Università degli Studi di Trento

Department of Industrial Engineering
Automatic Control
Prof.: Zaccarian Luca

Course Notes

Matteo Dalle Vedove
matteo.dallevedove@studenti.unitn.it

Academic Year 2021-2022
March 31, 2022

Contents

Chapter 1

Dynamical systems and linear systems theory

This course will treat the **automatic control** of **dynamical systems**, hence it's necessary to firstly understand what a *dynamical system* is and how it's described. In general such systems can be regarded as *black box* that for a given **input** $u(t)$ determines an **output** $y(t)$; systems can be **continuous** or **discrete-time** depending on the type of *time axis scale*, but also **hybrid** systems exists (as it will be studied).

Rabbits   In order to understand behaviours of dynamical system, let's consider the example of pairs of rabbit. If we describe with $y(t)$ the total number of rabbits couple that one person has (where t is the time expressed in months), we can consider that initially a person start with no animals ($y(0) = 0$) and goes at $t = 1$ at an animal shop to buy a couple of small rabbits (hence $y(1) = 1$). In the first month of life they grow up and from *babies* they become *adults*, but the person still has $y(2) = 1$ couple of rabbit. In the next month they procreate a new couple of rabbits and so we have $y(3) = 2$ rabbit pairs. Assuming that no rabbit dies, we can describe the evolution of the system as

$$y(n) = y(n - 2) + y(n - 1) \tag{1.1}$$

where the term $y(n - 2)$ relates to the elder population and $y(n - 1)$ describes the newly borne couple from each pair of elder pair. Equation 1.1 is a **finite difference equation** that allows to recursively define the number of rabbit pairs depending on the time t assuming to know the **initial conditions** $y(0)$ and $y(1)$.

For such system we can define it's **states** $x_1(t) = y(t)$ and $x_2 = y(t - 1)$ determining the vector $x(t)$:

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \tag{1.2}$$

Given so the states $x(t)$, we can compute their evolution $x(t + 1)$ as

$$x(t + 1) = \begin{pmatrix} x_1(t + 1) \\ x_2(t + 1) \end{pmatrix} = \begin{pmatrix} y(t + 1) \\ y(t) \end{pmatrix} = \begin{pmatrix} y(t - 1) + y(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x_2(t) + x_1(t) \\ x_1(t) \end{pmatrix}$$

Using linear algebra the evolution of the state can be regarded as a linear combination of the actual states, in fact

$$x(t + 1) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \mathcal{A}x(t) \tag{1.3}$$

Having the eigenvalues of the matrix $\mathcal{A}$ equal to $\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}$ we will prove that the continuous explicit function that express the rabbit pairs y as function of t is

$$y(t) = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^t - \left(\frac{1 - \sqrt{5}}{2} \right)^t \right]$$

Classification of system This example allowed to intrude the first classification of system: they in fact can be **linear** if the states variation and output can be regarded as linear combination of the actual states and inputs, meaning that they can be expressed as

$$\begin{cases} x(t+1) = \mathcal{A}x(t) + \mathcal{B}u(t) \\ y(t) = \mathcal{C}x(t) + \mathcal{D}u(t) \end{cases} \tag{1.4}$$

where $\mathcal{A}$ is the **dynamic matrix**, $\mathcal{B}$ the **input** m., $\mathcal{C}$ the **output** m. and $\mathcal{D}$ is the **feed-through** (instantaneous) m.; both u, x and y are in general vectors. To simplify the notation usually the dependency on the time t is dropped and the increment for **discrete-time** systems (where $t \in \mathbb{Z}$) is described as x^+ (values of the state at the next step), hence equation 1.4 is rewritten as

$$\begin{cases} x^+ = \mathcal{A}x + \mathcal{B}u \\ y = \mathcal{C}x + \mathcal{D}u \\ x(t_0) = x_0 \end{cases} \tag{1.5}$$

For **continuous-time** systems (where $t \in \mathbb{R}$) the concept of increment is represented by the derivative $\dot{x} = \frac{dx}{dt}$, hence

$$\begin{cases} \dot{x} = \mathcal{A}x + \mathcal{B}u \\ y = \mathcal{C}x + \mathcal{D}u \\ x(t_0) = x_0 \end{cases} \tag{1.6}$$

Note that for all the system is very important to define the **initial condition** $x(t_0) = x_0$ of the states because the response of the output is strictly related to them.

System that don't exhibits a linear behaviour are therefore called **non-linear** and the input/state/output relations are described by functions:

$$\begin{cases} \delta x = f(x, u) \\ y = h(x, y) \\ x(t_0) = x_0 \end{cases} \tag{1.7}$$

> **Note:** the term δx is the generalization of $\dot{x}$ for continuous-time or x^+ for discrete-time systems in order to make the notation more generalized and will be used throughout the whole book.

Until now the notation were referred to **time-invariant** systems where the matrix $\mathcal{A}, \mathcal{B}\mathcal{C}, \mathcal{D}$ (for linear sys.) or the functions f, h (for non-linear sys.) were not dependent on time, however **time-variant** systems also exists that exhibiting an explicit relation with time, hence can be regarded as

$$\begin{cases} \delta x = \mathcal{A}(t)x + \mathcal{B}(t)u \\ y = \mathcal{C}(t)x + \mathcal{D}(t)u \\ x(t_0) = x_0 \end{cases} \quad \text{or} \quad \begin{cases} \delta x = f(x, u, t) \\ y = h(x, y, t) \\ x(t_0) = x_0 \end{cases} \tag{1.8}$$

for both linear and non-linear case.

By a mathematical point of view a system if it satisfy that

$$\begin{cases} x(t_0 + T) = x_0 \\ u(t - T) \end{cases} \quad \Rightarrow \quad y(t - T) \quad \quad \forall t \geq t_0 + T$$

ESEMPI DEI DIVERSI TIPI DI SISTEMI

Example 1.1: linear time-variant dynamical system

Let's consider the model of a 3 year course of an university with the goal of study the behaviour of students. Considering the graphs that follows, we denote as u the number of students registering at the first year, while $\alpha_i \in [0, 1]$ is the relative value of students that

passes to the next year/graduates; y finally represent the number of student graduated. In this case the states x_i are representing the number of student in each year of the course. This models assumes that no student drops the course (more complex models can indeed consider such phenomena).
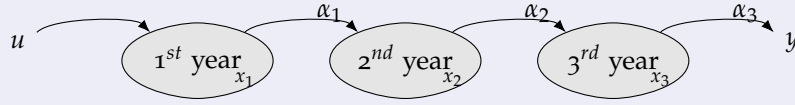
We can see that this dynamical system is discrete-time (the system is considered to evolved each year at the start of the academic year); analysing the graph we observe that each next year the number of student in each *class* (x_i) depends on the number of the one passed from the previous course ($\alpha_i x_{i-1}$) and the one that didn't passed ($(1 - \alpha_i)x_i$), hence we have that

$$\begin{cases} x_1^+ &= (1 - \alpha_1)x_1 + u \\ x_2^+ &= (1 - \alpha_2)x_2 + \alpha_1 x_1 \\ x_3^+ &= (1 - \alpha_3)x_3 + \alpha_2 x_2 \\ y &= \alpha_3 x_3 \end{cases}$$

In this case if we assume that α_i are constant value, the model that we obtain doesn't suit the real behaviour of the course where problems might happen (for example a *evil teacher* reduces the rate of passed students one year, reducing α_i), hence we have a system that's time-variant (α_i depends on time t) while the model is still linear, because it can be expressed as

$$x^+ = \begin{pmatrix} x_1^+ \\ x_2^+ \\ x_3^+ \end{pmatrix} = \begin{bmatrix} 1 - \alpha_1(t) & 0 & 0 \\ \alpha_1(t) & 1 - \alpha_2(t) & 0 \\ 0 & \alpha_2(t) & 1 - \alpha_3(t) \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} u = \mathcal{A}(t)x + \mathcal{B}(t)u$$

$$y = \begin{bmatrix} 0 & 0 & \alpha_3(t) \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{bmatrix} 0 \end{bmatrix} u = \mathcal{C}(t)x + \mathcal{D}(t)u$$

1.1 Linear time-invariant systems and transfer functions

As equations 1.5 and 1.6 stated, **linear time-invariant** LTI systems can be described in matrix form as

$$\begin{cases} \delta x = \mathcal{A}x + \mathcal{B}u & x(0) = x_0 \\ y = \mathcal{C}x + \mathcal{D}u \end{cases}$$

where the matrix $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ are constant (hence independent from time due to the time-invariancy requirement). Having the system time invariant the initial condition can be expressed as $x(0)$ (because it doesn't depend on the choice of t_0, but on any general initial condition).

Continuous-time LTI systems can be easily analysed in the domain of the complex variable s using the **Laplace transform** $\mathcal{L}$, an operator defined

$$X(s) = \mathcal{L}\{x(t)\} = \int_0^\infty x(t)e^{-st}\, dt \qquad\qquad s \in \mathbb{C}$$

An important property of such operator is that converts differential equations into algebraic ones in the variable s, in fact we can observe that

$$\dot{x}(t) \mapsto sX(s) - x(0)$$

The Laplace transform has also the useful property of being linear, hence the input-output relation of LTI system can be transformed according to Laplace, hence

$$(eq.1.6) \mapsto \begin{cases} sX(s) - x(0) = \mathcal{A}X(s) + \mathcal{B}U(s) \\ Y(s) = \mathcal{C}X(s) + \mathcal{D}U(s) \end{cases} \qquad\qquad (1.9)$$

The first equation can be solved in order to explicit the transform of the states $X(s) = \mathscr{L}\{x(s)\}$ as function of the transform of the input $U(s) = \mathscr{L}\{u(t)\}$ determining

$$sX(s) - \mathcal{A}X(s) = \mathcal{B}U(s) + x(0)$$
$$(s\mathcal{I} - \mathcal{A})X(s) = \mathcal{B}U(s) + x(0)$$
$$X(s) = (s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B}U(s) + (sI - \mathcal{A})^{-1}x(0)$$

> **Note:** In the first step in order to collect $sX(s)$ and $\mathcal{A}X(s)$ has been necessary to multiply s by the identity matrix $\mathcal{I}$: s is in fact a scalar while $\mathcal{A}$ is a matrix and no operation is compatible between this elements (addition is defined only for matrix with same sizes).

With this result obtained we can compute the transform of the output $Y(s) = \mathscr{L}\{y(t)\}$ as function of the lonely $U(s)$ as

$$Y(s) = \underbrace{\left(\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}\right)}_{=\hat{\mathcal{G}}(s)} U(s) + \underbrace{\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{D}}_{\hat{\psi}(s)} x(0) \qquad (1.10)$$

In this equation $\hat{\mathcal{G}}(s)$ represent the **transfer function** of the system (however more generally it's a matrix). Observing that the output in the Laplace domain can be so simplified to the form $Y(s) = \hat{G}(s)U(s) + \hat{\psi}(s)x(0)$ and so by using property of the (inverse) Laplace transform we can rewrite the output as the convolution of the input sequence and the **impulse response** $G(t) = \mathscr{L}^{-1}\{\hat{G}(s)\}$ of the system:

$$y(t) = (G * x)(t) + \psi(t)x(0) \qquad (1.11)$$

where the convolution is defined as

$$(G * x)(t) = \int_0^t G(t)x(t - \tau)\, d\tau$$

Considering the case of discrete-time LTI systems instead of the Laplace we have to use the **Z transform** $\mathscr{Z}$ that maps discrete time sequences into a domain in the complex variable z; knowing that $X(z) = \mathscr{Z}\{x(t)\}$ (with $t \in \mathbb{Z}$) the analogous property of differentiation in time is that $x(t + 1) = x^+ \mapsto zX(z) - zx(0)$, hence

$$(eq.1.5) \mapsto \begin{cases} zX(z) - zx(0) = \mathcal{A}X(z) + \mathcal{B}U(z) \\ Y(z) = \mathcal{C}X(z) + \mathcal{D}U(z) \end{cases} \qquad (1.12)$$

As in the continuous case it's possible to compute the transform of the output $Y(z) = \mathscr{Z}\{y(t)\}$ as function of the spectrum $U(z) = \mathscr{Z}\{x(t)\}$ as

$$Y(z) = \underbrace{\left(\mathcal{C}(z\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}\right)}_{=\hat{\mathcal{G}}(z)} U(z) + \underbrace{\mathcal{C}(z\mathcal{I} - \mathcal{A})^{-1}\mathcal{D}z}_{\hat{\psi}(z)} x(0) \qquad (1.13)$$

Theorem A fundamental theorem for continuous-time linear time-invariant systems states that *the transfer function* $\hat{\mathcal{G}}(s)$ *and so the* **impulse response** $\mathcal{G}(t)$ *of such systems is computed as*

$$\hat{\mathcal{G}}(s) = \left(\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}\right) \qquad \mathcal{G}(s) = \mathscr{L}^{-1}\left\{\left(\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}\right)\right\} \qquad (1.14)$$

1.2 Interconnected systems

The control of dynamical systems is based on **interconnections** between, and such connection can be summarized as: in parallel, in series, in feedback. In this section a proof of transfer function of interconnected system is given in the case of linear time-invariant system (where the concept of transfer function is *well defined*). Such system represented as

$$\begin{cases} \delta x = \mathcal{A}x + \mathcal{B}u \qquad x(0) = x_0 \\ y = \mathcal{C}x + \mathcal{D}u \end{cases}$$

are so characterized by the 4 matrix $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ by the *big matrix notation*

$$\begin{pmatrix} \delta x \\ y \end{pmatrix} = \left[\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array} \right] \begin{pmatrix} x \\ u \end{pmatrix}$$

Parallel connection

Considering the parallel connection of the systems Σ_1, Σ_2 (figure 1.1), the overall transfer function of the system Σ can be computed considering that the output $Y(\cdot)$ in the Laplace/Z domain is the sum of the contribution of the single stages that can be computed so independently

$$\begin{aligned} Y = Y_1 + Y_2 = \hat{\mathcal{G}}_1 U_1 + \hat{\mathcal{G}}_2 U_2 &= \hat{\mathcal{G}}_1 U + \hat{\mathcal{G}}_2 U \\ &= \underbrace{\left(\hat{\mathcal{G}}_1 + \hat{\mathcal{G}}_2 \right)}_{=\hat{\mathcal{G}}} U \end{aligned} \tag{1.15}$$

While connection two or more systems we have that the overall states x of the system Σ is the union of the individual systems x_i, hence in this case

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

More formally the we have that the variation of the states δx of the system Σ can be regarded as

$$\delta x = \mathcal{A}_1 x_1 + \mathcal{B}_1 u_1 + \mathcal{A}_2 u_2 + \mathcal{B}_2 u_2 = \begin{bmatrix} \mathcal{A}_1 & 0 \\ 0 & \mathcal{A}_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \end{bmatrix} u$$

and similarly the output in the time domain

$$y = \mathcal{C}_1 x_1 + \mathcal{D}_1 u_1 + \mathcal{C}_2 x_2 + \mathcal{D}_2 x_2 = \begin{bmatrix} \mathcal{C}_1 & \mathcal{C}_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (\mathcal{D}_1 + \mathcal{D}_2) u$$

The associated *big matrix* is so

$$\begin{pmatrix} \delta x_1 \\ \delta x_2 \\ y \end{pmatrix} = \left[\begin{array}{cc|c} \mathcal{A}_1 & 0 & \mathcal{B}_1 \\ 0 & \mathcal{A}_2 & \mathcal{B}_2 \\ \hline \mathcal{C}_1 & \mathcal{C}_2 & \mathcal{D}_1 + \mathcal{D}_2 \end{array} \right] \begin{pmatrix} x_1 \\ x_2 \\ u \end{pmatrix} \tag{1.16}$$

Series (cascade) connection

A series connection (figure 1.2) between two system is implemented by feeding the input of Σ_2 with the output of Σ_1; in this case the input-output of the overall system can be regarded in the Laplace/Z transform as

$$\begin{aligned} Y = \hat{\mathcal{G}}_2 U_2 = \hat{\mathcal{G}}_2 Y_1 &= \hat{\mathcal{G}}_2 \hat{\mathcal{G}}_1 U_1 \\ &= \hat{\mathcal{G}}_2 \hat{\mathcal{G}}_1 U = \hat{\mathcal{G}} U \end{aligned} \tag{1.17}$$
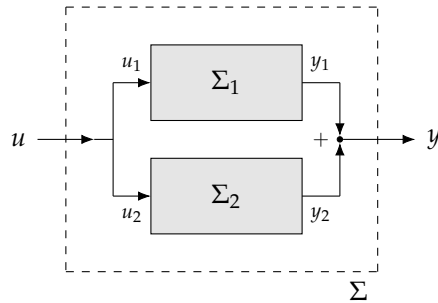

Figure 1.1: *black box representation of a parallel connection between two systems.*

Note that in this case the order of the transfer function $\hat{\mathcal{G}} = \hat{\mathcal{G}}_2\hat{\mathcal{G}}_1$ is important due to the multiplication of matrix; in fact in order to make the series connection possible we need to have that the length of the vector y_1 is equal to the one of x_2: this makes possible to have a matrix $\hat{\mathcal{G}}$ the can be multiplied by U and returns a proper Y.

Carefully representing this system in the space of the states, we obtain that

$$
\begin{aligned}
\delta x_1 &= \mathcal{A}_1 x_1 + \mathcal{B}_1 u \\
y_1 &= \mathcal{C}_1 x_1 + \mathcal{D}_1 u = u_2 \\
\delta x_2 &= \mathcal{A}_2 x_2 + \mathcal{B}_2 u_2 = \mathcal{A}_2 x_2 + \mathcal{B}_2(\mathcal{C}_1 x_1 + \mathcal{D}_1 u) \\
y = y_2 &= \mathcal{C}_2 x_2 + \mathcal{D}_2 u_2 = \mathcal{C}_2 x_2 + \mathcal{D}_2(\mathcal{C}_1 x_1 + \mathcal{D}_1 u)
\end{aligned}
$$

hence the *big matrix* is

$$
\left[
\begin{array}{cc|c}
\mathcal{A}_1 & 0 & \mathcal{B}_1 \\
\mathcal{B}_2\mathcal{C}_1 & \mathcal{A}_2 & \mathcal{B}_2\mathcal{D}_1 \\
\hline
\mathcal{D}_2\mathcal{C}_1 & \mathcal{C}_2 & \mathcal{D}_2\mathcal{D}_1
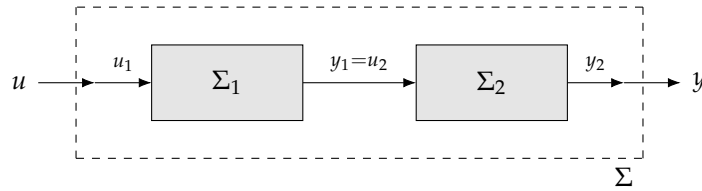\end{array}
\right]
\tag{1.18}
$$

Figure 1.2: black box representation of a series connection between two systems.

Feedback connection

A feedback connection (figure 1.3) of a system is realised by subtracting to the input u of the system it's output y. Performing the analysis in the frequency domain we have have in fact the solution

$$
Y = \hat{\mathcal{G}}_1 U_1 = \hat{\mathcal{G}}_1 (U - Y)
$$

We see that this expression is implicit in Y, but we can still obtain the solution

$$
\begin{aligned}
Y + \hat{\mathcal{G}}_1 Y = Y(\mathcal{I} + \hat{\mathcal{G}}_1) &= \hat{\mathcal{G}}_1 U \\
Y &= (\mathcal{I} + \hat{\mathcal{G}}_1)^{-1}\hat{\mathcal{G}}_1 U
\end{aligned}
\tag{1.19}
$$

This solution is obtained by inverting the matrix $\mathcal{I} + \hat{\mathcal{G}}_1$ and this operation is possible if the matrix is *well-posed*, condition that happens every time the matrix $\mathcal{I} + \mathcal{D}_1$ is invertible. By performing the analysis in the states space we have in fact

$$
\begin{aligned}
y = \mathcal{C}_1 x_1 + \mathcal{D}_1 u_1 &= \mathcal{C}_1 x_1 + \mathcal{D}_1(u - y) \\
(\mathcal{I} + \mathcal{D}_1)y &= \mathcal{C}_1 x_1 + \mathcal{D}_1 u \\
y &= (\mathcal{I} + \mathcal{D}_1)^{-1}\mathcal{C}_1 x_1 + (\mathcal{I} + \mathcal{D}_1)^{-1}\mathcal{D}_1 u
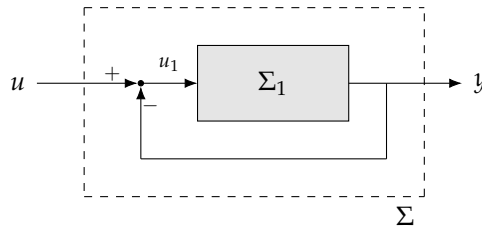\end{aligned}
$$

Figure 1.3: black box representation of a series feedback connection of a system.

From this expression is possible to observe why the well-posedness condition is related to having $\mathcal{I} + \mathcal{D}_1$ invertible. Expliciting the variation of the states we have instead we have

$$
\begin{aligned}
\delta x &= \mathcal{A}_1 x_1 + \mathcal{B}_1 u_1 = \mathcal{A}_1 x_1 + \mathcal{B}_1 (u - y) \\
&= \mathcal{A}_1 x_1 + \mathcal{B}_1 u - \mathcal{B}_1 \left((\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{C}_1 x_1 + (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{D}_1 u \right) \\
&= \left(\mathcal{A}_1 - \mathcal{B}_1 (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{C}_1 \right) x_1 + \left(\mathcal{B}_1 - \mathcal{B}_1 (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{D}_1 \right) u
\end{aligned}
$$

From this states space expansion we can define the *big matrix* associated to a feedback connection as

$$
\left[\begin{array}{c|c} \mathcal{A}_1 - \mathcal{B}_1 (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{C}_1 & \mathcal{B}_1 \left(\mathcal{I} - (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{D}_1 \right) \\ \hline (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{C}_1 & (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{D}_1 \end{array} \right] = \left[\begin{array}{c|c} \mathcal{A}_1 - \mathcal{B}_1 (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{C}_1 & \mathcal{B}_1 (\mathcal{I} + \mathcal{D}_1)^{-1} \\ \hline (\mathcal{I} + \mathcal{D}_1)^{-1} \mathcal{C}_1 & \mathcal{D}_1 (\mathcal{I} + \mathcal{D}_1)^{-1} \end{array} \right]
$$

$$(1.20)$$

1.3 Realization theory

Given a rational polynomial $g(s)$ so defined as the ratio between two polynomial $n(s)/d(s)$ (this expression indeed represent the case of a one dimensional transfer function), such function is said

- **strictly proper** if the degree $\#n(s)$ of the denominator is less then the degree $\#d(s)$ od the denominator; this means that
$$\lim_{s \to \infty} g(s) = 0$$

- **proper** if in general we have $\#n(s) \leq \#d(s)$, but more specifically it has to be
$$\lim_{s \to \infty} g(s) \neq \infty$$

- every time we have $\#n(s) > \#d(s)$, the rational polynomial is defined as **improper** and has
$$\lim_{s \to \infty} g(s) = \infty$$

A rational polynomial $G(s)$ is **biproper** if both g and it's inverse g^{-1} are propers. Extending this concepts to a transfer function $\hat{\mathcal{G}}(s)$ (that's in general a matrix), the system is (strictly) proper if all it's entries are (strictly) proper; if one entry is improper, then $\hat{\mathcal{G}}(s)$ is also improper.

In general it's proven that for linear time-invariant systems the transfer function $\hat{G}(s)$ is composed by a strictly proper and a proper term as

$$\hat{\mathcal{G}}(s) = \underbrace{\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B}}_{\text{strictly proper}} + \underbrace{\mathcal{D}}_{\text{proper}} \tag{1.21}$$

Realization Given a transfer function $\hat{\mathcal{G}}(s)$, we say that the linear time-invariant system in the form

$$
\begin{cases}
\delta x = \mathcal{A}x + \mathcal{B}u \qquad x(0) = x_0 \\
y = \mathcal{C}x + \mathcal{D}u
\end{cases}
$$

is a **realization** of the transfer function $\hat{G}$ if equation 1.21 holds.

This states that a LTI system characterized by the matrices $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ is the realization of $\hat{G}(s)$ if it happens that $\hat{G}(s) = \mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}$. We also define two realization as **zero state equivalent** if they realise the same transfer function $\hat{G}(s)$.

Realizations and transfer functions

An important theorem states that a transfer function $\hat{\mathcal{G}}(s)$ can be **realised** by an LTI system in the *standard form* **if and only if** the rational function $\hat{\mathcal{G}}(s)$ is **proper**.

Proof pt. 1 We can prove the *only if* conditions stating

$$\left[\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right] \mapsto \hat{\mathcal{G}}(s)$$

the proof is straightforward: as 1.21 states, given the constant matrices $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ the resulting transfer function is strictly proper for the term $\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B}$ and becomes proper only in the case when $\mathcal{D} \neq 0$.

Proof pt. 2 More complex is proving the *if* condition that starting from the transfer function computes the *characteristic* matrices of LTI systems:

$$\hat{\mathcal{G}}(s) \mapsto \left[\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right]$$

We want now to prove that if the transfer function $\hat{\mathcal{G}}(s)$ is proper then it state space representation exists. Determining the non-strictly proper element $\mathcal{D}$ is straightforward, in fact it can be computed simply as the limit

$$\mathcal{D} = \lim_{s \to \infty} \hat{\mathcal{G}}(s)$$

We can so build the strictly proper transfer function as

$$\hat{\mathcal{G}}_{sp}(s) = \mathcal{B}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{C} = \hat{\mathcal{G}}(s) - \mathcal{D}$$

Building the polynomial $d(s)$ as the least common denominator of all the entries in the matrix $\hat{\mathcal{G}}_{sp}(s)$ and writing him in a form

$$d(s) = s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_{n-1}s + \alpha_n$$

we can rewrite the strictly proper part of the transfer function as a combination of matrices N_i in the following way:

$$\hat{\mathcal{G}}_{sp}(s) = \frac{N(s)}{d(s)} = \frac{N_1 s^{n-1} + N_2 s^{n-2} + \cdots + N_{n-1}s + N_n}{s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_{n-1}s + \alpha_n}$$

With that said we have all the elements to compute the state space representation of the system: it can be proven in fact that the LTI has a state space representation in the form

$$\left[\begin{array}{cccc|c} -\alpha_1 \mathcal{I}_k & -\alpha_2 \mathcal{I}_k & \dots & -\alpha_n \mathcal{I}_k & \mathcal{I}_k \\ \mathcal{I}_k & 0 & \dots & 0 & 0 \\ 0 & \ddots & & 0 & \vdots \\ 0 & 0 & \mathcal{I}_k & 0 & 0 \\ \hline N_1 & N_2 & \dots & N_m & \mathcal{D} \end{array}\right] \tag{1.22}$$

where $\mathcal{I}_k$ is the $k \times k$ identity matrix; this notation is called **controllable canonical form**.

> **Example 1.2: state representation from a transfer function**
>
> Given a system characterized by a transfer function
>
> $$\hat{\mathcal{G}}(s) = \left[\begin{array}{cc} \frac{4s-10}{s+1} & \frac{3}{s+2} \\ \frac{1}{s+2} & \frac{4}{s+1} \end{array}\right]$$
>
> in order to compute one it's representation the first thing to do is extract the output matrix $\mathcal{D}$ as
>
> $$\mathcal{D} = \lim_{s \to \infty} \hat{\mathcal{G}}(s) = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

And so we can compute the strictly proper part as

$$\hat{G}_{sp}(s) = \hat{G}(s) - \mathcal{D} = \begin{bmatrix} \frac{-14}{s+1} & \frac{3}{s+2} \\ \frac{1}{s+2} & \frac{4}{s+1} \end{bmatrix}$$

$$= \frac{1}{(s+1)(s+2)} \begin{bmatrix} -14(s+2) & 3(s+1) \\ 1(s+1) & 4(s+2) \end{bmatrix} = \frac{1}{s^2+3s+2} \begin{bmatrix} -14s-28 & 3s+3 \\ s+1 & 4s+8 \end{bmatrix}$$

$$= \frac{\begin{bmatrix} -14 & 3 \\ 1 & 4 \end{bmatrix} s + \begin{bmatrix} -28 & 3 \\ 1 & 8 \end{bmatrix}}{s^2+3s+2}$$

Using equation 1.22 we can compute the state representation of the system using the yet computed denominator and matrices N_i providing the result

$$\left[\begin{array}{cccc|cc} -3 & 0 & -2 & 0 & 1 & 0 \\ 0 & -3 & 0 & -2 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline -14 & 3 & -28 & 3 & 4 & 0 \\ 1 & 4 & 1 & 8 & 0 & 0 \end{array} \right]$$

By computing $\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}$ using the matrices yet obtain we can indeed see that, after some simplifications, that we obtain the original transfer function $\hat{G}(s)$.

Proof pt. 3 All that's left now is to prove that equation 1.22 indeed is a realization of the proper transfer function $\hat{G}(s)$. In order to do so we consider we consider the sub-matrix $z(s)$ defined as

$$z(s) = \begin{bmatrix} z_1^t & z_2^t & \cdots & z_n^t \end{bmatrix} = (s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} \qquad \Rightarrow \qquad (s\mathcal{I} - \mathcal{A})z(s) = \mathcal{B}$$

Considering the *lower part* of the matrices $\mathcal{A}$ and $\mathcal{B}$ (the last 3 rows in equation 1.22), the multiplication results in the set of equalities

$$sz_2 - z_1 = 0 \qquad sz_3 - z_2 = 0 \qquad \cdots \qquad sz_n = z_{n-1}$$

Considering so that $z_n = \frac{1}{s}z_{n-1}$ but also $z_{n-1} = \frac{1}{s}z_{n-2}$ up to $z_2 = \frac{1}{s}z_1$ we can determine the generic vector z_k as

$$z_k = \frac{1}{s^{k-1}}z_1 \tag{1.23}$$

Considering so now the multiplication $(s\mathcal{I} - \mathcal{A})z(s) = \mathcal{B}$ using the top row of $\mathcal{A}, \mathcal{B}$ respect to the formulation in equation 1.22, we have that

$$(s + \alpha_1)z_1 + \alpha_2 z_2 + \cdots + \alpha_n z_n = \mathcal{I}_k$$

hence for relation 1.23

$$\left(s + \alpha_1 + \frac{\alpha^2}{s} + \cdots + \frac{\alpha_{n-1}}{s^{n-2}} + \frac{\alpha_n}{s^{n-1}} \right) z_1 = \mathcal{I}_k$$

Since the polynomial in the left-hand side is given by the ratio $\frac{d(s)}{s^{n-1}}$ we can define the sub-matrix z as

$$z(s) = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \frac{1}{d(s)} \begin{bmatrix} s^{n-1}\mathcal{I}_k \\ s^{n-2}\mathcal{I}_k \\ \vdots \\ \mathcal{I}_k \end{bmatrix}$$

hence

$$\mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B} = \mathcal{C}z(s) = \frac{1}{d(s)} \begin{bmatrix} N_1 & N_2 & \dots & N_n \end{bmatrix} \begin{bmatrix} s^{n-1}\mathcal{I}_k \\ s^{n-2}\mathcal{I}_k \\ \vdots \\ \mathcal{I}_k \end{bmatrix}$$

$$= \hat{\mathcal{G}}_{sp}(s)$$

LTI SISO system case and multiple realizations

A SISO (single-input single-output) LTI system is characterized by a transfer function that's simply a rational polynomial in the form

$$\hat{\mathcal{G}}(s) = \frac{\beta_1 s^{n-1} + \beta_2 s^{n-2} + \dots + \beta_{n-1}s + \beta_m}{s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \dots + \alpha_{n-1}s + \alpha_n}$$

Recalling the results of equation 1.22, a realization of this kind of system is characterized by the matrices

$$\mathcal{A} = \begin{bmatrix} -\alpha_1 & -\alpha_2 & \dots & -\alpha_{n-1} & -\alpha_n \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \qquad \mathcal{B} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \qquad \mathcal{C} \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_{n-1} & \beta_n \end{bmatrix}$$

Figure 1.4 shows how a system described in a controllable canonical representation (as in this case) can be graphed.
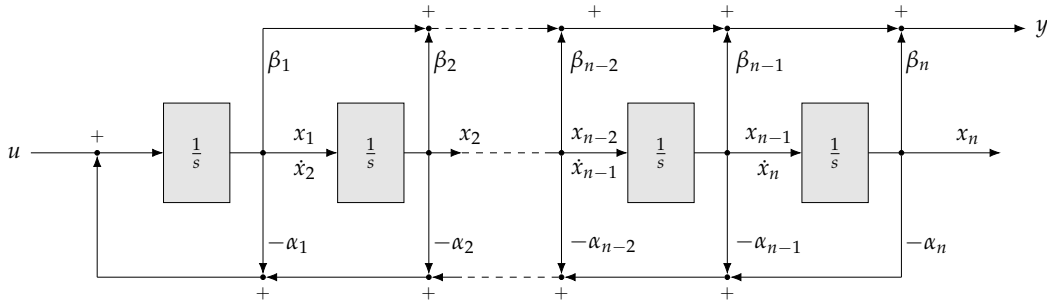

Figure 1.4: graph representation of an LTI SISO system expressed in the controllable canonical form; the $1/s$ blocks in practise are representing integrators of the input.

In general this is not the lonely allowable representation of the transfer function $\hat{\mathcal{G}}(s)$, but there exists infinitely many ones. Considering T as a non-singular matrix, we can define a new vector of the states $\tilde{x} = Tx$ (hence $x = T^{-1}\tilde{x}$) that allows us to determine a new state space representation of the system, in fact we can build

$$\begin{cases} \delta\tilde{x} = T\,\delta x = T\mathcal{A}x + T\mathcal{B}u = T\mathcal{A}T^{-1}\tilde{x} + T\mathcal{B}u = \tilde{\mathcal{A}}\tilde{x} + \tilde{\mathcal{B}}u \\ \tilde{y} = \mathcal{C}x + \mathcal{D}u = \mathcal{C}T^{-1}\tilde{x} + \mathcal{D}u = \tilde{\mathcal{C}}\tilde{x} + \mathcal{D}u \end{cases}$$

We can so state that the systems

$$\left[\begin{array}{c|c} \tilde{\mathcal{A}} & \tilde{\mathcal{B}} \\ \hline \tilde{\mathcal{C}} & \mathcal{D} \end{array} \right] = \left[\begin{array}{c|c} T\mathcal{A}T^{-1} & T\mathcal{B} \\ \hline \mathcal{C}T^{-1} & \mathcal{D} \end{array} \right] \qquad \text{and} \qquad \left[\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array} \right] \qquad (1.24)$$

are **algebraically equivalent** and are indeed representations of the same transfer function $\hat{\mathcal{G}}(s)$. In this case the non-singular matrix T is referred as ***similarity/equivalence transformation*** and allows to describe the system into a different set of states. This operation can be useful because it might allow to express a system in a more suitable way for numerical computations.

An example of algebraically equivalent representation is the **observable canonical form** described as

$$\left[\begin{array}{cccc|c} -\alpha_1 & 1 & 0 & 0 & \beta_1 \\ \vdots & 0 & \ddots & 0 & \vdots \\ -\alpha_{n-1} & 0 & 0 & 1 & \beta_{n-1} \\ -\alpha_n & 0 & \dots & 0 & \beta_n \\ \hline 1 & 0 & \dots & 0 & \end{array}\right] \tag{1.25}$$

The graph representation of such system is shown in figure 1.5.
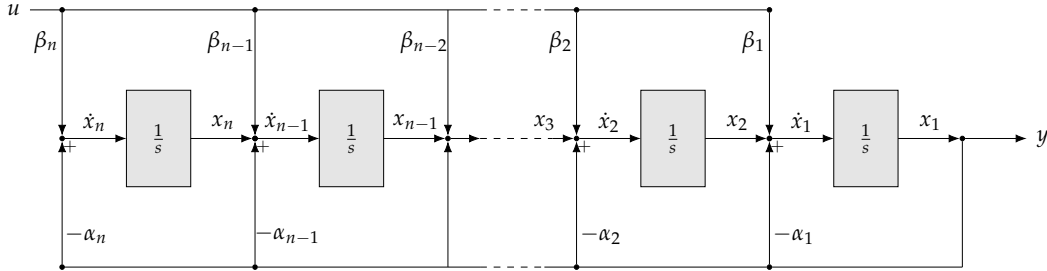

Figure 1.5: *graph representation of an LTI SISO system expressed in the controllable canonical form; the $1/s$ blocks in practise are representing integrators of the input.*

1.4 Solution of linear time-varying systems

Continuous-time linear time-variant systems

Given a continuous-time linear time-variant system (CT-LTV), it's solution must be a function of the time t and must be differentiable (in order to be the solution of the differential equations). In particular the zero input response of the system expressed as

$$\dot{x} = \mathcal{A}(t)x \qquad\qquad x(t_0) = x_0$$

can be computed using the **Peano-Baker series**, a matrix $\Phi(t, t_0) \in \mathbb{R}^{n\times n}$ defined as

$$\Phi(t, t_0) = \mathcal{I} + \int_{t_0}^t \mathcal{A}(s_1)\, ds_1 + \int_{t_0}^t \mathcal{A}(s_1) \int_{t_0}^{s_1} \mathcal{A}(s_2)\, ds_2\, ds_1 +$$
$$+ \int_{t_0}^t \mathcal{A}(s_1) \int_{t_0}^{s_1} \mathcal{A}(s_2) \int_{t_0}^{s_2} \mathcal{A}(s_3)\, ds_3\, ds_2\, ds_1 + \dots \tag{1.26}$$

The matrix Φ is mostly referred as the **state transition matrix** and is characterized by the following property:

i) it has that $\Phi(t, t) = \mathcal{I}$ for any $t \in \mathbb{R}$; this result come just by the definition of the state transition matrix: considering in fact that for the integrals $\int_a^a f(x)\, dx = 0$, then the evaluation for $t = t_0$ cancels all the terms unless the identity matrix;

ii) a second important property is associated to the derivation of Φ; knowing that $\frac{d}{dt} \int_a^t f(\tau)\, d\tau = f(\tau)$, them we can show that

$$\frac{d}{dt}\Phi(t, t_0) = 0 + \mathcal{A}(t) + \mathcal{A}(t) \int_{t_0}^t \mathcal{A}(s_2)\, ds_2 + \mathcal{A}(t) \int_{t_0}^t \mathcal{A}(s_2) \int_{t_0}^{s_2} \mathcal{A}(s_3)\, ds_3\, ds_2 + \dots$$
$$= \mathcal{A}(t)\Phi(t, t_0)$$

With this property states we can compute and prove that the zero input response solution of a continuous-time linear system is determined as

$$x(t) = \Phi(t, t_0)x_0 \qquad\qquad \forall t \tag{1.27}$$

This solution is also proven to be unique (but we won't prove that). Equation 1.27 can be proven considering that $x(t_0)$ gives indeed $\Phi(t_0, t_0)x_0 = \mathcal{I}x_0 = x_0$ using property *i)*. To generally prove that this formulation is also solution to the original state equation $\dot{x}(t) = \mathcal{A}(t)x(t)$ we can just take the derivative of the suggested solution $\Phi(t, t_0)x_0$ respect to time and using property *ii)* we indeed have

$$\dot{x}(t) = \frac{d}{dt}\Big(\Phi(t, t_0)x_0\Big) = \mathcal{A}(t)\phi(t, t_0)x_0 = \mathcal{A}(t)x(t)$$

In order now to further extend the solution in case of non-zero inputs it's mandatory to define other 2 properties for the state transition matrix:

iii) each i-th column of the matrix $\Phi(t, t_0)$ is the unique solution at the problem

$$\dot{x} = \mathcal{A}(t)x \qquad \text{with } x(t_0) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_i$$

where the *one* is placed in the i-th position of the vector x_0, meaning that's the i-th canonical base e_i. This property is very useful for computation because it allow to decompose the study of the solution $x(t)$ by independently solving all the states;

iv) it holds the **semi-group property** stating that

$$\Phi(t, s)\Phi(s, \tau) = \Phi(t, \tau) \qquad\qquad \forall t, s, \tau \in \mathbb{R}$$

We can in fact consider that, given the initial state x_0 at time t_0, the state at a second timespan t_1 is regarded as $x(t_1) = \Phi(t_1, t_0)x_0$. Considering now a second time t_2 we have that

$$x(t_2) = \Phi(t_2, t_1)x(t_1) = \Phi(t_2, t_1)\Phi(t_1, t_0)x_0$$
$$\Phi(t_2, t_0)x_0$$

Note that the semi-group property doesn't set any requirement in the *position* of the times t, s, τ that can in fact be also *back in time*.

v) due to the semi-group property yet stated we have that the state transition matrix $\Phi(t, t_0)$ is always invertible (hence is non-singular) and it's inverse is equal to $\Phi(t, s) = \Phi(s, t)$. We can in fact see that
$$\Phi(t, s)\Phi(s, t) = \Phi(t, t) = \mathcal{I} \qquad\qquad \forall t, s \in \mathbb{R}$$

Addition of the inputs The more general unique solution of a CT-LTV system defined by the state space representation

$$\begin{cases} \dot{x} = \mathcal{A}(t)x + \mathcal{B}(t)u \\ y = \mathcal{C}(t)x + \mathcal{D}(t)u \end{cases} \qquad\qquad x(t_0) = x_0$$

is determined by equation

$$x(t) = \overbrace{\Phi(t, t_0)x_0}^{\text{zero-input res.}} + \overbrace{\int_{t_0}^{t} \Phi(t, \tau)\mathcal{B}(\tau)u(\tau)\,d\tau}^{\text{zero-state response}}$$

$$y(t) = \underbrace{\mathcal{C}(t)\Phi(t, t_0)x_0}_{\text{homogeneous res.}} + \underbrace{\int_{t_0}^{t} \mathcal{C}(\tau)\Phi(t, \tau)\mathcal{B}(\tau)u(\tau)\,d\tau + \mathcal{D}(t)u(t)}_{\text{forced response}}$$

(1.28)

To prove that the provided solution works (we won't prove that's unique, but it happens so) we can simply derive in time the state solution and watch that matches the state equation $\dot{x} = \mathcal{A}(t)x +$

$\mathcal{B}(t)u$. In order to do so we have to consider the following property in the derivation of composed function in integration, in fact

$$\frac{d}{dt}\int_{t_0}^{t} f(t,\tau)\,d\tau = f(t,t) + \int_{t_0}^{t}\frac{d}{dt}f(t,\tau)\,d\tau$$

With this stated we have

$$\begin{aligned}
\dot{x}(t) &= \frac{d}{dt}\left(\Phi(t,t_0)x_0 + \int_{t_0}^{t}\Phi(t,\tau)\mathcal{B}(\tau)u(\tau)\,d\tau\right)\\
&= \mathcal{A}(t)\Phi(t,t_0)x_0 + \Phi(t,t)\mathcal{B}(t)u(t) + \int_{t_0}^{t}\frac{d\Phi}{dt}\mathcal{B}(\tau)u(\tau)\,d\tau\\
&= \mathcal{A}(t)\Phi(t,t_0)x_0\mathcal{B}(t)u(t) + \int_{t_0}^{t}\mathcal{A}(t)\Phi(t,\tau)\mathcal{B}(\tau)u(\tau)\,d\tau\\
&= \mathcal{A}(t)\underbrace{\left(\Phi(t,t_0) + \int_{t_0}^{t}\Phi(t,\tau) + \mathcal{B}(\tau)u(\tau)\,d\tau\right)}_{=x(t)} + \mathcal{B}(t)u(t)
\end{aligned}$$

Discrete-time linear time-variant systems

Given a discrete-time linear time-variant DT-LTV system, the unique solution of the homogeneous term $x^{+} = \mathcal{A}(t)x$ subject to an initial state $x(t_0) = x_0$ is determined by the expression

$$x(t) = \Phi(t,t_0)x_0 \qquad\qquad x_0 \in \mathbb{R}^n, \forall t \geq t_0 \qquad\qquad (1.29)$$

> **Note:** In this case the solution is valid for time t grater or equal to t_0; in the continuous time case, equation 1.27 is valid also for $t < t_0$ due to the semi-group property; for discrete-time system the matrix $\Phi(t,t_0)$ can sometimes be singular, meaning that it cannot be inverted as it will be shown.

In this case $\Phi(t,t_0)$ is the **discrete-time state transition matrix** defined as

$$\Phi(t,t_0) = \begin{cases} \mathcal{I} & t = t_0 \\ \mathcal{A}(t-1)\mathcal{A}(t-2)\ldots\mathcal{A}(t_0+1)\mathcal{A}(t_0) & t > t_0 \end{cases} \qquad (1.30)$$

Such definition can be inductively retrieved: considering in fact that the state at the step t_0+1 can be computed as $x(t_0+1) + \mathcal{A}(t_0)x_0$ but also $x(t_0+2) = \mathcal{A}(t_0+1)x(t_1) = \mathcal{A}(t_0+2)\mathcal{A}(t_0+1)x_0$. Inductively it's proven how the discrete-time state transition matrix was defined. This formulation should be extended also to the *backward times*, for $t < t_0$, in fact considering

$$x(\underbrace{t+1}_{=t_0}) = \mathcal{A}(t)x(t) \qquad\Rightarrow\qquad x(t) = \mathcal{A}^{-1}(t_0-1)x(t_0)$$

This operation is feasible if and only if the matrix $\mathcal{A}(t_0)$ is non-singular that hence can be inverted, but this condition isn't stated a priori.

As for the continuous time case, the general solution of the discrete-time linear time-variant systems of the form

$$\begin{cases} x^{+} = \mathcal{A}(t)x + \mathcal{B}(t)u \qquad\qquad x(t_0) = x_0 \\ y = \mathcal{C}(t)x + \mathcal{D}(t)u \end{cases}$$

is determined by the functions

$$x(t) = \Phi(t,t_0)x_0 + \sum_{\tau=0}^{t-1}\Phi(t,\tau+1)\mathcal{B}(\tau)u(\tau)$$

$$y(t) = \mathcal{C}(t)\Phi(t,t_0)x_0 + \sum_{\tau=0}^{t-1}\mathcal{C}(\tau)\Phi(t,\tau+1)\mathcal{B}(\tau)u(\tau) + \mathcal{D}(t)u(t)$$

$$(1.31)$$

1.5 Solutions to linear time-invariant systems

Discrete-time case

Starting with the case of discrete-time case, if the system is time invariant than it means that the matrices $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ does not depends on time. Considering the definition of Peano-Bakers series for discrete-time systems (equation 1.30), knowing that $\mathcal{A}(t)$ is constant (and is so regarded as $\mathcal{A}$), then it means that the state transition matrix can be regarded as

$$\Phi(t, t_0) = \mathcal{A}^{t-t_0} \qquad\qquad \forall t \geq t_0 \qquad\qquad (1.32)$$

where conventionally $\mathcal{A}^0 = \mathcal{I}$. With that we can rewrite the solutions of the DT-LTI system as

$$x(t) = \mathcal{A}^{t-t_0} x_0 + \sum_{\tau=0}^{t-1} \mathcal{A}^{t-\tau-1} \mathcal{B} u(\tau)$$

$$y(t) = \mathcal{C}\mathcal{A}^{t-t_0} + \sum_{\tau=0}^{t-1} \mathcal{C}\mathcal{A}^{t-\tau-1} \mathcal{B} u(\tau) + \mathcal{D} u(t) \qquad (1.33)$$

We refer to $\mathcal{A}^t$ (with $t \in \mathbb{Z}$) as the **matrix power** and it will be the goal in the definition of the solution is to identify a *clever* way to compute it's value (because numerically it's not feasible to perform matrix multiplication at each iteration); such problem will be fixed later with the definition of the Jordan normal form for matrices.

Continuous-time case

Considering now the continuos-time case for LTI systems, the definition of the Peano-Baker series (equation 1.26, page 11) can be simplified: the matrices $\mathcal{A}$ can in fact be taken outside the integral terms resulting in

$$\Phi(t, t_0) = \mathcal{I} + \mathcal{A}(t - t_0) + \mathcal{A}^2 \frac{(t - t_0)^2}{2} + \mathcal{A}^3 \frac{(t - t_0)^3}{2 \cdot 3} + \dots$$

$$= \sum_{k=0}^{\infty} \mathcal{A}^k \frac{(t - t_0)^k}{k!}$$

This definition resembles the exponential; considering in fact the Taylor series expansion of the exponential is $e^x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k$, we can determine the state transition matrix as the **exponential matrix**

$$\Phi(t, t_0) = e^{\mathcal{A}(t-t_0)} \qquad\qquad (1.34)$$

hence the unique solution of a CT-LTI system is provided by the expression

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

$$y(t) = \mathcal{C} e^{\mathcal{A}(t-t_0)} x_0 + \int_{t_0}^{t} \mathcal{C} e^{\mathcal{A}(t-t_0)} \mathcal{B} u(\tau) \, d\tau + \mathcal{D} u(t) \qquad (1.35)$$

Relevant properties of the exponential matrix are

i) that each i-th column of the matrix $e^{\mathcal{A}t}$ is the unique solution of the problem

$$\dot{x}(t) = \mathcal{A}x(t) \qquad\qquad \text{with } x(0) = e_i$$

ii) it still holds the semi-group property stating that $\Phi(t, \tau)\Phi(\tau, s) = \Phi(t, s)$ $(\forall t, s, \tau)$; by an application point of view it means that for the exponential matrix

$$e^{\mathcal{A}t} e^{\mathcal{A}s} = e^{\mathcal{A}(t+s)}$$

iii) it's always invertible, in fact from the semi-group property we have that $e^{\mathcal{A}t}e^{-\mathcal{A}t} = e^{\mathcal{A}0} = \mathcal{I}$ and so it means that $e^{\mathcal{A}t}$ is always non-singular (hence invertible) and its inverse is defined as

$$\left(e^{\mathcal{A}t}\right)^{-1} = e^{-\mathcal{A}t}$$

iv) for any matrix $\mathcal{A} \in \mathbb{R}^{n \times n}$ it holds the following *"commutative"* property:

$$\mathcal{A}e^{\mathcal{A}t} = e^{\mathcal{A}t}\mathcal{A} \qquad \forall t$$

v) there exists n functions $\alpha_i(t)$ such that

$$e^{\mathcal{A}t} = \sum_{k=0}^{n-1} \alpha_k(t)\mathcal{A}^k$$

Cayley-Hamilton theorem   The **Cayley-Hamilton theorem** states that *for each matrix $\mathcal{A} \in \mathbb{R}^{n \times n}$ with characteristic polynomial*

$$p_{\mathcal{A}}(s) = \det(s\mathcal{I} - \mathcal{A}) = s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n$$

then $\mathcal{A}$ is a matrix solution of this polynomial, meaning that

$$\mathcal{A}^n + a_1\mathcal{A}^{n-1} + \ldots a_{n-1}\mathcal{A} + a_n\mathcal{I} = 0$$

In this case 0 is a $n \times n$ matrix of zeros. By reversing this quantity we can indeed see that

$$\mathcal{A}^n = -\left(a_1\mathcal{A}^{n-1} + \ldots a_{n-1}\mathcal{A} + a_n\mathcal{I}\right)$$

and so by multiplying both sides of the equation by $\mathcal{A}$ allows to compute the exponential $\mathcal{A}^{n+1}$ as

$$\begin{aligned}
\mathcal{A}^{n+1} &= -\mathcal{A}^n a_1 - a_2\mathcal{A}^{n-1} - \cdots - a_{n-1}\mathcal{A}^2 - a_n\mathcal{A} \\
&= a_1\left(a_1\mathcal{A}^{n-1} + \cdots + za_{n-1}\mathcal{A} + a_n\mathcal{I}\right) - a_2\mathcal{A}^{n-1} - \cdots - a_{n-1}\mathcal{A}^2 - a_n\mathcal{A} \\
&= (a_1^2 - a_2)\mathcal{A}^{n-1} + (a_1a_2 - a_3)\mathcal{A}^{n-2} + \cdots + (a_1a_{n-1} - a_n)\mathcal{A} + a_1a_n\mathcal{I} - \cdots - a_{n-1}\mathcal{A}^2
\end{aligned}$$

In general this proposition can be extended and any exponential of a matrix $\mathcal{A} \in \mathbb{R}^{n \times n}$ can be regarded as a linear combination of the first n exponentials of matrix $\mathcal{A}$:

$$\mathcal{A}^h = \sum_{k=0}^{n-1} c_{h,k}\mathcal{A}^k \tag{1.36}$$

1.5.1 Jordan normal form

As was show in equation 1.24 (page 10), using a transformation matrix T a linear system characterized by a state space representation

$$\begin{cases} \delta x = \mathcal{A}x + \mathcal{B}u \\ y = \mathcal{C}x + \mathcal{D}u \end{cases}$$

can be transformed in a algebraically equivalent representation that could have resulted in a easier computation model. From an ideal point of view the *best* solution we can get after transformation is a **diagonal state matrix** in the form

$$T\mathcal{A}T^{-1} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

where λ_i are the eigenvalues of the *original* matrix $\mathcal{A}$ (values that are always shared in each algebraic equivalent representation of the same system).

A problem of this kind of formulation is that in practise not all matrices $\mathcal{A}$ are diagonalizable, and in particular this happens when the geometric multiplicity doesn't match the algebraic multiplicity of the related eigenvalues.

A solution of this problem is provided by Jordan with a theorem stating that *for each matrix* $\mathcal{A} \in \mathbb{C}^{n \times n}$ *with eigenvalues* $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ *there exists an non-singular (hence invertible) matrix* $T \in \mathbb{C}^{n \times n}$ *such that*

$$\mathcal{J} = T\mathcal{A}T^{-1} = \begin{bmatrix} J_1 & 0 & \ldots & 0 \\ 0 & J_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & J_l \end{bmatrix} \in \mathbb{C}^{n \times n} \tag{1.37}$$

where J_i are matrices in the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ 0 & 0 & \lambda_i & \ldots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_i \end{bmatrix} \in \mathbb{C}^{n_i \times n_i} \tag{1.38}$$

The matrix $\mathcal{J}$ is known as the **Jordan normal form** of $\mathcal{A}$ and is proven to be unique (upon rearrangement of the sub-matrices J_i). Note that in general sub-matrices J_i, J_j (with $i \neq j$) might be constructed with the same eigenvalue λ_k but what might change are their dimension ($n_i \neq n_j$). The number of *blocks* in the Jordan normal form are $l \leq n$ and are associated to each linearly independent eigenvectors of $\mathcal{A}$ (is proven that at least one eigenvector exists for each matrix $\mathcal{A}$ and there can be maximum n linearly independent ones).

A matrix is so called **semi-simple** (or **diagonalizable**) if its Jordan normal form is diagonal. In particular given a matrix $\mathcal{A} \in \mathbb{C}^{n \times n}$ the 3 following statements are equivalents:

- $\mathcal{A}$ is semi-simple;

- $\mathcal{A}$ has n linearly independent eigenvectors;

- $p(\mathcal{A}) \neq 0$ for all non-zero polynomials $p(s)$ having a degree less then n.

Matrix power

The power matrix value can be *easily* determined considering the Jordan normal form of the matrix on which we want to compute the power $\mathcal{A}^t$; in particular we can see that

$$\mathcal{A}^t = \underbrace{\overbrace{T^{-1}\mathcal{J}T}^{=\mathcal{A}} T^{-1}\mathcal{J}T \ldots T^{-1}\mathcal{J}T}_{t \text{ times}}$$

Observing that *in the middle* the terms TT^{-1} cancels out (they evaluate to the identity matrix $\mathcal{I}$) what we obtain is that

$$\mathcal{A}^t = T^{-1}\mathcal{J}^t T \tag{1.39}$$

where the power of a matrix in Jordan forms evaluates to

$$\mathcal{J}^t = \begin{bmatrix} J_1^t & & & \\ & J_2^t & & \\ & & \ddots & \\ & & & J_l^t \end{bmatrix} \tag{1.40}$$

where

$$J_i^t = \begin{bmatrix} \lambda_i^t & t\lambda_i^{t-1} & \frac{t!\lambda_i^{t-2}}{(t-2)!2!} & \frac{t!\lambda_i^{t-3}}{(t-3)!3!} & \cdots & \frac{t!\lambda_i^{t-n_i+1}}{(t-n_i+1)!(n_i-1)!} \\ 0 & \lambda_i^t & t\lambda_i^{t-1} & \frac{t!\lambda_i^{t-2}}{(t-2)!2!} & \cdots & \frac{t!\lambda_i^{t-n_i+2}}{(t-n_i+2)!(n_i-2)!} \\ \vdots & \ddots & \lambda_i^t & t\lambda_i^{t-1} & \cdots & \frac{t!\lambda_i^{t-n_i+3}}{(t-n_i+3)!(n_i-3)!} \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \lambda_i^t & t\lambda_i^{t-1} \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda_i^t \end{bmatrix} \tag{1.41}$$

Region of convergence of the system   Considering that the homogeneous response of the discrete time system $x(t+1) = \mathcal{A}x(t)$ (with $x(0) = x_0$) is determines as

$$x(t) = \mathcal{A}^t x_0$$

In order to have a convergent solution we want to ensure that $\lim_{t\to\infty} \mathcal{A}^t$ tends to zero (or at least it doesn't diverge), in the sense that all it's entry are zero (or non-diverging). Considering that the matrix power $\mathcal{A}^t$ can be regarded as $T^{-1}\mathcal{J}T$, the limit is intrinsecally transmitted to the Jordan normal form matrix $\mathcal{J}$ for which we can state

$$\lim_{t\to\infty} \mathcal{J}^t = 0 \qquad \Leftrightarrow \qquad \lim_{t\to\infty} J_i^t = 0 \quad \forall i$$

Considering the formal definition of the sub-matrix power J_i^t shown in equation 1.40, we observe that each entry of the matrix is made by the eigenvalue λ_i to the power of t and a polynomial of t. Considering that the exponential asymptotically grows/decreases faster then polynomial we observe that whenever $|\lambda_i| \leq 1$ then all entries converges to 0. We can so state that if all eigenvalues λ_i of the state matrix $\mathcal{A}$ have a magnitude less then 1, then the system converges to a null solution.

In contrary if we consider $|\lambda_i| > 1$ the exponential λ_i^t explodes to infinity for $t \to \infty$, hence the systems behaviour diverges. All that's left to analyse is the case where $\mathcal{J}^t$ (for $\to \infty$) doesn't diverge and this happens when $|\lambda_i| = 1$ (where $1^t \xrightarrow{t\to\infty} 1$), but in order to ensure that the matrix exponential doesn't diverge we have to impose that $n_i = 1$ (in this case we don't have entries with diverging polynomials). In general we can so state that

$$\lim_{t\to\infty} \mathcal{A}^t \neq \infty \qquad \Leftrightarrow \qquad \lim_{t\to\infty} J_i^t \neq \infty \; \forall i \qquad \Leftrightarrow \qquad |\lambda_i| \leq 1 \text{ and if } |\lambda_i| = 1 \Rightarrow n_i = 1$$

Fibonacci sequence   Recalling the Fibonacci state representation reported in equation 1.3 (page 1), we have that the problem was described as

$$x^+ = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} x \qquad\qquad y = \begin{bmatrix} 1 & 0 \end{bmatrix} x \qquad\qquad x(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

In this case the polynomial characteristic of the state matrix $\mathcal{A} \in \mathbb{R}^{2\times 2}$ evaluates as $p_\mathcal{A}(s) = s^2 - s - 1$ whose roots (hence eigenvalues of $\mathcal{A}$) are $s_{1,2} = \frac{1\pm\sqrt{5}}{2}$. Having two distinct eigenvalues this means that the Jordan normal form of $\mathcal{A}$ is the diagonal matrix

$$\mathcal{J} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

In order now to determine the transformation matrix $T \in \mathbb{R}^{2\times 2}$ we need to solve the linear system $T\mathcal{A} = \mathcal{J}T$, hence

$$\begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix} \qquad \Rightarrow \qquad \begin{cases} t_1 + t_2 & = \frac{1+\sqrt{5}}{2} t_1 \\ t_1 & = \frac{1+\sqrt{5}}{2} t_2 \\ t_3 + t_4 & = \frac{1-\sqrt{5}}{2} t_3 \\ t_3 & = \frac{1-\sqrt{5}}{2} t_4 \end{cases}$$

hence

$$t_1 = \frac{1+\sqrt{5}}{2}t_2 \qquad\qquad t_3 = \frac{1-\sqrt{5}}{2}t_4 \qquad\qquad \forall t_2, t_4$$

In order not to have a singular matrix we have to chose value $t_2, t_4 \neq 0$; picking for simplicity $t_2 = t_4 = 2$ the transformation matrix we obtain is

$$T = \begin{bmatrix} 1+\sqrt{5} & 2 \\ 1-\sqrt{5} & 2 \end{bmatrix} \qquad \Rightarrow \qquad T^{-1} = \frac{1}{4\sqrt{5}} \begin{bmatrix} 2 & -2 \\ -(1-\sqrt{5}) & 1+\sqrt{5} \end{bmatrix}$$

Using equation 1.39 we can compute the matrix power $\mathcal{A}^t$ as $T^{-1}\mathcal{J}T$; to ease the computation we can see that $1+\sqrt{5} = 2\lambda_1$ and $1-\sqrt{5} = 2\lambda_2$ we can write the equation as

$$\mathcal{A}^t = \frac{1}{4\sqrt{5}} \begin{bmatrix} 2 & -2 \\ -2\lambda_2 & 2\lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{bmatrix} \begin{bmatrix} 2\lambda_1 & 2 \\ 2\lambda_2 & 2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -1 \\ -\lambda_2 & \lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{bmatrix} \begin{bmatrix} \lambda_1 & 1 \\ \lambda_2 & 1 \end{bmatrix}$$

$$= \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -1 \\ -\lambda_2 & \lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1^{t+1} & \lambda_1^t \\ \lambda_2^{t+1} & \lambda_2^t \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1^{t+1} - \lambda_2^{t+1} & \lambda_1^t - \lambda_2^t \\ -\lambda_1^{t+1}\lambda_2 + \lambda_1\lambda_2^{1+t} & -\lambda_1^t\lambda_2 + \lambda_1\lambda_2^t \end{bmatrix}$$

Computing the homogeneous solution of the system as $y(t) = \mathcal{C}\mathcal{A}^t x_0$, given the particular formulation of the matrices $\mathcal{C}$ and x_0 means that the output $y(t)$ is determined by the second entry in the first column, hence

$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1^{t+1} - \lambda_2^{t+1} & \lambda_1^t - \lambda_2^t \\ -\lambda_1^{t+1}\lambda_2 + \lambda_1\lambda_2^{1+t} & -\lambda_1^t\lambda_2 + \lambda_1\lambda_2^t \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \left(\lambda_1^t - \lambda_2^t \right)$$

$$= \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^t - \left(\frac{1-\sqrt{5}}{2} \right)^t \right]$$

Matrix exponential

If we want to compute the exponential $e^{\mathcal{A}t}$ of the matrix $\mathcal{A}$ with Jordan normal form $\mathcal{J}$ we can use the fact that the exponential of such Jordan form is

$$e^{\mathcal{J}t} = \begin{bmatrix} e^{J_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{J_l t} \end{bmatrix} \tag{1.42}$$

where $e^{J_i t} \in \mathbb{R}^{n_i \times n_i}$ is defined as

$$e^{J_i t} = e^{\lambda_i t} \begin{bmatrix} 1 & t & \cdots & \frac{t^{n_i-1}}{(n_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & t \\ & & & 1 \end{bmatrix} \tag{1.43}$$

Considering now the Cayley-Hamilton theorem (equation 1.36, page 15) we have that the matrix exponential can be regarded as a linear combination of the first n matrix powers as

$$e^{\mathcal{A}t} = \sum_{k=0}^{n-1} \alpha_k(t)\mathcal{A}^k$$

Knowing so that $\mathcal{A} = T^{-1}\mathcal{J}T$ we can rewrite

$$e^{\mathcal{A}t} = \sum_{k=0}^{n-1} \alpha_k(t)T^{-1}\mathcal{J}^k T = T^{-1}\left(\sum_{k=0}^{n-1} \mathcal{J}^k \right) T \tag{1.44}$$

$$= T^{-1}e^{\mathcal{J}t}T$$

on which we can compute the formal definition of the Jordan form exponential as shown in equation 1.42.

Region of convergence Considering so that the exponential $e^{\mathcal{J}t}$ is the unique solution of the problem $\dot{x}(t) = \mathcal{J}x(t)$ with $x(0) = \mathcal{I}$, we can so analyse the asymptotic behaviour of the system (as was done for the discrete-time ones). In general the eigenvalues an be complex ($\lambda_i = \alpha_i + j\beta_i \in \mathbb{C}$) and we can show that the sub-matrix exponentials $e^{J_i t}$ converges for $t \to \infty$ if and only if the real part α_i of the eigenvalues is strictly negative; at contrary if $\alpha_i > 0$ what we obtain is that the term $e^{\lambda_i t}$ premultiplying the whole matrix diverges and hence $\mathcal{J}$.

In the case where $\alpha_i = 0$ we have the case of discrete-time systems: if the dimension n_i is unitary then e^{J_i} is a constant (non-infinite) 1-dimensional matrix, while if $n_i > 1$ the upper triangular terms polynomially explodes to ∞:

$$\lim_{t \to \infty} e^{At} \neq \infty \quad \Leftrightarrow \quad \lim_{t \to \infty} e^{J_i t} \neq \infty \, \forall i \quad \Leftrightarrow \quad \mathbb{Re}\,\{\lambda_i\} \leq 0 \text{ and if } \mathbb{Re}\,\{\lambda_i\} = 0 \Rightarrow n_i = 1$$

Example Given the continuous time LTI system characterized by the state matrix

$$\mathcal{A} = \begin{bmatrix} 5 & 3 \\ -6 & -4 \end{bmatrix}$$

in order to compute it's exponential e^{At} we have to firstly determine it's Jordan normal form $\mathcal{J}$. This can be achieved by firstly determining the eigenvalues of the matrix; knowing that the characteristic polynomial is $p(\lambda) = (5 - \lambda)(-4 - \lambda) + 6 \cdot 3 = \lambda^2 - \lambda - 2$ we can compute it's roots and hence determine $\mathcal{J}$:

$$\lambda_{1,2} = \frac{1 \pm \sqrt{9}}{2} \qquad \Rightarrow \qquad \mathcal{J} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$$

With this result we have to compute the transformation matrix T by imposing that $\mathcal{J}T = TA$ and solving the related linear system in the unknowns entries of T:

$$\begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix} = \begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix} \begin{bmatrix} 5 & 3 \\ -6 & -4 \end{bmatrix}$$

$$\begin{cases} 2t_1 & = 5t_1 - 6t_2 \\ 2t_2 & = 3t_1 - 4t_2 \\ -t_3 & = 5t_3 - 6t_4 \\ -t_4 & = 3t_3 - 4t_4 \end{cases} \qquad \Rightarrow \qquad \begin{cases} 2t_2 = t_1 \\ t_3 = t_4 \end{cases}$$

Choosing as free parameters $t_1 = t_3 = 1$ for simplicity we obtain the transformation matrix

$$T = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \qquad \Rightarrow \qquad T^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

We can finally compute the exponential matrix e^{At} as $T^{-1}e^{\mathcal{J}t}T$ (equation 1.44):

$$e^{At} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} e^{2t} & 0 \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2e^{2t} & e^{2t} \\ e^{-t} & e^{-t} \end{bmatrix}$$

$$= \begin{bmatrix} 2e^{2t} - e^{-t} & e^{2t} - e^{-t} \\ -2e^{2t} + 2e^{-t} & -e^{2t} + 2e^{-to} \end{bmatrix}$$

1.6   Internal / Lyapunov stability

Norms for vector In order to start the discussion regarding the stability of linear systems, we have to recall the norms for vector and matrices. Given a vector $x \in \mathbb{R}^n$, we define it's **norm** $x \mapsto |x|$ an operation that satisfy the following properties:

$$\begin{aligned} i) & \quad |x| \geq 0 \, \forall x \in \mathbb{R}^n \quad \text{and} \quad |x| = 0 \Leftrightarrow x = 0 \\ ii) & \quad |x + y| \leq |x| + |y| \qquad \forall x, y \in \mathbb{R}^n \\ iii) & \quad |ax| = |a|\,|x| \qquad \forall x \in \mathbb{R}^n, a \in \mathbb{R} \end{aligned}$$

Matching this definition we can define different kinds of norms:

- the *one norm* $|\cdot|_1$ defined as

$$|x|_1 = \sum_{i=1}^{n} |x_i| \qquad (1.45)$$

- the *two norm* $|\cdot|_2$ (euclidean norm) defined as

$$|x|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \qquad (1.46)$$

- the *infinity norm* $|\cdot|_\infty$ defined as

$$|x|_\infty = \max_{i=1,\dots,n} |x_i| \qquad (1.47)$$

- this previous concepts can be generalized by the concept of *p norm* $|\cdot|_p$ defined as

$$|x|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \qquad (1.48)$$

All this norms are **equivalents**: for any norm $p_1, p_2 \in [1, \infty)$ there exists two constants $c_1, c_2 \in \mathbb{R}$ such that

$$c_1 |x|_{p_1} \leq |x|_{p_2} \leq c_2 |x|_{p_1} \qquad \forall x \in \mathbb{R}^n$$

Norms for matrix   The definition of norm can be extended also in case of matrices; in this case the 3 property of the norm must be respected but given a matrix $A \in \mathbb{R}^{m \times n}$ with m rows and n columns we can compute its norm $\|A\|$ in different ways:

- *one norm* $\|\cdot\|_1$ defined as

$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |a_{ij}| \qquad (1.49)$$

  If we consider the matrix as the *concatenation* of n column vectors $v_i \in \mathbb{R}^m$, then this norm can be regarded as the maximum one norm $|v_i|_1$ among such vectors.

- *infinity norm* $\|\cdot\|_\infty$ defined as

$$\|A\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |a_{ij}| \qquad (1.50)$$

- *two norm* $\|\cdot\|_2$ defined as

$$\|A\|_2 = \sigma_{\max}\{A\} = \sqrt{\lambda_{max}\{A^T A\}} \qquad (1.51)$$

  We denote with $\sigma\{A\}$ the **singular values** of the matrix A and are computed as the square root of the eigenvalues of the square matrix $A^T A \in \mathbb{R}^{n \times n}$; in particular the two norm corresponds to the maximum singular value of the matrix A. Considering that a vector can be regarded as a matrix $\mathbb{R}^{m \times 1}$ this definitions coincide with the euclidean norm for vectors (equation 1.46).

- *Frobenius norm* $\|\cdot\|_F$ defined as

$$\|A\|_F = \sqrt{\sum_{i=0}^{m} \sum_{j=0}^{n} a_{ij}^2} = \sqrt{\sum_{i=0}^{n} \sigma_i^2\{A\}} \qquad (1.52)$$

  Also in this case if we consider a vector as a matrix $\mathbb{R}^{m \times 1}$ the computation of the Frobenius norm coincides with the euclidean/two norm (equation 1.46) but in general for arbitrarily large matrices $\|A\|_2 \leq \|A\|_F$.

> **Example 1.3: two vs Frobenius norm**
>
> Considering the matrix
>
> $$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$
>
> in order to determine both two and Frobenius norm we have to compute the singular values of the matrix, and to do so we have to compute the matrix $A^T A$ that results in
>
> $$A^T A = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$
>
> Being this matrix diagonal, the two eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = 1$ that, squared, results in the singular values $\sigma_1 = 2$, $\sigma_2 = 1$. We can so compute the two norms as
>
> $$\|A\|_2 = \sigma_{max} = \sigma_1 = 2 \qquad \neq \qquad \|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{4+1} = \sqrt{5}$$
>
> We just showed with an example that in general Frobenius and two norms doesn't usually coincides for matrices of arbitrary dimensions.

Also matrix norms are **equivalent**, meaning that for any norm $p_1, p_2 \in \{1, 2, F, \infty\}$ there exists two constants $c_1, c_2 \in \mathbb{R}$ such that

$$c_1 \|A\|_{p_1} \leq \|A\|_{p_2} \leq c_2 \|A\|_{p_1} \qquad \forall A \in \mathbb{V}$$

Matrix norms are also **sub-multiplicative**, meaning that

$$\|AB\|_p \leq \|A\|_p \|B\|_p \qquad \forall A, B \in \mathbb{V}, \ \forall p \in \{1, 2, F, \infty\}$$

Induced norm Given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $x \in \mathbb{R}^n$, the product Ax results in a $\mathbb{R}^n$ vector on which we can compute the norm. Considering so the sub-multiplicative property we have that

$$|Ax|_p \leq \|A\|_p |x|_p \qquad \Rightarrow \qquad \|A\|_p \geq \frac{|Ax|_p}{|x|_p}$$

For $p \in \{1, 2, \infty\}$ we have that the associated norms are **subordinate**, meaning that we can compute the norm $\|A\|_p$ as

$$\|A\|_p = \max_{x \neq 0} \frac{|Ax|_p}{|x|_p} \tag{1.53}$$

Note that the Frobenius norm isn't subordinate, hence this relation can't be applied.

Lyapunov stability for linear system

Given a linear system (both discrete/continuous-time, time-varying/invariant) described as

$$\begin{cases} \delta x = \mathcal{A}(t)x + \mathcal{B}(t)u & x(t_0) = x_0 \\ y = \mathcal{C}(t)x + \mathcal{D}(t)u \end{cases}$$

we say that

- is **Lyapunov** (or *marginally*) **stable** if for each initial state x_0 there exists a constant $M \in \mathbb{R}$ (that depends of course on x_0) such that the zero input response satisfies

$$|x(t)| \leq M \qquad \forall t \geq t_0, t_0 \geq 0$$

- is (Lyapunov) **asymptotically stable** if it's Lyapunov stable and all solutions satisfy

$$\lim_{t \to \infty} |x(t)| = 0 \qquad \forall x_0$$

- is (Lyapunov) **exponentially stable** if it's asymptotically stable and exists constants $c > 0, \lambda > 0, \mu \in [0,1)$ such that

$$|x(t)| \leq ce^{-\lambda(t-t_0)}|x_0| \qquad \forall x_0, t_0 \qquad \text{for continuous-time systems}$$

$$|x(t)| \leq c\mu^{t-t_0}|x_0| \qquad \forall x_0, t_0 \qquad \text{for discrete-time systems}$$

- a system is **unstable** if it's not Lyapunov stable.

We can see that this conditions are always more stringent, hence we can say that

$$\text{exponential stability} \quad \Rightarrow \quad \text{asymptotic stability} \quad \Rightarrow \quad \text{Lyapunov stability}$$

From what concerns Lyapunov, stability is analysed regarding the homogeneous (zero input) response and, as was previously described, the solution can be described using the state transition matrix Φ, hence we can consider:

$$|x(t)| = |\Phi(t,t_0)x_0| \leq \|\Phi(t,t_0)\| \, |x_0| \qquad \forall x_0$$

From what concerns the Lyapunov stability this means that the norm of the state transition matrix must be bounded, mathematically $\|\Phi(t,t_0)\| \leq M_\Phi$; the same idea holds for both the asymptotic ($\lim_{t\to\infty} \|\Phi(t,t_0)\| = 0$) and exponential stability, where $\|\Phi(t,t_0)\| \leq ce^{-\lambda(t-t_0)}$ (for continuous-time systems, but the same can be transliterated for discrete-time ones).

We indeed see that **Lyapunov stability** only depends on the state transition matrix $\Phi(t,t_0)$, and not the initial condition x_0, hence is an **intrinsic property of the system** independently on the inputs and can be ensured by the analysis of the behaviour of the state transition matrix.

Linear time invariant systems

Considering continuous-time time-invariant linear systems, we proved that the state transition matrix $\Phi(t,t_0)$ can be computed as the matrix exponential $e^{\mathcal{A}(t-t_0)}$ and, as was described, we showed that such matrix converged if $\mathrm{Re}\,\{\lambda_i\} < 0 \forall i$. In particular for continuous-time LTI systems we have

$$\begin{aligned} \text{asymptotic stability} \quad &\Leftrightarrow \quad \mathrm{Re}\,\{\lambda_i\{\mathcal{A}\}\} < 0 \;\forall i \\ \text{exponential stability} \quad &\Leftrightarrow \quad \mathrm{Re}\,\{\lambda_i\{\mathcal{A}\}\} < 0 \;\forall i \\ \text{Lyapunov stability} \quad &\Leftrightarrow \quad \mathrm{Re}\,\{\lambda_i\{\mathcal{A}\}\} \leq 0 \;\forall i \text{ and if } \mathrm{Re}\,\{\lambda_j\} = 0 \;\Rightarrow\; n_j = 0 \end{aligned}$$

We see so that for linear time-invariant systems asymptotic and exponential stability are *coincident* concepts, meaning that they both happen at the same time.

Considering instead discrete-time LTI systems the state transition matrix $\Phi(t,t_0)$ is computed as the power $\mathcal{A}^{t-t_0}$ and, as was previously discussed, we can conclude that

$$\begin{aligned} \text{asymptotic stability} \quad &\Leftrightarrow \quad |\lambda_i\{\mathcal{A}\}| < 1 \;\forall i \\ \text{exponential stability} \quad &\Leftrightarrow \quad |\lambda_i\{\mathcal{A}\}| < 1 \;\forall i \\ \text{Lyapunov stability} \quad &\Leftrightarrow \quad |\lambda_i\{\mathcal{A}\}| \leq 1 \;\forall i \text{ and if } |\lambda_j| = 1 \;\Rightarrow\; n_j = 0 \end{aligned}$$

1.6.1 Conditions for exponential stability

Positive and negative definite matrices In order to make assertions regarding the exponential stability of system, a recall of definition from linear algebra is required. Given a square matrix $Q \in \mathbb{R}^{n \times n}$ **symmetric** (so such that $Q = Q^T$), we say that it is

- **positive definite**, noted as $Q > 0$, if $x^T Q x > 0 \;\forall x \neq 0$;

- **negative definite** ($Q < 0$) if $x^T Q x < 0 \;\forall x \neq 0$;

- **semi-positive definite** ($Q \geq 0$) if $x^T Q x \geq 0 \;\forall x \in \mathbb{R}^n$;

- **semi-negative definite** ($Q \leq 0$) if $x^T Q x \leq 0 \;\forall x \in \mathbb{R}^n$.

> **Note:** This definition is set for symmetric matrices Q, however in the more general case a matrix $Q \in \mathbb{R}^{n \times n}$ is definite regarding only is symmetric part. Recalling the property for matrices calculations $(ABC)^T = C^T B^T A^T$ and observing that $x^T Q x$ is a scalar hence it's symmetric, we can see that
>
> $$x^T Q x = \frac{1}{2}\left(x^T Q x + (x^T Q x)^T\right) = \frac{1}{2}\left(x^T Q x + x^T Q^t x\right) = x^T \frac{Q + Q^T}{2} x = x^T Q_{symm} x$$
>
> where $Q_{symm} = \frac{Q + Q^T}{2}$ is the symmetrical part of Q; the matrix Q can in fact be consider as the sum $Q_{symm} + Q_{skew}$ where $Q_{skew} = \frac{Q - Q^T}{2}$ is the skew-symmetric part of Q that results always in a zero-definite part.

As a convention for any symmetric matrices Q_1, Q_2, with the notation $Q_1 > Q_2$ we imply that $Q_1 - Q_2 > 0$ or $Q_2 - Q_1 < 0$. A results from linear algebra for any symmetric matrix $Q = Q^T \in \mathbb{R}^{n \times n}$ the following statements are equivalents:

- Q is positive definite;

- all eigenvalues of Q are positive and real evaluated;

- the determinant of all the principle minors (the upper-left sub-matrices) are all positive definite. This is referred as the Sylvester criterion;

- there exists a non-singular matrix $H \in \mathbb{R}^{n \times n}$ such that $Q = H^T H$ as for the Cholesky upper-triangular factorization. This also implied that exists the root $K = \sqrt{Q}$, so $\exists K$ such that $Q = KK$.

As a consequence for **quadratic functions** (in the form $x^T Q x$) it holds

$$\lambda_{min}\{Q\}|x|_2^2 \leq x^T Q x \leq \lambda_{max}\{Q\}|x|_2^2 \qquad \forall x \in \mathbb{R}^n \tag{1.54}$$

Theorem For a continuous-time LTI system in the form $\dot{x} = \mathcal{A}x$ (hence only the zero input part) the following statements are equivalent:

(i) it is asymptotically stable;

(ii) it is exponentially stable;

(iii) the matrix $\mathcal{A}$ is *Hurwitz*, meaning that all the eigenvalues have negative real part ($\mathbb{Re}\{\lambda_i(A)\} < 0 \ \forall i$);

(iv) for each symmetric positive definite matrix $Q = Q^T > 0 \in \mathbb{R}^{n \times n}$ there exists a symmetric matrix $P \in \mathbb{R}^{n \times n}$ such that $P^T = P > 0$ and that satisfy the **Lyapunov equality**

$$\mathcal{A}^T P + P\mathcal{A} = -Q \tag{1.55}$$

Moreover (as an extra fact), such matrix P is also unique;

(v) there exists a symmetric matrix $P = P^T > 0 \in \mathbb{R}^{n \times n}$ that satisfies the **Lyapunov inequality**

$$\mathcal{A}^T P + P\mathcal{A} < 0$$

From a computational point of view the last statement *(v)*, associated to the so called *linear matrix inequality*, somehow certifies the exponential stability. Lyapunov equality *(iv)* is a stronger definition but more complex to compute and the inequality is usually preferred to assert the internal stability of the system.

In general the implications *(i)* $\Leftrightarrow$ *(ii)* $\Leftrightarrow$ *(iii)* were already previously discussed; relating *(iv)* $\Rightarrow$ *(v)* is somewhat trivial (while the reverse operation is mathematically impossible and requires to pass through *(i)* of *(ii)*), but in the next paragraph an analytical proof connecting *(ii)* $\Rightarrow$ *(iv)* and *(v)* $\Rightarrow$ *(ii)* will be presented.

Proof 1: exponential stability implies the Lyapunov equality We can show that one solution (and we will show that's also unique) of the system is characterized by the matrix P defined as

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

This matrix solves the Lyapunov equality and so we have to show that's finite; in order to to so we can compute it's norm obtaining

$$\|P\| = \left\| \int_0^\infty e^{A^T t} Q e^{At} \, dt \right\| \leq \int_0^\infty \left\| e^{A^T t} Q e^{At} \right\| \, dt$$

Using the sub-multiplicative property of the matrix norm we moreover obtain

$$\|P\| = \int_0^\infty \left\| e^{A^T t} \right\| \|Q\| \left\| e^{At} \right\| \, dt = \|Q\| \int_0^\infty \left\| e^{At} \right\|^2 \, dt$$

As a property for any square matrix $A \in \mathbb{R}^{n \times n}$, denoting with $\mu\{A\} = \max_i \operatorname{Re}\{\lambda_i(A)\}$ it's **spectral abscissa**, we have that $\forall \lambda > \mu\{A\}$ there exists a constant $k \in \mathbb{R}$ such that $\|e^{At}\| \leq k e^{\lambda t} \; \forall t > 0$. We can so rewrite the norm inequality as

$$\|P\| \leq \|Q\| \int_0^\infty k e^{2\lambda t} \, dt$$

Considering that we started with the assumption that the system is exponentially stable, then it means that all the eigenvalues λ_i have a real negative part (hence λ in the reported integral is negative) meaning that the integral (and so the norm) converges to a finite number.

We can now show that the so computed matrix P is the solution of the Lyapunov equality: by plugging such definition in equation 1.55 what we obtain is

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

$$= \int_0^\infty \left(\mathcal{A}^T e^{A^T t} Q e^{At} + e^{A^T t} Q e^{At} \mathcal{A} \right) \, dt$$

Considering that $e^{A^T t}$ can be regarded as $\left(e^{At}\right)^T$, the whole term in the parenthesis can be interpreted as the derivative $\frac{d}{dt}\left(e^{A^T t} Q e^{At}\right)$ then the evaluation of the integrals reduces to

$$\mathcal{A}^T P + P \mathcal{A} = \left[e^{A^T t} Q e^{At} \right] \Big|_{t=0}^{\infty} = 0 - \mathcal{I} Q \mathcal{I}$$

$$= -Q$$

That concluded the proof that such matrix P is indeed the solution of the Lyapunov equality but we also have to show that P is positive definite; this can be considered computing the product

$$z^T P z = \int_0^\infty z^T e^{A^T t} Q e^{At} z \, dt \qquad\qquad \forall z \in \mathbb{R}^n / \{0\}$$

Observing that the product $e^{At} z$ evaluates always in a non-zero vector $w(t)$ we can interpret the previous expression as

$$z^T P z = \int_0^\infty w^T(t) Q w(t) \, dt \geq \lambda_{min}\{Q\} \int_0^1 |w(t)|^2 \, dt > 0$$

where the inequality has been obtained considering that $w^T Q w(t)$ is a quadratic function and hence we can apply the *sandwich inequality* (equation 1.54).

Lastly we can show that the solution P is unique; if we assume that P_1, P_2 are two solution of the Lyapunov equation this means that they both satisfy

$$\begin{cases} \mathcal{A}^T P_1 + P_1 \mathcal{A} = -Q \\ \mathcal{A}^T P_2 + P_2 \mathcal{A} = -Q \end{cases}$$

Subtracting the second equation to the first evaluates to $\mathcal{A}^T(P_1 - P_2) + (P_1 - P_2)\mathcal{A} = 0$; pre-multiplying by a factor $e^{\mathcal{A}^T t} \neq 0$ and post-multiplying by $e^{\mathcal{A}t} \neq 0$ both terms we return to the case

$$e^{\mathcal{A}^T t} \mathcal{A}^T (P_1 - P_2) e^{\mathcal{A}t} + e^{\mathcal{A}^T t}(P_1 - P_2)\mathcal{A}e^{\mathcal{A}t} = 0$$

$$\mathcal{A}^T e^{\mathcal{A}^T t}(P_1 - P_2)e^{\mathcal{A}t} + e^{\mathcal{A}^T t}(P_1 - P_2)\mathcal{A}e^{\mathcal{A}t} =$$

$$\frac{d}{dt}\left(e^{\mathcal{A}^T t}(P_1 - P_2)e^{\mathcal{A}t}\right) =$$

This means that the product $e^{\mathcal{A}^T t}(P_1 - P_2)e^{\mathcal{A}t}$ is constant; knowing that for $t \to \infty$ the exponential $e^{\mathcal{A}t}$ evaluates to zero, in order to ensure that the term remains constant we need to have $P_1 - P_2 = 0$, hence the two solutions P_1, P_2 must coincide.

Proof 2: Lyapunov inequality implies exponential stability To show that a system withstand the Lyapunov inequality (so $\mathcal{A}^T P + T\mathcal{A} <$ and $P > 0$) we have to use the **Lyapunov function** $\mathcal{V}(x)$ defined as

$$\mathcal{V}(x) = x^T P x \quad \in \mathbb{R} \qquad \forall x \in \mathbb{R}^n \tag{1.56}$$

Observing that the result is a scalar, we can apply the *sandwich* relation (equation 1.54) to such formulation hence

$$\lambda_{min}\{P\}|x|^2 \leq x^T P x \leq \lambda_{max}\{P\}|x|^2$$

rewritten as

$$c_1|x|^2 \leq \mathcal{V}(x) \leq c_2|x|^2 \qquad \forall x \in \mathbb{R}^n$$

If we now take the directional derivative of the Lyapunov function with direction x what we obtain is

$$\frac{d\mathcal{V}(x)}{dt} = \frac{\partial \mathcal{V}}{\partial x}\frac{dx(t)}{dt} = \frac{\partial \mathcal{V}}{\partial x}\dot{x} = \frac{\partial \mathcal{V}}{\partial x}\mathcal{A}x = \nabla \mathcal{V}^T(x)\mathcal{A}x = \langle \nabla \mathcal{V}(x), \mathcal{A}x \rangle$$

it's shown that the gradient $\nabla V = \nabla x^T P x$ is equals to $2Px$ and hence we can say

$$\langle \nabla \mathcal{V}(x), \mathcal{A}x \rangle = 2P_x\mathcal{A}x = x^T P\mathcal{A}x + x^T P\mathcal{A}x = x^T P\mathcal{A}x + x^T \mathcal{A}^T P^T x$$

$$= x^T(P\mathcal{A} + \mathcal{A}^T P)x = -x^T \Sigma x$$

Knowing that for *(v)* the matrix $\mathcal{A}^T P + P\mathcal{A}$ is negative definite for the Lyapunov inequality, then it means that the matrix $\Sigma = -(P\mathcal{A} + \mathcal{A}^T P)$ is positive definite. Having from the *sandwich* inequality that $\lambda_{min}\{\Sigma\}|x|^2 \leq x^T \Sigma x$ then it's also true that $-\lambda_{min}\{\Sigma\}|x|^2 \geq -x^T \Sigma x$: we can finally state that

$$\langle \nabla \mathcal{V}(x), \mathcal{A}x \rangle \leq -\lambda_{min}\{\Sigma\}|x|^2 \leq -\lambda_{min}\{\Sigma\}\frac{1}{c_2}\mathcal{V}(x) = -c_3 \mathcal{V}^2(x)$$

where this relationships are commonly referred as *flow inequalities*. In order to complete the proof we have to use the **comparison theorem** stating *given a scalar function $v(t)$ that's differentiable and such that $\dot{v}(t) \leq -\mu v(t)$ for all $t \geq t_0$ where $\mu \in \mathbb{R}$ is a constant, then it happens that $v(t) \leq e^{-\mu(t-t_0)}v(t_0)$ for all $t \geq t_0$.* We can so state that for each solution $x(t)$ (for $t \geq t_0$) we can define the scalar function $v(t) = \mathcal{V}(x(t))$ what we obtain from the flow inequality is that

$$\dot{v}(t) = \langle \nabla \mathcal{V}(x(t)), \dot{x}(t) \rangle = \langle \nabla \mathcal{V}(x(t)), \mathcal{A}x(t) \rangle \leq -c_3 \mathcal{V}(x(t))$$

Applying the comparison theorem to such result we can express

$$\mathcal{V}(x(t)) = v(t) \leq e^{-c_3(t-t_0)}v(t_0) = e^{-c_3(t-t_0)}\mathcal{V}(x(t_0))$$

From the initial *sandwich* of the proof we furthermore obtain

$$c_1|x(t)|^2 \leq \mathcal{V}(x(t)) \leq e^{-c_3 t}\mathcal{V}(x(t_0)) \leq e^{-c_3(t-t_0)}c_2|x(t_0)|^2$$

We can so prove the exponential decay of the solution by making explicit the inequality respect to it's norm:

$$|x(t)|^2 \leq \frac{c_2}{c_1}e^{-c_3(t-t_0)}|x(t_0)|^2 \qquad \Rightarrow \qquad |x(t)| \leq \sqrt{\frac{c_2}{c_1}}e^{-\frac{c_3}{2}(t-t_0)}|x(t_0)|$$

Theorem for discrete-time system For discrete-time LTI systems in the form $x^+ = \mathcal{A}x$ we can express similar equivalent statements as for the continuous-time one:

(i) it is asymptotically stable;

(ii) it is exponentially stable;

(iii) all the eigenvalues of $\mathcal{A}$ have a magnitude strictly lower then one ($|\lambda_i\{\mathcal{A}\}| \leq 1 \ \forall i$);

(iv) for each symmetric positive definite matrix $Q = Q^T > 0 \in \mathbb{R}^{n \times n}$ there exists a symmetric matrix $P \in \mathbb{R}^{n \times n}$ such that $P^T A = PA > 0$ and that satisfy the **Lyapunov equality**

$$\mathcal{A}^T P \mathcal{A} - P = -Q$$

(v) there exists a symmetric matrix $P = P^T > 0 \in \mathbb{R}^{n \times n}$ that satisfies the **Lyapunov inequality**

$$\mathcal{A}^T P \mathcal{A} - P < 0$$

1.7 Linear quadratic regulator

Assuming a continuous-time LTI system where all the staates can be measured, the best control system that we can use it the **linear quadratic regulator** that it's based on the minimization of the functional

$$\mathcal{J} := \int_0^\infty y^T(\tau) Q y(t) + u^T(\tau) R u(\tau) \, d\tau \tag{1.57}$$

where Q, R are two symmetric positive definite matrices. In particular the linear quadratic regular is based on the solution of the following optimal control problem known as the **Bayes Rule**:

$$\mathcal{J}^* = \min_{u(t),\, t \geq 0} \mathcal{J} = \min_{u(t),\, t \geq 0} \int_0^\infty y^T(\tau) Q y(t) + u^T(\tau) R u(\tau) \, d\tau \tag{1.58}$$

The matrices Q, R are used respectively to evaluate (and penalize) the *output* and *control signal energy* of the system. **DA RICHIEDERE BENE QUESTA PARTE IL PERCHE**

With the problem as here state, we can (partially) prove that the solution of the optimal control problem is characterized by the input

$$u^*(t) = -Kx(t) \tag{1.59}$$

where K is a matrix that has to be *tuned* according to the chosen matrices Q, R (but also on $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ characterizing the system); note that the solution implies that all the states can be measured and used as input in the system.

Feedback invariant In order to prove that equation 1.59 is indeed the argument minimizing the cost functional $\mathcal{J}$ we have to introduce the functional

$$\mathcal{H}\big(x(\cdot), u(\cdot)\big) := -\int_0^\infty \Big(\mathcal{A}x(t) + \mathcal{B}u(t)\Big)^T Px(t) + x^T(t)P\Big(\mathcal{A}x(t) + \mathcal{B}u(t)\Big) dt \tag{1.60}$$

that is a **feedback invariant** property of the system (and depends only on the initial state) as long it holds that $\lim_{t \to \infty}(x) = 0$.

We can prove that such functional is invariant substituting $\mathcal{A}x + \mathcal{B}u = \dot{x}$ and showing that

$$\mathcal{H}\big(x(\cdot), u(\cdot)\big) = -\int_0^\infty \dot{x}^T Px + x^T P\dot{x} \, dt = -\int_0^\infty \frac{d}{dt}\left(x^T Px\right) dt = -\left[x^T(t)Px(t)\right]_0^\infty$$

$$= -x^T(\infty)Px(\infty) + x^T(0)Px(0) = x_0^T Px_0$$

where the quadratic term $x^T(\infty)Px(\infty)$ has been cancelled due to the preliminary assumption that $x(\infty) = 0$.

Proof   To prove that the solution $u^*(t) = -Kx(t)$ is indeed the solution of the optimal control problem, we consider a system whose output depends only on the system (hence $\mathcal{D}$ is identically null for simplicity in calculation). Assuming that the functional $\mathcal{J}$ can be rewritten as

$$\mathcal{J} = \mathcal{H}\big(x(\cdot), u(\cdot)\big) + \int_0^\infty \Lambda\big(x(t), u(t)\big)\, dt$$

where Λ is a function characterized by having

$$\min_u \Lambda\big(x, u\big) = 0 \qquad \forall x$$

To show that this rewrite of the functional $\mathcal{J}$ is equal to the original statement we can add and subtract from the original definition the functional $\mathcal{H}$:

$$\mathcal{J} = \int_0^\infty y^T(\tau) Q y(t) + u^T(\tau) R u(\tau)\, d\tau + \mathcal{H}(x, u) - \mathcal{H}(x, u)$$

Taking in the integral the expansion of $-\mathcal{H}$ and knowing that $y = \mathcal{C}x$ (that transposed evaluates to $(\mathcal{C}x)^T = x^T \mathcal{C}^T$) then we can rewrite

$$\mathcal{J} = \int_0^\infty x^T \mathcal{C}^T Q \mathcal{C} x + u^T R u + \big(\mathcal{A}x + \mathcal{B}u\big)^T P x + x^T P\big(\mathcal{A}x + \mathcal{B}u\big)\, d\tau + \mathcal{H}(x, u)$$
$$= \int_0^\infty x^T \big(\mathcal{C}^T Q \mathcal{C} + \mathcal{A}^T P + P \mathcal{A}\big) x + 2 x^T P \mathcal{B} u + u^T R u\, dt + \mathcal{H}(x, u)$$

Knowing that the expected solution is in the form $u + Kx$, performing it's quadratic form respect to R evaluates to

$$\Lambda(x, u) = \big(u^T + x^T K^T\big) R\big(u + Kx\big) = u^T R u + x^T K^T R K x + 2 x^T K^T R u$$

Matching the *mixed* quadratic form involving $x^T \cdot u$ in both $\mathcal{J}$ and Λ allows to compute the value of the matrix K as

$$P \mathcal{B} = K^T R \qquad \Rightarrow \qquad K^T = P \mathcal{B} R^{-1} \qquad \Rightarrow \qquad K = \big(P \mathcal{B} R^{-1}\big)^T = \big(R^{-1}\big)^T \mathcal{B}^T P^T = R^{-1} \mathcal{B}^T P$$

With such choice we can rewrite the functional as

$$\mathcal{J} = \mathcal{H}(x, u) + \int_0^\infty x^T \big(\mathcal{C}^T Q \mathcal{C} + \mathcal{A}^T P + P \mathcal{A}\big) x + 2 x^T P \mathcal{B} u + u^T R u\, dt$$
$$= \mathcal{H}(x, u) + \int_0^\infty x^T \big(\mathcal{C}^T Q \mathcal{C} + \mathcal{A}^T P + P \mathcal{A} + P \mathcal{B} R^{-1} \mathcal{B}^T P - P \mathcal{B} R^{-1} \mathcal{B}^T P\big) x + 2 x^T P \mathcal{B} u + u^T R u\, dt$$

We can so see that $x^T P \mathcal{B} R^{-1} \mathcal{B}^T P x$ matches the expansion of Λ evaluated at $x^T K^T R K x$. With that in order to have a functional $\mathcal{J}$ in the form $\mathcal{H}(x, u) + \int_0^\infty \Lambda(x, u)\, dt$ we have to impose the condition on the quadratic term $x^T \cdot x$ determining the **algebraic Riccati equation**:

$$\mathcal{A}^T P + P \mathcal{A} + \mathcal{C}^T Q \mathcal{C} - P \mathcal{B} R^{-1} \mathcal{B}^T P = 0 \tag{1.61}$$

The problem now is finding a symmetric positive definite matrix P that satisfies such equation and if we ensure that $\lim_{t \to \infty} x(t) = 0$ then the solution $u^* = -Kx$ is the solution of the linear quadratic regulator; it also happens that the optimal value $\mathcal{J}^+$ evaluates to $H(x, u) = x_0^T P x_0$.

RICONTROLLARE

1.8   Controllability

The main idea behind the *controllability* is to determine what the inputs u can *accomplish* in the output depending on the initial states. Recalling that the whole response of a continuous-time linear time-varying system is made by a free and a forced term in the form

$$x(t_1) = \Phi(t_1, t_0) x(t_0) + \int_{t_0}^{t_1} \Phi(t_1, \tau) \mathcal{B}(\tau) u(\tau)\, d\tau$$

In particular given two times $t_1 > t_0 \geq 0$, we define with $\mathcal{R}[t_0, t_1]$ the **reachable set** (that's indeed a linear subspace) defined as

$$\mathcal{R}[t_1, t_2] = \left\{ x_1 \in \mathbb{R}^n \ : \ \exists u \text{ such that } \int_{t_0}^{t_1} \Phi(t_1, \tau) \mathcal{B}(\tau) u(\tau) \, d\tau = x_1 \right\} \tag{1.62}$$

The idea of the reachable set is to know which states the system can reach in the interval $[t_0, t_1]$ starting from a zero initial state $x(t_0) = 0$. Conversely the **controllable set** (that's a linear subspace) $\mathcal{C}[t_0, t_1]$ as the set of all the initial states $x_0 = x(t_0)$ resulting in a zero final state $x(t_1) = 0$:

$$\begin{aligned} \mathcal{C}[t_1, t_2] &= \left\{ x_0 \in \mathbb{R}^n \ : \ \exists u \text{ such that } \int_{t_0}^{t_1} \Phi(t_0, \tau) \mathcal{B}(\tau) \big(-u(\tau) \big) \, d\tau = x_1 \right\} \\ &= \left\{ x_0 \in \mathbb{R}^n \ : \ \exists u \text{ such that } \int_{t_0}^{t_1} \Phi(t_0, \tau) \mathcal{B}(\tau) v(\tau) \, d\tau = x_1 \right\} \end{aligned} \tag{1.63}$$

Such integral (slightly different from the reachable set) has been obtained pre-multiplying the response by the inverse $\Phi(t_1, t_0)^{-1} = \Phi(t_0, t_1)$ what we obtain is

$$x_1 = 0 = \underline{\Phi(t_0, t_1) \Phi(t_1, t_0)} x_0 + \int_{t_0}^{t_1} \Phi(t_0, t_1) \Phi(t_1, \tau) \mathcal{B}(\tau) u(\tau) \, d\tau = x_0 + \int_{t_0}^{t_1} \Phi(t_0, \tau) \mathcal{B}(\tau) u(\tau) \, d\tau$$

$$-x_0 = \int_{t_0}^{t_1} \Phi(t_0, \tau) \mathcal{B}(\tau) u(\tau) \, d\tau$$

Linear algebra recall In order to ease the following definition of *Gramians* a recall on concept of linear algebra is required. Given a matrix $W \in \mathbb{R}^{m \times n}$ we define as *range* or, preferably, **image** of W the set

$$\text{Im} W = \{ y \in \mathbb{R}^m \ : \ y = Wx \text{ with } x \in \mathbb{R}^n \} \subseteq \mathbb{R}^m$$

The dimension of the image of W coincides with its **rank**, meaning the number of independent columns (and also rows) composing the matrix. The **kernel** (also referred as *null space*) of the matrix is defined by the set

$$\ker W = \{ x \in \mathbb{R}^n \ : \ Wx = 0 \} \subseteq \mathbb{R}^n$$

The dimension $\dim\{\ker W\}$ is sometimes referred as *nullity*.

An important theorem that we have to always keep in mind is the **fundamental theorem of linear algebra** stating that for any matrix $W \in \mathbb{R}^{m \times n}$ we have that

$$\dim\{\text{Im} W\} + \dim\{\ker W\} = n$$

Given a linear space $\mathcal{V} \subseteq \mathbb{R}^n$ we can determine its **orthogonal complement** $\mathcal{V}^\perp$ the set defines as

$$\mathcal{V}^\perp = \left\{ x \in \mathbb{R}^n \text{ such that } x^T z = 0 \ \forall z \in \mathcal{V} \right\}$$

Singular value decomposition The **singular value decomposition** **SVD** is a particular transformation of any matrix $W \in \mathbb{R}^{m \times n}$ that's defined as

$$W = USV^T \tag{1.64}$$

where U, V are special orthogonal matrices (so such that $U^T U = UU^T = \mathcal{I}$ and so the vectors composing such matrices determines an orthonormal base of the space they are representing) and S is a matrix containing diagonally the singular values $\sigma_i = \sqrt{\lambda_i \{W^T W\}}$ of the matrix W:

$$P = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_p & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

MANCANO DA SCRIVERE DELLE CONSIDERAZIONI

Gramians   Given two times $t_1 > t_0 \geq 0$, the **Gramian** **reachability** and **controllability** of a continuous-time linear time-varying system are described by the symmetric matrices

$$W_{\mathscr{R}}[t_0, t_1] = \int_{t_0}^{t_1} \Phi(t_1, \tau)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_1, \tau)\, d\tau \qquad \in \mathbb{R}^{n \times n}$$

$$W_{\mathscr{C}}[t_0, t_1] = \int_{t_0}^{t_1} \Phi(t_0, \tau)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_0, \tau)\, d\tau \qquad \in \mathbb{R}^{n \times n}$$

(1.65)

We in fact relate such matrices respectively with the reachable and controllable set as

$$\mathscr{R}[t_0, t_1] = \mathrm{Im}\, W_{\mathscr{R}}[t_0, t_1] \qquad\qquad \mathscr{C}[t_0, t_1] = \mathrm{Im}\, W_{\mathscr{C}}[t_0, t_1] \qquad (1.66)$$

Moreover denoting with $x_1 = W_{\mathscr{R}}[t_0, t_1]\eta_1$ the resulting reachable state (similarly for controllability $x_0 = W_{\mathscr{C}}[t_0, t_1]\eta_0$) we have that the optimal input $u(t)$ that gets to such value of the states is the one determined by

$$u^*(t) = \mathcal{B}^T(t)\Phi^T(t_1, t)\eta_1 \qquad \forall t \in [t_0, t_1] \qquad (1.67)$$

Similarly for the controllable set the optimal output is defined as $u^*(t) = -\mathcal{B}^T(t)\Phi^T(t_0, t)\eta_0$.

Proof   To prove the Gramian theorem (associated to equation 1.66) we can firstly check that $\mathrm{Im}\, W_{\mathscr{R}} \subseteq \mathscr{R}$ (the proof is made for the reachability of the system, but similar calculations can be performed for the observability) and then observing that $\mathscr{R} \subseteq \mathrm{Im}\, W_{\mathscr{R}}$: this two conditions happening simultaneously indeed implies that $\mathscr{R} = \mathrm{Im}\, W_{\mathscr{R}}$.

i) Taking $x_1 \in \mathrm{Im}\, W_{\mathscr{R}}$ a state belonging to the image of $W_{\mathscr{R}}$, using the optimal control solution presented in equation 1.67 in the variation of constant formula (equation 1.28 at page 12 where, for the reachability, x_0 is assumed to be zero):

$$x(t_1) = \int_{t_0}^{t_1} \Phi(t_1, \tau)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_1, \tau)\eta_1\, d\tau = \int_{t_0}^{t_1} \Phi(t_1, \tau)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_1, \tau)\, d\tau\; \eta_1$$

$$= W_{\mathscr{R}}[t_0, t_1]\eta_1 = x_1$$

ii) Considering now that for any state $x_1 \in \mathscr{R}[t_0, t_1]$ there exists at least one control input u such that $x_1 = \int_{t_0}^{t_1} \Phi(t_1, \tau)\mathcal{B}(\tau)u(\tau)\, d\tau$ (still from the variation of constants formula), then we can show that

$$x_1 \in \mathrm{Im}\, W_{\mathscr{R}} = \left(\ker W_{\mathscr{R}}^T \right)^{\perp} = \left(\ker W_{\mathscr{R}} \right)^{\perp}$$

For all vectors $\eta_{1k} \in \ker W_{\mathscr{R}}$ we can prove that $x_1^T \eta_{1k} = 0$; expanding the the definition of $x(t)$ from the variation of constants we have

$$x^T \eta_{1k} = \int_{t_0}^{t_1} \Phi(t_1, \tau)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_1, \tau)\eta_{1k}\, d\tau$$

Since η_{1k} belongs to the kernel we have that $(W_{\mathscr{R}}\eta_{1k})^T = \eta_{1k}^T W_{\mathscr{R}}$, meaning

$$\eta_{1k}^T W_{\mathscr{R}}[t_0, t_1]\eta_{1k} = \int_{t_0}^{t_1} \eta_{1k}\Phi(t_1, \tau)\Phi(t_1, \tau)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_1, \tau)\eta_{1k}\, d\tau$$

$$\eta_{1k}^T 0 = \int_{t_0}^{t_1} \left\| \mathcal{B}^T(\tau)\Phi^T(t_1, \tau)\eta_{1k} \right\|^2 d\tau = 0$$

Knowing that the element inside the integral is non-negative, in order to have a zero integral it must have $\mathcal{B}^T(\tau)\Phi^T(t_1, \tau)\eta_{1k} = 0$.

Theorem   Given two times $t_1 > t_0 \geq 0$:

- for each state $x_1 \in \mathscr{R}[t_0, t_1]$ in the reachable set, the control input u^* described in equation 1.67 transforms the state from $x_0 = 0$ to $x_1 = W_{\mathscr{R}}\eta_1$ with the minimum energy

$$\int_{t_0}^{t_1} |u(\tau)|^2\, d\tau$$

Moreover it has that $\min_u \int_{t_0}^{t_1} |u^*(\tau)|^2\, d\tau = \eta_1^T W_{\mathscr{R}}\eta_1$.

- similarly for every initial state $x_0 \in \mathscr{C}[t_0, t_1]$ the control $u^*(t) = -\mathcal{B}^T(t)\Phi^T(t_0, t)\eta_0$ transforms the initial state $x_0 \neq 0$ to the final state $x_1 = 0$ with the minimum energy and the minimum value is $\eta_0^T W_\mathscr{C} \eta_0$.

The main idea behind this theorem is that the matrices $W_\mathscr{R}, W_\mathscr{C}$ can be used to estimate the energy required for *moving the states*.

Proof   We can prove that u^* minimizes the energy required for moving the states from x_0 to $x_1 \in \mathscr{R}[t_0, t_1]$ (similar proof can be written for the controllability); considering $\tilde{u}(t)$ one other input that allows to *achieve the goal*, so such that $\int_{t_0}^{t_1} \Phi\mathcal{B}\tilde{u}\, d\tau = x_1$, we can compute the difference between such solution and the optimal one as

$$0 = \int_{t_0}^{t_1} \Phi(t_1, \tau)\mathcal{B}(\tau) \underbrace{\left(u^*(\tau) - \tilde{u}(\tau)\right)}_{=v(z)} d\tau$$

Writing with $v(z)$ the difference $u^* - \tilde{u}$ between the optimal input and another valid one. Studying the minimization of the energy of the alternative input $\tilde{u}$, we have that

$$\int_{t_0}^{t_1} |\tilde{u}(\tau)|^2\, d\tau = \int_{t_0}^{t_1} \tilde{u}^T(\tau)\tilde{u}(\tau)\, d\tau = \int_{t_0}^{t_1} \left(u^*(\tau) - v(\tau)\right)^T \left(u^*(\tau) - v(\tau)\right) d\tau$$

$$= \int_{t_0}^{t_1} u^{*T} u^*\, d\tau + \int_{t_0}^{t_1} v^T v\, d\tau + \int_{t_0}^{t_1} 2u^{*T} v\, d\tau$$

Knowing that the ideal input u^* is in the form $\mathcal{B}^T \Phi^T \eta_1$ (equation 1.67) we can expand the definition of the energy of the alternative input $\tilde{u}$ as

$$\int_{t_0}^{t_1} |\tilde{u}|\, d\tau = \eta_1^T \int_{t_0}^{t_1} \Phi\mathcal{B}\mathcal{B}^T\Phi\, d\tau\, \eta_1 + \int_{t_0}^{t_1} |v|^2\, d\tau - 2\eta_1^T \overset{0}{\cancel{\int_{t_0}^{t_1} \Phi\mathcal{B}v\, d\tau}}$$

$$= \eta_1^T W_\mathscr{R} \eta_1 + \int_{t_0}^{t_1} |v|^2\, d\tau$$

The mixed terms has been driven to zero considering the initial condition on the proof; considering that $|v|$ is a non-negative quantity, it means that integrating over time results in an increment on the energy spent and the optimal solution u^* minimizes the energy to the minimum value $\eta_1^T W_\mathscr{R} \eta_1$.

Continuous-time linear time-invariant case

Considering the case of continuous-time LTI systems the definitions of the Gramian reachability (equation 1.65) can be simplified: considering in fact that $\Phi(t_1, \tau) = e^{\mathcal{A}(t_1 - \tau)}$ we have that

$$W_\mathscr{R}[t_0, t_1] = \int_{t_0}^{t_1} e^{\mathcal{A}(t_1 - \tau)} \mathcal{B}\mathcal{B}^T e^{\mathcal{A}^T(t_1 - \tau)}\, d\tau \xrightarrow{\ t_1 - \tau = t\ } \int_0^{t_1 - t_0} e^{\mathcal{A}t} \mathcal{B}\mathcal{B}^T e^{\mathcal{A}^T t}\, dt \qquad (1.68)$$

Similarly for the controllability Gramian we obtain

$$W_\mathscr{C}[t_0, t_1] = \int_{t_0}^{t_1} e^{\mathcal{A}(t_0 - \tau)} \mathcal{B}\mathcal{B}^T e^{\mathcal{A}^T(t_0 - \tau)}\, d\tau \xrightarrow{\ t_0 - \tau = -t\ } \int_0^{t_1 - t_0} e^{-\mathcal{A}t} \mathcal{B}\mathcal{B}^T e^{-\mathcal{A}^T t}\, dt$$

To solve the dynamic of such system we can use the so called **controllability matrix** $R \in \mathbb{R}^{n \times (kn)}$ defined as the *concatenation* of matrices in the form

$$R = \begin{bmatrix} \mathcal{B} & \mathcal{A}\mathcal{B} & \mathcal{A}^2\mathcal{B} & \dots & \mathcal{A}^{n-1}\mathcal{B} \end{bmatrix} \qquad (1.69)$$

As consequence a theorem states that for any two times $t_1 > t_0 \geq 0$ for a continuous-time linear time-invariant system we have that

$$\mathscr{R}[t_0, t_1] = \mathrm{Im}W_\mathscr{R}[t_0, t_1] = \mathrm{Im}R = \mathrm{Im}W_\mathscr{C}[t_0, t_1] = \mathscr{C}[t_0, t_1]$$

meaning that,in this case, the controllability and reachability subspaces are the same, they coincides. As consequent property we can consider the *time reversibility*, meaning that a state is controllable if and only if its reachable, and also the *time scaling*, meaning that reachability/controllability does not depend on the time difference $t_1 - t_0$ (in fact there always exists a input $u(t)$ that allows to reach the desired state in a *arbitrarily small time*).

Proof   To prove that controllability and reachability set are coincident, we can prove that $\mathcal{R} \subseteq \mathrm{Im}R$ as a first step and then prove that $\mathrm{Im}R \subseteq \mathrm{Im}W_{\mathcal{R}}$ (considering that is already proven that $\mathcal{R} = \mathrm{Im}W_{\mathcal{R}}$); similarly it can be proven that $\mathrm{Im}R = \mathrm{Im}W_{\mathcal{C}} = \mathcal{C}$.

i) considering a state $x_1 \in \mathcal{R}[t_0, t_1]$ means that exists an input $u(t)$ such that

$$x_1 = \int_{t_0}^{t_1} e^{\mathcal{A}(t_1-\tau)} \mathcal{B} u(\tau)\, d\tau$$

Using the Cayley-Hamilton theorem (equation 1.36, page 15) we can rewrite this expression as

$$x_1 = \int_{t_0}^{t_1} \sum_{i=0}^{n-1} \alpha_i(t_1 - \tau) \mathcal{A}^i \mathcal{B} u(\tau)\, d\tau = \sum_{i=0}^{n-1} \mathcal{A}^i \mathcal{B} \int_{t_0}^{t_1} \alpha_i(t_1 - \tau) u(\tau)\, d\tau$$

This expression can be rewritten as a linear combination of the form

$$x_1 = \underbrace{\begin{bmatrix} \mathcal{B} & \mathcal{AB} & \dots & \mathcal{A}^{n-1}\mathcal{B} \end{bmatrix}}_{R} \underbrace{\begin{pmatrix} \int_{t_0}^{t_1} \alpha_0(t_1 - \tau) u(\tau)\, d\tau \\ \int_{t_0}^{t_1} \alpha_1(t_1 - \tau) u(\tau)\, d\tau \\ \vdots \\ \int_{t_0}^{t_1} \alpha_{n-1}(t_1 - \tau) u(\tau)\, d\tau \end{pmatrix}}_{\nu}$$

Having ν a vector we indeed have that x_1 is in the image set of R because it can be regarded as $R\nu$.

ii) it has been previously proven that $\mathcal{B}^T \Phi^T(t_1, \tau) \eta_{1k} = 0$ for any vector $\eta_{1k} \in \ker W_{\mathcal{R}}[t_0, t_1]$ and $\forall \tau \in [t_0, t_1]$ in the time-variant case, considering a LTI system we have that

$$(a) \qquad \mathcal{B}^T e^{\mathcal{A}^T(t_1-\tau)} \eta_{1k} = 0$$

Deriving with respect to time this expression transposed evaluates to $\frac{d}{dt}(a)^T = -\eta_{1k}^T \mathcal{A} e^{\mathcal{A}(t_1-\tau)} \mathcal{B}$ and more generally we can express

$$\frac{d^i}{dt^i}(a)^T = (-1)^i \eta_{1k}^T \mathcal{A}^i e^{\mathcal{A}(t_1-\tau)} \mathcal{B} = 0$$

Evaluating this derivatives for $\tau = t_1$ results in $\eta_{1k}^T \mathcal{A}^i \mathcal{B} = 0\ \forall i$; for any state $x_1 = R\nu \in \mathrm{Im}W_{\mathcal{R}}[t_0, t_1]$ we so have that

$$x_1^T \eta_{1k} = \eta_{1k}^T x_1 = \eta_{1k}^T R\nu = \eta_{1k}^T \begin{bmatrix} \mathcal{B} & \mathcal{AB} & \dots & \mathcal{A}^{n-1}\mathcal{B} \end{bmatrix} \nu$$

$$= \begin{bmatrix} \eta_{1k}^T \mathcal{B}^{\,0} & \eta_{1k}^T \mathcal{AB}^{\,0} & \dots & \eta_{1k}^T \mathcal{A}^{n-1}\mathcal{B}^{\,0} \end{bmatrix} \nu$$

This means that $x_1 \perp \eta_{1k}$ for any vector $\eta_{1k} \in \ker W_{\mathcal{R}}$, meaning that x_1 belongs to the orthogonal subspace $(\ker W_{\mathcal{R}})^\perp$. Knowing that in general $\mathrm{Im}W = (\ker W^T)^\perp$, but in this case $W_{\mathcal{R}}$ is symmetric we have that $\mathrm{Im}W_{\mathcal{R}} = (\ker W_{\mathcal{R}})^\perp$ and so we have that $\mathrm{Im}R \subset \mathrm{Im}W_{\mathcal{R}}$.

We can so state that in general the reachable/controllable subspaces, for LTI systems, are properties of the system itself and are determined only by the matrices $\mathcal{A}$ and $\mathcal{B}$.

Discrete-time systems

Discrete-time LTV   With the linear continuous-time system analysed (both time-variant and in-variant), the same concepts regarding the controllability (reachability) of system can be described also for the discrete-time case. Given the state-space representation $x^+ = \mathcal{A}(t)x + \mathcal{B}(t)u$, from the variation of constant formula we have

$$x(t_1) = \Phi(t_1, t_0)x_0 + \sum_{\tau=t_0}^{t_1-1} \Phi(t, \tau+1)\mathcal{B}(\tau)u(\tau)$$

Given so 2 times $t_1 > t_0 \geq 0$, we can define the $[t_0, t_1]$-reachable subspace and the $[t_0, t_1]$-controllable one the sets described as

$$\mathscr{R}[t_0, t_1] = \left\{ x_1 \text{ such that } \exists u(t) \text{ with } t = t_0, t_0+1, \ldots, t_1 - 1 \; : \; x(t_1) = \sum_{\tau=t_0}^{t_1-1} \Phi(t, \tau+1)\mathcal{B}(\tau)u(\tau) \right\}$$

$$\mathscr{C}[t_0, t_1] = \left\{ x_0 \text{ such that } \exists u(t) \text{ with } t = t_0, \ldots, t_1 - 1 \; : \; 0 = \Phi(t_1, t_0)x_0 + \sum_{\tau=t_0}^{t_1-1} \Phi(t, \tau+1)\mathcal{B}(\tau)u(\tau) \right\}$$

Moreover if the matrix $\mathcal{A}$ is non-singular $\forall t \in [t_0, t_1]$, then the state transition matrix Φ can be inverted and we can re-define the controllable set as

$$\mathscr{C}[t_0, t_1] = \left\{ x_0 \text{ such that } \exists v(t) = -u(t) \; : \; x_0 = \sum_{\tau=t_0}^{t_1-1} \Phi(t_0, \tau+1)\mathcal{B}(\tau)v(\tau) \right\}$$

The **discrete-time Gramians** of the system are defined as

$$W_{\mathscr{R}} = \sum_{\tau=t_0}^{t_1-1} \Phi(t_1, \tau+1)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_1, \tau+1) \qquad W_{\mathscr{C}} = \sum_{\tau=t_0}^{t_1-1} \Phi(t_0, \tau+1)\mathcal{B}(\tau)\mathcal{B}^T(\tau)\Phi^T(t_0, \tau+1)$$

Observe that the controllable Gramian $W_{\mathscr{C}}$ is valid only if the matrix $\mathcal{A}$ is non-singular in the domain $[t_0, t_1]$ (as for the second definition of the set itself).

It still holds that $\mathrm{Im}W_{\mathscr{R}}[t_0, t_1] = \mathscr{R}[t_0, t_1]$ and $\mathrm{Im}W_{\mathscr{C}}[t_0, t_1] = \mathscr{C}[t_0, t_1]$ for any times $t_1 > t_0 \geq 0$ and, moreover, if $x_1 = W_{\mathscr{R}}\eta_1 \in \mathrm{Im}W_{\mathscr{R}}$ is a state in the reachable set, then the control $u^*(t) = \mathcal{B}^T(t)\Phi^T(t_1, t+1)\eta_1$ moves $x_0 = 0$ to $x_1 \neq 0$ with the minimum energy; similarly for the control set if we have $x_0 = W_{\mathscr{C}}\eta_0$, them the optimal control is $u^*(t) = -B^T(t)\Phi^T(t_0, t+1)\eta_0$.

Discrete-time LTI   Considering a system with state-space representation $x^+ = \mathcal{A}x + \mathcal{B}u$, then if $\mathcal{A}$ is invertible and knowing that $\Phi(t_1, \tau) = \mathcal{A}^{t_1-\tau}$ performing some change of variables we can rewrite the Gramians of discrete-time LTI systems as

$$W_{\mathscr{R}} = \sum_{s=0}^{t_1-t_0-1} \mathcal{A}^s \mathcal{B}\mathcal{B}^T \left(\mathcal{A}^T\right)^s \qquad\qquad W_{\mathscr{C}} = \sum_{s=0}^{t_1-t_0-1} \mathcal{A}^{-s-1}\mathcal{B}\mathcal{B}^T\left(\mathcal{A}^T\right)^{-s-1}$$

Also in this case we can say that for any $t_1 > t_0 \geq 0$

$$\mathscr{R}[t_0, t_1] = \mathrm{Im}W_{\mathscr{R}}[t_0, t_1] = \mathrm{Im}R = \mathrm{Im}W_{\mathscr{C}}[t_0, t_1] = \mathscr{C}[t_0, t_1]$$

if we assume that $\mathcal{A}$ is invertible (for the controllability property) and $t_1 \geq t_0 + n$ (this is due to the fact that *we cannot scale the input in time as much as we want*).

1.8.1   State feedback and single input eigenvalue assignment

Given a linear time-invariant system that's so characterized by a state-space response $x^+/\dot{x} = \mathcal{A}x + \mathcal{B}u$, if we can measure all the states then we can assign as input u a linear combination of them, hence $u = -Kx$ (where K is a matrix): this means that the states variations corresponds to a form

$$\dot{x} = (\mathcal{A} - \mathcal{B}K)x$$

Single input eigenvalue assignment theorem   Considering a controllable (meaning that $\mathscr{R} = R^n$) single input LTI system characterized by the pair of matrices $\mathcal{A}, \mathcal{B}$, then for each set $\lambda_1, \ldots, \lambda_n$ of desired eigenvalues there always exists a matrix $K \in \mathbb{R}^{k \times n}$ such that

$$\mathcal{A}_{cl} = \mathcal{A} - \mathcal{B}K \qquad \text{has eigenvalues } \lambda_1, \ldots, \lambda_n \tag{1.70}$$

where $\mathcal{A}_{cl}$ is usually referred as the ***closed loop matrix***. The goal is now to find a *procedure* to determine such matrix K that allows us to determined the desired eigenvalues.

Recalling the standard controllable form of a system (seen in page 10) that's characterized by the matrices

$$\mathcal{A}_{ctr} = \left[\begin{array}{c|ccc} 0 & & \mathcal{I} & \\ \hline -\alpha_0 & -\alpha_1 & \cdots & -\alpha_{n-1} \end{array}\right] \qquad \mathcal{B}_{ctr} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

where the coefficients α_i are the coefficient of the characteristic polynomial of the original matrix $\mathcal{A}$ in the form

$$p_{\mathcal{A}}(s) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0$$

Having in this case a single input system, it means that the desired matrix K is a row vector in the form $K = \begin{bmatrix} k_0 & \cdots & k_{n-1} \end{bmatrix}$. In order to compute such matrix, we can start off by building the characteristic polynomial associated to the desired eigenvalues of the closed loop matrix in the form

$$p_{\mathcal{A}_{cl}}(s) = (s - \lambda_1) \ldots (s - \lambda_2) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0$$

By equating the controllable form of the closed loop matrix with the one obtained by $\mathcal{A}_{ctr} - \mathcal{B}_{ctr}K$ we have that

$$\left[\begin{array}{c|ccc} 0 & & \mathcal{I} & \\ \hline -\beta_0 & -\beta_1 & \cdots & -\beta_{n-1} \end{array}\right] = \left[\begin{array}{c|ccc} 0 & & \mathcal{I} & \\ \hline -\alpha_0 - k_0 & -\alpha_1 - k_1 & \cdots & -\alpha_{n-1} - k_{n-1} \end{array}\right]$$

meaning that the coefficients k_i of the matrix K can be evaluates as $k_i = \beta_i - \alpha_i$.

In the more general case where $\mathcal{A}, \mathcal{B}$ are not in controllable form but are *generic* matrices, we can show that if R is non-singular, then the product $R^{-1}\mathcal{A}R$ evaluates to the matrix $\mathcal{A}_{ob}$ in observable form and $R^{-1}\mathcal{B} = \mathcal{B}_{ob}$ (where R is the controllability matrix seen at page 30) described as

$$\mathcal{A}_{ob} = \left[\begin{array}{c|c} 0 & \begin{array}{c} -\alpha_0 \\ -\alpha_1 \end{array} \\ \hline \mathcal{I} & \begin{array}{c} \vdots \\ -\alpha_{n-1} \end{array} \end{array}\right] \qquad \mathcal{B}_{ob} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This relations can be proven considering that

$$R\mathcal{A}_{ob} = \begin{bmatrix} \mathcal{B} & \mathcal{A}\mathcal{B} & \cdots & \mathcal{A}^{n-1}\mathcal{B} \end{bmatrix} \left[\begin{array}{c|c} 0 & \begin{array}{c} -\alpha_0 \\ -\alpha_1 \end{array} \\ \hline \mathcal{I} & \begin{array}{c} \vdots \\ -\alpha_{n-1} \end{array} \end{array}\right]$$

$$= \left[\begin{array}{ccc|c} \mathcal{A}\mathcal{B} & \cdots & \mathcal{A}^{n-1}\mathcal{B} & x \end{array}\right]$$

where x is a vector determines as the linear combination $x = -\alpha_0\mathcal{B} - \alpha_1\mathcal{A}\mathcal{B} - \cdots - \alpha_{n-1}\mathcal{A}^{n-1}\mathcal{B}$ (where note that all the products $\mathcal{A}^k\mathcal{B}$ are column vectors). Collecting $\mathcal{B}$ we have that $x = \left(-\alpha_0\mathcal{I} - \alpha_1\mathcal{A} - \cdots - \alpha_{n-1}\mathcal{A}^{n-1}\right)\mathcal{B}$ and recalling the Cayley-Hamilton theorem (equation 1.36, page 15) we can consider that the term in the parenthesis exactly as $\mathcal{A}^n$, hence

$$R\mathcal{A}_{ob} = \begin{bmatrix} \mathcal{A}\mathcal{B} & \cdots & \mathcal{A}^{n-1}\mathcal{B} & \mathcal{A}^n\mathcal{B} \end{bmatrix} = \mathcal{A} \begin{bmatrix} \mathcal{B} & \mathcal{A}\mathcal{B} & \cdots & \mathcal{A}^{n-1}\mathcal{B} \end{bmatrix} = \mathcal{A}R$$

The proof of $R^{-1}\mathcal{B} = \mathcal{B}_{ob}$ is straightforward and can be obtained considering that $R\mathcal{B}_{ob}$ is indeed $\mathcal{B}$. We can so prove that for each observable form there exists a matrix M such that $M^{-1}\mathcal{A}_{ob}M = \mathcal{A}_{ctr}$ and $M^{-1}\mathcal{B}_{ob} = \mathcal{B}_{ctr}$ that allows to determine the canonical form. In particular M is the symmetric matrix defines as

$$
M = \begin{bmatrix}
\alpha_1 & \alpha_2 & \dots & \alpha_{n-1} & 1 \\
\alpha_2 & \iddots & & \iddots & \\
\vdots & & \iddots & & \\
\alpha_{n-1} & \iddots & & & \\
1 & & & &
\end{bmatrix}
$$

We can in fact show that

$$
M\mathcal{A}_{ctr} = \left[\begin{array}{c|c}
\alpha^T & 1 \\
& 0 \\
\hline
M_{21} & \vdots \\
& 0
\end{array}\right]
\left[\begin{array}{c|c}
0 & \\
\vdots & \mathcal{I} \\
0 & \\
\hline
-\alpha_0 & -\alpha^T
\end{array}\right]
= \left[\begin{array}{c|ccc}
-\alpha_0 & 0 & \dots & 0 \\
0 & & & \\
\vdots & & M_{21} & \\
0 & & &
\end{array}\right]
$$

where M_{21} is a symmetric submatrix of M and α is the column vector $(\alpha_1, \dots, \alpha_{n-1})$. Observing that the resulting matrix is symmetric we have that

$$
M\mathcal{A}_{ctr} = \left(M\mathcal{A}_{ctr}\right)^T = \mathcal{A}_{ctr}^T M^T = \mathcal{A}_{ob}M
$$

With all this proofs we have the ability now to explicit the general procedure that allows to determine the matrix K that determines the closed loop matrix $\mathcal{A}_{cl} = \mathcal{A} - \mathcal{B}K$ whose eigenvalues are the desired $\lambda_1, \dots, \lambda_n$. Choosing in particular $T = RM$ what we have is that

$$
T^{-1}\mathcal{A}T = \mathcal{A}_{ctr} \qquad \text{and} \qquad T^{-1}\mathcal{B} = \mathcal{B}_{ctr}
$$

Determined the controllable forms of both $\mathcal{A}$ and $\mathcal{B}$ we can compute the matrix K_{ctr} by computing the characteristic polynomials $p_{\mathcal{A}_{ctr}}(s)$ and $p_{\mathcal{A}_{cl}}(s)$ and by matching the coefficients we determine the entries of K_{ctr} as $k_i = \beta_i - \alpha_i$. In particular K associated to the initial matrices $\mathcal{A}, \mathcal{B}$ evaluates to

$$
K = K_{ctr}T^{-1}
$$

Ackerman Another formula used to compute such matrix K in order to assign the eigenvalues to $\mathcal{A} - \mathcal{B}K$ is by applying the **Ackerman equation**

$$
K = \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix} R^{-1} p_{\mathcal{A}_{cl}}(\mathcal{A})
$$

Chapter 2

Non-Linear Systems

Non-linear system are characterized by a description in state space that in both continuous and discrete time case are

$$CT: \begin{cases} \dot{x} = f(x,u) \\ y = h(x,u) \end{cases} \qquad DT: \begin{cases} x^+ = f(x,u) \\ y = h(x,u) \end{cases} \tag{2.1}$$

2.1 Approximations to linear systems

Studying the behaviour of non-linear system is usually a complex task and so a *good* idea is to somehow make the system linear in order to use the associated known results.

2.1.1 Feedback linearization

In general mechanical system are described by models in the form

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + K(q)q = u$$

where M, C, K are matrices depending on the current position and/or velocity of the parameters q of the system; u is in general the applied force on the dynamical system that make it moves. A way to *cancel out* the terms associated to C, K is by simply choosing as control the function $u = C(\dot{q}, q)\dot{q} + K(q)q + M(q)v$ that so results in

$$M(q)\ddot{q} + \underline{C(q,\dot{q})\dot{q}} + \underline{K(q)q} = \underline{C(\dot{q},q)\dot{q}} + \underline{K(q)q} + M(q)v \qquad \Rightarrow \qquad \ddot{q} = v$$

This result has been obtained considering that the matrix M is always invertible (and for mechanical system such condition is satisfied). Such process is so called **feedback linearization** because it exploit a *smart* choice on the control u in order to have a linear relation between $\ddot{q}$ and a desired value v.

2.1.2 Linearization

Considering a non-linear time invariant system (equation 2.1) we call $\varphi(t): \mathbb{R}^+ \to \mathbb{R}$ a **solution** of the system if it holds that

$$\varphi(0) = x_0$$

and

$$\left. \begin{aligned} \dot{\varphi}(t) \\ \varphi^+ \end{aligned} \right\} = f\big(\varphi(t), u(t)\big) \qquad \forall t > 0$$

Considering such definition we say that a pair state-input (x^{eq}, u^{eq}) is an **equilibrium** for a non-linear time-invariant system if

$$u(t) = u^{eq} \quad x(t) = x^{eq} \qquad \Rightarrow \qquad \varphi(t) = x^{eq} \quad \forall t > 0$$

This definition is so saying that a point to be an equilibrium must have a stationary solution that *doesn't move*. As implication we have that for a continuous-time NL-TI system equilibriums are all point satisfying $\dot{x} = f(x^{eq}, u^{eq}) = 0$; similarly for discrete-time system the equilibrium can be obtaining imposing $x^+ = f(x^{eq}, u^{eq}) = x^{eq}$.

Small perturbation condition and linear approximation   Considering now a *small perturbation* both on the initial condition $\delta_{x(0)}$ and on the input signal $\delta_{u(t)}$ from the equilibrium position as

$$x(0) = x^{eq} + \delta_{x(0)} \qquad\qquad u(t) = u^{eq} + \delta_{u(t)}$$

What we would like now to characterize is the *deviation* of the states $\delta_{x(t)} = x(t) - x^{eq}$ and consequently of the output $\delta_{y(t)} = y(t) - y^{eq}$. Using the Taylor series expansion up to the first order we can indeed generate a **linear approximation** about the equilibrium point (x^{eq}, u^{eq}):

$$f(x,u) = f(x^{eq}, y^{eq}) + \left.\frac{\partial f}{\partial x}\right|_{eq}(x - x^{eq}) + \left.\frac{\partial f}{\partial u}\right|_{eq}(u - u^{eq}) + \mathcal{O}\left(\|(\delta_x, \delta_u)\|\right)$$

$$h(x,u) = h(x^{eq}, y^{eq}) + \left.\frac{\partial h}{\partial x}\right|_{eq}(x - x^{eq}) + \left.\frac{\partial h}{\partial u}\right|_{eq}(u - u^{eq}) + \mathcal{O}\left(\|(\delta_x, \delta_u)\|\right)$$

(2.2)

where the general expression $\frac{\partial g}{\partial z}$ represents the jacobian of the map g respect to the variable vector z. In this expression we indeed that such jacobians evaluated at the equilibrium point are indeed the matrices of the linearized system: $\mathcal{A} = \frac{\partial f}{\partial x}$, $\mathcal{B} = \frac{\partial f}{\partial u}$, $\mathcal{C} = \frac{\partial h}{\partial x}$ and $\mathcal{D} = \frac{\partial h}{\partial u}$. We can prove that the linearity strictly depends on the integral part can be obtained considering that

$$\begin{array}{ll} CT: & \left.\begin{array}{l}\dot{\delta}_x = \dot{x} - \dot{x}^{eq} = f(x,u) \\ \delta_x^+ = x^+ - x^{eq+} = f(x,u) - x^{eq}\end{array}\right\} = f(x,u) - f(x^{eq}, u^{eq}) \\ DT: & \end{array}$$

Using Taylor series we have been able to approximate a non-linear system to an LTI system (for which analytical tool are acknowledged to compute the stability), however such operation can be performed to linearize a system with respect to a generic solution $\left(x^{sol}(t), u^{sol}(t)\right)$ of the non-linear system:

$$\left[\begin{array}{c|c}\mathcal{A}(t) & \mathcal{B}(t) \\ \hline \mathcal{C}(t) & \mathcal{D}(t)\end{array}\right] = \left[\begin{array}{c|c}\partial f/\partial x & \partial f/\partial u \\ \hline \partial h/\partial x & \partial h/\partial u\end{array}\right]\Bigg|_{x^{sol}(t), u^{sol}(t)}$$

Stability   In general the stability of a non-linear system strictly depends on the chosen equilibrium point (is not a system-related property as for linear systems); given a continuous time system characterized by the state equation $\dot{x} = f(x)$ and an equilibrium x^{eq}, then we say that that x^{eq} is **locally exponentially stable** if there exists positive constants $c, \lambda, r > 0$ such that

$$|x(0) - x^{eq}| \le r \qquad \Rightarrow \qquad |x(t) - x^{eq}| \le ce^{-\lambda t}|x(0) - x^{eq}| \qquad \forall t \ge 0$$

A similar definition holds for discrete-time non-linear system where instead we require

$$|x(0) - x^{eq}| \le r \qquad \Rightarrow \qquad |x(t) - x^{eq}| \le c\mu^t|x(0) - x^{eq}| \qquad \forall t \ge 0 \qquad \text{with } c, \mu, r > 0 \text{ and } \mu < 1$$

This mathematical definition express the *start-close stay-close* stability statement; note that for non-linear system this is just a local property, hence there's a bound on the perturbation that results in a solution that reaches the initial equilibrium position.

Theorem   Assuming that $f \in C^2$ is a function differentiable with continuity at least two time, if it's linearization about x^{eq} is exponentially stable then it means that x^{eq} is locally exponentially stable for $\delta x = f(x)$.
   RIVEDERE DIMOSTRAZIONE