Theory of Linear Systems
Lecture Notes

Luigi Palopoli and Daniele Fontanelli

Dipartimento di Ingegneria e Scienza dell'Informazione

Università di Trento

Contents

List of symbols and notations

- $\mathbb{R}$: set of real numbers
- $\mathbb{N}$: set of natural numbers
- $\mathbb{C}$: set of complex numbers
- $\mathcal{Z}(:)$ set of integer numbers
- $\mathcal{T}$: time space
- $\mathcal{U}$: class of input signals
- $\mathcal{Y}$: class of output signal
- $\mathcal{S}$: system
- $\mathbf{1}(:)$ step function defined as

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

- $\mathcal{L}(.)$: Laplace transform
- $\mathcal{L}^{-1}(.)$: inverse Laplace transform
- $\mathbf{Real}(z)$: real part of a complex number
- $\mathbf{Imag}(z)$: imaginary part of a complex number
- $|z|$: modulus of a complex number
- $\angle z$: phase of a complex number
- $\overline{z}$: complex conjugate.

CHAPTER 1

Introduction

This lecture notes are related to the course of "Theory of Linear Systems" taught in the undergraduate class of Telecommunications and Computing Engineering at the University of Trento. The presentation of Linear Systems proposed here is necessarily synthetic and introductory. The reader intested in additional details is referred to one of the books that exist on the the topic. Two excellent examples are offered by the book written by Joa Hespanha on Linear Systems [**Hes09**] or by the book writtern by Philipps et al. [**PPR95**] on linear systems.

CHAPTER 2

Definition of system

This course revolves around the notion of timed systems. For our purposes, a system is a physical or artificial entity (e.g., a computer program), where a number of meaningful quantities evolve in a way that can be described using a mathematical formalism. The evolution of our quantities of interest is described by a mathematical function that associates a value of a variable defined "time" with a value taken by the quantity which is defined in an appropriate set (e.g., voltage, position, velocity etc.). Such functions will be defined "signals".

In some cases we will have a convenience in defining a system as a relation between input signals and output signals. In other cases, a system is best thought as a relation between the evolution of some quantities. An example is offered by an economic system or by a human body. In this case, we could have a convenience in identifying input and output quantities or not (as an example, the system could evolve autonomously and have no input).

Our illustration of this notion is best given using examples.

EXAMPLE 2.1. Consider the RC circuit shown in Figure 2.1. In this system, we can easily identify several quantities of interest, for which it is possible to take measurements. The most evident ones are currents and voltages across the resistor and the capacitor. The laws of physics allow us to describe the evolution of the different quantities in time.

In this example, by time we mean exactly the "physical time", which flows with continuity and allows us to observe the evolution of physical phenomena.

Let us move to a completely different example.

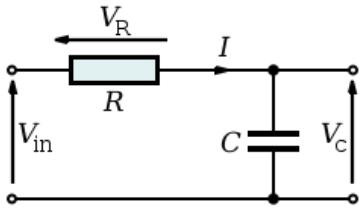


FIGURE 1. A simple RC network

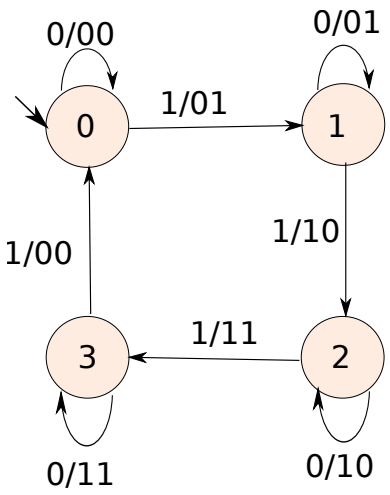


FIGURE 2. State machine implementing a modulo 4 counter

EXAMPLE 2.2. Consider a digital circuit that counts the number of times that a certain event occurs (e.g., input = 1) modulo 4 and that prints this number as an output on the occurrence of any event. Recalling basic notions of digital circuits, it is easy to see that a circuit like this can be modelled as the state machine shown in Figure 2.2. What we have just shown in the Figure is commonly referred to as a "state transition diagram", which is a graph where circles represent states and arcs are associated to transitions. A transition is labelled by an input/output pair, meaning that the transition is taken for the specified input and results in the emission of the specified output. In our example, if the system receives a 1 as an input when in state 0, it moves into state 1 producing as output 01. The label associated with the transitions is therefore "1/01", where the first number represents the input that triggers the transition and in the second number represents the output generated when the transition is taken. In this example, we are implicitly using a binary representation for numbers.

The system in the example could be implemented by a sequential electronic circuit or by a software programme and is different from the analogue circuit discussed in Example 2.1 in many respects. The most important one is that we are interested in describing the input/output relation (i.e., what each input sequence produces) without being necessarily interested in the point in time where each "reaction" of the machine takes place. The number of events that the system can respond to is clearly infinite, but given the amount of time that the system needs to produce a reaction we cannot generate events continuously in time. One possible way to formalise this is that the number of events that intervene between any two events [**LSV98**] is always finite.

As known from elementary courses of computer systems, it is possible to reduce our freedom and require that the system reacts only at precise points in time. Indeed, most sequential networks are implemented in a "synchronous" way, with reaction to events forced to take place upon the rising or falling edge of a clock signal[1]. In this case, the minimum time interval between two different reactions is bound to the clock period. However legitimate and reasonable, this is just one of the possible implementation of a state machine. In a different asynchronous implementations, the machine reacts to an event whenever it occurs. We could say that in Figure 2.2 we have offered an abstract specification of the state machine, which is open to different implementation options for the designer.

There are cases where the system is inherently synchronous. Consider the following example.

Example 2.3. Consider a banking account where losses or gains are capitalised every T time units (where T is a sub-multiple of one year). The customer is allowed to deposit or withdraw money every time an interval expires. If at the beginning of an interval the capital is positive, at the end the credit grows with an annual interest rate I_+. If it is negative, at the end the debt grows with an annual interest rate I_-.

Let $C(kT)$ be the capital at the beginning of the $k-$th observation interval and let $C((k+1)T)$ be the capital at the end. Finally, let $S(kT)$ be the amount of money (positive or negative) that the customer deposits or withdraws. The evolution of the capital is given by:

(2.1)
$$C((k+1)T) = \begin{cases} \left(1 + I_+ \frac{T}{12}\right)(C(kT) + S(kT)) & \text{Se } C(kT) + S(kT) \geq 0 \\ \left(1 + I_- \frac{T}{12}\right)(C(kT) + S(kT)) & \text{Se } C(kT) + S(kT) < 0 \end{cases}$$

At a fist sight, this system looks very similar to the state machine described earlier. Once again, it is important to know the exact sequence of deposits and withdrawals. Therefore, it is required that the time space $\mathcal{T}$ be an ordered set. However, it is not enough. This time, we *need* to assume that the time between intervals between events are of the same duration and we need to know how large this interval is (otherwise the computation of the interest is not possible). Contrary to the modulo counter in Example 2.2, the banking account example is not only synchronous in one of its possible implementation, but it is inherently so since its evolution is tied to the evolution of physical time. However, contrary to the RC network in Example 2.1, it is not required (nor possible) to have arbitrarily close events.

2.1. Signals

A common presence in all the systems mentioned above are "signals". A signal is simply defined as a function from a time space $\mathcal{T}$ to a set $\mathbb{U}$. At

[1]When no event occurs on a clock, the network does not react

this point, we need not make any particular assumption on the set $\mathbb{U}$ (we will have to later on).

We will frequently use calligraphic letter to denote classes of signals. For example,

$$\mathcal{U} = \{u(\cdot) : \mathcal{T} \to \mathbb{U}\}.$$

2.1.1. The notion of time. A first problem to face is to define the exact meaning for the notion of *time*, a set that we will henceforth denote by $\mathcal{T}$.

2.1.1.1. *Continuous Time.* Continuous time (CT) signals are based on a time space that has to be:

(1) totally ordered,
(2) metric,
(3) a continuum (in the mathematical sense).

The exact meaning of these concepts is recalled in Appendix A. At this point, it is more useful to elaborate on the intuition underneath and on why we need them to describe CT signals (and systems).

Let us start with an example.

EXAMPLE 2.4. Consider the RC circuit in Example 2.1, $V_R(t)$ can be used to denote the voltage drop across the resistor R as a function of time and is a CT signal. The same applies to the voltage $V_C(t)$ measured over the capacitor C, for $V_{in}(t)$ and for $I(t)$:

CT signals are used to represent the evolution of physical quantities. In this setting, if a quantity takes on two different values at different time it is relevant to know which one comes first and which one follows, and how far away the two events are. By the adjective "relevant" we mean that the evolution of the system will be quite different if the order of two events is inverted or if their distance is changed. To understand this, consider an electrical RC circuit as in Example 2.1. Suppose that a 5V power source is connected to the terminals at time t_1 and switched off at time t_2. The evolution of the system is understandably different if $t_2 - t_1 = 10^{-3}$s or $t_2 - t_1 = 10^5$s. Moreover, the value of a quantity like V_r can be measured at any time t and is potentially affected by the evolution of a different quantity at all possible times $t' \in \mathcal{T}$. These considerations explain why we need a set that is a continuum to mode CT signals. Therefore, the most obvious choice for the time space is to choose it as the real set: $\mathcal{T} = \mathbb{R}$.

2.1.1.2. *Discrete Events.* The opposite case to CT signals is when the connection between the time space $\mathcal{T}$and the physical time is very shallow. Consider the system shown in Example 2.2. If we are only interested in the evolution of the output sequence given the input sequence, all we need to know on the timing of the event is condensed in their order. For instance, the input sequence 011 will certainly produce a different sequence of states and of output than the sequence 101 which only differs from the first one for the order of the first two symbols.

An important point is that an input sequence (e.g., $0 \cdot 1 \cdot 1$) will produce the same output sequence whatever the spacing between any two events (be it one millisecond or one century). So if our goal is to specify the output sequence given an input sequence, we do not need continuity or metric structure for our time space $\mathcal{T}$.

EXAMPLE 2.5. For the modulo counter in Example 2.2, a sequence of input such as $0 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot \ldots$ is an example of a DE signal. So is the sequence of output $00 \cdot 01 \cdot 01 \cdot 10 \cdot \ldots$.

Discrete event (DE) signals are defined over time spaces that are totally ordered but are not required to be a continuum and to be metric. Some authors [**LSV98**] suggest that a in order to define a DE signal the time space has to satisfy the additional property that between any two events we can have a finite number of intervening events. This requirement is formally captured by imposing that events are *order-isomorphic* to the natural numbers. This subtle but important point is beyond the scope of these notes.

For our purposes, we can simply say that a DE signal consists of a sequence of totally ordered events. Each of them can be associated with an increasing natural number that reflects its position in the sequence. So, in the sequence 011, the first 0 will be associated with the event 0, the first 1 with event 1 and the third 1 with event 3.

2.1.1.3. *Discrete Time.* As mentioned above, DE signals are ordered sequences of events each one associated with a time instant. For an important subclass of these signals, events have to be *synchronous*, meaning that they occur/are registered at specified time instants. This means that the time variable the signal is defined over can take on only specific values. A typical (but not mandatory choice) is that such instants be multiple of a specified period T.

Generally speaking, we define a discrete–time (DT) signal a DE signal where events are constrained to be synchronous.

EXAMPLE 2.6. In the banking account described in Example 2.3, the evolution of the capital $C(kT)$ and the sum $S(kT)$ deposited or withdrawn are examples of DT signal.

As in the case of DE signals, we can order the different time instants using the natural numbers. But, for DT signals the algebraic structure of the time space (i.e., its being an abelian group) is used in full for sums and differences. hence, we will use the set $\mathcal{Z}$ (t)o represent DT systems.

2.1.2. Sampled Data. Finally, it is useful to mention the presence of a different type of systems, compounded of a collection of heterogeneous subsystems, each one associated with a different time space. A classic example is offered by sample data systems, which are DT systems obtained from CT systems restricting the points in time where certain quantities can be measured (sampled) or certain input variables be changed.

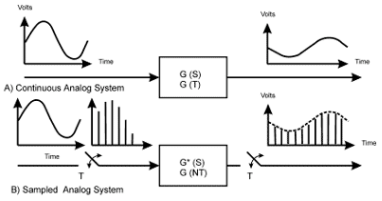


FIGURE 3. Example of a sampled data system

An example of this type is shown in Figure 2.1.2. A given physical quantity that evolves in continuous time is sampled at some points in time that are spaced by a fixed amount of time (sampling interval). The outcome is a sequence of numbers that is a DT system. We can have this signal processed through a DT system that generates a new DT signal (sequence of samples), which can in its turn be converted back into a CT signal through a digital to analogue converter. This is typically done by sampling the signal value at time kT and holding it constant throughout the interval $[kT,\, (k+1)T]$ (Zero Order Hold).

2.2. Systems

In this course, we will concentrate on systems defined on CT signals and DT signals. We start by giving an abstract definition general enough to encompass most systems of interest. Let $\mathcal{U}$ represent a class of input signals taking value in the set U and $\mathcal{Y}$ represent a class of output signals taking value in the set Y. We could define the system as a binary relation between $\mathcal{U}$ and $\mathcal{Y}$: $\mathcal{S} \subseteq \mathcal{U} \times \mathcal{Y}$.

We recall that a binary relation is a set of pairs. In this case one element of the pair is the input signal and the other one is the output signal. Importantly, the relation can associate more than one output signal to the same input.

We remark that this definition of system is "oriented", i.e., it assumes the distinction between an input (cause) and of an output (effect). In a more general setting, we could simply define it as a relation between tuples of signals, no-matter the role each one plays. This is a "speculative" viewpoint which is of interest for such areas as theoretical physics or biology. In this course, on the contrary, we will take an "engineering perspective" where a distinction is made between quantities that can be acted upon and other quantities that evolve as a result of these actions.

EXAMPLE 2.7. Consider the system in Example 2.1 and suppose that the input is given by $V_{in}(t)$ and the output is given by $V_c(t)$. Suppose we apply a *step input* $V_f \cdot \mathbf{1}(t)$ where the step $\mathbf{1}(t)$ is defined as

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

It can be seen (we will) that $V_C(t) = V_c(0)e^{-t/RC} + (1 - e^{-t/RC})V_f$. Therefore, depending on the initial value $V_c(0)$ we obtain a different output for the same input.

EXAMPLE 2.8. Consider the system in Example 2.3. Suppose that at time $t = 0$ we start with a capital C_0 and that we deposit a constant amount of money S at the beginning of every capitalisation period. It is possible to see that the capital at time kT is given by

$$c(kT) = (1 + I_+ \cdot \frac{T}{12})^k C_0 + S\frac{(1 + I_+ \cdot \frac{T}{12})^{k+1} - 1}{I_+}.$$

Depending on the initial capital, the evolution can be much different.

As we have just seen, we could have a different output associated to the same input both for the case of CT systems and of DT systems.

In the examples above, we also underscore the choice of an initial point in time to start the observation of the evolution of the system. Generally speaking, we can choose an instant t_0. Denote by $\mathcal{T}(t_0)$ the subset of the time space containing instants

$$\mathcal{T}(t_0) = \{t \in \mathcal{T} : t \geq t_0\}.$$

We can denote by $\mathcal{W}^{\mathcal{T}(t_0)}$ the set of functions defined over $\mathcal{T}(t_0)$ that take value in W:

$$\mathcal{W}^{\mathcal{T}(t_0)} = \{w_0(\cdot) : \forall t \geq t_0, t \to w_0(t) \in W\}.$$

By $w0|_{T(t_1)}$ we will denote the truncation of the function, which is evaluated from $t_1 > t_0$ onward.

Now we can formally define a system as follows:

DEFINITION 2.9. An abstract dynamic system is a 3-tuple $\{\mathcal{T}, \mathcal{U} \times \mathcal{Y}, \Sigma\}$ where

- $\mathcal{T}$ is the time space
- $\mathcal{U}$ is the set of input functions
- $\mathcal{Y}$ is the set of output functions

and

$$\Sigma = \left\{\Sigma(t_0) \subset \mathcal{U}^{T(t_0)} \times \mathcal{Y}^{T(t_0)} : t_0 \in \mathcal{T} \text{and CRT is satisfied}\right\},$$

where CRT stands for closure with respect to truncation: i.e., $\forall t_1 \geq t_0$

$$(u_0,\, y_0) \in \Sigma(t_0) \implies \left(u_0|_{\mathcal{T}(t_1)}, y_0|_{\mathcal{T}(t_1)}\right) \in \Sigma(t_1).$$

In plain words, the CRT property means that if a couple of functions belong to the system from t_0 onward, so will their truncation from t_1 onward.

2.2.1. Parametric representation. The binary relation $R = \mathcal{U} \times \mathcal{Y}$ is made of ordered pairs $(u,\, y)$, we denote by $D(R)$ the domain of the relation (i.e., $\mathcal{U}$) and by $R(R)$ its range (i.e., $\mathcal{Y}$).

A general result applicable to binary relations is the following [**RI79**]

LEMMA 2.10. *Given a binary relation R, it is possible to define a set P and a function* $\pi : P \times D(R) \to R(R)$ *such that*

$$(a,b) \in R \implies \exists p : b = \pi(p,a) \tag{2.2}$$

$$p \in P, a \in D(R) \implies (a, \pi(p,a)) \in R \tag{2.3}$$

π *is said* parametric representation *and* (P,π) *is said* parametrisation *of the relation.*

This essentially means that it is possible to represent a relation as the union of a set of function graphs parametrised by a suitable choice of parameters. The main interest of this result lies in the following:

THEOREM 2.11. *Consider a system defined as in Definition 2.9. It is possible to identify a parametrisation* (X_{t_0}, π) *such that*

$$\pi = \{\pi_{t_0} : X_{t_0} \times D(\Sigma(t_0)) \to R(\Sigma(t_0))/t_0 \in \mathcal{T}\} \tag{2.4}$$

satisfying the following properties:

$$(u_0, y_0) \in \Sigma(t_0) \implies \exists x_0 : y_0 = \pi_{t_0}(x_0, y_0) \tag{2.5}$$

$$x_0 \in X_{t_0}, u_0 \in D(\Sigma(t_0)) \implies (u_0, \pi_{t_0}(x_0, u_0)) \in \Sigma(t_0). \tag{2.6}$$

This result is easily understood looking at Example 2.7 and 2.8, where the $V_c(0)$ and C_0 are used as parameters.

2.2.2. Causality. We are now in condition to state the notion of causality.

DEFINITION 2.12. Let $u|_{[t_0,\bar{t}]}$ be the restriction of the function u to the closed interval $[t_0,\, \bar{t}]$. A system is causal if it has a representation (X_{t_0}, π) such that

$$\forall t_0 \in \mathcal{T}, \forall x_0 \in X_{t_0}, \forall \bar{t} \in \mathcal{T} \tag{2.7}$$

$$u_{[t_0,\,\bar{t}]} = u'_{[t_0,\,\bar{t}]} \implies [\pi_{t_0}(x_0,u)](\bar{t}) = [\pi_{t_0}(x_0,u')](\bar{t}). \tag{2.8}$$

A parametric representation of this type is said causal. If instead of the closed interval $[t_0,\, \bar{t}]$ we use the semi-open interval $[t_0,\, \bar{t})$, the parametric representation and the system is said strictly causal.

Observe that π is a functional, i.e., a function defined over space of functions. So $\pi_{t_0}(x_0, u)$ is the output function associated to the parameter x_0 and to the function u, and $[\pi_{t_0}(x_0,u)](\bar{t})$ is the value it takes at time $\bar{t}$. In plain words, the definition just introduced means that the values of u beyond $\bar{t}$ do not affect the value of the output at time $\bar{t}$. A simple way to put it is that a causal system does not foresee the future. If the system is strictly the output at time $\bar{t}$ is only affected by the input at time *strictly* smaller than $\bar{t}$.

2.2.3. Number of input and output signals. In the discussion above, we have not made specific assumptions on the range U of the input functions $\mathcal{U}$ and on the range Y of the output functions $\mathcal{Y}$. In our course, we will have cases where these quantities are scalars, and other cases in which they are vectors.

The first case is when both input and output functions are scalar. We call this type of systems Single Input Single Output (SISO) systems.

If the input is not a scalar function we will refer to the system as Multiple Input (MI); likewise, if the output is not scalar we will use the definition Multiple Output (MO). Clearly, we can have all different combinations: Single Input Single Output (SISO), Single Input Multiple Output (SIMO), Multiple Input Single Output (MISO), Multiple Input Multiple Output (MIMO).

As an example, the RC network in Example 2.1 is a SISO system because its input is given by the the input voltage V_{in} and the output V_c are both scalar functions. However, if our output functions are both V_{in} and I the system should be considered as SIMO. As for the banking account described in Example 2.3, the amount of money $s(kT)$ deposited or withdrawn is the scalar input and the capital $c(kT)$ is the scalar output. So the system is SISO. However, if the interest rates I_+ and I_- are allowed to change in time, the system will be MISO.

2.3. I/O Representation

In the previous sections, we have seen that a system is essentially a relation between spaces of functions defined over a common time space. In view of a general results any relation can be described as the union of the graph of a set of functions parametrised by a suitable set. This lead us to the parametric representation of system that applies to any type of system. We are now left with the problem of understanding how such a parametric description can be found.

We are now going to delve more deeply inside the subclass of continuous time ($\mathcal{T}= \mathbb{R}$) and of discrete time systems ($\mathcal{T}=\mathbb{Z}$).

One possible way to describe a system is by expressing a mathematical property that relates input to output. We will now turn our attention to systems that can be described by differential equations (CT systems) or difference equations (DT systems) of an appropriate order n:

$$F(y(t), \mathfrak{D}y(t), \mathfrak{D}^2y(t), \dots, \mathfrak{D}^n y(t), u(t), \mathfrak{D}u(t), \dots, \mathfrak{D}^p u(t), t) = 0, \tag{2.9}$$

where the operator $\mathfrak{D}$, when applied to a generic function f, is defined as:

$$\mathfrak{D}^k f = \begin{cases} \frac{d^k f}{dt^k} & \text{for CT systems} \\ y(t+k) & \text{for DT systems.} \end{cases} \tag{2.10}$$

For *well posed* systems, it is possible to write the equation in a normal form

$$\mathfrak{D}^n y(t) = F(y(t), \mathfrak{D}y(t), \mathfrak{D}^2y(t), \dots, \mathfrak{D}^n y(t), u, \mathfrak{D}u(t), \dots, \mathfrak{D}^p u(t), t), \tag{2.11}$$

EXAMPLE 2.13. Consider the differential equation:

$$\dot{v}(t) = -\gamma v(t) - g, \tag{2.12}$$

which could describe a rocket moving in the atmosphere under the action of gravity. Let us It is possible to solve the differential equation as follows:

$$\begin{aligned}
\frac{dv}{dt} &= -\gamma v - g \\
\frac{dv}{\gamma v + g} &= -dt \\
\int \frac{dx}{\gamma v + g} &= \int -t \\
\frac{1}{\gamma} \log |\gamma v + g| &= -t + c \\
\gamma v + g &= e^{-\gamma t + c\gamma} \\
v(t) &= -g/\gamma + Ae^{-\gamma t}
\end{aligned}$$

The last equation tells us that the solution is determined *modulo* a constant (A), which can be found imposing the initial condition. As we can see we have found a parametrisation of the system (although applicable only to constant input function).

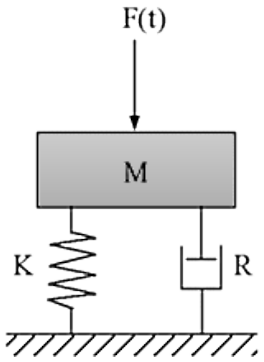


FIGURE 4. Example of mass-spring-damper system

EXAMPLE 2.14. Consider the mass-spring-damper system in Figure 4. Assume that the x-axis is perpendicular to the ground and points upward. The Newton equation leads us to following differential equation

$$m\ddot{x} = F(t) - mg - K(x - x_{rest}) - R\dot{x} \tag{2.13}$$

where the term $-K(x - x_{rest})$ accounts for the elastic reaction of the spring and $-R\dot{x}$ accounts for Coulomb friction. This is an instance of the more general expression in Equation (2.11). For this example, the solution of the equation is not so straightforward with standard means (although perfectly feasible). We will see generic ways to find a parametrisation of this system for any input function $F(t)$ (if they are analytically "nice" enough).

EXAMPLE 2.15. Let us go back to our differential equation

$$\dot{v}(t) = -\gamma v(t) + F(t), \tag{2.14}$$

which could describe a rocket moving in the atmosphere under the action of its thrusters ($F(t)$). Suppose we want to integrate the equation numerically using a fixed integration interval. We can integrate both sides of the equation

$$\int_{kT}^{(k+1)T} \dot{v}dt = \int_{kT}^{(k+1)T} (-\gamma v(t) + F(t))dt.$$

If the integration interval is small enough, we can assume that both $x(t)$ and $F(t)$ remain constant throughout the interval. We can also use the Euler approximation for the derivative:

$$\dot{v}(t) \approx \frac{v((k+1)T) - v(kT)}{T}.$$

Therefore the integral above can be written as:

$$v((k+1)T) - v(kT) \approx -\gamma v(kT)T + F(kT)T.$$

As we can see, we have obtained a DT time system in the normal form in (2.11). Given an initial velocity v_0, the equation above allows us to iteratively construct the parametrisation of the system.

In the examples before we have seen that CT systems are described by differential equations, DT systems are described by difference equations. We have also seen that for the latter that the difference equation is in fact an algorithmic way to obtain a parametrisation of the system. In this parametric representation we need to initialise the input u and the output y up to the for a number of steps equal than the maximum forward step with which they appear in the equation minus 1. For instance for the equation

$$y(k+2) = y(k+1) + y(k) + u(k+1) + u(k)$$

in order to start the algorithm, we need $y(0), y(1), u(0)$. We will also see how to derive explicit forms for this representation.

For CT systems the same result is obtained solving a differential equation (a task that is not always easy). A couple of very well known results (Peano and Lipschitz Theorems) specify conditions (e.g., Lipschitz continuity) under which a differential equation has a unique solution. This allows us to construct a parametrisation, where we have to consider the initial conditions for the input and output function and for their derivatives with an order lower than the maximum order with which they appear in the equation. For instance for the equation:

$$\ddot{y} = -\dot{y} + \ddot{u} + 3y - u$$

we will need $y(0), \dot{y}(0), u(0), \dot{u}(0)$. The complex discussion on existence and uniqueness of solutions will be left out of these notes, and we will assume that the differential equation is well behaved and that it admits a unique solutions for the input function of our interest.

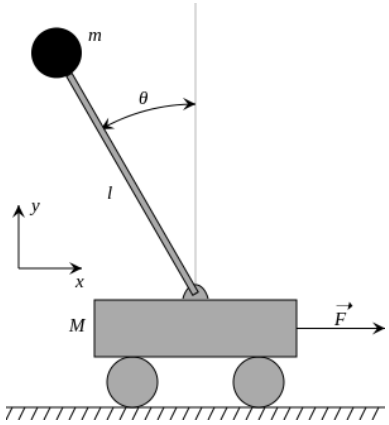


FIGURE 5. Inverted Pendulum

2.3.1. Causality. In order for the system to be causal, we have to have $n \geq p$. Indeed, if we consider a simple DT example like:

$$y(k+1) = y(k) - 3u(k) + 4u(k+2),$$

in which $n = 1$ and $p = 2$, we easily see that the output at a given time depends on the future value of the input.

Consider now an even simpler and apparently inoffensive example of CT system like

$$y(t) = \dot{u}.$$

In order to define a derivative, we have to compute the limit of the left and of the right increment as see that they are equal. But knowing the right increment is tantamount to peering into the future (although by an infinitesimal amount).

2.3.2. MIMO systems. The definition in Equation (2.11) refers to SISO systems. It is easy to generalise to a MIMO systems as follows:

(2.15)

$$\mathfrak{D}^n y_j(t) = F(y_1(t), \dots, \mathfrak{D}^n y_1(t), \dots, y_m(t), \dots, \mathfrak{D}^n y_m(t), u_1, \mathfrak{D}u_1(t), \dots, \dots, \mathfrak{D}^p u_1(t), \dots u_h, \mathfrak{D}u_m(t), \dots, \mathfrak{D}^p u_h(t), t), j = 1, \dots m$$

for a system with m output functions and h input functions.

EXAMPLE 2.16. Consider the inverted pendulum in Figure 2.16. Let M be the mass of the cart, m the mass of the pendulum, l the length of the pendulum, y_1 the position x of the centre of mass of the cart, y_2 the angle θ of the pendulum with the vertical line, and u be the force F applied to the cart. The equation of motion can be found as follows:

$$\ddot{y}_1 = \frac{ml}{M+m}\left(\ddot{y}_2 \cos y_2 - \dot{y}_2^2 \sin y_2\right) + \frac{1}{M+m}u$$
$$\ddot{y}_2 = \frac{1}{l}\left(g \sin y_2 + \ddot{y}_1 \cos y_2\right)$$

2.4. State Space representation

In the sections above we have seen that: 1. for a system the same input function can be associated with multiple output functions, 2. for each t_0, we can introduce a parametric representation such that, given an input function, the associated output function can be "disambiguated" using a parameter x_0, 3. for a class of systems (causal systems), this parametric representation is causal meaning that, for a given value of the parameter x_0, the output at any given time $t' > t_0$ only depends on the values taken by the input function in the interval $[t_0, t')$ (or $[t_0, t']$ if the causality is strict). Given that t_0 can be chosen freely, this means that: 1. a causal system cannot "foresee" the future, 2. the parameter x_0 summarises all the story of the system previous to t_0. These are qualitative arguments. In order to turn them into a consistent and mathematically sound definition of state we need a few more steps.

In particular, by changing the initial instant t_0 we will find a different value for the parameters x_0. Our first assumption is that whatever the choice of t_0 such parameters live in the same set X: $\forall t_0$, $x_0 \in X$. This is obviously the case if we adopt the IO representation of the system; as discussed in Section 2.3, if a CT system is described by means of a differential equations, and if this is mathematically well behaved (such as to admit a unique solution), the parameters are conveniently chosen as the set of all initial conditions for the derivatives of input and output function up to a certain order. The same reasoning can be applied to DT system where the parameters are the initial values. Therefore, whatever the initial time t_0, the parameters x_0 live in the space given by the initial conditions.

In order to have a complete identification between x_0 at time t_0 and the past history of the system (up to t_0) we need some form of consistency between the parameter x_0 at time t_0 and the parameter x_1 at time any other time t_1. We will assume the existence of a function ϕ, such that

$$x_1 = \phi(t_1, t_0, x_0, u),$$

where u is restricted to its values in the set $[t_0, t_1)$. Roughly speaking, ϕ tells us how to transform the story up until t_0 into the story up until t_1, given the knowledge of the input in between t_0 and t_1. Let $(T \times T)^*$ be defined as:

$$(T \times T)* = \{(t, t_0) : t, t_0 \in \mathcal{T} \text{ and } t \geq t_0\}.$$

The function ϕ is defined as

$$\begin{aligned} \phi : (T \times T) * \times X \times \mathcal{U} &\to X \\ x(t) &= \phi(t, t_0, x_0, u). \end{aligned} \tag{2.16}$$

We will require that it respects the following three properties:

Consistency:

$$\forall t \in T, \forall u \in \mathcal{U}, \phi(t, t, x, u) = x,$$

Causality:

$$\forall t, t_0 \in \mathcal{T}, \forall x_0 \in X \; u|_{[t_0,t)} = u'\big|_{[t_0,t)} \implies \phi(t, t_0, x_0, u) = \phi(t, t_0, x_0, u')$$

Separation:

$$\forall (t, t_0), \forall x_0 \in X, \forall u \in \mathcal{U} \tag{2.17}$$

$$t > t_1 > t_0 \implies \phi(t, t_0, x_0, u) = \phi(t, t_1, \phi(t_1, t_0, x_0, u), u). \tag{2.18}$$

The first two properties are obvious; the third one formalises the idea that the state captures the story of the system.

The notion of state we are introducing is rooted in that of a parametric representation introduced earlier. The same notion leads to the link between input state and output. Indeed, assuming a causal system, the function In Equation (2.4) leads us to:

$$\forall t_0, y_0(t) = \left[\pi_{t_0}(x_0, u_0)\right](t) \forall t \geq t_0,$$

where u_0 is restricted to $[t_0, t]$. Assuming $t_0 = t$ we can introduce the definition of a function η as follows:

$$y(t) = \left[\pi_t(x, u)\right](t) = \eta(t, x(t), u(t)).$$

The function η is described as

$$\begin{aligned} \eta : \mathcal{T} \times X \times U &\to Y \\ y(t) &= \eta(t, x(t), u(t)). \end{aligned} \tag{2.19}$$

2.4.1. Input signals, output signals and states. The variable $x(t)$ is called state of the system, and the set X it lives in is called state space. While systems with finite state space exist and are relevant (e.g., see Example 2.2), we will focus on systems with dense state space.

Generally speaking, a system has a vector of input signals, part of which are directly controllable and can be used as command variables to drive the system toward a desired behaviour. Other input signals are not under direct control of the user and act as external disturbance or noise.

As an example, if we model a car the throttle and the steering wheels are command variables, whereas the inclination of the road can be seen as an external disturbance.

Output signals are measurable quantities and can be used to formulate design specifications. For instance, in the model of a car, one can formulate a specification like *if the driver open the throttle completely, the speed reaches* $100km/h$ *in* 11 *seconds*, in which the speed plays the role of our controlled variable.

In other cases, an output signal conveys useful elements of information that allow reconstructing the system state when the latter is not directly measurable. For instance, the speed of a vehicle is typically a state variable that cannot be measured directly. It is possible to mount encoder on the wheels that measure the number of rotations n and from this derive an estimate of the velocity. The most obvious way is to compute a numeric

"derivative" using the left increments. This strategy typically produce noisy estimates. We will see in the course that better results can be obtained in systematic ways using the so called "observer", a system that builds on the knowledge of the system model and of the story of the output signal to reconstruct an estimate of the state.

2.4.2. Existence of a state. The functions ϕ and η that we have qualitatively introduced above allow us to generate a system:

$$\Sigma(t_0) = \Big\{ (u_0,\, y_0) \in \mathcal{U}^{\mathcal{T}(t_0)} \times \mathcal{Y}^{\mathcal{T}(t_0)}, \tag{2.20}$$
$$u_0 = u|_{\mathcal{T}(t_0)}, y_0 : y_0(t) = \eta(t,\ \phi(t, t_0, x_0, u), u(t)),\ \text{with } u_0 \in \mathcal{U} \text{ and } x_0 \in X \Big\}.$$

An important point is when is the converse true? In other words, when is it possible to construct a state representation as outlined above for generic abstract system? It turns out that this is indeed possible for an abstract system under mild conditions. Essentially, what is required is that input functions have the same range and that they be a closed set under concatenation. The existence and uniqueness of a state representation for abstract systems is beyond the scope of this course and we refer the interested reader to specialised text books [**RI79**] or to the scientific literature. In this course, we will always assume that a representation of this kind exists for our systems of interest.

2.4.3. Implicit Representation. The η and ϕ functions that we have introduced in Equation (2.16) and (2.19) compound the so-called *explicit* representation of a system. By "explicit" we mean that given an input function and an initial state the two function enable to compute immediately the evolution of the state and of the output.

In contrast, an *implicit* representation is one in which the computation the output is not immediate but it requires the solution of a system of difference or of differential equations.

For DT systems, an implicit representation can be found by by simply evaluating ϕ across a single step:

$$x(t+1) = \phi(t+1, t, x(t), u(t)) = f(t, x(t), u(t)) \tag{2.21}$$

The function $f(\ldots)$ is said generator function. The representation (X, f, η) is said implicit representation. Clearly the function ϕ can be recovered from f through a simple iteration.

Example 2.17. Consider a DT system with scalar state (i.e., state with dimension 1) and suppose that its generator function is given by the linear function:

$$x(t+1) = a(t)x(t) + b(t)u(t).$$

If we want to compute the evolution from a state x_0 at time t_0 to a state x_1 at time t_1, we can do it by unrolling the following iteration:

$$\begin{aligned}
x(t_0+1) &= a(t_0)x_0 + b(t_0)u(t_0)\\
x(t_0+2) &= a(t_0+1)x(t_0+1) + b(t_0+1)u(t_0+1) =\\
&= a(t_0+1)a(t_0)x_0 + a(t_0+1)b(t_0)u(t_0) + b(t_0+1)u(t_0+1)\\
x(t_0+3) &= a(t_0+2)x(t_0+2) + b(t_0+2)u(t_0+2) =\\
&= a(t_0+2)a(t_0+1)a(t_0)x_0 + a(t_0+2)a(t_0+1)b(t_0)u(t_0) +\\
&+ a(t_0+2)b(t_0+1)u(t_0+1) + b(t_0+2)u(t_0+2)\\
&\cdots\cdots
\end{aligned}$$

We can prove by induction for a general H that:

$$\begin{aligned}
x(t_0+H) &= \phi(t_0+H, t_0, x_0, u) =\\
&= \psi(t_0+H, t_0)x_0 + \sum_{t=t_0}^{t_0+H-1} G(t_0+H-1, t)u(t)
\end{aligned}$$

where

$$\psi(t,\tau) = \begin{cases} \prod_{t'=\tau}^{t-1} a(t') & \text{if } t > \tau \\ 1 & \text{if } t \le \tau \end{cases}$$

$$G(t,\tau) = \psi(t+1, \tau+1)b(\tau)$$

When we move to the CT domain, things are not so immediate. It is reasonable to think of the evolution of the state in real time as described by means of a differential equation. This requires a sufficient regularity (at the very least differentiability) of the function ϕ. If we admit that ϕ is the solution to the equation:

$$\frac{\partial}{\partial t}\phi = f(t, \phi, u(t)),$$

then our implicit representation is given by:

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

Generally speaking the state is a vector and f is a vector valued function. This is easy to understand looking at the example below.

Example 2.18. Consider the Mass Spring system in Figure 4. Let $\tilde{x}$ represent the vertical position of the mass. If we choose the state as

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \tilde{x} - x_{rest} \\ \frac{d(\tilde{x} - x_{rest})}{dt} \end{bmatrix}$$

the use of standard laws of mechanics leads us to

$$\dot{x}(t) = \begin{bmatrix} \dot{x_1}(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} x_2 \\ F(t) - Kx_1 - Rx_2 \end{bmatrix}.$$

2.4.4. Link between state and I/O representation. Looking at Example 2.14 and at Example 2.18, the reader can infer an easy way to derive a state representation from an I/O representation. Suppose our system is described by an a differential (or by a difference) equation in Equation (2.11). Consider the simplified case in which $p = 1$. Suppose we set $x_i = \mathfrak{D}^{i-1}y$, with the obvious implication that

$$\mathfrak{D}x_i = \mathfrak{D}\mathfrak{D}^{i-1}y = \mathfrak{D}^{iy\beta}y = x_{i+1}$$

. We can write Equation (2.11) as:

$$\mathfrak{D}\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ \dots \\ F(x_1, x_2, \dots, x_n, u, t) \end{bmatrix} \tag{2.22}$$

$$y = x_1. \tag{2.23}$$

This way, we have found the expressions for the vector valued function $f(x, u, t)$ and for the function $h(x, u, t)$:

$$f(x, u, t) = \begin{bmatrix} x_2 \\ x_3 \\ \dots \\ F(x_1, x_2, \dots, x_n, u, t) \end{bmatrix}$$
$$\eta(x, u, t) = x_1,$$

where x is defined as $x = [x_1, x_2, \dots, x_n]^T$. If $n \geq p > 1$, the construction can be generalised by setting:

$$\begin{aligned} x_i &= \mathfrak{D}^{i-1}y && \text{for } i \in [1, n] \\ x_{i+n} &= \mathfrak{D}^{i-1}u && \text{for } i \in [1, p]. \end{aligned} \tag{2.24}$$

This strategy leads to us to the result. But, sometimes there are alternative routes that lead to more efficient results (especially when $p > 1$). Consider the following example:

Example 2.19. Consider a system whose I/O representation is given by

$$\dot{y} = y + 5\dot{u} - u.$$

Suppose, that we set $x_1 = y - 5u$.

$$\begin{aligned} \dot{x_1} &= \dot{y} - 5\dot{u} \\ &= y - u = \\ &= y - 5u + 4u \\ &= x_1 + 4u \\ y &= x_1 + 5u \end{aligned}$$

But, let us now introduce an additional state variable x_2 such that

$$\begin{aligned} \dot{x_1} &= x_1 + 4u \\ \dot{x_2} &= 3u \\ y &= x_1 + 5u \end{aligned}$$

From an IO perspective the behaviours generated by the two systems are absolutely the same.

The example above reveals that: 1) there are more than one state space representation for the same I/O description, 2) some systems (that we will learn to qualify as linear) allow for a much more compact state space than the one using the canonical set of variables in Equation (2.24)

2.5. Linear Systems

As we discussed above, an abstract system $\Sigma(t_0)$ is defined as a set of pairs $(u_0(t), y_0(t))$ with $u_0(t) \in \mathcal{U}^{\mathcal{T}(t_0)}$ and $y_0(t) \in \mathcal{Y}^{\mathcal{T}(t_0)}$. Suppose that the sets $\mathcal{U}$ and $\mathcal{Y}$ are both vector spaces defined over the real set $\mathbb{R}$. Clearly the range sets U and Y where the functions take value have to possess the same algebraic structure. In plain words this means that if $u_1(t) \in \mathcal{U}$ and $u_2(t) \in \mathcal{U}$, then $\alpha u_1(t) + \beta u_2(t) \in U$ for any $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ (and the same for any pair $y_1(t)$ and $y_2(t)$ belonging to $\mathcal{Y}$).

This being said, we can introduce the following definition

DEFINITION 2.20. The system $\Sigma(t_0)$ is linear if given any two pairs $(u_1(t), y_1(t))$, $(u_2(t), y_2(t))$, if they both belong to the system $\Sigma(t_0)$, so will their linear combinationa: $\alpha\,(u_1(t), y_1(t)) + \beta\,(u_2(t), y_2(t)) \in \Sigma(t_0)$, where $\alpha, \beta \in \mathbb{R}$.

Consider now a parametric description of the system with a parameter space X_{t_0} and a function π. For a given x_0, the system responds to an input u_1 with the output $y_1 = \pi_{t_0}(x_0, u_1)$ and to the input u_2 with the output $y_2 = \pi_{t_0}(x_0, u_2)$. The intuitive meaning of linearity is:

- *Superimposition*: the response to the combination $u_1(t) + u_2(t)$ will be the superimposition of the effect of $u_1(t)$ and $u_2(t)$: $y_1(t) + y_2(t)$
- *Scaling*: the response to $\alpha u_1(t)$ is $\alpha y_1(t)$.

EXAMPLE 2.21. Consider the mass spring system in Example 2.14. Suppose that starting from an initial condition $x_1(t)$ is the evolution of the system for $F_1(t)$ and $x_2(t)$ is the solution to corresponding to the force $F_2(t)$ (both are evaluated starting from the rest conditions). We can write

$$\begin{aligned} m\ddot{x}_1 &= F_1(t) - Kx_1 - R\dot{x}_1 \\ m\ddot{x}_2 &= F_2(t) - Kx_2 - R\dot{x}_2 \end{aligned}$$

If we multiply the first equation by α, the second equation by β and sum up the result we will see that $\alpha x_1(t) + \beta x_2(t)$ is a solution to $\alpha F_1(t) + \beta F_2(t)$.

The considerations in the example before apply because the differential equation describing the system is linear and the derivative is a linear operator: $\mathfrak{D}(\alpha f(t) + \beta g(t)) = \alpha\mathfrak{D}(f(t)) + \beta\mathfrak{D}(g(t))$.

If we repeat the same line of reasoning with the system in Example 2.16, we will fail because the equation is not linear. More simply, if we consider the mass spring example but we impose a maximum value for the force $F(t)$ the scaling principle will not apply and the linearity will be lost.

2.5.1. Linear I/O representations. The same argument of Example 2.21 can be repeated for any system described by a linear difference or differential equation of:

$$\mathfrak{D}^n y(t) = \sum_{i=0}^{n-1} \alpha_i(t)\mathfrak{D}^i y(t) + \sum_{j=0}^{p} \beta_j(t)\mathfrak{D}^j u(t). \tag{2.25}$$

Generally speaking, when this equation has a unique solution (which is generally true for "well behaved" functions $\alpha(t)$ and $\beta(t)$) and for causal systems ($p \leq n$), such a solution is a function of u and of the initial conditions:

$$y(t) = \phi\left(t, t_0, y(0), \mathfrak{D}y(0), \ldots, \mathfrak{D}^{n-1}y(0), u(0), \ldots, \mathfrak{D}^p u(0), u|_{[t_0,t]}\right).$$

There are two known facts stated in the following:

LEMMA 2.22. *Given the Equation* (2.25)*, define its associated homogeneous equation as follows:*

$$\mathfrak{D}^n y(t) = \sum_{i=0}^{n-1} \alpha_i(t)\mathfrak{D}^i y(t) \tag{2.26}$$

The solution to this equation form a vector space and the dimension of this space is n. This applies to both differential and difference equations.

LEMMA 2.23. *Suppose that $\tilde{y}$ is a particular solution of Equation* (2.25) *related to the input $\tilde{u}$. If we take solution of Equation* (2.26) *$\hat{y}$ and we sum it to $\tilde{y}$ we still obtain a solution of Equation* (2.25)*.*

By choosing any basis $\{\hat{y}_1, \ldots, \hat{y}_n\}$ of the vector space of the solutions of Equation (2.26), we can write a general solution of Equation (2.25) as

$$y(t) = \tilde{y}(t) + H_1\hat{y}_1(t) + \ldots + H_n\hat{y}_n(t)$$

where the coefficients H_i, $i \in 1, \ldots, n$, can be found by imposing n conditions derived from the initial conditions. In essence, the linearity of the system leads us to a break down of the solution in two pieces. The first one, $\hat{y}$, is called forced evolution because it depends on the input term $\hat{u}$. The second one is $H_1\hat{y}_1(t) + \ldots + H_n\hat{y}_n(t)$ and is called free evolution because it depends on the initial conditions.

An obvious but fundamental fact is that a linear representation as described above generates a linear system as in Definition 2.20.

2.5.2. Linear State Space Representations. A state space explicit representation is given by a couple of functions ϕ and η, defined as:

$$\begin{aligned} \phi : \mathcal{T} \times \mathcal{T} \times X \times \mathcal{U} \to X \qquad & x(t) = \phi(t, t_0, x_0, u) \\ \eta : \mathcal{T} \times X \times U \to X \qquad & y(t) = \eta(t, x, u) \end{aligned}$$

A representation of this type is linear if both functions are linear with respect to their arguments $X \times \mathcal{U}$. In plain words, if we consider ϕ we have:

$$\begin{cases} \phi(t, t_0, x_0, u_1|_{[t_0,t)}) = x_1 \\ \phi(t, t_0, x_0, u_2|_{[t_0,t)}) = x_2 \end{cases} \implies \phi(t, t_0, x_0, \alpha u_1|_{[t_0,t)} + \beta u_2|_{[t_0,t)}) = \alpha x_1 + \beta x_2$$

$$\begin{cases} \phi(t, t_0, x_{0,1}, u|_{[t_0,t)}) = x_1 \\ \phi(t, t_0, x_{0,2}, u|_{[t_0,t)}) = x_2 \end{cases} \implies \phi(t, t_0, \alpha x_{0,1} + \beta x_{0,2}, u) = \alpha x_1 + \beta x_2.$$

We can write similar equations for η,

$$\begin{cases} \eta(t, x_0, u_1) = y_1 \\ \eta(t, x_0, u_2) = y_2 \end{cases} \implies \eta(t, x_0, \alpha u_1 + \beta u_2) = \alpha y_1 + \beta y_2$$

$$\begin{cases} \eta(t, x_{0,1}, u) = y_1 \\ \eta(t, x_{0,2}, u) = y_2 \end{cases} \implies \eta(t, \alpha x_{0,1} + \beta x_{0,2}, u) = \alpha y_1 + \beta y_2.$$

The implicit representation can be obtained from the explicit one by differentiation (assuming that ϕ is differentiable) or by considering one step of evolution.

$$\begin{aligned} f : \mathcal{T} \times X \times U \to X \qquad & \mathfrak{D}x(t) = f(t, x, u) \\ \eta : \mathcal{T} \times X \times U \to X \qquad & y(t) = \eta(t, x, u) \end{aligned}$$

In this case f has to be linear with respect to $X \times U$. Observe the difference. For ϕ we require linearity with respect to $X \times \mathcal{U}$ (where $\mathcal{U}$ is a space of signals), while for f and for η we require linearity with respect to $X \times U$ (where U is the range of the input signals $\mathcal{U}$).

The linearity of f and η can be expressed in matrix notation as:

$$\begin{aligned} \mathfrak{D}x(t) &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)x(t) + D(t)u(t). \end{aligned} \tag{2.27}$$

2.6. Time Invariance

We can introduce a general definition of time invariance for an abstract system $\Sigma(t_0)$ defined as a set of pairs $(u_0(t), y_0(t))$.

Definition 2.24. A system $\Sigma(t_0)$ is time invariant iff

$$(u_0(t), y_0(t)) \in \Sigma(t_0) \implies (u_0(t - t_0), y_0(t - t_0)) \in \Sigma(t_0 - t_0).$$

In plain world, what we require is that if a behaviour $(u_0(t), y_0(t))$ is generated by the system, so will the same behaviour with u_0 and y_0 shifted forward or backward in time of the same amount.

The definition of time invariance applies also to the representation of the system.

2.6.1. IO representation. An IO representation of a system is time invariant if the differential or the difference equation do not explicitly depend on t.

In order to understand this definition, consider Example 2.3. If we allow the interest rates I_+, I_- to change in time the system is not obviously time invariant. If they are kept constant, the system is time invariant. Likewise, for the mass-spring system in Example 2.14, we can say that the system is time invariant. But if one of the parameters (e.g., the Hooke constant of the spring) changes with time, then the system is time-varying.

Generally speaking, if we consider the linear IO representation in Equation (2.25), we can say that the system is time–invariant if the parameters $\alpha_i(t)$ and $\beta_j(t)$ are constant over time. In this case, the equation becomes

$$\mathfrak{D}^n y(t) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i y(t) + \sum_{j=0}^{p} \beta_j \mathfrak{D}^j u(t). \tag{2.28}$$

2.6.2. State Representation. If we consider a state representation in explicit form, it is time invariant if and only if *phi* does not depends singularly on t and t_0, but it depends on their difference, and if η does not depend on t

$$\begin{aligned} \phi : \mathcal{T} \times \mathcal{T} \times X \times \mathcal{U} \to X \qquad & x(t) = \phi(t, t_0, x_0, u) \qquad = \phi(t - t_0, x_0, u) \\ \eta : \mathcal{T} \times X \times U \to X \qquad & y(t) = \eta(t, x, u) \qquad\qquad = \eta(x, u) \end{aligned}$$

If we derive the implicit form, it is easy to see that this implies that f and η do not depend on t.

$$\begin{aligned} f : \mathcal{T} \times X \times U \to X \qquad & \mathfrak{D}x(t) = f(t, x, u) \qquad = f(x, u) \\ \eta : \mathcal{T} \times X \times U \to X \qquad & y(t) = \eta(t, x, u) \qquad = \eta(x, u) \end{aligned}$$

In the case of linear system, time invariance is translated into a set of matrices A, B, C, D that do not depend on time:

$$\begin{aligned} \mathfrak{D}x(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t). \end{aligned} \tag{2.29}$$

It is possible to see that linear IO and state space representations are associated with linear systems

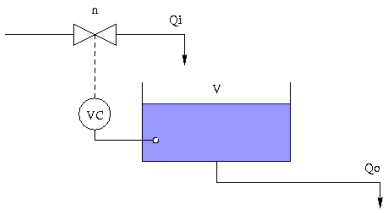


FIGURE 6. Example of an Hybrid System

2.7. Hybrid Systems

Different hybrid systems are obtained combining a DES event that determines changes in the "physical laws" that govern the evolution of a CT system. Consider the example in Figure 2.7. A tank is filled with a fluid, which flows out with a rate Q_0 from an outlet in the bottom of the tank. When the level of the fluid reaches a threshold, a sensor detects the event and opens a valve which produces an incoming flow $Q_1 > Q_2$ until the tank is full. In this example, we have a state machine that determines switches between two different evolutions of the system, each of them governed by CT dynamics. This type of systems are beyond the scope of this course.

CHAPTER 3

Impulse response and convolutions for linear and time invariant IO Representations

In this chapter, we restrict our focus on IO representation of SISO linear and time invariant systems. Generally speaking, the evolution of the system is described by the following equation

$$\mathfrak{D}^n y(t) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i y(t) + \sum_{j=0}^{p} \beta_j \mathfrak{D}^j u(t), \tag{3.1}$$

where time invariance requires that α_i and β_i be constant. The system is causal if $p \leq n$ and strictly causal if $p < n$. We have seen that the initial conditions:

$$y(0), \mathfrak{D}y(0), \ldots, \mathfrak{D}^{n-1}y(0), \ldots, \mathfrak{D}^p u(0), \ldots, \mathfrak{D}u(0)$$

and the input $u|_{[t_0,t]}$ allows us to generally find a unique solutions. We have finally seen that the linearity of the system allows us to split the evolution depending on the initial conditions and the evolution depending on the input in two separate terms:

$$\begin{aligned} y(t) &= y_{\text{free}}(t) + y_{\text{forced}}(t) \\ &\text{with} \\ y_{\text{free}}(t) &= \mathcal{F}_{\text{free}}(y(0), \mathfrak{D}y(0), \ldots, \mathfrak{D}^{n-1}y(0), \ldots, \mathfrak{D}^p u(0), \ldots, \mathfrak{D}u(0)) \\ y_{\text{forced}}(t) &= \mathcal{F}_{\text{forced}}(u|_{[t_0,\,t]}). \end{aligned}$$

Let us start with the computation of the forced evolution.

3.1. Forced Evolution of discrete–time systems

It is useful to introduce a special function, which we will call impulse function.

3.1.1. Kronecker δ. For discrete time system the impulse function is also known as Kronecker Delta and is defined as follows:

$$\delta(t) = \begin{cases} 1 & t = 0 \\ 0 & t \neq 0. \end{cases} \tag{3.2}$$

If we consider any discrete–time signal we have the following property:

$$f(t)\delta(t-t_0) = \begin{cases} f(t_0) & t = t_0 \\ 0 & t \neq t_0 \end{cases} \tag{3.3}$$

In plain words we will say that multiplying a Kronecker δ shifted to t_0 by a signal generates a new signal that is zero at all times except for t_0 (where it is $f(t_0)$). This is equivalent to saying that multiplying by δ has the effect of extracting a sample ("sampling") of the signal in the point where δ is centred. Another property is:

$$\sum_{\tau=-\infty}^{\tau=\infty} f(\tau)\delta(t-\tau) = f(t) \tag{3.4}$$

In other words we can express any signal as a linear combination of infinite δ signal, each translated by a different amount.

3.1.2. Impulse Response. Consider the DT system in Equation (3.1). Our goal is to compute the forced response to a generic signal $u(t)$.

Let us define $h(t)$ the forced response to the signal $\delta(t)$.

$$h(t) = \mathcal{F}_{\text{forced}}(\delta).$$

As we discussed above we can write $u(t)$ as

$$u(t) = \sum_{\tau=-\infty}^{\tau=\infty} u(\tau)\delta(t-\tau)$$

Thanks to the system's linearity we can write:

$$\begin{aligned} \mathcal{F}_{\text{forced}}(u) &= \mathcal{F}_{\text{forced}}\Big(\sum_{\tau=-\infty}^{\tau=\infty} u(\tau)\delta(t-\tau)\Big) \\ &= \sum_{\tau=-\infty}^{\tau=\infty} u(\tau)\mathcal{F}_{\text{forced}}(\delta(t-\tau)), \end{aligned}$$

and using time–invariance we get.

$$\mathcal{F}_{\text{forced}}(u) = \sum_{\tau=-\infty}^{\tau=\infty} u(\tau)h(t-\tau)$$

The summation above is called discrete–time convolution (also denoted as $u(t) * h(t)$. What we just said is that for a DT system it is sufficient to know the impulse response $h(t)$ to be able to reconstruct the response to any signal.

The meaning of the convolution sum is quite simple. Suppose we want to find $y(t) = u(t) * h(t)$. What we need to do is the following:

(1) Compute the sequence $h(-\tau)$, which is the reflection of $h(\tau)$ through $t = 0$;
(2) Shift the obtained sequence by t and compute the product element–wise;

(3) Sum up the products, and this produces $y(t)$.

This is shown in the two examples below.

EXAMPLE 3.1. Suppose that:

$$h(t) = \begin{cases} 1 & t = 0 \\ 2 & t = 1 \\ -2 & t = 2 \\ 0 & \text{Otherwise} \end{cases}$$

and

$$u(t) = \begin{cases} 5 & t = 0 \\ -3 & t = 1 \\ 0 & \text{Otherwise} \end{cases}$$

compute the system's response. We can see

$$y(t) = \sum_{\tau=-\infty}^{\tau=\infty} u(\tau)h(t-\tau)$$

$$= 5h(t) - 3h(t-1) = \qquad = \begin{cases} 5 & t = 0 \\ 10-3 & t = 1 \\ -10-6 & t = 2 \\ 6 & t = 3. \end{cases}$$

From this example we can clearly see that if the support of $h(t)$ is $[0, M]$ and the support of $u(t)$ is $[0, N]$, the support of $h(t) * u(t)$ will be $[0, N+M]$.

EXAMPLE 3.2. Suppose we know that the impulse response to a system is given by $1(t)a^t$, let us compute the forced response to the input function $u(t) = 1(t)b^t$, where

$$1(t) = \begin{cases} 1 & t > 0 \\ 0 & t \leq 0. \end{cases}$$

We have:

$$\begin{aligned} \mathcal{F}_{\text{forced}}(u) &= \sum_{\tau=-\infty}^{\tau=\infty} 1(\tau)b^\tau 1(t-\tau)a^{t-\tau} = \\ &= \sum_{\tau=1}^{\tau=t-1} b^\tau a^{t-\tau} = \\ &= a^t \sum_{\tau=1}^{\tau=t-1} \left(\frac{b}{a}\right)^\tau = \\ &= a^t \frac{(b/a) - (b/a)^t}{1-(b/a)} = \\ &= \frac{b}{a-b}a^t - \frac{a}{a-b}b^t. \end{aligned}$$

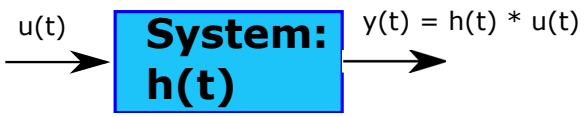


FIGURE 1. Example block scheme

Our previous discussion can be summarised in the block scheme in Figure 3.1.2. The box represent the system and the arrows the input and the output signals.

3.1.3. Properties of the convolution sum. Three important properties of the convolution sum are stated in the following.

THEOREM 3.3. *The convolution sum enjoys the following properties:*

(1) *Commutative Property:* $h(t) * u(t) = h(t) * u(t)$.
(2) *Distributive Property:* $(h_1(t) + h_2(t)) * u(t) = h_1(t) * u(t) + h_2(t) * u(t)$
(3) *Associative Property:* $h_1(t) * (h_2(t) * u(t)) = (h_1(t) * h_2(t)) * u(t)$.

PROOF. (1) Commutative Property: considering that $h(t) * u(t) = \sum_{\tau=-\infty}^{+\infty} h(\tau)u(t-\tau)$, by changing the variable as $t - \tau = \tau_1$ we can write :

$$\begin{aligned}\sum_{\tau=-\infty}^{+\infty} h(\tau)u(t-\tau) &= \sum_{\tau_1=\infty}^{-\infty} h(t-\tau_1)u(\tau_1) \\ &= u(t) * h(t).\end{aligned}$$

(2) Distributive Property: this is a straightforward implication of the linearity of the convolution operator:

$$\begin{aligned}(h_1(t) + h_2(t)) * u(t) &= \sum_{\tau=-\infty}^{+\infty} (h_1(\tau) + h_2(\tau))\, u(t-\tau) = \\ &= \sum_{\tau_1=\infty}^{-\infty} h_1(\tau)u(t-\tau) + \sum_{\tau_1=\infty}^{-\infty} h_2(\tau)u(t-\tau) = \\ &= h_1(t) * u(t) + h_2(t) * u(t)\end{aligned}$$

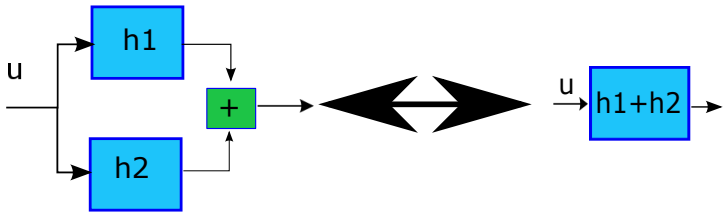


FIGURE 2. Parallel composition

FIGURE 3. Series composition

(3) Associative Property:

$$
\begin{aligned}
(h_1(t) * h_2(t)) * u(t) = \sum_{\tau_2=-\infty}^{\infty} u(t-\tau_2) \sum_{\tau_1=-\infty}^{\infty} h_1(\tau_1) h_2(\tau_2-\tau_1) \quad &= \\
= \sum_{\tau_2=-\infty}^{\infty} \sum_{\tau_1=-\infty}^{\infty} u(t-\tau_2) h_1(\tau_1) h_2(\tau_2-\tau_1) \quad &= \\
= \sum_{\tau_1=-\infty}^{\infty} h_1(\tau_1) \sum_{\tau_2'=-\infty}^{\infty} u(\tau_2') h_2(t-\tau_2'-\tau_1) \quad &= \\
= \sum_{\tau_1=-\infty}^{\infty} h_1(\tau_1)(u(t) * h_2(t))|_{t-\tau_1} & \\
= h_1(t) * (u(t) * h_2(t)) &
\end{aligned}
$$

□

The first property has merely an operational value: if we want to process $u(t)$ through a system $h(t)$ we can compute the most convenient between $h(t) * u(t)$ and $u(t) * h(t)$.

The second and the third property have a more substantial meaning related to the composition of two systems. The distributive property reveals that the parallel composition of two systems produces the same result as having one system whose impulse response is given by the sum of the two impulse responses (see Figure 2).

The associative property is related to the series composition of two systems, which is equivalent to one system having an impulse response given by the convolution of the two impulse responses (see Figure 3).

3.1.4. Eigenfunctions. There is particular type of signals that are treated in special way by LTI system. Consider signal $u(t) = z^t$ and compute

the forced evolution of a LTI system. The response to this input is given by:

$$\begin{aligned}\sum_{\tau=-\infty}^{+\infty} h(\tau)u(t-\tau) &= \sum_{\tau=-\infty}^{\infty} u(t-\tau)h(\tau) \\ &= \sum_{\tau=-\infty}^{\infty} z^{t-\tau}h(\tau) \\ &= z^t \sum_{\tau=-\infty}^{\infty} z^{-\tau}h(\tau) \\ &= z^t H(z)\end{aligned}$$

$$\text{where } H(z) = \sum_{\tau=-\infty}^{\infty} z^{-\tau}h(\tau).$$

The property just shown can be stated by saying that whenever a signal z^t (i.e., an exponential of t with basis z) gets processed by a DT LTI system, the outcome is the same signal scaled by a constant $H(z)$ which is the limit of the series $\sum_{\tau=-\infty}^{\infty} z^{-\tau}h(\tau)$, as far as the series converge. For this reason z^t is called an *eigenfunction*. An eigenfunction is essentially similar to an eigenvector of a standard linear application, with $H(z)$ being its eigenvalue. This applies both to real and complex z.

3.1.4.1. *A key property of eigenvalues.* We should recall that eigenvectors enjoy an extremely useful property. Consider a linear application defined from $\mathbb{R}^n$ to $\mathbb{R}^n$. Suppose that it is associated to a matrix A:

$$y = Ax.$$

Suppose that we can identify n independent eigenvectors: $\{u_1, u_2, \ldots, u_n\}$ related to the eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. These vectors form a basis. Let $M = [u_1 u_2 \ldots u_n]$ be the matrix composed using these vectors. Let $\hat{x}$ be the coordinates in this basis of a generic vector x expressed in the canonical basis. We have:

$$\hat{x} = M^{-1}x.$$

Similarly the transformed version of y is given by:

$$\hat{y} = M^{-1}y.$$

By combining the two conditions we find:

$$\begin{aligned}\hat{y} = M^{-1}y &= \\ &= M^{-1}Ax = \\ &= M^{-1}AM\tilde{x}. \qquad (3.5)\end{aligned}$$

It can easily be seen that $M^{-1}AM$ is a diagonal matrix. In simple words, this means that if we express a vector using a basis of eigenvectors, the system operates on each component in a decoupled way.

The same property holds if we express any signal as a linear combination of Eigenfunctions. The convenience of this choice will become apparent in the next few pages.

3.2. Forced Evolution of continuous–time systems

The line of arguments used for DT systems can be applied to CT systems as well. The key point is the definition of an appropriate impulse function.

3.2.1. Dirac δ. The notion of impulse is not as obvious for CT systems as it is for DT systems. Indeed, the type of functions we normally deal with are at the very least continuous and often differentiable. The impulse function does not fall within this category, but it is useful to think of it as the limit case of a relatively well behaved signal. The simplest choice that can be made is to define:

$$\delta(t) = \lim_{\Delta \to 0} \delta_\Delta(t) \tag{3.6}$$

$$\delta_\Delta(t) = \begin{cases} 0 & t \notin \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \\ \frac{1}{\Delta} & t \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \end{cases} \tag{3.7}$$

A few properties that are easy to show are the following:

Property 1: If we compute the integral of Dirac δ on any interval enclosing the origin, we get 1.0:

$$\forall a > 0,\, b > 0 \int_{-a}^{b} \delta(\tau) d\tau = 1.$$

Property 2: Multiplying a $\delta(t-\tau)$ by any function has the effect of "sampling" the value of the function in τ, meaning that we obtain an impulse function multiplied by τ:

$$f(t)\delta(t-\tau) = f(\tau)\delta(t-\tau).$$

Property 3: Any function can be expressed as an integral of impulse functions.

$$\forall \epsilon > 0,\, f(t) = \int_{t-\epsilon}^{t+\epsilon} f(\tau)\delta(t-\tau)d\tau.$$

Indeed,

$$\begin{aligned} \int_{t-\epsilon}^{t+\epsilon} f(\tau)\delta(t-\tau)d\tau &= \int_{t-\epsilon}^{t+\epsilon} f(t)\delta(t-\tau)d\tau \\ &= f(t)\int_{t-\epsilon}^{t+\epsilon} \delta(t-\tau)d\tau \\ &= f(t). \end{aligned}$$

Property 4: The integral from any negative number to a generic instant t produces a step function:

$$\forall \epsilon > 0, \int_{-\epsilon}^{t} \delta(\tau) d\tau = 1(t),$$

where $1(t)$ is defined as

$$1(t) \begin{cases} 1 & t > 1 \\ 0 & t \leq 0 \end{cases}$$

Property 5: We could define

$$\delta(t) = \frac{d}{dt} 1(t).$$

This is an obvious abuse of notation because we know that the step function is not differentiable in $t = 0$.

3.2.2. Impulse Response. We can repeat the same line of arguments followed in Section 3.1.2 to compute the forced response of system (3.1) starting from the impulse response (i.e., the forced response to $\delta(t)$):

$$h(t) = \mathcal{F}_{\text{forced}}(\delta).$$

Indeed, we can write

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

Using linearity and time–invariance:

$$\begin{aligned} y(t) &= \mathcal{F}_{\text{forced}}(u) \\ &= \mathcal{F}_{\text{forced}}(\int_{\tau=-\infty}^{\tau=\infty} u(\tau)\delta(t-\tau)d\tau) \\ &= \int_{\tau=-\infty}^{\tau=\infty} u(\tau)\mathcal{F}_{\text{forced}}(\delta(t-\tau))d\tau, \\ &= \int_{\tau=-\infty}^{\tau=\infty} u(\tau)h(t-\tau)d\tau. \end{aligned}$$

This allows us to conclude that once we know $h(t)$, we can compute the response to any signal. This is called "convolution integral" and is denoted by the same symbol as the convolution sum: $u(t) * h(t)$. The computation of $y(t)$ requires the following steps:

(1) Compute the "reflection" of $h(\tau)$ through $\tau = 0$,
(2) Translate the result to the right of t (to the left if t is negative).
(3) Compute the product by $u(\tau)$ and then the integral of the function thus obtained.

These steps can be appreciated through an example.

EXAMPLE 3.4. Suppose that $h(t) = 1(t)e^{-3t}$. Compute the response of the system to $u(t) = 1(t)$. The result can be found as follows:

$$\begin{aligned} y(t) &= \mathcal{F}_{\text{forced}}(u) \\ &= \int_{\tau=-\infty}^{\tau=\infty} u(\tau)h(t-\tau)d\tau \\ &= \int_{\tau=-\infty}^{\tau=\infty} 1(t)1(t-\tau)e^{-3t+3\tau}d\tau \\ &= \int_{0}^{\tau=t} e^{-3t+3\tau}d\tau \\ &= e^{-3t}\frac{1}{3}e^{3\tau}|_{\tau=0}^{t} \\ &= \frac{1-e^{-3t}}{3}. \end{aligned}$$

3.2.3. Properties of the convolution integral. The convolution integral has the same properties as the convolution sum for DT system. Three important properties of the convolution sum are stated in the following.

THEOREM 3.5. *The convolution integral enjoys the following properties:*

(1) *Commutative Property:* $u(t) * h(t) = h(t) * u(t)$.
(2) *Distributive Property:* $(h_1(t) + h_2(t)) * u(t) = h_1(t) * u(t) + h_2(t) * u(t)$
(3) *Associative Property:* $h_1(t) * (h_2(t) * u(t)) = (h_1(t) * h_2(t)) * u(t)$.

The proof is very similar to one we have given for the convolution sum.

3.2.4. Eigenfunctions. Suppose that we use as input an exponential function $u(t) = e^{st}$.

$$\begin{aligned} y(t) &= \int_{\tau=-\infty}^{+\infty} h(\tau)u(t-\tau)d\tau \\ &= \int_{\tau=-\infty}^{+\infty} h(\tau)e^{s(t-\tau)}d\tau \\ &= e^{st}\int_{\tau=-\infty}^{+\infty} h(\tau)e^{-s\tau}d\tau. \end{aligned}$$

Let $H(s)$ be the integral $\int_{\tau=-\infty}^{+\infty} h(\tau)e^{-s\tau}d\tau$, assuming that it converges. We can summarise the computation above by saying that e^{st} is an Eigenfunctions related to the eigenvalue $H(s)$.

The above results applies to both real and complex exponentials. This leads us to an interesting fact if we analyse the output to an harmonic function: $u(t) = \cos\omega t$. Using Euler's expression, we have

$$\cos\omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2}.$$

Applying linearity we find:

$$\begin{aligned} y(t) &= \int_{\tau=-\infty}^{+\infty} h(\tau)u(t-\tau)d\tau \\ &= \int_{\tau=-\infty}^{+\infty} h(\tau)\frac{e^{j\omega(t-\tau)}+e^{-j\omega(t-\tau)}}{2}d\tau \\ &= \frac{1}{2}\int_{\tau=-\infty}^{+\infty} h(\tau)e^{j\omega(t-\tau)}d\tau + \frac{1}{2}\int_{\tau=-\infty}^{+\infty} e^{-j\omega(t-\tau)}d\tau \\ &= \frac{1}{2}e^{j\omega t}H(j\omega) + \frac{1}{2}e^{-j\omega t}H(-j\omega), \end{aligned}$$

where $H(j\omega) = \int_{\tau=-\infty}^{+\infty} h(\tau)e^{-j\omega\tau}d\tau$. Let $\overline{z}$ represent the complex conjugate of a complex number z. If z_1 and z_2 are two complex numbers and α is a real, we can easily show that:

(1) $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$
(2) $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$
(3) $\overline{\alpha z_1} = \alpha\overline{z_1}$.

Applying these properties, we can see:

$$\begin{aligned} \overline{H(j\omega)} &= \overline{\int_{\tau=-\infty}^{+\infty} h(\tau)e^{-j\omega\tau}d\tau} \\ &= \int_{\tau=-\infty}^{+\infty} \overline{h(\tau)e^{-j\omega\tau}}d\tau \\ &= \int_{\tau=-\infty}^{+\infty} h(\tau)\overline{e^{-j\omega\tau}}d\tau \\ &= H(-j\omega) \end{aligned}$$

Therefore,

$$\begin{aligned} y(t) &= \frac{1}{2}e^{j\omega t}H(j\omega) + \frac{1}{2}e^{-j\omega t}H(-j\omega) \\ &= \frac{1}{2}e^{j\omega t}H(j\omega) + \frac{1}{2}e^{-j\omega t}\overline{H(j\omega)}. \end{aligned}$$

If we use for $H(j\omega)$ its modulus/phase representation:

$$H(j\omega) = |H(j\omega)|\, e^{j\angle H(j\omega)},$$

we find:

$$\begin{aligned} y(t) &= \frac{1}{2}e^{j\omega t}H(j\omega) + \frac{1}{2}e^{-j\omega t}\overline{H(j\omega)} \\ &= \frac{1}{2}\left|H(j\omega)\right| e^{j\angle H(j\omega)}e^{j\omega t} + \frac{1}{2}e^{-j\omega t}\left|H(j\omega)\right| e^{-j\angle H(j\omega)} \\ &= \frac{|H(j\omega)|}{2}\left(e^{j(\angle H(j\omega)+\omega t)} + e^{-j(\angle H(j\omega)+\omega t)}\right) \\ &= |H(j\omega)|\cos\left(\omega t + \angle H(j\omega)\right). \end{aligned}$$

The result above can be summarised in the following:

THEOREM 3.6. *Consider a TC LTI system. If* $\int_{\tau=-\infty}^{+\infty} h(\tau)e^{-j\omega\tau}d\tau$ *converges to a value* $H(j\omega)$*, then the system responds to an* harmonic input function $\cos\omega t$ *with an harmonic output function having the same frequency.*

3.3. Properties of the impulse response

Many important properties of the system can be inferred by looking at the analytical features of $h(t)$, both for DT and CT systems.

3.3.1. Causality. The first important problem that can be read from $h(t)$ is causality, as specified in the following Theorem.

THEOREM 3.7. *Let* $h(t)$ *be the impulse response of a system* Σ*. The system is causal if and only if* $h(t) = 0$ *for* $t < 0$*.*

PROOF. Let us focus for simplicity on CT systems. Necessity derives from the observation that if we choose $u(t) = \delta(t)$, causality requires that $h(t) = 0$ for $t < 0$.

Sufficiency is shown considering that if $h(t) = 0$ for $t < 0$ then

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

As shown in the formula above, the computation of y at time t is only affected by $u(\cdot)$ until t. The proof for DT systems is absolutely similar. □

3.3.2. BIBO Stability. Key to this course is the notion of stability. Roughly speaking, a stable system is one for which small perturbations of various kind on the evolution of the output signal. We will see several possible way in which this intuitive notions can be formalised in mathematical terms. The first one applies to IO representations and is the so-called Bounded-Input-Bounded-Output (BIBO) stability formalised below.

DEFINITION 3.8 (BIBO stability). A system is BIBO stable iff for all $\epsilon > 0$ there exists a positive real $\delta > 0$ such that

$$|u(t)| \leq \epsilon \implies |y(t)| < \delta.$$

The idea is that if we apply "small" input signals produce "small" output signals.

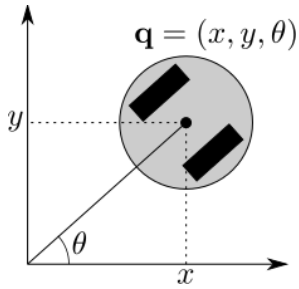


FIGURE 4. Scheme of a unicycle mobile robot.

EXAMPLE 3.9. Consider a cylindrical robot that moves along a path. Suppose we are able to control the vehicle forward speed $v(t)$ and the angular speed $\omega(t)$. We can easily see that the differential equations describing the kinematics of the system are

$$\dot{y} = v \sin\theta$$
$$\dot{\theta} = \omega.$$

Suppose that our output is y We can see that if apply the following signal

$$\omega = \begin{cases} \epsilon & t \in [0, 0.1s] \\ 0 & t \notin [0, 0.1s] \end{cases}$$

the system changes slightly its orientation, but from that moment onward y starts to grow unbounded even if ϵ is very small. This is the perfect example of a BIBO unstable system.

The BIBO stability property is particularly easy to study for LTI system, for which we can prove the following.

THEOREM 3.10. *Consider a LTI system Σ with impulse response $h(t)$. If the system is DT then it is BIBO stable if and only if there exists a constant S such that $\sum_{-\infty}^{\infty} |h(t)| = S < \infty$. If the system is CT then it is BIBO stable if and only if there exist a constant S such that $\int_{-\infty}^{\infty} |h(\tau)| d\tau = S < \infty$.*

PROOF. We give the proof for the case of DT systems. The output of the system is given by

$$y(t) = \sum_{\tau=-\infty}^{\infty} h(\tau) u(t - \tau)$$

To prove sufficiency, assume that for some $\epsilon > 0$ we have $u(t) \leq \epsilon, \forall t$. We have:

$$\begin{aligned} y(t) &= \sum_{\tau=-\infty}^{\infty} h(\tau)u(t-\tau) \\ &\leq \sum_{\tau=-\infty}^{\infty} |h(\tau)u(t-\tau)| \\ &\leq \sum_{\tau=-\infty}^{\infty} |h(\tau)||u(t-\tau)| \\ &\leq \sum_{\tau=-\infty}^{\infty} |h(\tau)|\epsilon \\ &\leq \epsilon \sum_{\tau=-\infty}^{\infty} |h(\tau)| \\ &\leq S\epsilon. \end{aligned}$$

Hence,

$$|u(t)| \leq \epsilon \implies |y(t)| \leq \delta = S\epsilon,$$

which proves sufficiency.

To prove necessity, it is sufficient to consider the input signal $u(t) = \epsilon \text{sign}(-t)$. If we compute $y(0)$, we have:

$$\begin{aligned} y(0) &= \sum_{\tau=-\infty}^{\infty} h(\tau)u(-\tau) \\ &\leq \sum_{\tau=-\infty}^{\infty} \epsilon h(\tau)\text{sign}(h(\tau)) \\ &\leq \epsilon \sum_{\tau=-\infty}^{\infty} |h(\tau)| \end{aligned}$$

Therefore, if $\sum_{\tau=-\infty}^{\infty} |h(\tau)|$ diverges, so will $y(0)$ in the face of a bounded signal $u(t)$. $\square$

CHAPTER 4

Laplace and z-Transform

In the past chapter we have seen that LTI system respond to exponential signals (e^{st} for CT systems and z^t for DT systems) in a special way. Such functions are "eigenfunctions", meaning that the corresponding output is the same function multiplied by an eigenvalue, which is given by:

$$H(s) = \int_0^\infty h(\tau)e^{-s\tau}d\tau \qquad \text{For CT systems}$$

$$H(z) = \sum_0^\infty h(\tau)z^{-\tau} \qquad \text{For DT systems.}$$

This is the basis for a solution strategy based on the so-called Laplace and Z transform. Before delving into this, it is useful to re-cap some basic facts about complex exponentials.

4.1. Complex exponentials

An exponential function has a fundamental property:

$$e^a e^b = e^{a+b}$$
$$e^a / e^b = e^{a-b}$$
$$(e^a)^n = e^{na}.$$

The very same properties are also enjoyed by a strange complex valued function:

$$e^{j\theta} = \cos\theta + j\sin\theta,$$

which we call the Euler exponential. We can see this for the first property:

$$\begin{aligned} e^{j\theta}e^{j\psi} &= (\cos\theta + j\sin\theta)(\cos\psi + j\sin\psi) \\ &= (\cos\theta\cos\psi - \sin\theta\sin\psi) + j(\sin\theta\cos\psi + \sin\psi\cos\theta) \\ &= (\cos(\theta+\psi) + j(\sin\theta + \psi) \\ &= e^{j(\theta+\psi)}. \end{aligned}$$

It is possible define exponential with real and imaginary part:

$$\begin{aligned} e^{\sigma+j\theta} &= e^\sigma e^{j\theta} \\ &= e^\sigma(\cos\theta + j\sin\theta). \end{aligned}$$

Another important property has to do with the computation of a power of a complex number z. If we express $z = \rho e^{j\theta}$, we have $z^n = \rho^n e^{jn\theta}$.

4.2. Complex Exponential functions

Based on the definition of complex exponential, we can introduce complex exponential functions.

In the CT domain, given any complex number $s = \sigma + j\omega$, with $\sigma \in \mathbb{R}$, $\omega \in \mathbb{R}$, we define CT complex exponential the function:

$$\begin{aligned} e^{st} : \mathbb{R} \in = e^{(\sigma+j\omega)t} = \\ = e^{\sigma t} \left(\cos \omega t + j \sin \omega t\right). \end{aligned}$$

Both the real and the imaginary part of this function are sinusoidal oscillations whose amplitude is modulated by an exponential function (increasing if $\sigma > 0$ and decreasing if $\sigma < 0$).

Likewise, if we consider any complex variable z, we can express $z = \rho e^{j\theta}$. In this setting z^t is given by:

$$z^t = \rho^t e^{jt\theta}.$$

4.3. Definition of the Laplace Transform

The Laplace transform is an integral operator defined as

$$\mathcal{L}\left(f(t)\right) = F(s) = \int_0^\infty f(\tau) e^{-s\tau} d\tau.$$

The idea is to associate each signal f (which is a function of t) with a new function $F(s)$, which is a function of the complex variable s.

Example 4.1. Let us compute the Laplace transform of $\mathbf{1}(t)$.

$$\begin{aligned} \mathcal{L}\left(\mathbf{1}(t)\right) = \int_0^\infty \mathbf{1}(\tau) e^{-s\tau} d\tau = \\ = \int_0^\infty e^{-s\tau} d\tau = \\ = \frac{1}{s}\left(1 - \lim_{t\to\infty} e^{-st}\right). \end{aligned}$$

We can easily see that

$$lim_{t\to\infty} e^{-st} = \begin{cases} 0 & \mathbf{Real}\left(s\right) > 0 \\ \text{Does not converge} & \mathbf{Real}\left(s\right) \leq 0. \end{cases}$$

Therefore,

$$\mathcal{L}\left(\mathbf{1}(t)\right) = \begin{cases} \frac{1}{s} & \text{if } \mathbf{Real}\left(s\right) > 0 \\ \text{undefined} & \text{otherwise} \end{cases}$$

EXAMPLE 4.2. Let us compute the Laplace transform of $\mathbf{1}(t)e^{at}$, with $a \in \mathbb{R}$.

$$\begin{aligned}\mathcal{L}\left(\mathbf{1}(t)e^{at}\right) &= \int_0^\infty \mathbf{1}(t)e^{a\tau}e^{-s\tau}d\tau = \\ &= \int_0^\infty e^{-(s-a)\tau}d\tau = \\ &= \frac{1}{s-a}\left(1 - \lim_{t\to\infty} e^{-(s-a)t}\right).\end{aligned}$$

We can easily see that

$$lim_{t\to\infty}e^{-(s-a)t} = \begin{cases}0 & \mathbf{Real}\,(s) = \sigma > a \\ \text{Does not converge} & \mathbf{Real}\,(s) = \sigma \leq a.\end{cases}$$

Hence, the Laplace transform is given by:

$$\mathcal{L}\left(\mathbf{1}(t)e^{at}\right) = \begin{cases}\frac{1}{s-a} & \text{if } \mathbf{Real}\,(s) > a \\ \text{undefined} & \text{otherwise}\end{cases}$$

The same line of reasoning also applies to exponential functions with complex exponent: $e^{(\sigma+\jmath\omega)t}$ with $\sigma \in \mathbb{R}$ and $\omega \in \mathbb{R}$. In this case the result is $1/(s-\sigma-j\omega)$ and the convergence condition is $\mathbf{Real}\,(s) \geq \sigma$.

EXAMPLE 4.3. Let us compute the Laplace transform of the DIRAC $\delta(t)$.

$$\begin{aligned}\mathcal{L}\left(\delta(t)\right) &= \int_0^\infty \delta(t)e^{-s\tau}d\tau = \\ &= \int_0^\infty \delta(t)e^{-s0}d\tau = \\ &= \int_0^\infty \delta(t)d\tau = \\ &1\end{aligned}$$

Contrary to the examples before this transform is defined for all possible s.

Looking at the examples above, we can derive the following conclusions:

- The Laplace transform of a CT signal is a function of a complex variable s,
- The Transform is not defined for all possible values. In general, we have a region of convergence (ROC) where the Transform makes sense. For the step function $\mathbf{1}(t)$ the ROC is $\mathbf{Real}\,(s) \geq 0$. For an exponential signal $\mathbf{1}(t)e^{at}$ the ROC is $\mathbf{Real}\,(s) > a$.

There are several points that remain to be addressed. The first one is if the relation between a function and its Laplace transform is a bijection. In the affirmative case, another open point is how to invert the Laplace

transform. The third point is: what is the meaning and the practical use of this function.

4.4. Existence and Uniqueness of the Laplace transform

For the discussion on existence and uniqueness of the Laplace transform, we need the following definitions.

DEFINITION 4.4. A function $f(t)$ is said of exponential order γ if there exist a constant A such that

$$f(t) \leq Ae^{\gamma t}.$$

It is easy to show the following.

THEOREM 4.5. *Consider a function $f(t)$ and assume that: 1) $f(t)$ is continuous, 2) $f(t)$ is of exponential order γ. Then the Laplace transform:*

$$\mathcal{L}\left(f(t)\right) = \int_0^\infty f(\tau)e^{-s\tau}d\tau,$$

exists and the ROC contains the half-space $\mathbf{Real}\,(s) > \gamma$.

This theorem gives us conditions for the existence of the Laplace transform for some values of s.

The condition on invertibility requires the same definition.

DEFINITION 4.6. Two functions $f(t)$ and $g(t)$ are said almost equal, $f(t) \approx g(t)$, if $f(t)$ and $g(t)$ are equal for all t except for a set of points of null measure.

A simple way to think of two functions that are almost equal is in the following example.

EXAMPLE 4.7. Consider the two functions $f(t) = e^{3t}$ and

$$g(t) = \begin{cases} 0 & \text{if } t = 2, 4, 6, 8 \dots \\ e^{3t} & \text{otherwise.} \end{cases}$$

The two functions are almost equal because the set of isolated points given by $t = 2k$, with $k \in \mathbb{N}$ has null measure.

We are now able to state the following:

THEOREM 4.8 (Lerch's Theorem). *Suppose f(t) and g(t) are continuous except for a countable number of isolated points, and that they are of exponential order γ. Then if $\mathcal{L}\left(f(t)\right) = \mathcal{L}\left(g(t)\right)$ for all $s > \gamma$ the two functions are almost equal: $f(t) \approx g(t)$.*

From an engineering viewpoint saying that two functions are almost equal is tantamount to saying that they are the same, therefore with a slight abuse of notation we will write $f(t) = g(t)$.

4.4.1. Inverse Laplace Transform. The importance of Lerch's theorem is lies in the fact that under reasonable and mild conditions the Laplace transform is actually invertible. The inverse is given by:

$$f(t) = \mathcal{L}^{-1}\left(F(s)\right) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} F(s)e^{st}ds,$$

where $\mathbf{Real}\,(s) = \sigma$ belongs to the ROC. This complex integral is generally difficult to compute and we will find more manageable means to compute the inverse transform. However, the expression is quite insightful, because it reveals that the Laplace transform is actually a decomposition of a vector $f(t)$ using a basis given by the family of exponential functions, with $F(s)$ being in some sense the coordinate related to the basis element e^{st}.

4.4.2. A first application of the Laplace transform. As we discussed above, the Laplace transform $H(s)$ of the impulse response has actually a very precise meaning: it is the eigenvalue of the eigenfunctions e^{st}. The function $H(s)$ is commonly called *transfer function*. This allows us to immediate answer the query in the following example.

EXAMPLE 4.9. Suppose that a system has impulse response $1(t)e^{-t}$. Find the forced response to $e^{\alpha t}$. For the reasons that we have seen $H(s) = 1/(s+1)$ with ROC $\mathbf{Real}\,(s) > -1$. The eigenvalue is given by $H(\alpha)$ and is is defined only if $\alpha > -1$. In this case the forced response is given by $y(t) = e^{\alpha t}/(\alpha+1)$.

With a moderate effort, we can also answer the query in the following example.

EXAMPLE 4.10. Suppose that a system has impulse response $\mathbf{1}(t)e^{-3t}$. Find the forced response to $\cos 4t$. As discussed in Chapter 2 (Theorem 3.6), the response is well defined if $H(j4)$ converges, and in this case it is given by:

$$y(t) = |H(j4)|\cos(4t + \angle H(j4).$$

$H(s)$ is given by $1/(s+3)$ and its ROC is $\mathbf{Real}\,(s) > -3$. Since $j4$ is on the imaginary axis $H(j4)$ converges and the response above is well defined. We can find:

$$\begin{aligned} |H(j4)| &= \left|\frac{1}{j4+3}\right| = \\ &= \frac{|1|}{|3+j4|} = \\ &= \frac{1}{\sqrt{9+16}} = \\ &= \frac{1}{5}, \end{aligned}$$

and

$$\begin{aligned}\angle H(j4) = \angle \frac{1}{j4+3} = \\ = \angle 1 - \angle 3 + 4j \\ = -\arctan 4/3.\end{aligned}$$

4.5. Properties of the Laplace transform

We have seen expressions for both the direct and the inverse Laplace transform. The easiest way to accomplish these tasks is not the application of these formulas, but the use of a number of properties that we will discuss below and the knowledge of the expressions of a few "elementary" transforms. In this section we will quickly review some of the properties.

4.5.1. Linearity. Given the linearity of the integral operator, it is straightforward to prove the following.

THEOREM 4.11. *Let* $F_1(s) = \mathcal{L}(f_1(t))$ *and* $F_2(s) = \mathcal{L}(f_2(t))$. *Then,*

$$\mathcal{L}(h_1 f_1(t) + h_2 f_2(t)) = h_1 F_1(s) + h_2 F_2(s).$$

We immediately see an application of this result.

EXAMPLE 4.12. Let us compute the Laplace transform of $\cos 3t$.

$$\begin{aligned}\mathcal{L}(\cos(3t)) &= \mathcal{L}\left(\frac{e^{j\omega t} + e^{-j\omega t}}{2}\right) \\ &= \frac{1}{2}\left(\mathcal{L}\left(e^{j\omega t} + e^{-j\omega t}\right)\right) \\ &= \frac{1}{2}\left(\frac{1}{s - j\omega} + \frac{1}{s + j\omega}\right) \\ &= \frac{1}{2}\left(\frac{2s}{s^2 - (j\omega)^2}\right) \\ &= \frac{s}{s^2 + \omega^2}.\end{aligned}$$

4.5.2. Time Shifting. This property applies when we delay or anticipate a causal function.

THEOREM 4.13. *Let* $F(s) = \mathcal{L}(\mathbf{1}(t)f(t))$. *Then* $F(s)e^{-st_0} = \mathcal{L}(\mathbf{1}(t - t_0)f(t - t_0))$.

PROOF. We can prove this by a simple change of variable:

$$\begin{aligned}
\mathcal{L}\left(\mathbf{1}(t-t_0)f(t-t_0)\right) &= \int_0^\infty \mathbf{1}(t-t_0)f(t-t_0)e^{-st}dt \\
&= \int_0^\infty \mathbf{1}(t-t_0)f(t-t_0)e^{-st}e^{-st_0}e^{st_0}dt \\
&= e^{-st_0}\int_0^\infty \mathbf{1}(t-t_0)f(t-t_0)e^{-s(t-t_0)}dt \\
&= e^{-st_0}\int_{-t_0}^\infty \mathbf{1}(t')f(t')e^{-st'}dt' \\
&= e^{-st_0}\int_0^\infty \mathbf{1}(t')f(t')e^{-st'}dt' \\
&= e^{-st_0}F(s).
\end{aligned}$$

□

This property is shown in the following example.

EXAMPLE 4.14. Find the Laplace transform of $f(t)$ given by the following expression:

$$f(t) = \begin{cases} 1 & \text{If } t \in [0,\, 1] \\ 0 & \text{otherwise.} \end{cases}$$

We can easily see that $f(t) = \mathbf{1}(t) - \mathbf{1}(t-1)$. Hence,

$$\begin{aligned}
\mathcal{L}\left(f(t)\right) &= \mathcal{L}\left(\mathbf{1}(t)\right) - \mathcal{L}\left(\mathbf{1}(t-1)\right) \\
&= \frac{1}{s} - \frac{e^{-s}}{s} \\
&= \frac{1-e^{-s}}{s}.
\end{aligned}$$

4.5.3. Shifting in the Laplace domain. The dual property of time shifting applies if we apply a shift in the Laplace complex variable, as shown in the following.

THEOREM 4.15. *Left* $F(s) = \mathcal{L}\left(f(t)\right)$. *Then* $F(s-s_0) = \mathcal{L}\left(f(t)e^{s_0 t}\right)$.

PROOF. Using the formula for Laplace transform, we have:

$$\begin{aligned}
\mathcal{L}\left(f(t)e^{s_0 t}\right) &= \int_0^\infty f(\tau)e^{s_0\tau}e^{-s\tau}d\tau \\
&= \int_0^\infty f(\tau)e^{-(s-s_0)\tau}d\tau \\
&= F(s-s_0),
\end{aligned}$$

where the last step derives from the application of the definition of Laplace transform. □

EXAMPLE 4.16. Suppose we know that $\mathcal{L}(\mathbf{1}(t)) = 1/s$, then it is immediate to deduct that $\mathcal{L}\left(\mathbf{1}(t)e^{at}\right) = 1/(s-a)$. Likewise, if we know that $\mathcal{L}(\cos\omega t) = s/(s^2+\omega^2)$, it is immediate to deduct that $\mathcal{L}\left(e^{at}\cos\omega t\right) = \frac{s-a}{(s-a)^2+\omega^2}$.

4.5.4. Time Scaling. It is interesting to see what happens if we expand or shrink the time scale.

THEOREM 4.17. *Let $F(s) = \mathcal{L}(f(t))$ and let $a \in \mathbb{R}^+$. Then $\mathcal{L}(f(at)) = \frac{1}{a}F(s/a)$.*

PROOF. We can apply the definition:

$$\begin{aligned}\mathcal{L}(f(at)) &= \int_0^\infty f(at)e^{-st}dt \\ &= \int_0^\infty f(t')e^{-st'/a}\frac{dt'}{a} \\ &= \frac{1}{a}F(s/a).\end{aligned}$$

□

EXAMPLE 4.18. Suppose we know that $\mathcal{L}(\cos t) = s/(s^2+1)$ then

$$\begin{aligned}\mathcal{L}(\cos\omega t) &= \frac{1}{\omega}\frac{s/\omega}{s^2/\omega^2+1} \\ &= \frac{s}{s^2+\omega^2}.\end{aligned}$$

4.5.5. Convolution. Now we come to one of the most important properties of the Laplace Transform for its implications on Linear Systems.

THEOREM 4.19. *Let $f(t)$ and $h(t)$ be two causal functions and let $\mathcal{L}(f(t)) = F(s)$ and $\mathcal{L}(h(t)) = H(s)$. Then,*

$$\mathcal{L}(f(t) * h(t)) = F(s)H(s).$$

PROOF. Let $g(t) = f(t) * h(t)$. We have

$$\begin{aligned}\mathcal{L}(g(t)) &= \int_0^\infty e^{-st} f(t) * h(t)dt \\ &= \int_0^\infty e^{-st}\left(\int_0^t h(\tau)f(t-\tau)\tau\right)dt \\ &= \int_0^\infty \int_0^t e^{-st}h(\tau)f(t-\tau)d\tau dt\end{aligned}$$

The integration is carried out in the triangle $0 \le \tau \le t$. We can change the order of integration:

$$\mathcal{L}(g(t)) = \int_{\tau=0}^\infty \int_{t=\tau}^\infty e^{-st}h(\tau)f(t-\tau)dtd\tau.$$

Let us make the following change of variable $\bar{t} = t - \tau$.

$$\begin{aligned}\mathcal{L}(g(t)) &= \int_{\tau=0}^{\infty}\int_{t=0}^{\infty} e^{-s(\bar{t}+\tau)t}h(\tau)f(\bar{t})d\bar{t}d\tau \\ &= \left(\int_{\tau=0}^{\infty} e^{-s\tau}h(\tau)d\tau\right)\left(\int_{\tau=0}^{\infty} e^{-s\bar{t}}f(\bar{t})d\bar{t}\right) \\ &= F(s)H(s).\end{aligned}$$

□

We have just discovered a key fact: convolution in the time domain becomes algebraic product in the Laplace domain. This suggests the following procedure to find the forced response to any signal $u(t)$.

(1) compute the *Transfer Function* $H(s) = \mathcal{L}(h(t))$,
(2) compute $U(s)$,
(3) compute the inverse transform of $H(s)U(s)$.

The practical execution of these steps requires: 1. a method for the computation of $H(s)$ given a system described by a differential equation, 2. an effective strategy to compute the inverse transform. Such problems will be addressed below.

4.5.6. Differentiation rule. The computation of the Laplace transform of the derivative is key to the practical solution of differential equations.

THEOREM 4.20. *Let $F(s) = \mathcal{L}(f(t))$. Then,*

$$\mathcal{L}\left(\frac{df(t)}{dt}\right) = sF(s) - f(0).$$

PROOF. We can apply integration by part to the definition of the Laplace transform.

$$\begin{aligned}\mathcal{L}\left(\frac{df(t)}{dt}\right) &= \int_0^{\infty}\frac{df(t)}{dt}e^{-st}dt \\ &= e^{-st}f(t)|_0^{\infty} - \int_0^{\infty} f(t)(-se^{-st})dt \\ &= \lim_{t\to\infty} e^{-st}f(t) - f(0) + s\int_0^{\infty} f(t)e^{-st}dt \\ &= sF(s) - f(0),\end{aligned}$$

the last step is justified because in the ROC we must have $\lim_{t\to\infty} e^{-st}f(t) = 0$. □

We now have a clear avenue to the solution of any differential equation. Consider the following example.

EXAMPLE 4.21. Consider the RC network in discussed in Example 2.1. Set $u(t) = V_{in}(t)\mathbf{1}(t)$ and $y(t) = V_c(t)$. As we now its evolution is described

by the differential equation:

$$\dot{y} = -\frac{y}{RC} + \frac{u(t)}{RC}.$$

Let $\tau = RC$. We can apply the properties of the Laplace transform as follows:

$$\begin{aligned}
\mathcal{L}(\dot{y}) &= \mathcal{L}\left(-\frac{y}{RC} + \frac{u(t)}{RC}\right) \\
sY(s) - y(0) &= -\mathcal{L}\left(-\frac{y}{RC}\right) + \mathcal{L}\left(\frac{u(t)}{RC}\right) \\
sY(s) - y(0) &= - - \frac{Y(s)}{\tau} + \frac{U(s)}{\tau} \\
Y(s) &= \frac{U(s)}{\tau(s+\frac{1}{\tau})} + \frac{y(0)}{s+\frac{1}{\tau}} = \\
Y(s) &= \frac{1}{\tau s(s+\frac{1}{\tau})} + \frac{y(0)}{s+\frac{1}{\tau}}.
\end{aligned}$$

We can observe a few facts of interest.

- The decomposition between forced evolution ($\frac{1}{\tau s(s+\frac{1}{\tau})}$) and free evolution ($\frac{y(0)}{s+\frac{1}{\tau}}$) comes for free.
- Given that the forced response is given by $\frac{U(s)}{\tau(s+\frac{1}{\tau})}$ and that we know, from the convolution rule, that $Y(s) = H(s)U(s)$, where $H(s)$ is the transfer function, we can deduct that $H(s) = \frac{1}{\tau(s+\frac{1}{\tau})}$
- We can further write

$$\begin{aligned}
\frac{1}{\tau s(s+\frac{1}{\tau})} &= \frac{1}{\tau}\left(\frac{\tau}{s} - \frac{\tau}{s+\frac{1}{\tau}}\right) \\
&= \frac{1}{s} - \frac{1}{s+\frac{1}{\tau}}
\end{aligned}$$

Under our assumptions, the Laplace transform is invertible. Hence, we can write:

$$y(t) = \mathbf{1}(t)(1 - e^{t/\tau}) + y(0)e^{-t/\tau}.$$

We can apply the differentiation recursively. For the case of the second derivative:

$$\mathcal{L}\left(\mathfrak{D}^2 f(t)\right) = s\mathcal{L}\left(\mathfrak{D}f(t)\right) - \mathfrak{D}f(0) = s^2F(s) - sf(0) - \mathfrak{D}f(0).$$

In the general case:

$$\mathcal{L}\left(\mathfrak{D}^n f(t)\right) = s^nF(s) - s^{n-1}f(0) - s^{n-2}\mathfrak{D}f(0) - \ldots \mathfrak{D}^{n-1}f(0).$$

4.5.7. Integration Rule. The Laplace transform of the integral of a function is easily derived, as shown in the following:

THEOREM 4.22. *Le* $F(s) = \mathcal{L}(f(t))$*, then*

$$\mathcal{L}\left(\int_0^t f(\tau)d\tau\right) = \frac{F(s)}{s}.$$

PROOF. We can simply observe that $\int_0^t f(\tau)d\tau = f(t) * \mathbf{1}(t)$ and apply the convolution rule taking into account that $\mathcal{L}(\mathbf{1}(t)) = \frac{1}{s}$. □

The convolution, the integration and the differentiation rule simply tell us that integral and differential operations in the time domain become simple algebraic operations in the Laplace domain.

4.5.8. Final Value and Initial Value. There is much we can read from the expression of the Laplace transform without computing the inverse transform. This is shown in the following:

THEOREM 4.23. *Let* $F(s)$ *be the known expression of* $\mathcal{L}(f(t))$*. Then: I) if* $\lim_{t\to 0} f(t)$ *exists then* $\lim_{t\to 0} f(t) = \lim_{s\to\infty} sF(s)$*, II) if* $\lim_{t\to\infty} f(t)$ *exists and is finite, then* $\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$*.*

PROOF. Let us first prove the first claim. From the differentiation rule, we know:

$$\mathcal{L}(\mathfrak{D}f(t)) = \int_0^\infty \frac{d}{dt}f(t)e^{-st}dt = sF(s) - f(0). \tag{4.1}$$

If we compute $\int_0^\infty \frac{d}{dt}f(t)e^{-st}dt$ for $s \to \infty$, we find:

$$\int_0^\infty \frac{d}{dt}f(t)e^{-s\infty}dt = 0,$$

Therefore, we find $f(0) = \lim_{s\to\infty} sF(s)$. Now, let us move to the second claim. If we evaluate it for $s \to 0$ of $\int_0^\infty \frac{d}{dt}f(t)e^{-st}dt$, we have:

$$\lim_{s->0}\int_0^\infty \frac{d}{dt}f(t)e^{-st}dt = \int_0^\infty \lim_{s->0}\frac{d}{dt}f(t)e^{-st}dt = \\ = \frac{d}{dt}f(t)\Big|_0^\infty = f(\infty) - f(0).$$

Consider once more Equation 4.1, we find for $s \to 0$:

$$\lim_{s\to 0} sF(s) - f(0) = f(\infty) - f(0),$$

which leads us straight to the claim. □

EXAMPLE 4.24. Suppose that the Laplace Transform of a function is given by $F(s) = \frac{s}{s(s+2)}$. Then we have $f(0) = \lim_{s\to\infty} sF(s) = 1$, and $f(\infty) = \lim_{s\to 0} sF(s) = 0$.

4.5.9. Differentiation in the Laplace domain. The dual property for the differentiation in time domain is shown in the following.

THEOREM 4.25. *Let* $\mathcal{L}(f(t)) = F(s)$. *Then* $\mathcal{L}(-tf(t)) = dF(s)/ds$.

PROOF. We can write:

$$\begin{aligned}\frac{dF(s)}{ds} &= \frac{d}{ds}\int_0^\infty f(t)e^{-st}dt \\ &= \int_0^\infty \frac{df(t)e^{-st}}{ds}dt \\ &= \int_0^\infty f(t)\frac{de^{-st}}{ds}dt \\ &= \int_0^\infty (-t)\cdot f(t)e^{-st}dt \\ &= \mathcal{L}(-tf(t))\end{aligned}$$

□

We can generalise the result in the following corollary:

COROLLARY 4.26. Let $\mathcal{L}(f(t)) = F(s)$. Then $\mathcal{L}((-1)^n t^n f(t)) = d^n F(s)/ds^n$.

PROOF. It descends from an iterative application of Theorem 4.25. □

4.6. Inversion of the Laplace Transform

The application of the properties described in the previous section leads to the following consideration.

FACT 4.27. Consider a CT LTI system described by the differential equation

$$\mathfrak{D}^n y(t) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i y(t) + \sum_{j=0}^{p} \beta_j \mathfrak{D}^j u(t), \tag{4.2}$$

with $p \leq n$ and initial conditions:

$$y(0), \mathfrak{D}y(0), \ldots, \mathfrak{D}^{n-1}y(0), \ldots, \mathfrak{D}^p u(0), \ldots, \mathfrak{D}u(0).$$

Then the application of the differentiation rule leads to

$$Y(s) = Y_{\text{forced}}(s) + Y_{\text{free}}(s) \qquad \text{with}$$

$$Y_{\text{forced}}(s) = \frac{\sum_{j=0}^{p} \beta_j s^j}{s^n - \sum_{i=0}^{n-1} \alpha_i s^i} U(s)$$

$$Y_{\text{free}}(s) = \frac{N_0(s)}{s^n - \sum_{i=0}^{n-1} \alpha_i s^i}$$

where $N_0(s)$ is a polynomial of degree $n-1$ whose coefficients are functions for the initial conditions.

Looking at the form of $Y_{\text{forced}}(s)$, we also conclude that the expression of the transfer functions $H(s)$ (i.e., the transform of the impulse response) is given by:

$$H(s) = \mathcal{L}(h(t)) = \frac{\sum_{j=0}^{p} \beta_j s^j}{s^n - \sum_{i=0}^{n-1} \alpha_i s^i}.$$

Observing that: 1. for standard functions $u(t)$ its Laplace Transform is given itself by a fraction of polynomials, 2. we operate in conditions where the uniqueness of the inverse Laplace Transform is guaranteed (except for a number of isolated points), we conclude that *the solution of a differential equation of a CT LTI system requires the inversion of two Laplace functions* given by fractions of polynomial in the s variable.

4.6.1. Inversion of fractions of polynomials. The general expression of fraction of polynomials can be written as:

$$A\frac{s^p + a_{p-1}s^{p-1} + a_{p-2}s^{p-2} + \dots a_0}{s^n + b_{n-1}s^{n-1} + \dots + b_0},$$

where $p \leq n$ for the causality of the system. A fundamental result of algebra is the following:

THEOREM 4.28. *A polynomial of degree n with complex coefficients has exactly n complex roots.*

This allows us to write the fraction of polynomials as

$$A\frac{(s-z_1)(s-z_2)\dots(s-z_p)}{(s-p_1)(s-p_2)\dots(s-p_n)}.$$

The roots z_i of the numerator are called *zeros* while the roots p_i of the denominator are called *poles*.

We now consider different cases.

4.6.1.1. *The case of real and distinct poles.* The first case we consider is when poles are real and distinct. In this case it is possible to see that our function can be expressed as

$$\begin{aligned} F(s) &= A\frac{(s-z_1)(s-z_2)\dots(s-z_p)}{(s-p_1)(s-p_2)\dots(s-p_n)} \\ &= \frac{A_1}{s-p_1} + \frac{A_2}{s-p_2} + \dots + \frac{A_n}{s-p_n} \end{aligned}$$

Each term $\frac{A_i}{s-p_i}$ is said a *partial fraction* and this is called partial fraction expansion. We can show the following:

PROPOSITION 4.29. Consider a function

$$F(s) = A\frac{(s-z_1)(s-z_2)\dots(s-z_p)}{(s-p_1)(s-p_2)\dots(s-p_n)}$$

with distinct and real roots. Then the coefficients of its partial fraction expansion

$$F(s) = \frac{A_1}{s-p_1} + \frac{A_2}{s-p_2} + \ldots + \frac{A_n}{s-p_n},$$

are given by:

$$A_i = F(s)(s-p_i)|_{s=p_i}.$$

PROOF. Without loss of generality we show the proposition for A_1. We can write

$$F(s) = A\frac{(s-z_1)(s-z_2)\ldots(s-z_p)}{(s-p_1)(s-p_2)\ldots(s-p_n)} = \frac{A_1}{s-p_1} + \frac{A_2}{s-p_2} + \ldots + \frac{A_n}{s-p_n}$$

and multiply both sides by $(s-p_1)$.

$$F(s)(s-p_1) = A_1 + \frac{A_2(s-p_1)}{s-p_2} + \ldots + \frac{A_n(s-p_1)}{s-p_n}.$$

If we evaluate the result at $s = p_1$, each of the terms $\frac{A_j(s-p_1)}{s-p_j}$ is null (being the roots distinct). Therefore, the right hand side produces A_1. The right hand side is finite because the term $s-p_1$ and the numerator cancels the *only* $s-p_1$ term at the denominator. □

EXAMPLE 4.30. Consider the system:

$$\ddot{y} = -5\dot{y} - 4y - 4\dot{u} + u.$$

We want to study the system's evolution for $u(t) = \mathbf{1}(t)$, $y(0) = 1$, $\dot{y}(0) = 0$, $u(0) = 2$. Computing the Laplace transform we get:

$$\begin{aligned}
s^2Y(s) - sy(0) - \dot{y}(0) &= -5sY(s) - 5y(0) - 4Y(s) - 4sU(s) + 4u(0) + U(s) \\
\left(s^2 + 5s + 4\right)Y(s) &= (1-4s)\,U(s) + sy(0) - 5y(0) + \dot{y}(0) + 4u(0) \\
Y(s) &= \frac{1-4s}{s^2+5s+4}U(s) + \frac{sy(0) + 5y(0) + \dot{y}(0) + 4u(0)}{s^2+5s+4} \\
Y(s) &= \frac{1-4s}{s^2+5s+4}\frac{1}{s} + \frac{s+5+8}{s^2+5s+4}
\end{aligned}$$

Observe that $s^2 + 5s + 4 = (s+4)(s+1)$ and

$$Y(s) = \frac{1-4s}{s(s+4)(s+1)} + \frac{s+13}{(s+4)(s+1)}$$

If we compute the partial fraction expansion, we find:

$$Y(s) = \frac{A_1}{s} + \frac{A_2}{s+1} + \frac{A_3}{s+4} + \\ + \frac{B_1}{s+1} + \frac{B_2}{s+4}$$

where

$$A_1 = \left.\frac{1-4s}{(s+4)(s+1)s} s\right|_{s=0} = \frac{1}{4}$$
$$A_2 = \left.\frac{1-4s}{(s+4)(s+1)s}(s+1)\right|_{s=-1} = \frac{-5}{3}$$
$$A_3 = \left.\frac{1-4s}{(s+4)(s+1)s}(s+4)\right|_{s=-4} = \frac{17}{12}$$
$$B_1 = \left.\frac{s+13}{(s+4)(s+1)}(s+1)\right|_{s=-1} = 4$$
$$B_2 = \left.\frac{s+13}{(s+4)(s+1)}(s+4)\right|_{s=-4} = -3.$$

and

$$y(t) = \mathbf{1}(t)\left(\frac{1}{4} - \frac{5}{3}e^{-t} + \frac{17}{12}e^{-3t}\right) + \\ + \mathbf{1}(t)\left(4e^{-t} - 3e^{-3t}\right).$$

4.6.1.2. *The case of complex conjugate poles.* Before considering the case of complex poles it is useful to state the following results:

LEMMA 4.31. *Consider a polynomial $P(s)$ in a complex variable s with real coefficient. Then $P(\overline{s}) = \overline{P(s)}$.*

PROOF. We can write:

$$\begin{aligned} P(\overline{p}) &= \overline{p}^n + a_{n-1}\overline{p}^{n-1} + a_{n-2}\overline{p}^{n-2} + \ldots + a_0 \\ &= \overline{p}^n + \overline{a_{n-1}p}^{n-1} + \overline{a_{n-2}p}^{n-2} + \ldots + \overline{a_0} \\ &= \overline{p^n + a_{n-1}p^{n-1} + a_{n-2}p^{n-2} + \ldots + a_0} \qquad = \overline{P(p)}. \end{aligned}$$

where the first step is applicable because the coefficients a_i are real and are not affected by the conjugation and the second one because the conjugate of a sum is the sum of the conjugates. □

THEOREM 4.32. *Consider a polynomial in a complex variable s with real coefficient. If $s = p$ is a root of the polynomial, then also its conjugate $s = \overline{p}$ is.*

PROOF. Consider the polynomial

$$P(s) = s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \ldots + a_0,$$

with $a_i \in \mathbb{R}$. If p is a root, then

$$P(p) = p^n + a_{n-1}p^{n-1} + a_{n-2}p^{n-2} + \ldots + a_0 = 0.$$

In view of Lemma 4.31 we have: $\overline{P(p)} = P(\overline{p})$. If we now observe that $P(p) = 0 \implies \overline{P(p)} = 0$ our claim easily follows. □

With this result in mind let us consider the case of fractions of polynomial with real coefficients. This situation always occurs when dealing with ordinary differential equations. Without loss of generality we will consider a single pair of complex conjugate roots. We can show the following:

PROPOSITION 4.33. Consider the following function:

$$F(s) = A\frac{(s - z_1)(s - z_2)\ldots(s - z_p)}{(s - p_1)(s - p_2)\ldots(s - p_n)}$$

given by a ratio of complex polynomials with real coefficients. Assume that p_1 and $\overline{p_1}$ are a pair of complex conjugate poles.

$$\begin{aligned} F(s) &= \frac{n(s)}{d(s)} \\ &= \frac{n(s)}{(s - p_1)(s - p_1^*)\ldots}. \end{aligned}$$

Let

$$F(s) = \frac{A_1}{s - p_1} + \frac{{A_1}'}{s - \overline{p_1}} + \ldots$$

be its partial fraction expansion. Then $A_1' = \overline{A_1}$. In other words the coefficient of the partial fraction related to the complex conjugate of the pole is the complex conjugate of the partial fraction related to the pole.

PROOF. Proposition 4.35 holds independently of whether poles are real or complex. Therefore, we can write:

$$A_1' = \lim_{s \to \overline{p_1}} \frac{n(s)}{d(s)}(s - \overline{p_1}).$$

The denominator $d(s)$ can be written as $d(s) = (s - p_1)(s - \overline{p_1})d_1(s)$. It is easy to see that $d_1(s)$ is a polynomial with real coefficients. By Applying

Lemma 4.31, we find

$$\begin{aligned}
\overline{A_1'} &= \overline{\frac{n(s)}{d(s)}(s-\overline{p_1})\Big|_{s=\overline{p_1}}} \\
&= \frac{n(\overline{p_1})}{d_1(\overline{p_1})(\overline{p_1}-p_1)} \\
&= \frac{\overline{n(p_1)}}{\overline{d_1(p_1)}(-2j\mathbf{Imag}\,(p_1))} \\
&= \overline{\frac{n(p_1)}{d_1(p_1)(2j\mathbf{Imag}\,(p_1))}} \\
&= \overline{A_1}.
\end{aligned}$$

□

An important consequence of the above can be seen when we compute the inverse transform of the two partial fractions related to a complex conjugate pair. If

$$F(s) = \frac{A_1}{s-p_1} + \frac{{A_1}'}{s-\overline{p_1}} + F_1(s),$$

with $p_1 = \sigma_1 + j\omega_1$, then

$$\begin{aligned}
\mathcal{L}^{-1}\left(F(s)\right) &= \mathcal{L}^{-1}\left(\frac{A_1}{s-p_1} + \frac{{A_1}'}{s-\overline{p_1}} + F_1(s)\right), \\
&= \mathcal{L}^{-1}\left(\frac{A_1}{s-p_1}\right) + \mathcal{L}^{-1}\left(\frac{{A_1}'}{s-\overline{p_1}}\right) + \mathcal{L}^{-1}\left(F_1(s)\right) \\
&= \mathcal{L}^{-1}\left(\frac{A_1}{s-p_1}\right) + \mathcal{L}^{-1}\left(\frac{\overline{A_1}}{s-\overline{p_1}}\right) + \mathcal{L}^{-1}\left(F_1(s)\right) \\
&= \mathbf{1}(t)\left(A_1 e^{p_1 t} + \overline{A_1} e^{\overline{p_1} t}\right) + f_1(t) \\
&= \mathbf{1}(t)\left(A_1 e^{p_1 t} + \overline{A_1 e^{p_1 t}}\right) + f_1(t) \\
&= 2\mathbf{1}(t)\mathbf{Real}\left(A_1 e^{p_1 t}\right) + f_1(t) \\
&= 2\mathbf{1}(t)\left|A_1\right| e^{\sigma_1 t}\cos(\omega_1 t + \angle A_1)) + f_1(t).
\end{aligned}$$

In other words, a complex conjugate pair gives rise to a sinusoidal oscillation with amplitude modulate by a real exponential.

Example 4.34. Consider the differential equation:

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

Compute the system's response to $\mathbf{1}(t)$ for null initial conditions. Neglecting the initial conditions the differential Laplace transform is given by:

$$\begin{aligned} Y(s)(s^2 - s + 1) &= U(s) \\ Y(s) &= \frac{1}{(s^2 - s + 1)s} \\ &= \frac{1}{(s - \frac{1+\sqrt{-3}}{2})(s - \frac{1-\sqrt{-3}}{2})s} \\ &= \frac{1}{(s - \frac{1+j\sqrt{3}}{2})(s - \frac{1-j\sqrt{3}}{2})s} \\ &= \frac{A_1}{s} + \frac{A_2}{s - \frac{1+j\sqrt{3}}{2}} + \frac{\overline{A_2}}{s - \frac{1-j\sqrt{3}}{2}} \end{aligned}$$

with

$$\begin{aligned} A_1 &= \left.\frac{1}{(s^2 - s + 1)}\right|_{s=0} = 1 \\ A_2 &= \left.\frac{1}{(s - \frac{1-j\sqrt{3}}{2})s}\right|_{s=\frac{1+j\sqrt{3}}{2}} \\ &= \frac{1}{j\sqrt{3}\frac{1+j\sqrt{3}}{2}} \\ &= \frac{1}{-\frac{3}{2} + j\frac{\sqrt{3}}{2}} \\ &= \frac{-\frac{3}{2} - j\frac{\sqrt{3}}{2}}{\frac{9}{4} + \frac{3}{4}} \\ &= -\frac{1}{2} - j\frac{\sqrt{3}}{6} \end{aligned}$$

Therefore we can compute the inverse as:

$$\begin{aligned} y(t) &= \mathbf{1}(t)\left(1 + 2\left|A_2\right| e^{\frac{1}{2}t}\cos(\frac{\sqrt{3}}{2}t + \angle A_2)\right) \\ |A_2| &= \sqrt{\frac{1}{4} + \frac{3}{36}} \\ &= \frac{\sqrt{3}}{3} \\ \angle A_2 &= \operatorname{atan2}(-\frac{\sqrt{3}}{6}, -\frac{1}{2}) \\ &= -0.8571 \end{aligned}$$

4.6.1.3. *The case of multiple roots.* There are cases when a pole occurs as a multiple roots of the denominator of the transfer function. For instance,

we could have $d(s)$ factored as $(s+3)^2(s^2-s+1)$. For the moment being let us focus on the case of just one multiple and real pole. Without loss of generality, let us assume that the multiple pole is p_1 and that its multiplicity is h:

$$F(s) = A\frac{n(s)}{(s-p_1)^h d_1(s)}$$

where $d_1(s)$ does not divide $(s-p_1)$.

PROPOSITION 4.35. Consider the function in Equation (4.3), where the pole p_1 is real and with multiplicity h. It is possible to come up with the following partial fraction expansion:

$$\begin{aligned} F(s) &= A\frac{n(s)}{(s-p_1)^h d_1(s)} \\ &= \frac{A_{1,1}}{s-p_1} + \frac{A_{1,2}}{(s-p_1)^2} + \ldots + \frac{A_1,h}{(s-p_1)^h} + F_1(s) \end{aligned} \tag{4.3}$$

(4.4)

where $F_1(s)$ is found using the same rules that apply to single poles as we discussed above and $A_{1,i}$ is given by:

$$A_{1,h-r} = \frac{1}{(h-r)!}\frac{d^r}{ds^r}\left[F(s)(s-p_i)^h\right]\Big|_{s=p_i}.$$

PROOF. Multiplying both sides of Equation (4.4) by $(s-p_1)^h$ we get:

$$\frac{n(s)}{d_1(s)} = A_{1,1}(s-p_1)^{h-1} + \ldots + A_{1,h-1}(s-p_1) + A_{1,h} + F_1(s)(s-p_1)^h.$$

Since neither $d_1(s)$ nor any denominator in the partial fraction expansion of $F_1(s)$ divide $(s-p_1)$, when we evaluate in $s=p_1$, we will have:

$$\frac{n(p_1)}{d_1(p_1)} = A_{1,h}.$$

Similarly, if we differentiate r times we get:

$$\begin{aligned} \frac{d^r}{ds^r}\left[\frac{n(s)}{d_1(s)}\right] =& A_{1,1}(h-1)(h-2)\ldots(h-r)(s-p_1)^{h-1-r} + A_{1,2}(h-2)\ldots(h-r-1)(s-p_1)^{h-2-r} + . \\ &\ldots + (h-r)(h-r-1)\ldots 1\cdot A_{1,h-r} + \\ &+ h(h-1)\ldots(h-r)(s-p_1)^{h-r}F_1(s) + \frac{d^r}{ds^r}\left[F_1(s)\right](s-p_1)^h. \end{aligned}$$

If we evaluate in $s=p_1$ we obtain our claim. □

Another important fact is in the following:

PROPOSITION 4.36. Let $F(s) = \frac{1}{(s-p)^h}$. Then $\mathcal{L}^{-1}\left(F(s)\right) = \frac{t^{h-1}}{h-1!}e^{pt}$.

PROOF. It descends from the application of Corollary 4.26 to $\mathcal{L}\left(\mathbf{1}(t)e^{at}\right) = \frac{1}{s-a}$ □

This leads us to:

THEOREM 4.37. *Consider the function in Equation* (4.3)*, where the pole* p_1 *is real and with multiplicity* h*. Let*

$$F(s) = \frac{A_{1,1}}{s-p_1} + \frac{A_{1,2}}{(s-p_1)^2} + \ldots + \frac{A_1, h}{(s-p_1)^h} + F_1(s)$$

then

$$\mathcal{L}^{-1}\left(F(s)\right) = \mathbf{1}(t)\left(A_{1,1}e^{p_1 t} + A_{1,2}te^{p_1 t} + A_{1,3}\frac{t^2}{2}e^{p_1 t} + \ldots + A_{1,h}\frac{t^{h-1}}{(h-1)!}e^{p_1 t}\right) + \mathcal{L}^{-1}\left(F_1(s)\right)$$

EXAMPLE 4.38. Consider the following differential equation

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

Let us compute the free evolution for $y(0) = -4$, $\dot{y}(0) = 2$. Neglecting the input function $u(t)$, the Laplace function becomes:

$$(s^2 - sy(0) - \dot{y}(0))Y(s) - (2s - 2y(0))Y(s) + Y(s) = 0,$$

which is easily written as

$$\begin{aligned} Y(s) &= \frac{sy(0) + \dot{y}(0) - 2y(0)}{s^2 - 2s + 1} \\ &= \frac{-4s + 10}{(s-1)^2} \\ &= \frac{A_{1,1}}{s-1} + \frac{A_{1,2}}{(s-1)^2} \\ A_{1,2} &= (-4s + 10)|_{s=1} = 6 \\ A_{1,1} &= \frac{d}{ds}\left[(-4s + 10)\right]\Big|_{s=1} = -4 \end{aligned}$$

As a consequence,

$$y(t) = \mathbf{1}(t)e^t\left(-4 + 6t\right).$$

The case of complex conjugate poles with multiplicity greater than 1 is not significantly different. Following the same line of reasoning as in Section We can treat each pole in the complex conjugate pair as in Section 4.6.1.2: 1. finding the coefficients of the partial fraction expansion, 2. observing that the coefficient related to the complex conjugate are complex conjugate, 3. recombining the pairs related to the partial fraction with the same power and reducing them to real functions. This is shown in the following example:

EXAMPLE 4.39. Let us compute the inverse of

$$F(s) = \frac{1}{s(s^2 - s + 1)^2}.$$

Let $p_1 = \frac{1+j\sqrt{3}}{2}$ be one of the roots of $(s^2 - s + 1)$ (the other one being its conjugate $\overline{p}_1$. We can proceed as follows:

$$\begin{aligned}
F(s) &= \frac{1}{s(s-p_1)^2(s-\overline{p_1})^2} \\
&= \frac{A_{1,1}}{s-p_1} + \frac{\overline{A_{1,1}}}{s-\overline{p_1}} + \\
&+ \frac{A_{1,2}}{(s-p_1)^2} + \frac{\overline{A_{1,2}}}{(s-\overline{p_1})^2} + \\
&+ \frac{A_2}{s} \\
A_2 &= F(s)s\big|_{s=0} = 1 \\
A_{1,2} &= F(s)(s-p_1)^2\big|_{s=p_1} = \frac{1}{p_1(p_1-\overline{p_1})^2} \\
A_{1,1} &= \frac{d}{ds}\left[F(s)(s-p_1)^2\right]\Big|_{s=p_1} \\
&= \frac{-((s-\overline{p_1})^2 + 2s(s-\overline{p}_1)}{s^2(s-\overline{p_1})^4}\Big|_{s=p_1} \\
&= \frac{-(3s-\overline{p_1})}{s^2(s-\overline{p_1})^3}\Big|_{s=p_1} \\
&= \frac{-3p_1 + \overline{p_1}}{p_1^2(p_1-\overline{p_1})^3}
\end{aligned}$$

Observing that $p_1 = e^{j\pi/3}$ and $p_1 - \overline{p_1} = 2j\sqrt{3}2 = j\sqrt{3}$, we find

$$\begin{aligned}
A_{1,2} &= -\frac{1}{\frac{1+j\sqrt{3}}{2}3} \\
&= -\frac{1}{3e^{j\pi/3}} \\
&= \frac{1}{3}e^{-j\pi/3}
\end{aligned}$$

and

$$\begin{aligned}
A_{1,1} &= \frac{e^{-j\pi/3} - 3e^{j\pi/3}}{e^{j2\pi/3}(-j3\sqrt{3})} \\
&= \frac{e^{-j\pi/3} - 3e^{j\pi/3}}{e^{j2\pi/3}e^{-j\pi/2}3\sqrt{3}} \\
&= \frac{e^{-j\pi/3} - 3e^{j\pi/3}}{e^{j\pi/6}3\sqrt{3}} \\
&= \frac{e^{-j\pi/2} - 3e^{j\pi/6}}{3\sqrt{3}} \\
&= \frac{-j - 3\frac{\sqrt{3}}{2} - \frac{3}{2}j}{3\sqrt{3}} \\
&= \frac{-3\frac{\sqrt{3}}{2} - \frac{5}{2}j}{3\sqrt{3}} \\
&= -\frac{1}{2} - \frac{5}{6\sqrt{3}}j \\
&= \frac{1}{3}\sqrt{\frac{13}{3}}e^{j\text{atan2}(-5/(6\sqrt{3}),-1/2)} \\
&= \frac{1}{3}\sqrt{\frac{13}{3}}e^{-j2.3754}
\end{aligned}$$

We can combine the results as follows:

$$\begin{aligned}
f(t) &= A_2\mathbf{1}(t) + \\
&\quad + \mathbf{1}(t)\left(A_{1,1}e^{p_1 t} + \overline{A_{1,1}}e^{\overline{p_1}t}\right) + \\
&\quad + \mathbf{1}(t)\left(A_{1,2}te^{p_1 t} + \overline{A_{1,2}}te^{\overline{p_1}t}\right) \\
&\quad = \mathbf{1}(t)\left(1 + 2\mathbf{Real}\left(A_{1,1}e^{p_1 t}\right) + 2\mathbf{Real}\left(A_{1,2}te^{p_1 t}\right)\right) = \\
&\quad = \mathbf{1}(t)\left(1 + 2\frac{1}{3}\sqrt{\frac{13}{3}}e^{1/2t}\cos\left(\frac{\sqrt{3}}{2}t - 2.3754\right) + \frac{2}{3}te^{1/2t}\cos\left(\frac{\sqrt{3}}{2}t - \pi/3\right)\right)
\end{aligned}$$

4.7. BIBO Stability for CT systems

The discussion above has lead us to the conclusion that a pole in a CT system gives rise to a complex exponential possibly multiplied by a power of t (if the root is multiple). In view of Theorem 3.10, this can be translated into the following.

THEOREM 4.40. *Consider a CT LTI system with transfer function:*

$$H(s) = \frac{n(s)}{d(s)}.$$

Assume that no cancellation between poles and zero takes place: $\nexists p : n(p) = d(p) = 0$. *The system is BIBO stable if and only if all of its poles have negative real part.*

PROOF. In view of Theorem 3.10, the system is stable if and only if $\int_0^\infty |h(t)|dt = L < +\infty$. We now start by proving sufficiency. If $H(s) = \mathcal{L}(h(t))$ has all poles with negative real part then $h(t) = \sum_{h=1}^{H} F_h h_h(t)$, where the functions $h_h(t)$ can be one of the following:

$$h_h(t) = \begin{cases} e^{p_h t} & \text{Real root (single or multiple)} \\ e^{\mathbf{Real}(p_h)t}\cos\left(\mathbf{Imag}\left(p_h\right)t + \phi_h\right) & \text{Complex single or multiple root} \\ t^{n_h}e^{p_h t} & \text{Real multiple root} \\ t^{n_h}e^{\mathbf{Real}(p_h)t}\cos\left(\mathbf{Imag}\left(p_h\right)t + \phi_h\right) & \text{Complex multiple root,} \end{cases}$$

which can be interpreted as

$$|h_h(t)| \leq H_h(t) = \begin{cases} e^{p_h t} & \text{Real root (single or multiple)} \\ e^{\mathbf{Real}(p_h)t} & \text{Complex single or multiple root} \\ t^{n_h}e^{p_h t} & \text{Real multiple root} \\ t^{n_h}e^{\mathbf{Real}(p_h)t} & \text{Complex multiple root,} \end{cases}$$

and

$$\int_0^\infty H_h(t)dt = \begin{cases} \lim_{K\to\infty} \frac{e^{p_h K}-1}{p_h} & \text{For real roots} \\ \lim_{K\to\infty} \frac{e^{\mathbf{Real}(p_h)K}-1}{\mathbf{Real}(p_h)} & \text{For complex roots} \\ \lim_{K\to\infty} \sum_{k=0}^{n_h}(-1)^{n_h-k}\frac{n_h!}{k!p_h^{n_h-k}}K^k e^{p_h K} - (-1)^{n_h}\frac{1}{p_h^{n_h}} & \text{Real multiple root} \\ \lim_{K\to\infty} \sum_{k=0}^{n_h}(-1)^{n_h-k}\frac{n_h!}{k!\mathbf{Real}(p_h)^{n_h-k}}K^k e^{\mathbf{Real}(p_h)K} - (-1)^{n_h}\frac{1}{\mathbf{Real}(p_h)^{n_h}} & \text{Complex multiple root} \end{cases}$$

If $\mathbf{Real}\left(p_h\right)$ is smaller than 0 for all poles, this leads to

$$\int_0^\infty H_h(t)dt = \begin{cases} \frac{-1}{p_h} & \text{For real roots} \\ \frac{-1}{\mathbf{Real}(p_h)} & \text{For complex roots} \\ -\frac{(-1)^{n_h}}{p_h^{n_h}} & \text{Real multiple root} \\ -\frac{(-1)^{n_h}}{\mathbf{Real}(p_h)^{n_h}} & \text{Real multiple root.} \end{cases}$$

In other words we can find a constant F that upper bounds all $H_h(t)$ for all h.

$$\int_0^\infty |h(t)|dt \leq |F_h| \int_0^\infty \sum |H_h(t)|dt \leq FM.$$

To prove necessity, Consider a single pole p_i with non negative real part exists and assume for simplicity that its multiplicity is1. Then in the partial fraction expansion of $H(s)$ we will have a term $\frac{A_i}{s-p_i}$ with non null A_i. Indeed,

$$A_i = \frac{n(p_i)}{(p_i - p_1)\dots(p_i - p_{i-1})(p_i - p_{i+1})\dots(p_i - p_n)},$$

and by our assumption $n(p_i) \neq 0$. This fraction will give rise to an exponential function in $h(t)$ whose integral grows in an unbounded way. □

EXAMPLE 4.41. Let us consider the following transfer function:

$$H(s) = \frac{s-1}{s^2+3s+2}.$$

Its poles are given by

$$p_{1,2} = \frac{-3 \pm \sqrt{9-8}}{2} = \begin{cases} -2 \\ -1 \end{cases}.$$

Both roots are negative. Hence, the system is BIBO stable.

This example is easy to analyse because the roots of a second order equation can be found using a closed-form expression. This is possible also for order 3 and 4, although the expressions are way less intuitive to deal with.

DEFINITION 4.42. The polynomial $d(s)$ at the denominator of the transfer function is said *characteristic polynomial* and the equation $d(s) = 0$ is called *characteristic equation*.

4.7.1. The Routh-Hurwitz Criterion. If we have a mere interest in system's BIBO stability the explicit computation of the roots is not required as long as we are able to compute the sign of their real part. This can be done using the Routh-Hurwitz criterion. Suppose the denominator $d(s)$ of the transfer function is:

$$a_n s^n + a_{n-1} s^{n-1} + \dots a_0.$$

Suppose that n is even. The first step is to form the Routh table as follows:

s^n	a_n	a_{n-2}	a_{n-4}	$\dots$	a_2	a_0
s^{n-1}	a_{n-1}	a_{n-3}	a_{n-5}	$\dots$	a_3	a_1
s^{n-2}	b_1	b_2	b_3	$\dots$	b_{n_2}	0
s^{n-2}	c_1	c_2	c_3	$\dots$	0	0
$\dots$	$\dots$	$\dots$				
s^0	q					

where computation of each row is made using the results of the two rows immediately above. For instance:

$$\begin{aligned}
b_1 &= \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}} \\
b_2 &= \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}} \\
b_3 &= \frac{a_{n-1}a_{n-6} - a_n a_{n-7}}{a_{n-1}} \\
&\dots \\
c_1 &= \frac{b_1 a_{n-3} - b_2 a_{n-1}}{b_1} \\
c_2 &= \frac{b_1 a_{n-5} - b_3 a_{n-1}}{b_1} \\
c_2 &= \frac{b_1 a_{n-7} - b_4 a_{n-1}}{b_1} \\
&\dots.
\end{aligned}$$

After building this table, we can state the following.

THEOREM 4.43. *Suppose that the Routh table can be built as:*

$$\begin{array}{c|cccccc}
s^n & a_n & a_{n-2} & a_{n-4} & \dots & a_2 & a_0 \\
s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & \dots & a_3 & a_1 \\
s^{n-2} & b_1 & b_2 & b_3 & \dots & b_{n_2} & 0 \\
\dots & \dots & \dots & & & & \\
s^0 & q & & & & &
\end{array}$$

Assume that the first column $(a_n, a_{n1}, b_1, c_1, \dots, q)$ does not contain 0 elements. Then the number of sign changes in the first column $(a_n, a_{n1}, b_1, c_1, \dots, q)$ of the Routh table corresponds to the number of poles in the right half of the complex plan. Moreover, a necessary and sufficient condition for all roots to be in the left a half plan is that all coefficients a_i be positive and that all the elements of the first column be positive.

EXAMPLE 4.44. Consider the classic second order equation:

$$a_2 s^2 + a_1 s + a_0.$$

The Routh table is given by

$$\begin{array}{c|cc}
s^2 & a_2 & a_0 \\
s & a_1 & 0 \\
s^0 & a_0. &
\end{array}$$

Hence the requirements that all poles are in the left half plan reduces to the fact that all coefficients have to be positive. This can be seen by considering that if the roots are negative, we can write the polynomial as

$$(s + r_1)(s + r_2) = s^2 + (r_1 + r_2)s + r_1 r_2$$

So if r_1 and r_2 are positive (meaning that the roots are negative) if and only if $(r_1 + r_2)$ and $r_1 r_2$ are positive in their turn.

EXAMPLE 4.45. We can now consider the case of polynomial of degree three.

$$a_3 s^3 + a_2 s^2 + a_1 s + a0 = 0.$$

The Routh table is given by:

$$\begin{array}{c|ccc} s^3 & a_3 & a_1 & 0 \\ s^2 & a_2 & a_0 & 0 \\ s & \frac{a_2 a_1 - a_3 a_0}{a_2} & 0 & 0 \\ s^0 & a_0 & 0 & 0 \end{array}$$

Hence, along with $a_i > 0$, BIBO stability will also require $a_2 a_1 - a_3 a_0 > 0$.

To simplify the construction of the table, we can use the following result:

PROPOSITION 4.46. If we multiply an entire row of the Routh table by a positive constant Theorem 4.43 still holds.

There are two important special cases:

(1) The first element of a row is 0, thus preventing to derive the following row (which requires the division by the first element),
(2) An entire row is 0.

Both cases reveal the presence of roots on the imaginary axis or in the positive half plan and therefore the loss of BIBO stability. Still we can use Routh criterion to identify the number of stable and unstable roots.

The first problem is solved by perturbing the 0 element (setting it to ϵ) and evaluating the number of changes and permanence both for positive and negative ϵ.

EXAMPLE 4.47. Consider the polynomial

$$s^3 + s - 1.$$

Its associated table is

$$\begin{array}{c|ccc} s^3 & 1 & 1 & 0 \\ s^2 & 0 & -1 & 0 \\ s & ??? & & \\ s^0 & ??? & & \end{array}$$

In order to compute the missing elements of the Routh table, we can introduce ϵ:

$$\begin{array}{c|ccc} s^3 & 1 & 1 & 0 \\ s^2 & \epsilon & -1 & 0 \\ s & \frac{\epsilon+1}{\epsilon} & 0 & 0 \\ s^0 & -1 & 0 & \end{array}$$

If we consider $\epsilon \to 0^+$ we can see observe one sign change in the first column. The same is we consider the changes for negative values: $\epsilon \to 0^-$. We infer the presence of one root in the right half plan and two roots in the negative half plan.

When an entire row is null, this means that two rows are proportional and that the characteristic polynomial $d(s)$ can be divided by the polynomial associated with the row immediately above the null row. The divisor polynomial is called "auxiliary polynomial" and call it $p(s)$. Having a null row also indicates that the auxiliary polynomial has roots that are symmetrical with respect to the imaginary axis. The number of changes above the line with zeros accounts for the stability of the roots of the "remainder" polynomial. One possibility to complete the construction of the Routh table is to replace the null line with $\frac{dp(s)}{ds}$ and then continue. Looking the first column of the modified Routh table, the number of sign changes is still equal to the number of roots with positive real part. From the null row down, each sign change indicates a root with positive real part. Since roots are symmetric, there is be a corresponding root in the negative half plan. All roots not accounted in this way (i.e., no sign changes) are on the imaginary axis.

EXAMPLE 4.48. Consider the polynomial

$$\begin{aligned} d(s) &= (s^2+1)(s+1)(s+3) \\ &= s^4 + 4s^3 + 4s^2 + 4s + 3 \end{aligned}$$

In this case we know up-front that the polynomial has two stable roots and a pair of imaginary roots. Let us see how the modified Roth criterion allows us to find this information. The first two rows of the Routh table are:

$$\begin{array}{c|ccc} s^4 & 1 & 4 & 3 \\ s^3 & 4 & 4 & 0 \\ \dots & & & \end{array}$$

We can divide the second row by 4 and compute the third row as

$$\begin{array}{c|ccc} s^4 & 1 & 4 & 3 \\ s^3 & 1 & 1 & 0 \\ s^2 & 3 & 3 & 0 \\ \dots & & & \end{array}$$

We divide the third row by 3 and find:

$$\begin{array}{c|ccc} s^4 & 1 & 4 & 3 \\ s^3 & 1 & 1 & 0 \\ s^2 & 1 & 1 & 0 \\ s & 0 & 0 & 0 \end{array}$$

The three positive elements of the first rows reveal two roots with negative real part. We can now complete the construction. The auxiliary polynomial is given by $s^2 + 1$ and its derivative is $2s$. Therefore we can complete the

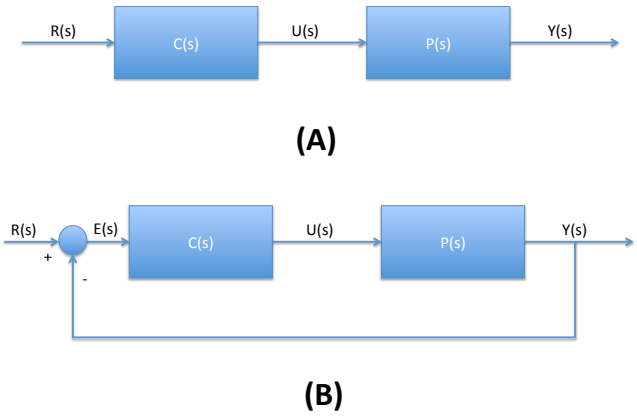


FIGURE 1. (A) series connection, (B) feedback connection

construction as:

$$\begin{array}{c|ccc} s^4 & 1 & 4 & 3 \\ s^3 & 1 & 1 & 0 \\ s^2 & 1 & 1 & 0 \\ s & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{array}$$

The two new elements are positive. This means no changes and two roots on the imaginary axis.

4.8. Use of Laplace Transform for Control Design

The Laplace transform is a very useful tool for designing systems that control other systems. This is done taking advantage of some properties of the connections of systems. The two simplest ways of connecting two systems are shown in Figure 1. For series connection (Figure 1.(A)), we know already that, given the two impulse response $c(t)$ and $p(t)$ the output is $y(t) = p(t) * c(t) * r(t)$. In the Laplace domain this property can be translated as:

$$Y(s) = P(s)C(s)R(s),$$

because convolutions correspond to products in the Laplace domain. This connection can be used in different ways as in the following examples.

EXAMPLE 4.49. Consider a system

$$P(s) = \frac{1}{s+5}.$$

Suppose we want it to track with 0 error constant reference signals. If our reference signal is $r(t) = A\mathbf{1}(t)$, this requirement could be stated as:

$$\lim_{t\to\infty} y(t) = A.$$

If make a series connection of $P(s)$ and $C(s)$, we have:

$$Y(s) = P(s)C(s)\frac{A}{s}.$$

Applying the final value theorem

$$\begin{aligned}\lim_{t\to\infty} y(t) &= \lim_{s\to 0} sP(s)C(s)\frac{A}{s}\\ &= AP(0)C(0).\end{aligned}$$

Hence, the specification requires: $C(0) = \frac{1}{P(0)} = 5$. The simplest way of having this is by obviously choosing $C(s) = 5$. Clearly, if the pole is known with some error Δ the steady state error will be affected by an error of the same amount.

Suppose, now, that we also require a faster response than the one associated with e^{-5t}, setting it, for instance, to e^{-10t}. One possibility is to choose the controller

$$C(s) = 10\frac{s+5}{s+10}.$$

Notice that gain has been chosen equal to 10 in order to have $P(0)C(0) = 1$.

As shown in the previous example, using series connection we can adjust the steady state response (up to a certain precision) and even change the dynamics of the system cancelling some of the poles with zeros in the controller. This is not possible if some of the poles are unstable, as shown next.

Example 4.50. Consider the system

$$P(s) = \frac{1}{s-5}.$$

Suppose we want it to track with 0 error constant the reference signal is $r(t) = A\mathbf{1}(t)$. The evident problem here is that the system is unstable. So our first problem is to "stabilise" it. We could think of a series connection with a controller system like $C(s) = p\frac{s-5}{s+p}$ with $p > 0$. The problem becomes evident if the pole is known with some precision. Suppose that instead of having

$$P(s) = \frac{1}{s-5+\Delta}.$$

with $\Delta << 1$. The overall response to $U(s) = \frac{1}{s}$ is given by:

$$\begin{aligned}P(s)C(s)R(s) &= p\frac{s-5}{s(s+p)(s-5+\Delta)}\\ &= \frac{5}{5-\Delta}\frac{1}{s} - \frac{p+5}{(p+5-\Delta)}\frac{1}{s+p} - p\frac{-\Delta}{(5-\Delta)(5-\Delta+p)}\frac{1}{s-5+\Delta}\end{aligned}$$

The term

$$H\frac{1}{s-5+\Delta},$$

with $H = -p\frac{-\Delta}{(5-\Delta)(5-\Delta+p)}$ is not null. This means that however small a perturbation Δ on the pole can be, it gives rise to an exponential term $e^{(5-\Delta)t}$, which diverges compromising the system's stability.

From the example above, we have seen that controlling a system by only using a cascaded connection with a controller is a solution that successfully solves only parts of the possible range of problems that a designer might encounter. In particular, compensating the error of the system by a "gain" (i.e., multiplication by a constant) can be a viable solution but it is affected adversely by the limited knowledge of the system parameters. More importantly, an unstable system cannot be made stable by simply "cancelling" the unstable poles with zeros in the controlling system. A more useful solution for this cases is the feedback connection shown in Figure 1.(B) For this connection we can write:

$$\begin{aligned} Y(s) &= P(s)U(s) \\ U(s) &= C(s)E(s) \\ E(s) &= R(s) - Y(s). \end{aligned}$$

By combining the three equations we easily find:

$$Y(s) = \frac{P(s)C(s)}{1+P(s)C(s)}R(s). \tag{4.5}$$

The equation above is called "closed–loop equation" and the term $\frac{P(s)C(s)}{1+P(s)C(s)}$ is the "closed–loop transfer function".

If we want to enforce a 0 steady state error to the step response ($U(s) = \frac{A}{s}$), we can once again apply the final value theorem:

$$\begin{aligned} \lim_{t\to\infty} y(t) &= \lim_{s\to 0} sY(s) \\ &= \lim_{s\to 0} sY(s) \\ &= \lim_{s\to 0} s\frac{P(s)C(s)}{1+P(s)C(s)}\frac{A}{s} \end{aligned}$$

Now, if $\lim_{s\to 0} P(s)C(s) = \infty$ we have that

$$\lim_{t\to\infty} y(t) = A,$$

no matter how imprecise the knowledge of $P(0)$ is. This means that *either* $P(0) = \frac{1}{s}P_1(s)$ *or* $C(0) = \frac{1}{s}C_1(s)$. To put it in other words: "if we want to use a feedback connection to follow a step function with steady state error 0, an integrator $\frac{1}{s}$ has to be found in $P(s)$ or in $C(s)$". This is a particular example of principle known as "internal model principle", whereby if we want

to generate a signal it has to be contained in the system to be controlled or we have to insert it into the loop through the controller.

The fact that $C(s)$ appears in the denominator of the closed loop transfer function allows us to shift the position of the pole. Unlike what we did with a simple pole-zero cancellation, this technique is much more reliable in terms of parametric uncertainty. This is shown in the following.

EXAMPLE 4.51. Consider the system

$$P(s) = \frac{10 + \Delta_0}{s - 5 - \Delta_1}$$

where Δ_0 and Δ_1 are bounded but unknown. Suppose we choose the feedback controller:

$$C(s) = \frac{10(s+5)}{s}.$$

Let us see whether or not it fulfils two requirements: 1. closed loop stability, 2. zero steady state error.

In order to fulfil the requirement of having zero steady state error, we need to have an integrator in the loop, which is verified in our case because $C(s)$ has a pole in $s = 0$. Therefore the requirement is met as long as the system is stable. The closed loop response is given by

$$\begin{aligned}\frac{P(s)C(s)}{1 + P(s)C(s)} &= \frac{\frac{10+\Delta_0}{s-5-\Delta_1}\frac{10(s+5)}{s}}{1 + \frac{10+\Delta_0}{s-5-\Delta_1}\frac{10(s+5)}{s}} \\ &= \frac{(10 + \Delta_0)(10(s+5))}{(s - 5 - \Delta_1)s + 10(s+5)(10 + \Delta_0)} \\ &= \frac{(10 + \Delta_0)(10(s+5))}{s^2 + (10\Delta_0 - \Delta_1 + 95)s + 500 + 50\Delta_0}\end{aligned}$$

In order to verify when the system is stable, we can apply Routh Criterion, and require that all coefficients of the characteristic equation be positive, which is verified if:

$$\begin{aligned}\Delta_0 &> -10 \\ \Delta_1 &< 10\Delta_0 + 95.\end{aligned}$$

As long as these conditions are met the system is both BIBO stable and responds with null error to a step input. We emphasise that if we use an "open loop" (series connection) scheme, it is sufficient to have small Δ_1 to destroy stability and that having a $\Delta_0 \neq 0$ determines errors in he steady state behaviour.

4.9. The z-Transform

The z-transform is the discrete time counter–part of the Laplace tranform. In the same way as the Laplace transform, the z-Tranform, provides analytical methods for the solution of a difference equation, it offers direct insight into the transient and steady state behaviour of a DT signal, and can be used to evaluate the stability of a system.

The definition of the z-Transform is the following:

$$\mathcal{Z}\left(f(t)\right) = \sum_{0}^{\infty} f(t) z^{-t}.$$

This time we associate a DT signal with a function of the complex variable z. We will find it convenient to use the polar representation: $z = \rho e^{j\phi}$.

EXAMPLE 4.52. Let us compute the z-Transform of $\mathbf{1}(t)$:

$$\begin{aligned}
\mathcal{Z}\left(\mathbf{1}(t)\right) &= \sum_{0}^{\infty} \mathbf{1}(t) z^{-t} \\
&= \sum_{0}^{\infty} z^{-t} \\
&= \lim_{H\to\infty} \sum_{0}^{H} z^{-t} \\
&= \lim_{H\to\infty} \frac{1 - z^{-H}}{1 - z^{-1}}.
\end{aligned}$$

Setting $z = \rho e^{j\theta}$, we have $z^{-H} = \rho^{-H} e^{-jH\theta}$. We have two cases:

$$\lim_{H\to\infty} z^{-H} = \begin{cases} 0 & \text{if } \rho = |z| > 1 \\ \infty & \text{if } \rho = |z| < 1 \\ e^{-jH\theta} & \text{if } \rho = |z| = 1 \end{cases}$$

$$\mathcal{Z}\left(\mathbf{1}(t)\right) = \lim_{H\to\infty} \frac{1 - z^{-H}}{1 - z^{-1}} = \begin{cases} \frac{1}{1-z^{-1}} & \text{if } |z| > 1 \\ \text{is not defined} & \text{otherwise} \end{cases}$$

EXAMPLE 4.53. Let us compute the z-Transform of $\mathbf{1}(t)a^t$:

$$\begin{aligned}\mathcal{Z}\left(\mathbf{1}(t)a^t\right) &= \sum_0^\infty \mathbf{1}(t)a^t z^{-t} \\ &= \sum_0^\infty z^{-t} \\ &= \lim_{H\to\infty} \sum_0^H \left(\frac{a}{z}\right)^t \\ &= \lim_{H\to\infty} \frac{1-\left(\frac{a}{z}\right)^H}{1-\frac{a}{z}}.\end{aligned}$$

Suppose $a > 0$. Setting $z = \rho e^{j\theta}$, we have $\left(\frac{a}{z}\right)^H = \left(\frac{a}{\rho}\right)^H e^{-jH\theta}$. If $a < 0$, we have $\left(\frac{a}{z}\right)^H = \left(\frac{a}{\rho}\right)^H e^{j\pi - jH\theta}$. This allows us to conclude:

$$\mathcal{Z}\left(\mathbf{1}(t)a^t\right) = \begin{cases} \frac{z}{z-a} & \text{if } |z| > |a| \\ \text{is not defined} & \text{otherwise} \end{cases}$$

From this discussion we can conclude the following facts.

- The z-Transform of a DT signal is a function of a complex variable z (in the same way as the Laplace transform is a function of the complex variable s).
- As for the Laplace transform, it is not typically possibile to define the z-Transform for all values of z, but only for a subset that we define Region of Convergence (ROC).
- Whereas for the Laplace transform the ROC is typically an half space ($\mathbf{Real}(s) > \alpha$) for the z-Transform it is an anulus $|z| \geq \rho$.

4.10. Existence and Uniqueness of z-Transform

On the existence of the z - Transform, we can show the following result:

THEOREM 4.54. *Consider a function $f(t)$ and assume that one of the following limits exists:*

$$\begin{aligned}R_f &= \lim_{t\to\infty} |f(t)|^{1/t} \\ R_f &= \lim_{t\to\infty} \frac{f(t+1)}{f(t)}.\end{aligned}$$

Then:

1. *the z-Transform $\mathcal{Z}(f(t))$ exisats and converges for $|z|$ $geqR_f$.*
2. *the z-Trasform is analytic, i.e., continouus and infinitely differentiable w.r.t. z, for $|z| \geq R_f$.*

Showing this result is complex and beyond the scope of this course. We only remark that the existence of a limit $R_f = \lim_{t\to\infty} |f(t)|^{1/t}$ can be rephrased as the existence of an exponential upper bound for the function

$$f(t) \leq AR_f^t,$$

for some A.

We can also show the following inverse result:

THEOREM 4.55. *If $F(z)$ and $G(z)$ are z-Transofrms of two functions $f(t)$ and $g(t)$ and if $F(z) = G(z)$ for all $|z| > R$, for some $R > 0$ then $f(t) = g(t)$ for $t = 0, 1, 2, \ldots$.*

PROOF. If $G(z) = F(z)$ for all $|z| > R$ then

$$\sum_{t=0}^{\infty} f(t)z^{-t} = \sum_{t=0}^{\infty} g(t)z^{-t} \leftrightarrow$$

$$\sum_{t=0}^{\infty} \left(f(t) - g(t)\right) z^{-t} = 0 \leftrightarrow \sum_{t=0}^{\infty} a_t w^t = 0$$

where we have set $w = 1/z$ and $a_t = f(t) - g(t)$. It is known that if we have a power series $\sum_{t=0}^{\infty} a_t w^t$ then if $\sum_{t=0}^{\infty} a_t w^t = 0$ for all $|w| \leq W$ then $a_t = 0$. Using this result, we conclude $f(t) = g(t)$. □

4.10.1. Inverse z-Transform. It is possible to show that the inverse transform of the z-Transform is defined by the following line integral.

$$f(t) = \mathcal{Z}^{-1}\left(F(z)\right) = \oint_{|z|=R} F(z)z^{t-1}dz$$

where the line integral is computed along a circle included in the ROC. As shown below, this formula is impractical and the inverse z-Transform is best computed using a different procedure. As for the Laplace transform this formula is insightful in that it reveals that the z-Transform is in fact a decomposition of a function using a basis of functions of type z^n.

4.11. Properties of the z-Transform

Not surprisingly, the z-Transform satisfies a number of properties that bear a close resemblance with those that we have studied for the Laplace transform. This is shown in the following.

THEOREM 4.56. *The z-Transform has the following properties (we will use upper-case letters for the z-Transform and lower case letters for the corresponding function. Let $X(z)$ have ROC R, $X_1(z)$ have ROC R_1 and $X_2(z)$ have ROC R_2.*

Linearity: $\mathcal{Z}\left(\alpha_1 x_1(t) + \alpha_2 x_2(t)\right) = \alpha_1 X_1(z) + \alpha_2 X_2(z)$ *(ROC $R' = R_1 \cap R_2$)*

Time-shifting: *if $x(t)$ is a causal signal and $k > 0$ then $\mathcal{Z}(x(t-k)) = z^{-k}X(z)$ (ROC $R' \supset R \cap \{0 < |z| < \infty\}$, and $\mathcal{Z}(x(t+k)) = z^k X(z) - z^k x(0) - z^{k-1}x(1) - \ldots - zx(K-1)$ (ROC $R' = R$)*
Multiplication by exponential: $\mathcal{Z}\left(z_0^t x(t)\right) = X(\frac{z}{z_0})$ *(ROC $R' = |z_0|\, R$*
Multiplication by t: $\mathcal{Z}(tx(t)) = -z\frac{dX(z)}{dz}$ *(ROC $R' = R$)*
Time Scaling: $\mathcal{Z}(x(t/t_0)) = X(z^{t_0})$ *for t_0 positive integer.*
Convolution: $\mathcal{Z}(x_1(t) * x_2(t)) = X_1(z)X_2(z)$. *(ROC $R' \supset R_1 \cap R_2$).*
Accumulation: $\mathcal{Z}\left(\sum_{\tau=0}^{t} x(\tau)\right) = \frac{z}{z-1}X(z)$ *(ROC $R' = R \cap \{|z| > 1\}$)*
Initial value: $x(0) = \lim_{z\to\infty} X(z)$
Final value: $x(\infty) = \lim_{z\to 1}(z-1)X(z)$, *if $x(\infty)$ exists.*

PROOF. The proof is quit similar to the proofs we have given for the analogous properties of the Laplace transform. Let us take two examples 1. Time shifting (with positive shift):

$$\begin{aligned}\mathcal{Z}(x(t+k)) &= \sum_{t=0}^{\infty} x(t+k)z^{-t} = \sum_{t=0}^{\infty} x(t+k)z^{-(t+k)}z^k = \\ &= z^k \sum_{t'=k}^{\infty} x(t')z^{-t'} = z^k\left(\sum_{t'=0}^{\infty} x(t')z^{-t'}\right) - x(0)z^k - x(1)z^{k-1} - \ldots zx(k-1) \\ &= z^k X(z) - x(0)z^k - x(1)z^{k-1} - \ldots zx(k-1)\end{aligned}$$

2. Convolution (case of causal signals)

$$\begin{aligned}\mathcal{Z}(x_1(t) * x_2(t)) &= \sum_{t=0}^{\infty} x_1(t) * x_2(t)z^{-t} = \sum_{t=0}^{\infty}\sum_{\tau=0}^{\infty} x_1(\tau)x_2(t-\tau)z^{-t} = \\ &= \sum_{\tau=0}^{\infty}\sum_{t=0}^{\infty} x_1(\tau)x_2(t-\tau)z^{-t} = \sum_{\tau=0}^{\infty} x_1(\tau)\sum_{t=0}^{\infty} x_2(t-\tau)z^{-t} = \\ &= \sum_{\tau=0}^{\infty} x_1(\tau)\sum_{t=0}^{\infty} x_2(t-\tau)z^{-(t-\tau)}z^{-\tau} = \sum_{\tau=0}^{\infty} x_1(\tau)z^{-\tau}\sum_{t=-\tau}^{\infty} x_2(t')z^{-t'} = \\ &= X_1(z)X_2(z)\end{aligned}$$

3. Final value theorem

$$\begin{aligned}\mathcal{Z}(x(t+1) - x(t)) &= \sum_{t=0}^{\infty}(x(t+1) - x(t))\, z^{-t} \to \\ z(X(z) - x(0)) - X(z) &= \sum_{t=0}^{\infty}(x(t+1) - x(t))\, z^{-t} \to \\ (z-1)X(z) - zx(0) &= \sum_{t=0}^{\infty}(x(t+1) - x(t))\, z^{-t}\end{aligned}$$

Computing the limit for $z \to 1$ of both sides we have

$$\lim_{z\to 1}(z-1)X(z) - zx(0) = \sum_{t=0}^{\infty}(x(t+1)-x(t)) \to$$
$$\lim_{z\to 1}(z-1)X(z) - x(0) = \lim_{t\to\infty} x(t+1) - x(0) \to$$
$$\lim_{z\to 1}(z-1)X(z) = \lim_{t\to\infty} x(t)$$

□

EXAMPLE 4.57. We can use the properties to build the tranform of more complex signals. For instance, consider the signal $s(t) = \mathbf{1}(t)\cos\Omega t$. We can use the properties as follows:

$$\mathcal{Z}\left(\mathbf{1}(t)\cos\Omega t\right) = \frac{1}{2}\mathcal{Z}\left(\mathbf{1}(t)e^{j\Omega t}\right) + \frac{1}{2}\mathcal{Z}\left(\mathbf{1}(t)e^{-j\Omega t}\right) = \frac{1}{2}\frac{z}{z-e^{j\Omega}} + \frac{1}{2}\frac{z}{z-e^{j\Omega}} =$$
$$= \frac{1}{2}\frac{z(z-e^{j\Omega}+z-e^{-j\Omega})}{z^2 - z(e^{j\Omega}+e^{-j\Omega})+1} = \frac{z(z-\cos\Omega)}{z^2-2z\cos\Omega+1}$$

By using the properties, we can find the following results:

z-Transform of known Signals

Signal	z-Tranform	ROC
$\delta(t)$	1	$z \in \mathbb{C}$
$\delta(t-t_0)$	z^{-t_0}	$z \in \mathbb{C}\backslash 0$
$\mathbf{1}(t)$	$\frac{z}{z-1}$	$\lvert z\rvert > 1$
t	$\frac{z}{(z-1)^2}$	$\lvert z\rvert > 1$
t^2	$\frac{z(z+1)}{(z-1)^3}$	$\lvert z\rvert > 1$
a^t	$\frac{z}{(z-a)}$	$\lvert z\rvert > \lvert a\rvert$
a^t	$\frac{z}{(z-a)}$	$\lvert z\rvert > \lvert a\rvert$
ta^t	$\frac{az}{(z-a)^2}$	$\lvert z\rvert > \lvert a\rvert$
t^2a^t	$\frac{az(z+a)}{(z-a)^3}$	$\lvert z\rvert > \lvert a\rvert$
$\cos\Omega t$	$\frac{z(z-\cos\Omega)}{z^2-2z\cos\Omega+1}$	$\lvert z\rvert > 1$
$\sin\Omega t$	$\frac{z\sin\Omega}{z^2-2z\cos\Omega+1}$	$\lvert z\rvert > 1$
$a^t\cos\Omega t$	$\frac{z(z-a\cos\Omega)}{z^2-2az\cos\Omega+a^2}$	$\lvert z\rvert > a$
$a^t\sin\Omega t$	$\frac{za\sin\Omega}{z^2-2az\cos\Omega+a^2}$	$\lvert z\rvert > a$

A more interesting application of the properties introduced above is in the following.

EXAMPLE 4.58. Consider the following difference equation.

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

Let us find $Y(z)$ for $u(t) = \mathbf{1}(t)$, $y(1) = 1$, $y(0) = -1$, $u(0) = 0$. We can find the z-Transform as follows:

$$\mathcal{Z}\left(y(t+2)\right) = \mathcal{Z}\left(3y(t+1) - 2y(t) + u(t+1) - u(t)\right).$$

The application of the time shifting rule produces.

$$z^2Y(z) - z^2y(0) - zy(1) = 3zY(z) - 3zy(0) - 2Y(z) + zU(z) - zu(0) - 3U(z)$$

which can be written as

$$\begin{aligned} Y(z) &= \frac{U(z)(z-3)}{z^2-3z+2} + \frac{z^2y(0) + z(y(1) - 3y(0) - u(0))}{z^2-3z+2} \\ &= \frac{z(z-2)}{(z-1)(z^2-3z+2)} + \frac{z^2y(0) + z(y(1) - 3y(0) - u(0))}{z^2-3z+2}. \end{aligned}$$

4.12. Inverse z-Transform

As shown by Example 4.58, also for LTI DT systems the evolution of the system is compounded by a free evolution and by a forced evolution. In both cases we have to deal with a ratio of polynomial with the numerator that is typically proportional to z. One of the possible ways to deal with this case is by using the same technique (partial fraction expansion) that we have used for the Laplace transform. This is shown in the following sequence of examples.

Example 4.59. Let us compute the free evolution Example 4.58 for $y(1) = 1$, $y(0) = -1$, $u(0) = 0$.

$$\begin{aligned} Y(z) &= \frac{z^2y(0) + z(y(1) - 3y(0) - u(0))}{z^2-3z+2} \\ &= \frac{-z^2+4z}{(z-2)(z-1)} \end{aligned}$$

It is convenient to divide by z and then proceed with partial fraction expansion:

$$\frac{Y(z)}{z} = \frac{-z+4}{(z-2)(z-1)} = \frac{2}{z-2} - \frac{3}{z-1}$$

and

$$\begin{aligned} Y(z) &= \frac{2z}{z-2} - \frac{3z}{z-1} \\ y(t) &= \mathbf{1}(t)\left(2 \cdot 2^t - 3\right). \end{aligned}$$

Example 4.60. Let us compute the forced evolution of Example 4.58 for $u(t) = \mathbf{1}(t)$.

$$Y(z) = \frac{z(z-3)}{(z-1)^2(z-2)}.$$

It is convenient to divide by z and then proceed with partial fraction expansion:

$$\begin{aligned}\frac{Y(z)}{z} &= \frac{(z-3)}{(z-1)^2(z-2)} = \\ &= \frac{A_{1,1}}{(z-1)^2} + \frac{A_{1,2}}{z-1} + \frac{A_2}{z-2} = \\ &= \frac{A_{1,1}}{(z-1)^2} + \frac{A_{1,2}}{z-1} + \frac{A_2}{z-2} = \\ &= \frac{2}{(z-1)^2} + \frac{1}{z-1} - \frac{1}{z-2} =\end{aligned}$$

which leads to

$$\begin{aligned}Y(z) &= \frac{2z}{(z-1)^2} + \frac{z}{z-1} - \frac{z}{z-2} \\ y(t) &= \mathbf{1}(t)\left(t+1-2^t\right)\end{aligned}$$

EXAMPLE 4.61. Let us compute the forced step response of the following:

$$y(k+3) + 0.2y(k+2) - 0.12y(k+1) + 0.04y(k) = u(k).$$

The z-Transform produces:

$$z^3Y(z) + 0.2z^2Y(z) - 0.12zY(z) + 0.04Y(z) = U(z)$$

$$Y(z) = \frac{1}{z^3 + 0.2z^2 - 0.12z + 0.04}U(z) = \frac{1}{z^3 + 0.2z^2 - 0.12z + 0.04}\frac{z}{z-1}$$

The partial fraction expansion is as follows:

$$\begin{aligned}\frac{Y(z)}{z} &= \frac{1}{(z+0.5)(z-1)(z-0.2-0.2j)(z-0.2+0.2j)} \\ &= \frac{1}{(z+0.5)(z-1)(z-0.2-0.2j)(z-0.2+0.2j)} \\ &= \frac{A_1}{z-1} + \frac{A_2}{z+0.5} + \frac{A_3}{z-0.2-0.2j} + \frac{\overline{A_3}}{z-0.2+0.2j}\end{aligned}$$

with

$$\begin{aligned}A_1 &= \frac{1}{(1+0.5)(1-0.2-0.2j)(1-0.2+0.2j)} &&= 0.9804 \\ A_2 &= \frac{1}{(-0.5-1)(-0.5-0.2-0.2j)(-0.5-0.2+0.2j)} &&= -1.2579 \\ A_3 &= \frac{1}{(0.5+0.2+0.2j)(0.2+0.2j-1)((0.2+0.2j-0.2+0.2j)} = \\ &= \frac{1}{0.008-0.24j} = \frac{0.08+0.24j}{0.0577} = 0.1386 + 4.1594j.\end{aligned}$$

Therefore,

$$y(t) = \mathbf{1}(t)\left(A_1 + A_2(-0.5)^t + A_3(0.2+0.2j)^t + \overline{A_3}(0.2-0.2j)^t\right).$$

Since

$$0.2 + 0.2j = 0.2828e^{j\frac{\pi}{4}}$$
$$0.2 - 0.2j = 0.2828e^{-j\frac{\pi}{4}}$$

we have

$$\begin{aligned} y(t) &= \mathbf{1}(t)\left(A_1 + A_2(-0.5)^t + A_3 0.2828^t e^{j\frac{\pi}{4}t} + \underline{A_3} 0.2828^t e^{-j\frac{\pi}{4}t}\right) \\ &= \mathbf{1}(t)\left(A_1 + A_2(-0.5)^t + |A_3|\, 0.2828^t \left(e^{j\frac{\pi}{4}t + j\angle A_3} + e^{-j\frac{\pi}{4}t - j\angle A_3}\right)\right) \\ &= \mathbf{1}(t)\left(A_1 + A_2(-0.5)^t + 2\,|A_3|\, 0.2828^t \cos\left(\frac{\pi}{4}t + \angle A_3\right)\right) \end{aligned}$$

4.12.1. Natural modes. For CT systems we have seen that poles determine the evolution of the system. Each pole determines an evolution of the system (natural modes) described by an exponential function. The same can be said for DT system as shown in the following table.

Natural modes associated with the different ples		
Pole	CT modes	DT modes
Single real pole p	e^{pt}	p^t
Multiple real pole p (multipl. m)	$e^{pt}, te^{pt}, \ldots, t^{m-1}e^{pt}$	$p^t, tp^t, \ldots, t^{m-1}p^t$
Single complex pair $p, \overline{p}$	$e^{\mathbf{Real}(p)t}\cos\left(\mathbf{Imag}\,(p)\,t + \phi\right)$	$\lvert p\rvert^t \cos\left(\angle pt + \phi\right)$

It is worth observing that for DT systems a real pole p, when negative, gives rise to an exponential mode p^t that oscillates. For CT systems, on the contrary, oscillating behaviours are only possible for complex conjugate pairs.

4.13. BIBO stability of DT systems

The discussion above can be translated into the following results.

THEOREM 4.62. *Consider a DT LTI system with transfer function:*

$$H(z) = \frac{n(z)}{d(z)}.$$

and assume that no zero pole cancellation takes place. Then the system is BIBO stable if and only if all poles have modules smaller than 1*:* $\forall ps.t.\ d(p) = 0$, *we have* $|p| < 1$.

The proof of this result is the same as the proof of the similar results that we have given for CT systems. The stability could be checked with a criterion very similar to the Rout-Hurwitz criterion (called Jury criterion), but this is out of the scope of this course.

EXAMPLE 4.63. Consider the system:

$$y(k+2) - 3y(k+1) + 2y(k) = 3u(k+1) - u(k).$$

Let us verify if the system is BIBO stable. The transfer function of the system is

$$Y(z)(z^2 - 3z + 2) = 3zU(z) - U(z)$$
$$Y(z) = \frac{3z-1}{z^2 - 3z + 2}.$$

The roots of the denominator are given by:

$$p = \frac{3 \pm \sqrt{9-8}}{2} = \{2, 1\}.$$

Both poles are outside the unit circle. Hence, the system is BIBO unstable.

4.14. z-Transform of sampled data signals

It is interesting to see how the z-Transform and Laplace transform are in fact close relatives. Consider a CT signal $f(t)$, its Laplace transform is given by: $F(s) = \int_0^\infty f(\tau)e^{-s\tau}d\tau$. Suppose we transform the signal into a DT sequence by taking a sample every T time units. This can be done by multiplying the signal by a sequence of Dirac δ, each one extracting a sample: $f_D(t) = f(t)\sum_{k=0}^{\infty}\delta(t-kT)$. If we compute the Laplace transform of this signal, we find:

$$\begin{aligned}
\mathcal{L}\left(f_D(t)\right) &= \int_0^\infty f_D(\tau)e^{-s\tau}d\tau = \int_0^\infty f(t)\sum_{k=0}^{\infty}\delta(t-kT)e^{-s\tau}d\tau \\
&= \int_0^\infty \sum_{k=0}^{\infty} f(kT)\delta(t-kT)e^{-s\tau}d\tau = \int_0^\infty \sum_{k=0}^{\infty} f(kT)\delta(t-kT)e^{-skT}d\tau \\
&= \sum_{k=0}^{\infty}\int_{(k-1)T}^{kT} f(kT)\delta(t-kT)e^{-skT}d\tau = \sum_{k=0}^{\infty} f(kT)e^{-skT}\int_{(k-1)T}^{kT}\delta(t-kT)d\tau \\
&= \sum_{k=0}^{\infty} f(kT)e^{-skT} = \sum_{k=0}^{\infty} f(kT)(e^{-sT})^k
\end{aligned}$$

If we set $e^{sT} = z$ we find

$$F(z) = \sum_{k=0}^{\infty} f(kT)z^k,$$

we find the definition of the z-Transform.

CHAPTER 5

State space Analysis

In this chapter, we will focus our attention on the analytic tools that can be used to compute the time response of time-invariant linear systems in the *state space form*. This type of form was introduced in Section 2.4 and is reported below for the reader's convenience:

$$\begin{aligned} \mathfrak{D}x(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t). \end{aligned} \tag{5.1}$$

The operator $\mathfrak{D}$ has been defined in (2.10) and represents the differential (CT systems) or difference (DT systems) operator. For a linear system in state space form, the $x \in \mathbb{R}^n$ is a vector representing the n *states* of the system, $y \in \mathbb{R}^p$ is a vector representing the p *outputs* of the system, while $u \in \mathbb{R}^m$ is a vector representing the m *inputs* of the system. One can immediately recognise how the state space form is a richer description of the system dynamic than the I/O representation (expressed in terms of of the transfer functions) presented in Chapter 4. Indeed, it allows for the description of MIMO systems and, more importantly, it is able also to capture system dynamics that are not visible in the I/O relation.

5.1. Control Canonical Form

As discussed above, for a SISO linear and time invariant system described in terms of its IO relation, the evolution is generally expressed by the differential (or difference) equation (3.1), which we report below for clarity

$$\mathfrak{D}^n y(t) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i y(t) + \sum_{j=0}^{p} \beta_j \mathfrak{D}^j u(t), \tag{5.2}$$

where time invariance requires that α_i and β_i be constant. The system is causal if $p \leq n$ and strictly causal if $p < n$. We have seen that the initial conditions:

$$y(0), \mathfrak{D}y(0), \ldots, \mathfrak{D}^{n-1}y(0), \ldots, \mathfrak{D}^p u(0), \ldots, \mathfrak{D}u(0)$$

and the input $u|_{[t_0,t]}$ determine the evolution of the output in the interval $[t_0, t]$. In this section we will show how to obtain the state space description (5.1) from (5.2).

Let us start from the case of $p = 0$, i.e.,

$$\mathfrak{D}^n y(t) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i y(t) + u(t).$$

The immediate solution to this problem is to consider as state space variables $x_1 = y$, $x_2 = \mathfrak{D}y$, etc., that is associated with:

$$A = \left[\begin{array}{c|cccc} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \hline \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} \end{array}\right], \text{ and } B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$$C = \left[\begin{array}{c|ccc} 1 & 0 & 0 & \cdots & 0 \end{array}\right], \text{ and } D = 0.$$

with initial conditions

$$x_i(0) = \mathfrak{D}^{i-1} y(0).$$

The matrix A is in *lower horizontal companion form*, since it is the *companion* of its own *characteristic polynomial* $\mathcal{P}(A)$ (see below), whose coefficients appear (with opposite sign) in the last row.

In the general case, i.e., $p \geq 0$, it is still possible to compute a similar description using the *superposition principle* and the fact that the output due to the k-th time derivative of the input signal is the k-th time derivative of the output. Indeed, let us call $x(t)$ the solution considering only $u(t)$, i.e.,

$$\mathfrak{D}^n x(t) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i x(t) + u(t), \tag{5.3}$$

$$y(t) = x(t). \tag{5.4}$$

If we apply the $\mathfrak{D}$ operator to both sides of the equation, we find:

$$\mathfrak{D}\mathfrak{D}^n x(t) = \mathfrak{D}\left(\sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i x(t) + u(t)\right),$$

and applying the linearity of $\mathfrak{D}$, we find

$$\mathfrak{D}^n\left(\mathfrak{D}x(t)\right) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i\left(\mathfrak{D}x(t)\right) + \mathfrak{D}u(t).$$

Therefore $\mathfrak{D}x$ is solution to the equation with input $\mathfrak{D}u$ with output given by $y(t) = \mathfrak{D}x(t)$. The same applies to $\mathfrak{D}^2 u, \mathfrak{D}^3 u, \ldots, \mathfrak{D}^{n-1} u$.

It follows that the output corresponding to an input given by the linear combination of the time derivatives of $u(t)$

$$\omega(t) = \sum_{j=0}^{p} \beta_j \mathfrak{D}^j u(t),$$

is given by

$$y(t) = \sum_{j=0}^{p} \beta_j \mathfrak{D}^j x(t).$$

If we now set $\mathfrak{D}^i x(t) = x_i(t)$, we find the same matrices A and B as for the trivial case $p = 0$, while we get for *strictly causal* systems ($p < n$), we have $y(t) = \sum_{j=1}^{p} \beta_j x_j(t)$. In matrix form, we can write:

$$A = \left[\begin{array}{c|cccc} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \hline \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} \end{array}\right], \text{ and } B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$$C = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_p & 0 & \cdots & 0 \end{bmatrix}, \text{ and } D = 0.$$

If instead $p = n$ (*non-strictly causal* systems), we can repeat the same argument that we made above on the application of the superimposition for the different derivatives that we made above. We can still say that $\mathfrak{D}^n x$ is the response to $\mathfrak{D}^n u$. However, $\mathfrak{D}^n x$ can be expressed from the differential equation as:

$$\mathfrak{D}^n x(t) = \sum_{i=0}^{n-1} \alpha_i \mathfrak{D}^i x(t) + u(t).$$

Therefore, the output $y(t)$ to

$$\omega(t) = \sum_{j=0}^{p} \beta_j \mathfrak{D}^j u(t),$$

will be

$$\begin{aligned} y(t) &= \sum_{j=0}^{n} \beta_j \mathfrak{D}^j x(t) \\ &= \beta_n \mathfrak{D}^n x(t) + \sum_{j=0}^{n} \beta_j \mathfrak{D}^j x(t) \\ &= \sum_{i=0}^{n-1} \beta_n \alpha_i \mathfrak{D}^i x(t) + u(t) + \sum_{j=0}^{n} \beta_j \mathfrak{D}^j x(t). \end{aligned}$$

This leads to state space form with

$$A = \left[\begin{array}{c|cccc} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \hline \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} \end{array}\right], \text{ and } B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$$C = \begin{bmatrix} \beta_0 + \beta_n\alpha_0 & \beta_1 + \beta_n\alpha_1 & \beta_2 + \beta_n\alpha_2 & \cdots & \beta_{n-1} + \beta_n\alpha_{n-1} \end{bmatrix}, \text{ and } D = \beta_n.$$

EXAMPLE 5.1. Consider the equation

$$a\ddot{y}(t) - by(t) = c\ddot{u}(t) - d\dot{u}(t) - eu(t),$$

where a, b, c, d, e are constant parameters. The canonical control form is given by

$$A = \left[\begin{array}{c|c} 0 & 1 \\ \hline \frac{b}{a} & 0 \end{array}\right], \text{ and } B = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$C = \begin{bmatrix} \frac{cb-ea}{a^2} & -\frac{d}{a} \end{bmatrix}, \text{ and } D = \frac{c}{a}.$$

□

EXAMPLE 5.2. Consider the circuit with a resistor R and an inductor L in Figure 1. The dynamic equations are given by (Ohm and Henry laws)

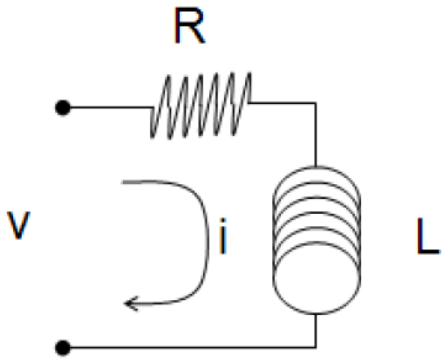


FIGURE 1. RL example.

$$L\frac{di(t)}{dt} + Ri(t) = v(t),$$

for which we assume the output $y(t) = i(t)$ and the input $u(t) = v(t)$. Therefore:

$$\dot{y}(t) = -\frac{R}{L}y(t) + \frac{u(t)}{L},$$

and, hence, in state space

$$\begin{aligned}\dot{x}_1(t) &= -\frac{R}{L}x_1(t) + \frac{u(t)}{L},\\ y(t) &= x_1(t)\end{aligned}$$

□

EXAMPLE 5.3. A mass m in position $p(t)$ subjected to an external force $f(t)$ and constrained to a wall by a spring with elasticity K and a damper B, as depicted in Figure 2. The dynamic equations are given by (Newton,

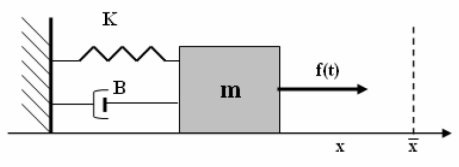


FIGURE 2. Mass-Spring-Damper example.

Hooke and Rayleigh laws)

$$\frac{d^2p(t)}{dt^2}m + K(p(t) - \hat{p}) + B\frac{dp(t)}{dt} = f(t),$$

hence, choosing $y(t) = p(t)$ and $u(t) = f(t)$, yields to

$$\ddot{y}(t) = -\frac{K}{m}(y(t) - \hat{y}) - \frac{B}{m}\dot{y}(t) + \frac{u(t)}{m}.$$

Without loss of generality we assume the spring relaxed position be $\hat{y} = 0$, which yields to the following canonical control form

$$\begin{aligned}\dot{x}(t) &= \left[\begin{array}{c|c} 0 & 1 \\ \hline -\frac{K}{m} & -\frac{B}{m} \end{array}\right] x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t),\\ y(t) &= \begin{bmatrix} \frac{1}{m} & 0 \end{bmatrix} x(t).\end{aligned}$$

□

EXAMPLE 5.4. The following discrete time equation

$$ay(k-2) - by(k-1) + cy(k) = -du(k) + eu(k-1),$$

where a, b, c, d, e are constant parameters, has the following canonical control form

$$A = \left[\begin{array}{c|c} 0 & 1 \\ \hline -\frac{a}{c} & \frac{b}{c} \end{array}\right], \text{ and } B = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$C = \begin{bmatrix} \frac{da}{c^2} & \frac{ac-db}{c^2} \end{bmatrix}, \text{ and } D = -\frac{d}{c}.$$

□

5.2. Coordinates Transformation

The state space description is in its essence a dynamic relation between quantities, i.e., the states x, the inputs u and the outputs y, expressed in different vector spaces. As shown in Section 2.4.4, the inputs and the outputs can be defined looking at the differential or difference equations describing the system dynamics. Moreover, one possible way to define the states is by using the outputs and their derivatives (or differences). However, this is not always possible and, more importantly, somethimes is not necessarily the best thing to do.

While the inputs u and the outputs y are, in some respect, constrained by the homogeneous representation of the equations, the states x are to some extent arbitrary. In general, it is always possible to define a *coordinates transformation* that binds a state variable x to another state variable, say z, using a *coordinates mapping*:

$$z = \Phi(x).$$

A mapping $\Phi(\cdot)$ able to correctly return an alternative description of the states x has to be chosen with some care. Indeed, it has to be bijective, so that any x is associated with only one z. Under these conditions, it is possible to define the inverse mapping $x = \Phi^{-1}(z)$.

Of particular interest are the linear mappings, satisfying the following linear property:

$$\Phi(\alpha x_1 + \beta x_2) = \alpha\Phi(x_1) + \beta\Phi(x_2).$$

By recalling that any linear trasformation from a vector space of dimension a to a vector space of dimension b is represented by a matrix in $b \times a$, it follows that a (time invariant) linear mapping between two state space representations is given by

$$z = Tx, T \in \mathbb{R}^{n\times n},$$

satisfying $\det(T) \neq 0$ in order to be bijective.

In the case of a linear time invariant system, it is possible to write the dynamic in z given (5.1). Indeed, by substituting in (5.1) the $x = T^{-1}z$ and recalling that T is time invariant, we find

$$\begin{aligned} \mathfrak{D}z(t) &= A_z z(t) + B_z u(t), \\ y(t) &= C_z z(t) + D_z u(t), \end{aligned}$$

where $A_z = TAT^{-1}$, $B_z = TB$, $C_z = CT^{-1}$ and $D_z = D$. All the possible state coordinates are *similar* to each other and the system dynamic obtained are *equivalent*. As a consequence, there is not a representation that is "better" than another one, which is true also for nonlinear systems, and the choice depends on the particular problem to solve.

Moreover, we also say that A_z is *similar* to A.

5.3. Continous Time Linear System Solution

For the linear system the solution of the differentail equation involved in (5.1) can be computed explicitly.

Let us start from the scalar differential equation

$$\dot{x}(t) = ax(t) + bu(t).$$

We will then see how to extend to the corresponding multidimensional case

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

From elementary math courses, we know that the solutions of a linear differential equation can be found by *summing* a particular solution of the differential equation with all the solutions of the homogeneous equations (the one obtained setting $u(t) = 0$).

5.3.1. Solution of the homogeneous equation. The *homogeneous* solution of the differential scalar equation, i.e., assuming $u(t) = 0$, is simply given by

$$x_u(t) = e^{at}x(0),$$

where $x(0)$ is the *initial condition* fo the system. This can be see by simple substitution. Using the Taylor expansion around $t_0 = 0$, the exponential function can be rewritten as

$$x_u(t) = \sum_{i=0}^{+\infty} \frac{a^i t^i}{i!} x(0).$$

Suppose we define in a similar way the solution of the multidimensional system:

$$x_u(t) = \left(I + At + \frac{A^2t^2}{2} + \dots \right) x(0) = \sum_{i=0}^{+\infty} \frac{A^i t^i}{i!} x(0),$$

where I is the *identity* matrix of dimension n. It remains to be seen if such a solution is *in fact* a solution of the homogeneous differential equation $\dot{x}(t) = Ax(t)$. This is easily seen by deriving the solution:

$$\begin{aligned} \dot{x}_u(t) &= \left(0 + A + At + \frac{A^2t^2}{2} + \dots \right) x(0) = \\ &= A \left(I + At + \frac{A^2t^2}{2} + \dots \right) x(0) = Ax_u(t). \end{aligned}$$

An obvious choice is to define the *matrix exponential* as

$$e^{At} = \sum_{i=0}^{+\infty} \frac{A^i t^i}{i!}, \tag{5.5}$$

and then write

$$x_u(t) = e^{At}x(0) = \Phi(t)x(0). \tag{5.6}$$

$\Phi(t)$ is referred to as the *state transition matrix*, which maps the initial state $x(0)$ in the state $x(t)$ with a linear transformation.

EXAMPLE 5.5. Let us compute the homogeneous solution of

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

when the initial conditions are $x_1(0) = 1$ and $x_2(0) = 2$.

The direct computation of the sum in (5.5) produces

$$e^{At} = I + At + I\frac{t^2}{2!} + A\frac{t^3}{3!} + I\frac{t^4}{4!} + A\frac{t^5}{5!} + \dots$$

and hence

$$e^{At} = \begin{bmatrix} 1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \frac{t^4}{4!} + \frac{t^5}{5!} + \dots & 0 \\ -2\left(t + \frac{t^3}{3!} + \frac{t^5}{5!} + \dots\right) & 1 - t + \frac{t^2}{2} - \frac{t^3}{3!} + \frac{t^4}{4!} - \frac{t^5}{5!} + \dots \end{bmatrix}.$$

Observing that

$$\begin{aligned} 1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \frac{t^4}{4!} + \frac{t^5}{5!} + \dots &= e^t,\\ 1 - t + \frac{t^2}{2} - \frac{t^3}{3!} + \frac{t^4}{4!} - \frac{t^5}{5!} + \dots &= e^{-t},\\ t + \frac{t^3}{3!} + \frac{t^5}{5!} + \dots &= \frac{e^t - e^{-t}}{2}, \end{aligned}$$

we can write

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

Wrapping up:

$$x(t) = \begin{bmatrix} e^t & 0 \\ e^{-t} - e^t & e^{-t} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} e^t \\ 3e^{-t} - e^t \end{bmatrix}.$$

□

5.3.2. Computation of a particular solution of the differential equation. It is now necessary to compute the *particular* solution of the differential equation in order to derive the evolution of the state variables of the system. Indeed, the overall response will be given by the *superposition principle*, i.e., the sum of the homogeneous and of the particular responses.

There are several ways to compute this. One possibility is to compute *directly* this sum by considering again the scalar system is the following:

$$\dot{x}(t) = ax(t) + bu(t) \Rightarrow \dot{x}(t) - ax(t) = bu(t).$$

Since

$$\frac{de^{-at}x(t)}{dt} = e^{-at}\left(\dot{x}(t) - ax(t)\right),$$

it follows that

$$\frac{de^{-at}x(t)}{dt} = e^{-at}bu(t).$$

Integrating both sides

$$\int_0^t \frac{de^{-a\tau}x(\tau)}{d\tau}d\tau = e^{-at}x(t) - x(0) = \int_0^t e^{-a\tau}bu(\tau)d\tau,$$

which leads to

$$x(t) = e^{at}x(0) + \int_0^t e^{a(t-\tau)}bu(\tau)d\tau.$$

As a consequence, we have that the particular solution is given by

$$x_f(t) = \int_0^t e^{a(t-\tau)}bu(\tau)d\tau.$$

In order to derive the solution for the multidimensional case, we first prove some properties of the matrix exponential in (5.5):

- $e^{A_1}e^{A_2} = e^{A_2}e^{A_1} = e^{A_1+A_2}$ iff $A_1A_2 = A_2A_1$. The proof follows from the definition in (5.5);
- From the previous property follows that $e^Ae^{-A} = e^{A-A} = e^0 = I$, where the last relation derives again directly from (5.5). Notice that the inverse of an exponential matrix e^A is e^{-A}, which *always* exists;
- Finally, $\frac{de^{At}}{dt} = Ae^{At} = e^{At}A$, which again follows directly from (5.5).

We are now in a position to follow the same steps of the monodimensionale case. Indeed, let us start with

$$\dot{x}(t) = Ax(t) + Bu(t) \Rightarrow \dot{x}(t) - Ax(t) = Bu(t).$$

Since

$$\frac{de^{-At}x(t)}{dt} = -Ae^{-At}x(t) + e^{-At}\dot{x}(t) = -e^{-At}Ax(t) + e^{-At}\dot{x}(t) = e^{-At}\left(\dot{x}(t) - Ax(t)\right),$$

it follows that

$$\frac{de^{-At}x(t)}{dt} = e^{-At}Bu(t).$$

Integrating both sides

$$\int_0^t \frac{de^{-A\tau}x(\tau)}{d\tau}d\tau = e^{-At}x(t) - e^{A0}x(0) = \int_0^t e^{-A\tau}bu(\tau)d\tau,$$

that finally leads to

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau. \tag{5.7}$$

To double check this result, it is possible to compute

$$\begin{aligned}\frac{dx(t)}{dt} &= Ae^{At}x(0) + \frac{de^{At}\int_0^t e^{-A\tau}Bu(\tau)d\tau}{dt} \\ &= Ae^{At}x(0) + Ae^{At}\int_0^t e^{-A\tau}Bu(\tau)d\tau + e^{At}e^{-At}Bu(t) \\ &= Ax(t) + Bu(t).\end{aligned}$$

EXAMPLE 5.6. Compute the *response* of the system

$$\begin{aligned} \dot{x}_1 &= x_1 \\ \dot{x}_2 &= -2x_1 - x_2 + u \end{aligned} \quad xo$$

when the initial conditions are $x_1(0) = 1$ and $x_2(0) = 2$ and the input $u(t) = 10$, $\forall t \geq 0$.

From Example 5.5, we know that

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

The unforced response is

$$x_u(t) = \begin{bmatrix} e^t & 0 \\ e^{-t} - e^t & e^{-t} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} e^t \\ 3e^{-t} - e^t \end{bmatrix}.$$

For the forced response, we have to compute

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

since the input is a step function. Therefore

$$\int_0^t e^{A(t-\tau)} d\tau B u(t) = \begin{bmatrix} \int_0^t e^{t-\tau} d\tau & 0 \\ \int_0^t e^{\tau-t} - e^{t-\tau} d\tau & \int_0^t e^{\tau-t} d\tau \end{bmatrix} B u(t) = \begin{bmatrix} e^t - 1 & 0 \\ 2 - e^t - e^{-t} & 1 - e^{-t} \end{bmatrix} B u(t),$$

. Observing that $B = [0 \ 1]^T$, we have

$$x(t) = x_u(t) + x_f(t) = \begin{bmatrix} e^t \\ 3e^{-t} - e^t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 - e^{-t} \end{bmatrix} u(t) = \begin{bmatrix} e^t \\ 3e^{-t} - e^t + 10(1 - e^{-t}) \end{bmatrix}.$$

□

REMARK 5.7. The state space evolution (5.7), solution of the linear time invariant system (5.1), comprises two terms. The first depends only on the *initial condition* $x(0)$ of the system and hence it is dubbed *unforced response*. The second term instead depends on the inputs $u(t)$ but not on the initial condition and therefore it is termed *forced response*. □

REMARK 5.8. Since e^{At} is a constant matrix for fixed t is fixed. Thereby, the states are combined through a linear transformation. Moreover, since the integral is a linear operator, the combination of forced and unforced response is an application of the *superposition principle* between the states and the inputs. □

REMARK 5.9. There are other definitions of the matrix exponential that may be useful sometimes, for example:

$$e^{At} = \lim_{k \to +\infty} \left(I + \frac{At}{k} \right)^k.$$

□

5.3.3. Connection between the State Space Form and the Laplace Transform. To explain the connection between the Laplace Transform and the state space description, let us first consider the dynamic of a *autonomous system*:

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

by applying the *differentiation rule*, we find the Laplace Transform:

$$\mathcal{L}\left(\dot{x}(t)\right) = sX(s) - x(0), \text{ and } \mathcal{L}\left(Ax(t)\right) = AX(s).$$

Hence,

(5.8)

$$sX(s) - x(0) = AX(s) \Rightarrow (sI - A)X(s) = x(0) \Rightarrow X(s) = (sI - A)^{-1}x(0).$$

The matrix $(sI - A)^{-1}$ is called the *resolvent* of A. The resolvent is defined for any $s \in \mathbb{C}$ except for the *eigenvalues* of A, which are the points in which $\det(sI - A) = 0$.

By applying the inverse Laplace Transform to (5.8), we have

$$x(t) = \mathcal{L}^{-1}\left((sI - A)^{-1}\right)x(0),$$

that, recalling (5.6), implies that the inverse Laplace Transform of the resolvent of A *is* the matrix exponential e^{At} or, equivalently, the transition matrix $\Phi(t)$.

REMARK 5.10. An alternative way to show that $\mathcal{L}^{-1}\left((sI - A)^{-1}\right) = e^{At}$ is by considering the power series expansion of the inverse of a generic non-singular matrix $I - M$, i.e.,

$$(I - M)^{-1} = I + M + M^2 + M^3 + \ldots,$$

that, in the case of the resolvent turns to

$$(sI - A)^{-1} = \left[s\left(I - \frac{A}{s}\right)\right]^{-1} = \frac{1}{s}\left(I - \frac{A}{s}\right)^{-1} = \frac{I}{s} + \frac{A}{s^2} + \frac{A^2}{s^3} + \frac{A^3}{s^4} + \ldots$$

It is now possible to apply the inverse Lapace Transform and the superposition principle to obtain

$$\mathcal{L}^{-1}\left((sI - A)^{-1}\right) = I + At + \frac{A^2t^2}{2} + \frac{A^3t^3}{3!} + \cdots \triangleq e^{At}.$$

□

EXAMPLE 5.11. Consider the following *harmonic oscillator*

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x = Ax.$$

Since

$$sI - A = \begin{bmatrix} s & -1 \\ 1 & s \end{bmatrix},$$

we have that the eigenvalues are $\pm j$ and hence

$$(sI - A)^{-1} = \frac{1}{s^2 + 1}\begin{bmatrix} s & 1 \\ -1 & s \end{bmatrix}.$$

So, by the inverse Laplace Transform we have

$$x(t) = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} x(0),$$

which is a *rotation matrix*. Notice that, by directly computing the exponential matrix, we have

$$e^{At} = \sum_{i=0}^{+\infty} \frac{A^i t^i}{i!} = \begin{bmatrix} 1 - \frac{t^2}{2} + \frac{t^4}{4!} + \dots & t - \frac{t^3}{3!} + \frac{t^5}{5!} + \dots \\ -t + \frac{t^3}{3!} - \frac{t^5}{5!} + \dots & 1 - \frac{t^2}{2} + \frac{t^4}{4!} + \dots \end{bmatrix},$$

whose elements are the Taylor expansions of $\cos t$ and $\sin t$ around $t_0 = 0$.

EXAMPLE 5.12. Consider the following *double integrator*

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x = Ax.$$

Since

$$sI - A = \begin{bmatrix} s & -1 \\ 0 & s \end{bmatrix},$$

we have that the eigenvalues are 0 and 0, and hence

$$(sI - A)^{-1} = \frac{1}{s^2} \begin{bmatrix} s & 1 \\ 0 & s \end{bmatrix}.$$

So, by the inverse Laplace Transform we have

$$x(t) = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} x(0).$$

Notice that, by directly computing the exponential matrix, we have

$$e^{At} = \sum_{i=0}^{+\infty} \frac{A^i t^i}{i!} = I + At,$$

as expected. Notice that the matrix A is nilpotent, i.e., there exists a number q such that $A^q \neq 0$ and $A^{\bar{q}} = 0 \ \forall \bar{q} > q$.

We are now able to note that the time evolution of the output is given by substituting the solution of $x(t)$ given in (5.7) in the output description given in (5.1), i.e.,

$$y(t) = C\left(e^{At}x(0) + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau\right) + Du(t). \tag{5.9}$$

Notice that the forced response involves the extension to matrices of the *convolution integral* between the states and the inputs. In particular, for a SISO system and assuming that $u(\bar{t}) = 0 \ \forall \bar{t} < 0$, we can write

$$\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau = e^{At} * Bu(t),$$

and, as a consequence,

$$\mathcal{L}\left(e^{At} * Bu(t)\right) = (sI - A)^{-1}BU(s).$$

Therefore, computing the Laplace Transform of (5.9) and applying the superposition principle, we get

$$\mathcal{L}\left(y(t)\right) = C(sI - A)^{-1}x(0) + C(sI - A)^{-1}BU(s) + DU(s).$$

Of course, if the system starts at rest, as it is usually assumed for the frequency domain approach, we have $x(0) = 0$ and hence

$$\frac{Y(s)}{U(s)} = C(sI - A)^{-1}B + D.$$

REMARK 5.13. A similar result can be also obtained by computing the Laplace tranform of the linear time invariant dynamic in (5.1) as

$$\begin{aligned} sX(s) - x(0) &= AX(s) + BU(s), \\ Y(s) &= CX(s) + DU(s), \end{aligned}$$

that, by solving the first equation in $X(s)$ and then substituting in the second produces

$$Y(s) = C(sI - A)^{-1}x(0) + \left[C(sI - A)^{-1}B + D\right]U(s),$$

as desired. □

5.3.4. The Role of the Eigenvalues in the Transfer Function. Let us consider a SISO system. As discussed above, there is a clear connection between the Laplace Transform and the state space representation. Indeed, we know that for a transfer function

$$\frac{Y(s)}{U(s)} = G(s),$$

there is a prominent role for the poles, which are the roots of the denominator. For example, the poles determine if a system is BIBO stable or not by applying the inverse Laplace Transform to the output given the input.

If we assume $x(0) = 0$, just as we did for the analysis of the the transfer function, we have

$$Y(s) = C(sI - A)^{-1}BU(s) + DU(s),$$

where the direct coupling between the input and the output given by D is now neglected since it does not involve the system characteristics. Therefore, we can restrict to studying $G(s) = C(sI - A)^{-1}B$. Let us now notice that the (i, j) element of the resolvent $(sI - A)^{-1}$ can be computed using the Cramer's rule

$$(-1)^{i+j}\frac{\det \mathrm{Adj}_{i,j}}{\mathcal{P}\left(A\right)},$$

where $\mathrm{Adj}_{i,j}$ is the *adjoint* matrix of $sI - A$, i.e., the matrix $sI - A$ with the j-th row and i-th column deleted. $\mathcal{P}\left(A\right)$ is instead the *characteristic polynomial* of A, defined as

$$\mathcal{P}\left(A\right) = \det(sI - A),$$

and satisfying the following properties:

- $\mathcal{P}(A)$ is a polynomial of degree n, with leading (i.e., s^n) coefficient one;
- The roots of $\mathcal{P}(A)$ are the eigenvalues of A;
- $\mathcal{P}(A)$ has real coefficients, so eigenvalues are either real or occur in conjugate pairs;
- There are n eigenvalues (if we count multiplicity as roots of $\mathcal{P}(A)$).

It follows that $\det \mathrm{Adj}_{i,j}$ has degree less than n. Moreover, we also notice that for SISO systems all the roots of the denominator of $G(s)$ are also the roots $\mathcal{P}(A)$, and, hence, are the eigenvalues of A. The converse is not true, since there may be some *cancellation* between the roots of $\mathcal{P}(A)$ and the roots of $\det \mathrm{Adj}_{i,j}$. This is a problem that will have a direct impact on the analysis of the structural properties of a system.

5.4. The Role of the Eigenvalues for Continuous Time Systems

Let us introduce some properties of the exponential matrix and of its eigenvalues that will be very useful in the following developments.

- If $v \in \mathbb{R}^n$ is an *eigenvector* of a matrix A associated with the eigenvalue λ, i.e., $Av = \lambda v$, then the same v is also an eigenvector of the matrix A^k associated with the eigenvalue λ^k, i.e., $A^k v = \lambda^k v$. Hence, by recalling the definition of the exponential matrix as an infinite summation given in (5.5), we have that v is also the eigenvector of e^{At} associated to the eigenvalue $e^{\lambda t}$, i.e., $e^{At}v = e^{\lambda t}v$;
- For any given A and for any transformation matrix T, we have that $(T^{-1}AT)^k = T^{-1}A^kT$. Hence, it follows again from (5.5) that

$$e^{T^{-1}AT} = T^{-1}e^AT.$$

Bearing these properties in mind, we can now relate the exponential matrix to its eigenvalues.

5.4.1. Exponential Matrix for Diagonal Matrices. Let us first recall the definition of *diagonalisable matrix*.

Definition 5.14. A square matrix A is called *diagonalisable* if it is *similar* to a diagonal matrix, i.e., if there exists an invertible matrix T such that $T^{-1}AT$ is a *diagonal matrix*.

A few preliminary results come in handy. We start by a useful definition:

Definition 5.15. Consider a matrix A and let λ be an eigenvalue. Then the set of all eigenvectors given by $(A - \lambda I)v = 0$ is a subspace, which we define "eigenspace" related to λ.

Now a few lemmas.

Lemma 5.16. *The eigenvectors related to two distinct eigenvalues are independent.*

PROOF. Let λ_1, λ_2 be distinct eigenvalues and v_1, v_2 be their associated eigenvectors. Suppose by contradiction that v_1 is linearly dependent from v_2: $v_1 = \alpha v_2$. By definition of eigenvector, we have:

$$Av_1 = \lambda_1 v_1$$

From the assumed dependency, we also have:

$$\begin{aligned} Av_1 &= A(\alpha v_2) \\ &= \alpha\lambda_2 v_2 \\ &= \lambda_2 v_1 \end{aligned}$$

Comparing the two equations, we find $\lambda_1 = \lambda_2$ which contradicts the hypotheses that they be distinct. □

LEMMA 5.17. *The eigenvectors related to distinct eigenvalues are independent.*

PROOF. From Lemma 5.16, we know that the claim applies to two eigenvalues. Suppose by contradiction that v_1, v_2, ..., v_{i-1} are independent and that v_i is linearly dependent from them: $v_i = \alpha_1 v_1 + \ldots + \alpha_n v_n$. We can write:

$$\begin{aligned} Av_i &= \lambda_i v_i = \\ &= \lambda_i\alpha_1 v_1 + \ldots + \lambda_i\alpha_n v_n \\ Av_i &= A(\alpha_1 v_1 + \ldots + \alpha_n v_n) = \\ &= \alpha_1\lambda_1 v_1 + \ldots + \alpha_n\lambda_n v_n \end{aligned}$$

By equating the two terms,

$$\lambda_i\alpha_1 v_1 + \ldots + \lambda_i\alpha_n v_n = \alpha_1\lambda_1 v_1 + \ldots + \alpha_n\lambda_n v_n$$

By the assumed independence of v_1, ..., v_n, we get: $\lambda_i = \lambda_1 = ...\lambda_n$, which contradicts the hypotheses of distinct eigenvalues. □

Now we can state the following:

THEOREM 5.18. *A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalisable* if and only if *the sum of the dimensions of its* eigenspaces *is equal to n. In this case a basis exists made of eigenvectors, we can form a matrix T having the eigenvectors as columns, and $T^{-1}AT = \Lambda$ with Λ diagonal. The diagonal entries of Λ are the eigenvalues of A.*

PROOF. We just prove sufficiency. In view of Lemma 5.17 the eigenvectors form a basis of $\mathbb{R}^n$. The matrix $T = [v_1 v_2 \ldots v_n]$ is invertible because

the eigenvectors are independent. We can easily see:

$$\begin{aligned} AT &= [Av_1 Av_2 \dots Av_n] = \\ &= [\lambda_1 v_1 \lambda_2 v_2 \dots \lambda_n v_n] = \\ &= T\mathrm{diag}\,\{\lambda_1, \lambda_2, \dots, \lambda_n\} = \\ &= T\Lambda, \end{aligned}$$

which leads us to the claim. □

In the theorem above the different eigenvalues λ_i are not assumed to be distinct. We could easily have $\lambda_2 = \lambda_3$ and the results would still apply. If the eigenvalues are dsitinct, we can easily prove the following:

COROLLARY 5.19. If the matrix $A \in R^{n\times n}$ has n distinct eigenvalues then it is dianlisable.

PROOF. For each eigenvlaue we have to have at least one eigenvector. Being the eigenvector independent, they form a basis of $\mathbb{R}^n$. Thereby, we can apply the theorem above. □

If a matrix is diagonalisable, the exponential matrix can be easily obtained. Indeed, let

$$T^{-1}AT = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

we have by the definition (5.5)

$$e^{\Lambda t} = \sum_{i=0}^{+\infty} \frac{\Lambda^i t^i}{i!} = \begin{bmatrix} \sum_{i=0}^{+\infty} \frac{\lambda_1^i t^i}{i!} & 0 & \dots & 0 \\ 0 & \sum_{i=0}^{+\infty} \frac{\lambda_2^i t^i}{i!} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_{i=0}^{+\infty} \frac{\lambda_n^i t^i}{i!} \end{bmatrix},$$

or, equivalently

$$e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n t} \end{bmatrix},$$

and finally

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

5.4.2. Exponential Matrix for Diagonal Matrices with Complex Eigenvalues. If the matrix A has complex eigenvalues, the associated eigenvector will be complex in its turn, and so the columns of the matrix T. However, since the matrix A is real by definition, complex eigenvalues will have to appear in conjugate pairs and, hence, the matrix exponential will be real. Therefore, it is possible in this case to resort to a coordinates transformation that transforms a diagonalisable matrix on the complex set into a *block diagonal matrix*, whose block dimension is at most 2. The diagonal blocks are associated to each complex and conjugated pairs.

Let us consider a matrix A having a pairs of eigenvalues $\lambda_{1,2} = \sigma \pm j\omega$, i.e.,

$$A[v_r + jv_c] = (\sigma + j\omega)[v_r + jv_c],$$
$$A[v_r - jv_c] = (\sigma - j\omega)[v_r - jv_c],$$

with v_r and v_c being the real and complex part of the eigenvectors, respectively. In matrix form, we have

$$A[v_r + jv_c\,|v_r - jv_c] = [v_r + jv_c\,|v_r - jv_c] \begin{bmatrix} \sigma + j\omega & 0 \\ 0 & \sigma - j\omega \end{bmatrix}.$$

It then follows immediately

$$e^{At}[v_r + jv_c\,|v_r - jv_c] = [v_r + jv_c\,|v_r - jv_c]e^{\sigma t} \begin{bmatrix} e^{j\omega t} & 0 \\ 0 & e^{-j\omega t} \end{bmatrix}.$$

By the Euler formula, we have that

$$e^{\sigma + j\omega} = e^{\sigma}[\cos(\omega) + j\sin(\omega)],$$
$$e^{\sigma - j\omega} = e^{\sigma}[\cos(\omega) - j\sin(\omega)].$$

Moreover, we notice that, using the invertible matrix

$$H = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}, \quad H^{-1} = -j\begin{bmatrix} j & j \\ -1 & 1 \end{bmatrix},$$

we have

$$[v_r + jv_c\,|v_r - jv_c]H = [v_r|v_c],$$

and

$$H^{-1}\begin{bmatrix} \sigma + j\omega & 0 \\ 0 & \sigma - j\omega \end{bmatrix} H = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix},$$

and, finally,

$$H^{-1}\begin{bmatrix} e^{j\omega t} & 0 \\ 0 & e^{-j\omega t} \end{bmatrix} H = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix}.$$

Therefore, noticing that if $AV = V\Lambda$ we have $AVH = V\Lambda H = VHH^{-1}\Lambda H$, and

$$A[v_r + jv_c\,|v_r - jv_c]H = A[v_r|v_c] = [v_r|v_c]\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix},$$

and

$$e^{At}[v_r + jv_c\,|v_r - jv_c]H = e^{At}[v_r|v_c] = [v_r|v_c]e^{\sigma t}\begin{bmatrix}\cos(\omega t) & \sin(\omega t)\\ -\sin(\omega t) & \cos(\omega t)\end{bmatrix}.$$

For instance, let $A \in \mathbb{R}^{4\times 4}$ with λ_1 and λ_4 being real and $\lambda_{2,3} = \sigma \pm j\omega$, we have

$$e^{At} = Te^{\Lambda t}T^{-1} = T\begin{bmatrix}e^{\lambda_1 t} & 0 & 0 & 0\\ 0 & e^{\sigma t}\cos(\omega t) & e^{\sigma t}\sin(\omega t) & 0\\ 0 & -e^{\sigma t}\sin(\omega t) & e^{\sigma t}\cos(\omega t) & 0\\ 0 & 0 & 0 & e^{\lambda_4 t}\end{bmatrix}T^{-1}.$$

5.4.3. The Jordan Canonical Form for Defective Matrices. The definition of a *defective matrix* is given below.

DEFINITION 5.20. A square matrix A is called *defective* if it is not diagonalisable.

Let us make this point clearer. Recalll that a matrix $A \in \mathbb{R}^{n\times n}$ is diagonalisable *if and only if* there exists a basis of $\mathbb{R}^n$ consisting of eigenvectors of A. This is certainly true if the matrix has n *distinct* eigenvalues, i.e.,

$$\mathcal{P}(A) = (\lambda - \lambda_1)(\lambda - \lambda_2)\dots(\lambda - \lambda_n),$$

where $\mathcal{P}(A)$ is the characteristic polynomial of the matrix A. However, this is only a sufficient condition, since it may happen that a matrix has only $h \leq n$ distinct eigenvalues λ_i, $i = 1, \dots, h$. The polynomial characteristic is in this case

$$\mathcal{P}(A) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2}\dots(\lambda - \lambda_h)^{r_h},$$

where r_i is the *algebraic multiplicity* of the eigenvalue λ_i. In such a case, it is important to recall the definition of *geometric multiplicity* of the eigenvalue λ_i.

DEFINITION 5.21. The *geometric multiplicity* of the eigenvalue λ_i is the number of linearly independent eigenvectors $v_{i,k}$ associated to λ_i.

As a direct application of Theorem 5.18, ff the *geometric multiplicity*, i.e., the number of linearly independent solutions of

$$Av_{i,k} = \lambda_i v_{i,k} \Rightarrow (\lambda_i I - A)v_{i,k} = 0,$$

is equal to the *algebraic multiplicity* r_i, i.e., there exists

$$(\lambda_i I - A)v_{i,k} = 0, \text{ with } k = 1, \dots, r_i,$$

linearly independent solutions, the matrix is still diagonalisable. In such a case, we have:

$$T = [v_{1,1}|v_{1,2}|\dots|v_{1,r_1}|v_{2,1}|v_{2,2}|\dots|v_{2,r_2}|v_{3,1}|\dots|v_{h,1}|v_{h,2}|\dots|v_{1,r_h}],$$

and

$$\Lambda = T^{-1}AT = \begin{bmatrix} \lambda_1 I_{r_1} & 0 & \dots & 0 \\ 0 & \lambda_2 I_{r_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_h I_{r_h} \end{bmatrix}.$$

If this is not the case, there exists at least one eigenvalue λ_i such that

$$(\lambda_i I - A)v_{i,k} = 0, \text{ with } k = 1, \dots, q_i < r_i, \tag{5.10}$$

or, equivalently, for which the number of linearly independent eigenvectors is less then r_i: $q_1 < r_i$. In other words, the algebraic multiplicity is greater than the geometric multiplicity. In such a case, a basis for the coordinates transformation can still be defined using the notion of *generalised eigenvectors*.

A basis of generalised eigenvectors is constructed forming a chain of generalised eigenvectors starting from each independent eigenvector $v_{i,k}$. This is done through a sequence of linear trasformations. Set $v_{i,k}^{(1)} = v_{i,k}$. a first element of the chain $v_{i,k}^{(2)}$ can be constructed using the following relations:

$$(A - \lambda_i I)v_{i,k}^{(2)} \neq 0 \text{ and } (A - \lambda_i I)^2 v_{i,k}^{(2)} = 0. \tag{5.11}$$

Such an eigenvector is simply given by

$$(A - \lambda_i I)v_{i,k}^{(2)} = v_{i,k}^{(1)} = v_{i,k}.$$

Indeed, by premultiplying the previous equation by $A - \lambda_i I$ we have that (5.11) is satisfied by means of (5.10). The first thing we can show is:

LEMMA 5.22. *The vector $v_{i,k}^{(2)}$ is independent from $v_{i,k}^{(1)}$.*

PROOF. Assume by contradiction that $v_{i,k}^{(2)} = \alpha v_{i,k}^{(1)}$. Then

$$(A - \lambda_i I)v_{i,k}^{(2)} = (A - \lambda_i I)\alpha v_{i,k}^{(1)} = 0,$$

which contradicts the hypotheses that

$$(A - \lambda_i I)v_{i,k}^{(2)} = v_{i,k}^{(1)} = v_{i,k}.$$

□

The construction is then iterated until $m_{i,k}$, i.e.,

$$\begin{aligned} (A - \lambda_i I)v_{i,k}^{(2)} &= v_{i,k}^{(1)} = v_{i,k}, \\ (A - \lambda_i I)v_{i,k}^{(3)} &= v_{i,k}^{(2)}, \\ (A - \lambda_i I)v_{i,k}^{(4)} &= v_{i,k}^{(3)}, \\ &\vdots \\ (A - \lambda_i I)v_{i,k}^{(m_{i,k}+1)} &= v_{i,k}^{(m_{i,k})}. \end{aligned}$$

We can generalise the previous result and state the following:

LEMMA 5.23. *Each element $v_{i,k}^{(i)}$ is independent from the others.*

PROOF. This is certainly true for i=2, in view of Lemma 5.22. Now assume, by contradiction, that this is true up until h bot not for $h+1$. Then we can express $v_{i,k}^{h+1} = \alpha_1 v_{i,k}^{(1)} + \ldots + \alpha_h v_{i,k}^{(h)}$. By definition:

$$(A - \lambda_i I)v_{i,k}^{(h+1)} = v_{i,k}^{(h)}$$
$$Av_{i,k}^{(h+1)} = \lambda_i v_{i,k}^{(h+1)} - v_{i,k}^{(h)} \quad = \alpha_1\lambda_i v_{i,k}^{(1)} + \alpha_2\lambda_i v_{i,k}^{(2)} + \ldots + (\lambda_i\alpha_h - 1)v_{i,k}^{(h)}$$

By our hypotheses, we can write:

$$Av_{i,k}^{(h+1)} = A\alpha_1 v_{i,k}^{(1)} + \ldots + A\alpha_j v_{i,k}^{(h)}$$
$$= (\alpha_1\lambda_i - \alpha_2)v_{i,k}^{(1)} + (\alpha_2\lambda_i - \alpha_3)v_{i,k}^{(2)} \ldots + (\alpha_{h-1}\lambda_i - \alpha_h)\lambda_i v_{i,k}^{(h-1)} + \lambda_i\alpha_h v_{i,k}^{(h)}$$

By equating the two terms:

$$\alpha_1\lambda_i v_{i,k}^{(1)} + \alpha_2\lambda_i v_{i,k}^{(2)} + \ldots + (\lambda_i\alpha_h - 1)v_{i,k}^{(h)} =$$
$$(\alpha_1\lambda_i - \alpha_2)v_{i,k}^{(1)} + (\alpha_2\lambda_i - \alpha_3)v_{i,k}^{(2)} \ldots + (\alpha_{h-1}\lambda_i - \alpha_h)\lambda_i v_{i,k}^{(h-1)} + \lambda_i\alpha_h v_{i,k}^{(h)}$$

that is

$$-\alpha_2 v_{i,k}^{(1)} - \alpha_3\lambda_i v_{i,k}^{(2)} + \ldots - v_{i,k}^{(h)} = 0,$$

for which $v_{i,k}^{(h)}$ would be linearly dependent from $v_{i,k}^{(1)}$, $v_{i,k}^{(2)}$, ..., $v_{i,k}^{(h-1)}$, which contradicts the hypotheses. □

Finally, we can prove the following:

LEMMA 5.24. *For each eigenvector λ_i and for each eigenvector $v_{i,k}$ let $m_{i,k}$ denote the length of the chain generated from $m_{i,k}$ and r_i be the algebraic multiplicity of λ_i. Then we can show:*

$$r_i = \sum_{k=1}^{q_i} m_{i,k},$$

there exists a set of r_i generalised eigenvectors that are linearly independent.

If we consider distinct eigenvectors, it is possible to prove the following:

LEMMA 5.25. *The generalised eignevector $v_{i,k}^{(f)}$ of the chain originated from each eigenvector $v_{i,k}$ is independent from the generalised eigenvalues of the chains generated by other eigenvectors (both related to the same and to different eigenvalues).*

All these partial results lead us to the following:

THEOREM 5.26. *Let $A \in \mathbb{R}^{n\times n}$ be a matrix. Then the following facts are true: I) there exist a complete basis of $\mathbb{R}^n$ made of generalised eigenvectors, II) By using the transformation matrix*

$$T = [v_{1,1}^{(1)}|v_{1,1}^{(2)}|\dots|v_{1,1}^{(m_{1,1})}|v_{2,1}^{(1)}|v_{2,1}^{(2)}|\dots|v_{2,1}^{(m_{2,1})}|\dots|v_{h,q_h}^{(m_{h,q_h})}],$$

we obtain the following block-diagonal matrix

$$J = T^{-1}AT = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_h \end{bmatrix},$$

where the J matrix is dubbed Jordan Canonical Form*, and each block J_i is the* Jordan block *associated to λ_i and has dimension r_i. Moreover, every Jordan block is a block-diagonal matrix of type*

$$J_i = \begin{bmatrix} J_{i,1} & 0 & \dots & 0 \\ 0 & J_{i,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{i,q_i} \end{bmatrix},$$

where $J_{i,k}$, $k = 1, \dots, q_i$, is a Jordan miniblock*. The number q_i of miniblocks is equal to the* geometric multiplicity *of the eigenvalue λ_i. Each miniblock is of the form*

$$J_{i,k} = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_i & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_i \end{bmatrix},$$

and its dimension is given by the number $m_{i,k}$ of linearly independent generalised eigenvectors of the k-th chain.

REMARK 5.27. From the previous analysis it follows that necessary and sufficient conditions for diagonalisability are:

- There exists n linearly independent eigenvectors;
- The algebraic multiplicity r_i of λ_i equal to the geometric multiplicity q_i;
- The dimension of each Jordan miniblock $J_{i,k}$ is unitary.

5.4.4. Exponential Matrix of Defective Matrices. We start by noticing that for block-diagonal matrices, the exponential is given by the block-diagonal of the exponential of each block:

$$\begin{bmatrix} A_1t & 0 & \dots & 0 \\ 0 & A_2t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_nt \end{bmatrix}^k = \begin{bmatrix} (A_1t)^k & 0 & \dots & 0 \\ 0 & (A_2t)^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (A_nt)^k \end{bmatrix},$$

that immediately leads to:

$$e^{\left(\begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_n \end{bmatrix} t\right)} = \begin{bmatrix} e^{A_1 t} & 0 & \dots & 0 \\ 0 & e^{A_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{A_n t} \end{bmatrix}.$$

From the analysis of the defective matrices, we have noticed that each matrix can be transformed into a block-diagonal matrix where each block comprises a *Jordan miniblock*. Therefore, to compute the matrix exponential in the general case, it is sufficient to compute the exponential of a Jordan miniblock, i.e.,

$$e^{J_{i,k}t} = e^{\left(\begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_i & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_i \end{bmatrix} t\right)} = e^{\Lambda_i t + \bar{J}_{i,k} t} = e^{\Lambda_i t} e^{\bar{J}_{i,k} t},$$

since the two matrices commute (indeed, a diagonal matrix commute with any matrix). It is now possible to notice that $\bar{J}_{i,k}$ is a *nilpotent matrix* of order $p = m_{i,k}$, i.e., $\bar{J}_{i,k}^{p} = 0$ but $\bar{J}_{i,k}^{\bar{p}} \neq 0 \ \forall p > \bar{p} \geq 0$. Hence,

$$e^{\bar{J}_{i,k}t} = I + \bar{J}_{i,k} t + \bar{J}_{i,k}^2 \frac{t^2}{2!} + \dots + \bar{J}_{i,k}^2 \frac{t^{p-1}}{(p-1)!}.$$

Finally, we have

$$e^{J_{i,k}t} = e^{\Lambda_i t} e^{\bar{J}_{i,k} t} = e^{\Lambda_i t} \begin{bmatrix} 1 & t & \frac{t^2}{2!} & \dots & \frac{t^{p-2}}{(p-2)!} & \frac{t^{p-1}}{(p-1)!} \\ 0 & 1 & t & \dots & \frac{t^{p-3}}{(p-3)!} & \frac{t^{p-2}}{(p-2)!} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & t \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

5.4.5. Exponential Matrix of Defective Matrices with Complex Eigenvalues. If the matrix A has complex eigenvalues associated with generalised eigenvectors, i.e., there exists $p = m_{i,k} > 1$, it is still possible to obtain a suitable representation. Indeed, suppose $p = 2$ and let us define

$$V = [v_r^{(1)} + j v_c^{(1)} \, | v_r^{(2)} + j v_c^{(2)} \, | v_r^{(1)} - j v_c^{(1)} \, | v_r^{(2)} - j v_c^{(2)}],$$

and the invertible matrix

$$H = \frac{1}{2} \begin{bmatrix} 1 & -j & 0 & 0 \\ 0 & 0 & 1 & -j \\ 1 & j & 0 & 0 \\ 0 & 0 & 1 & j \end{bmatrix}, \quad H^{-1} = -j \begin{bmatrix} j & 0 & j & 0 \\ -1 & 0 & 1 & 0 \\ 0 & j & 0 & j \\ 0 & -1 & 0 & 1 \end{bmatrix},$$

we have

$$VH = [v_r^{(1)}|v_c^{(1)}|v_r^{(2)}|v_c^{(2)}].$$

Moreover, $AV = VJ$ where

$$J = \begin{bmatrix} \sigma + j\omega & 1 & 0 & 0 \\ 0 & \sigma + j\omega & 0 & 0 \\ 0 & 0 & \sigma - j\omega & 1 \\ 0 & 0 & 0 & \sigma - j\omega \end{bmatrix}.$$

Hence,

$$J_r = H^{-1}JH = \begin{bmatrix} \sigma & \omega & 1 & 0 \\ -\omega & \sigma & 0 & 1 \\ 0 & 0 & \sigma & \omega \\ 0 & 0 & -\omega & \sigma \end{bmatrix} = \begin{bmatrix} W & I \\ 0 & W \end{bmatrix}.$$

For the exponential, we first notice that

$$J_r^k = \begin{bmatrix} W^k & kW^{k-1} \\ 0 & W^k \end{bmatrix},$$

hence

$$e^{J_r t} = \begin{bmatrix} I + Wt + W^2\frac{t^2}{2!} + \dots & 0 + It + 2W\frac{t^2}{2!} + 3W^2\frac{t^3}{3!} + \dots \\ 0 & I + Wt + W^2\frac{t^2}{2!} + \dots \end{bmatrix} = \begin{bmatrix} e^{Wt} & te^{Wt} \\ 0 & e^{Wt} \end{bmatrix}.$$

By recalling that

$$e^{Wt} = e^{\sigma t}\begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix},$$

we finally have

$$e^{J_r t} = \begin{bmatrix} e^{Wt} & te^{Wt} \\ 0 & e^{Wt} \end{bmatrix} = e^{\sigma t}\begin{bmatrix} \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} & t\begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \\ 0 & \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \end{bmatrix}.$$

In general, if $p > 2$, we have that following the same steps we can express

$$J_r = H^{-1}JH = \begin{bmatrix} W & I & 0 & \dots & 0 & 0 \\ 0 & W & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & W & I \\ 0 & 0 & 0 & \dots & 0 & W \end{bmatrix},$$

and hence

$$e^{J_r t} = \begin{bmatrix} e^{Wt} & te^{Wt} & \frac{t^2}{2!}e^{Wt} & \dots & \frac{t^{p-2}}{(p-2)!}e^{Wt} & \frac{t^{p-1}}{(p-1)!}e^{Wt} \\ 0 & e^{Wt} & te^{Wt} & \dots & \frac{t^{p-3}}{(p-3)!}e^{Wt} & \frac{t^{p-2}}{(p-2)!}e^{Wt} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & e^{Wt} & te^{Wt} \\ 0 & 0 & 0 & \dots & 0 & e^{Wt} \end{bmatrix}.$$

5.4.6. Some Examples.

EXAMPLE 5.28. Let us consider the matrix

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

which is the real representation of a complex matrix. Indeed, let us first compute the eigenvalues of the matrix A, which are the roots of the following scalar equation

$$\mathcal{P}(A) = \det(\lambda I - A) = \det\left(\begin{bmatrix} \lambda & -\omega \\ \omega & \lambda \end{bmatrix}\right) = (\lambda^2 + \omega^2) = (\lambda - j\omega)(\lambda + j\omega) = 0.$$

The eigenvector associated to $\lambda_1 = j\omega$ is

$$(\lambda_1 I - A)v_1 = 0 \Rightarrow \begin{bmatrix} j\omega & -\omega \\ \omega & j\omega \end{bmatrix} v_1 = 0 \Rightarrow v_1 = \begin{bmatrix} a \\ ja \end{bmatrix} \Rightarrow v_1 = \begin{bmatrix} 1 \\ j \end{bmatrix}.$$

The eigenvector associated to $\lambda_2 = -j\omega$ will be the complex conjugated of v_1, indeed

$$(\lambda_2 I - A)v_2 = 0 \Rightarrow \begin{bmatrix} -j\omega & -\omega \\ \omega & -j\omega \end{bmatrix} v_2 = 0 \Rightarrow v_2 = \begin{bmatrix} a \\ -ja \end{bmatrix} \Rightarrow v_2 = \begin{bmatrix} 1 \\ -j \end{bmatrix}.$$

We can now define the transformation matrix

$$T = [v_1|v_2] = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}, \; T^{-1} = -\frac{1}{2j}\begin{bmatrix} -j & -1 \\ -j & 1 \end{bmatrix},$$

which produces

$$T^{-1}AT = \Lambda = \begin{bmatrix} j\omega & 0 \\ 0 & -j\omega \end{bmatrix},$$

since the matrix A is diagonalisable having two distinct complex and conjugated roots or equivalently

$$AT = \Lambda T = T\Lambda,$$

with Λ diagonal. Using the invertible matrix

$$H = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}, \; H^{-1} = -j\begin{bmatrix} j & j \\ -1 & 1 \end{bmatrix},$$

we notice that $H = T^{-1}$ and hence

$$TH = I,$$

since the columns of I are the real and complex part, respectively, of the vector v_1. Therefore

$$ATH = T\Lambda H \Rightarrow A = THH^{-1}\Lambda H \Rightarrow A = H^{-1}\Lambda H \Rightarrow A = T\Lambda T^{-1},$$

as expected, i.e., the matrix A is already in real form. As a consequence

$$e^{At} = Te^{\Lambda t}T^{-1} = T\begin{bmatrix} e^{j\omega t} & 0 \\ 0 & e^{-j\omega t} \end{bmatrix}T^{-1} = \begin{bmatrix} \frac{e^{j\omega t}+e^{-j\omega t}}{2} & \frac{e^{j\omega t}-e^{-j\omega t}}{2j} \\ -\frac{e^{j\omega t}-e^{-j\omega t}}{2j} & \frac{e^{j\omega t}+e^{-j\omega t}}{2} \end{bmatrix} = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix},$$

as shown previously.

EXAMPLE 5.29. Let us compute e^{At} for

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

The first step is to compute the eigenvalues of A, i.e., the roots of

$$\det(\lambda I - A) = 0 \Rightarrow \lambda(\lambda^2 - 4\lambda + 5) = \lambda[(\lambda - 2)^2 + 1] = 0,$$

that has roots $\lambda_{1,2} = 2 \pm j$ and $\lambda_3 = 0$. For the associated eigenvector, we have

$$(\lambda_1 I - A)v_1 = 0 \Rightarrow v_1 = \begin{bmatrix} 2 \\ 2 - j \\ 1 + j \end{bmatrix},$$

then

$$v_2 = \begin{bmatrix} 2 \\ 2 + j \\ 1 - j \end{bmatrix},$$

and

$$(\lambda_3 I - A)v_3 = 0 \Rightarrow v_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

We can now define the transformation matrix

$$T = [v_{1,r}|v_{1,c}|v_3] = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & -1 \\ 1 & 1 & -1 \end{bmatrix}, \; T^{-1} = \frac{1}{5}\begin{bmatrix} 2 & 1 & 1 \\ 1 & -2 & 3 \\ 3 & -1 & -1 \end{bmatrix},$$

that produces the following similar matrix for A

$$T^{-1}AT = \Lambda = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

hence

$$e^{At} = Te^{\Lambda t}T^{-1} = T\begin{bmatrix} e^{2t}\cos(t) & e^{2t}\sin(t) & 0 \\ -e^{2t}\sin(t) & e^{2t}\cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix} T^{-1}.$$

5.5. Discrete Time Linear System Solution

x For the linear system the solution of the difference equation involved in (5.1) can be computed explicitly using an iterative solution. Indeed,

$$x(1) = Ax(0) + Bu(0),$$

$$x(2) = Ax(1) + Bu(1) = A^2x(0) + ABu(0) + Bu(1),$$

$$\vdots$$

$$x(k) = Ax(k-1) + Bu(k-1) = A^kx(0) + A^{k-1}Bu(0) + \cdots + ABu(k-2) + Bu(k-1),$$

that leads to

$$
\begin{aligned}
x(k) &= A^k x(0) + \sum_{i=0}^{k-1} A^{k-1-i} B u(i), \\
y(k) &= CA^k x(0) + C \sum_{i=0}^{k-1} A^{k-1-i} B u(i) + Du(k).
\end{aligned}
\tag{5.12}
$$

REMARK 5.30. As for the continuous-time systems, the state space evolution, solution of the discrete-time linear time-invariant system (5.1), comprises two terms. The first depends only on the *initial condition* $x(0)$ of the system and hence it is dubbed *unforced response*. The second term instead depends on the inputs $u(t)$ but not on the initial condition and therefore it is termed *forced response*. □

REMARK 5.31. As for the continuous-time systems, since A^k is a constant once k is fixed, it turns out that the states are combined through a time invariant linear transformation. Moreover, the combination of forced and unforced response is an application of the *superposition principle* between the states and the inputs. □

5.5.1. Connection between the State Space Representation and the Z-Transform. To explain the connection between the Z-Transform and the state space description, let us first consider the dynamic of a *autonomous system*:

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

whose Z-Transform are given by

$$\mathcal{Z}\left(x(k+1)\right) = zX(z) - zx(0), \text{ and } \mathcal{Z}\left(Ax(k)\right) = AX(z),$$

that has been obtained by applying the *time-shifting rule*. Hence,

$$zX(z) - zx(0) = AX(z) \Rightarrow (zI - A)X(z) = zx(0) \Rightarrow X(z) = (zI - A)^{-1} zx(0). \tag{5.13}$$

The matrix $(zI - A)^{-1}$ is defined for any $z \in \mathbb{C}$ except for the *eigenvalues* of A, which are the points in which $\det(zI - A) = 0$.

By applying the inverse Z-Transform to (5.13), we have

$$x(k) = \mathcal{Z}^{-1}\left(z(zI - A)^{-1}\right) x(0),$$

that, recalling (5.12), implies that the inverse Z-Transform of $z(zI - A)^{-1}$ *is* the matrix power A^k or, equivalently, the transition matrix $\Phi(k)$.

EXAMPLE 5.32. Consider the following *harmonic oscillator*

$$x(k+1) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} x(k) = Ax.$$

Since

$$zI - A = \begin{bmatrix} z & 1 \\ -1 & z \end{bmatrix},$$

we have that the eigenvalues are $\pm j$ and hence

$$(zI - A)^{-1} = \frac{1}{z^2 + 1} \begin{bmatrix} z & -1 \\ 1 & z \end{bmatrix}.$$

The computation of the forced response requires the extension to matrices of the *discrete convolution sum* between the states and the inputs. In particular, for a SISO system and assuming that $u(\bar{k}) = 0 \ \forall \bar{k} < 0$, we can write

$$\sum_{i=0}^{k-1} A^{k-1-i} B u(i) = A^{k-1} * Bu(k),$$

and, as a consequence,

$$\mathcal{Z}\left(A^{k-1} * Bu(k)\right) = (zI - A)^{-1} BU(z).$$

Therefore, computing the Z-Transform of (5.9) and applying the superposition principle, we get

$$\mathcal{Z}\left(y(k)\right) = Cz(zI - A)^{-1} x(0) + C(zI - A)^{-1} BU(z) + DU(z).$$

Of course, if the system starts at rest, as it is usually assumed for the frequency domain approach, we have $x(0) = 0$ and hence

$$\frac{Y(z)}{U(z)} = C(zI - A)^{-1} B + D.$$

REMARK 5.33. For a SISO system and assuming that $u(\bar{k}) = 0 \ \forall \bar{k} < 0$, we can get to the same result through a direct application of the Z-Tranform of the linear time invariant dynamic in (5.1) as

$$\begin{aligned} zX(z) - zx(0) &= AX(z) + BU(z), \\ Y(z) &= CX(z) + DU(z), \end{aligned}$$

that, by solving the first equation in $X(z)$ and then substituting in the second leads to

$$Y(z) = Cz(zI - A)^{-1} x(0) + \left[C(zI - A)^{-1} B + D\right] U(z),$$

as desired. □

EXAMPLE 5.34. The following linear system

$$\begin{aligned} x(k+1) &= \begin{bmatrix} 0.5 & 1 \\ 0 & -0.5 \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k), \\ y(k) &= \begin{bmatrix} 1 & -1 \end{bmatrix} x(k), \end{aligned}$$

is represented by the following transfer function

$$G(z) = \frac{-z + 1.5}{z^2 - 0.25}.$$

□

5.5.2. The Role of the Eigenvalues in the Discrete Time Transfer Function. Let us consider a SISO system. As for the continuous time systems, there is a clear connection between the Z-Transform and the state space representation. Indeed, we know that for a transfer function

$$\frac{Y(z)}{U(z)} = G(z),$$

are of prominent role the poles, which are the roots of the denominator. Hence, assuming $x(0) = 0$ as in the transfer function analysis case, we have

$$Y(z) = C(zI - A)^{-1}BU(z) + DU(z),$$

where the direct coupling between the input and the output given by D is now neglected since it does not involve the system characteristics. In such a case, we have $G(z) = C(zI - A)^{-1}B$. As we did for the continuous time case, the (i, j) element of $(zI - A)^{-1}$ can be computed using the Cramer's rule

$$(-1)^{i+j}\frac{\det \mathrm{Adj}_{i,j}}{\mathcal{P}(A)},$$

where $\mathcal{P}(A)$ is the *characteristic polynomial* of A, defined as

$$\mathcal{P}(A) = \det(zI - A).$$

$\det \mathrm{Adj}_{i,j}$ has degree less than n and there is still the possibility of having a *cancellation* between the roots of $\mathcal{P}(A)$ and the roots of $\det \mathrm{Adj}_{i,j}$.

5.6. The Role of the Eigenvalues for Discrete Time Systems

Let us start with some properties of the matrix power and its eigenvalues that turn to be useful in the next development.

- $A_1^k A_2^k = A_2^k A_1^k = (A_1 A_2)^k$, $\forall k$, iff $A_1 A_2 = A_2 A_1$;
- If A is invertible, then $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k = A^{-k}$;
- If $v \in \mathbb{R}^n$ is an *eigenvector* of a matrix A associated with the eigenvalue λ, i.e., $Av = \lambda v$, than it immediately follows that v is also an eigenvector of the matrix A^k associated with the eigenvalue λ^k, i.e., $A^k v = \lambda^k v$;
- For any given A and for any transformation matrix T, we have that $(T^{-1}AT)^k = T^{-1}A^kT$;
- $\det(A^k) = [\det(A)]^k$. From this property, and recalling that $\det(AB) = \det(A)\det(B)$ (if of course A and B are square matrices), it follows that $\det(AA^{-1}) = \det(A)\det(A^{-1}) = 1$, which is of course $\det(I) = 1$. Hence, the *eigenvalues does not change for similarity transformation*, which is a property we have given for granted, but never proved.

Bearing these properties in mind, we can now relate the matrix power to its eigenvalues.

5.6.1. Matrix Powers for Diagonal Matrices. If a matrix is diagonalisable, the matrix power can be easily obtained. Indeed, let

$$T^{-1}AT = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

we have that

$$A^k = T\Lambda^k T^{-1} = T \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix} T^{-1}.$$

5.6.2. Matrix Powers for Diagonal Matrices with Complex Eigenvalues. If the matrix A has complex eigenvalues, it is possible to resort to a coordinates transformation that transforms a diagonalisable matrix on the complex set into a *block diagonal matrix*, whose block dimension is at most 2. The diagonal blocks are associated to each complex and conjugated pairs. As derived for the continuous time case and since $AV = V\Lambda$ we have $AVH = V\Lambda H = VHH^{-1}\Lambda H$, we have

$$A[v_r + jv_c \,|v_r - jv_c]H = A[v_r|v_c] = [v_r|v_c] \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix},$$

that can be further transformed using the modulus $\rho = \sqrt{\sigma^2 + \omega^2}$ and phase $\theta = \arctan\left(\frac{\omega}{\sigma}\right)$, i.e.,

$$A[v_r|v_c] = [v_r|v_c]\rho \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Therefore

$$A^k[v_r|v_c] = [v_r|v_c]\rho^k \begin{bmatrix} \cos(k\theta) & \sin(k\theta) \\ -\sin(k\theta) & \cos(k\theta) \end{bmatrix}.$$

For instance, let A be a matrix defined in $\mathbb{R}^{4\times 4}$, with λ_1 and λ_4 real and $\lambda_{2,3} = \sigma \pm j\omega$, we have

$$A^k = T\Lambda^k T^{-1} = T \begin{bmatrix} \lambda_1^k & 0 & 0 & 0 \\ 0 & \rho^k\cos(k\theta) & \rho^k\sin(k\theta) & 0 \\ 0 & -\rho^k\sin(k\theta) & \rho^k\cos(k\theta) & 0 \\ 0 & 0 & 0 & \lambda_4^k \end{bmatrix} T^{-1}.$$

5.6.3. Exponential Matrix of Defective Matrices. In this case, it is still possible to use the Jordan form. Indeed, we still have that

$$\begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_n \end{bmatrix}^k = \begin{bmatrix} A_1^k & 0 & \dots & 0 \\ 0 & A_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_n^k \end{bmatrix}.$$

Hence, the analysis is related to the behaviour of the *Jordan miniblocks*, i.e.,

$$J_{i,l}^k = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_i & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_i \end{bmatrix}^l = (\Lambda_i + \bar{J}_{i,l})^k = \sum_{q=0}^{k} C_q^k \lambda_i^{k-q} \bar{J}_{i,l}^q,$$

where

$$\begin{cases} C_q^k = \begin{pmatrix} k \\ q \end{pmatrix} = \dfrac{k!}{q!(k-q)!} & \text{if } k \geq q, \\ C_q^k = 0 & \text{if } k < q, \end{cases},$$

that is a polynomial function of k of degree q, i.e., the *binomial coefficient*. Therefore, for a miniblock of dimension $p = m_{i,k}$, we have

$$J_{i,l}^k = \begin{bmatrix} \lambda_i^k & C_1^k \lambda_i^{k-1} & C_2^k \lambda_i^{k-2} & \dots & C_{q-2}^k \lambda_i^{k-q+2} & C_{q-1}^k \lambda_i^{k-q+1} \\ 0 & \lambda_i^k & C_1^k \lambda_i^{k-1} & \dots & C_{q-3}^k \lambda_i^{k-q+3} & C_{q-2}^k \lambda_i^{k-q+2} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i^k & C_1^k \lambda_i^{k-1} \\ 0 & 0 & 0 & \dots & 0 & \lambda_i^k \end{bmatrix}.$$

5.6.4. Matrix Powers of Defective Matrices with Complex Eigenvalues. If the matrix A has complex eigenvalues associated with generalised eigenvectors, i.e., there exists $p = m_{i,k} > 1$, it is still possible to obtain a suitable representation. Indeed, suppose $p = 2$ and following the same rationale of the continuous time case, we have

$$J_r = H^{-1} J H = \begin{bmatrix} \sigma & \omega & 1 & 0 \\ -\omega & \sigma & 0 & 1 \\ 0 & 0 & \sigma & \omega \\ 0 & 0 & -\omega & \sigma \end{bmatrix} = \begin{bmatrix} W & I \\ 0 & W \end{bmatrix}.$$

For the matrix power, we recall that

$$J_r^k = \begin{bmatrix} W^k & kW^{k-1} \\ 0 & W^k \end{bmatrix} = \begin{bmatrix} W^k & C_1^k W^{k-1} \\ 0 & W^k \end{bmatrix},$$

and thus

$$J_r^k = \begin{bmatrix} \rho^k \begin{bmatrix} \cos(k\theta) & \sin(k\theta) \\ -\sin(k\theta) & \cos(k\theta) \end{bmatrix} & k\rho^{k-1} \begin{bmatrix} \cos((k-1)\theta) & \sin((k-1)\theta) \\ -\sin((k-1)\theta) & \cos((k-1)\theta) \end{bmatrix} \\ 0 & \rho^k \begin{bmatrix} \cos(k\theta) & \sin(k\theta) \\ -\sin(k\theta) & \cos(k\theta) \end{bmatrix} \end{bmatrix}.$$

In general, if $p > 2$, we have that following the same steps we can express

$$J_r = H^{-1}JH = \begin{bmatrix} W & I & 0 & \dots & 0 & 0 \\ 0 & W & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & W & I \\ 0 & 0 & 0 & \dots & 0 & W \end{bmatrix},$$

and hence

$$J_r^k = \begin{bmatrix} W^k & C_1^k W^{k-1} & C_2^k W^{k-2} & \dots & C_{q-2}^k W^{k-q+2} & C_{q-1}^k W^{k-q+1} \\ 0 & W^k & C_1^k W^{k-1} & \dots & C_{q-3}^k W^{k-q+3} & C_{q-2}^k W^{k-q+2} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & W^k & C_1^k W^{k-1} \\ 0 & 0 & 0 & \dots & 0 & W^k \end{bmatrix}.$$

5.6.5. Some Examples.

EXAMPLE 5.35. Let us consider the matrix

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

It follows that

$$A^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{k(k-1)}{2}\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{bmatrix}.$$

□

5.7. Modes Analysis

The unforced response of a linear system given by the equations (5.1) is just a linear composition of *only* the functions that can be found in the real Jordan form matrix. Such functions are called the *modes* of the system.

5.7.1. Continuous Time Systems. As previously obtained, the modes of a continuous time system are given by:

(1) Simple exponential functions $e^{\lambda t}$ given by Jordan miniblocks of dimension one with real eigenvalue λ. These modes are convergent to zero, constant or exponentially divergent if the eigenvalue λ is less, equal or greater than zero, respectively;
(2) Composite exponential functions $t^k e^{\lambda t}$ given by Jordan miniblocks of dimension $m > 1$ (and hence $k = 0, \dots, m-1$) with real eigenvalue λ. These modes are convergent to zero if $\lambda < 0$, polynomially divergent if $\lambda = 0$ and $k > 0$ or exponentially divergent if $\lambda > 0$;

(3) Oscillating functions of the type $e^{\sigma t}\cos(\omega t)$ and $e^{\sigma t}\sin(\omega t)$ given by two miniblocks of dimension 1 associated to complex and conjugated eigenvalues $\lambda = \sigma \pm j\omega$. For the real Jordan form, the miniblock is only one. These modes are convergent to zero if $\sigma < 0$, persistently oscillating if $\sigma = 0$ or exponentially divergent if $\sigma > 0$;
(4) Oscillating functions of the type $t^k e^{\sigma t}\cos(\omega t)$ and $t^k e^{\sigma t}\sin(\omega t)$ given by Jordan miniblocks of dimension $m > 1$ (and hence $k = 0, \ldots, m-1$) with complex and conjugated eigenvalues $\lambda = \sigma \pm j\omega$. These modes are convergent to zero if $\sigma < 0$, polynomially divergent if $\sigma = 0$ and $k > 0$ or exponentially divergent if $\sigma > 0$.

5.7.2. Discrete Time Systems. As previously obtained, the modes of a discrete time system are given by:

(1) Powers λ^k given by Jordan miniblocks of dimension one with real eigenvalue λ. These modes are convergent to zero, constant or divergent if the eigenvalue $|\lambda|$ is less, equal or greater than one, respectively. If $\lambda < 0$, the series comprises an oscillating behaviour with alternate signs;
(2) Series of functions $C_q^k \lambda^{k-q}$ given by Jordan miniblocks of dimension $m > 1$ (and hence $k = 0, \ldots, m-1$) with real eigenvalue λ. These modes are convergent to zero if $|\lambda| < 1$, polynomially divergent if $|\lambda| = 1$ or exponentially divergent if $|\lambda| > 1$. Again the oscillating behaviour is given by $\lambda < 0$;
(3) Oscillating sequence of the type $\rho^k\cos(k\theta)$ and $\rho^k\sin(k\theta)$ given by two miniblocks of dimension 1 associated to complex and conjugated eigenvalues $\lambda = \sigma \pm j\omega = \rho e^{\pm j\theta}$. For the real Jordan form, the miniblock is only one. These modes are convergent to zero if $\rho < 1$, persistently oscillating if $\rho = 1$ or exponentially divergent if $\rho > 1$. The frequency of the oscillations is proportional to θ: the greater is θ, the higher would be the oscillation frequency. The oscillating behaviour that previously was given by $\lambda < 0$, it is now given for the maximum value of $\theta = \pi$;
(4) Oscillating series of the type $C_q^k \rho^{k-q}\cos((k-q)\theta)$ and $C_q^k rho^{k-q}\sin((k-q)\theta)$ given by Jordan miniblocks of dimension $m > 1$ (and hence $k = 0, \ldots, m-1$) with complex and conjugated eigenvalues $\lambda = \sigma \pm j\omega = \rho e^{\pm j\theta}$. These modes are convergent to zero if $\rho < 1$, polynomially divergent if $\rho = 1$ and $k > 0$ or exponentially divergent if $\rho > 1$.

CHAPTER 6

Structural Properties of Systems

The unforced response with the initial condition $x(0)$ expresses if the states will converge or diverge regardless of the input. This is a property that can be inferred from the modes analysis presented in 5.7. Hence the eigenvalues plays a prominent role for the overall system, having close connection with BIBO stability. However we will see that the concepts of convergence/divergence, more formally defined as *stability*, is a system-wide property, rather than an input-output relation as the BIBO stability. It will be clear in the following that the system modes determines the BIBO stability.

6.1. Stability

Consider a generic system, not necessarily linear, on a vector space $x \in \mathbb{R}^n$

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

If the system is time-invariant, we can drop the variable t and write

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

The trajectory of the system is given by $x(t) = \Phi_x(x_0, u(t))$, that is a function of the initial conditions $x(t_0) = x_0$ and of the input vector $u(t)$. For linear systems, these two components can be computed separately and then summed up using the superposition principle, i.e., the unforced and forced response.

An important concept for the stability analysis is given by the *equilibrium point*.

Definition 6.1. The *equilibrium point* is a constant solution to a differential equation or a *fixed point* to a difference equation. □

In practice, $\bar{x} \in \mathbb{R}^n$ is an *equilibrium point* for the differential equation

$$\frac{dx(t)}{dt} = \mathbf{f}(x(t), t),$$

if

$$\mathbf{f}(\bar{x}, t) = 0, \forall t.$$

Similarly, $\bar{x} \in \mathbb{R}^n$ is an *equilibrium point* (or *fixed point*) for the difference equation

$$x(t+1) = \mathbf{f}(x(t), t),$$

if

$$\bar{x} = \mathbf{f}(\bar{x}, t), \forall t = 0, 1, 2, \dots$$

EXAMPLE 6.2. Let us consider the RL circuit of Example 5.2 in Figure 1, whose dynamic equations are given by (Ohm and Henry laws)

$$L\frac{di(t)}{dt} + Ri(t) = v(t).$$

Assuming the output $y(t) = i(t)$ and the input $u(t) = v(t)$, we have the following state space realisation

$$\begin{aligned} \dot{x}_1(t) &= -\frac{R}{L}x_1(t) + \frac{u(t)}{L}, \\ y(t) &= x_1(t). \end{aligned}$$

The equilibrium point is given by

$$\dot{\bar{x}}_1(t) = 0 = -\frac{R}{L}\bar{x}_1(t) + \frac{u(t)}{L} \Rightarrow \bar{x}_1 = \frac{u}{R}.$$

□

EXAMPLE 6.3. Consider the mass-dumper-spring of Example 5.3 in Figure 2, whose dynamic equations are given by (Newton, Hooke and Rayleigh laws)

$$\frac{d^2p(t)}{dt^2}m + K(p(t) - \hat{p}) + B\frac{dp(t)}{dt} = f(t),$$

where $p(t)$ is the position of the cart, $\hat{p}$ is the relax point of the spring and $f(t)$ is the external force acting on the system. Choosing $y(t) = p(t)$, $u(t) = f(t)$ and assuming $\hat{p} = \hat{y} = 0$, the following canonical control form is obtained

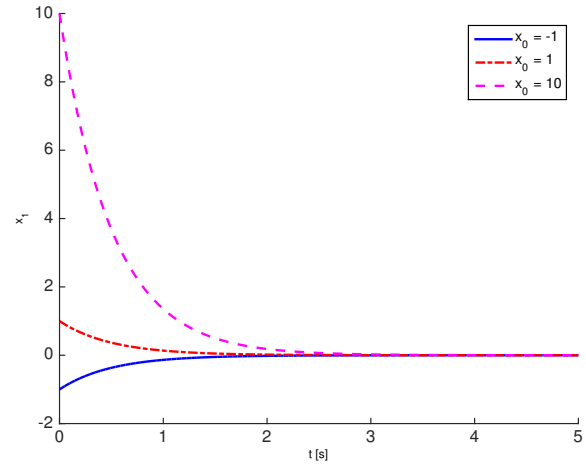
$$\begin{aligned} \dot{x}(t) &= \left[\begin{array}{c|c} 0 & 1 \\ \hline -\frac{K}{m} & -\frac{B}{m} \end{array}\right] x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t), \\ y(t) &= \begin{bmatrix} \frac{1}{m} & 0 \end{bmatrix}. \end{aligned}$$

The equilibrium point is given by

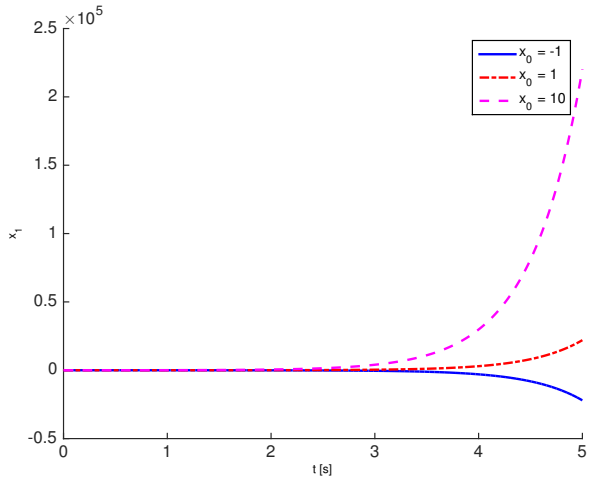
$$\begin{aligned} \bar{x}_1 &= \frac{m}{K}u, \\ \bar{x}_2 &= 0. \end{aligned}$$

□

From the previous examples it is clear how the equilibrium point can be a function also of the input. Moreover, in order to reach a constant state value, the input should be constant as well. Furthermore, by the modes analysis it turns out that if all the modes are convergent, then the unforced response tends towards zero, which means that for $u(t) = 0$ the equilibrium point will be the origin.



(a)



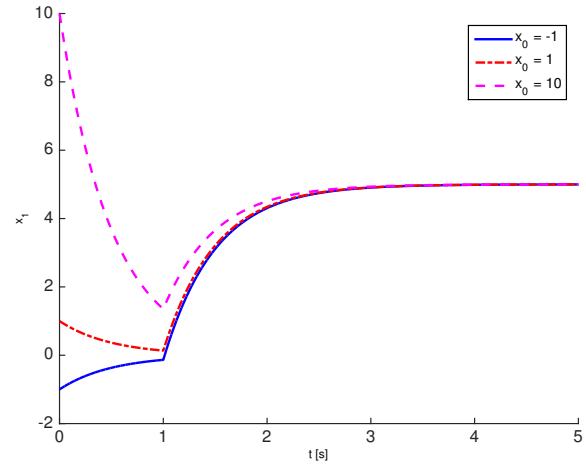
(b)

FIGURE 1. RL example: a) Unforced response for $R = 2$ Ohm and $L = 1$ Henry; b) Unforced response for $R = -2$ Ohm and $L = 1$ Henry.

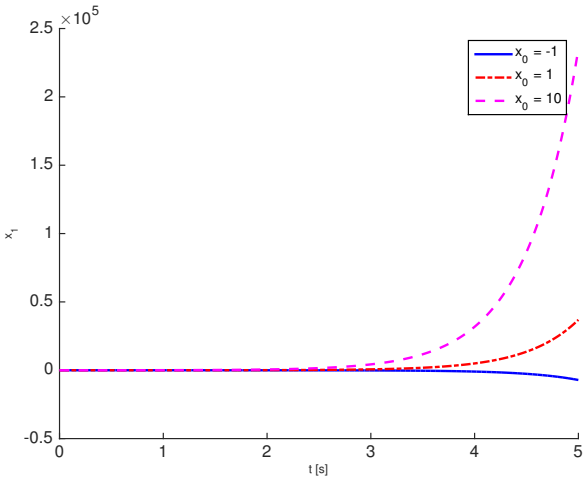
EXAMPLE 6.4. The unforced response for the Example 6.2 is reported in Figure 1-a when $R = 2$ Ohm and $L = 1$ Henries. In this case the modes are convergent, indeed there is only one eigenvalue equals to

$$\lambda_1 = -\frac{R}{L},$$

hence exponential convergence. The figure reports the time evolution for three different initial conditions. Irrespective of the initial condition, the system exponentially converges towards the origin. Similarly, if $R = -2$ Ohm



(a)

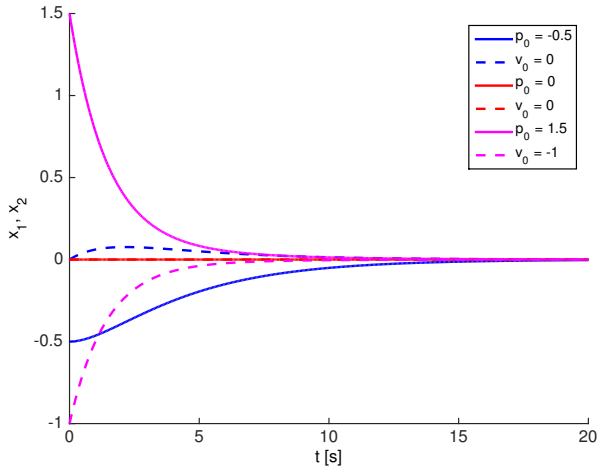


(b)

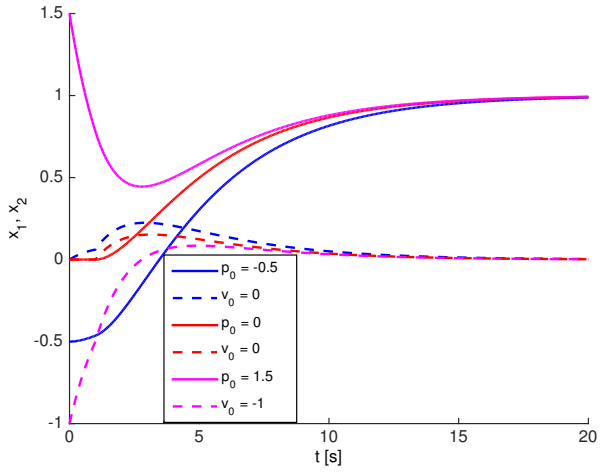
FIGURE 2. RL example: a) Response for $R = 2$ Ohm and $L = 1$ Henry; b) Response for $R = -2$ Ohm and $L = 1$ Henry. The constant input $u = 10$ Volts is applied after one second.

and $L = 1$ Henries (Figure 1-b), which is not physically feasible, otherwise the resistor would inject energy in the system instead of dissipating it, the behaviour is divergent irrespective of the initial condition. In fact, in this case the eigenvalue λ_1 previously computed would be positive.

If an input is added equal to $u = 10$ Volts after one second, the responses change to the sum of the unforced and forced responses, as reported in Figure 2. The equilibrium point reached is a function of the constant input value. Notice how the presence of a constant input does not change the



(a)



(b)

FIGURE 3. Mass-Spring-Damper example for $m = 10$ Kg, $K = 2$ N/m and $B = 10$: a) Unforced response; b) Overall response assuming a constant input $u = 2$ N applied after one second.

divergent behaviour (compare Figure 1-b and Figure 2-b). Moreover, as for the unforced response, the steady state value reached after the transient (i.e., the equilibrium point) is not affected by the initial condition. This is trivial by recalling the definition of the equilibrium point previously retrieved. □

EXAMPLE 6.5. The unforced response for the Example 6.3 is reported in Figure 3-a when $m = 10$ Kg, $K = 2$ N/m and $B = 10$. In this case the

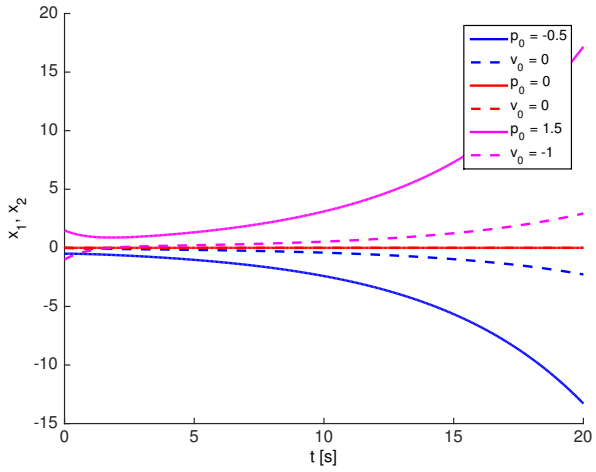


FIGURE 4. Mass-Spring-Damper example: Unforced response for $m = 10$ Kg, $K = -2$ N/m and $B = 10$.

modes are convergent, indeed there are two eigenvalue equals to

$$\lambda_1 = \frac{-B + \sqrt{B^2 - 4Km}}{2m} \text{ and } \lambda_2 = \frac{-B - \sqrt{B^2 - 4Km}}{2m},$$

that, given the described choices of the parameters, are both negative and real. As a consequence, we have exponential convergence. The figure reports the time evolution for three different initial conditions for both the positions and the velocity $v(t) = \dot{p}(t)$. Irrespective of the initial conditions, the system exponentially converges towards the origin. If an input (the external force) is added equal to $u = 2$ N after one second, the responses change to the sum of the unforced and forced responses, as reported in Figure 3-b. The equilibrium point reached is a function of the constant input value. Moreover, as for the unforced response, the steady state value reached after the transient (i.e., the equilibrium point) is not affected by the initial conditions. This is trivial by recalling the definition of the equilibrium point previously computed.

If $m = 10$ Kg, $K = -2$ N/m and $B = 10$ (Figure 4), which is not physically feasible, otherwise the spring would push the mass when it is extended beyond its relax point, the behaviour of the unforced response is divergent irrespective of the initial conditions. In fact, in this case the eigenvalue λ_1 previously computed would be positive. Also in this case, the presence of a constant input would not change the divergent behaviour. The important fact here to notice is that without *perturbation* the system does not exhibit an unstable behaviour in the unforced response, as highlighted by the red line corresponding to $x_1(t_0) = p(t_0) = p_0 = 0$ and $x_2(t_0) = v(t_0) = v_0 = 0$. This is very important and will be used extensively in the stability analysis.

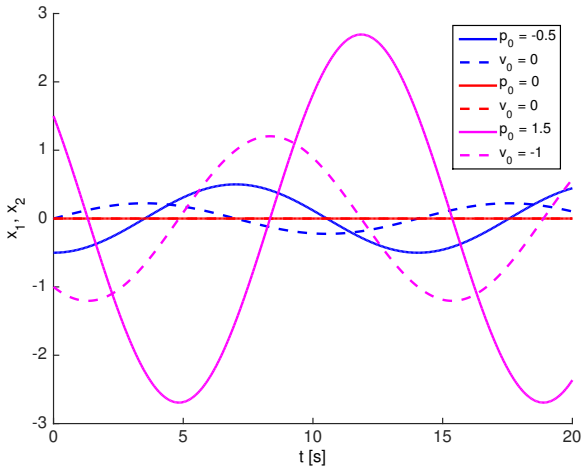


FIGURE 5. Mass-Spring-Damper example: Unforced response for $m = 10$ Kg, $K = 2$ N/m and $B = 0$.

An interesting case is when $m = 10$ Kg, $K = 2$ N/m and $B = 0$, as reported in Figure 5. The unforced response now exhibits a persistent oscillations, whose amplitude do depend on the initial conditions. In this case the eigenvalues are given by

$$\lambda_1 = \frac{\sqrt{-Km}}{m} \text{ and } \lambda_2 = \frac{-\sqrt{-Km}}{m},$$

which are two imaginary conjugated roots. In this case, the presence of a constant input would change the DC gain of the oscillations but bot its persistent oscillations. Again, without *perturbation* the system does not exhibit a persistent oscillation in the unforced response, as highlighted by the red line corresponding to $x_1(t_0) = p(t_0) = p_0 = 0$ and $x_2(t_0) = v(t_0) = v_0 = 0$. □

From the concept of perturbation, it is clear that the equilibrium point makes only sense if the system has converging behaviour, being the point in the state space that the system reaches starting from any initial condition.

6.1.1. The role of the Eigenvectors in the Unforced Response. We saw in Chapter 5 that each matrix $A \in \mathbb{R}^{n\times n}$ has n (generalised) linearly independent eigenvectors v_i associated to the eigenvalues λ_i. To clearly understand the role of the eigenvectors, let us consider a diagonalisable matrix A. In such a case, there exists n independent solutions

$$Av_i = \lambda_i v_i, \forall i = 1, \dots, n.$$

Moreover, if the system is continuous time (similar arguments can be followed for the discrete counterparts), we know that

$$e^{At}v_i = e\lambda_i t v_i, \forall i = 1, \dots, n.$$

Furthermore, we are also aware that having v_i linearly independent yields to

$$x = \sum_{i=1}^{n} \alpha_i v_i, \forall x \in \mathbb{R}^n,$$

where $\alpha_i \in \mathbb{R}$.

Since the unforced response is given by

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

the state space dynamic *aligns to the eigenvector directions*. Without loss of generality, let us suppose that $\lambda_i \in \mathbb{R}$ and that $\lambda_i < 0$, $\forall i = 1, \ldots, n$. Moreover, let us assume that $\lambda_n < \lambda_{n-1} < \cdots < \lambda_1 < 0$. It turns out that the unforced solution will converge towards the origin along the direction given by v_1. Similar considerations can be derived if $\lambda_i < 0$, $\forall i = 2, \ldots, n$, and $\lambda_1 > 0$: the system will diverge along the direction given by v_1.

EXAMPLE 6.6. The *phase portrait*, i.e., the solution of the differential equation in the state space as a function of time, of the unforced response for the Example 6.3 is reported in Figure 6-a when $m = 10$ Kg, $K = 2$ N/m and $B = 10$. In this case the modes are convergent, and the phase portrait clearly shows that the origin is approached, for any initial condition, along v_1, since the two eigenvalues are real and negative and equals to

$$\lambda_1 = \frac{-B + \sqrt{B^2 - 4Km}}{2m} \text{ and } \lambda_2 = \frac{-B - \sqrt{B^2 - 4Km}}{2m}.$$

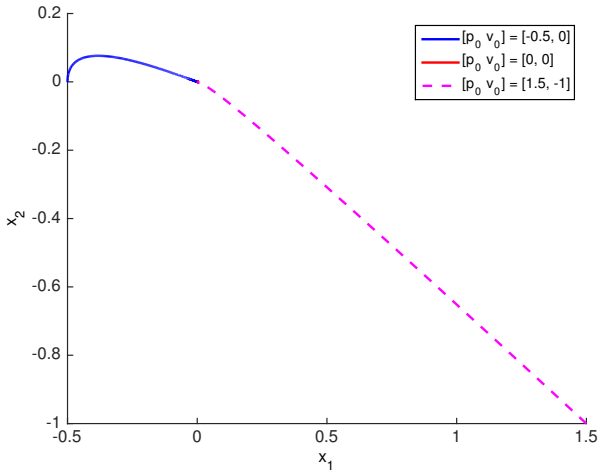
Instead, for $m = 10$ Kg, $K = -2$ N/m and $B = 10$ the system is unstable and the system aligns, for any initial condition, along the eigenvector v_1, since $\lambda_1 > 0$ and $lambda_2 < 0$ (see Figure 6-b).

Finally, if $m = 10$ Kg, $K = 2$ N/m and $B = 0$, the system has a persistent oscillation with imaginary conjugated eigenvalues

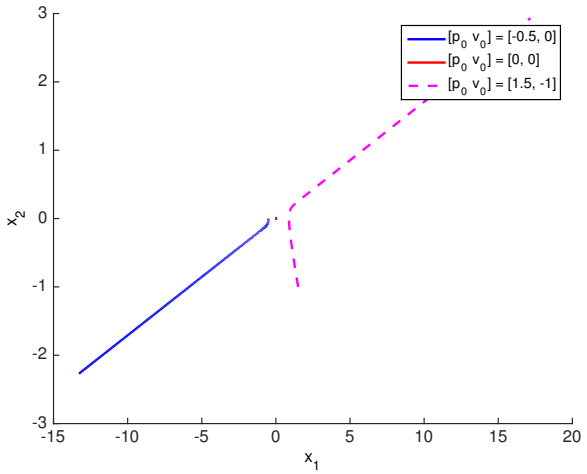
$$\lambda_1 = \frac{\sqrt{-Km}}{m} \text{ and } \lambda_2 = \frac{-\sqrt{-Km}}{m}.$$

Recalling the results obtained in Chapter 5, we have that the exponential matrix is a rotation matrix, therefore the phase portrait depicts circles (see Figure 7). □

6.1.2. Lyapunov Stability Definition. It is clear from the previous analysis that what matters for system is its behaviour in time. In particular it can be converging, diverging or oscillating depending on the eigenvalues (i.e., modes analysis). Moreover, the natural intuition leads to the following statement: if the behaviour is converging or oscillating the system will be *stable*, if it is diverging it will be *unstable*. However, in order to give an answer to the stability property of the system, we had to compute the time evolution of the system, that is we had to solve the differential or difference equations. Of course, this is a solution that is always feasible for linear



(a)



(b)

FIGURE 6. Mass-Spring-Damper example: phase portrait for the system having a) a converging behaviour with $m = 10$ Kg, $K = 2$ N/m and $B = 10$; b) a diverging behaviour with $m = 10$ Kg, $K = -2$ N/m and $B = 10$.

systems, since it is only requested to compute the eigenvalues of the dynamic matrix. Nonetheless, this can be a very difficult problem for a nonlinear system. So, is there any possibility to solve this problem in the general case and then specialise it in the linear case?

The main idea is to express the *distance* from the equilibrium point: if the distance decreases with time, the system is *asymptotically stable*; if it is bounded, the system is *stable*; if it increases with time, the system is

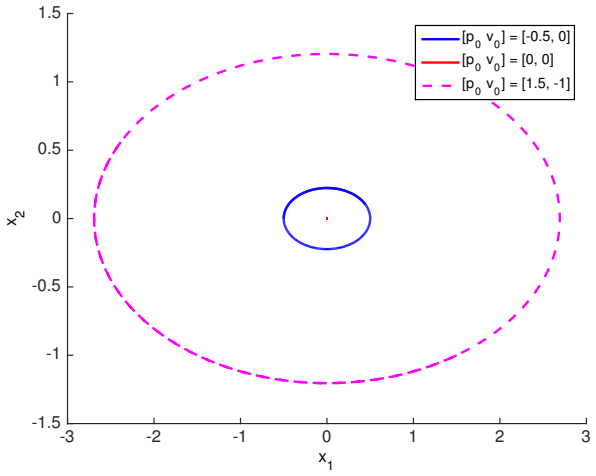


FIGURE 7. Mass-Spring-Damper example: phase portrait for the system showing a persistent oscillation with $m = 10$ Kg, $K = 2$ N/m and $B = 0$.

unstable. An easy way to express the distance, or *vector length*, in $\mathbb{R}^n$ is by means of a *vector norm*.

DEFINITION 6.7. Given a vector space V defined on a field $\mathbf{F}$, a *vector norm* ρ satisfies the following properties:

- *Positive homogeneity*: $\rho(a\mathbf{v}) = |a|\rho(\mathbf{v})$, $\forall a \in \mathbf{F}$ and $\forall \mathbf{v} \in V$;
- *Triangle inequality*: $\rho(\mathbf{v} + \mathbf{u}) \leq \rho(\mathbf{v}) + \rho(\mathbf{u})$;
- *Zero vector*: $\rho(\mathbf{v}) = 0$, if and only if $\mathbf{v} = \mathbf{0}$.

□

We will usually consider $\rho(\cdot)$ as $\|\cdot\|$.

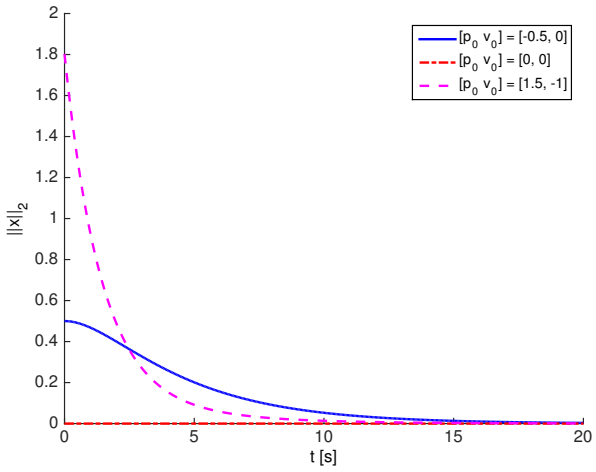
DEFINITION 6.8. A vector space with a norm $\|\cdot\|$ is called a *normed vector space*. □

For example, the *Euclidean norm* (or *2-norm*) is $\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = \sqrt{\mathbf{v}^T\mathbf{v}}$.

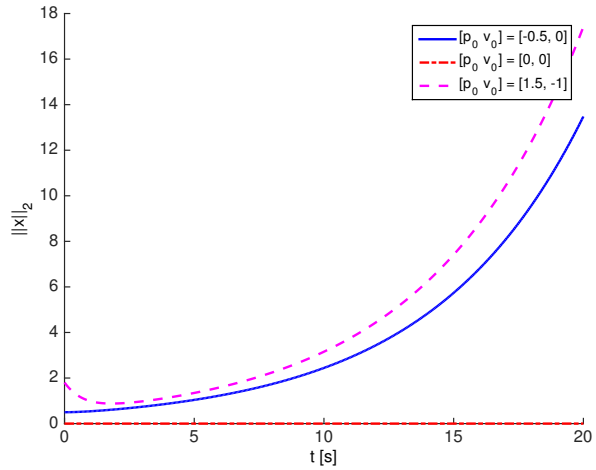
EXAMPLE 6.9. Let us consider again the unforced response for the Example 6.3. In particular, the Euclidean 2-norm of the state vector $x(t) - \bar{x}$ when $m = 10$ Kg, $K = 2$ N/m and $B = 10$ is reported in Figure 8-a In this case the modes are convergent, and, hence, the norm is converging irrespective of the initial conditions. Instead, for $m = 10$ Kg, $K = -2$ N/m and $B = 10$ the system is unstable and the norm is diverging (see Figure 6-b).

If $m = 10$ Kg, $K = 2$ N/m and $B = 0$, the system has a persistent oscillation, hence the norm is bounded (see Figure 9). □

With this idea in mind, we can now presents the definitions that underlie the *Lyapunov Stability Theory*.



(a)



(b)

FIGURE 8. Mass-Spring-Damper example: norm the system having a) a converging behaviour with $m = 10$ Kg, $K = 2$ N/m and $B = 10$; b) a diverging behaviour with $m = 10$ Kg, $K = -2$ N/m and $B = 10$.

DEFINITION 6.10. An equilibrium point $\bar{x}$ is *stable* if all the trajectories $x(t) = \Phi x(x', u, t)$ that starts from initial points x' *sufficiently close* to $\bar{x}$ remain *arbitrarily closed* to $\bar{x}$. □

More precisely: if $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. $\|x' - \bar{x}\| < \delta$, then $\|\Phi x(x', u, t) - \bar{x}\| < \varepsilon$, $\forall t$. This case is represented by Figure 10-a.

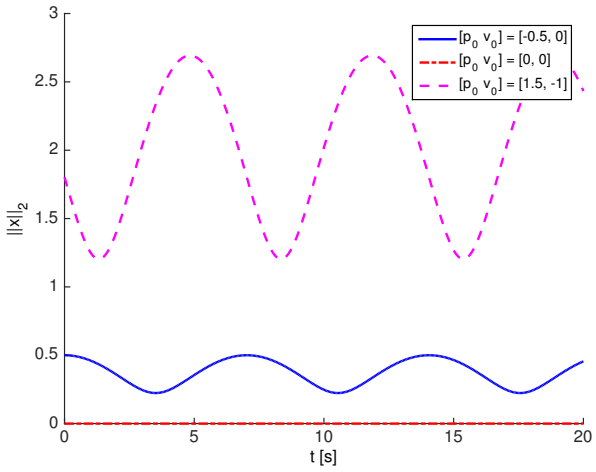
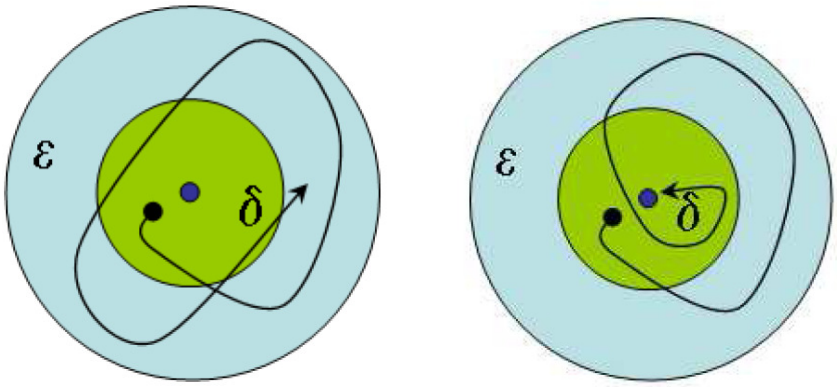


FIGURE 9. Mass-Spring-Damper example: norm for the system showing a persistent oscillation with $m = 10$ Kg, $K = 2$ N/m and $B = 0$.



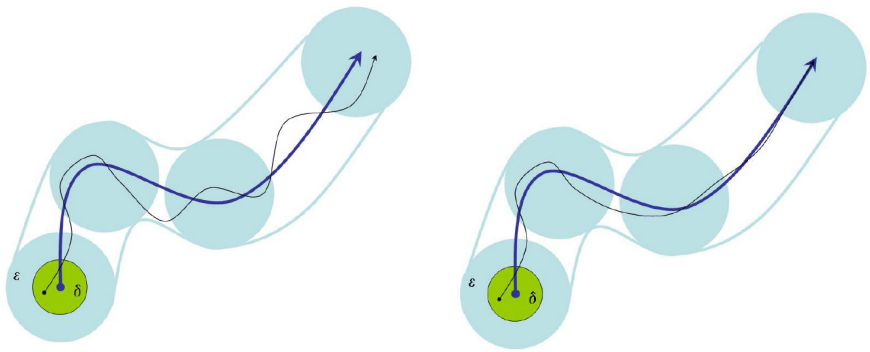
(a) (b)

FIGURE 10. a) Trajectory originating from a stable equilibrium point. b) Trajectory originating from an asymptotically stable equilibrium point.

DEFINITION 6.11. An equilibrium point $\bar{x}$ is *attractive* if all the trajectories $x(t) = \Phi x(x', u, t)$ that starts from initial points x' *sufficiently close* to $\bar{x}$ *converge* towards $\Phi x(\bar{x}, u, t)$, for $t \to +\infty$. □

More precisely: if $\exists \delta > 0$ s.t. $\|x' - \bar{x}\| < \delta$, then $\lim_{t\to+\infty} \|\Phi x(x', u, t) - \bar{x}\| = 0$.

DEFINITION 6.12. An equilibrium point $\bar{x}$ is *asymptotically stable* if it is stable and attractive. □



(a) (b)

FIGURE 11. a) Stable trajectory. b) Asymptotically stable trajectory.

The example of an asymptotically stable point is depicted in Figure 10-b.

DEFINITION 6.13. An equilibrium point $\bar{x}$ that is not stable is *unstable*. □

REMARK 6.14. The same concept of stability fro an equilibrium point $\bar{x}$ can be applied also to *reference trajectories* by substituting in the previous definitions the $\bar{x}$ with $\bar{x}(t) = \Phi x(x_0, u, t)$. □

Graphical representations of stable and asymptotically stable trajectories are given in Figure 11.

From the previous definitions, it is clear that the norm plays a crucial role. In particular, the theory relies on the definition of *positive definiteness* of a function.

DEFINITION 6.15. A function $V(x) : \mathbb{R}^n \mapsto \mathbb{R}$ is *positive definite* (p.d.) if and only if it is $V(x) > 0$, $\forall x \in B_r \backslash 0$ and $V(0) = 0$. It is globally p.d. if and only if it is $V(x) > 0$, $\forall x \in \mathbb{R}^n \backslash 0$ and $V(0) = 0$. □

DEFINITION 6.16. A function $V(x) : \mathbb{R}^n \mapsto \mathbb{R}$ is *positive semi definite* (p.s.d.) if and only if it is $V(x) \geq 0$, $\forall x \in B_r \backslash 0$ and $V(0) = 0$ (e.g. a quadratic form with $P \geq 0$). It is globally p.s.d. if and only if it is $V(x) \geq 0$, $\forall x \in \mathbb{R}^n \backslash 0$ and $V(0) = 0$. □

DEFINITION 6.17. A function $V(x) : \mathbb{R}^n \mapsto \mathbb{R}$ is *negative definite* (n.d.) if and only if it is $V(x) < 0$, $\forall x \in B_r \backslash 0$ and $V(0) = 0$ (e.g. a quadratic form with $P < 0$). It is globally n.d. if and only if it is $V(x) < 0$, $\forall x \in \mathbb{R}^n \backslash 0$ and $V(0) = 0$. □

DEFINITION 6.18. A function $V(x) : \mathbb{R}^n \mapsto \mathbb{R}$ is *negative semi definite* (n.s.d.) if and only if it is $V(x) \leq 0$, $\forall x \in B_r \backslash 0$ and $V(0) = 0$ (e.g. a quadratic form with $P \leq 0$). It is globally n.s.d. if and only if it is $V(x) \leq 0$, $\forall x \in \mathbb{R}^n \backslash 0$ and $V(0) = 0$. □

DEFINITION 6.19. A function $V(x) : \mathbb{R}^n \mapsto \mathbb{R}$ is not definite if it is not p.d., p.s.d, n.d. nor n.s.d. □

A trivial example of a p.d. function is the Euclidean 2-norm. Similarly, an example of a n.d. function is minus the Euclidean 2-norm.

DEFINITION 6.20. Given a function $V(x) : \mathbb{R}^n \mapsto \mathbb{R}$, the region of points in which $V(x) =$ constant is called a *level surface*. Usually, the level surfaces are ellipsoids in $\mathbb{R}^n$. □

DEFINITION 6.21. Let $V(x) : \mathbb{R}^n \mapsto \mathbb{R}$ be a continuous function with continuous first partial derivatives in a neighborhood of the origin of $\mathbb{R}^n$. If $V(x)$ is p.d., then $V(x)$ is called a *Lyapunov function candidate* for the system $\dot{x}(t) = \mathbf{f}(x(t), t)$ (where $u(t)$ is supposed to be zero). □

According to our intuition and to the stability definitions of an equilibrium point, if $V(x)$ is decreasing for increasing t, than the given solution trajectory must be converging toward the equilibrium point. This is the underlying idea of *Lyapunov stability* theorems. In order to understand if the Lyapunov function candidate $V(x)$ is decreasing or not, it is sufficient to determine $\dot{V}(x)$, which is given by the *total derivative* of $V(x)$, i.e.,

$$\begin{aligned}\dot{V}(x) &= \frac{\partial V(x)}{\partial x}\dot{x}(t) = \frac{\partial V(x)}{\partial x}\mathbf{f}(x,t) = \\ &= \frac{\partial V(x)}{\partial x_1}f_1(x,t) + \frac{\partial V(x)}{\partial x_2}f_2(x,t) + \cdots + \frac{\partial V(x)}{\partial x_n}f_n(x,t) = \\ &= \nabla V(x)\mathbf{f}(x,t) = L_{\mathbf{f}}V(x),\end{aligned}$$

where we made the implicit assumption that

$$\dot{x}(t) = \begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \vdots \\ \dot{x}_n(t)\end{bmatrix} = \mathbf{f}(x(t),t) = \begin{bmatrix}f_1(x,t)\\ f_2(x,t)\\ \vdots \\ f_n(x,t)\end{bmatrix}.$$

It is now quite easy to introduce the Lyapunov stability theorems.

THEOREM 6.22. *The equilibrium point of* $\dot{x}(t) = \mathbf{f}(x(t), t)$ *is* stable *if there exists a Lyapunov function candidate* $V(x)$ *with* $L_{\mathbf{f}}V(x) \leq 0$. □

THEOREM 6.23. *The equilibrium point of* $\dot{x}(t) = \mathbf{f}(x(t), t)$ *is* unstable *if there exists a Lyapunov function candidate* $V(x)$ *with* $L_{\mathbf{f}}V(x) > 0$. □

THEOREM 6.24. *The equilibrium point of* $\dot{x}(t) = \mathbf{f}(x(t), t)$ *is* asymptotically stable *if there exists a Lyapunov function candidate* $V(x)$ *with* $L_{\mathbf{f}}V(x) < 0$. □

We immediately notice that the stability property is a function of the system whatever is the input applied, indeed the theorems do not depend on the input. The following example clarify this point.

EXAMPLE 6.25. Let us recall the RL Example 6.2, whose equilibrium point is

$$\bar{x}_1 = \frac{u}{R}.$$

By defining the error variable $\tilde{x}_1 = x_1 - \bar{x}_1$, it is possible to define the following Lyapunov candidate

$$v(\tilde{x}_1) = \tilde{x}_1^2,$$

which is the distance of the state variable x_1 to the equilibrium point. Let us compute its time derivative

$$\dot{V}(\tilde{x}_1) = L_{\mathbf{f}} V(\tilde{x}_1) = 2\tilde{x}_1 \dot{\tilde{x}}_1 = 2\tilde{x}_1 \dot{x}_1,$$

since $\dot{\tilde{x}}_1 = \dot{x}_1 + \dot{\bar{x}}_1 = \dot{x}_1$ since $\bar{x}_1$ is constant. Then

$$\dot{V}(\tilde{x}_1) = 2\tilde{x}_1 \dot{x}_1 = 2\tilde{x}_1 \left(-\frac{R}{L} x_1 + \frac{u}{L} \right) = -2\frac{R}{L} \tilde{x}_1 \left(x_1 - \frac{u}{R} \right) = -2\frac{R}{L} \tilde{x}_1^2.$$

When $R = 2$ Ohm and $L = 1$ Henries, we have that $\dot{V}(\tilde{x}_1)$ is n.d., hence the system is asymptotically stable. Instead, when $R = -2$ Ohm and $L = 1$ Henries, we have that $\dot{V}(\tilde{x}_1)$ is p.d., hence the system is unstable. □

The most powerful characteristic of the Lyapunov theory is that stability can be proved *without* explicitly computing the solution of the differential equations.

REMARK 6.26. For discrete time systems, the time derivative is substituted with the *difference* of the Lyapunov function, i.e.,

$$\Delta_{\mathbf{f}} V(x) = V(\mathbf{f}(x)) - V(x).$$

□

6.1.3. Quadratic Forms. For the choice of the Lyapunov candidate, one standard choice comes from the *quadratic forms*.

DEFINITION 6.27. Given a matrix $P \in \mathbb{R}^{n \times n}$, the scalar function

$$V(x) = x^T P x$$

is said to be a *quadratic form*. □

For the quadratic form, the antisymmetric part of P is not relevant. Indeed, any matrix P can be written as the sum of its symmetric and antisymmetric part, i.e.,

$$P = \frac{P + P^T}{2} + \frac{P - P^T}{2}.$$

It then follows immediately that

$$V(x) = x^T P x = x^T \left(\frac{P + P^T}{2} + \frac{P - P^T}{2} \right) x = x^T \frac{P + P^T}{2} x.$$

We should emphasise that $V(x)$ is p.d. if P is p.d. (i.e., $P > 0$). Moreover, $V(x)$ is p.s.d. if P is p.s.d. (i.e., $P \geq 0$).

DEFINITION 6.28. A real, symmetric matrix $P \in \mathbb{R}^{n\times n}$ is *positive definite* if (all the following statements are equivalent):

(1) $x^T P x > 0$ for every nonzero $x \in \mathbb{R}^n$;
(2) all eigenvalues of P are positive;
(3) $P = R^T R$ for some nonsingular R;
(4) the leading principal minors of P are positive;
(5) all principal minors of P are positive.

□

It is easy to see that if P is p.d. (or p.s.d.), then $-P$ will be n.d. (or n.s.d.).

The condition $P = R^T R$ highlights the fact that a quadratic form can be seen as the Euclidean 2-norm of a vector in a specific set of coordinates. Indeed,

$$x^T P x = x^T R^T R x = y^T y = \|y\|_2^2,$$

where $y = Rx$.The choice of R is not unique.

6.2. Stability of Linear Systems

6.2.1. Continuous Time Systems. Let us consider the following linear system

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

and the Lyapunov function candidate

$$V(x) = x^T P x.$$

The time derivative is then given by

$$\dot{V}(x) = 2x^T P\dot{x} = 2x^T PAx = x^T(A^T P + PA)x \triangleq -x^T Q x.$$

In other words, $-Q$ is the symmetric part of $2PA$. The equation

$$A^T P + PA = -Q$$

is termed the *continuous time Lyapunov equation.*

Then, it follows immediately that:

(1) If P is p.d. and Q is p.s.d., then the system is *stable*;
(2) If P is p.d. and Q is p.d., then the system is *asymptotically stable.*

In general, if P is arbitrarily chosen, then Q is not defined. A solution instead exists in the reverse order: fixing Q p.d. and assuming that A is asymptotically stable, than the solution is given by

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

From the previous equation it follows that the modes analysis is equally complex to finding a solution for the Lyapunov function.

6.2.2. Discrete Time Systems. Let us consider the following linear system

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

and the Lyapunov function candidate

$$V(x) = x^T Px.$$

The Lyapunov function difference is then given by

$$\Delta V(x) = x(k+1)^T Px(k+1) - x(k)^T Px(k) = x(k)^T (A^T PA + P)x(k) \triangleq -x^T Qx.$$

The equation

$$A^T PA + P = -Q$$

is termed the *discrete time Lyapunov equation*.

Then, it follows immediately that:

(1) If P is p.d. and Q is p.s.d., then the system is *stable*;
(2) If P is p.d. and Q is p.d., then the system is *asymptotically stable*.

Again, if P is arbitrarily chosen, then Q is not defined. A solution instead exists in the reverse order: fixing Q p.d. and assuming that A is asymptotically stable, than the solution is given by

$$P = \sum_{i=0}^{+\infty} (A^T)^i Q A^i.$$

As for the continuous time case, from the previous equation it follows that the modes analysis is equally complex to finding a solution for the Lyapunov function.

APPENDIX A

Some mathematical definition of interest

We now list some mathematical definition of interest. In the text below we will refer to a set denoted as $\mathcal{T}$and to its elements, denoted by lower capitals letters such as $t, t_1, t_2, \ldots$.

DEFINITION A.1. Un insieme $\mathcal{T}$ si dice totalmente ordinato se su di essa è definita una relazione $\leq$ di ordinamento totale definita su coppie degli elementi dell'insieme. Una definizione di ordinamento totale gode delle seguenti proprietà:

(1) Riflessività: $\forall t \in \mathcal{T} : t \leq t$;
(2) Antisimmetria: $t_1 \leq t_2 \wedge t_2 \leq t_1 \rightarrow t_1 = t_2$
(3) Transitività: $t_1 \leq t_2 \wedge t_2 \leq t_3 \rightarrow t_1 \leq t_3$
(4) Confrontabilità (tricotomia): $\forall t_1, t_2 \in \mathcal{T}$ o $t_1 \leq t_2$ oppure $t_2 \leq t_1$.

Se valgono solo le prime tre proprietà si parla di ordinamento parziale.

Come esempio tutti i possibili spazi rappresentati il temp nel nostro corso sono almeno parzialmente ordinati.

Diamo ora la definizione di spazio metrico.

DEFINITION A.2. Consideriamo una funzione $d : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$, detta distanza, definita sulle coppie di elementi di un insieme $\mathcal{T}$ che associa a ciascuna coppia di punti un numero reale. Tale funzione definisce una metrica se:

- Positività: $\forall t_1, t_2 \in \mathcal{T}$, $d(t_1, t_2) \geq 0$ e $d(t_1, t_2) = 0$ se e solo se $t_1 = t_2$.
- Simmetria: $\forall t_1, t_2 \in T$: $d(t_1, t_2) = d(t_2, t_1)$.
- Disuguaglianza triangolare: $\forall t_1, t_2, t_2 \in \mathcal{T}$: $d(t_1, t_3) \leq d(t_1, t_2) + d(t_2, t_3)$.

DEFINITION A.3. Uno spazio metrico è un insieme su cui sia definita una metrica.

DEFINITION A.4. A point t is a limit point of a set $\mathcal{T}$if for any $\epsilon > 0$ there exist another point t_1 different from t such that $d(t, t_1) \leq \epsilon$, where $d(.,.)$ is a metric.

DEFINITION A.5. The space $\mathcal{T}$is continuum if it is a closed connected set. A connected set is one where no matter how it is divided in two disjoint set, at least one of them will contain limit points of the other. A closed set is one that contains the limit poitns of all its subsets.

The time space $\mathcal{T}$ for CT systems is a continuum because we cannot pose constraints on when events and observations occurs, just as it happens in the physical world.

Bibliography

[Hes09] João P Hespanha, *Linear systems theory*, Princeton university press, 2009.

[LSV98] Edward A Lee and Alberto Sangiovanni-Vincentelli, *A framework for comparing models of computation*, Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on **17** (1998), no. 12, 1217–1229.

[PPR95] Charles L Philips, John M Parr, and E Riskin, *Signals, systems, and transforms*, Prentice Hall, 1995.

[RI79] Antonio Ruberti and Alberto Isidori, *Teoria dei sistemi*, Boringhieri, 1979.