\mathrm{i}\alpha)|\cos(\alpha t + \arg(H(\mathrm{i}\alpha)))$. (Example 5.5.2.)

Example 5.5.4 (Hypnotist). A hypnotist will sometimes swing a watch on a chain to hypnotize the public; see Figure 5.12. We denote the horizontal position of hand by u and the angle the pendulum makes with the vertical axis by ϕ. We state without proof that the linearized equation of motion is

$$m\ell\ddot{\phi} + k\ell\dot{\phi} + mg\phi = -m\ddot{u} - k\dot{u}.$$

The coefficients that occur in this equation are ℓ (length of the chain on which the watch hangs), g (the gravitational acceleration), and k (a friction coefficient),

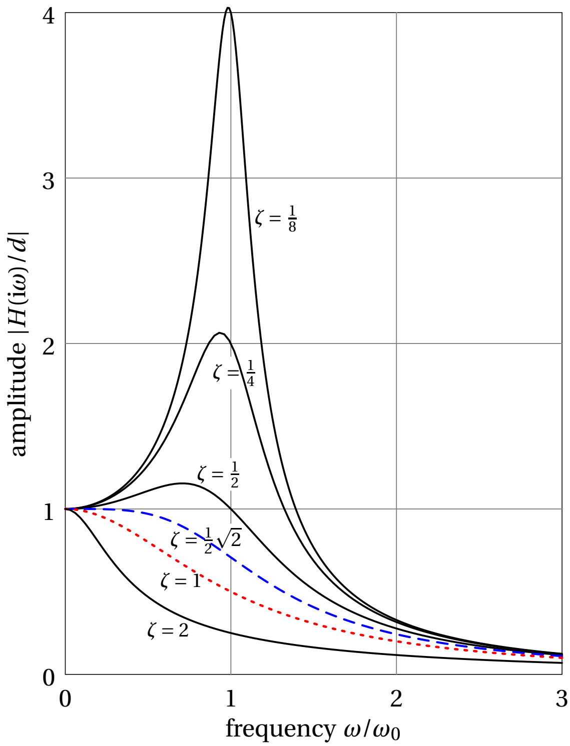


FIGURE 5.11: See Example 5.5.3. Amplitude of $H(i\omega) = 1/((i\omega)^2 + 2\zeta\omega_0(i\omega) + \omega_0^2)$ for $\zeta = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2$ (the solid lines; $\zeta = 1$ is red dotted) and also for $\zeta = \frac{1}{2}\sqrt{2}$ (the blue dashed line). The absolute value $|H(i\omega)|$ for $\zeta \le \frac{1}{2}\sqrt{2}$ is monotonically decreasing as a function of positive frequency. For $0 < \zeta < \frac{1}{2}\sqrt{2}$, the frequency response $H(i\omega)$ has a positive resonance frequency around $\omega/\omega_0 = 1$. The peak increases as the relative damping ζ decreases.

all positive. Because they are all positive, this system is asymptotically stable (see Exercise 2.12). The frequency response is

$$H(i\omega) = \frac{-m(i\omega)^2 - k(i\omega)}{m\ell(i\omega)^2 + k\ell(i\omega) + mg}.$$

Figure 5.13 gives an example of a frequency response. The amplitude $|H(i\omega)|$ takes its maximum in the neighborhood of $\omega \approx 5$. Now, if the hypnotist is smart, he will make a sinusoidal motion with his hand, with frequency $\omega \approx 5$. For this frequency, it will be easy to keep the watch in motion with barely visible movement of the hand. The amplitude of the pendulum will then be about 25 times that of the hand. The frequency ω for which $|H(i\omega)|$ is maximal is called the *resonance frequency*. For the resonance frequency, the phase of $H(i\omega)$ here is about -1.7. This means that the watch is not in phase with the motion of the hand but lags behind

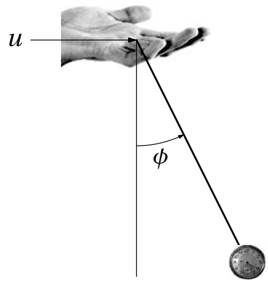


FIGURE 5.12: The hypnotist (Example 5.5.4).

the latter,

$$u(t) = \epsilon \cos(5t) \quad \Longrightarrow \quad \phi(t) \approx 25\epsilon \cos(5t - 1.7).$$

Let the hypnosis begin. □

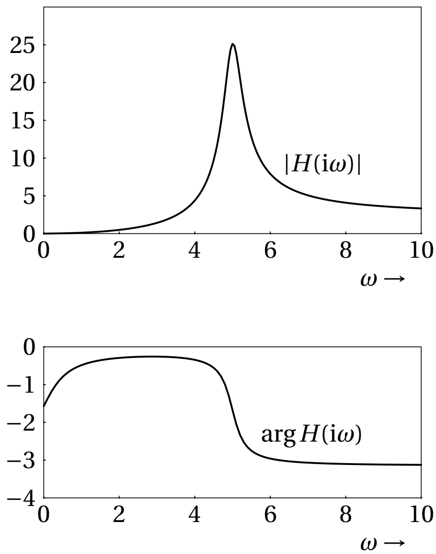


FIGURE 5.13: Amplitude and phase of the frequency response $H(\mathrm{i}\omega)$ of the hypnotist (for $m = 0.1, k = 0.05, \ell = 0.4, g = 10$). (Example 5.5.4.)

5.6 Transfer Function

The Laplace transform can be seen as an extension of the Fourier transform by simply replacing $\mathrm{i}\omega$ with s. Since you are all familiar with the Laplace transform we can be brief in our discussion here. By replacing the $i\omega \in i\mathbb{R}$ in

$$H(i\omega) = \int_{-\infty}^{\infty} h(t)e^{-i\omega t}\, dt$$

with $s \in \mathbb{C}$ we obtain the (two-sided) *Laplace transform* of $h(t)$,

$$H(s) = \int_{-\infty}^{\infty} h(t)e^{-st}\, dt.$$

If $h(t)$ is the impulse response of a system then $H(s)$ is called the *transfer function* of that system. Many results of the previous two sections carry over to this case, sometimes with slight modifications:

1. If $\dot{x} = Ax + Bu, y = Cx + Du$ and if it is initially at rest then

   $$H(s) = C(sI - A)^{-1}B + D$$

   and it is well defined if $\operatorname{Re}(s) > \max_i \operatorname{Re}(\lambda_i(A))$.

2. If the system is described by a differential equation of the form $P(\frac{d}{dt})y(t) = Q(\frac{d}{dt})u(t)$ then, when initially at rest, the transfer function equals

   $$H(s) = Q(s)/P(s)$$

   and it is well defined if $\operatorname{Re}(s) > \max_i \operatorname{Re}(s_i)$ where s_i are the zeros of $P(s)$.

3. If $y = \mathscr{H}(u)$ is LTI and the Laplace transforms of $u(t)$ and $h(t)$ exist for $\operatorname{Re}(s) > \gamma$ then also the Laplace transform $Y(s)$ of the output exists for $\operatorname{Re}(s) > \gamma$ and we have

   $$Y(s) = H(s)U(s).$$

LTI systems that do not have a frequency response (e.g. unstable systems) typically *do* have a transer function, be it only defined on a subset of the complex numbers s. Interesting is the fact that eigenvalues of the A-matrix return as poles of the transfer function (well, some might cancel). This means that typically we can read the stability properties of the system from its transfer function. In asymptotically stable systems the eigenvalues of the A-matrix have negative real part and consequently all poles of $H(s) = C(sI - A)^{-1}B + D$ have negative real part as well. Because of this we define:

Definition 5.6.1 (Asymptotically stable transfer function). A rational transfer function $H(s)$ is *asymptotically stable* if $\operatorname{Re}(s_i) < 0$ for all poles s_i of $H(s)$. □

The zeros and poles of $H(s)$ have a clear interpretation (see Exercise 5.27).

Example 5.6.2 (Poles and zeros). The transfer function $H(s)$ of

$$\begin{aligned} \dot{x} &= -3x + u \\ y &= -5x + u \end{aligned}$$

is

$$\begin{aligned}H(s) &= -5(s+3)^{-1}1+1\\ &= \frac{-5}{s+3}+\frac{s+3}{s+3}\\ &= \frac{s-2}{s+3}.\end{aligned}$$

This has one pole, $s=-3$, and is therefore asymptotically stable. It has one zero, $s=+2$. The pole of $H(s)$ corresponds to the eigenvalue -3 of the A-matrix: $A=-3$. This is often the case. It can, however, also happen that an eigenvalue is divided out. A simple example of this is the unstable system

$$\begin{aligned}\dot{x} &= +2x+0u\\ y &= x+u.\end{aligned}$$

This has transfer function

$$\begin{aligned}H(s) &= 1(s+2)^{-1}0+1\\ &= 1.\end{aligned}$$

This has no poles. As a transfer function, it is asymptotically stable, even though it is derived from an unstable state representation. □

We have seen that not all eigenvalues of A need to return as poles of the transfer function $H(s)$. However, if the system is controllable and observable, then they must return, and in that case stability of $H(s)$ is the same as stability of the underlying state representation:

Theorem 5.6.3 (Minimal realization). *If (A,B) is controllable and (A,C) is observable, then the poles of $H(s)=C(sI-A)^{-1}B+D$ are exactly the eigenvalues of A.*

Proof. We prove it for the case that $n_u=n_y=1$. Suppose (A,B) is controllable and (A,C) is observable. Without loss of generality, we assume that $D=0$, because the poles of $H(s):=C(sI-A)^{-1}B+D$ do not depend on D. Since the system is observable, by Lemma 3.5.4 it is isomorphic to a system of the form

$$\begin{aligned}\dot{z} &= \begin{bmatrix}0 & \cdots & \cdots & 0 & -p_0\\ 1 & \ddots & & \vdots & -p_1\\ 0 & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots\\ 0 & \cdots & 0 & 1 & -p_{n-1}\end{bmatrix} z+\begin{bmatrix}q_0\\ q_1\\ \vdots\\ \vdots\\ q_{n-1}\end{bmatrix} u,\\ y &= \begin{bmatrix}0 & \cdots & \cdots & 0 & 1\end{bmatrix} z.\end{aligned}$$

Let $P(s)=s^n+p_{n-1}s^{n-1}+\cdots+p_1s+p_0$ and $Q(s)=q_{n-1}s^{n-1}+\cdots+q_1s+q_0$. Incidentally, $P(s)=\det(sI-A)$ because A is a companion matrix (Lemma 2.5.6). From the controllability it now follows that $P(s)$ and $Q(s)$ do not have a common zero (see the derivation after Lemma 3.5.4). Moreover, Lemma 2.5.2 says that $P(\frac{\mathrm{d}}{\mathrm{d}t})y(t)=Q(\frac{\mathrm{d}}{\mathrm{d}t})u(t)$ and therefore that $H(s)=Q(s)/P(s)=Q(s)/\det(sI-A)$. All zeros of $P(s)$ are poles of $H(s)$ because $P(s)$ and $Q(s)$ have no common zeros. ■

Example 5.6.4. Let α be some real number and consider the system depending on α,

$$\dot{x} = \begin{bmatrix} -2 & \alpha \\ 0 & 3 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$
$$y = \begin{bmatrix} 1 & 1 \end{bmatrix} x.$$

The controllability and observability matrices are

$$\mathscr{C} = \begin{bmatrix} 0 & \alpha \\ 1 & 3 \end{bmatrix}, \qquad \mathscr{O} = \begin{bmatrix} 1 & 1 \\ -2 & \alpha+3 \end{bmatrix}.$$

It is not controllable if $\alpha = 0$, and it is not observable if $\alpha = -5$. The eigenvalues of the A-matrix are -2 and 3 and normally these are the poles of $H(s)$. We have

$$\begin{aligned} H(s) &= C(sI - A)^{-1}B \\ &= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} s+2 & -\alpha \\ 0 & s-3 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \end{bmatrix} \frac{1}{(s+2)(s-3)} \begin{bmatrix} s-3 & \alpha \\ 0 & s+2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{s+2+\alpha}{(s+2)(s-3)}. \end{aligned}$$

There is a cancellation of common factors only if $\alpha = 0$, in which case we have

$$H(s) = \frac{1}{s-3},$$

or if $\alpha = -5$, in which case we have

$$H(s) = \frac{1}{s+2}.$$

These cancellations occur precisely if the system is either uncontrollable or unobservable. □

Example 5.6.5 (Integrator and differentiator). The transfer function of the integrator $\dot{y} = u$ is $H(s) = \frac{1}{s}$. This has a pole at $s = 0$ and is therefore defined for all $s \in \mathbb{C}$ with $\operatorname{Re}(s) > 0$. This system is the integrator.

The transfer function of the differentiator $y = \dot{u}$ is $H(s) = s/1 = s$. This does not have any poles and is therefore defined for all $s \in \mathbb{C}$. □

Example 5.6.6 (Second-order system—DC gain & poles). Consider one more time the system from Examples 5.3.2 and 5.5.3,

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = d\omega_0^2 u(t).$$

Its transfer function is

$$H(s) = \frac{d\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}.$$

If we divide the numerator and denominator by ω_0^2, this becomes

$$H(s) = \frac{d}{(s/\omega_0)^2 + 2\zeta(s/\omega_0) + 1}.$$

The DC gain $H(0)$ is equal to d. We already saw this in Figure 5.8. Note that s/ω_0 is dimensionless and that $H(s)$ is nothing else than a scaled version of the dimensionless $\frac{1}{s^2+2\zeta s+1}$. The two poles of $1/(s^2+2\zeta s+1)$ are depicted in Figure 5.14 for several $\zeta > 0$.

Notice that the poles for $\zeta = 1/8$ are close to the imaginary axis, at approximately $\pm \mathrm{i}$, so it is not surprising that for $\zeta = 1/8$, the absolute value $|H(\mathrm{i}\omega)|$ has a high peak near $\omega/\omega_0 = 1$ (that is, $\omega = \omega_0$), see Fig. 5.11. For this reason, ω_0 is called the *natural resonance frequency*. □

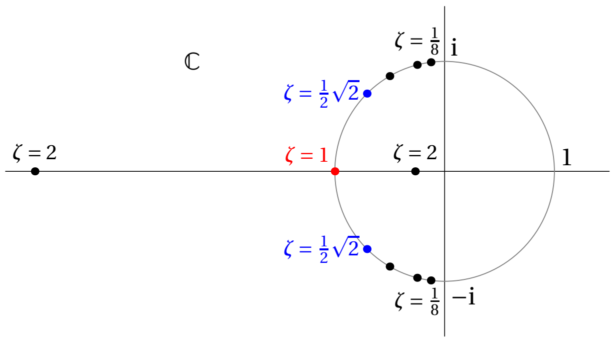
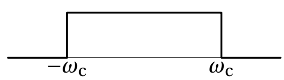


FIGURE 5.14: See Example 5.6.6. Poles $s_{1,2}$ of $1/(s^2+2\zeta s+1)$ in the complex plane for $\zeta = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, \frac{1}{2}\sqrt{2}, 1$ and 2. For $0 < \zeta < 1$ the two poles $s_{1,2}$ are complex and lie close to the imaginary axis, and for $\zeta = \frac{1}{2}\sqrt{2}$ the poles lie at an angle of 45 degrees from the imaginary axis (blue). For $\zeta = 1$ both poles are at $s = -1$ (red) and for $\zeta > 1$ they are distinct and real.

5.6.1 Design of Butterworth Filters

A standard problem in signal processing is to remove undesired components from a signal. In this setting, the system $y = \mathscr{H}(u)$ that removes the undesired components is usually called a *filter*. Filters are classified by the frequency bands they remove. The *ideal low-pass filter* is a system $y = \mathscr{H}(u)$ with frequency response

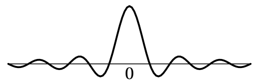
$$H(\mathrm{i}\omega) = \begin{cases} 1 & |\omega| < \omega_c \\ 0 & \text{elsewhere.} \end{cases}$$



This completely removes all harmonic signals with frequency greater than ω_c, and lets all harmonic signals with frequency less than ω_c through unchanged. The ω_c

is called the *cut-off frequency*. However, as we know from Fourier Analysis, the inverse Fourier transform of this frequency response is the function

$$h(t) = \frac{\sin(\omega_c t)}{\pi t},$$



and since $h(t)$ is not zero for $t < 0$, the system $y = h * u$ is not causal. The practical compromise is to use rational frequency responses that *are* causal (namely, described by differential equations) and that approximate ideal low-pass filters. The fact that asymptotic stability is easy to characterize in the frequency domain—the poles must have negative real part—makes it possible to design such practical filters in the frequency domain. Here, we design *Butterworth filters.* We first design Butterworth filters of order 2.

Example 5.6.7 (Second-order Butterworth filter). We are looking for an asymptotically stable transfer function of order 2

$$H_2(s) = \frac{1}{s^2 + bs + c} \tag{5.12}$$

whose frequency response satisfies

$$|H_2(\mathrm{i}\omega)|^2 = \frac{1}{1+\omega^4}.$$

Given the form (5.12) we find

$$\begin{aligned}
|H_2(\mathrm{i}\omega)|^2 &= \frac{1}{|(\mathrm{i}\omega)^2 + b\mathrm{i}\omega + c|^2} \\
&= \frac{1}{|(c-\omega^2) + \mathrm{i}(b\omega)|^2} \\
&= \frac{1}{c^2 - 2c\omega^2 + \omega^4 + b^2\omega^2} \\
&= \frac{1}{c^2 + (b^2 - 2c)\omega^2 + \omega^4}.
\end{aligned}$$

It equals $1/(1+\omega^4)$ if $c^2 = 1$ and $b^2 = 2c$, that is, $c = 1$ and $b = \pm\sqrt{2}$. Since we are after an asymptotically stable system we need $b > 0$ (see Exercise 2.12). So the only option left is $c = 1, b = \sqrt{2}$:

$$H_2(s) = \frac{1}{s^2 + \sqrt{2}s + 1}.$$

This the Butterworth filter of order 2. □

Butterworth filters $\mathscr{H}_n$ of order n are stable systems whose frequency responses satisfy

$$|H_n(\mathrm{i}\omega)|^2 = \frac{1}{1+\omega^{2n}}. \tag{5.13}$$

These approximate the ideal low-pass filter, because for n large, we have

$$\begin{aligned} 0 \le \omega < 1 &\implies \omega^{2n} \approx 0 &&\implies \quad H_n(\omega) \approx 1 \\ 1 < \omega &\implies \omega^{2n} \approx \infty &&\implies \quad H_n(\omega) \approx 0. \end{aligned}$$

The amplitude $|H_n(\mathrm{i}\omega)|$ therefore shows some similarity with that of the ideal low-pass filter with cut-off frequency $\omega_\mathrm{c} = 1$. Figure 5.15 shows $|H_n(\mathrm{i}\omega)|$ for $n = 1$ through $n = 4$.

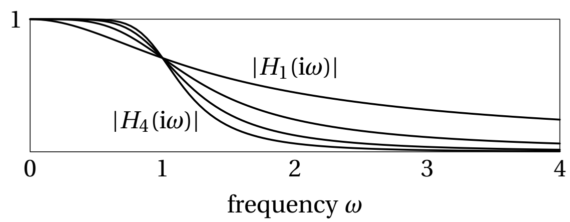


FIGURE 5.15: Amplitude of $H_n(\mathrm{i}\omega)$ for $n = 1, 2, 3, 4$.

To find an asymptotically stable system that satisfies (5.13), we must rewrite (5.13) in terms of transfer functions. For this, write the left-hand side of (5.13) in terms of $s = \mathrm{i}\omega$ as

$$\begin{aligned} |H_n(\mathrm{i}\omega)|^2 &= H_n(\mathrm{i}\omega)\overline{H_n(\mathrm{i}\omega)} \\ &= H_n(\mathrm{i}\omega)H_n(-\mathrm{i}\omega) \\ &= H_n(s)H_n(-s) \end{aligned}$$

and write the right-hand side of (5.13) in terms of $s = \mathrm{i}\omega$,

$$\frac{1}{1+\omega^{2n}} = \frac{1}{1+(s/\mathrm{i})^{2n}} = \frac{1}{1+(-s^2)^n}.$$

In terms of transfer functions, the equality (5.13) is therefore

$$H_n(s)H_n(-s) = \frac{1}{1+(-s^2)^n}. \tag{5.14}$$

We use this to determine an asymptotically stable $H_n(s)$, that is, all its poles have negative real part. To achieve this, we factor the right-hand side of (5.14)

$$\frac{1}{1+(-s^2)^n} = (-1)^n \prod_{k=1}^{2n} \frac{1}{s-s_k},$$

in which s_k are the zeros of $1+(-s^2)^n$, that is, the solutions of

$$(-s_k^2)^n = -1.$$

These s_k satisfy

$$(-s_k^2) = \sqrt[n]{-1} = \sqrt[n]{\mathrm{e}^{\mathrm{i}(2k-1)\pi}} = \mathrm{e}^{\frac{\mathrm{i}(2k-1)\pi}{n}},$$

for $k = 1, 2, \dots, n$, so

$$s_k = \mathrm{i}\mathrm{e}^{\frac{\mathrm{i}(k-1/2)\pi}{n}}, \quad k = 1, 2, \dots, 2n.$$

These poles s_k are distributed uniformly over the unit circle, with the first half $s_1, \dots, s_n$ in the left half-plane (that is, to the left of the imaginary axis) and the second half $s_{n+1}, \dots, s_{2n}$ in the right half-plane; see Figure 5.17. Because of the stability, the poles of $H_n(s)$ must have negative real part, so we take the first half of the poles to define $H_n(s)$,

$$H_n(s) = \prod_{k=1}^{n} \frac{1}{s - s_k}.$$

Consequently, $H_n(-s)$ is a rational function with poles to the right of the imaginary axis, and we have

$$H_n(s) H_n(-s) = (-1)^n \prod_{k=1}^{2n} \frac{1}{s - s_k},$$

as desired. For $n = 1, 2, 3, 4$ we thus obtain

$$\begin{aligned}
H_1(s) &= \frac{1}{s+1}, \\
H_2(s) &= \frac{1}{s^2 + \sqrt{2}s + 1}, \\
H_3(s) &= \frac{1}{(s+1)(s^2 + s + 1)}, \\
H_4(s) &= \frac{1}{(s^2 + \sqrt{2+\sqrt{2}}s + 1)\,(s^2 + \sqrt{2-\sqrt{2}}s + 1)}.
\end{aligned}$$

See Figure 5.15. The filter (system) with transfer function $H_n(s)$ is called the nth order *Butterworth filter*, and because $H_n(s)$ is rational, this system can be modeled by a differential equation. The second-order Butterworth filter can, for example, is described by

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

Because of this, the behavior of Butterworth filters can be simulated (for example in MATLAB), and analog models of these systems can be built as electrical circuits using capacitors, resistors, and inductors. We illustrate the above second-order Butterworth filter in MATLAB:

```
s=tf('s');
H2=1/(s^2+sqrt(2)*s+1);        % H_2(s) = 1/(s^2 + sqrt(2)s + 1)
Ts=0.01;
t=0:Ts:20;
u=sin(t/2)+randn(size(t))/5; % u(t) = sin(t/2) + "noise"
y=lsim(H2,u,t);               % y = H_2(u)
plot(t,u,t,y);
```

Here the input $u(t)$ is a sinusoid with angular frequency $1/2$ corrupted with random noise. The output $y(t)$ of the Butterworth filter retains most of the sinusoidal part of the input (but with slightly smaller amplitude, and a bit delayed, see § 5.5), and most of the noise is gone, see Fig. 5.16.

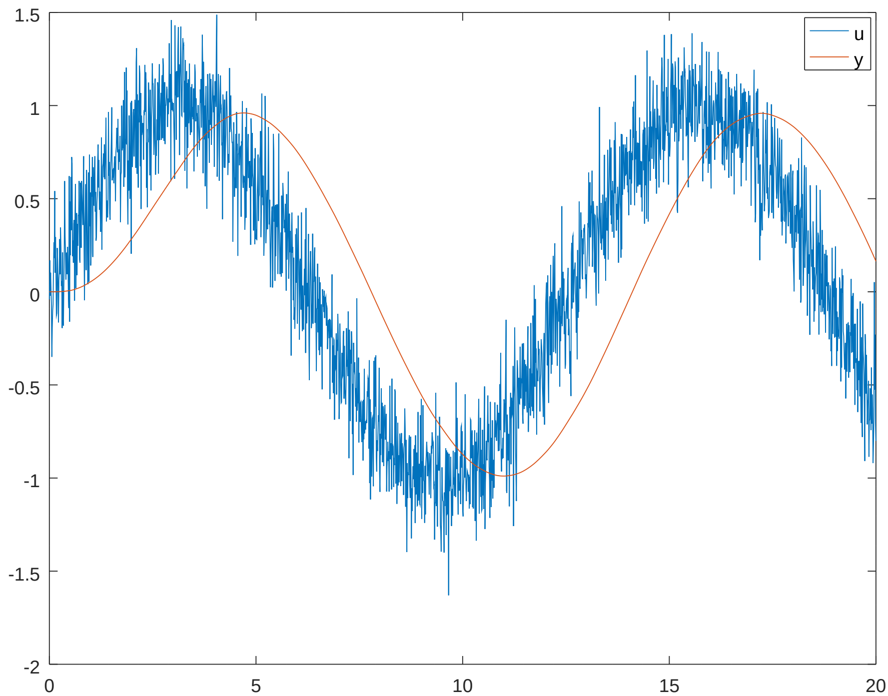


FIGURE 5.16: A noisy $u(t)$ and the response $y(t)$ of the 2nd order Butterworth filter to this input $u(t)$, see § 5.6.1

5.7 Interconnections

Transfer functions and frequency responses are well suited to describe systems formed by interconnections of a several subsystems. This is important for modeling complex systems where it can be useful to split up the complex total system

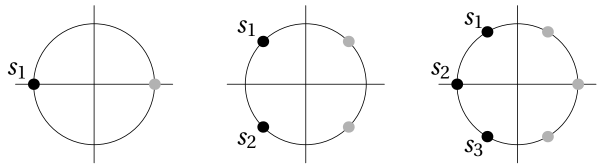


FIGURE 5.17: The stable poles $s_1,\ldots,s_n \in \mathbb{C}$ for the Butterworth filters of order $n = 1,2,3$.

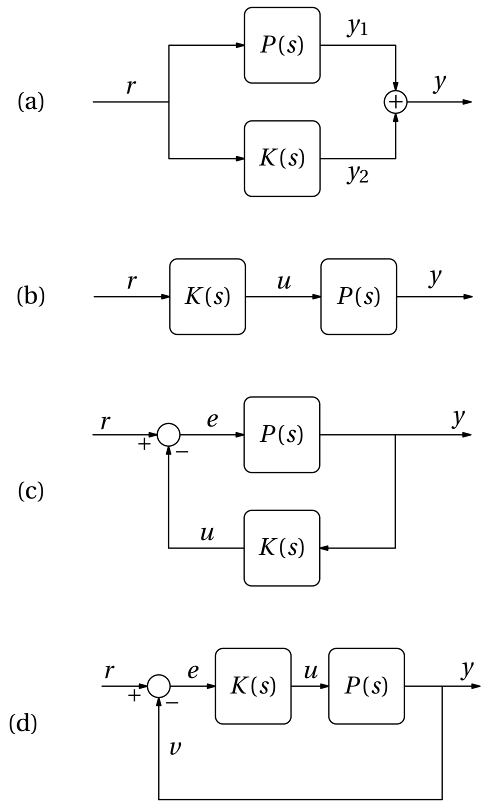


FIGURE 5.18: Four types of interconnections.

and later connect the models of the subsystems. It is also important for control engineering where the aim is to design one or more subsystems to impose a certain behavior on the total system—we will elaborate on this in the next section. We first present four much-used interconnections of two subsystems with transfer functions $P(s)$ and $K(s)$.

Parallel interconnection. Figure 5.18(a) shows the parallel circuit. In the parallel circuit, the output $Y(s)$ is the sum of $P(s)R(s)$ and $K(s)R(s)$. The transfer function $H(s)$ from r to y is therefore

$$H_{y/r}(s) = P(s) + K(s).$$

Series interconnection. The series interconnection is shown in Figure 5.18(b). The output of $K(s)$ is taken as input for $P(s)$. This way, $Y(s) = P(s)U(s) = P(s)K(s)R(s)$. The transfer function $H(s)$ from r to y is now given by

$$H_{y/r}(s) = P(s)K(s).$$

Feedback 1. There are several variants of feedback, but in essence they are of the form indicated in Figure 5.18(c,d). In (c) we see that $E(s) = R(s) - U(s) = R(s) - K(s)P(s)E(s)$. That is, $(1 + K(s)P(s))E(s) = R(s)$. So

$$E(s) = \frac{1}{1 + K(s)P(s)} R(s).$$

The transfer function from r to y is therefore

$$H_{y/r}(s) = \frac{P(s)}{1 + K(s)P(s)}.$$

Note that it uses a *negative* feedback: e is r *minus* u. This is a convention, but is otherwise not significant.

Feedback 2. Another feedback method is depicted in part (d) of Figure 5.18. Verify for yourself that the transfer function from r to y is now given by

$$H_{y/r}(s) = \frac{P(s)K(s)}{1 + P(s)K(s)}.$$

An important application of negative feedback is Black's negative feedback amplifier, which made it possible, for the first time, to realize reliable high-quality amplifiers (see Exercise 5.21 for a stylized version).

Example 5.7.1 (Juggler). In Example 4.4.2, we designed a dynamical controller for the "juggler". The juggler was described by

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ -2 \end{bmatrix} u, \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x. \end{aligned}$$

This has transfer function

$$\begin{aligned}P(s) &= \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}s & -1\\ -2 & s\end{bmatrix}^{-1}\begin{bmatrix}0\\ -2\end{bmatrix}\\ &= \frac{-2}{s^2-2}.\end{aligned}$$

The controller was

$$\begin{aligned}\dot{\hat{x}} &= \begin{bmatrix}-4 & 1\\ -7 & -2\end{bmatrix}\hat{x} + \begin{bmatrix}4\\ 6\end{bmatrix}y,\\ u &= \begin{bmatrix}3/2 & 1\end{bmatrix}\hat{x}\end{aligned}$$

and this has transfer function

$$\tilde{K}(s) = \begin{bmatrix}3/2 & 1\end{bmatrix}\begin{bmatrix}s+4 & -1\\ 7 & s+2\end{bmatrix}^{-1}\begin{bmatrix}4\\ 6\end{bmatrix} = \frac{12s+17}{s^2+6s+15}.$$

Now, it is common in control engineering to feed the output y back with a *minus* sign; see Figure 5.19. This means that the input of the controller is not y, but $e := -y$. The controller $K(s)$ from $e = -y$ to u then gains a minus sign,

$$K(s) = -\tilde{K}(s) = \frac{-(12s+17)}{s^2+6s+15}.$$

The block diagram in Figure 5.19 also includes an external signal w. This models a perturbing signal, such as the effect of the wind on the position y. Because of this perturbing signal w, the controlled system will not work perfectly. We can analyze the effect of w on y using the transfer function from w to y. This is (verify this yourself)

$$\begin{aligned}H_{y/w}(s) &= \frac{1}{1+K(s)P(s)} \qquad (5.15)\\ &= \frac{(s^2-2)(s^2+6s+15)}{s^4+6s^3+13s^2+12s+4}.\end{aligned}$$

See Figure 5.20. From this figure we see that a constant w causes a deviation from the attained position y that is $|H_{y/w}(0)| = |-30/4| = 7.5$ times as large. This is unwanted. High frequency disturbances w return in y unaltered because $H_{y/w}(\mathrm{i}\omega) \approx 1$ for large values of ω. We also see this from Fig. 5.20. In the next section we design better controllers (Example 5.8.5). □

5.7.1 Stability of Closed Loops

For systems described by ordinary differential equations, there are various methods for verifying the asymptotic stability of a closed loop. One method uses state-space representations of the subsystems to construct a state-space representation of the closed loop. We used this in the previous chapters. Another method uses transfer functions (not covered in this course), and a third method uses differential equations. This we explore now.

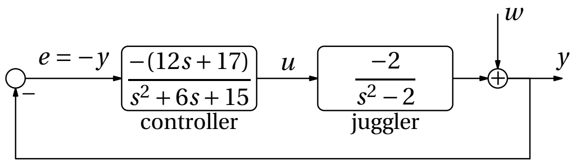


FIGURE 5.19: Block diagram of Example 5.7.1.

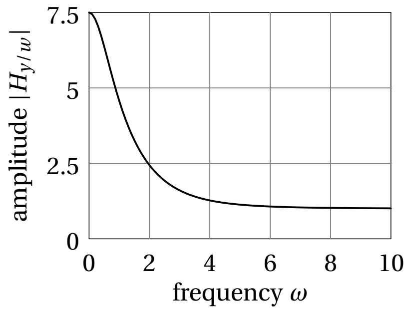


FIGURE 5.20: Amplitude $|H_{y/w}(\mathrm{i}\omega)|$ of the transfer function from w to y of Example 5.7.1.

Example 5.7.2 (Juggler). In Example 4.4.2 about the juggler, a controller is constructed that places the eigenvalues of the closed-loop system in -1 (double; these are the eigenvalues of $A-BF$) and -2 (double; these are the eigenvalues of $A-LC$). We can also describe the juggler and controller also using a DE. From $P(s) = -2/(s^2-2)$, we see that the juggler can be described by the DE

$$\ddot{y} - 2y = -2u,$$

and from $\tilde{K}(s) = \frac{(12s+17)}{s^2+6s+15}$ follows that the differential equation of the controller is

$$\ddot{u} + 6\dot{u} + 15u = 12\dot{y} + 17y.$$

The combination of juggler and controller is therefore described by the system of DEs

$$\begin{bmatrix} \frac{\mathrm{d}^2}{\mathrm{d}t^2} - 2 & 2 \\ -(12\frac{\mathrm{d}}{\mathrm{d}t} + 17) & \frac{\mathrm{d}^2}{\mathrm{d}t^2} + 6\frac{\mathrm{d}}{\mathrm{d}t} + 15 \end{bmatrix} \begin{bmatrix} y(t) \\ u(t) \end{bmatrix} = 0.$$

By Lemma 2.5.8, this system, with outputs u, y, is asymptotically stable if and only if the determinant of the corresponding polynomial matrix

$$\begin{bmatrix} s^2 - 2 & 2 \\ -(12s+17) & s^2 + 6s + 15 \end{bmatrix}$$

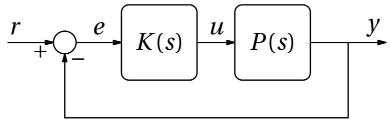
is asymptotically stable. This determinant is

$$\chi_{\text{closed}}(s) = s^4 + 6s^3 + 13s^2 + 12s + 4.$$

This is called the closed-loop characteristic polynomial. It should not come as a surprise that the zeros of this polynomial are -1 and -2 (both double), exactly the eigenvalues of the closed-loop system in the state-space representation. We now also see that the transfer function $H_{y/w}(s)$ from w to y given in (5.15) has this characteristic polynomial $\chi_{\text{closed}}(s)$ as denominator. This $H_{y/w}(s)$ therefore has its poles at -1 and -2 (both double) and is consequently asymptotically stable. □

The following theorem is the central result regarding closed loop stability.

Theorem 5.7.3 (Closed loop asymptotic stability). *Consider the closed-loop system of Figure 5.18(d), copied here:*



Suppose that $P(s)$ is described by the DE

$$D_P(\tfrac{\mathrm{d}}{\mathrm{d}t})y = N_P(\tfrac{\mathrm{d}}{\mathrm{d}t})u,$$

and that the controller $K(s)$ is described by the DE

$$D_K(\tfrac{\mathrm{d}}{\mathrm{d}t})u = N_K(\tfrac{\mathrm{d}}{\mathrm{d}t})e.$$

Then the closed loop is described by the system of DE's

$$\begin{bmatrix} D_P(\frac{\mathrm{d}}{\mathrm{d}t}) & -N_K(\frac{\mathrm{d}}{\mathrm{d}t}) \\ N_K(\frac{\mathrm{d}}{\mathrm{d}t}) & D_K(\frac{\mathrm{d}}{\mathrm{d}t}) \end{bmatrix} \begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ D_K(\frac{\mathrm{d}}{\mathrm{d}t}) \end{bmatrix} r, \quad e = r - y. \tag{5.16}$$

The closed loop with input r and output (e, u, y) is asymptotically stable if and only if the closed-loop characteristic polynomial defined by

$$\chi_{closed}(s) = D_P(s)D_K(s) + N_P(s)N_K(s)$$

is asymptotically stable.

Proof. See Exercise 5.36. ■

Note that in this case the transfer functions are

$$K(s) = \frac{N_K(s)}{D_K(s)}, \qquad P(s) = \frac{N_P(s)}{D_P(s)}.$$

The closed-loop characteristic polynomial is nothing more than the product of the denominators of $K(s)$ and $P(s)$ plus the product of the numerators of $K(s)$ and $P(s)$.

Example 5.7.4. Let

$$K(s) = \frac{3}{s}, \qquad P(s) = \frac{1}{s+2}.$$

The implicit assumption here is that $u = K(s)e$ is described by the DE $\dot{u} = 3e$, and that $y = P(s)u$ is described by the DE $\dot{y} + 2y = u$. The closed-loop characteristic polynomial is now

$$\chi_{\text{closed}}(s) = s(s+2) + 3 \times 1 = s^2 + 2s + 3.$$

This is asymptotically stable (the closed loop poles are $s_{1,2} = -1 \pm \sqrt{2}$). Hence $K(s) = 3/s$ is a so-called *stabilizing* controller. Note that $K(s)$ itself is not asymptotically stable.

The transfer function $H_{y/r}(s)$ from r to y then is

$$H_{y/r}(s) = \frac{P(s)K(s)}{1 + P(s)K(s)} = \frac{3/(s(s+2))}{1 + 3/(s(s+2))} = \frac{3}{\chi_{\text{closed}}(s)}.$$

The closed loop poles, $s_{1,2} = -1 \pm \sqrt{2}$, are poles of $H_{y/r}(s)$. We claim that this transfer function is well defined for all s with $\operatorname{Re}(s) > \operatorname{Re}(s_{1,2}) = -1$. So closed loop stability in particular implies that the frequency response $H_{r/y}(\mathrm{i}\omega)$ is well defined[4].

□

Example 5.7.5. Let

$$K(s) = \frac{s+2}{s-1}, \qquad P(s) = \frac{s-1}{s+1}.$$

Then the closed-loop characteristic polynomial is

$$\begin{aligned}\chi_{\text{closed}}(s) &= (s-1)(s+1) + (s+2)(s-1) \\ &= (s-1)(2s+3).\end{aligned}$$

This is not asymptotically stable because of the factor $s - 1$. □

In the next section, we will see that using Theorem 5.7.3, we can construct a well-functioning controller fairly easily.

---

[4] Notice that the above $K(s)$ is not defined at $s = 0$, but that the closed loop transfer function $H_{y/r}(s)$ is claimed to be defined for all complex s with $\operatorname{Re}(s) > -1$ (so also defined at $s = 0$). How can that be? The short answer is: do not take the Laplace transform of the separate systems (K and P), but take the Laplace transform of the closed loop system (5.16). Laplace transformation of this leads to $\begin{bmatrix} D_P(s) & -N_K(s) \\ N_K(s) & D_K(s) \end{bmatrix}\begin{bmatrix} Y(s) \\ U(s) \end{bmatrix} = \begin{bmatrix} 0 \\ D_K(s) \end{bmatrix} R(s)$ which is now defined for all s with $\operatorname{Re}(s) > \operatorname{Re}(s_{1,2})$ (assuming $R(s)$ is defined when $\operatorname{Re}(s) > \operatorname{Re}(s_{1,2})$). Then $Y(s) = H_{y/r}(s)R(s)$ for the same $H_{y/r}(s)$ but now properly defined for all s with $\operatorname{Re}(s) > -1$.

5.8 Error Feedback

A natural way to keep signals constant or influence them in another way is through *error feedback*. We examine this important principle in this section. We will also show that feedback is eminently suitable for making the behavior of a system *more robust*, that is, less sensitive to external influences and dynamical uncertainties. Feedback with this effect is common in biological systems and, for example, in the human body. A classic example is the regulation of blood sugar, but cruise control systems and the ordinary thermostat are also examples of this principle.

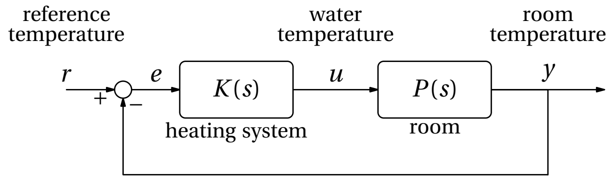


FIGURE 5.21: Temperature control based on the error, $e = r - y$, (Example 5.8.1).

Example 5.8.1 (Error-feedback temperature control). Suppose that you set the desired temperature to r degrees Celsius. A simple temperature control mechanism is as follows: if the room temperature is $(r+1)$°C or more, then the thermostat gives a sign that the heating should be shut off. If the room temperature is $(r-1)$°C or less, then the sign is given that the heating should be turned on. In other words:

> The control law (heating on or off) is determined solely on the basis of the *error* $e = r - y$ between the desired temperature r and the current temperature y.

Figure 5.21 depicts this control mechanism. The "room" $P(s)$ is the given system with input u the water temperature in the heating pipes, and output y the room temperature. The signal r is the reference temperature, that is, the desired temperature that we impose. The signal $e = r - y$ in Figure 5.21 is therefore the error between the desired and current temperature. Only this error is available to the controller $K(s)$, so the controller has only e as input (and u as output). □

Error feedback is the most applied form of feedback. In the majority of the cases (as in this example), the given system $P(s)$ accounts for most of the volume, price, service, etc., and the controller is only a small part (often a chip, or a small box; see Figure 5.23). The given system $P(s)$ in engineering is often called the *plant*, hence the notation $P(s)$. The aim of the controller, in its simplest form, is to

- make the closed-loop system asymptotically stable;

- make the closed-loop system less sensitive to changes in $P(s)$;
- make the system less sensitive to certain disrupting signals; and
- make the output y automatically follow a reference signal r.

5.8.1 PID Controller

The *PID* controllers are extremely popular and successful. These are controllers of the form

$$K_{\mathrm{PID}}(s) = k_{\mathrm{P}} + \frac{k_{\mathrm{I}}}{s} + k_{\mathrm{D}}s.$$

They consist of three parts with each a clear function and interpretation. We treat the three parts separately.

Proportional (P). For $k_{\mathrm{I}} = k_{\mathrm{D}} = 0$, the PID controller $u = K(s)e$ reduces to a constant amplifier

$$u(t) = k_{\mathrm{P}}e(t).$$

In this case the signal $u(t)$ is always proportional to the error. A large value of k_{P} means an aggressive controller: even a small error $e(t)$ can lead to large inputs $u(t)$. The controller is static, that is, the input at time t depends only on the error signal at that same time.

Integral (I). For $k_{\mathrm{P}} = k_{\mathrm{D}} = 0$, the PID controller $u = K(s)e$ reduces to a pure integrator $u = \frac{k_{\mathrm{I}}}{s}e$; that is, in the time domain,

$$u(t) = \int_{-\infty}^{t} k_{\mathrm{I}}e(\tau)\mathrm{d}\tau$$

or, equivalently,

$$\dot{u}(t) = k_{\mathrm{I}}e(t).$$

This form is also easy to understand. It says: the greater the error $e(t)$, the more quickly $u(t)$ must change. Applied to our thermostat problem (and with $k_{\mathrm{i}} > 0$), this says: if the temperature is too low, $y(t) < r(t)$, then the error $e(t) = r(t) - y(t)$ is greater than zero, and the control law $\dot{u}(t) = k_{\mathrm{I}}e(t) > 0$ will turn up the heater. This also works in the other direction: if $e(t) = r(t) - y(t)$ is negative (so $y(t)$ is too great), then the control law $\dot{u}(t) = k_{\mathrm{I}}e(t) < 0$ will turn down the heater. Only if the error is zero is the heater left alone.

Differentiation (D). For $k_P = k_I = 0$, the PID controller $u = K(s)e$ reduces to a pure differentiator

$$u(t) = k_D \dot{e}(t).$$

This form is also easy to understand. It says: the more quickly the error decreases, the smaller the input. Once again applied to our heating system: if $y(t)$ increases in the direction of $r(t)$, then $y(t)$ might overshoot. In that case, it is not a bad idea to already turn the heater down a bit. This is what the differentiator does, because an increasing $y(t)$ (w.r.t. $r(t)$) means a decreasing $e(t) = r(t) - y(t)$, hence negative $\dot{e}(t)$ and therefore negative $u(t) = k_D \dot{e}(t)$ (if $k_D > 0$). The differentiator in a way anticipates; see Figure 5.22.

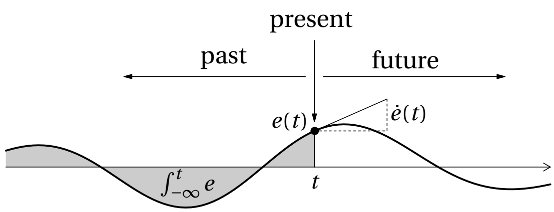


FIGURE 5.22: A possible error signal $e(t)$. The integrator of a PID controller takes into account the “past”, the proportional term only the “present”, and the differentiator somewhat the “future”. (This figure is based on a figure in [1].)

FIGURE 5.23: A PID controller from NIPPON TECHNOLOGIES.

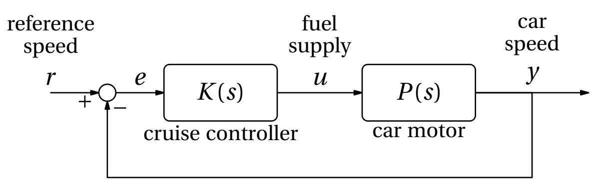


FIGURE 5.24: Cruise control system (see Examples 5.8.2 and 5.8.3).

Figure 5.24 shows the block diagram of a car with cruise control. The cruise control system's job is to fully automatically bring the car to the desired reference speed r and keep it there. The speed y of the car depends on the throttle opening u that determines the fuel supply. This throttle opening is the only quantity the cruise control system can manipulate, and it does so based only on the difference e between the reference and measured speeds. A simple model of the car linearized around a given throttle opening is

$$\textit{motor:} \quad \dot{y} = -ay + bu, \tag{5.17}$$

for some $a, b > 0$ (see Appendix A.10 for a derivation). The transfer function of the motor from u to y is therefore

$$P(s) = \frac{b}{s+a}.$$

In the following two examples, we illustrate controllers with proportional and integrating action.

Example 5.8.2 (P-controller). Let $u = Ke$ be a constant amplifier

$$K(s) = k_{\mathrm{P}}.$$

This stabilizes $P(s) = b/(s+a)$ provided $k_{\mathrm{P}} \geq -a/b$, because $\chi_{\text{closed}}(s) = (s+a) + bk_{\mathrm{P}} = s + (a + bk_{\mathrm{P}})$ and $a, b > 0$. The transfer function from r to y is

$$\begin{aligned} H_{y/r}(s) &= \frac{K(s)P(s)}{1 + K(s)P(s)} \\ &= \frac{k_{\mathrm{P}} \frac{b}{s+a}}{1 + k_{\mathrm{P}} \frac{b}{s+a}} \\ &= \frac{k_{\mathrm{P}} b}{s + a + k_{\mathrm{P}} b}. \end{aligned}$$

The DC-gain is the transfer function at $s = 0$,

$$H_{y/r}(0) = \frac{k_{\mathrm{P}} b}{a + k_{\mathrm{P}} b} = \frac{1}{a/(k_{\mathrm{P}} b) + 1}.$$

This is equal to 1 only if $k_{\mathrm{P}} = \infty$. So only if the amplification k_{P} is insanely large, do we have, practically speaking, that $y(t) \to r$ if $r(t)$ is constant. □

Example 5.8.3 (I-controller). We again take $P(s) = b/(s+a)$ and now propose a controller with integrating action,

$$K(s) = \frac{k_{\mathrm{I}}}{s}.$$

The closed-loop characteristic polynomial is now

$$\begin{aligned} \chi_{\text{closed}}(s) &= s(s+a) + k_{\mathrm{I}} b \\ &= s^2 + as + k_{\mathrm{I}} b. \end{aligned}$$

According to the Routh–Hurwitz test, the closed loop is asymptotically stable iff $k_\mathrm{I} > 0$ (since $a, b > 0$). Figure 5.25 shows the response y to the step $r(t) = \mathbb{1}(t)$ for certain values of a, b and k_I. What is remarkable, is that $y(t)$ converges to $r(t) = 1$ for every choice of $k_\mathrm{I} > 0$. Apparently, this controller ensures that the measured speed converges to the desired speed. This also follows from the transfer function. The transfer function from r to y is

$$\begin{aligned} H_{y/r}(s) &= \frac{K(s)P(s)}{1 + P(s)K(s)} \\ &= \frac{\frac{k_\mathrm{I}}{s}\frac{b}{s+a}}{1 + \frac{k_\mathrm{I}}{s}\frac{b}{s+a}} \\ &= \frac{k_\mathrm{I} b}{s(s+a) + k_\mathrm{I} b}. \end{aligned}$$

It immediately follows that $H_{y/r}(0) = 1$; that is, the DC gain is one, i.e. the stationary values of $r(t)$ and $y(t)$ are equal if $r(t)$ is constant. This does not depend on the values of a, b, k_I (provided that they are positive). So the simple control law $\dot{u} = k_\mathrm{I} e$ will *always* ensure that the speed $y(t)$ goes to the desired constant speed r, regardless of how strong or weak the motor $P(s) = b/(s + a)$ is. The only condition is that the closed loop is asymptotically stable (so $a, b, k_\mathrm{I} > 0$). □

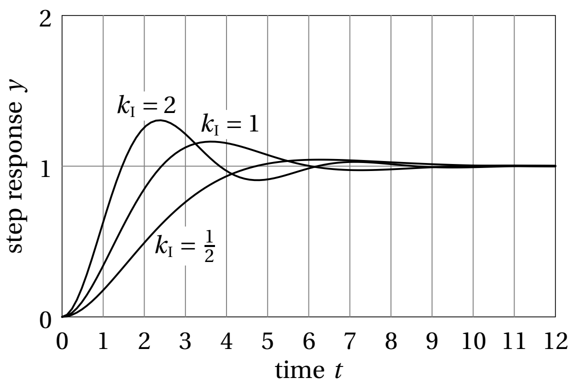


FIGURE 5.25: Step response of $H_{y/r}(s)$ in Example 5.8.3 for $a = b = 1$ and $k_\mathrm{I} = \frac{1}{2}$, 1, 2.

In an asymptotically stable LTI system driven by constant signals, all signals converge to a constant value. If the controller is an integrator,

$$\dot{u}(t) = k_\mathrm{I} e(t), \qquad k_\mathrm{I} \neq 0,$$

then something remarkable happens: because $u(t)$ converges to a constant, it is tempting to conclude that $e(t) = \dot{u}(t)/k_\mathrm{I}$ converges to zero. The conclusion of the previous example that $\lim_{t\to\infty} e(t) = 0$—and therefore $\lim_{t\to\infty} y(t) = r$ for every constant r—is perhaps no coincidence? Indeed, this always happens, for every $P(s)$, provided that the closed loop is asymptotically stable and that the controller or plant has an integrator! This is worth a theorem:

Theorem 5.8.4 (Integrating action—zero steady-state error). *Consider the system of Figure 5.18(d). Let $r(t) = r$ or $r(t) = r\mathbb{1}(t)$. If the closed loop is asymptotically stable and $P(s)K(s)$ has a pole at $s = 0$, then $y(t)$ converges to r as $t \to \infty$.*

Proof. Because of the asymptotic stability of the closed loop, the transfer function from r to $e = r - y$ is also asymptotically stable. This transfer function is $H_{e/r} := 1/(1+PK)$. Since PK has a pole at $s = 0$, the DC gain is $H_{e/r}(0) = 1/\infty = 0$. So $e(t)$ converges to zero if r is constant. Hence $y(t) = r - e(t)$ converges to r. ■

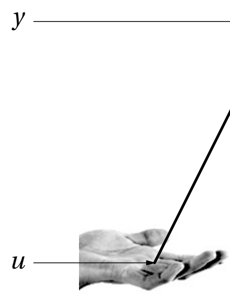


FIGURE 5.26: The walking juggler (Example 5.8.5).

We can now design a controller for the *walking* juggler.

Example 5.8.5 (Walking juggler). Consider, once more, the block diagram of Figure 5.18(d) with the juggler system with transfer function

$$P_{y/u}(s) = \frac{-2}{s^2 - 2}.$$

See Figure 5.26. It turned out that the stabilizing controller of Example 5.7.1 was not very robust against disturbances w, and the controller was also not designed to send $y(t)$ to a value other than zero.

We now propose the second-order controller

$$K(s) = \frac{-4s^2 - 6s - 1/2}{s(s+4)}.$$

This has an integrating action (that is, it has a pole at $s = 0$, which is good), and the closed-loop characteristic polynomial is asymptotically stable, because

$$\begin{aligned} \chi_{\text{closed}}(s) &= s(s+4)(s^2-2) + (-4s^2 - 6s - 1/2)(-2) \\ &= s^4 + 4s^3 + 6s^2 + 4s + 1 \\ &= (s+1)^4. \end{aligned}$$

(K was designed this way; see Exercise 5.38.) All closed-loop poles therefore are at -1. The transfer function $H(s)$ from r to y is now (verify this yourself)

$$H_{y/r}(s) = \frac{8s^2 + 12s + 1}{(s+1)^4}. \tag{5.18}$$

Figure 5.27 shows the amplitude of the frequency response $H_{y/r}(\mathrm{i}\omega)$. Because of the integrator in $K(s)$, the DC gain of $H_{y/r}$ is equal to 1. The output y will therefore follow the input r (provided that r is constant). Figure 5.28 shows the response to $r(t) = \mathbb{1}(t)$. There is quite some overshoot, but in the end the tip of the pendulum $y(t)$ reaches the desired $r(t) = 1$. In this figure, the input $u(t)$ (the position of the hand) has also been plotted. This input is first negative (hand to the left), making the inverted pendulum fall (to the right). Shortly after that, the hand catches up with the pendulum and brings it to a halt, but now at the desired final position $r(t) = 1$. A juggler would do the same! □

In this example we just set a controller that—surprise, surprise—seemed to work well. A more systematic method to design controllers is by first adding an integrator to $P(s)$ and then constructing a stabilizing controller $\tilde{K}(s)$ for the resulting $\frac{1}{s}P(s)$, using, for example, state feedback and observers. As a last step. the integrator is transferred to the controller, $K(s) = \tilde{K}(s)/s$. See Figure 5.29. This controller has an integrator and is also stabilizing. Good.

There is much more to say about system and control theory. It has many applications and it is part of almost all electrical, mechanical and biomedical studies. There are also many interesting connections with signal processing. And where methods are being developed, a deeper knowledge of the mathematics involved is needed, a fine task for the applied mathematician.

5.9 Exercises

5.1 Comprehension questions (on the whole chapter). Prove or give a counterexample.

(a) $y(t) = u(t+1)$ is LTI?

(b) $y(t) = u(t+1)$ is BIBO stable?

(c) If $y = \mathscr{H}(u)$ is a nonlinear system and $\|y\|_\infty < 1$ for all inputs, then the maximal peak-to-peak gain is finite?

(d) If $y = \mathscr{H}(u)$ is LTI and $\|y\|_\infty < 1$ for all inputs, then $y = 0$ for all inputs?

(e) If the step response does not exist, then the system is not BIBO stable?

(f) If for every $\omega_0 \in \mathbb{R}$, the response to $u(t) = \sin(\omega_0 t)$ is a harmonic signal with angular frequency ω_0, then the system is LTI?

(g) If two systems $\dot{x}_1 = A_1x_1 + B_1u, y = C_1x_1$ and $\dot{x}_2 = A_2x_2 + B_2u, y = C_2x_2$ are isomorphic, then they have the same transfer function?

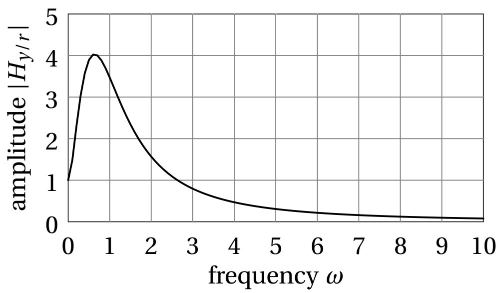


FIGURE 5.27: Amplitude $|H_{y/r}(\mathrm{i}\omega)|$ of the transfer function from r to y of Example 5.8.5.

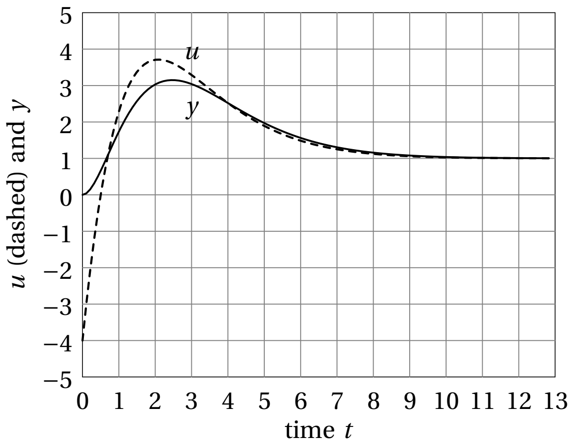


FIGURE 5.28: The response of the hand $u(t)$ (dashed) and the tip of the pendulum $y(t)$ to the step $r(t) = \mathbb{1}(t)$ of the walking juggler (Example 5.8.5).

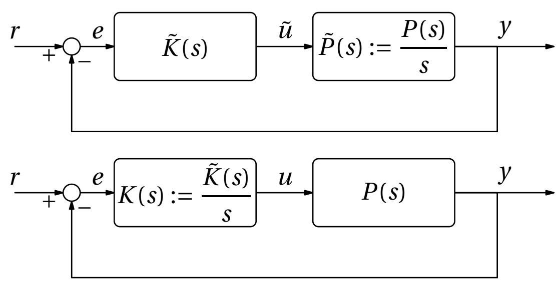


FIGURE 5.29: By designing a stabilizing controller $\tilde{K}(s)$ for $\tilde{P}(s) := \frac{P(s)}{s}$, we obtain the stabilizing controller $K(s) := \frac{\tilde{K}(s)}{s}$ for $P(s)$ with a pole at the origin.

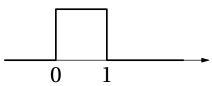
(h) If the transfer functions of $\dot{x}_1 = A_1 x_1 + B_1 u, y = C_1 x_1$ and $\dot{x}_2 = A_2 x_2 + B_2 u, y = C_2 x_2$ are equal, then the systems are isomorphic?

(i) If $H(\mathrm{i}\omega) = C(\mathrm{i}\omega I - A)^{-1} B + D$, then $\lim_{\omega\to\infty} H(\mathrm{i}\omega) = D$?

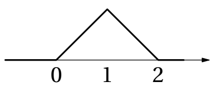
(j) If $H(s)$ is an asymptotically stable transfer function, then the DC gain of $1/(1+H(s))$ is never 0?

5.2 Suppose that $y = \mathscr{H}(u)$ is time-invariant (but not necessarily linear). Assuming that the response to constant inputs is well defined show that the response y is constant if u is constant.

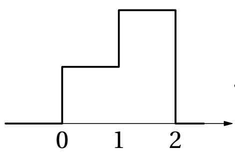
5.3 Suppose $y = \mathscr{H}(u)$ is LTI and that we know that the response $y_0(t)$ to

$u_0(t) =$ 

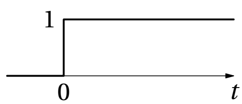
is equal to

$y_0(t) =$ 

Sketch the response $y(t)$ to

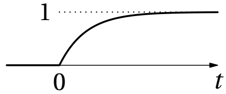
$u(t) =$  .

5.4 Suppose $y = \mathscr{H}(u)$ is LTI and that we know that the output for the unit step function

$u_0(t) = \mathbb{1}(t)$ 

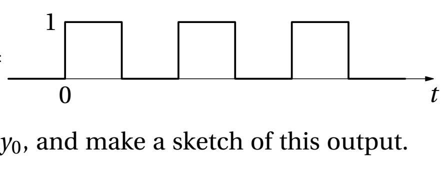
is equal to

$$y_0(t) = \begin{cases} 0 & t < 0 \\ 1 - \mathrm{e}^{-2t} & t \geq 0 \end{cases}$$



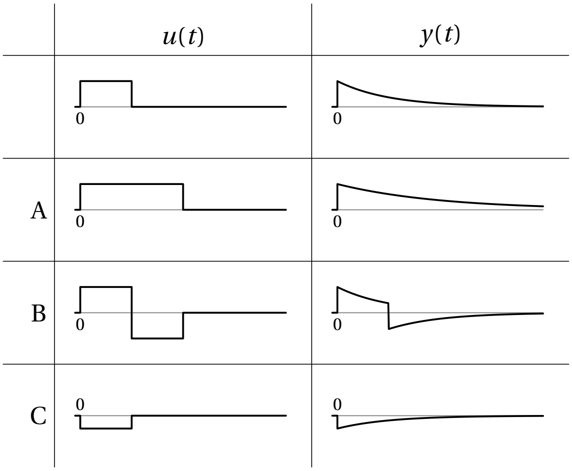
Express the output $y(t)$ for the square wave

$$u(t) = u_0(t) - u_0(t-1) + u_0(t-2) - u_0(t-3) + \cdots$$

$=$ 

in terms of y_0, and make a sketch of this output.

5.5 Suppose $y = \mathscr{H}(u)$ is LTI. The first row of the table below gives a possible input/output pair (u, y) of this system.



Which pairs (u, y) in the rows A, B, C correspond to the given pair? Indicate for each row A, B, C what you use: linearity, time invariance, or both?

5.6 *Linearity and time invariance*

(a) Is the system $y(t) = u^2(t)$ linear?

(b) Is the system $y(t) = u^2(t)$ time invariant?

(c) Is the system $y(t) = t^2 u(t)$ linear?

(d) Is the system $y(t) = t^2 u(t)$ time invariant?

(e) Is the system $y(t) = u(t^2)$ linear?

(f) Is the system $y(t) = u(t^2)$ time invariant?

5.7 Consider the system $y(t) = u(\alpha t)$ with $u, y : \mathbb{R} \to \mathbb{R}$ and with α a given real number.

(a) Is the system linear?

(b) For which $\alpha \in \mathbb{R}$ is the system time invariant?

5.8 Show that the system $y(t) = u(t) + \frac{1}{2}u(t-1) + \frac{1}{4}u(t-2) + \frac{1}{8}u(t-3) + \cdots$ is LTI.

5.9 The following systems are LTI. Determine which of the following systems are BIBO stable:

(a) $y(t) = u(t) + u(t-1)$.

(b) $y(t) = \sum_{n=1}^{\infty} \frac{1}{n} u(t-n)$. (Assume that u is initially-at-rest.)

(c) $y(t) = \frac{1}{P} \int_{t-P/2}^{t+P/2} u(\tau)\, \mathrm{d}\tau$.

(d) initially-at-rest system $\dot{y}(t) = y(t)/2 + u(t)$.

The next system is not LTI (not even linear). Assuming we define BIBO-stability the same for nonlinear systems, is this system BIBO-stable:

(e) $y(t) = \sin(\sqrt{|u(t)|})$. [This one might be tricky.]

5.10 Determine the maximal peak-to-peak gain of the systems of Exercise 5.9. (They can be ∞.)

5.11 *Preservation of mass.* Suppose that $y = h * u$ is a BIBO-stable LTI system with impulse response $h(t)$. Show that the following four statements are equivalent:

(a) $\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t = 1$.

(b) $y = u$ if u is constant.

(c) $H(0) = 1$.

(d) *Total mass in equals total mass out*: if u and y are absolutely integrable[5] then

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

[Hint: Use the Fourier transform property $Y(\mathrm{i}\omega) = H(\mathrm{i}\omega)U(\mathrm{i}\omega)$.]

5.12 Assume the following systems are initially-at-rest. Then they are LTI as well. Determine which of the systems are BIBO stable, and if it is, determine the DC gain:

(a) $y(t) = u(t-1)$.

(b) $\dot{y}(t) + y(t)/2 = u(t)$.

(c) $\dot{y}(t) - y(t)/2 = u(t)$.

(d) $\ddot{y}(t) + 2\dot{y}(t) + 3y(t) = \dot{u}(t) + 2u(t)$.

(e) $\ddot{y}(t) - 2\dot{y}(t) - 3y(t) = -\dot{u}(t) - 2u(t)$. [Tricky.]

5.13 Suppose that the system $y = h * u$ is BIBO stable.

(a) Show that $\|\mathscr{H}\|_1 \geq |H(\mathrm{i}\omega)|$ for all $\omega \in \mathbb{R}$.

(b) Show that $\|\mathscr{H}\|_1 = H(0)$ if $h(t) \geq 0$ for all t.

5.14 *Delay.* Let $t_0 > 0$. Consider the system $y(t) = u(t - t_0)$. This system is a pure delay.

(a) Show that the system is LTI.

[5] Meaning: $\int_{-\infty}^{\infty} |u(t)|\,\mathrm{d}t < \infty$ and $\int_{-\infty}^{\infty} |y(t)|\,\mathrm{d}t < \infty$. By the way, it can be proved that y is absolutely integrable if both u and h are absolutely integrable.

(b) Show that the system is BIBO stable.

(c) Show that $H(\mathrm{i}\omega)$ exists.

(d) Show that $H(s) = \mathrm{e}^{-st_0}$.

5.15 *Transfer Function.* Determine the transfer function from u to y for the following initially-at-rest systems:

(a) $\dot{y}(t) + y(t) = u(t)$.

(b) $\dot{y}(t) + y(t) = \dot{u}(t)$.

(c) $\ddot{y}(t) + 2\dot{y}(t) + 3y(t) = \dot{u}(t)$.

(d) $(\frac{\mathrm{d}}{\mathrm{d}t} + 1)^5 y(t) = u(t)$.

(e) $\dot{y}(t) - y(t) = \dot{u}(t) - u(t)$.

(f) $y(t) = u(t-1) + u(t-2)$.

(g) $y(t) = u(t) + \frac{1}{2}u(t-1)$.

5.16 *Proper transfer functions.* Let $H(s) = C(sI - A)^{-1}B + D$ for certain $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times 1}, C \in \mathbb{R}^{1\times n}, D \in \mathbb{R}$. From linear algebra, we know that the (i,j)th element of the inverse of a matrix T can be written as

$$(T^{-1})_{ij} = \frac{\det(T_i)}{\det(T)}$$

with T_i the matrix T in which the ith *column* has been replace by the jth unit vector $e_j = (0,\dots,0,1,0,\dots,0)$ (column vector). This result is know as *Cramer's rule*.

(a) Show that $(sI - A)^{-1}$ is a matrix of rational functions in s with degree of the numerator less than that of the denominator.

(b) Show that $\lim_{s\to\infty} H(s) = D$.

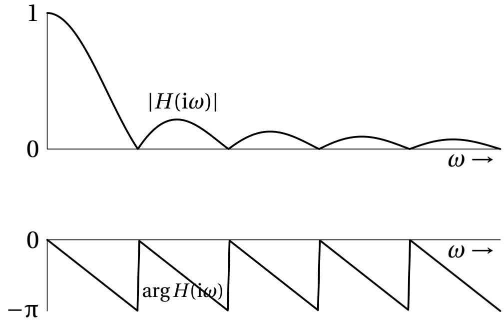
(c) Show that $H(s)$ is a rational function, $H(s) = Q(s)/P(s)$, in which $P(s), Q(s)$ are polynomials and the degree of $P(s)$ is at least that of $Q(s)$.

5.17 Determine the transfer function of

(a) initially-at-rest $\ddot{y}(t) + 3y(t) = \dot{u}(t)$.

(b) initially-at-rest echo system $y(t) = \frac{1}{2}y(t-5) + u(t)$ (with $t \in \mathbb{R}$).

5.18 *Moving-average system.* Consider the moving average system and its frequency response as determined in Example 5.4.3. A plot of the absolute value and argument of $H(\mathrm{i}\omega)$ are



(The labels on the frequency axis are missing for a reason.)

(a) Determine the DC gain.

(b) Express the frequencies $\omega_k \geq 0$ for which $H(\mathrm{i}\omega_k) = 0$ in terms of P, and what can you say about the response $y(t)$ if $u(t) = \cos(\omega_k t)$?

(c) Do the plots corroborate that high frequency signals are reduced by this system?

5.19 Show that the two systems of Lemma 3.5.6 have the same transfer function and use this to prove Lemma 3.5.6.

5.20 Section 5.6 correctly claims that the transfer function $H(s) = C(sI - A)^{-1}B + D$ exists if $\operatorname{Re}(s) > \max_i \operatorname{Re}(\lambda_i(A))$. In this exercise we relax this condition.

Suppose that the system is in the Kalman controllability form

$$\dot{x} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} x + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u,$$
$$y = \begin{bmatrix} C_1 & C_2 \end{bmatrix} x + Du.$$

(a) Use the definition of a matrix exponential to show that e^{At} is block upper triangular:

$$\mathrm{e}^{At} = \begin{bmatrix} \mathrm{e}^{A_{11}t} & ? \\ 0 & \mathrm{e}^{A_{22}t} \end{bmatrix}$$

with ? an unspecified block.

(b) Show that the impulse response does not depend on A_{12}, A_{22}, C_2.

(c) Show that $H(s)$ exists if $\operatorname{Re}(s) > \max_i \operatorname{Re}(\lambda_i(A_{11}))$.

(d) Show that transfer function of $\dot{x} = Ax + Bu, y = Cx + Du$ exists if $\operatorname{Re}(s) > \max_i \operatorname{Re}(s_i)$ where s_i are the poles of the rational function $H(s) = C(sI - A)^{-1}B + D$. [Tricky question.]

5.21 *Feedback with high amplification.* It is 1927 and Black is tasked with making a reliable amplifier. All he has at his disposal is

- an inaccurate constant amplifier with amplification factor somewhere between 5000 and 50000
- a very accurate constant amplifier with amplification factor 1/10 (this weakens the signal).

With which of the four block diagrams in Fig. 5.18 was Black able to make an accurate amplifier from r to y with amplification factor 10 (with an error of less than 1%)? Explain.

(This is Black's negative feedback amplifier in a nutshell. Its importance cannot be overestimated.)

5.22 Consider the system of Figure 5.18(c) with transfer functions

$$P(s) = \frac{s^2+3}{(s+1)^2}, \qquad K(s) = \frac{1}{2}.$$

(a) Determine the transfer function from r to y.

(b) Is there a sinusoid $r(t) = \cos(\omega_0 t)$ for which the response $y(t)$ is always zero?

5.23 Consider the system of Figure 5.18(a). Assume that the two systems are described by

$$\Sigma_P : \dot{y}_1 = r, \qquad \Sigma_K : \dot{y}_2 = -y_2 + r.$$

(a) Determine the transfer functions $P(s), K(s)$.

(b) Determine the transfer function from r to y.

(c) Use part (b) to formulate a differential equation $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})r$ for the system with input r and output y.

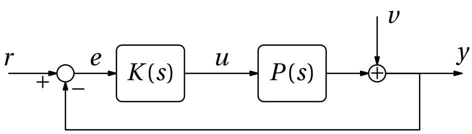


FIGURE 5.30: A closed-loop system.

5.24 Check the $H_{y/w}(s)$ in (5.18).

5.25 *Closed-loop transfer functions.* Consider the system of Figure 5.30 with $K(s) = 1/s, P(s) = M > 0$.

(a) Determine the transfer function from v to y.

(b) Determine the transfer function from r to y.

(c) Is the closed loop asymptotically stable?

(d) Determine the response $y(t)$ to the constant input v (with $r = 0$).

(e) Determine the response $y(t)$ to the constant input r (with $v = 0$).

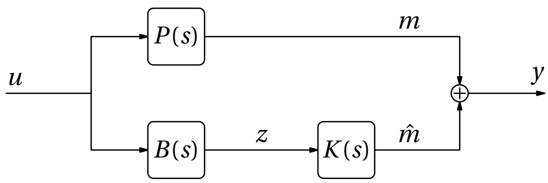


FIGURE 5.31: Filtering.

5.26 *Filtering and noise reduction.* Consider the configuration in Figure 5.31. It is an interconnection of three systems with transfer functions P, B, K.

(a) Determine the transfer function from u to y.

(b) Suppose that u is the noise produced by, say, the propulsion system of an airplane. The sound m a passenger hears is a function of u, that is, $m = P(s)u$. To suppress m, we must construct a "counter noise" $\hat{m}$ based on another measured sound signal $z = B(s)u$. In other words, we must design a filter $K(s)$. Suppose

$$P(s) = \frac{1}{s+1}, \qquad B(s) = \frac{s-\beta}{s+2}.$$

For which $\beta \in \mathbb{R}$ can an asymptotically stable system $K(s)$ be designed for which the steady-state response $y(t)$ is zero for all u (so, for which the passenger does not hear the propulsion system at all).

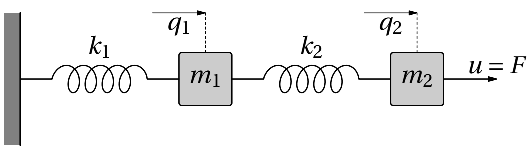


FIGURE 5.32: Another spring-mass system.

5.27 *Interpretation of zeros and poles.* Consider the spring-mass system of Figure 5.32. For simplicity, we assume that the two masses are equal to 1. The positions q_1 and q_2 of the two masses are described by the equations

$$\ddot{q}_1 = k_2(q_2 - q_1) - k_1 q_1,$$
$$\ddot{q}_2 = u - k_2(q_2 - q_1).$$

The input u is an external force that can be exerted on the second mass. Take $y = q_2$ as output.

(a) Determine the transfer function $H_{y/u}(s)$.

(b) What is the connection between the zeros of $H_{y/u}(s)$ and the input signals u for which the output (the second mass) remains at rest for all t?

(c) What is the connection between the poles of $H_{y/u}(s)$ and the motion of the second mass (the output) if we take $u \equiv 0$?

5.28 Let

$$H(s) = \frac{(s-1)}{(s-2)(s+3)}.$$

(a) Determine a differential equation $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ with $H(s)$ as transfer function.

(b) Does there exist an asymptotically stable $P(\frac{\mathrm{d}}{\mathrm{d}t})y = Q(\frac{\mathrm{d}}{\mathrm{d}t})u$ with $H(s)$ as transfer function?

5.29 *Interconnection of systems.* Consider two systems Σ_P, Σ_K with state-space representations

$$y = P(s)u: \qquad \begin{cases} \dot{x} &= Ax + Bu, \\ y &= Cx \end{cases}$$

and

$$\hat{y} = K(s)\hat{u}: \qquad \begin{cases} \dot{z} &= \hat{A}z + \hat{B}u, \\ \hat{y} &= \hat{C}z. \end{cases}$$

Determine state-space representations with state (x, z), input r, and output y for

(a) the parallel circuit (see Figure 5.18(a))

(b) the series circuit (see Figure 5.18(b))

(c) the feedback (see Figure 5.18(d))

[Remark: this exercise shows that interconnections with transfer functions are simpler than those with state-space representations.]

5.30 *RLC circuit.* Consider the *RLC* circuit of Example 2.1.3 (page 36) with u the initial source voltage and y the output voltage across the capacitor.

(a) Determine the transfer function $H_{y/u}(s)$ and frequency response $H_{y/u}(\mathrm{i}\omega)$.

(b) What is the response $y(t)$ to $u(t) = \sin(\omega_0 t)$?

(c) What is the response $y(t)$ to $u(t) = 1$?

5.31 Consider the block diagram in Figure 5.18(d). In Figure 5.25, we saw that the step response does not always converge *monotonically* to the chosen final value. This is unwanted, but cannot always be prevented. Suppose that $P(s)$ has a double pole at the origin, so

$$P(s) = \frac{1}{s^2} P_0(s)$$

with $P_0(0) \neq 0$. Also assume that the closed loop is asymptotically stable.

(a) Show that $K(0) \neq 0$.

(b) Show that the product $K(s)P(s)$ has a double pole at $s = 0$.

(c) Suppose that $r(t) = \mathbb{1}(t)$. Show that the Laplace transform $E(s)$ of e is zero at zero: $E(0) = 0$.

(d) Suppose that $r(t) = \mathbb{1}(t)$. Show that $\int_{-\infty}^{\infty} e(t)\,\mathrm{d}t = 0$.

(e) Suppose that $r(t) = r\mathbb{1}(t)$ for some $r \neq 0$, and that $y(0) = 0$. Use the previous part to show that $y(t)$ can not converge monotonically to r!

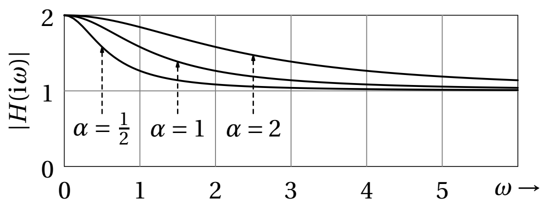
5.32 *Loudness.* The human ear is not very sensitive to low frequencies at a low volume. To remedy this, radios and TVs sometimes have a *loudness* system. The loudness system amplifies low-frequency signals.

(a) Suppose that $\alpha, \beta, \gamma > 0$. Determine $H(0)$ and $\lim_{\omega\to\infty} H(\mathrm{i}\omega)$ for the system described by

$$\dot{y} + \alpha y = \beta \dot{u} + \gamma u.$$

(b) Determine the relation between positive α, β, γ, such that very low frequencies are amplified with a factor 2, and very high frequencies are not amplified (i.e. amplified with a factor 1).

Remark: in this relation we can still choose α. The α determines the frequency band over which effective amplification takes place. For a couple of α's the graphs of $|H(\mathrm{i}\omega)|$ are



5.33 *RC circuit.* Consider the *RC* circuit of Example 5.5.2. Determine $|H\mathrm{i}\omega)|$ and $\arg H(\mathrm{i}\omega)$ as real-valued functions of $\omega > 0$ and verify (5.11).

5.34 *Feedback.* Consider the block diagram in Figure 5.33 with

$$K(s) = k \in \mathbb{R}, \qquad P(s) = \frac{s-1}{s+3}.$$

(a) For which $k \in \mathbb{R}$ is the closed-loop system asymptotically stable?

(b) Let r be constant and nonzero, and suppose that the closed-loop system is asymptotically stable for some $K(s) = k$. Do we then have $\lim_{t\to\infty} y(t) = r$?

5.35 Consider the cruise control problem of Example 5.8.3 (so with $P(s)$ and $K(s)$ as in that example), but now with an external signal v (see Figure 5.33). This term can represent the influence of wind.

(a) Determine the transfer function $H_{y/v}(s)$ from v to y.

(b) Is $H_{y/v}(0) = 0$?

(c) Let r and v be constant. Does a constant wind v influence the final velocity $\lim_{t\to\infty} y(t)$?

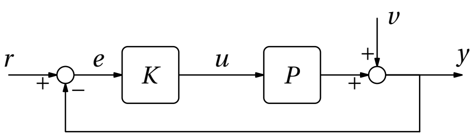
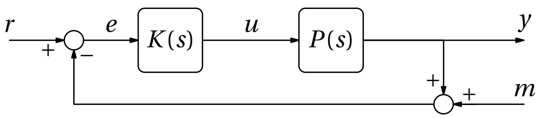


FIGURE 5.33: Cruise control system with noise v.

5.36 Prove Theorem 5.7.3. [Hint: have a look at Lemma 2.5.8.]

5.37 Consider the closed loop system



with $K(s)$ and $P(s)$ rational transfer functions $K(s) = \frac{N_K(s)}{D_K(s)}$, $P(s) = \frac{N_P(s)}{D_P(s)}$. The signal m models a measurement error (it is the difference between the measurement of y, on which the control is based, and the actual y.)

(a) Determine the transfer function $H_{y/m}$.

(b) Suppose $P(s) = 1/s^2$. Determine a controller $K(s)$ that stabilizes the closed loop system.

(c) Suppose $P(s) = 1/s^2$ and that $K(s)$ stabilizes the closed loop system. Let $r(t) = 0$ and $m(t) = m_0$ (a constant). Determine $\lim_{t\to\infty} y(t)$. Does this limit depend on the choice of stabilizing controller?

5.38 *Youla parameterization of the juggler.* Consider Figure 5.18(d). Let $Q(s)$ and $R(s)$ be two polynomials. Show that the controller

$$K(s) = \frac{Q(s) + R(s)(s^2 - 2)}{2R(s)}$$

stabilizes the system $P(s) = -2/(s^2 - 2)$ if and only if $Q(s)$ is asymptotically stable.

5.39 Consider Figure 5.18(d) and set $P(s) = 1/(s-1)$. Determine all controllers $K(s) = N_K(s)/D_K(s)$ for which the closed-loop characteristic polynomial is equal to

(a) $\chi_{\text{closed}}(s) = 1$

(b) $\chi_{\text{closed}}(s) = s+2$

(c) $\chi_{\text{closed}}(s) = (s+2)^2$

For which of these three case is there a proper[6] K and for which is there a strictly proper K?

5.40 *Second-order system.* Suppose that p_2, p_1, p_0 are nonzero and have the same sign. Show that every system

$$p_2\ddot{y}(t) + p_1\dot{y}(t) + p_0 y(t) = q_0 u(t)$$

can be written as

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = d\omega_0^2 u(t)$$

for some $\zeta > 0, \omega_0 > 0$ and $d \in \mathbb{R}$.

Tougher Exercises

5.41 Is the system with impulse response $h(t) = \sin(t)/t$ BIBO stable?

5.42 *Relative degree.* Consider a system with transfer function $H(s) = Q(s)/P(s) = C(sI - A)^{-1}B + D$, and define the relative degree of $H(s)$ as $r := \deg(P) - \deg(Q)$. Suppose $r \geq 1$. Prove that

$$CA^iB = \begin{cases} 0 & \text{if } i = 0, 2, \ldots, r-2 \\ \neq 0 & \text{if } i = r-1. \end{cases}.$$

5.43 *Causality of convolutions.* Let $u, h, y : \mathbb{R} \to \mathbb{R}$. Consider the convolution system $y = h * u$,

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

Show that it is causal if and only if $h(t) = 0$ for all $t < 0$. (Causality is defined in Definition 1.6.6.)

5.44 *Smith predictor.* Determine the transfer function from r to y of the system in Figure 5.34.

[6] *Proper* means that the degree of the numerator is less than or equal to the degree of the denominator. *Strictly* proper means that the degree of the numerator is less than that of the denominator.

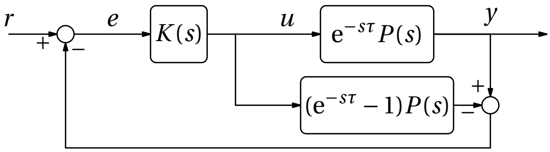


FIGURE 5.34: Smith predictor.

5.45 *Stability of an infinite-dimensional system.* Is the system $y(t)+y^{(1)}(t)-\frac{1}{2}y(t-1)=u(t)$ BIBO stable?

5.46 Consider the system

$$\dot{y}(t)+y(t-\pi/2)=-\dot{u}(t-\pi/2)+u(t)$$

and assume it is initially-at-rest up to time $t=0$. Determine the step response. [Hint: Use that $y(t)=0$ for all $t<0$.]

5.47 Now for a tricky and confusing question! Let $\mathbb{U}$ be the vector space defined as

$$\mathbb{U}=\{u:\mathbb{R}\to\mathbb{R}\mid \lim_{t\to\infty}u(t)\text{ exists}\}.$$

This is our space of inputs. We take the space of outputs to be the same, $\mathbb{Y}=\mathbb{U}$. Consider the system $y=\mathscr{H}(u)$ defined as

$$y(t)=\lim_{\tau\to\infty}u(\tau).$$

Notice that every output is constant!

(a) Show that $\mathscr{H}:\mathbb{U}\to\mathbb{Y}$ is linear.

(b) Show that $\mathscr{H}:\mathbb{U}\to\mathbb{Y}$ is time invariant.

(c) So our system is LTI. The general wisdom is that the system hence is a convolution. Show that $\mathscr{H}$ is not a convolution!

What is weird and non-practical about this system is that the value of $y(t)$ at any time t depends on the "infinite future" of the input. This system is absurdly non-causal. Also the vector spaces are rather special.

Appendix A

Some proofs and derivations

This appendix contains a number of proofs and derivations of results used in this course. They are included here because they fall beyond the scope of this course. We also give an overview of the MATLAB scripts we used.

A.1 Heated Beam (Example 1.7.5)

For completeness, we briefly, and without derivations, summarize the standard mathematical model. The heat balance at every position r over an infinitesimal segment dr of the beam gives

$$\frac{\partial T(t,r)}{\partial t} = \frac{\sigma}{\rho C} \frac{\partial^2 T(t,r)}{\partial r^2} \tag{A.1}$$

with σ the heat-transfer coefficient, ρ the density, and C the specific heat of the beam. At the left extremity of the beam, we have for the incoming heat

$$-A \left. \frac{\partial T(t,r)}{\partial r} \right|_{r=0} = u(t), \tag{A.2}$$

where A is the area of the cross section of the beam. Because the right extremity is isolated, we have

$$\left. \frac{\partial T(t,r)}{\partial r} \right|_{r=L} = 0. \tag{A.3}$$

In this case, the differential equation is therefore given by the *partial differential equation* (A.1), together with the boundary conditions (A.2) and (A.3). Note that in this example, the input $u(t)$ only comes in through the boundary condition (A.2). In this course, we do not discuss partial differential equations.

A.2 Weak Solution (Thm. 2.2.4)

If u is bounded but discontinuous, then the x defined in (2.14) is well defined and continuous, but not differentiable in the classic sense. What does $\dot{x} = Ax + Bu$ then mean? The usual solution of this paradox is to say that x is a *weak* solution of $\dot{x} = Ax + Bu$ if x satisfies the integrated equation: $x(t) = x(t_0) + \int_{t_0}^{t} Ax(\tau) + Bu(\tau)\,\mathrm{d}\tau$. The choice of t_0 is irrelevant. One can verify that the following holds.

Lemma A.2.1 (Weak solution). *For every locally integrable*[1] *u, the x in (2.14) is well defined, is continuous, and is a weak solution of $\dot{x} = Ax + Bu$.*

Proof (sketch). Since $u(\tau)$ is locally integrable and all components of $\mathrm{e}^{A(t-\tau)}B$ are continuous as functions of τ, every component of $\mathrm{e}^{A(t-\tau)}Bu(\tau)$ is also locally integrable as a function of τ. Consequently, the $x(t)$ defined in (2.14) exists. This $x(t)$ is even continuous because

$$
\begin{aligned}
&\lim_{h\to 0} x(t+h) - x(t) \\
&= \lim_{h\to 0} (\mathrm{e}^{Ah} - I)x(t) + \int_0^h \mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau \\
&= \underbrace{\lim_{h\to 0} (\mathrm{e}^{Ah} - I)x(t)}_{=0} + \underbrace{\lim_{h\to 0} \int_0^h \mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau}_{=0} \\
&= 0.
\end{aligned}
$$

The last limit is zero by the dominant convergence theorem from the theory of Lebesgue integration. This theorem says that $\lim_{n\to\infty} \int f_n(\tau)\,\mathrm{d}\tau = \int f(\tau)\,\mathrm{d}\tau$ if the f_n are (locally) integrable and $f(t) = \lim_{n\to\infty} f_n(t)$ almost everywhere. If we take $f_h(\tau) := \mathrm{e}^{A(t-\tau)}Bu(\tau)$ for $\tau \in [0,h]$ and $f_h(\tau)$ zero elsewhere, and define $f(\tau) := \lim_{h\to 0} f_h(\tau)$, then $f(\tau) = 0$ for all $\tau \neq 0$. In this case, the dominant convergence theorem gives

$$
\lim_{h\to 0} \int_0^h \mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau = \lim_{h\to 0} \int f_h(\tau)\,\mathrm{d}\tau = \int f(\tau)\,\mathrm{d}\tau = 0.
$$

■

In particular, the x in (2.14) is a well-defined (weak) solution for every locally *bounded* but possibly discontinuous u. The notion of weak solution may look like a trick, but it is not. During the modeling process—when the differential equations are determined—the signals are often assumed to be differentiable purely for convenience. However, they are not necessarily that smooth. In such cases, the weak solutions with their restricted smoothness condition are more natural and more general.

[1] We say that u is *locally integrable* if $\int_a^b |u_i(t)|\,\mathrm{d}t < \infty$ for all $a, b \in \mathbb{R}$ and all components u_i of u.

A.3 Routh–Hurwitz (Thm. 2.3.6)

The Routh–Hurwitz test (Thm. 2.3.6) is stated in terms of the Routh table because the test is then easier to carry out. To understand the test, it is better to state it in terms of polynomials.

Theorem A.3.1 (Routh–Hurwitz). *A non-constant polynomial* $p(s) = p_0 s^n + p_1 s^{n-1} + \cdots + p_n$ $(p_i \in \mathbb{R}, p_0 \neq 0)$ *is asymptotically stable if and only if*

1. p_0 *and* p_1 *have the same sign (in particular,* $p_1 \neq 0$*);*
2. *the degree* $n-1$ *polynomial*

$$q(s) := p(s) - \frac{p_0}{p_1}(p_1 s^n + p_3 s^{n-2} + p_5 s^{n-4} + \cdots)$$

is asymptotically stable.

Proof. Let λ_i be the zeros of $p(s)$. Expanding the factorization

$$p(s) = p_0 \prod_{i=1}^{n}(s - \lambda_i) \tag{A.4}$$

shows that $p_1 = -p_0 \sum_i \lambda_i$ (this is one of *Vieta's formulas*). If p is asymptotically stable then all its zeros λ_i have negative real part so then by Vieta's formula p_1 has the same sign as p_0. This proves that part 1 holds for asymptotically stable polynomials.

Now assume that p_1 is nonzero and define the family of polynomials r_η by

$$r_\eta(s) := p(s) - \eta \tfrac{p_0}{p_1}(p_1 s^n + p_3 s^{n-2} + \cdots), \qquad \eta \in [0,1].$$

For $\eta = 0$ we have $r_\eta = p$, and for $\eta = 1$ we have $r_\eta = q$. A special property of this family of polynomials is that the *imaginary* zeros (including their multiplicity) do not depend on η. (The proof will be given later.) This means that when we vary η, none of the zeros of r_η can cross or land on the imaginary axis. The only way the number of stable zeros of r_η can change is if the degree drops. The polynomial r_η equals

$$r_\eta(s) = p_0(1-\eta)s^n + p_1 s^{n-1} + \cdots.$$

From this form it follows that the only value of η for which the degree drops is $\eta = 1$ (so when $r_\eta = q$) and then it drops precisely one degree because we assumed $p_1 \neq 0$. For all $0 \leq \eta < 1$, the polynomial r_η therefore has as many stable zeros as $r_0 = p$. By Vieta's formula, the zeros $\lambda_{i,\eta}$ (stable and unstable) of r_η add up to

$$\sum_i \lambda_{i,\eta} = \frac{-p_1}{p_0(1-\eta)}.$$

In the limit $\eta \uparrow 1$, precisely $n-1$ of these zeros go to the zeros of $r_1 = q$ and the remaining zero therefore goes to $-p_1/(p_0(1-\eta))$ minus those $n-1$ zeros. This remaining zero hence goes to $\pm\infty$. This remaining zero is stable if and only if $-p_1/(p_0(1-\eta))$ is negative for $\eta \uparrow 1$, in other words, if and only if p_0 and p_1 have the same sign. Done.

Remains to prove that the imaginary zeros of r_η are independent of η: We prove it for even n (the proof for odd n is analogous). For even n, we can write r_η as

$$r_\eta(s) = [p_{\text{even}}(s) - \eta \tfrac{p_0}{p_1}\, s\, p_{\text{odd}}(s)] + p_{\text{odd}}(s).$$

(The even part of a polynomial is the sum of the even powers s^{2k} and the odd part is the sum of the odd powers s^{2k+1}.) An imaginary $s = \mathrm{i}\omega$ is a zero of r_η of multiplicity k if and only if it is a zero of multiplicity k of both the even part $[p_{\text{even}} - \eta \frac{p_0}{p_1} s p_{\text{odd}}]$ and the odd part p_{odd}. This is because for imaginary $s = \mathrm{i}\omega$, the even part is real and the odd part is imaginary. So $\mathrm{i}\omega$ is a zero of r_η of multiplicity k if and only if it is a zero of multiplicity k of both p_{odd} and $p_{\text{even}} - \eta \frac{p_0}{p_1} s p_{\text{odd}}$, but that is the case iff it is a zero of multiplicity k of both p_{odd} and of p_{even}. This is independent of η. ■

Now, we need to connect this theorem to the Routh table. It is not difficult (but also no fun) to verify that the Routh table of $q(s)$ is exactly that of $p(s)$ minus the first row. Since the Routh–Hurwitz test is correct for first-degree polynomials (verify), it follows by induction that it is correct for every degree n.

A.4 Linearization (Thm. 2.4.2)

Proof of Thm. 2.4.2. Without loss of generality, we assume $x^* = 0$ and therefore $x = \delta_x$. We prove that asymptotic stability of $\dot{\delta}_x = A\delta_x$ implies asymptotic stability of $\dot{x}(t) = f(x(t))$. A difference between the multidimensional case (Thm. 2.4.2) and the scalar case (Lemma. 2.4.1) is that in the multidimensional case, asymptotic stability does not necessarily imply *monotone* convergence of the components to the equilibrium point. The state can first increase (in norm) before it converges to the origin. We now define a quantity V that *does* decrease monotonically and from which we can deduce the convergence of x. Let

$$V_{x_0}(t) = \int_t^\infty x^{\mathrm{T}}(\tau)x(\tau)\,\mathrm{d}\tau$$

with x the solution of the linearization $\dot{x} = Ax$ and $x(0) = x_0$. This Lyapunov function V, as it is called, is quadratic in $x(t)$:

$$\begin{aligned} V_{x_0}(t) &= \int_t^\infty x^{\mathrm{T}}(\tau)x(\tau)\,\mathrm{d}\tau \\ &= \int_0^\infty x^{\mathrm{T}}(t)\mathrm{e}^{A^{\mathrm{T}}\tau}\mathrm{e}^{A\tau}x(t)\,\mathrm{d}\tau \\ &= x^{\mathrm{T}}(t)Px(t) \end{aligned}$$

with

$$P := \int_0^\infty \mathrm{e}^{A^\mathrm{T}\tau}\mathrm{e}^{A\tau}\,\mathrm{d}\tau.$$

This P exists because the integrand $\mathrm{e}^{A^\mathrm{T}\tau}\mathrm{e}^{A\tau}$ consists of exponentially decreasing functions. The matrix P is also invertible because otherwise, there would exist an $x_0 \neq 0$ such that $V_{x_0}(t) = 0$, which is not the case. By construction, $V_{x_0}(t)$ is decreasing as a function of time,

$$\dot V_{x_0}(t) = -x^\mathrm{T}(t)x(t) = -\|x(t)\|^2.$$

Now, consider a solution $x(t)$ of the nonlinear system $\dot x = f(x)$. Then $V_{x_0}(t) := x^\mathrm{T}(t)Px(t)$ is also decreasing – at least in some small neighborhood of the equilibrium – because

$$\begin{aligned}\dot V_{x_0}(t) &= 2x^\mathrm{T}(t)P\dot x(t)\\ &= 2x^\mathrm{T}(t)P(Ax + \mathrm{o}(x(t)))\\ &= -\|x(t)\|^2 + \mathrm{o}(\|x(t)\|^2)\\ &< -\frac{1}{2}\|x(t)\|^2 \text{ for some } \gamma > 0\\ &\qquad\qquad \text{and all } 0 < \|x(t)\| < \gamma.\end{aligned}$$

Since P is invertible, for every $\epsilon > 0$ there exists a $\delta > 0$ such that $x^\mathrm{T}Px < \delta$ implies $\|x\| < \min(\gamma,\epsilon)$. Consequently, for every $x_0^\mathrm{T}Px_0 < \delta$, the function V_{x_0} is decreasing,

$$\dot V_{x_0}(0) < -\frac{1}{2}\|x(0)\|^2,$$

and therefore $x^\mathrm{T}(t)Px(t) < \delta$. Hence $\|x(t)\| < \min(\gamma,\epsilon)$; that is, *for all* $t > 0$,

$$\dot V_{x_0}(t) < -\frac{1}{2}\|x(t)\|^2.$$

The right-hand side can be bounded in terms of V_{x_0}: there exists an $M > 0$ such that $\|x\|^2 > Mx^\mathrm{T}Px$. We then have

$$\dot V_{x_0}(t) < -\frac{M}{2}V_{x_0}(t) \qquad \forall t > 0.$$

The function $V_{x_0}(t)$ is therefore bounded from above by $\mathrm{e}^{-M/2t}V_{x_0}(0)$ and consequently converges to zero for $t \to \infty$. Finally, because of the invertibility of P, we then also have $\lim_{t\to\infty} x(t) = 0$. ■

A.5 Model of the Inverted Pendulum (Example 3.2.2)

The following is a straight-forward derivation. Let $\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right] \in \mathbb{R}^2$ be the position of the tip of the pendulum. Then we have

$$\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}u + \ell\sin(\phi)\\ \ell\cos(\phi)\end{bmatrix}.$$

Differentiating with respect to time gives

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \begin{bmatrix}\dot{u} + \ell\cos(\phi)\dot{\phi}\\ -\ell\sin(\phi)\dot{\phi}\end{bmatrix}.$$

Differentiating again, we obtain

$$\begin{bmatrix}\ddot{x}\\ \ddot{y}\end{bmatrix} = \begin{bmatrix}\ddot{u} - \ell\sin(\phi)\dot{\phi}^2 + \ell\cos(\phi)\ddot{\phi}\\ -\ell\cos(\phi)\dot{\phi}^2 - \ell\sin(\phi)\ddot{\phi}\end{bmatrix}. \tag{A.5}$$

By Newton's second law, this is equal to the force $F \in \mathbb{R}^2$ divided by the mass. The force consists of the gravity $-mg$ in the y-direction and a compression force λ in the direction of the stick. So the total force is

$$\begin{bmatrix}0\\ -mg\end{bmatrix} + \lambda\begin{bmatrix}\sin(\phi)\\ \cos(\phi)\end{bmatrix},$$

for some compression force λ. This, divided by m, equals (A.5), so

$$\begin{bmatrix}0\\ -g\end{bmatrix} + \tfrac{1}{m}\lambda\begin{bmatrix}\sin(\phi)\\ \cos(\phi)\end{bmatrix} = \begin{bmatrix}\ddot{u} - \ell\sin(\phi)\dot{\phi}^2 + \ell\cos(\phi)\ddot{\phi}\\ -\ell\cos(\phi)\dot{\phi}^2 - \ell\sin(\phi)\ddot{\phi}\end{bmatrix}.$$

This is a linear equation in $(\lambda, \ddot{\phi})$. The solution is

$$\begin{aligned}\tfrac{1}{m}\lambda &= \cos(\phi)g + \sin(\phi)\ddot{u} - \ell\dot{\phi}^2\\ \sin(\phi)g &= \cos(\phi)\ddot{u} + \ell\ddot{\phi}.\end{aligned}$$

The first of these two is of no importance to us (it determines λ, so the compression force of the stick). The second is the model we are were looking for.

A.6 Canonical Form (Formula (3.29))

We know that the desired T transforms the matrix A to

$$T^{-1}AT = A_z := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & & \vdots\\ \vdots & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\\ -p_0 & -p_1 & \cdots & \cdots & -p_{n-1}\end{bmatrix}.$$

For $S = T^{-1}$, this becomes

$$SA = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & & \vdots\\ \vdots & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\\ -p_0 & -p_1 & \cdots & \cdots & -p_{n-1}\end{bmatrix} S.$$

We write the rows of S as $s_1, \dots, s_n$. The first row of the equation above says that $s_1 A = s_2$; the second says that $s_2 A = s_3$; etc. In other words,

$$S = \begin{bmatrix} s_1 \\ s_1 A \\ s_1 A^2 \\ \vdots \\ s_1 A^{n-1} \end{bmatrix}.$$

Of the controllability matrix $\mathscr{C}_z$, we only use that it has the following structure (verify this yourself):

$$\mathscr{C}_z = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \iddots & \iddots & * \\ 0 & \iddots & \iddots & \vdots \\ 1 & * & \cdots & * \end{bmatrix}.$$

Since $S = \mathscr{C}_z \mathscr{C}_x^{-1}$, we see that the first row s_1 is equal to $\begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \mathscr{C}_x^{-1}$. Hence $s_1 = \eta$, as defined in (3.29).

The T from Thm. 3.5.4 can be determined in a similar way. We have $AT = TA_z$. We denote the columns of T by $T_1, T_2, \dots, T_n$, and for A_z we use the companion matrix. Then $AT = TA_z$ becomes

$$A \begin{bmatrix} T_1 & T_2 & \cdots & T_n \end{bmatrix} = \begin{bmatrix} T_1 & T_2 & \cdots & T_n \end{bmatrix} \begin{bmatrix} 0 & \cdots & \cdots & 0 & -p_0 \\ 1 & \ddots & & \vdots & -p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -p_{n-1} \end{bmatrix}.$$

The first column of this equation says that $AT_1 = T_2$; the second says that $AT_2 = T_3$; etc. Hence

$$T = \begin{bmatrix} T_1 & AT_1 & A^2 T_1 & \cdots & A^{n-1} T_1 \end{bmatrix}.$$

We take T_1 from the observability matrix $\mathscr{O}_z$. It has the following structure (verify):

$$\mathscr{O}_z = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \iddots & \iddots & * \\ 0 & \iddots & \iddots & \vdots \\ 1 & * & \cdots & * \end{bmatrix}$$

Since $T = \mathscr{O}_x^{-1} \mathscr{O}_z$, the first column T_1 of T must be

$$T_1 = \mathscr{O}_x^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

This is the η from Lemma 3.5.4.

A.7 Heymann's Lemma (Thm. 4.2.3)

The proof of Thm. 4.2.3 uses Heymann's lemma.

Lemma A.7.1 (Heymann's lemma). *If (A,B) is controllable, then for every $u_0 \in \mathbb{R}^{n_u}$ with $b := Bu_0 \neq 0$, there exists an $\tilde{F}$ such that $(A - B\tilde{F}, b)$ is controllable.*

Proof. This is a technical proof. We first show that for $x_0 = 0$ and some suitable input $u_0, u_1, \dots, u_{k-1}$, the *discrete*-time system $x_{k+1} = Ax_k + Bu_k$ produces a series of states $x_1, x_2, \dots, x_k$ that are linearly independent for all $k = 1, \dots, n$. We prove this using induction on k: for $k = 1$, the result obviously holds because $x_1 = Bu_0 \neq 0$. Now, suppose that $x_1, x_2, \dots, x_k$, for $k < n$, are linearly independent. Then there exists a u_k such that $x_{k+1} = Ax_k + Bu_k \notin \operatorname{span}\{x_1, x_2, \dots, x_k\}$. Indeed, suppose that this is not the case, so that

$$x_{k+1} := Ax_k + Bu_k \in \underbrace{\operatorname{span}\{x_1, x_2, \dots, x_k\}}_{\mathscr{L}_k :=} \quad \forall u_k.$$

This implies that $Ax_k \in \mathscr{L}_k$ (take $u_k = 0$) and therefore also that $\operatorname{im}(B) \subseteq \mathscr{L}_k$. But this then implies that $Ax_j = x_{j+1} - Bu_j \in \mathscr{L}_k$ for all $j \leq k$. This means that

$$A\mathscr{L}_k \subseteq \mathscr{L}_k.$$

Together with the inclusion $\operatorname{im}(B) \subseteq \mathscr{L}_k$, this gives

$$\operatorname{im}\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \subseteq \mathscr{L}_k.$$

Because of the controllability, this says $\mathbb{R}^n \subseteq \mathscr{L}_k$, or $\mathscr{L}_k = \mathbb{R}^n$. This implies that $k \geq n$, which is a contradiction. So for all $k \leq n$ there *does* exist a u_k such that $x_1, x_2, \dots, x_k$ are linearly independent.

Given such u_i, x_i and an arbitrary u_n define F_0 as

$$F_0 = -\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{-1}.$$

By construction it satisfies $-F_0 x_i = u_i$. We then have

$$x_{k+1} := Ax_k + Bu_k = (A - BF_0)x_k,$$

or $x_{k+1} = (A - BF_0)^k x_1 = (A - BF_0)^k b$. The controllability matrix of $(A - BF_0, b)$ is equal to the *invertible* matrix $\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$, and therefore the pair $(A - BF_0, b)$ is controllable. ∎

A.8 Linear Quadratic Control & and a note on Kalman Filtering

Ackermann's formula assumes that the input u has one entry only. If u has more than one entry then stabilizing inputs $u = -Fx$ are harder to find. One very popular and elegant such method is the *Linear Quadratic Regulator (LQR)* problem.

The LQR problem in its simplest form is to minimize a sum of squared states and inputs

$$\int_0^\infty x_1^2(t) + \cdots + x_n^2(t) + u_1^2(t) + \cdots + u_{n_u}^2(t)\ \mathrm{d}t$$

for a given system $\dot{x} = Ax + Bu$ and given initial state $x(0)$ over all inputs $u = (u_1, \ldots, u_{n_u})$. Minimization of this integral means that we want all our states to be "small" and our inputs "small" as well. The famous result is that the optimal input is a state feedback $u = -Fx$ and that this feedback makes the closed loop asymptotically stable if the system is stabilizable.

One can choose different weights: given positive definite matrices[2] $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{n_u \times n_u}$ the LQR problem is to minimize

$$\int_0^\infty x^{\mathrm{T}}(t)Qx(t) + u^{\mathrm{T}}(t)Ru(t)\ \mathrm{d}t$$

for a given system $\dot{x} = Ax + Bu$ and given initial state $x(0)$ over all inputs. The optimal solution $u = -Fx$ can be computed in MATLAB as follows:

```
A=[1 1; 1 -2];  % some A matrix
B=[1;2];        % some B matrix
Q=eye(2);       % some positive definite matrix
R=1/2;          % some positive definite matrix
F=lqr(A,B,Q,R); % SOLVE THE LQR PROBLEM
eig(A-B*F)      % check: asymptotically stable
```

In practical applications one often fixes the Q-matrix (to, say, $Q = I$) and the R-matrix to $R = \rho I$ which still depends on a single parameter ρ. By trying different ρ's the hope is to achieve a good compromise between fast convergence of the state (i.e. a "fast" closed loop system; this often requires large inputs) and small magnitude of the inputs (which typically makes the closed loop slow).

By transposing the entire system, the LQR method can be used to design observers for systems with more than one output y:

```
A=[1 1; 1 -2];    % some A matrix
B=[1;2];          % some B matrix
Q=eye(2);         % some positive definite matrix
R=1/2;            % some positive definite matrix
L=lqr(A',C',Q,R); % TRANSPOSE BOTH A AND C
L=L'
eig(A-L*C)        % check: asymptotically stable
```

In practical applications one often fixes the Q-matrix (to, say, $Q = I$) and the R-matrix to $R = \sigma I$ which still depends on a single parameter σ. By trying different σ's the hope is to achieve a good compromise between fast convergence of the observer state to the plant state (i.e. "fast" observer poles `eig(A-L*C)`; this often

[2] A symmetric matrix Q is positive definite if $x^{\mathrm{T}}Qx > 0$ for all nonzero $x \in \mathbb{R}^n$.

requires large L) and effect of measurement noise (on y) on the quality of the estimated state (which often requires a "slow" observer, i.e. "small" L). The so constructed L also has an elegant optimality interpretation but it would require too much to include it here. The observer constructed this way is an example of the famous *Kalman filter*.

A.9 Two Water Tanks (Example 5.2.8)

Consider the system

$$\dot{x} = \begin{bmatrix} -a_1 & 0 \\ a_1 & -a_2 \end{bmatrix} x + \begin{bmatrix} \beta \\ 1-\beta \end{bmatrix} u,$$
$$y = \begin{bmatrix} 0 & a_2 \end{bmatrix} x,$$

where a_1, a_2 are positive numbers and $\beta \in [0,1]$. Why is $h(t) \geq 0$? A mathematical answer is: Take $u(t) = \delta(t) \geq 0$ and note that $u(t) \geq 0$ for all t. If $x_1(t_0) = 0$, then at that time, x_1 can only increase because $\dot{x}_1(t_0) = \beta u(t_0) \geq 0$. Since $x(t_0) = 0$, we have $x_1(t) \geq 0$ for all $t \geq t_0$. For the same reason, if $\dot{x}_2(t_0) = 0$, then at that time, x_2 can only increase because $\dot{x}_2(t_0) = a_1 x(t_0) + (1-\beta)u(t_0) \geq 0$. Given $x_2(0) = 0$, we thus see that $x_2(t)$ also remains greater than or equal to zero for all $t > 0$. Hence we also have $h(t) = y(t) = a_2 x_2(t) \geq 0$ for all $t \geq 0$.

A "physical proof" goes as follows: This is the model for the two water tanks in Figure A.1. Here u is the mass that flows into the two tanks per unit of time, β is the fraction that flows into tank 1 (and so $1-\beta$ is the fraction that flows into tank 2), x_1 is the mass in tank 1. From tank 1, $a_1 x_1$ (mass per unit of time) flows into tank 2. So the net mass flow in tank 1 is $\dot{x}_1 = -a_1 x_1 + \beta u$. In tank 2, $a_1 x_1 + (1-\beta)u$ flows in and $y = a_2 x_2$ flows out (mass per unit of time). Hence $\dot{x}_2 = a_1 x_1 - a_2 x_2 + (1-\beta)u$. The impulse response is by definition $y(t)$ if $u(t) = \delta(t)$ and if everything is zero (empty tanks) for $t < 0$. It is clear for physical reasons that $h(t) \geq 0$.

(The equality $\|\mathscr{H}\|_1 = 1$ can be viewed as conservation of mass (see Exercise 5.11).)

A.10 Model of a Car (Example 5.8.2)

This model comes from [3, § 11.2]. The simple model (5.17) of a car can be explained as follows. By Newton's law,

$$m\dot{y}(t) = F_{\text{total}}(t), \tag{A.6}$$

where m is the mass of the car and F_{total} is the sum of the forces exerted on the car in forward direction. This sum equals

$$F_{\text{total}}(t) = -\rho y^2(t) + cu(t). \tag{A.7}$$

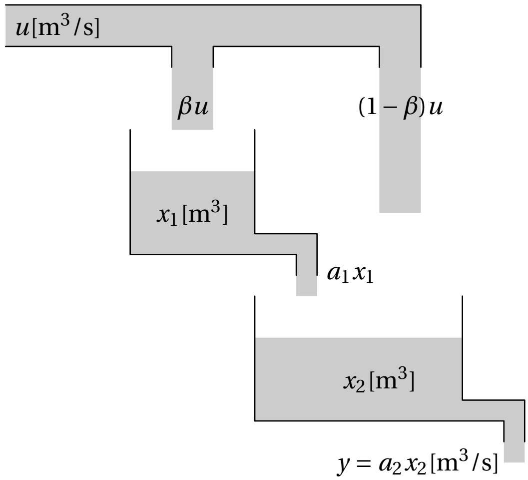


FIGURE A.1: Two water tanks.

The term $cu(t)$ is the propulsion force of the engine, and for simplicity we assume that it is proportional to the throttle opening. The other term, $-\rho y^2$, represents a nonlinear friction term (provided $y(t) > 0$). If the throttle is fully open, $u(t) = 1$, then the car will end up reaching a maximum constant speed. This maximum speed $y_{\max}$ satisfies $0 = -\rho y_{\max}^2 + c$. So $y_{\max} = \sqrt{c/\rho}$. If we now define

$$\tilde{y} := \frac{y}{y_{\max}},$$

then $m\dot{y}(t) = -\rho y^2(t) + cu(t)$ normalizes to

$$T\dot{\tilde{y}}(t) = -\tilde{y}^2(t) + u(t), \qquad T = \frac{m}{\sqrt{\rho c}}. \tag{A.8}$$

A typical value is $T = 10[\mathrm{s}]$. We linearize (A.8) around a constant throttle opening $u = u_* \in [0,1]$ and resulting stationary speed $\tilde{y}_*$. The stationary speed satisfies $0 = -\tilde{y}_*^2 - u_*$, that is, $\tilde{y}_* = \sqrt{u_*}$. This gives the linearization

$$T\dot{\delta}_y(t) = -2\tilde{y}_*\delta_y(t) + \delta_u(t). \tag{A.9}$$

In other words,

$$\dot{\delta}_y = -\frac{1}{\theta}\delta_y + \frac{1}{T}\delta_u, \tag{A.10}$$

with

$$\theta = \frac{T}{2y_*} > 0. \tag{A.11}$$

Equation (A.10) is of the form (5.17). The time constant θ strongly depends on the choice of u_*.

A.11 Some MATLAB Scripts

Matrix exponential. In MATLAB, we can determine the matrix exponential e^A for a matrix $A \in \mathbb{R}^{n\times n}$ using `expm`. For example, if $A = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$, we use

```
A=[1 2; 3 4];
expm(A)
```

We can obtain an eigendecomposition as follows:

```
A=[-3 5; -4 6];
[T,D]=eig(A);
T*D/T           % should be A
```

MATLAB can also (to a degree) do symbolic manipulations. For example,

```
syms t          % t is now a variable
A=[3 -4; -4 -3];
expm(A*t)       % e^{At} as a function of t
```

Simulation of LTI systems. For the simulation of linear time-invariant systems of the form (2.4), we can use `lsim`. For example, Figure 2.3 (Example 2.1.1) can be generated as follows:

```
k=2;                  % spring constant
r=1/2;                % damping coefficient
m=1;                  % mass

A=[0 1; -k/m -r/m];   % state data
B=[0; 1/m];
C=[1 0];
D=0;
x0=[1; 0];            % initial state

h=0.1;                % step size
t=0:h:40;             % time
u=stepfun(t,20);      % u(t)=1 from t=20 on

sys=ss(A,B,C,D);      % define the system
y=lsim(sys,u,t,x0);   % do the simulation
plot(t,y,t,u);
```

Simulation of nonlinear time-varying systems. For the simulation of more general nonlinear systems, SIMULINK can be useful, but we can also do this with MAT-

LAB. The pendulum (copied from (2.29), page 59) is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -\sin(x_1) \end{bmatrix}$$

It can be simulated as follows. First, we must make a file that gives $\dot{x}$ as a function of t and $x(t)$, for example

```
function xdot=fff(t,x)
xdot=[x(2); -sin(x(1))];
```

We save this file under the name `fff.m`. We can then carry out the simulation:

```
x0=[1;0];                   % initial state
tspan=[0,30];               % time interval
[t,x]=ode45(@fff,tspan,x0);
plot(t,x(:,1));             % show x1
```

In this example there is no input.

Simulation of the controlled juggler. In Chapter 4 we constructed a stabilizing controller for the juggler. To simulate the closed-loop behavior we can do:

```
n=2;                 % number of states of system %%
g=10;                % gravitational acceleration %%
l=0.5;               % whatever length of pendulum %

A=[0 1; g/l 0];      % A matrix for x=[q;v] %%%%%%%
B=[0; -g/l];         % B matrix for u=position hand%
C=[1 0];             % y=q %%%%%%%%%%%%%%%%%%%%%%%%%
F=[-1-l/g -2*l/g];   % so A-BF has eigenv -1 (twice)
L=[4; 4+g/l];        % so A-LC has eigenv -2 (twice)

eigABF=eig(A-B*F)    % check the stability of A-BF %
eigALC=eig(A-L*C)    % and of A-LC %%%%%%%%%%%%%%%%

% x'=f(x,u) for this f (both nonlinear & linearized)
f =@(x,u) [x(2); +g/l*sin(x(1))]+[0; -g/l*u]; % nonl
fl=@(x,u) [A*x+B*u]; % linearized %%%%%%%%%%%%%%%%%%

% Let z=[x;xhat] be the combined state (4 entries) %
% The nonlinear closed loop system then is z'=F(z) %
FWX=@(t,z) [f(z(1:n),-F*z(n+(1:n)));
            L*C*z(1:n)+(A-L*C-B*F)*z(n+(1:n))];
% The linearized closed loop system has this F: %%%%
FLW=@(t,z) [fl(z(1:n),-F*z(n+(1:n)));
            L*C*z(1:n)+(A-L*C-B*F)*z(n+(1:n))];

% Let z0=[q;v;qhat;vhat] at time 0: %%%%%%%%%%%%%%%%
```

```
%z0=[0;0.2;0;0];      % okay %%%%%%%%%%%%%%%%%%%%%%%%
%z0=[.3;.3;0;0];      % too big (for nonlinear case)%
z0=[0;0.3;0;0];       % scary   (for nonlinear case)%
tspan =[0 20];
[tt,z]=ode45(FWX,tspan,z0);  % simulate nonlinear %%
%[tt,z]=ode45(FLW,tspan,z0); % simulate linear %%%%%

% Plot all four state components against time: %%%%%
figure(1)
plot(tt,z');
legend('q','v','qh','vh')
grid
set(gca,'Fontsize',14);
```

The plot is shown in Fig. A.2. The estimates $(\hat{q}, \hat{v})$ seem to converges quite fast to (q, v) and all of them converge to zero exponentially fast.

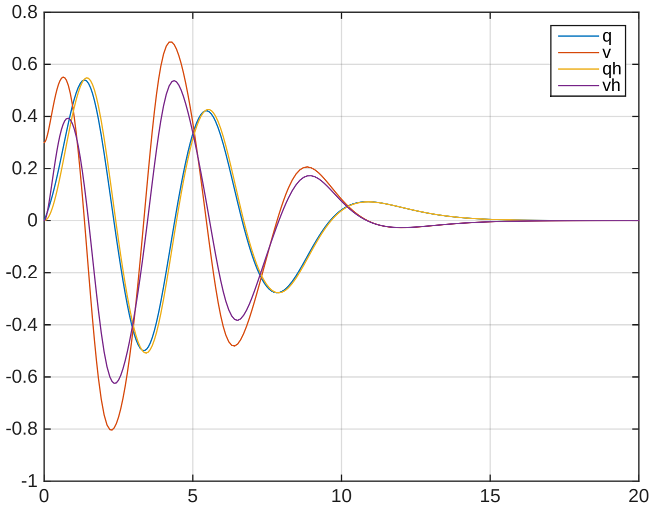


FIGURE A.2: Actual state $x = (q, v)$ of the juggler, and estimated state $\hat{x} = (\hat{q}, \hat{v})$ obtained by a dynamic controller that stabilizes the closed loop system. See the code on the previous page.

If input is available as samples only. What to do if only samples $u(kh), k \in \mathbb{N}$ of the input are available? Using `ode23` or `ode45` can then be tricky. In view of the inaccuracy in our knowledge of the input (we only know samples), a manual Runge–Kutta method (with fixed step size h) is in general sufficient. Define the function `nsim.m` as follows:

```
function x=nsim(ff,u,t,x0)
%NSIM  solve x'(t)=f(x(t),u(t),t) for sampled inputs
%
% SYNTAX: x=nsim(@fff,u,t,x0)
% fff is a function, xdot=fff(x,u,t)
% u and t are the input and time vectors
% x0 is the vector of the initial conditions
% x is a length(x0)*length(t) matrix

x=x0*t(:)'*0; % initialize x. This can...
x(:,1)=x0;    % ... take up much memory space
for k=2:length(t)
  h=t(k)-t(k-1);
  um=(u(k)+u(k-1))/2;
  x1=ff(x(:,k-1),u(k-1),t(k-1));
  x2=ff(x(:,k-1)+h/2*x1,um,t(k-1)+h/2);
  x3=ff(x(:,k-1)+h/2*x2,um,t(k-1)+h/2);
  x4=ff(x(:,k-1)+h*x3,um,t(k));
  x(:,k)=x(:,k-1)+h/6*(x1+2*x2+2*x3+x4);
end
```

To simulate the system

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

for a given input with samples, say, $u(k0.01) = \cos(k0.01)$, make a file `ffu.m` with code

```
function xdot=ffu(x,u,t)
xdot=[x(2); -x(1)-x(2) +u./(1+t.*t)]
```

and then do

```
t=0:.01:20;          % time interval
u=cos(t);            % the input samples
x0=[1;0];            % initial state
x=nsim(@ffu,u,t,x0);
plot(t,x(1,:));      % show x_1
```

Transfer functions and step responses. With the command

```
s=tf('s')
```

we tell MATLAB that s is the "Laplace variable". This allows us to define and simulate systems very efficiently. To define and simulate the walking juggler of Example 5.8.5 (page 189), the following suffices:

```
P=-2/(s^2-2);                  % juggler P_{y/u}(s)
K=(-4*s^2-6*s-1/2)/s/(s+4);
Hy=feedback(P*K,1);            % H_{y/r}(s)
Hu=feedback(K,P);              % H_{u/r}(s)
step(Hy,Hu);                   % responses y and u
                               % for r(t) = 1(t)
```

A.11.1 GNU Octave (an Alternative to Matlab)

GNU OCTAVE has virtually the same functionality and syntax as MATLAB, but is free. OCTAVE does not have commercial tool boxes, such as SIMULINK, although there are reasonable alternatives for some tool boxes.

For example, simulating $\dot{x}(t) = -2tx(t)$ goes as follows:

```
f=@(x,t) -2*t*x;  % define f(x,t) = -2tx
x0=1;             % initial condition
tspan=[0 3];
[t,x]=ode45(f,tspan,x0);
plot(t,x);
```

The MATLAB code is identical.

Bibliography

[1] K.J. Åström and R.M. Murray. *Feedback Systems: an introduction for scientists and engineers.* Princeton University Press, 2011.

[2] R.C. Dorf and R.H. Bishop. *Modern Control Systems.* Prentice Hall, 12 edition, 2010.

[3] H. Kwakernaak and R. Sivan. *Modern Signals and Systems.* Prentice Hall, 1990.

Index