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
July 15, 2022

Contents

Chapter 1

Dynamical systems

This course treats the **automatic control** of **dynamical system** hence it's necessary to firstly understand what's a *dynamical system* is indeed and how to describe them. In general we can regard dynamical systems as ***black box*** that, given an **input** $\boldsymbol{u}(t)$, determines and **output** $\boldsymbol{y}(t)$. This signals are function of the time t that can be either discrete ($t \in \mathbb{Z}$), and so we refer them as **discrete-time systems**, or continuous ($t \in \mathbb{R}$), namely **continuous-time** systems. Note that also **hybrid systems** exists (where the system can evolve both in continuous and discrete-time).

The Fibonacci sequence An example of discrete-time system is the **Fibonacci sequence**

$$0, 1, 1, 2, 3, 5, 7, 12, 19 \dots$$

The mathematical law defining such recurrence is a finite difference equation that given the **initial conditions** $y(0)$ and $y(1)$ allows to compute any other value following the rule

$$y(t+1) = y(t-1) + y(t)$$

We can observe that the output y at any time step t is **linearly** dependent from the previous two output (thus we can say that the system is *linear*, as we will define). For this system we can define the vector of the so called **states** $\boldsymbol{x}$ made by the component $x_1(t) = y(t)$ and $x_2(t) = y(t-1)$:

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

Being the system discrete-time we can try to use the states at a given time t (that are known to us) to compute the state at the next time step, hence at $t+1$; from what we have defined we have indeed

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

Recalling the formal definition of the Fibonacci sequence we further have

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

We have now been able to determine the values of the states at time $t+1$ knowing only the states at t; exploiting a matrix notation we can more compactly see that the new states are a linear combination (through a matrix $\mathbf{A}$) of the current states:

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

We observe that given the **initial condition** for the state $\boldsymbol{x}_0 = (1, 0)$ allows to compute the sequence considering at each iteration the second state $y(t) = x_2(t)$.

This recursive definition of the Fibonacci sequence makes hard to determine the generic t-th value (because it requires the knowledge of all the past states), however as we will discover we can

link the eigenvalues of $\mathbf{A}$ (that are $\frac{1\pm\sqrt{5}}{2}$) to a law determining explicitly the output as function of the time as follows:

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

1.1 Classification of dynamical systems

Firstly we distinguished **continuous** from **discrete-time** system based on the the time domain of the variable t, but the Fibonacci sequence example introduced a new classification criteria: *linearity*. We say in fact that a system is **linear** if it's **state-space representation** is in the form

$$\begin{cases} \boldsymbol{x}(t+1) &= \mathbf{A}(t)\boldsymbol{x}(t) + \mathbf{B}(t)\boldsymbol{u}(t) \\ \boldsymbol{y}(t) &= \mathbf{C}(t)\boldsymbol{x}(t) + \mathbf{D}(t)\boldsymbol{u}(t) \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0 \end{cases} \tag{1.1}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are commonly referred as **dynamic matrix**, **input** m., **output** m. and **feed-through** (or instantaneous) m. respectively. As we can see input $\boldsymbol{u} \in \mathbb{R}^m$, states $\boldsymbol{x} \in \mathbb{R}^n$ and outputs $\boldsymbol{y} \in \mathbb{R}^p$ are vectors and all this elements are *interconnected* through linear combination. Usually to simplify the notation the dependency from the time t of $\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{y}$ is dropped and the discrete-time increment $\boldsymbol{x}(t+1)$ is described by the notation $\boldsymbol{x}^+$ (representing so the value of the state at the next state), thus we have

$$\begin{cases} \boldsymbol{x}^+ &= \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u} \\ \boldsymbol{y} &= \mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u} \qquad t \in \mathbb{Z} \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0 \end{cases} \tag{DT-LTI}$$

For continuous-time system (with $t \in \mathbb{R}$) the concept of the time increment is not well defined and so a differential notation is used relating the variation of the states $\dot{\boldsymbol{x}} = \frac{d\boldsymbol{x}}{dt}$:

$$\begin{cases} \dot{\boldsymbol{x}} &= \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u} \\ \boldsymbol{y} &= \mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u} \qquad t \in \mathbb{R} \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0 \end{cases} \tag{CT-LTI}$$

In opposition we have **non-linear systems** for which the relations for both the state variation and output are not linear and are described by generic functions $\boldsymbol{f}, \boldsymbol{h}$:

$$\begin{cases} \dot{\boldsymbol{x}}/\boldsymbol{x}^+ &= \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \\ \boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}) \qquad t \in \mathbb{R}, \mathbb{Z} \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0 \end{cases} \tag{NL-TI}$$

We also observe the absolute importance of specifying in the state-space representation the **initial conditions** $\boldsymbol{x}_0$ of the states: their values are of absolute importance because are storing all the past history of the system; changing the initial condition might lead to completely different behaviour of the system (specially for the non-linear ones).

Up to now in both (DT-LTI) and (CT-LTI), but also in (NL-TI) we removed the time dependence also on the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ (and in the functions $\boldsymbol{f}, \boldsymbol{h}$), but this might not always be the case. Where such matrix are characterized by constant-coefficient (independent from time but also from inputs or states), then the system is said **time-invariant**, thus the acronym **LTI** for *linear time-invariant system* is used. The more general case is characterized by the **time-varying** **linear** systems in the form

$$\begin{cases} \dot{\boldsymbol{x}}/\boldsymbol{x}^+ &= \mathbf{A}(t)\boldsymbol{x} + \mathbf{B}(t)\boldsymbol{u} \\ \boldsymbol{y} &= \mathbf{C}(t)\boldsymbol{x} + \mathbf{D}(t)\boldsymbol{u} \qquad t \in \mathbb{R}, \mathbb{Z} \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0 \end{cases} \tag{LTV}$$

of **time-varying non-linear** systems

$$\begin{cases} \dot{\boldsymbol{x}}/\boldsymbol{x}^+ &= \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},t) \\ \boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{x},\boldsymbol{u},t) \qquad t \in \mathbb{R},\mathbb{Z} \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0 \end{cases} \tag{NL-TV}$$

In general for time-invariant system the definition of the initial time t_0 is not relevant (what matters are just the entries of the vector $\boldsymbol{x}_0$) and for this reason intuitively we chose $t_0 = 0$; the same is not true in general for time-varying system where the choice of t_0 affects the time-evolution of the system (due to the time variance of the dynamical laws).

1.2 Transfer functions for LTI systems

Linear time-invariant (**LTI**) systems can be summarized by the state-space representation

$$\begin{cases} \dot{\boldsymbol{x}}/\boldsymbol{x}^+ &= \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u} \\ \boldsymbol{y} &= \mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u} \qquad t \in \mathbb{R},\mathbb{Z} \\ \boldsymbol{x}(0) &= \boldsymbol{x}_0 \end{cases} \tag{LTI}$$

where the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are constant (in order to respect the time invariance of the system). Continuous-time LTI system can be easily analysed in the domain of the complex variable $s \in \mathbb{C}$ exploiting the **Laplace transform** $\mathscr{L}$, the operator defined as

$$X(s) = \mathscr{L}\{x(t)\} = \int_0^\infty x(t)e^{-st}\,dt \tag{1.2}$$

An important property of such operation is that allows to convert differential equations in t into algebraic ones in the variables s: we in fact have that $\dot{x}(t) \mapsto sX(s) - x(0)$. Applying the Laplace transform to (CT-LTI) determines

$$\begin{cases} s\boldsymbol{X}(s) - \boldsymbol{x}(0) &= \mathbf{A}\boldsymbol{X}(s) + \mathbf{B}\boldsymbol{U}(s) \\ \boldsymbol{Y}(s) &= \mathbf{C}\boldsymbol{X}(s) + \mathbf{D}\boldsymbol{U}(s) \end{cases} \tag{1.3}$$

The first equation can be used to explicitly have the transform of the state $\boldsymbol{X}(s) = \mathscr{L}\{\boldsymbol{x}(t)\}$ as function of the input transform $\boldsymbol{U}(s) = \mathscr{L}\{u(t)\}$:

$$\begin{aligned} s\boldsymbol{X}(s) - \mathbf{A}\boldsymbol{X}(s) &= \mathbf{B}\boldsymbol{U}(s) + \boldsymbol{x}(0) \\ \big(s\mathbf{I} - \mathbf{A}\big)\boldsymbol{X}(s) &= \mathbf{B}\boldsymbol{U}(s) + \boldsymbol{x}(0) \\ \boldsymbol{X}(s) &= \big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B}\boldsymbol{U}(s) + \big(s\mathbf{I} - \mathbf{A}\big)^{-1}\boldsymbol{x}(0) \end{aligned}$$

Putting this result in the second equation of (1.3) allows to explicitly compute the transform of the output $\boldsymbol{Y}(s) = \mathscr{L}\{y(t)\}$ as function of the input transform and the initial state lonely:

$$\boldsymbol{Y}(s) = \underbrace{\Big(\mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B} + \mathbf{D}\Big)}_{\hat{\mathbf{G}}(s)}\boldsymbol{U}(s) + \underbrace{\mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}}_{\hat{\boldsymbol{\psi}}(s)}\boldsymbol{x}(0) \tag{1.4}$$

In this equation $\hat{\mathbf{G}}(s)$ represents the so called **transfer function** of the system, that being precise is a matrix and reduces to a function only for the **SISO** (single-input single-output) case. Observing that the output $\boldsymbol{Y}(s)$ in the Laplace domain can be simplified to the form $\hat{\mathbf{G}}(s)\boldsymbol{U}(s) + \hat{\boldsymbol{\psi}}\boldsymbol{x}(0)$ exploiting the properties of the **inverse Laplace transform** we have that the output in the time domain can be regarded as the convolution of the input with the **impulse response** $\mathbf{G}(t) = \mathscr{L}^{-1}\{\hat{\mathbf{G}}(s)\}$, so

$$\boldsymbol{y}(t) = \big(\mathbf{G} * \boldsymbol{x}\big)(t) + \boldsymbol{\psi}(t)\boldsymbol{x}(0) \tag{1.5}$$

where the convolution is the operation defined as

$$\big(\mathbf{G} * \boldsymbol{x}\big)(t) = \int_0^t \mathbf{G}(t)\boldsymbol{x}(t-\tau)\,d\tau$$

In (1.4) we can also observe the presence of the function $\hat{\boldsymbol{\psi}}(s)$ whose transform $\boldsymbol{\psi}(t)$ represents the **free response** of the system, hence the output evolution determined uniquely by the initial state $\boldsymbol{x}_0$ of the system. Usually this term is neglected because, for simplicity, we assume an initial state $\boldsymbol{x}_0 = 0$ where so the product becomes null.

We can define the concept of **transfer function** also for discrete-time LTI system, but in this case we need to use the **Z transform** $\mathscr{Z}$ mapping discrete-time sequences into a domain of the complex variable $z \in \mathbb{C}$. Having $X(z) = \mathscr{Z}\{x(t)\}$ and knowing the similar property $x^+ \mapsto zX(z) - zx(0)$, applying the transform to (DT-LTI) determines

$$\begin{cases} z\boldsymbol{X}(z) - z\boldsymbol{x}(0) & = \mathbf{A}\boldsymbol{X}(z) + \mathbf{B}\boldsymbol{U}(z) \\ \boldsymbol{Y}(z) & = \mathbf{C}\boldsymbol{X}(z) + \mathbf{D}\boldsymbol{U}(z) \end{cases}$$

Solving the system in order to obtain the transform of the output $\boldsymbol{Y}(z) = \mathscr{Z}\{y(t)\}$ as function of the spectrum $\boldsymbol{U}(z) = \mathscr{Z}\{u(t)\}$ still determines the same transfer function:

$$\boldsymbol{Y}(z) = \underbrace{\Big(\mathbf{C}\big(z\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B} + \mathbf{D}\Big)}_{\hat{\mathbf{G}}(z)}\boldsymbol{U}(z) + \underbrace{\mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}}_{\hat{\boldsymbol{\psi}}(z)}\boldsymbol{x}(0)$$

Theorem 1.1: For continuous-time LTI system its **transfer function** $\hat{\mathbf{G}}(s)$ and the **impulse response** $\mathbf{G}(t)$ can be directly computed as

$$\hat{\mathbf{G}}(s) = \mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B} + \mathbf{D} \qquad \mathbf{G}(t) = \mathscr{L}^{-1}\Big\{\mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B} + \mathbf{D}\Big\} \tag{1.6}$$

1.3 Interconnected systems

THIS PART HAS TO BE REDONE

1.4 Realization theory

Given a **rational polynomial** $\hat{g}(s) = n(s)/d(s)$ representing the transfer function of a SISO (single-input single-output) LTI system we say that it is

- **strictly proper** if the degree of the numerator is leen then the degree of the denominator. This implies $\lim_{s\to\infty} \hat{g}(s) = 0$;
- **proper** if $\#n(s) \leq \#d(s)$ and more specifically $\lim_{s\to\infty} \hat{g}(s) \neq \infty$ (*the limit do not diverge*);
- **improper** if $\#n(s) > \#d(s)$ meaning that $\lim_{s\to\infty} \hat{g}(s) = \pm\infty$.

If it happens that both the rational polynomial $\hat{g}$ and it's inverse $\hat{g}^{-1}$ are proper, then $\hat{g}(s)$ is also said **biproper**.

This definitions can be also extended to a more general system of higher dimension: the **transfer matrix** $\hat{\mathbf{G}}(s)$ is (strictly) proper if and only if all it's entries are (strictly) proper. This means that if just one entry $g_{ij}(s) \in \hat{\mathbf{G}}(s)$ is improper, then $\hat{\mathbf{G}}$ is improper.

It is proven that for LTI system all transfer functions are proper; in particular for every one of them we can recognize a strictly proper part (related to the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$) and a proper part (related only to $\mathbf{D}$) as follows

$$\hat{\mathbf{G}}(s) = \underbrace{\mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B}}_{\text{strictly proper}} + \underbrace{\mathbf{D}}_{\text{proper}} \tag{1.7}$$

Realization Given a transfer function $\hat{\mathbf{G}}(s)$, we say that (LTI) is a **realization** of $\hat{\mathbf{G}}(s)$ if equation (1.7) holds.
This means that an LTI system characterized by the matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ is the realization of $\hat{\mathbf{G}}(s)$ if it holds that $\hat{\mathbf{G}}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$. From this definition we say that two realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ and $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{D}})$ are **zero state equivalent** if they realize the same transfer function $\hat{\mathbf{G}}(s)$. We observe that still this realization might have different free responses $\boldsymbol{\psi}(t) \neq \overline{\boldsymbol{\psi}}(t)$, but starting from a zero state $\boldsymbol{x}_0 = 0$ the behaviour of the output is the same (if the input is the same), as we can see from (1.4).

Theorem 1.2: A transfer function $\hat{\mathbf{G}}(s)$ can be realised by a LTI system in the *standard form* if and only if the rational function $\hat{\mathbf{G}}(s)$ is proper.

Proof 1.1: The proof can be performed in 3 steps:

a) firstly we can show that a realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ determines a proper transfer function: this is a direct consequence of (1.7). In particular if $\mathbf{D} \neq 0$ the transfer function is proper, otherwise if $\mathbf{D} = 0$ then $\hat{\mathbf{G}}(s)$ is strictly proper.

b) determining instead a realization from the transfer function is instead more difficult, but we can now present a possible realization of $\hat{\mathbf{G}}(s)$. To simplify this step we can preliminary compute the non-strictly proper element $\mathbf{D}$ of the transfer function simply with the limit

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

This allows to construct only the strictly-proper part $\hat{\mathbf{G}}_{sp}(s)$ of the transfer function as

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

We can build now the polynomial $d(s)$ as the least common denominator of all the entries in the matrix $\hat{\mathbf{G}}_{sp}(s)$ and we can put him in the form

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

Collecting now all the numerators $n_i(s)$ of the-strictly proper transfer function into a matrix $\mathbf{N}(s)$ and collecting all terms $s^0, \ldots, s^{n-1}$ (being $\hat{\mathbf{G}}_{sp}(s)$ strictly-proper we know that at the numerator the maximum degree of s that can appear is $n-1$) we have

$$\hat{\mathbf{G}}_{sp}(s) = \frac{\mathbf{N}(s)}{d(s)} = \frac{\mathbf{N}_1 s^{n-1} + \mathbf{N}_2 s^{n-2} + \cdots + \mathbf{N}_{n-1} s + \mathbf{N}_n}{s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_{n-1} s + \alpha_n} \tag{†}$$

From this expression we can build the so called **controllable canonical form** defined by the following elements:

$$\left[\begin{array}{cccc|c} -\alpha_1 \mathbf{I}_k & -\alpha_2 \mathbf{I}_{k\times k} & \cdots & -\alpha_n \mathbf{I}_{k\times k} & \mathbf{I}_{k\times k} \\ \mathbf{I}_{k\times k} & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & 0 & \vdots \\ 0 & 0 & \mathbf{I}_{k\times k} & 0 & 0 \\ \hline \mathbf{N}_1 & \mathbf{N}_2 & \cdots & \mathbf{N}_m & \mathbf{D} \end{array}\right] \tag{1.8}$$

c) as last step we can show that (1.8) is indeed a realization of the transfer function. Let us consider the sub-matrix $\mathbf{Z}(s)$ of the transfer function characterized by the vector $\boldsymbol{z}_i$ as

$$\mathbf{Z}(s) = \begin{bmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_n \end{bmatrix} = \big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B}$$

where the components $\mathbf{z}_i$ are considered expanded as rows. By this definition we have that $(s\mathbf{I}-\mathbf{A})\mathbf{Z}(s)=\mathbf{B}$. Considering now the *lower part* (bottom 3 rows) of the matrices $\mathbf{A},\mathbf{B}$ of (1.8), the multiplications evaluates to the following set of equalities:

$$s\mathbf{z}_2-\mathbf{z}_1=0 \qquad s\mathbf{z}_3-\mathbf{z}_2=0 \qquad \dots \qquad s\mathbf{z}_n=\mathbf{z}_{n-1}$$

Observing that it holds $\mathbf{z}_j=\frac{1}{s}\mathbf{z}_{j-1}$ for $j=2,\dots,n$ we can determine the generic vector $\mathbf{z}_k$ as

$$\mathbf{z}_k=\frac{1}{s^{k-1}}\mathbf{z}_1 \tag{$\star$}$$

Considering the top row in (1.8) applied to $(s\mathbf{I}-\mathbf{A})\mathbf{Z}(s)=\mathbf{B}$ determines $(s+\alpha_1)\mathbf{z}_1+\alpha_2\mathbf{z}_2+\cdots+\alpha_n\mathbf{z}_n=\mathbf{I}_{k\times k}$; substituting $(\star)$ determines

$$\left(s+\alpha_1+\frac{\alpha_2}{s}+\cdots+\frac{\alpha_n}{s^{n-1}}\right)\mathbf{z}_1=\frac{d(s)}{s^{n-1}}\mathbf{z}_1=\mathbf{I}_{k\times k}$$

This implies so

$$\mathbf{Z}(s)=\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\\ \vdots\\ \mathbf{z}_n\end{bmatrix}=\frac{1}{d(s)}\begin{bmatrix}s^{n-1}\mathbf{I}_{k\times k}\\ s^{n-2}\mathbf{I}_{k\times k}\\ \vdots\\ \mathbf{I}_{k\times k}\end{bmatrix}$$

From this we can conclude that

$$\mathbf{C}\big(s\mathbf{I}-\mathbf{A}\big)^{-1}=\mathbf{C}\mathbf{Z}(s)=\frac{1}{d(s)}\begin{bmatrix}\mathbf{N}_1 & \mathbf{N}_2 & \dots & \mathbf{N}_n\end{bmatrix}\begin{bmatrix}s^{n-1}\mathbf{I}_{k\times k}\\ s^{n-2}\mathbf{I}_{k\times k}\\ \vdots\\ \mathbf{I}_{k\times k}\end{bmatrix}=\hat{\mathbf{G}}_{sp}(s)$$

The last equation holds because it's equal to (†).

1.4.1 SISO case and multiple realizations: algebraic equivalence

A **single-input single-output** (SISO) **LTI** system is characterized by a transfer function that reduces to a rational polynomial in the form

$$\hat{G}(s)=\frac{\beta_1 s^{n-1}+\beta_2 s^{n-2}+\cdots+\beta_{n-1}s+\beta_n}{s^n+\alpha_1 s^{n-1}+\alpha_2 s^{n-2}+\cdots+\alpha_{n-1}s+\alpha_n} \tag{SISO-LTI}$$

Recalling the **controllable canonical form** in (1.8), a realization of such transfer function is characterized by the matrices

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

This is just one of the infinitely many realization of the same transfer function $\hat{G}(s)$; given a **non-singular** matrix $\mathbf{T}\in\mathbb{R}^{n\times n}$ representing a linear transformation, we can define a new vector of states $\tilde{\boldsymbol{x}}=\mathbf{T}\boldsymbol{x}$ (hence $\boldsymbol{x}=\mathbf{T}^{-1}\tilde{\boldsymbol{x}}$): this allow to describe in a new *coordinate system* the state-space representation of the system. We can in fact consider

$$\begin{cases}\dot{\tilde{\boldsymbol{x}}}=\mathbf{T}\dot{\boldsymbol{x}} &= \mathbf{T}\mathbf{A}\boldsymbol{x}+\mathbf{T}\mathbf{B}\boldsymbol{u}=\mathbf{T}\mathbf{A}\mathbf{T}^{-1}\tilde{\boldsymbol{x}}+\mathbf{T}\mathbf{B}\boldsymbol{u}=\tilde{\mathbf{A}}\tilde{\boldsymbol{x}}+\tilde{\mathbf{B}}\boldsymbol{u}\\ \tilde{\boldsymbol{y}} &= \mathbf{C}\boldsymbol{x}+\mathbf{D}\boldsymbol{u}=\mathbf{C}\mathbf{T}^{-1}\tilde{\boldsymbol{x}}+\mathbf{D}\boldsymbol{u}=\tilde{\mathbf{C}}\tilde{\boldsymbol{x}}+\mathbf{D}\boldsymbol{u}\end{cases}$$

More compactly we can state that the following systems

$$\left[\begin{array}{c|c}\tilde{\mathbf{A}} & \tilde{\mathbf{B}}\\ \hline \tilde{\mathbf{C}} & \tilde{\mathbf{D}}\end{array}\right]=\left[\begin{array}{c|c}\mathbf{T}\mathbf{A}\mathbf{T}^{-1} & \mathbf{T}\mathbf{B}\\ \hline \mathbf{C}\mathbf{T}^{-1} & \tilde{\mathbf{D}}\end{array}\right] \qquad \text{and} \qquad \left[\begin{array}{c|c}\mathbf{A} & \mathbf{B}\\ \hline \mathbf{C} & \mathbf{D}\end{array}\right] \tag{1.9}$$

are **algebraically equivalent**, meaning that they are representing the same transfer function $\hat{G}(s)$. In this case the non-singular matrix **T** is referred as **similarity** (or equivalence) **transformation** and allows to describe an initial 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 computation. We also observe that this equivalence hold for the generic case of **MIMO** (multi-input multi-output) systems.

An example of **algebraically equivalent representation** is the **observable canonical form** that for (SISO-LTI) is characterized by the state-space representation in the form

$$\left[\begin{array}{cccc|c} -\alpha_1 & 1 & & 0 & \beta_1 \\ \vdots & & \ddots & & \vdots \\ -\alpha_{n-1} & 0 & & 1 & \beta_{n-1} \\ -\alpha_n & 0 & \dots & 0 & \beta_n \\ \hline 1 & 0 & \dots & 0 & \end{array}\right] \tag{1.10}$$

Chapter 2

Solution of linear systems

2.1 Linear time-varying systems

Continuous-time LTV systems

Given an **autonomous continuous-time LTV** system characterized by the state-space representation

$$\begin{cases} \dot{\boldsymbol{x}} & = \mathbf{A}(t)\boldsymbol{x} \\ \boldsymbol{x}(t_0) & = \boldsymbol{x}_0 \end{cases} \tag{ACT-LTV}$$

the solution to such problem is a differentiable signal $\boldsymbol{x}(t)$. This function can be computed by firstly defining the **Peano-Baker series**, a matrix $\boldsymbol{\Phi}(t, t_0) \in \mathbb{R}^{n \times n}$ defined as

$$\begin{aligned} \boldsymbol{\Phi}(t, t_0) = \mathbf{I} + \int_{t_0}^{t} \mathbf{A}(s_1)\, ds_1 + \int_{t_0}^{t} \mathbf{A}(s_1) \int_{t_0}^{s_1} \mathbf{A}(s_2)\, ds_2\, ds_1 + \\ + \int_{t_0}^{t} \mathbf{A}(s_1) \int_{t_0}^{s_1} \mathbf{A}(s_2) \int_{t_0}^{s_2} \mathbf{A}(s_3)\, ds_3\, ds_2\, ds_1 + \dots \end{aligned} \tag{2.1}$$

The matrix $\boldsymbol{\Phi}$ is mostly referred as the **state transition matrix** and is characterized by the following properties:

i) $\phi(t, t) = \mathbf{I}$ for any $t \in \mathbb{R}$. This results is a direct consequence of the definition of the state transition matrix: knowing that $\int_a^a f(x)\, dx$ always evaluates to zero, then all terms (except the first identity matrix) are evaluating to zero;

ii) $\frac{d}{dt}\boldsymbol{\Phi}(t, t_0) = \mathbf{A}(t)\boldsymbol{\Phi}(t, t_0)$. This comes from the integral property stating that $\frac{d}{dt}\int_a^t f(\tau)\, d\tau = f(\tau)$: applying this property to (2.1) determines

$$\begin{aligned} \frac{d}{dt}\boldsymbol{\Phi}(t, t_0) &= 0 + \mathbf{A}(t) + \mathbf{A}(t) \int_{t_0}^{t} \mathbf{A}(s_2)\, ds_2 + \mathbf{A}(t) \int_{t_0}^{t} \mathbf{A}(s_2) \int_{t_0}^{s_2} \mathbf{A}(s_3)\, ds_3\, ds_2 + \dots \\ &= \mathbf{A}(t)\boldsymbol{\Phi}(t, t_0) \end{aligned}$$

This property is relevant because it allows-us to explicitly compute the **zero-input response** solution of (ACT-LTV) as

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 \quad \forall t \tag{2.2}$$

This solution is proven to be the unique solving the problem $\dot{\boldsymbol{x}} = \mathbf{A}(t)\boldsymbol{x}$ (however no proof for the uniqueness is reported here and must be taken for granted). To show that (2.2) is indeed a solution of (ACT-LTV) we can consider property *i)* to see that $\boldsymbol{\Phi}(t_0, t_0)\boldsymbol{x}_0 = \mathbf{I}\boldsymbol{x}_0 = \boldsymbol{x}_0$ (satisfying so the initial condition for the states); exploiting property *ii)* we also observe that also the dynamics of the system is satisfied, in fact:

$$\dot{\boldsymbol{x}}(t) = \frac{d}{dt}\left(\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0\right) = \mathbf{A}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 = \mathbf{A}(t)\boldsymbol{x}(t)$$

If we want now to expand the definition of the solution also for a generic LTV system (characterized so by a non-zero input $\boldsymbol{u}$) it's mandatory to define other 2 properties of the state transition matrix:

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

$$\dot{\boldsymbol{x}} = \mathbf{A}(t)\boldsymbol{x}$$

where $\boldsymbol{x}(t_0) = \boldsymbol{e}_i$, meaning that the initial condition is the i-th component of the canonical basis of $\mathbb{R}^n$. This property is useful for practical computation because it allows us to decompose the study of the solution $x(t)$ by independently solving the states; after this process we can use linearity to obtain more solutions to the LTV systems;

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

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

We can in fact consider that, given the initial state x_0 at the time instant t_0, the states $\boldsymbol{x}_1$ in a second time t_1 can be regarded as $\boldsymbol{x}(t_1) = \mathbf{\Phi}(t_1, t_0)\boldsymbol{x}_0$. Considering now another time t_2 we moreover have

$$\boldsymbol{x}(t_2) = \mathbf{\Phi}(t_2, t_1)\boldsymbol{x}(t_1) = \mathbf{\Phi}(t_2, t_1)\mathbf{\Phi}(t_1, t_0)\boldsymbol{x}_0 = \mathbf{\Phi}(t_2, t_0)\boldsymbol{x}_0$$

Note that the statement of the semi-group property do not set any requirement in the *relative position* of the times, meaning that is not necessary to have $t > s > \tau$;

v) direct consequence of the semi-group property is that $\mathbf{\Phi}(t, s)$ is always invertible (is never singular) with inverse equal to $\mathbf{\Phi}(s, t)$. Combining property *iv)* with *i)* shows that

$$\mathbf{\Phi}(t, s)\mathbf{\Phi}(s, t) = \mathbf{\Phi}(t, t) = \mathbf{I} \quad \forall t, s \in \mathbb{R} \qquad \Leftrightarrow \qquad \mathbf{\Phi}^{-1}(t, s) = \mathbf{\Phi}(s, t)$$

Theorem 2.1: Considering a general continuous-time LTV system characterized by the state-space representation

$$\begin{cases} \dot{\boldsymbol{x}} & = \mathbf{A}(t)\boldsymbol{x} + \mathbf{B}(t)\boldsymbol{u} \\ \boldsymbol{y} & = \mathbf{C}(t)\boldsymbol{x} + \mathbf{D}(t)\boldsymbol{u} \\ \boldsymbol{x}(t_0) & = \boldsymbol{x}_0 \end{cases} \tag{CT-LTV}$$

the general solution, proven to be unique, for this problem is determined by the **variation of constant formula** defined as

$$\begin{aligned} \boldsymbol{x}(t) &= \overbrace{\mathbf{\Phi}(t, t_0)\boldsymbol{x}_0}^{\text{zero-input res.}} + \overbrace{\int_{t_0}^{t} \mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)\boldsymbol{u}(\tau)\, d\tau}^{\text{zero-state response}} \\ \boldsymbol{y}(t) &= \underbrace{\mathbf{C}(t)\mathbf{\Phi}(t, t_0)\boldsymbol{x}_0}_{\text{homogeneous res.}} + \underbrace{\int_{t_0}^{t} \mathbf{C}(\tau)\mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)\boldsymbol{u}(\tau)\, d\tau + \mathbf{D}(t)\boldsymbol{u}(t)}_{\text{forced response}} \end{aligned} \tag{2.3}$$

Proof 2.1: To prove this theorem we have to exploit the differential property stating that $\frac{d}{dt}\int_{t_0}^{t} f(t, \tau)\, d\tau = f(t, t) + \int_{t_0}^{t} \frac{d}{dt} f(t, \tau)\, d\tau$: considering the state solution $\boldsymbol{x}(t)$ of (2.3) we can show that it's differentiation in time satisfies the dynamic of (CT-LTV):

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \frac{d}{dt}\left(\mathbf{\Phi}(t, t_0)\boldsymbol{x}_0 + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)u(\tau)\, d\tau\right) \\ &= \mathbf{A}(t)\mathbf{\Phi}(t, t_0)\boldsymbol{x}_0 + \mathbf{\Phi}(t, t)\mathbf{B}(t)\boldsymbol{u}(t) + \int_{t_0}^{t} \frac{d\mathbf{\Phi}(t, \tau)}{dt}\mathbf{B}(\tau)\boldsymbol{u}(\tau)\, d\tau \\ &= \mathbf{A}(t)\mathbf{\Phi}(t, t_0)\boldsymbol{x}_0 + \mathbf{I}\mathbf{B}(t)\boldsymbol{u}(t) + \int_{t_0}^{t} \mathbf{A}(t)\mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)\boldsymbol{u}(\tau)\, d\tau \\ &= \mathbf{A}(t)\underbrace{\left(\mathbf{\Phi}(t, t_0)\boldsymbol{x}_0 + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)u(\tau)\, d\tau\right)}_{=\boldsymbol{x}(t)} + \mathbf{B}(t)\boldsymbol{u}(t) \end{aligned}$$

Discrete-time LTV systems

Given an **autonomous discrete-time LTV** system characterized by the state-space representation

$$\begin{cases} \boldsymbol{x}^+ &= \mathbf{A}(t)\boldsymbol{x} \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0 \end{cases} \tag{ADT-LTV}$$

the homogenous response of this system, similarly for the continuous-time counterpart, is determined by the equation

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 \qquad \forall t \geq t_0 \tag{2.4}$$

The main difference with respect to (2.2) is that in this case the definition is correct only for $t \geq t_0$. This is due to the new definition of the **state transition matrix** defined as

$$\boldsymbol{\Phi}(t, t_0) = \begin{cases} \mathbf{I} & \text{if } t = t_0 \\ \mathbf{A}(t-1)\mathbf{A}(t-2)\dots\mathbf{A}(t_0+1)\mathbf{A}(t_0) & \text{if } t > t_0 \end{cases} \tag{2.5}$$

Intuitively we can think that the states at $t_1 = t_0 + 1$ can be computed as $\boldsymbol{x}(t_0+1) = \mathbf{A}(t_0)\boldsymbol{x}_0$; moreover $\boldsymbol{x}(t_0+2) = \mathbf{A}(t_0+1)\boldsymbol{x}(t_1) = \mathbf{A}(t_0+2)\mathbf{A}(t_0+1)\boldsymbol{x}_0$: iterating this process allows us to directly obtain the definition of the state transition matrix in (2.5).

If we want now to extend the definition of $\boldsymbol{\Phi}$ for times $t < t_0$ we have to consider that (given $t+1 = t_0$) it must hold $\boldsymbol{x}(t_0) = x(t+1) = \mathbf{A}(t)x(t)$, hence

$$\boldsymbol{x}(t) = \mathbf{A}^{-1}(t_0 - 1)\boldsymbol{x}_0$$

This operation however is feasible if and only if the matrix $\mathbf{A}(t_0)$ is non-singular (allowing its inversion), however this condition isn't satisfied *a-priori*.

Addition of the input Having defined the state transition matrix for discrete-time system and the homogeneous solution for the autonomous case, we can consider the general case of a **discrete-time LTV** system whose solution is characterized by the following **variation of constants** formula:

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 + \sum_{\tau=0}^{t-1} \boldsymbol{\Phi}(t, \tau+1)\mathbf{B}(\tau)\boldsymbol{u}(\tau) \\ \boldsymbol{y}(t) &= \mathbf{C}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 + \sum_{\tau=0}^{t-1} \mathbf{C}(\tau)\boldsymbol{\Phi}(t, \tau+1)\mathbf{B}(\tau)\boldsymbol{u}(\tau) + \mathbf{D}(t)\boldsymbol{u}(t) \end{aligned} \tag{2.6}$$

2.2 Linear time-invariant systems

Discrete-time LTI systems

Considering a (DT-LTI) system (page 2), the time invariance determines that all matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ present constant coefficients (not depending on time): this allows to simplify the definition of the Peano-Baker series (2.5) into the form

$$\boldsymbol{\Phi}(t, t_0) = \mathbf{A}^{t-t_0} \qquad \forall t \geq t_0 \tag{2.7}$$

where conventionally is chosen $\mathbf{A}^0 = \mathbf{I}$. This allows to simplify also the **variation of constants formula** having the solutions of (DT-LTI) as

$$\begin{aligned} \boldsymbol{x}(t) &= \mathbf{A}^{t-t_0}\boldsymbol{x}_0 + \sum_{\tau=0}^{t-1} \mathbf{A}^{t-\tau-1}\mathbf{B}(\tau)\boldsymbol{u}(\tau) \\ \boldsymbol{y}(t) &= \mathbf{C}\mathbf{A}^{t-t_0}\boldsymbol{x}_0 + \sum_{\tau=0}^{t-1} \mathbf{C}(\tau)\mathbf{A}^{t-\tau-1}\mathbf{B}(\tau)\boldsymbol{u}(\tau) + \mathbf{D}\boldsymbol{u} \end{aligned} \tag{2.8}$$

We usually refer to the operation $\mathbf{A}^t$ (with $t \in \mathbb{Z}$) as the **matrix power** whose computation will be explained properly in page 13.

Continuous-time LTI systems

Considering now a (CT-LTI) system, the definition of the Peano-Baker series (2.1) can also be simplified considering that all matrices have constant coefficient; in particular we can observe that

$$\begin{aligned}\mathbf{\Phi}(t,t_0) &= \mathbf{I} + \mathbf{A}(t-t_0) + \mathbf{A}^2\frac{(t-t_0)^2}{2} + \mathbf{A}^3\frac{(t-t_0)^3}{2\cdot 3} + \dots \\ &= \sum_{k=0}^{\infty} \mathbf{A}^k \frac{(t-t_0)^k}{k!}\end{aligned}$$

One can so find a strict similarity with the Taylor series expansion of an exponential (that's indeed $e^x = \sum_{k=0}^{\infty} \frac{1}{k!}x^k$): for this reason for continuous-time LTI system the state transition matrix is represented by the **exponential matrix** as

$$\mathbf{\Phi}(t,t_0) = e^{\mathbf{A}(t-t_0)} \tag{2.9}$$

This simplifies the definition of the unique solution of (CT-LTI) to the form

$$\begin{aligned}\boldsymbol{x}(t) &= e^{\mathbf{A}(t-t_0)}\boldsymbol{x}_0 + \int_{t_0}^{t} e^{\mathbf{A}(t-\tau)}\mathbf{B}\boldsymbol{u}(\tau)\,d\tau \\ \boldsymbol{y}(t) &= \mathbf{C}e^{\mathbf{A}(t-t_0)}\boldsymbol{x}_0 + \int_{t_0}^{t} \mathbf{C}e^{\mathbf{A}(t-\tau)}\mathbf{B}\boldsymbol{u}(\tau)\,d\tau + \mathbf{D}\boldsymbol{u}(t)\end{aligned} \tag{2.10}$$

Properties of the matrix exponential Relevant properties of the matrix exponential are:

i) each i-th column of the matrix $e^{\mathbf{A}t}$ is the unique solution of the problem $\dot{\boldsymbol{x}}(t) = \mathbf{A}\boldsymbol{x}(t)$ where the initial condition is set to the i-th component of the canonical base $\boldsymbol{x}(0) = \boldsymbol{e}_i$ (recalling property *iii)* of the Peano-Baker series at page 8);

ii) because it still holds the semi-group property $\mathbf{\Phi}(t,\tau)\mathbf{\Phi}(\tau,s) = \mathbf{\Phi}(t,s)$ for all t,s,τ, then it means that

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

iii) the matrix exponential is always invertible, in fact still for the semi-group property $e^{\mathbf{A}t}e^{-\mathbf{A}t} = e^{\mathbf{A}0} = \mathbf{I}$; moreover we have that the inverse of the exponential matrix is

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

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

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

v) given $\mathbf{A} \in \mathbb{R}^{n\times n}$ there exists n function $\alpha_i(t)$ such that

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

Theorem 2.2: For each matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with characteristic polynomial

$$p_{\mathbf{A}}(s) = \det\left(s\mathbf{I} - \mathbf{A}\right) = s^n + a_1 s^{n-1} + \dots + a_{n-1}s + a_n$$

then $\mathbf{A}$ is a matrix that's a solution of such polynomial, meaning that

$$p_{\mathbf{A}}(\mathbf{A}) = \mathbf{A}^n + a_1\mathbf{A}^{n-1} + \dots + a_{n-1}\mathbf{A} + a_n\mathbf{I} = 0_{n\times n}$$

This is referred as the **Cayley-Hamilton theorem**.

A direct consequence from this theorem (that it will be used also for later proofs) is that by reverting the last equality we can state the power $\mathbf{A}^n$ as a linear combination of *lower-order powers*:

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

Multiplying both sides by $\mathbf{A}$ allows also to compute the power $\mathbf{A}^{n+1}$ as a linear combination of all powers of $\mathbf{A}$ up to the order $n-1$, in fact

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

This principle can be extended to any power matrix of $\mathbf{A}$ and each one of them can be regarded as a particular linear combination of the first n matrix powers of $\mathbf{A}$:

$$\mathbf{A}^h = \sum_{k=0}^{n-1} c_{h,k}\mathbf{A}^k \qquad \forall h \geq n \tag{2.11}$$

2.2.1 Recall on the Jordan normal form

Equation (1.9), page 6, showed how system can be rewritten in algebraically equivalent form by means of similarity transformation regarded as a non-singular matrix $\mathbf{T}$. Ideally a *good* transformation, one that can determine a simpler state-space representation, is the one that put the dynamic matrix in a diagonal form

$$\overline{\mathbf{A}} = \mathbf{T}\mathbf{A}\mathbf{T}^{-1} = \begin{bmatrix}\lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix} \tag{2.12}$$

This in fact allows to have dynamic equations where the variation of the states are related only to themselves, so in the form $\dot{x}_i / x_i^+ = \lambda_i x_i$ for any i. Moreover the elements λ_i in the diagonal matrix $\overline{\mathbf{A}}$ are the **eigenvalues** of the $\mathbf{A}$ (this values are always preserved through any similarity transformation).

Computing the required matrix $\mathbf{T}$ that satisfies (2.12) requires solving the eigenvector-eigenvalues problem: while doing so what might happen is that the algebraic multiplicity of some eigenvalues are bigger then the associated geometric multiplicity. As direct consequence of this, no matrix $\mathbf{T}$ exists that diagonalizes $\mathbf{A}$ into $\overline{\mathbf{A}}$.

Acknowledged that (2.12) might not hold in general, a solution to our problem is provided by **Jordan** stating:

Theorem 2.3: For each matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ there always exists a non-singular (hence invertible) matrix $\mathbf{T} \in \mathbb{C}^{n\times n}$ such that

$$\boldsymbol{J} = \mathbf{T}\mathbf{A}\mathbf{T}^{-1} = \begin{bmatrix}\mathbf{J}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{J}_l\end{bmatrix} \in \mathbb{C}^{n\times n} \tag{2.13}$$

where the matrices $\mathbf{J}_i$ are in the form

$$\mathbf{J}_i = \begin{bmatrix}\lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i\end{bmatrix} \in \mathcal{C}^{n_i\times n_i} \tag{2.14}$$

with n_i the algebraic multiplicity of the i-th eigenvalue

Matrix $\boldsymbol{J}$ in (2.13) is known as the **Jordan normal form** of the matrix $\mathbf{A}$ and is proven to be unique (upon rearrangements of the sub-matrices $\mathbf{J}_i$). Note that in general sub-matrices $\mathbf{J}_i, \mathbf{J}_j$ (with $i \neq j$) might be constructed with the same eigenvalue λ_k but what might change are their dimension $n_i \neq n_j$ (while still the sum of all $n_i, n_j, \dots$ associated to a specific eigenvalue must add up to its algebraic multiplicity).
The number of *blocks* $\mathbf{J}_i$ in the Jordan normal form are $l \leq n$ and are associated to each linearly independent eigenvector of $\mathbf{A}$ (is proven in fact that at least one eigenvector exists for each matrix $\mathbf{A}$ and that can exits at maximum n linearly independent ones).

From this, a square matrix is called **semi-simple**, or **diagonalizable**, if its Jordan normal form (2.13) is diagonal as in (2.12). In particular given $\mathbf{A} \in \mathbb{C}^{n \times n}$, the following 3 statements are equivalent:

i) $\mathbf{A}$ is semi-simple;

ii) $\mathbf{A}$ has n linearly independent eigenvectors;

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

2.2.2 Matrix power

With the definition of the Jordan normal form the computation of the matrix power of $\mathbf{A}$ is quite straightforward. Reverting (2.13) gives that $\mathbf{A} = \mathbf{T}^{-1}\boldsymbol{J}\mathbf{T}$; expanding so the matrix power what we see is that

$$\mathbf{A}^t = \underbrace{\mathbf{A}\mathbf{A}\dots\mathbf{A}}_{t \text{ times}} = \underbrace{\mathbf{T}^{-1}\boldsymbol{J}\mathbf{T}\mathbf{T}^{-1}\boldsymbol{J}\mathbf{T}\dots\mathbf{T}^{-1}\boldsymbol{J}\mathbf{T}}_{t \text{ times}}$$

By definition of matrix inverse we have that all the products $\mathbf{T}\mathbf{T}^{-1}$ *in the middle* are evaluating to the identity matrix (hence they can be *forgotten* in the matrix multiplication), so what remains is the equation that we can exploit for the computation of the **matrix power**:

$$\mathbf{A}^t = \mathbf{T}^{-1}\boldsymbol{J}^t\mathbf{T} \tag{2.15}$$

In particular it's known that the matrix power of the Jordan normal form evaluates to

$$\boldsymbol{J}^t = \begin{bmatrix} \mathbf{J}_1^t & & 0 \\ & \ddots & \\ 0 & & \mathbf{J}_l^t \end{bmatrix} \tag{2.16}$$

where

$$\mathbf{J}_i^t = \begin{bmatrix} \lambda_i^t & t\lambda_i^{t-1} & \frac{t!}{(t-2)!2!}\lambda_i^{t-2} & \dots & \frac{t!}{(t-n_i+1)!(n_i-1)!}\lambda_i^{t-n_i+1} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{t!}{(t-2)!2!}\lambda_i^{t-2} \\ & & & \ddots & t\lambda_i^{t-1} \\ 0 & & & & \lambda_i^t \end{bmatrix} \tag{2.17}$$

This is the general definition of the matrix $\mathbf{J}_i^t$, however we observe that the exponentials $k\lambda_i^{t-\tau}$ are *activated* only when the coefficient of the exponent is greater or equal to zero, in the sense that

$$\mathbf{J}_i^t = \begin{bmatrix} \lambda_i^t & & 0 \\ & \ddots & \\ 0 & & \lambda_i^t \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \quad \text{for } t = 0$$

$$\mathbf{J}_i^t = \begin{bmatrix} \lambda_i^t & t\lambda_i^{t-1} & & 0 \\ & \ddots & \ddots & \\ & & \ddots & t\lambda_i^{t-1} \\ 0 & & & \lambda_i^t \end{bmatrix} = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix} \quad \text{for } t = 1 \qquad \dots$$

Region of convergence Let us consider now the discrete-time autonomous system whose solution is the sequence

$$\boldsymbol{x}(t) = \mathbf{A}^t \boldsymbol{x}_0$$

In order to have a **convergent solution** we want to ensure that $\lim_{t\to\infty} \mathbf{A}^t = 0$, or that at least that such limit do not diverge; from a practical point of view what we would like is that all entries in $\mathbf{A}^t$, with $t \to \infty$, are zero (or non-diverging). Considering that the matrix power $\mathbf{A}^t$ can be regarded as the products $\mathbf{T}^{-1}\boldsymbol{J}\mathbf{T}$, then such limit is intrinsically transmitted to the Jordan normal form matrix $\boldsymbol{J}$. In particular to have the **convergence** we want

$$\lim_{t\to\infty} \boldsymbol{J}^t = 0 \qquad \Leftrightarrow \qquad \lim_{t\to\infty} \mathbf{J}_i^t = 0 \ \forall i$$

Considering the formal definition of $\mathbf{J}_i^t$ in (2.17) we can observe that each entry in the matrix is characterized by a power of the eigenvalue λ_i (with an exponent dependent from time) multiplied by a polynomial in t. Knowing that the exponential function asymptotically grows (or decreases) faster then the polynomial term we observe that whenever the magnitude of the eigenvalue is less then zero, then the sub-matrix $\mathbf{J}_i^t$ converges to zero:

$$\lim_{t\to\infty} \mathbf{J}_i^t = 0 \qquad \Leftrightarrow \qquad |\lambda_i| < 1 \tag{2.18}$$

In contrary, considering $|\lambda_i| > 1$ determines a diverging power λ_i^t for $t \to \infty$.
One last condition that we have to check is whenever $|\lambda_i| = 1$: in this case the power remains constant (indeed $1^t \xrightarrow{t\to\infty} 1$) and so the behaviour of the solution is ruled by the polynomial terms. In this case if we impose that $n_i = 1$ then there are no polynomial terms out of the diagonal and $\mathbf{J}_i^t$ approaches the identity matrix, while in all other cases the entries are diverging.

We can summarize the **convergence** of the system (requiring so that $\mathbf{A}$ *do not explode to infinity*) considering the following implications:

$$\lim_{t\to\infty} \mathbf{A}^t \neq \infty \qquad \Leftrightarrow \qquad \lim_{t\to\infty} \mathbf{J}_i^t \neq \infty \ \forall i \qquad \Leftrightarrow \qquad |\lambda_i| \leq 1 \text{ and if } |\lambda_i| = 1 \ \Rightarrow \ n_i = 1 \tag{2.19}$$

Fibonacci sequence Recalling the very first example in this book, the Fibonacci sequence, we determined it's state-space representation that was in the form

$$\begin{cases} \boldsymbol{x}^+ &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} &= \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}(0) &= (0,1) \end{cases} \qquad \text{with } \boldsymbol{x} \in \mathbb{R}^2$$

In this case the characteristic polynomial of $\mathbf{A}$ evaluates to $p_{\mathbf{A}}(s) = s^2 - s - 1$ determining two distinct eigenvalues $s_{1,2} = \frac{1\pm\sqrt{5}}{2}$. Having two distinct eigenvalues (and being $\mathbf{A}$ a 2×2 matrix) the only Jordan normal form allowable is the semi-simple one that's

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

In order to compute the matrix power $\mathbf{A}^t$ (used to determine the solution of the discrete-time system) we need to determine the transformation matrix $\mathbf{T}$ satisfying (2.13): this can be achieved by solving the linear problem $\mathbf{TA} = \boldsymbol{J}\mathbf{T}$, hence

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

Among all the possible solutions to this problem we can chose the matrix

$$\mathbf{T} = \begin{bmatrix} 1+\sqrt{5} & 2 \\ 1-\sqrt{5} & 2 \end{bmatrix}$$

whose inverse is

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

Exploiting now (2.15) we have that the matrix power of $\mathbf{A}^t$ is

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

Finally computing the solution of the output $y(t)$ of the system as $\mathbf{C}\mathbf{A}^t\boldsymbol{x}_0$ gives the same result presented at the beginning of this book:

$$\begin{aligned} y(t) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1^{t+1} - \lambda_2^{t+1} & \lambda_1^t - \lambda_2^t \\ \lambda_1\lambda_2^{t+1} - \lambda_1^{t+1}\lambda_2 & \lambda_1\lambda_2^t - \lambda_1^t\lambda_2 \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}}(\lambda_1^t - \lambda_2^t) \\ &= \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^t - \left(\frac{1-\sqrt{5}}{2} \right)^t \right] \end{aligned}$$

2.2.3 Matrix exponential

Similarly for what has been done for the matrix power, we will show that's possible to link the matrix exponential of $\mathbf{A}$ with the exponential of it's Jordan normal form $\boldsymbol{J}$. In particular we define the **exponential of the Jordan normal form** as

$$e^{\boldsymbol{J}t} = \begin{bmatrix} e^{\mathbf{J}_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\mathbf{J}_l t} \end{bmatrix} \tag{2.20}$$

where

$$e^{\mathbf{J}_i t} = e^{\lambda_i t} \begin{bmatrix} 1 & t & \dots & \frac{t^{n_i-1}}{(n_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & t \\ 0 & & & 1 \end{bmatrix} \tag{2.21}$$

In order now to compute the **matrix exponential** of $\mathbf{A}$ we have to use the Cayley-Hamilton theorem 2.3 (page 12) that allows us to consider the exponential as a linear combination of the first n powers of $\mathbf{A}$. Considering indeed $\mathbf{A} = \mathbf{T}^{-1}\boldsymbol{J}\mathbf{T}$ we have that

$$\begin{aligned} e^{\mathbf{A}t} &= \sum_{k=0}^{n-1} \alpha_k(t)\mathbf{A}^k = \sum_{k=0}^{n-1} \alpha_k(t)\mathbf{T}^{-1}\boldsymbol{J}^k\mathbf{T} = \mathbf{T}^{-1} \left(\sum_{k=0}^{n-1} \boldsymbol{J}^k \right) \mathbf{T} \\ &= \mathbf{T}^{-1}e^{\boldsymbol{J}t}\mathbf{T} \end{aligned} \tag{2.22}$$

Region of convergence Also in this case is important to define the **region of convergence** as the set of eigenvalues λ_i for which the exponential $e^{\mathbf{A}t}$ converges to zero (or at least it doesn't diverge). Transferring the property of converge to the exponential $e^{\boldsymbol{J}t}$, we observe that whenever $\mathbb{R}\mathrm{e}\{\lambda_i\} < 0$ the sub-block $e^{\mathbf{J}_i t}$ converges to zero, while if $\mathbb{R}\mathrm{e}\{\lambda_i\}$ we have the divergence to infinity. Moreover in the case of $\mathbb{R}\mathrm{e}\{\lambda_i\} = 0$ we have that the exponential remains constant and so in order not to have a polynomial divergence we must ensure that $n_i = 1$.

As for the matrix power we can summarize the **convergence of the matrix exponential** as

$$\lim_{t\to\infty} e^{\mathbf{A}t} \neq \infty \quad \Leftrightarrow \quad \lim_{t\to\infty} e^{\mathbf{J}_i t}\ \forall i \quad \Leftrightarrow \quad \mathbb{R}\mathrm{e}\{\lambda_i\} \leq 0 \text{ and if } \mathbb{R}\mathrm{e}\{\lambda_i\} = 0 \ \Rightarrow\ n_i = 1 \tag{2.23}$$

Chapter 3

Stability in Lyapunov sense

Norms for vector This chapter will concentrate to the definition of ***Lyapunov stability*** for linear system, however prior to do so a recall on the concept of *norm* for both vectors and matrices should be done.

Given a vector $\boldsymbol{x} \in \mathbb{R}^n$ we define it's **norm** $\boldsymbol{x} \mapsto |\boldsymbol{x}|$ an operation satisfying the following properties:

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

Satisfying this conditions a various amount of norms have been defined, such

- the ***one norm*** $|\cdot|_1$ defined as
$$|\boldsymbol{x}|_1 = \sum_{i=1}^{n} |x_i| \tag{3.1}$$
- the common ***two norm*** $|\cdot|_2$, also referred as *euclidean norm*, defined as
$$|\boldsymbol{x}|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \tag{3.2}$$
- the ***infinity norm*** $|\cdot|_\infty$ defined as
$$|\boldsymbol{x}|_\infty = \max_{i=1,\dots,n} |x_i| \tag{3.3}$$
- the generalized concept of ***p norm*** $|\cdot|_p$ defined as
$$|\boldsymbol{x}|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \qquad \text{with } p \in \mathbb{R}^+ \tag{3.4}$$

All this **norms** are **equivalent**: given in fact any two norms $p_1, p_2 \in [1, \infty)$, then where always exists two constants $c_1, c_2 \in \mathbb{R}$ such that

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

Norms for matrices The concept of norm as a *measure of the dimension* can also be extended also for matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$. Considering similar properties to the one of vector norms, example of **norms** of **matrices** $\|\mathbf{A}\|$ are:

- the ***one norm*** $\|\cdot\|_1$ defined as
$$\|\mathbf{A}\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |a_{ij}| \tag{3.5}$$
This can be regarded as the maximum norm of the row vectors composing the matrix $\mathbf{A}$;

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

$$\|\mathbf{A}\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |a_{ij}| \tag{3.6}$$

  This can be regarded as the maximum norm of the column vectors composing the matrix $\mathbf{A}$;

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

$$\|\mathbf{A}\|_2 = \sigma_{\max}\{\mathbf{A}\} = \sqrt{\lambda_{\max}\{\mathbf{A}^T\mathbf{A}\}} \tag{3.7}$$

  We denote in general with $\sigma\{\mathbf{A}\}$ the *singular values* of the matrix $\mathbf{A}$ and such terms are computed (as shown) as the square root of the eigenvalues of the square matrix $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$; in particular the two norm corresponds to the maximum singular value of $\mathbf{A}$. Considering that a vector can be regarded as a $m \times 1$ matrix, then this definition coincides with the euclidean norm (3.2);

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

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{\sum_{i=1}^{n} \sigma_i^2\{\mathbf{A}\}} \tag{3.8}$$

  Also in this case regarding a vector as a $m \times 1$ matrix the output of the Frobenius norm coincides with the euclidean norm (3.2), however in the general case we have that $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$.

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

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

Moreover norms are also **sub-multiplicative**, meaning so

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

Induced norm and subordination Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and a vector $\boldsymbol{x} \in \mathbb{R}^n$, then their products $\mathbf{A}\boldsymbol{x}$ results in a $\mathbb{R}^n$ vector on which we can compute the norm. Considering the yet stated sub-multiplicative property what we can observe is that

$$|\mathbf{A}\boldsymbol{x}|_p \leq \|\mathbf{A}\|_p|\boldsymbol{x}|_p \qquad \Rightarrow \qquad \|\mathbf{A}\|_p \geq \frac{|\mathbf{A}\boldsymbol{x}|_p}{|\boldsymbol{x}|_p} \qquad \forall p \in \{1, 2, F, \infty\}$$

An interesting properties for the norms $p \in \{1, 2, \infty\}$ is that they are **subordinate**: this allows to compute the norm of the matrix $\mathbf{A}$ as

$$\|\mathbf{A}\|_p = \sup_{\boldsymbol{x}\neq 0} \frac{|\mathbf{A}\boldsymbol{x}|_p}{|\boldsymbol{x}|_p}$$

3.1 Lyapunov stability for linear systems

Given a **linear system** (LTV), both continuous or discrete-time, we say that such system is

- **Lyapunov**, or **marginally**, **stable** if for each initial state $\boldsymbol{x}_0$ there exists a constant $M \in \mathbb{R}$ (depending from $\boldsymbol{x}_0$ in general) such that the zero-input response satisfies

$$|\boldsymbol{x}(t)| \leq M \qquad \forall t \geq t_0 \geq 0$$

- is **Lyapunov asymptotically stable** if it is Lyapunov marginally stable and all solutions satisfies
$$\lim_{t\to\infty} |\boldsymbol{x}(t)| = 0 \qquad \forall \boldsymbol{x}_0$$
- is **Lyapunov exponentially stable** if it is asymptotically stable and:
  (a) in case of continuous-time system exists $c, \lambda > 0$ such that
  $$|\boldsymbol{x}(t)| \leq ce^{-\lambda(t-t_0)}|\boldsymbol{x}_0| \qquad \forall x_0, t_0$$
  (b) in case of discrete-time system exists $c > 0, \mu \in [0,1)$ such that
  $$|\boldsymbol{x}(t)| \leq c\mu^{t-t_0}|\boldsymbol{x}_0| \qquad \forall \boldsymbol{x}_0, t_0$$
- is **unstable** if it is not Lyapunov stable.

We observe so a certain hierarchy in the conditions for the stability, being always more stringent. For linear system we can say that exponential stability directly implies asymptotic stability, and marginal stability. The contrary, of course, might not hold (a system can be asymptotic but not with exponential law).

Note that for what concerns **Lyapunov stability** what matters is just the homogeneous (zero-input) response of the system. Recalling (2.2), the study of the Lyapunov stability is essentially related to the study of the state transition matrix because for linear systems

$$|\boldsymbol{x}(t)| = |\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0| \leq \|\boldsymbol{\Phi}(t_0, t)\||\boldsymbol{x}_0| \qquad \forall \boldsymbol{x}_0$$

The classification of **stability** is performed by analysing the properties of the state transition matrix $\boldsymbol{\Phi}$, a **intrinsic characteristic of the system**, and do not depend in any way from the initial condition $|\boldsymbol{x}_0|$ (that can be regarded as a scaling factor that can always be bounded by a *big enough* constant c).

3.2 LTI systems and conditions for exponential stability

Continuous time Equation (2.9) proved that the state transition matrix $\boldsymbol{\Phi}(t, t_0)$ of a continuous-time LTI system collapsed to the exponential $e^{\mathbf{A}t}$. Recalling the idea and proof for the region of convergence of the homogeneous response of (ACT-LTI) we can say that

$$\begin{aligned} \text{asymptotic stability} &\quad\Leftrightarrow\quad \mathbb{R}\mathrm{e}\{\lambda_i\{\mathbf{A}\}\} < 0 \ \forall i \\ \text{exponential stability} &\quad\Leftrightarrow\quad \mathbb{R}\mathrm{e}\{\lambda_i\{\mathbf{A}\}\} < 0 \ \forall i \\ \text{Lyapunov stability} &\quad\Leftrightarrow\quad \mathbb{R}\mathrm{e}\{\lambda_i\{\mathbf{A}\}\} \leq 0 \ \forall i \text{ and if } \mathbb{R}\mathrm{e}\{\lambda_i\} = 0 \Rightarrow n_i = 1 \end{aligned} \tag{3.9}$$

We observe so that for linear time-invariant system the concept of asymptotic and exponential stability are coincident (meaning that one cannot happen without the other one happening too).

Discrete time Similarly to the continuous time case, equation (2.7) proved that the state transition matrix $\boldsymbol{\Phi}(t, t_0)$ for discrete-time LTI systems coincides with the power $\mathbf{A}^{t-t_0}$ and so recalling the associated region of convergence we have

$$\begin{aligned} \text{asymptotic stability} &\quad\Leftrightarrow\quad |\lambda_i\{\mathbf{A}\}| < 1 \ \forall i \\ \text{exponential stability} &\quad\Leftrightarrow\quad |\lambda_i\{\mathbf{A}\}| < 1 \ \forall i \\ \text{Lyapunov stability} &\quad\Leftrightarrow\quad |\lambda_i\{\mathbf{A}\}| \leq 1 \ \forall i \text{ and if } |\lambda_i| = 1 \Rightarrow n_i = 1 \end{aligned} \tag{3.10}$$

Positive and negative-definite matrices In order to make more general assertion concerning the exponential stability of LTI system, a recall of some definition from linear algebra is required. Given a square matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ that's **symmetric**, meaning that $\mathbf{Q}^T = \mathbf{Q}$, then it's said

- **positive-definite**, written as $\mathbf{Q} > 0$, if it holds $\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} > 0$ for all vectors $\boldsymbol{x} \in \mathbb{R}^n \neq 0$;

- **negative-definite** $\mathbf{Q} < 0$ if $\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} < 0 \; \forall \boldsymbol{x} \neq 0$;
- positive semi-definite $\mathbf{Q} \geq 0$ if $\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} \geq 0 \; \forall \boldsymbol{x}$;
- negative semi-definite $\mathbf{Q} \neq 0$ if $\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} \leq 0 \; \forall \boldsymbol{x}$.

As a convention for any symmetric matrices $\mathbf{Q}_1, \mathbf{Q}_2$ with the notation $\mathbf{Q}_1 > Q_2$ we are implicitly assuming $\mathbf{Q}_1 - \mathbf{Q}_2 > 0$ or $\mathbf{Q}_2 - \mathbf{Q}_1 < 0$.

Note: The definition requires that the matrix $\mathbf{Q}$ must be symmetric, however if we consider the more general case of a non-symmetric matrix $\mathbf{Q}$ we observe that such element can be decomposed in a symmetric part $\mathbf{Q}_{symm} = \frac{\mathbf{Q}+\mathbf{Q}^T}{2}$ and a skew-symmetric $\mathbf{Q}_{skew} = \frac{\mathbf{Q}-\mathbf{Q}^T}{2}$ (indeed we have $\mathbf{Q}_{symm} + \mathbf{Q}_{skew} = \mathbf{Q}$). Recalling the property for matrix multiplication and transposition stating that $(\mathbf{ABC})^T = \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T$, we can rewrite the the condition for positive (negative)-definitiveness as

$$\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} = \frac{1}{2}\left(\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \left(\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x}\right)^T\right) = \frac{1}{2}\left(\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \boldsymbol{x}^T\mathbf{Q}^T\boldsymbol{x}\right) = \boldsymbol{x}^T\frac{\mathbf{Q}+\mathbf{Q}^T}{2}\boldsymbol{x} = \boldsymbol{x}^T\mathbf{Q}_{symm}\boldsymbol{x}$$

An important result from linear algebra is that for any symmetric matrix $\mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{n\times n}$ the following statements are all equivalents:

i) $\mathbf{Q}$ is positive definite;

ii) all eigenvalues of $\mathbf{Q}$ are positive and real evaluates;

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

iv) there exist a non-singular matrix $\mathbf{H} \in \mathbb{R}^{n\times n}$ such that $\mathbf{Q} = \mathbf{H}^T\mathbf{H}$ due to the Cholesky upper-triangular factorization. This also implies that exists the *root* $\mathbf{K} = \sqrt{\mathbf{Q}}$, so a matrix such that $\mathbf{K}^T\mathbf{K} = \mathbf{Q}$.

As a consequence we determine that for all **quadratic functions** in the form $\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x}$ holds the so called ***sandwich inequality***

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

Theorem 3.1: Given an autonomous continuous-time LTI system with dynamics so in the form $\dot{\boldsymbol{x}} = \mathbf{A}\boldsymbol{x}$, the following statements are equivalent:

i) the system is asymptotically stable;

ii) the system is exponentially stable;

iii) the matrix $\mathbf{A}$ is ***Hurwitz***, meaning that all eigenvalues have a negative real part: $\mathbb{R}\mathrm{e}\{\lambda_i\} < 0 \; \forall i$;

iv) for each symmetric positive-definite matrix $\mathbf{Q} = \mathbf{Q}^T > 0 \in \mathbb{R}^{n\times n}$ there exists a matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ that's symmetric and positive definite that satisfy the following **Lyapunov equality**:

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q} \tag{3.11}$$

Moreover the solution of the matrix $\mathbf{P}$ is unique;

v) there exists a symmetric positive-definite matrix $\mathbf{P} > 0 \in \mathbb{R}^{n\times n}$ satisfying the **Lyapunov inequality**:

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} < 0 \tag{3.12}$$

From a computational point of view, statement *v)* is preferred to asses exponential stability through the use of the so called ***Linear Matrix Inequalities*** **LMIs**. The Lyapunov equality *iv)* is a stronger definition then the inequality, but the solution is more complex to compute (thus the choice of LMIs to solve such problems).

Proof 3.1: The equivalence of statements *i)*, *ii)* and *iii)* is *obvious* and the link between them has been already discussed in the previous pages. Intuition can easily relate *iv)* with *v)*, however the backward implication (*v)* $\Rightarrow$ *iv)*) is mathematically impossible and requires to pass through statements *i)* and *ii)*. The goal now is to prove the relation of *iv)* and *v)* with the other statements:

a) exponential stability *ii)* implies the Lyapunov equality *iv)*: we can show that the unique solution to (3.11) is determined by the matrix $\mathbf{P}$ defined as

$$\mathbf{P} = \int_0^\infty e^{\mathbf{A}^T t}\mathbf{Q}e^{\mathbf{A}t}\,dt \qquad (\star)$$

To show the *finiteness* of the solution we have to compute the norm on such matrix

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

and for the sub-multiplicative property

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

Denoting with $\mu\{\mathbf{A}\} = \max_i \mathbb{R}\mathrm{e}\{\lambda_i\}$ the **spectral abscissa** of the matrix $\mathbf{A}$ (where λ_i are the eigenvalues of such matrix), we can exploit a property stating that $\forall \lambda > \mu\{\mathbf{A}\}$ there exists a constant $k \in \mathbb{R}$ such that $\|e^{\mathbf{A}t}\| \leq ke^{\lambda t}$ for any positive time t. Applying this rule to the previous equation determines

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

Having hypothesized the exponential stability of the system, then it means that all eigenvalues λ_i of $\mathbf{A}$ have a negative real part, implying that the integral yet reported converges to a finite number, hence the norm of $\mathbf{P}$ is bounded.

We can show now that $(\star)$ is a solution of (3.11) by substituting the proposed $\mathbf{P}$ in the Lyapunov equality:

$$\begin{aligned}\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} &= \int_0^\infty \mathbf{A}^T e^{\mathbf{A}^T t}\mathbf{Q}e^{\mathbf{A}t}\,dt + \int_0^\infty e^{\mathbf{A}^T t}\mathbf{Q}e^{\mathbf{A}t}\mathbf{A}\,dt \\ &= \int_0^\infty \left(\mathbf{A}^T e^{\mathbf{A}^T t}\mathbf{Q}e^{\mathbf{A}t} + e^{\mathbf{A}^T t}\mathbf{Q}e^{\mathbf{A}t}\mathbf{A}\right) dt\end{aligned}$$

Considering that $e^{\mathbf{A}^T t}$ can be regarded as $\left(e^{\mathbf{A}t}\right)^T$, the whole term in the parenthesis can be regarded as the derivative $\frac{d}{dt}\left(e^{\mathbf{A}^T t}\mathbf{Q}e^{\mathbf{A}t}\right)$: this implies that the evaluation of the previous integral reduces to

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

Proven so that $(\star)$ is a solution of (3.11), all we need to show is that $\mathbf{P}$ is positive-definite. By computing the product

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

we can observe that the vector $\boldsymbol{w}(t)$ defined as $e^{\mathbf{A}t}\boldsymbol{z}$ is always non-zero and allows to express the previous equation as

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

where the inequality has been determined considering the sandwich (S); in this way we proved that $\mathbf{P}$ is positive definite.

As a last argument we can show that $(\star)$ is the unique solution of (3.11): by contradiction let's assume there exists two distinct matrices $\mathbf{P}_1, \mathbf{P}_2$ solving the Lyapunov equality, meaning that

$$\begin{cases} \mathbf{A}^T\mathbf{P}_1 + \mathbf{P}_1\mathbf{A} = -\mathbf{Q} \\ \mathbf{A}^T\mathbf{P}_2 + \mathbf{P}_2\mathbf{A} = -\mathbf{Q} \end{cases}$$

Subtracting the second equation from the first evaluates to $\mathbf{A}^T(\mathbf{P}_1 - \mathbf{P}_2) + (\mathbf{P}_1 - \mathbf{P}_2)\mathbf{A} = 0$; pre-multiplying this by the factor $e^{\mathbf{A}^T t} \neq 0$ and post-multiplying by $e^{\mathbf{A}t}$ both terms gives

$$\begin{aligned} 0 &= e^{\mathbf{A}^T t}\mathbf{A}^T(\mathbf{P}_1 - \mathbf{P}_2)e^{\mathbf{A}t} + e^{\mathbf{A}^T t}(\mathbf{P}_1 - \mathbf{P}_2)\mathbf{A}e^{\mathbf{A}t} = \mathbf{A}^T e^{\mathbf{A}^T t}(\mathbf{P}_1 - \mathbf{P}_2)e^{\mathbf{A}t} + e^{\mathbf{A}^T t}(\mathbf{P}_1 - \mathbf{P}_2)\mathbf{A}e^{\mathbf{A}t} \\ &= \frac{d}{dt}\left(e^{\mathbf{A}^T t}(\mathbf{P}_1 - \mathbf{P}_2)e^{\mathbf{A}t}\right) \end{aligned}$$

b) We can now show that the solution to the Lyapunov inequality (3.12) lead to an exponentially stable system *ii)*. To do so we have to define the concept of **Lyapunov function** as a function $V : \mathbb{R}^n \to \mathbb{R}$ characterized by specific properties that we will discuss in the chapter of hybrid systems. However in the specific case of LTI system we can regard as **Lyapunov function** the quadratic equation

$$V(\boldsymbol{x}) = \boldsymbol{x}^T\mathbf{P}\boldsymbol{x} \ \in \mathbb{R} \qquad\qquad \forall x \in \mathbb{R}^n, \mathbf{P} > 0 \tag{3.13}$$

Observing that the result is indeed a scalar, we can apply the sandwich inequality (S) and obtain

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

that can be rewritten as

$$c_1|\boldsymbol{x}|^2 \leq V(\boldsymbol{x}) \leq c_2|\boldsymbol{x}|^2$$

If we now ocnsider the directional derivative of the Lyapunov function what we obtain is

$$\frac{dV(\boldsymbol{x})}{dt} = \frac{\partial V}{\partial \boldsymbol{x}}\frac{d\boldsymbol{x}}{dt} = \frac{\partial V}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}} = \nabla V(\boldsymbol{x})\,\mathbf{A}\boldsymbol{x} = \langle \nabla V(\boldsymbol{x}), \mathbf{A}\boldsymbol{x}\rangle$$

where $\nabla V(\boldsymbol{x})$ is the gradient of the Lyapunov function considered as a row vector that evaluates to $2\mathbf{P}\boldsymbol{x}$ and is representing the **directional derivative** of $V(\boldsymbol{x})$. Substituting in the previous equation this gives

$$\begin{aligned} \nabla V(\boldsymbol{x})\,\mathbf{A}\boldsymbol{x} = 2\mathbf{P}\boldsymbol{x}\mathbf{A}\boldsymbol{x} &= \boldsymbol{x}^T\mathbf{P}\mathbf{A}\boldsymbol{x} + \boldsymbol{x}^T\mathbf{P}\mathbf{A}\boldsymbol{x} = \boldsymbol{x}^T\mathbf{P}\mathbf{A}\boldsymbol{x} + \boldsymbol{x}^T\mathbf{A}^T\mathbf{P}^T\boldsymbol{x} \\ &= \boldsymbol{x}^T(\mathbf{P}\mathbf{A} + \mathbf{A}^T\mathbf{P})\boldsymbol{x} = -\boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x} \end{aligned}$$

As hypothesis in *v)* we have that the matrix $\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A}$ is negative-definite, meaning that the matrix $\boldsymbol{\text{œ}}$ defined as $-(\mathbf{P}\mathbf{A} + \mathbf{A}^T\mathbf{P})$ must be positive-definite. From the sandwich inequality (S) we have $\lambda_{\min}\{\boldsymbol{\Sigma}\}|\boldsymbol{x}|^2 \leq \boldsymbol{x}^T\boldsymbol{\text{œ}}\boldsymbol{x}$ but also reverting $-\lambda_{\min}\{\boldsymbol{\Sigma}\} \geq -\boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x}$ we can show that

$$\nabla V(\boldsymbol{x})\,\mathbf{A}\boldsymbol{x} \leq -\lambda_{\min}\{\boldsymbol{\Sigma}\}|\boldsymbol{x}|^2 \leq -\lambda_{\min}\{\boldsymbol{\Sigma}\}\frac{1}{c_2}V(\boldsymbol{x}) = -c_3V(\boldsymbol{x}) \tag{$\circ$}$$

Intuitively we have that the Lyapunov function in (3.13) is positive definite and with proof until now we shown that it's derivative is always negative for any state $\boldsymbol{x}$: this means that, as the time increases, the states are always going closer to zero (in fact in order to have a reduction of $V(\boldsymbol{x})$ the vector $\boldsymbol{x}$ must get *smaller*); in particular $(\circ)$ is commonly referred as the **flow inequality**. Formally to ensure exponential stability we have to use the **comparison theorem** stating that *given a scalar function $v(t)$ that'd 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 hold $v(t) \leq e^{-\mu(t-t_0)}v(t_0)$ for all $t \geq t_0$.* In $(\circ)$ we indeed have the form

$$\dot{V}(t) = \nabla V(\boldsymbol{x})\,\mathbf{A}\boldsymbol{x} \leq -c_3V(\boldsymbol{x}(t))$$

and so for the comparison theorem the solution satisfies

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

The exponential decay is more evident if we explicitly relate $|\boldsymbol{x}|$ with $|\boldsymbol{x}(t_0)|$ as

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

Theorem 3.1 can also be extended to discrete-time systems as

Theorem 3.2: Given an autonomous discrete-time LTI system with dynamics so in the form $\boldsymbol{x}^+ = \mathbf{A}\boldsymbol{x}$, the following statements are equivalent:

i) the system is asymptotically stable;

ii) the system is exponentially stable;

iii) the matrix $\mathbf{A}$ is *Schur*, meaning that all eigenvalues are characterized by a magnitude smaller than one: $|\lambda_i| < 1 \; \forall i$;

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

$$\mathbf{A}^T \mathbf{P} \mathbf{A} - \mathbf{P} = -\mathbf{Q} \tag{3.14}$$

Moreover the solution of the matrix $\mathbf{P}$ is unique;

v) there exists a symmetric positive-definite matrix $\mathbf{P} > 0 \in \mathbb{R}^{n \times n}$ satisfying the **Lyapunov inequality**:

$$\mathbf{A}^T \mathbf{P} \mathbf{A} - \mathbf{P} < 0 \tag{3.15}$$

3.3 Linear quadratic regulator

Assuming now to have a continuous-time LTI plant where all it's states can be measured (note that this in general is a very strong hypothesis), then the best control system we can design is the **linear quadratic regulator LQR** and is based on the minimization of the functional

$$\mathcal{J} = \int_0^\infty \boldsymbol{y}^T(\tau) \mathbf{Q} \boldsymbol{y}(\tau) + \boldsymbol{u}^\tau \mathbf{R} \boldsymbol{u}(\tau) \, d\tau \tag{3.16}$$

Intuitively we can regard $\mathbf{Q}$ as a matrix that's used to describe *how much we trust the sensed value of the output* and $\mathbf{R}$ *how much we trust the actuators in input*. As a general rule of thumb $\mathbf{Q}, \mathbf{R}$ can be chosen as diagonal with their elements *measuring the confidence* of the related input/output, however in the more general case we have to ensure that both $\mathbf{Q}$ and $\mathbf{R}$ are symmetric positive-definite matrices; this, of course, are the main design parameters that have to be tuned and determined in order to obtained the desired result.

In particular the linear quadratic regulator is based on the solution of the problem known as ***Bayes rule*** defined as

$$\mathcal{J}^* = \min_{u(t), t \geq 0} \mathcal{J} = \min_{u(t), t \geq 0} \int_0^\infty \boldsymbol{y}^T(\tau) \mathbf{Q} \boldsymbol{y}(\tau) + \boldsymbol{u}^\tau \mathbf{R} \boldsymbol{u}(\tau) \, d\tau \tag{3.17}$$

The solution of such optimal control problem is characterized by an input $\boldsymbol{u}^*$ that can be regarded as a linear combination of the states $\boldsymbol{x}$ performed through a matrix $\mathbf{K} \in \mathbb{R}^{m \times n}$:

$$\boldsymbol{u}^*(t) = -\mathbf{K} \boldsymbol{x}(t) \tag{3.18}$$

The main goal in the design of the linear quadratic regular is to tune $\mathbf{K}$ as function of the chosen $\mathbf{Q}$ and $\mathbf{R}$ and the result heavily depends on the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ characterizing the system itself.

Feedback invariant In order to prove that (3.18) is indeed the optimal solution of (3.17) we have to introduce a **feedback invariant** $\mathcal{H}$, a property of the system depending only on the initial states that's defined as

$$\mathcal{H}(\boldsymbol{x}(\cdot),\boldsymbol{u}(\cdot)) = -\int_0^\infty \Big(\mathbf{A}\boldsymbol{x}(t)+\mathbf{B}\boldsymbol{u}(t)\Big)^T \mathbf{P}\boldsymbol{x}(t) + \boldsymbol{x}^T(t)\mathbf{P}\Big(\mathbf{A}\boldsymbol{x}(t)+\mathbf{B}\boldsymbol{u}(t)\Big)\,dt \tag{3.19}$$

Of course this integral converges to a value as long as $\lim_{t\to\infty}\boldsymbol{x}(t)=0$.

Proof 3.2: We can prove that (3.19) is indeed an invariant considering that $\mathbf{A}\boldsymbol{x}+\mathbf{B}\boldsymbol{u}$ is just $\dot{\boldsymbol{x}}$, so

$$\begin{aligned}\mathcal{H}(\boldsymbol{x}(\cdot),\boldsymbol{u}(\cdot)) &= -\int_0^\infty \dot{\boldsymbol{x}}^T\mathbf{P}\boldsymbol{x} + \boldsymbol{x}^T\mathbf{P}\dot{\boldsymbol{x}}\,dt = -\int_0^\infty \frac{d}{dt}(\boldsymbol{x}^T\mathbf{P}\boldsymbol{x})\,dt = -\Big[\boldsymbol{x}(t)^T\mathbf{P}\boldsymbol{x}(t)\Big]_0^\infty \\ &= -\cancel{\boldsymbol{x}^T(\infty)\mathbf{P}\boldsymbol{x}(\infty)} + \boldsymbol{x}^T(0)\mathbf{P}\boldsymbol{x}(0) = \boldsymbol{x}_0^T\mathbf{P}\boldsymbol{x}_0\end{aligned}$$

so as long as $\lim_{t\to\infty}\boldsymbol{x}(t)=0$ holds the functional $\mathcal{H}$ depend only on the initial state $\boldsymbol{x}_0$ (and on the positive-definite matrix $\mathbf{P}$).

Proof 3.3: We can now prove that (3.18) is indeed the solution of the optimal control problem (3.17). Let us consider for simplicity a system whose output depends only on the states of the system (so $\mathbf{D}$ is identically null), then we can rewrite the functional (3.16) as

$$\mathcal{J} = \mathcal{H}(\boldsymbol{x},\boldsymbol{u}) + \int_0^\infty \Lambda(\boldsymbol{x},\boldsymbol{u})\,dt \tag{$\star$}$$

where the function Λ is characterized by having

$$\min_{\boldsymbol{u}} \Lambda(\boldsymbol{x},\boldsymbol{u}) = 0 \qquad \forall \boldsymbol{x}$$

To show that $(\star)$ is indeed equivalent to (3.16) we can add and subtract from the formal definition of $\mathcal{J}$ the functional $\mathcal{H}$, so

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

Expanding $\mathcal{H}$ as reported in (3.19) and knowing that the output $\boldsymbol{y}$ depend only on the states as $\mathbf{C}\boldsymbol{x}$ (and so it means $\boldsymbol{y}^T = (\mathbf{C}\boldsymbol{x})^T = \boldsymbol{x}^T\mathbf{C}^T$), then we can rewrite

$$\begin{aligned}\mathcal{J} &= \int_0^\infty \boldsymbol{x}^T\mathbf{C}^T\mathbf{Q}\mathbf{C}\boldsymbol{x} + \boldsymbol{u}^T\mathbf{R}\boldsymbol{u} + \big(\mathbf{A}\boldsymbol{x}+\mathbf{B}\boldsymbol{u}\big)^T\mathbf{P}\boldsymbol{x} + \boldsymbol{x}^T\mathbf{P}\big(\mathbf{A}\boldsymbol{x}+\mathbf{B}\boldsymbol{u}\big)\,d\tau + \mathcal{H}(\boldsymbol{x},\boldsymbol{u}) \\ &= \int_0^\infty \underbrace{\boldsymbol{x}^T\big(\mathbf{C}^T\mathbf{Q}\mathbf{C}+\mathbf{A}^T\mathbf{P}+\mathbf{P}\mathbf{A}\big)\boldsymbol{x} + 2\boldsymbol{x}^T\mathbf{P}\mathbf{B}\boldsymbol{u} + \boldsymbol{u}^T\mathbf{R}\boldsymbol{u}}_{\Lambda(\boldsymbol{x},\boldsymbol{u})}\,dt + \mathcal{H}(\boldsymbol{x},\boldsymbol{u})\end{aligned} \tag{$\dagger$}$$

The main goal now is to define a positive-definite matrix $\mathbf{M}$ that allows to rewrite Λ as $(\boldsymbol{u}+\mathbf{K}\boldsymbol{x})^T\mathbf{M}(\boldsymbol{u}+\mathbf{K}\boldsymbol{x})$: if such matrix exists then it's straightforward to see that the input $\boldsymbol{u}$ that minimize Λ is indeed $\boldsymbol{u}=-\mathbf{K}\boldsymbol{x}$ (because eventually $\boldsymbol{u}+\mathbf{K}\boldsymbol{x}=0$ meaning that Λ reaches it's minimum value that's zero).
For this scope let us consider $\mathbf{R}$ in order to solve the quadratic form: computing in fact $(\boldsymbol{u}+\mathbf{K}\boldsymbol{x})^T\mathbf{R}(\boldsymbol{u}+\mathbf{K}\boldsymbol{x})$ evaluates to $\boldsymbol{u}\mathbf{R}\boldsymbol{u}+\boldsymbol{x}^T\mathbf{K}^T\mathbf{R}\mathbf{K}\boldsymbol{x}+2\boldsymbol{x}^T\mathbf{K}^T\mathbf{R}\boldsymbol{u}$. By matching the resulting solution with the result proposed in (†) we see that $\mathbf{P}\mathbf{B}$ must equate $\mathbf{K}^T\mathbf{R}$, meaning so that $\mathbf{K}^T = \mathbf{P}\mathbf{B}\mathbf{R}^{-1}$: this determines the optimal solution of $\mathbf{K}$ as

$$\mathbf{K} = \big(\mathbf{P}\mathbf{B}\mathbf{R}^{-1}\big)^T = \big(\mathbf{R}^{-1}\big)^T\mathbf{B}^T\mathbf{P}^T = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} \tag{3.20}$$

Substituting the definition of $\mathbf{K}$ in (†) allow to rewrite the functional as

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

where the terms $\boldsymbol{x}^T\mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}\boldsymbol{x}$ inside the integral correctly matches the term $\boldsymbol{x}^T\mathbf{K}^T\mathbf{R}\mathbf{K}\boldsymbol{x}$ obtained by expanding Λ as quadratic form. Showed that $(\star)$ holds, in order to have the asymptotic convergence of the system we have to impose a condition on the quadratic term $\boldsymbol{x}^T\dots\boldsymbol{x}$: this can be achieved by solving the **algebraic Riccati equation** defined as

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} + \mathbf{C}^T\mathbf{Q}\mathbf{C} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} = 0 \tag{3.21}$$

Solving this linear equality for a positive-definite matrix $\mathbf{P}$, if we can ensure that $\lim_{t\to\infty}\boldsymbol{x}(t) = 0$ then we are sure that $\boldsymbol{u}^* = -\mathbf{K}\boldsymbol{x}$ is the solution of the linear quadratic regulator. Moreover the optimal value $\mathcal{J}^*$ of the problem is characterized by $\mathcal{H}(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{x}_0^T\mathbf{P}\boldsymbol{x}_0$.

Chapter 4

Controllability and reachability of linear systems

Chapter 2 (page 8) described the **solutions** of **linear system** (in both continuous and discrete-time case as well as time variant or invariant), but what we want to understand now is *what we can achieve* by the system.

Given two times $t_1 > t_0 \geq 0$, we denote with $\mathcal{R}[t_0, t_1]$ the $[t_0, t_1]$**-reachable subset** (or **subspace**), a linear subspace defined as

$$\mathcal{R}[t_0, t_1] = \left\{ \boldsymbol{x}_1 \in \mathbb{R}^n \ : \ \exists \boldsymbol{u} \text{ such that } \int_{t_0}^{t_1} \boldsymbol{\Phi}(t_1, \tau)\mathbf{B}(\tau)u(\tau)\, d\tau = \boldsymbol{x}_1 \right\} \tag{4.1}$$

Intuitively this set contains all states $\boldsymbol{x}_1$ that can be achieved by a linear system considering a zero-state initial condition. Each different input, using the variation of constants formula (2.3) at page 9, might generate a different final state $\boldsymbol{x}(t_1)$ and $\mathcal{R}$ contains all possible values of those values. We observe that the **reachable subset** is a **linear subspace**: given in fact an input $\overline{\boldsymbol{u}}$ leading to a certain state $\overline{\boldsymbol{x}}_1$, then in order to reach the state $k\overline{\boldsymbol{x}}_1$ (with $k \in \mathbb{R}$) all we need to do is to chose the input $\boldsymbol{u} = k\overline{\boldsymbol{u}}$.

Conversely we denote with $\mathcal{C}[t_0, t_1]$ the $[t_0, t_1]$ **controllable set** (**subspace**), the linear subspace defined as

$$\mathcal{C}[t_0, t_1] = \left\{ \boldsymbol{x}_0 \in \mathbb{R}^n \ : \ \exists \boldsymbol{u} \text{ such that } \boldsymbol{\Phi}(t_1, t_0)\boldsymbol{x}_0 + \int_{t_0}^{t_1} \boldsymbol{\Phi}(t_1, \tau)\mathbf{B}(\tau)\boldsymbol{u}(\tau)\, d\tau = 0 \right\} \tag{4.2}$$

In this case the definition is *revered* and the controllable set contains all possible initial states $\boldsymbol{x}_0$ that can be driven, through a specific input $\boldsymbol{u}(t)$, to a final state $\boldsymbol{x}_1$ that's zero: the definition contains in fact also the free response of the system that's affected by the initial condition only.
Another equivalent definition of the controllable set can be obtained by pre-multiplying the variation of constants formula by $\boldsymbol{\Phi}(t_1, t_0)^{-1}$, leading to the definition

$$\mathcal{C}[t_0, t_1] = \left\{ \boldsymbol{x}_0 \in \mathbb{R}^n \ : \ \exists \boldsymbol{u} \text{ such that } \int_{t_0}^{t_1} \boldsymbol{\Phi}(t_0, \tau)\mathbf{B}(\tau)\boldsymbol{u}(\tau)\, d\tau = -\boldsymbol{x}_0 \right\}$$

We say that a **system** is **controllable** if the dimension of the controllable subspace coincides with the dimension of the states, so mathematically

$$\text{controllable system} \qquad \Leftrightarrow \qquad \mathcal{C}[t_0, t_1] = \mathbb{R}^n \tag{4.3}$$

However all this discussion will be further clarified in section 4.5, page 33.

Linear algebra recall In order to ease the following definition of *Gramians*, a recall of concept of linear algebra is recommended. Given a matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$ we call **image** of $\mathbf{W}$ the linear subspace defined as

$$\text{Im}\{\mathbf{W}\} = \{\boldsymbol{y} \in \mathbb{R}^m \ : \ \boldsymbol{y} = \mathbf{W}\boldsymbol{x} \text{ with } \boldsymbol{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m \tag{4.4}$$

As an important note, we have that the **dimension** of the image of any matrix coincides with its **rank**, the number of linearly independent column vectors composing the matrix. We define the **kernel** of $\mathbf{W}$, also known as **null space**, the set defined as

$$\text{Ker}\{\mathbf{W}\} = \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \mathbf{W}\boldsymbol{x} = 0\} \subseteq \mathbb{R}^n \tag{4.5}$$

In the following pages it will be often used the **fundamental theorem of linear algebra** stating that for any matrix $\mathbf{W} \in \mathbb{R}^{m\times n}$ it holds

$$\dim\{\text{Im}\{\mathbf{W}\}\} + \dim\{\text{Ker}\{\mathbf{W}\}\} = n \tag{4.6}$$

Lastly, given a linear subspace $\mathbb{V} \subseteq \mathbb{R}^n$, we call **orthogonal complement** of $\mathbb{V}$ the set defined as

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

As a remark, for every $m \times n$ matrix $\mathbf{W}$ it holds that

$$\text{Im}\{\mathbf{W}\} = (\text{Ker}\left\{\mathbf{W}^T\right\})^\perp \qquad\qquad \text{Ker}\{\mathbf{W}\} = (\text{Im}\left\{\mathbf{W}^T\right\})^\perp$$

4.1 Gramians

Given two times $t_1 > t_0 \geq 0$, we denote with $\mathbf{W}_\mathcal{R}, \mathbf{W}_\mathcal{C}$ respectively the **reachability Gramians** and **controllability Gramians** of a continuous-time LTIV system the $n \times n$ symmetric matrices

$$\begin{aligned}\mathbf{W}_\mathcal{R}[t_0, t_1] &= \int_{t_0}^{t_1} \mathbf{\Phi}(t_1, \tau)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\mathbf{\Phi}^T(t_1, \tau)\, d\tau \\ \mathbf{W}_\mathcal{C}[t_0, t_1] &= \int_{t_0}^{t_1} \mathbf{\Phi}(t_0, \tau)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\mathbf{\Phi}^T(t_0, \tau)\, d\tau\end{aligned} \tag{4.7}$$

The absolute importance of such matrices is that they are directly linked with the reachability and controllability set: we will show in fact that

$$\mathcal{R}[t_0, t_1] = \text{Im}\{\mathbf{W}_\mathcal{R}[t_0, t_1]\} \qquad\qquad \mathcal{C}[t_0, t_1] = \text{Im}\{\mathbf{W}_\mathcal{C}[t_0, t_1]\} \tag{4.8}$$

Furthermore determining $\boldsymbol{\eta}_1$ ($\boldsymbol{\eta}_0$) for which it holds $\boldsymbol{x}_1 = \mathbf{W}_\mathcal{R}[t_0, t_1]\boldsymbol{\eta}_1$ ($\boldsymbol{x}_0 = \mathbf{W}_\mathcal{C}[t_0, t_1]\boldsymbol{\eta}_0$), then the **optimal input** $\boldsymbol{u}^*(t)$ minimizing the energy of the system is in the form

$$\boldsymbol{u}^*(t) = \mathbf{B}^T(t)\mathbf{\Phi}^T(t_1, t)\boldsymbol{\eta}_1 \qquad\qquad \forall t \in [t_0, t_1] \tag{4.9}$$

(4.9) holds considering reachability, but in case of controllability we have the similar result

$$\boldsymbol{u}^*(t) = -\mathbf{B}^T(t)\mathbf{\Phi}^T(t_0, t)\boldsymbol{\eta}_0 \qquad\qquad \forall t \in [t_0, t_1] \tag{4.10}$$

Proof 4.1: The key point of what has just been presented is equation (4.8) linking the reachable (controllable) set with the image of the respective Gramian; the proof presented now is made for the reachable case, but is dual for the controllable set.
In particular we will firstly prove that $\text{Im}\{\mathbf{W}_\mathcal{R}\} \subseteq \mathcal{R}$ and secondly $\mathcal{R} \subseteq \text{Im}\{\mathbf{W}_\mathcal{R}\}$: having that the two must hold together, then it means that $\mathcal{R} = \text{Im}\{\mathbf{W}_\mathcal{R}\}$.

a) Taking $\boldsymbol{x}_1$ a state belonging to the image of the reachability gramian $\boldsymbol{x}_1 \in \text{Im}\{\mathbf{W}_\mathcal{R}\}$, using the optimal control solution (4.9) plugged in the variation of constants formula (2.3), page 9, considering a initial zero-state evaluates to

$$\begin{aligned}\boldsymbol{x}(t_1) &= \int_{t_0}^{t_1} \mathbf{\Phi}(t_1, \tau)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\mathbf{\Phi}^T(t_1, \tau)\boldsymbol{\eta}_1\, d\tau = \left(\int_{t_0}^{t_1} \mathbf{\Phi}(t_1, \tau)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\mathbf{\Phi}^T(t_1, \tau)\, d\tau\right)\boldsymbol{\eta}_1 \\ &= \mathbf{W}_\mathcal{R}[t_0, t_1]\boldsymbol{\eta}_1 = \boldsymbol{x}_1\end{aligned}$$

b) For any state $\boldsymbol{x}_1 \in \mathcal{R}[t_0, t_1]$ it's known by definition that at least one control input $\boldsymbol{u}^*$ exists

satisfying $\boldsymbol{x}_1 = \int_{t_0}^{t_1} \boldsymbol{\Phi}(t_1,\tau)\mathbf{B}(\tau)\boldsymbol{u}^*(\tau)\,d\tau$. Showing that

$$\boldsymbol{x}_1 \in \operatorname{Im}\{\mathbf{W}_{\mathcal{R}}\} = \left(\operatorname{Ker}\left\{\mathbf{W}_{\mathcal{R}}^T\right\}\right)^{\perp} = (\operatorname{Ker}\{\mathbf{W}_{\mathcal{R}}\})^T$$

implies $\boldsymbol{x}_1^T\boldsymbol{\eta}_{1k} = 0$ for any vector $\boldsymbol{\eta}_{1k} \in \operatorname{Ker}\{\mathbf{W}_{\mathcal{R}}\}$.
Expanding so the definition of $\boldsymbol{x}(t)$ from the variation of constants formula gives

$$\boldsymbol{x}^T\boldsymbol{\eta}_{1k} = \int_{t_0}^{t_1} \boldsymbol{\Phi}(t_1,\tau)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\boldsymbol{\Phi}^T(t_1,\tau)\boldsymbol{\eta}_{1k}\,d\tau$$

Since $\boldsymbol{\eta}_{1k}$ belongs to the kernel and having $\left(\mathbf{W}_{\mathcal{R}}\boldsymbol{\eta}_{1k}\right)^T = \boldsymbol{\eta}_{1k}^T\mathbf{W}_{\mathcal{R}}$ we have

$$\boldsymbol{\eta}_{1k}^T\mathbf{W}_{\mathcal{R}}[t_0,t_1]\boldsymbol{\eta}_{1k} = \int_{t_0}^{t_1} \boldsymbol{\eta}_{1k}\boldsymbol{\Phi}(t_1,\tau)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\boldsymbol{\Phi}^T(t_1,\tau)\boldsymbol{\eta}_{1k}\,d\tau$$
$$\boldsymbol{\eta}_{1k}\,0 = \int_{t_0}^{t_1} \left\|\mathbf{B}^T(\tau)\boldsymbol{\Phi}^T(t_1,\tau)\boldsymbol{\eta}_{1k}\right\|^2 d\tau = 0$$

Having an integrand that's always non-negative, in order to have a zero integral we must ensure $\mathbf{B}^T(\tau)\boldsymbol{\Phi}^T(t_1,\tau)\boldsymbol{\eta}_{1k} = 0$ for any time τ, proving so that $\boldsymbol{x}_1$ is orthogonal to $\boldsymbol{\eta}_{1k}$.

Theorem 4.1: Given two times $t_1 > t_0 \geq 0$:

i) for each state $\boldsymbol{x}_1 \in \mathcal{R}[t_0,t_1]$ in the reachable set, the control input $\boldsymbol{u}^*$ described in (4.9) transform the state from $\boldsymbol{x}_0 = 0$ to $\boldsymbol{x}_1 = \mathbf{W}_{\mathcal{R}}[t_0,t_1]\boldsymbol{\eta}_1$ with the **minimum energy control**

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

Moreover the minimum control energy evaluates to

$$\min_{\boldsymbol{u}} \int_{t_0}^{t_1} |\boldsymbol{u}(\tau)|^2\,d\tau = \boldsymbol{\eta}_1^T\mathbf{W}_{\mathcal{R}}\boldsymbol{\eta}_1 \tag{4.11}$$

ii) similarly, for each state $\boldsymbol{x}_0 \in \mathcal{C}[t_0,t_1]$, the control $\boldsymbol{u}^* = -\mathbf{B}^T(t)\boldsymbol{\Phi}^T(t_0,t)\boldsymbol{\eta}_0$ transform the initial state $\boldsymbol{x}_0 \neq 0$ to the final state $\boldsymbol{x}_1 = 0$ with the minimum energy control that evaluates to $\boldsymbol{\eta}_0^T\mathbf{W}_{\mathcal{C}}\boldsymbol{\eta}_0$.

The main advantage presented by this theorem is that Gramians $\mathbf{W}_{\mathcal{R}}, \mathbf{W}_{\mathcal{C}}$ can be used to estimate the energy required for *moving the states*.

Proof 4.2: We can prove now *i)*, but the same process can be applied with analogy to *ii)*. Given $\boldsymbol{u}^*(t)$ the minimum energy control moving the initial state $\boldsymbol{x}_0 = 0$ into $\boldsymbol{x}_1 \in \mathcal{R}[t_0,t_1]$ (defined in 4.9) and $\tilde{\boldsymbol{u}}(t)$ any other non-optimal input *achieving the same goal*, so such that $\int_{t_0}^{t_1} \boldsymbol{\Phi}\mathbf{B}\tilde{\boldsymbol{u}}\,d\tau = \boldsymbol{x}_1$, their relative difference evaluates to

$$0 = \int_{t_0}^{t_1} \boldsymbol{\Phi}(t_1,\tau)\mathbf{B}(\tau)\left(\boldsymbol{u}^*(\tau) - \tilde{\boldsymbol{u}}(\tau)\right)d\tau \tag{†}$$

Defining the difference $\boldsymbol{u}^*(t) - \tilde{\boldsymbol{u}}(t)$ as $\boldsymbol{v}(t)$, we can compute the control energy of $\tilde{\boldsymbol{u}}$ as function of both $\boldsymbol{u}^*$ and $\boldsymbol{v}$ as

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

Knowing the optimal input $\boldsymbol{u}^*$ that's in the form $\mathbf{B}^T\boldsymbol{\Phi}^T\boldsymbol{\eta}_1$, then we can expand the definition

of the energy of the control $\tilde{\boldsymbol{u}}$ as

$$\begin{aligned}\int_{t_0}^{t_1} |\tilde{\boldsymbol{u}}(\tau)|^2 \, d\tau &= \boldsymbol{\eta}_1^T \int_{t_0}^{t_1} \boldsymbol{\Phi}\mathbf{B}\mathbf{B}^T\boldsymbol{\Phi}^T \, d\tau \, \boldsymbol{\eta}_1 + \int_{t_0}^{t_1} |\boldsymbol{v}|^2 \, d\tau + 2\boldsymbol{\eta}_1^T \cancel{\int_{t_0}^{t_1} \boldsymbol{\Phi}\mathbf{B}\boldsymbol{v} \, d\tau} \\ &= \boldsymbol{\eta}_1^T \mathbf{W}_{\mathcal{R}} \boldsymbol{\eta}_1 + \int_{t_0}^{t_1} |\boldsymbol{v}|^2 \, d\tau\end{aligned}$$

where the mixed term has been cancelled because of (†). Considering that $|\boldsymbol{v}|$ is a non-negative quantity it means that integrating this value over time results in an increment on the energy spent by the control with respect to the minimum value $\boldsymbol{\eta}_1^T \mathbf{W}_{\mathcal{R}} \boldsymbol{\eta}_1$ determine by the optimal control.

4.2 Continuous-time LTI systems: controllability and reachability matrix

Considering the specific case of a **continuous-time LTI** system, we had that the state transition matrix $\boldsymbol{\Phi}(t, t_0)$ reduced to the matrix exponential $e^{\mathbf{A}(t-t_0)}$: this allows us to simplify the definition of the **Gramians** (4.7) for both reachable and controllable set to the form

$$\begin{aligned}\mathbf{W}_{\mathcal{R}}[t_0, t_1] &= \int_{t_0}^{t_1} e^{\mathbf{A}(t_1-\tau)}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T(t_1-\tau)} \, d\tau = \int_0^{t_1-t_0} e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T t} \, d\tau \\ \mathbf{W}_{\mathcal{C}}[t_0, t_1] &= \int_{t_0}^{t_1} e^{\mathbf{A}(t_0-\tau)}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T(t_0-\tau)} \, d\tau = \int_0^{t_1-t_0} e^{-\mathbf{A}t}\mathbf{B}\mathbf{B}^T e^{-\mathbf{A}^T t} \, d\tau\end{aligned} \tag{4.12}$$

The study of the dynamics for what concerns reachability and observability of LTI system is furthermore simplified and can be performed through the definition of the **controllability** (or equivalently **reachability**) **matrix** $\mathbf{R} \in \mathbb{R}^{n \times (kn)}$ defined as the *concatenation* of the elements as follows:

$$\mathbf{R} = \begin{bmatrix}\mathbf{B} & \mathbf{A}\mathbf{B} & \mathbf{A}^2\mathbf{B} & \dots & \mathbf{A}^{n-1}\mathbf{B}\end{bmatrix} \tag{4.13}$$

Theorem 4.2: Given any two times $t_1 > t_0 \geq 0$, then for a continuous-time LTI system it holds that

$$\mathcal{R}[t_0, t_1] = \operatorname{Im}\{\mathbf{W}_{\mathcal{R}}[t_0, t_1]\} = \operatorname{Im}\{\mathbf{R}\} = \operatorname{Im}\{\mathbf{W}_{\mathcal{C}}[t_0, t_1]\} = \mathcal{C}[t_0, t_1] \tag{4.14}$$

(4.14) implicitly states that for time-invariant system the concept of *controllability* and *reachability* are coinciding. A direct consequence of this theorem are the properties of **time reversibility**, meaning that a state is controllable if and only if it's reachable, and **time scaling** for which controllability/reachability do not depend on the time difference $t_1 - t_0$ because there will always exists an input $\boldsymbol{u}(t)$ that allows to reach the desired state in an arbitrary small time.

Proof 4.3: To prove (4.14) we can show that $\mathcal{R} \subseteq \operatorname{Im}\{\mathbf{R}\}$ and $\operatorname{Im}\{\mathbf{R}\} \subseteq \operatorname{Im}\{\mathbf{W}_{\mathcal{R}}\}$, thus for (4.8) we can ensure $\operatorname{Im}\{\mathbf{W}_{\mathcal{R}}\} = \mathcal{R} = \operatorname{Im}\{\mathbf{R}\}$:

a) Given a reachable state $\boldsymbol{x}_1 \in \mathcal{R}[t_0, t_1]$, then it's already ensured that exists an input $\boldsymbol{u}(t)$ for which $\int_{t_0}^{t_1} e^{\mathbf{A}(t_1-\tau)}\mathbf{B}\boldsymbol{u}(\tau) \, d\tau = \boldsymbol{x}_1$. Exploiting Cayley-Hamilton theorem, page 12, we can rewrite this integral as

$$\boldsymbol{x}_1 = \int_{t_0}^{t_1} \sum_{i=0}^{n-1} \alpha_i(t_1-\tau)\mathbf{A}^i\mathbf{B}\boldsymbol{u}(\tau) \, d\tau = \sum_{i=0}^{n-1} \mathbf{A}^i\mathbf{B} \int_{t_0}^{t_1} \alpha_i(t_1-\tau)\boldsymbol{u}(\tau) \, d\tau$$

The last equality can be rewritten as a linear combination in the form

$$\boldsymbol{x}_1 = \underbrace{\begin{bmatrix}\mathbf{B} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B}\end{bmatrix}}_{=\mathbf{R}} \underbrace{\begin{pmatrix}\int_{t_0}^{t_1} \alpha_0(t_1-\tau)\boldsymbol{u}(\tau)\,d\tau \\ \int_{t_0}^{t_1} \alpha_1(t_1-\tau)\boldsymbol{u}(\tau)\,d\tau \\ \vdots \\ \int_{t_0}^{t_1} \alpha_{n-1}(t_1-\tau)\boldsymbol{u}(\tau)\,d\tau\end{pmatrix}}_{\boldsymbol{v}}$$

From this expression we can clearly see that exists a vector $\boldsymbol{v}$ that pre-multiplied by the reachability matrix $\mathbf{R}$ gives the desired state $\boldsymbol{x}_1$, meaning that any reachable state $\boldsymbol{x}_1 \in \mathcal{R}[t_0, t_1]$ is also inside the image of $\mathbf{R}$.

b) In proof 4.1 we showed that $\mathbf{B}^T\boldsymbol{\Phi}^T(t_1,\tau)\boldsymbol{\eta}_{1k} = 0$ for any vector $\boldsymbol{\eta}_{1k} \in \operatorname{Ker}\{\mathbf{W}_{\mathcal{R}}[t_0,t_1]\}$ and for any time $\tau \in [t_0, t_1]$: applying this to the specific case of continuous-time system reduces to

$$\mathbf{B}^T e^{\mathbf{A}^T(t_1-\tau)}\boldsymbol{\eta}_{1k} = 0 \qquad (\dagger)$$

Deriving in time this expression transposed evaluates to $\frac{d}{dt}(\dagger)^T = -\boldsymbol{\eta}_{1k}^T\mathbf{A}e^{\mathbf{A}(t_1-\tau)}\mathbf{B} = 0$; generalizing the concept to the i-th derivative in time gives

$$\frac{d^i}{dt^i}(\dagger)^T = (-1)^i\boldsymbol{\eta}_{1k}^T\mathbf{A}^i e^{\mathbf{A}(t_1-\tau)}\mathbf{B} = 0$$

Evaluating the derivative for $\tau = t_1$ results in $\boldsymbol{\eta}_{1k}^T\mathbf{A}^i\mathbf{B} = 0\ \forall i$. For any reachable state $\boldsymbol{x}_i = \mathbf{R}\boldsymbol{v} \in \operatorname{Im}\{\mathbf{W}_{\mathcal{R}}\}[t_0,t_1]$ we can see that

$$\begin{aligned}\boldsymbol{x}_1^T\boldsymbol{\eta}_{1k} = \boldsymbol{\eta}_{1k}\boldsymbol{x}_1 = \boldsymbol{\eta}_{1k}\mathbf{R}\boldsymbol{v} &= \boldsymbol{\eta}_{1k}^T\begin{bmatrix}\mathbf{B} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B}\end{bmatrix}\boldsymbol{v} \\ &= \begin{bmatrix}\cancel{\boldsymbol{\eta}_{1k}\mathbf{B}} & \cancel{\boldsymbol{\eta}_{1k}\mathbf{AB}} & \dots & \cancel{\boldsymbol{\eta}_{1k}\mathbf{A}^{n-1}\mathbf{B}}\end{bmatrix}\boldsymbol{v}\end{aligned}$$

Having $\boldsymbol{x}_1 \perp \boldsymbol{\eta}_{1k}$ for any vector $\boldsymbol{\eta}_{1k} \in \operatorname{Ker}\{\mathbf{W}_{\mathcal{R}}\}$ then it means that $\boldsymbol{x}_1$ belong to the orthogonal subspace $\left(\operatorname{Ker}\{\mathbf{W}_{\mathcal{R}}\}\right)^{\perp}$; having from linear algebra $\operatorname{Im}\{\mathbf{W}\} = \left(\ker\mathbf{W}^T\right)^{\perp}$, but knowing also that $\mathbf{W}_{\mathcal{R}}$ is a symmetric matrix, when we have that $\left(\operatorname{Ker}\{\mathbf{W}_{\mathcal{R}}\}\right)^{\perp} = \operatorname{Im}\{\mathbf{W}_{\mathcal{R}}\}$, this $\operatorname{Im}\{\mathbf{R}\} \subseteq \operatorname{Im}\{\mathbf{W}_{\mathcal{R}}\}$.

As a matter of fact, combining (4.13) with (4.14) tells us that for LTI system both controllability and reachability are coincident sets that are characteristic of the dynamic of the system, in particular is determine just by the matrices $\mathbf{A}, \mathbf{B}$ (due to the characterization of the controllability matrix).

Moreover considering (4.3) we can further simplify the definition of **controllable system** by considering that the dimension of the controllable subspace coincides with the dimension of the image of the reachability matrix:

$$\text{controllable sys.} \quad \Leftrightarrow \quad \text{reachable sys.} \quad \Leftrightarrow \quad \dim\left(\operatorname{Im}\{\mathbf{R}\}\right) = n \qquad (4.15)$$

4.3 Extension to the discrete-time case

Time-varying case Until now only the linear continuous-time case has been considered, but similar tools can be developed also for the discrete-time counter part, but in this case we have to take care of more subtle details in the analysis.

Knowing the solution (2.6), page 10, of such system, given two times $t_1 > t_0 \geq 0$ we can define the $[t_0, t_1]$**-reachable** and $[t_0, t_1]$**-controllable subspaces** of the discrete-time LTV system the sets

defined as

$$\mathcal{R}[t_0,t_1] = \left\{ \boldsymbol{x}_1 \ : \ \exists \boldsymbol{u}(t) \text{ with } t \in [t_0,t_1) \text{ such that } \boldsymbol{x}(t_1) = \sum_{\tau=0}^{t_1-1} \boldsymbol{\Phi}(t_1,\tau+1)\mathbf{B}(\tau)\boldsymbol{u}(\tau) \right\}$$
$$\mathcal{C}[t_0,t_1] = \left\{ \boldsymbol{x}_0 \ : \ \exists \boldsymbol{u}(t) \text{ with } t \in [t_0,t_1) \text{ such that } 0 = \boldsymbol{\Phi}(t_1,t_0)\boldsymbol{x}_0 + \sum_{\tau=0}^{t_1-1} \boldsymbol{\Phi}(t_1,\tau+1)\mathbf{B}(\tau)\boldsymbol{u}(\tau) \right\} \tag{4.16}$$

Indeed the concept that the sets $\mathcal{R}$ and $\mathcal{C}$ are representing are fundamentally the same: the first subspace contains all the possible states $\boldsymbol{x}_1$ that can be achieved by controls starting from a zero-state configuration, while in the second case the set contains all initial states $\boldsymbol{x}_0$ that can be driven to zero.

Moreover we can observe that if the matrix $\mathbf{A}$ of the system is non singular for all times $t \in [t_0,t_1]$, then the state transition matrix in invertible and we can alternatively define the controllable subspace as

$$\mathcal{C}[t_0,t_1] = \left\{ \boldsymbol{x}_0 \ : \ \exists \boldsymbol{v}(t) = -\boldsymbol{u}(t) \text{ such that } \boldsymbol{x}_0 = \sum_{\tau=0}^{t_1-1} \boldsymbol{\Phi}(t_0,\tau+1)\mathbf{B}(\tau)\boldsymbol{v}(\tau) \right\}$$

The **discrete-time Gramians** are equivalent to the continuous-time ones, with the exception of interchanging the integral with a summation, resulting in

$$\begin{aligned} \mathbf{W}_{\mathcal{R}}[t_0,t_1] &= \sum_{\tau=0}^{t_1-1} \boldsymbol{\Phi}(t_1,\tau+1)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\boldsymbol{\Phi}^T(t_1,\tau+1) \\ \mathbf{W}_{\mathcal{C}}[t_0,t_1] &= \sum_{\tau=0}^{t_1-1} \boldsymbol{\Phi}(t_0,\tau+1)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\boldsymbol{\Phi}^T(t_0,\tau+1) \end{aligned} \tag{4.17}$$

We have to pay attention to the fact that the controllability Gramian uses the *backward in time* definition of the state transition matrix $\boldsymbol{\Phi}$: this operation can be performed if and only if $\mathbf{A}$ is non-singular in the domain $[t_0,t_1]$, otherwise the Gramian cannot be determined (and as consequence the state cannot be controlled).

As in the continuous-time case, it still holds

$$\text{Im}\left\{\mathbf{W}_{\mathcal{R}}[t_0,t_1]\right\} = \mathcal{R}[t_0,t_1] \qquad \text{Im}\left\{\mathbf{W}_{\mathcal{C}}[t_0,t_1]\right\} = \mathcal{C}[t_0,t_1] \qquad\qquad \forall t_1 > t_0 \geq 0$$

and considering a reachable (controllable) state $\boldsymbol{x}_1 = \mathbf{W}_{\mathcal{R}}\boldsymbol{\eta}_1 \in \mathcal{R}$ ($\boldsymbol{x}_0 = \mathbf{W}_{\mathcal{C}}\boldsymbol{\eta}_0 \in \mathcal{C}$), then the control $\boldsymbol{u}^*(t) = \mathbf{B}^T(t)\boldsymbol{\Phi}^T(t_1,t+1)\boldsymbol{\eta}_1$ ($\boldsymbol{u}^*(t) = -\mathbf{B}^T\boldsymbol{\Phi}^T(t_0,t+1)\boldsymbol{\eta}_0$) moves $\boldsymbol{x}_0 = 0$ ($\boldsymbol{x}_0 \neq 0$) into $\boldsymbol{x}_1 \neq 0$ ($\boldsymbol{x}_1 = 0$) with the minimum energy.

Time-invariant case Considering now a discrete-time LTI case, we have that the state transition matrix $\boldsymbol{\Phi}(t_1,\tau)$ collapses to the computation of a matrix power $\mathbf{A}^{t_1-\tau}$: this so simplifies the discrete-time Gramians to

$$\mathbf{W}_{\mathcal{R}}[t_0,t_1] = \sum_{s=0}^{t_1-t_0-1} \mathbf{A}^s\mathbf{B}\mathbf{B}^T\left(\mathbf{A}^T\right)^s \qquad\qquad \mathbf{W}_{\mathcal{C}}[t_0,t_1] = \sum_{s=0}^{t_1-t_0-1} \mathbf{A}^{-s-1}\mathbf{B}\mathbf{B}^T\left(\mathbf{A}^T\right)^{-s-1} \tag{4.18}$$

Theorem 4.2 still holds if and only if we assume that $\mathbf{A}$ is always invertible in $[t_0,t_1]$ (in order to have a proper definition of the controllability set) and if we impose that $t_1 \geq t_0 + n$: this is a very important bound limiting the time scaling property and is due to the fact that to reach all possible state we have to ensure, by Cayley-Hamilton, that all matrix power up to the n-th order exists.

4.4 Full-state feedback and single-input eigenvalue assignment

Given a **LTI** system of which we can measure all his state $\boldsymbol{x}$, then we can assign as input $\boldsymbol{u}$ a linear combination of them in the form $\boldsymbol{u} = -\mathbf{K}\boldsymbol{x}$, where $\mathbf{K}$ is a $m \times n$ matrix; note that this preamble is

the same as the one of the linear quadratic regulator described in page 22. With this imposition we obtain a new **dynamic equation** of the system that's called *in **full-state feedback***:

$$\dot{\boldsymbol{x}} = \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u} = \mathbf{A}\boldsymbol{x} - \mathbf{B}\mathbf{K}\boldsymbol{x} = \big(\mathbf{A} - \mathbf{B}\mathbf{K}\big)\boldsymbol{x} \tag{4.19}$$

Considering now a **single-input controllable system** (for which $\mathcal{R} = \mathcal{C} = \mathbb{R}^n$) characterized by the pair of matrices $\mathbf{A}, \mathbf{B}$ (that are the lonely one relevant for the computation of the reachability matrix $\mathbf{R}$), then for each set of **desired eigenvalues** $\lambda_1, \dots, \lambda_n$ there always exists a $1 \times n$ matrix $\mathbf{K}$ such that

$$\mathbf{A}_{cl} = \mathbf{A} - \mathbf{B}\mathbf{K}$$

has those eigenvalues; usually $\mathbf{A}_{cl}$ is referred as the **closed-loop matrix**. The goal of the next paragraph is to describe a *procedure* for the so called **single-input eigenvalue assignment**.

Controllable canonical form Given a controllable LTI system with characteristic polynomial of the form

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

at page 5 we introduced the controllable canonical form whose matrices $\mathbf{A}_{ctr}, \mathbf{B}_{ctr}$ realizing the system were

$$\mathbf{A}_{ctr} = \begin{bmatrix} 0 & & & \\ \vdots & & \mathbf{I}_{(n-1)\times(n-1)} & \\ 0 & & & \\ -\alpha_0 & -\alpha_1 & \dots & -\alpha_{n-1} \end{bmatrix} \qquad \mathbf{B}_{ctr} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Knowing that the characteristic polynomial of the closed-loop matrix can be regarded as

$$p_{\mathbf{A}_{cl}}(s) = \big(s - \lambda_1\big) \ddots \big(s - \lambda_n\big) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0$$

Equating the controllable canonical form of the closed-loop matrix $\mathbf{A}_{cl}$ with the system $\mathbf{A} - \mathbf{B}\mathbf{K}$ allows to obtain a simple relation to compute the matrix $\mathbf{K}$, because as we can see expanding the product $\mathbf{B}\mathbf{K}$ all we obtain is

$$\begin{bmatrix} 0 & & & \\ \vdots & & \mathbf{I} & \\ 0 & & & \\ -\beta_0 & -\beta_1 & \dots & -\beta n - 1 \end{bmatrix} = \begin{bmatrix} 0 & & & \\ \vdots & & \mathbf{I} & \\ 0 & & & \\ -\alpha_0 - k_0 & -\alpha_1 - k_1 & \dots & -\alpha_{n-1} - k_{n-1} \end{bmatrix}$$

If we considered so the controllable form of the system we could have easily determined the matrix $\mathbf{K}$ as

$$\mathbf{K} = \begin{bmatrix} \beta_0 - \alpha_0 & \beta_1 - \alpha_1 & \dots & \beta_n - \alpha_n \end{bmatrix} \tag{4.20}$$

More general case In the more general case $\mathbf{A}, \mathbf{B}$ are not in controllable canonical form, but are *generic* full matrices; however we can observe that the **reachability matrix** acts as a **similarity transformation** to transform any system into it's **observable canonical form**

$$\mathbf{A}_{ob} = \begin{bmatrix} 0 & \dots & 0 & -\alpha_0 \\ & & & -\alpha_1 \\ & \mathbf{I}_{(n-1)\times(n-1)} & & \vdots \\ & & & -\alpha_{n-1} \end{bmatrix} \qquad \mathbf{B}_{ob} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where the transformed matrices $\mathbf{A}_{ob}, \mathbf{B}_{ob}$ are obtained through the algebraic equivalence similar to (1.9) in the form

$$\mathbf{A}_{ob} = \mathbf{R}^{-1}\mathbf{A}\mathbf{R} \qquad \mathbf{B}_{ob} = \mathbf{R}^{-1}\mathbf{B} \tag{4.21}$$

Proof 4.4: To prove the effectiveness of (4.21) we can show that indeed $\mathbf{RA}_{ob} = \mathbf{AR}$. Computing the product $\mathbf{RA}_{ob}$ considering the block matrices in the form

$$\mathbf{RA}_{ob} = \left[\begin{array}{cccc} \mathbf{B} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B} \end{array}\right] \left[\begin{array}{c|c} 0 & -\alpha_0 \\ \hline & -\alpha_1 \\ \mathbf{I} & \vdots \\ & -\alpha_{n-1} \end{array}\right] = \left[\begin{array}{cccc|c} \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B} & \boldsymbol{x} \end{array}\right]$$

where the vector $\boldsymbol{x}$ is the linear combination defined as $-\alpha_0\mathbf{B} - \alpha_1\mathbf{AB} - \cdots - \alpha_{n-1}\mathbf{A}^{n-1}\mathbf{B}$; observe that for SISO system all products $\mathbf{A}^i\mathbf{B}$ evaluates to a vector. Collecting $\mathbf{B}$ we have $\boldsymbol{x} = (-\alpha_0\mathbf{I} - \alpha_1\mathbf{A} - \cdots - \alpha_{n-1}\mathbf{A}^{n-1})\mathbf{B}$ and recalling the Cayley-Hamilton theorem (page 12) we can consider the term in parenthesis exactly as $\mathbf{A}^n$, so we can rewrite

$$\mathbf{RA}_{ob} = \begin{bmatrix}\mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B} & \mathbf{A}^n\mathbf{B}\end{bmatrix} = \mathbf{A}\begin{bmatrix}\mathbf{B} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B}\end{bmatrix} = \mathbf{AR}$$

Similarly, by performing the matrix multiplication, we can show that $\mathbf{B}_{ob} = \mathbf{R}^{-1}\mathbf{B}$ by simply proving that $\mathbf{RB}_{ob} = \mathbf{B}$.

Moreover for any observable canonical form exists a similarity transformation matrix $\mathbf{M}$ that allows to obtain the controllable canonical form as $\mathbf{M}^{-1}\mathbf{A}_{ob}\mathbf{M} = \mathbf{A}_{ctr}$ and $\mathbf{M}^{-1}\mathbf{B}_{ob} = \mathbf{B}_{ctr}$. In particular $\mathbf{M}$ is a symmetric matrix in the form form

$$\mathbf{M} = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{n-1} & 1 \\ \alpha_2 & & & \iddots & \\ \vdots & & \iddots & & \\ \alpha_{n-1} & \iddots & & & \\ 1 & & & & \end{bmatrix} \tag{4.22}$$

Proof 4.5: To show that (4.22) is the similarity transformation transforming the observable into the canonical form we can start computing the product $\mathbf{MA}_{ctr}$ considering the block matrices

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

where $\mathbf{M}_{21}$ is the $(n-1)\times(n-1)$ submatrix of $\mathbf{M}$ and $\boldsymbol{\alpha}$ is the column vector $(\alpha_1, \dots, \alpha_{n-1})$. We observe that the result of the product is a symmetric matrix, so we can say that

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

proving the algebraic equivalence between the two representations.

Combining what has been said so far, considering the transformation matrix $T = \mathbf{RM}$ allows to directly compute the controllable canonical form of any pair $(\mathbf{A}, \mathbf{B})$ as

$$\mathbf{T}^{-1}\mathbf{AT} = \mathbf{A}_{ctr} \qquad \mathbf{T}^{-1}\mathbf{B} = \mathbf{B}_{ctr}$$

We can so summarize the procedure for the **single-input eigenvalue assignment** as

1. given the original pair $(\mathbf{A}, \mathbf{B})$ and the desired eigenvalues λ_i of the closed-loop system, compute their characteristic polynomials in the form

$$p_{\mathbf{A}}(s) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0 \qquad p_{\mathbf{A}_{cl}}(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0$$

2. compute the feedback matrix $\mathbf{K}_{ctr}$ in the controllable canonical form as

$$\mathbf{K}_{ctr} = \begin{bmatrix}\beta_0 - \alpha_0 & \dots & \beta_n - \alpha_n\end{bmatrix}$$

3. build the controllability matrix $\mathbf{R}$ (4.13) and the transformation $\mathbf{M}$ as in (4.22) and compute $\mathbf{T} = \mathbf{RM}$ and it's inverse;
4. determine the feedback matrix $\mathbf{K}$ for the initial pair $(\mathbf{A}, \mathbf{B})$ as

$$\mathbf{K} = \mathbf{K}_{ctr}\mathbf{T}^{-1}$$

As final check one can verify that indeed $\mathbf{A} - \mathbf{BK}$ gives the desired eigenvalues.

Ackerman method Another way to compute the feedback matrix $\mathbf{K}$ is through the **Ackerman equation** defined as

$$\mathbf{K} = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix} \mathbf{R}^{-1} p_{\mathbf{A}_{cl}}(\mathbf{A}) \tag{4.23}$$

4.5 Controllable systems

Given a **linear system** Σ with dynamics $\dot{\boldsymbol{x}}/\boldsymbol{x}^+ = \mathbf{A}(t)\boldsymbol{x} + \mathbf{B}(t)\boldsymbol{u}$, then given two times $t_1 > t_0 \geq 0$ we say that the pair $(\mathbf{A}, \mathbf{B})$ is **reachable** on the time interval $[t_0, t_1]$ if the reachable subspace has *maximum dimension*, meaning $\mathcal{R}[t_0, t_1] = \mathbb{R}^n$; similarly the pair $(\mathbf{A}, \mathbf{B})$ is **controllable** if $\mathcal{C}[t_0, t_1] = \mathbb{R}^n$. The underlying idea of this definition is that for such system we can choose particular inputs $u(t)$ that allows us to reach/control all possible state of Σ in $\mathbb{R}^n$.

For now on, where not otherwise specified, this section will consider the simplified condition of **linear time-invariant** system (and for discrete-time sys. we also assume that $\mathbf{A}$ is invertible) in order to have the coincidence between controllability and reachability.

For LTI systems the condition of controllability/reachability collapses to the study of the rank of the controllability matrix, in particular

$$\text{controllable sys.} \quad \Leftrightarrow \quad \text{reachable sys.} \quad \Leftrightarrow \quad \text{rank}\{\mathbf{R}\} = n \tag{4.24}$$

Proof 4.6: Knowing that $\mathbf{R}$ is a $n \times mn$ matrix, if it happens that n of it's rows are linearly independent then we are sure that $\text{Im}\{\mathbf{R}\} = \mathbb{R}^n$: this furthermore implies $\text{Im}\{\mathbf{W}_{\mathcal{R}}[t_0, t_1]\} = \mathbb{R}^n$ probing the controllability of the system.

The intuition behind the idea of ***controllable system*** is that *the input can reach the dynamic of all the states*. Considering as example the system

$$\begin{cases} \dot{x}_1 = x_1 \\ \dot{x}_2 = -2x_2 + x_1 + u \end{cases}$$

we can intuitively say that the system is not controllable: the input u in fact affects only the dynamic of the second state $\boldsymbol{x}_1$, while there's no way to act on the state x_1 (that's indeed autonomous). We can rigorously prove this intuition by build the matrices $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & -2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$: once we compute $\mathbf{R} = \begin{bmatrix} 0 & 0 \\ 1 & -2 \end{bmatrix}$ we see that the controllability matrix is not full rank, hence the system is not controllable.

Considering now instead the similar system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -2x_2 + x_1 + u \end{cases}$$

we see that u cannot directly modify the dynamic of the first state, however such dynamic can be driven controlling the state x_2 (appearing in the right-hand side of the first dynamic equation), thus there's a way to compute both states. Having this time $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ the reachability matrix $\mathbf{R} = \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix}$ has full rank, proving that the system is controllable.

4.5.1 Controllability tests

The scope now is to determine some *tests* that can be used to prove the controllability/reachability (or not) of the system without the need of computing the reachability matrix $\mathbf{R}$ (that in general for *big* system can be numerically expensive).

In order to do so, let's recall one concept from linear algebra: given a $n \times n$ matrix $\mathbf{A}$, the linear subspace $\mathcal{V} \subseteq \mathbb{R}^n$ is said **A-invariant** if $\forall \boldsymbol{v} \in \mathcal{V}$ it holds that $\mathbf{A}\boldsymbol{v} \in \mathcal{V}$. Subsequent properties are that

i) given a subspace $\mathcal{V} \neq \{0\}$ for which we can define a matrix $\mathbf{V} \in \mathbb{R}^{n \times k}$ whose k columns are a basis of $\mathbb{V}$, then there exists a matrix $\overline{\mathbf{A}} \in \mathbb{R}^{k \times k}$ such that

$$\mathbf{A}\boldsymbol{V} = \boldsymbol{V}\overline{\mathbf{A}}$$

Calling in fact $\boldsymbol{v}_i \in \mathbb{R}^n, i = 1 \dots k$, the columns of $\mathbf{V}$, is the associate linear space $\mathcal{V}$ is $\mathbf{A}$-invariant then it means that also $\mathbf{A}\boldsymbol{v}_i \in \mathcal{V}$ for any i. At the same time we have that $\mathbf{A}\boldsymbol{v}_i$ can be regarded as a linear combination $\overline{a}_{1i}\boldsymbol{v}_i + \overline{a}_{2i}\boldsymbol{v}_2 + \dots + \overline{a}_{ki}\boldsymbol{v}_k$, so

$$\mathbf{A}\begin{bmatrix}\boldsymbol{v}_1 & \dots & \boldsymbol{v}_k\end{bmatrix} = \begin{bmatrix}\boldsymbol{v}_1 & \dots & \boldsymbol{v}_k\end{bmatrix}\begin{bmatrix}\overline{a}_{11} & \dots & \overline{a}_{1k} \\ \vdots & \ddots & \\ \overline{a}_{k1} & & \overline{a}_{kk}\end{bmatrix}$$

ii) given $\mathbf{A} \in \mathbb{R}^{n \times n}$ and an $\mathbf{A}$-invariant linear subspace $\mathcal{V}$, then the matrix $\mathbf{V}$ (made by the basis of $\mathcal{V}$) contains at least one eigenvalue of $\mathbf{A}$.

Considering in fact the eigenvalue-eigenvector pair $(\lambda, \overline{\boldsymbol{v}})$ of the matrix $\overline{\mathbf{A}}$, so satisfying $\overline{\mathbf{A}}\overline{\boldsymbol{v}} = \lambda\overline{\boldsymbol{v}}$, then if $\mathcal{V}$ is $\mathbf{A}$-invariant for *i)* we equivalently have $\mathbf{A}\mathbf{V}\overline{\boldsymbol{v}} = \mathbf{V}\overline{\mathbf{A}}\overline{\boldsymbol{v}}$; since $\overline{\boldsymbol{v}}$ is an eigenvector of $\overline{\mathbf{A}}$, then such equality can be rewritten as

$$\mathbf{A}\mathbf{V}\overline{\boldsymbol{v}} = \lambda\mathbf{V}\overline{\boldsymbol{v}}$$

Substituting the vector $\mathbf{V}\overline{\boldsymbol{v}}$ with the variable $\boldsymbol{x}$, we finally reach the *standard form* of the eigenvalue statement $\mathbf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, proving so that λ is also an eigenvector of $\mathbf{A}$.

Eigenvector test The **eigenvector test for controllability** states that *a LTI system is controllable if and only if all the eigenvectors $\boldsymbol{x}$ of the matrix $\mathbf{A}^T$ are not in the kernel of $\mathbf{B}^T$.*

Proof 4.7: The proof of such test can be performed in two step: firstly by showing that if the pair $(\mathbf{A}, \mathbf{B})$ is controllable, then the eigenvector of $\mathbf{A}^T$ are not in the kernel of $\mathbf{B}^T$ and then the reversed statement:

a) By contradiction let's assume that exists an eigen-pair $(\boldsymbol{x}, \lambda) \neq 0$ such that $\mathbf{A}^T\boldsymbol{x} = \lambda\boldsymbol{x}$ and for which $\mathbf{B}^T\boldsymbol{x} = 0$, exploiting the definition (4.13) of the controllability matrix transposed, then we have that

$$\mathbf{R}^T\boldsymbol{x} = \begin{bmatrix}\mathbf{B}^T \\ \mathbf{B}^T\mathbf{A}^T \\ \vdots \\ \mathbf{B}^T\left(\mathbf{A}^{n-1}\right)^T\end{bmatrix}\boldsymbol{x} = \begin{pmatrix}\mathbf{B}^T\boldsymbol{x} \\ \mathbf{B}^T\mathbf{A}^T\boldsymbol{x} \\ \vdots \\ \mathbf{B}^T\left(\mathbf{A}^{n-1}\right)^T\boldsymbol{x}\end{pmatrix} = \begin{pmatrix}0 \\ \mathbf{B}^T\lambda\boldsymbol{x} \\ \vdots \\ \mathbf{B}^T\lambda^{n-1}\boldsymbol{x}\end{pmatrix} = \begin{pmatrix}0 \\ \lambda\mathbf{B}^T\boldsymbol{x} \\ \vdots \\ \lambda^{n-1}\mathbf{B}^T\boldsymbol{x}\end{pmatrix} = \begin{pmatrix}0 \\ 0 \\ \vdots \\ 0\end{pmatrix}$$

Shown that $\boldsymbol{x}$ is in the kernel of $\mathbf{R}^T$, then we are sure that $\dim \mathrm{Ker}\left\{\mathbf{R}^T\right\} \geq 1$ thus for the fundamental theorem of linear algebra we can also say that

$$\dim \mathrm{Im}\left\{\mathbf{R}^T\right\} < n$$

However as initial hypothesis, having the pair $(\mathbf{A}, \mathbf{B})$ controllable requires a reachability matrix that's full rank, so it must have $\dim \mathrm{Im}\left\{\mathbf{R}^T\right\} = \dim \mathrm{Im}\left\{\mathbf{R}\right\} = \mathrm{rank}\{\mathbf{R}\}$: this arise the contradiction proving so that if $\boldsymbol{x}$ in an eigenvector of $\mathbf{A}^T$, then surely it cannot belong to the kernel of $\mathbf{B}^T$.

b) Let us consider now the eigenvector $\boldsymbol{x}$ satisfying $\mathbf{A}^T\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\mathbf{B}^T\boldsymbol{x} \neq 0$, then we can prove by contradiction that the pair $(\mathbf{A}, \mathbf{B})$ is controllable. Assuming in fact that $(\mathbf{A}, \mathbf{B})$ is not controllable, by (4.24) we can say that $\text{rank}\{\mathbf{R}\} = \text{rank}\{\mathbf{R}^T\} < n$, this by the fundamental theorem of linear algebra $\dim \text{Ker}\left\{\mathbf{R}^T\right\} > 0$.
It happens that $\ker \mathbf{R}^T$ is $\mathbf{A}^T$-invariant (meaning $\mathbf{R}^T\boldsymbol{x} = \mathbf{R}^T\mathbf{A}^T\boldsymbol{x} = 0$), because considering

$$\mathbf{R}^T\boldsymbol{x} = \begin{bmatrix} \mathbf{B}^T \\ \mathbf{B}^T\mathbf{A}^T \\ \vdots \\ \mathbf{B}^T\left(\mathbf{A}^{n-1}\right)^T \end{bmatrix}\boldsymbol{x} = \begin{pmatrix} \mathbf{B}^T\boldsymbol{x} \\ \mathbf{B}^T\mathbf{A}^T\boldsymbol{x} \\ \vdots \\ \mathbf{B}^T\left(\mathbf{A}^{n-1}\right)^T\boldsymbol{x} \end{pmatrix} = 0$$

if further implies

$$\mathbf{R}^T\mathbf{A}^T\boldsymbol{x} = \begin{pmatrix} \mathbf{B}^T\mathbf{A}^T\boldsymbol{x} \\ \mathbf{B}^T(\mathbf{A}^2)^T\boldsymbol{x} \\ \vdots \\ \mathbf{B}^T\left(\mathbf{A}^n\right)^T\boldsymbol{x} \end{pmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{B}^T(\mathbf{A}^n)^T \end{bmatrix}\boldsymbol{x} = 0$$

Using Cayley-Hamilton theorem, page 12, we can in fact decompose the last block of the matrix as

$$\mathbf{B}^T(\mathbf{A}^n)^T\boldsymbol{x} = \mathbf{B}^T\left(\sum_{k=0}^{n-1}\alpha_k\mathbf{A}^k\right)^T\boldsymbol{x} = \sum_{k=0}^{n-1}\alpha_k\mathbf{B}^T(\mathbf{A}^k)^T\boldsymbol{x} = 0$$

Showed that $\text{Ker}\left\{\mathbf{R}^T\right\}$ is $\mathbf{A}^T$-invariant, then $\mathbf{A}^T\boldsymbol{x} \in \text{Ker}\left\{\mathbf{R}^T\right\}$ contains al least one eigenvector of $\mathbf{A}^T$ (due to property *ii)* of the $\mathbf{A}$-invariant definition). Since $\mathbf{R}^T\boldsymbol{x} = 0$ implies $\mathbf{B}^T\boldsymbol{x} = 0$, this contradicts the initial statement for which we should have had $\mathbf{B}^T\boldsymbol{x} \neq 0$, meaning that $(\mathbf{A}, \mathbf{B})$ cannot be non-controllable (hence must be controllable).

Popov-Belevitch-Hotus PBH test Another method that can be used to prove controllability of the system is the so called **PBH test**, named by its invetor Popov-Belevitch-Hotus, stating that *the pair* $(\mathbf{A}, \mathbf{B})$ *is controllable if and only if the rank of the matrix* $\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I} & \mathbf{B}\end{bmatrix}$ *is equal to n for any value of* λ*:*

$$(\mathbf{A}, \mathbf{B}) \text{ controllable} \quad \Leftrightarrow \quad \text{rank}\{\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I} & \mathbf{B}\end{bmatrix}\} = n \quad \forall\lambda \in \mathbb{C} \tag{4.25}$$

Note that for almost all values of λ condition (4.25) is always satisfied (because $\mathbf{A}$ is always full-rank), but the only *problematic points* are the eigenvalues λ of $\mathbf{A}$ that are making the matrix $\mathbf{A} - \lambda\mathbf{I}$ rank-deficient.

Proof 4.8: The PBH test can be proven by applying the fundamental theorem of linear algebra to the matrix $\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I} & \mathbf{B}\end{bmatrix}^T$ and requiring that the dimension of it's kernel must be null:

$$\dim \text{Ker}\left\{\begin{bmatrix}\mathbf{A}^T - \lambda\mathbf{I} \\ \mathbf{B}^T\end{bmatrix}\right\} = n - \text{rank}\{\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I} & \mathbf{B}\end{bmatrix}\} = 0$$

This means that

$$\begin{bmatrix}\mathbf{A}^T - \lambda\mathbf{I} \\ \mathbf{B}^T\end{bmatrix}\boldsymbol{x} = \begin{bmatrix}\mathbf{A}^T\boldsymbol{x} - \lambda\mathbf{I}\boldsymbol{x} \\ \mathbf{B}^T\boldsymbol{x}\end{bmatrix} \neq 0 \qquad \forall\boldsymbol{x} \neq 0, \forall\lambda \in \mathbb{C}$$

In the particular case when $(\lambda, \boldsymbol{x})$ is a eigen-pair of $\mathbf{A}^T$, so such that $\mathbf{A}^T\boldsymbol{x} = \lambda\boldsymbol{x}$, in order to have a non-zero matrix we must have $\mathbf{B}^T\boldsymbol{x} \neq 0$: each eigenvector of $\mathbf{A}^T$ so do not belong to $\ker \mathbf{B}^T$ and so for the eigenvalue test previously shown implies controllability.

Lyapunov controllability test Given a continuous/discrete-time LTI system characterized by a Hurwitz/Schur matrix $\mathbf{A}$ (so assuming that the system is exponentially stable), then the pair $(\mathbf{A}, \mathbf{B})$ is controllable if and only if exists a symmetric positive-definite matrix $\mathbf{W} = \mathbf{W}^T > 0$ such that

$$(CT): \quad \mathbf{A}\mathbf{W} + \mathbf{W}\mathbf{A}^T = -\mathbf{B}\mathbf{B}^T \qquad\qquad (DT): \quad \mathbf{A}\mathbf{W}\mathbf{A}^T - \mathbf{W} = -\mathbf{B}\mathbf{B}^T \tag{4.26}$$

Moreover the unique solution $\mathbf{W}$ can be computes as $\lim_{t_1-t_0\to\infty}\mathbf{W}_{\mathcal{R}}[t_0,t_1]$, so

$$(CT): \quad \mathbf{W}=\int_0^\infty e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T t}\,dt \qquad\qquad (DT): \quad \mathbf{W}=\sum_{\tau=0}^{\infty}\mathbf{A}^\tau\mathbf{B}\mathbf{B}^T(\mathbf{A}^T)^\tau \tag{4.27}$$

Proof 4.9: To prove the Lyapunov test for controllability we have to recall that the norm squared of a complex vector $\boldsymbol{v}\in\mathbb{C}^n$ can be computed as $|\boldsymbol{v}|^2=\boldsymbol{v}^{*T}\boldsymbol{v}$. With this premise, we can firstly show that if (4.26) holds then $(\mathbf{A},\mathbf{B})$ is controllable and then vice-versa:

a) We show now that if exists $\mathbf{W}$ positive-definite satisfying (4.26), then $(\mathbf{A},\mathbf{B})$ is controllable. Exploiting the eigenvector test, if $(\lambda,\boldsymbol{x})$ is an eigen-pair of $\mathbf{A}^T$ then it implies $\mathbf{B}^T\boldsymbol{x}=0$. By pre-multiplying by $\boldsymbol{x}^*$ and post-multiplying by $\boldsymbol{x}$ equation 4.26 results in

$$\begin{aligned}\boldsymbol{x}^*\left(\mathbf{A}\mathbf{W}+\mathbf{W}\mathbf{A}^T\right)\boldsymbol{x}=\boldsymbol{x}^*\mathbf{A}\mathbf{W}\boldsymbol{x}+\boldsymbol{x}^*\mathbf{W}\mathbf{A}^T\boldsymbol{x}=\left(\mathbf{A}^T\boldsymbol{x}\right)^*\mathbf{W}\boldsymbol{x}+\boldsymbol{x}^*\mathbf{W}\left(\mathbf{A}^T\boldsymbol{x}\right)=-\boldsymbol{x}^*\mathbf{B}\mathbf{B}^T\boldsymbol{x}\\ \left(\lambda\boldsymbol{x}\right)^*\mathbf{W}\boldsymbol{x}+\boldsymbol{x}^*\mathbf{W}\left(\lambda\boldsymbol{x}\right)=\lambda^*\boldsymbol{x}^*\mathbf{W}\boldsymbol{x}+\lambda\boldsymbol{x}^*\mathbf{W}\boldsymbol{x}=-|\mathbf{B}^T\boldsymbol{x}|^2\\ 2\mathbb{R}\mathrm{e}\{\lambda\}\boldsymbol{x}^*\mathbf{W}\boldsymbol{x}=\end{aligned}$$

With the hypothesis of $\mathbf{A}$ being Hurwitz, then it means that $\mathbb{R}\mathrm{e}\{\lambda\}<0$; moreover having $\mathbf{W}$ positive-definite tells us that the overall term $2\mathbb{R}\mathrm{e}\{\lambda\}\boldsymbol{x}^*\mathbf{W}\boldsymbol{x}$ is negative for each eigenvector $\boldsymbol{x}\neq 0$ of $\mathbf{A}^T$ (with eigenvalue λ). This means that the norm $|\mathbf{B}^T\boldsymbol{x}|^2$ is non-zero for all $\boldsymbol{x}\neq 0$, so by the eigenvector test this implies that the pair $(\mathbf{A},\mathbf{B})$ is controllable.

b) We show now that if the pair $(\mathbf{A},\mathbf{B})$ is controllable, then exists a positive-definite matrix $\mathbf{W}$ satisfying (4.26), Calling the product $\mathbf{B}\mathbf{B}^T$ as $\overline{\mathbf{Q}}$ and $\mathbf{A}^T=\overline{\mathbf{A}}$, what we obtain is a formulation in the form

$$\mathbf{W}\overline{\mathbf{A}}+\overline{\mathbf{A}}^T=-\overline{\mathbf{Q}}$$

This formulation coincides with the Lyapunov equality (3.11), page 19, from which we can compute the solution $\mathbf{W}$ as

$$\mathbf{W}=\int_{t_0}^{t_1}e^{\overline{\mathbf{A}}t}\overline{\mathbf{Q}}e^{\overline{\mathbf{A}}^T t}\,dt$$

This solution is based upon the fact that $\overline{\mathbf{Q}}$ is positive-definite, condition that until now wasn't ensured but it holds so by the assumption that $(\mathbf{A},\mathbf{B})$ is a controllable pair. Considering in fact

$$\boldsymbol{x}^T\mathbf{W}\boldsymbol{x}=\int_0^\infty\boldsymbol{x}^T e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T t}\boldsymbol{x}\,dt=\int_0^\infty\left|\mathbf{B}^T e^{\mathbf{A}^T t}\boldsymbol{x}\right|^2 dt\geq 0$$

proving so that the so obtained matrix $\mathbf{W}$ is positive definite.

Theorem 4.3: Given an autonomous LTI system $\dot{\boldsymbol{x}}/\boldsymbol{x}^+=\mathbf{A}\boldsymbol{x}$m then the following condition-s/statements are equivalent:

i) the system is asymptotically stable;

ii) the system is exponentially stable;

iii) matrix $\mathbf{A}$ is Hurwitz/Schur (meaning that there aren't *bad eigenvalues*);

iv) for every symmetric positive-definite matrix $\mathbf{Q}=\mathbf{Q}^T>0$ there exists a matrix $\mathbf{P}$ such that

$$(CT): \quad \mathbf{A}^T\mathbf{P}+\mathbf{P}\mathbf{A}=-\mathbf{Q} \qquad\qquad (DT): \quad \mathbf{A}^T\mathbf{P}\mathbf{A}-\mathbf{P}=\mathbf{Q}$$

Moreover the unique solution of the problem is described by $\mathbf{P}=\int_0^\infty e^{\mathbf{A}^T t}\mathbf{Q}e^{\mathbf{A}t}\,dt$ for continuous-time systems and $\mathbf{P}=\sum_{k=0}^\infty\left(\mathbf{A}^T\right)^k\mathbf{Q}\mathbf{A}^k$ for the discrete-time case;

v) always exists a symmetric positive-definite matrix $\mathbf{P}=\mathbf{P}^T>0$ such that

$$(CT): \quad \mathbf{A}^T\mathbf{P}+\mathbf{P}\mathbf{A}<0 \qquad\qquad (DT): \quad \mathbf{A}^T\mathbf{P}\mathbf{A}-\mathbf{P}<0$$

vi) for each matrix $\mathbf{B}$ such that $(\mathbf{A}, \mathbf{B})$ is a controllable pair exists a positive-definite matrix $\mathbf{W} = \mathbf{W}^T > 0$ solving

$$(CT): \quad \mathbf{AW} + \mathbf{WA}^T = -\mathbf{BB}^T \qquad\qquad (DT): \quad \mathbf{AWA}^T - \mathbf{W} = -\mathbf{BB}^T$$

The unique solution of $\mathbf{W}$ is presented in (4.27).

4.5.2 Feedback stabilization with Lyapunov test

Considering a controllable continuous-time LTI characterized by a controllable pair $(\mathbf{A}, \mathbf{B})$, we can show also the system $(\mathbf{B}, -\mu\mathbf{I} - \mathbf{A})$ is controllable for any given μ. We have in fact that both $-\mu\mathbf{I} - \mathbf{A}$ and $\mathbf{A}$ are sharing an eigenvector: given the eigen-pair $(\lambda, \boldsymbol{x})$ of $\mathbf{A}$ for which $\mathbf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, subtructing $-\mu\mathbf{I}\boldsymbol{x}$ on both sides evaluates to the expression

$$(-\mu\mathbf{I} - \mathbf{A})\boldsymbol{x} = (-\lambda - \mu)\boldsymbol{x}$$

This shows that $\boldsymbol{x}$ is an eigenvector of both $\mathbf{A}$ and $-\mu\mathbf{I} - \mathbf{A}$ (while their eigenvalues are different and are λ and $-\mu - \lambda$ respectively) and thus in both cases it satisfies the requirement $\mathbf{B}^T\boldsymbol{x} \neq 0$ (having initially impose that $(\mathbf{A}, \mathbf{B})$ is controllable). Picking a coefficient μ *large enough* allows us to *shift* all the eigenvalues of $\mathbf{A}$ into the *good region* for the eigenvalues (in particular there are infinite values of μ for which $-\mu\mathbf{I} - \mathbf{A}$ is Hurwitz); having that this system must satisfy the (4.26), so $(-\mu\mathbf{I} - \mathbf{A})\mathbf{W} + \mathbf{W}(-\mu\mathbf{I} - \mathbf{A})^T = -\mathbf{BB}^T$, we can also observe by computing the products that

$$\mathbf{AW} + \mathbf{WA}^T + \mathbf{BB}^T = -2\mu\mathbf{W}$$

Defining $\mathbf{P} = \mathbf{W}^{-1}$ and pre- and post-multiplying this expression by $\mathbf{P}$ gives

$$\mathbf{PA} + \mathbf{A}^T\mathbf{P} - \mathbf{PBB}^T\mathbf{P} = -2\mu\mathbf{P}$$

Defining now with $\mathbf{K} = \frac{\mathbf{B}^T\mathbf{P}}{2}$ we can observe that the term $\mathbf{PBB}^T\mathbf{P}$ can be rewritten as $\frac{(\mathbf{B}^T\mathbf{P})^T}{2}\mathbf{B}^T\mathbf{P} + \mathbf{PB}\frac{\mathbf{B}^T\mathbf{P}}{2} = \mathbf{K}^T\mathbf{B}^T\mathbf{P} + \mathbf{PBK}$, thus collecting $\mathbf{P}$ in the previous equation gives

$$\mathbf{P}(\mathbf{A} - \mathbf{BK}) + (\mathbf{A}^T - \mathbf{K}^T\mathbf{B}^T)\mathbf{P} = \mathbf{P}(\mathbf{A} - \mathbf{BK}) + (\mathbf{A}^T - \mathbf{BK})^T\mathbf{P} = -2\mu\mathbf{P} \qquad (\star)$$

Having both μ and $\mathbf{P}$ positive-definite, when it means that $\mathbf{Q} = 2\mu\mathbf{P}$ is also positive-definite: with this assumption we see that $(\star)$ is indeed in the same form of the Lyapunov equality (3.11) for which by theorem 3.1, page 19, implies that $\mathbf{A} - \mathbf{BK}$ determine an exponentially stable system.

Theorem 4.4: Given a continuous-time LTI system $\dot{\boldsymbol{x}} = \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}$ based on the controllable pair $(\mathbf{A}, \mathbf{B})$, then $\forall\mu > 0$ there exists a full-state feedback $\boldsymbol{u} = -\mathbf{K}\boldsymbol{x}$ that places all eigenvalues of the closed-loop system $\dot{\boldsymbol{x}} = (\mathbf{A} - \mathbf{BK})\boldsymbol{x}$ on the complex semi-plane characterized by $\mathbb{R}\mathrm{e}\{s\} \leq -\mu$. In particular we call μ the **convergence rate** of the system.
Similarly for discrete-time LTI plants $\boldsymbol{x}^+ = \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}$ a pair $(\mathbf{A}, \mathbf{B})$ is controllable if and only if $\forall\mu \in (0, 1)$ there exists a full-state feedback $\boldsymbol{u} = -\mathbf{K}\boldsymbol{x}$ that makes the closed-loop matrix $\mathbf{A}_{cl} = \mathbf{A} - \mathbf{BK}$ Schur (so $|\lambda_i| < 1$ for all eigenvalues).

Note that this two conditions are a double implication, meaning that if the pair is controllable, then we can asses whichever convergence rate we want through a full-state feedback, but also if we can ensure any convergence rate then the pair $(\mathbf{A}, \mathbf{B})$ must have been controllable in the first place.

4.5.3 Controllable decomposition

What we would to do now is to use the **similarity transformations** seen at page 6 to *split* the **controllable states** from the **non-controllable** ones. In this case we consider as transformation the *mapping* of the states $\tilde{\boldsymbol{x}} = \mathbf{T}^{-1}\boldsymbol{x}$ leading to a dynamics in the form

$$\dot{\tilde{\boldsymbol{x}}}/\tilde{\boldsymbol{x}}^+ = \tilde{\mathbf{A}}\tilde{\boldsymbol{x}} + \tilde{\mathbf{B}}\boldsymbol{u} = \mathbf{T}^{-1}\mathbf{AT}\tilde{\boldsymbol{x}} + \mathbf{T}^{-1}\mathbf{B}\boldsymbol{u}$$

Theorem 4.5: The pair $(\mathbf{A}, \mathbf{B})$ is controllable if and only if the pair $(\mathbf{T}^{-1}\mathbf{B}, \mathbf{T}^{-1}\mathbf{AT})$ is controllable for any non-singular $n \times n$ matrix $\mathbf{T}$.

The main takeaway from this theorem is that controllability is a property of the system and do not depend on the realization chosen (if a system is not controllable it doesn't exists a state remapping function that turns it controllable).

Proof 4.10: To prove this theorem we simply need to compute the reachability matrix of the equivalent system that's

$$\begin{aligned}\tilde{\mathbf{R}} &= \begin{bmatrix}\tilde{\mathbf{B}} & \tilde{\mathbf{A}}\tilde{\mathbf{B}} & \dots & \tilde{\mathbf{A}}^{n-1}\tilde{\mathbf{B}}\end{bmatrix} = \begin{bmatrix}\mathbf{T}^{-1}\mathbf{B} & \mathbf{T}^{-1}\mathbf{ATT}^{-1}\mathbf{B} & \dots & \mathbf{T}^{-1}\mathbf{ATT}^{-1}\mathbf{AT}\dots\mathbf{T}^{-1}\mathbf{B}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{T}^{-1}\mathbf{B} & \mathbf{T}^{-1}\mathbf{AB} & \dots & \mathbf{T}^{-1}\mathbf{A}^{n-1}\mathbf{B}\end{bmatrix} = \mathbf{T}^{-1}\begin{bmatrix}\mathbf{A} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B}\end{bmatrix} \\ &= \mathbf{T}^{-1}\mathbf{R}\end{aligned}$$

Being $\mathbf{T}$ (and so $\mathbf{T}^{-1}$) non-singular, then the rank of the matrices after the similar transformations remains unchanged, in particular $\text{rank}\{\tilde{\mathbf{R}}\} = \text{rank}\{\mathbf{R}\} = n$: having that both reachability matrix have maximum rank (being $(\mathbf{A}, \mathbf{B})$ controllable), then surely system $(\mathbf{T}^{-1}\mathbf{B}, \mathbf{T}^{-1}\mathbf{AT})$ is controllable.

Given now a generic LTI system $\dot{\boldsymbol{x}}/\boldsymbol{x}^+ = \mathbf{A}\boldsymbol{x} + \mathbf{B}u$: constructing $\mathbf{V}$ as the $n \times \overline{n}$ matrix show columns are forming a base of the reachable subspace $\mathcal{R}$ of the system (where in particular $\overline{n}$ is the the dimension of the reachable set), then the following holds:

i) the reachable set $\mathcal{R} = \text{Im}\{\mathbf{R}\}$ is $\mathbf{A}$-invariant, meaning that exists a $\overline{n} \times \overline{n}$ matrix $\mathbf{A}_c$ such that

$$\mathbf{AV} = \mathbf{VA}_c \tag{†}$$

ii) the image of the matrix $\mathbf{B}$ is contained in the reachable set, mathematically $\text{Im}\{\mathbf{B}\} \subset \mathcal{R}$, and the columns of $\mathbf{B}$ can be regarded as a linear combination of the columns of $\mathbf{V}$ determined with the multiplication by the $\overline{n} \times m$ matrix $\mathbf{B}_c$:

$$\mathbf{B} = \mathbf{VB}_c$$

Proof 4.11:

a) To prove *i)*, if we consider a vector $\boldsymbol{x} \in \text{Im}\{\mathbf{R}\}$, then it means that exists a vector $\boldsymbol{\beta} \in \mathbb{R}^n$ satisfying $\boldsymbol{x} = \mathbf{R}\boldsymbol{\beta}$, thus

$$\boldsymbol{x} = \mathbf{R}\boldsymbol{\beta}\begin{bmatrix}\mathbf{B} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B}\end{bmatrix}\boldsymbol{\beta} = \sum_{k=0}^{n-1}\mathbf{A}^k\mathbf{B}\beta_k$$

Pre-multiplying everything by $\mathbf{A}$ gives

$$\mathbf{A}\boldsymbol{x} = \mathbf{AR}\boldsymbol{\beta} = \sum_{k=0}^{n-1}\mathbf{A}^{k+1}\mathbf{B}\beta_k = \sum_{k=0}^{n-2}\mathbf{A}^{k+1}\mathbf{B}\beta_k + \mathbf{A}^n\mathbf{B}\beta_{n-1}$$

Exploiting Cayley-Hamilton theorem, page 12, we can rewrite the last term as

$$\mathbf{A}^n\mathbf{B}\beta_{n-1} = \sum_{i=0}^{n-1}\alpha_i\mathbf{A}^i\mathbf{B}\beta_{n-1}$$

Calling now $\tilde{\boldsymbol{\beta}}$ the difference $\boldsymbol{\beta} - \boldsymbol{\alpha}$ allows us to rewrite the product $\mathbf{A}\boldsymbol{x}$ as

$$\mathbf{A}\boldsymbol{x} = \sum_{k=0}^{n-1}\mathbf{A}^k\mathbf{B}\tilde{\beta}_k = \mathbf{R}\tilde{\boldsymbol{\beta}}$$

This proves that for any vector $\boldsymbol{x}$ in the reachable sub-set, than also $\mathbf{A}\boldsymbol{x}$ lies in the same set, proving the $\mathbf{A}$-invariance of the matrix $\mathbf{V}$ (that's representing in fact $\text{Im}\{\mathbf{R}\}$).

b) The proof of *ii)* is straightforward: it's obvious that $\text{Im}\{\mathbf{B}\} \subseteq \mathcal{R}$ by observing that

$$\text{Im}\{\mathbf{B}\} \subseteq \text{Im}\{\mathbf{R}\} = \text{Im}\left\{\begin{bmatrix}\mathbf{B} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B}\end{bmatrix}\right\} = \mathcal{R}$$

thus $\mathbf{B}$ can always be regarded as a linear combination of a basis of the reachable subset, hence as the product $\mathbf{VB}_c$.

Given so the basis of $\mathbf{R}$ determining the matrix $\boldsymbol{V}$ and calling $\mathbf{V}_c \in \mathbb{R}^{n\times(n-\overline{n})}$ it's **completion** of $\mathbb{R}^n$ (so such that the columns of $\begin{bmatrix}\mathbf{V} & \mathbf{V}_c\end{bmatrix}$ are a basis of $\mathbb{R}^n$), then the **transformation matrix** $\mathbf{T}$ defined as

$$\mathbf{T} = \begin{bmatrix}\mathbf{V} & \mathbf{V}_c\end{bmatrix} \tag{4.28}$$

allows us to perform the **controllable decomposition**, so to separate the controllable states from the unobservable ones.

Computing in fact the product

$$\mathbf{AT} = \begin{bmatrix}\mathbf{AV} & \mathbf{AV}_c\end{bmatrix} \overset{(\dagger)}{=} \begin{bmatrix}\mathbf{VA}_c & \mathbf{AV}_c\end{bmatrix}$$

Recalled that $\mathbf{A}_c$ is a $\overline{n}\times\overline{n}$ matrix, we can also observe that the product $\mathbf{AT}$ can be further decomposed as

$$\mathbf{AT} = \left[\begin{bmatrix}\mathbf{V} & \mathbf{V}_c\end{bmatrix}\begin{bmatrix}\mathbf{A}_c \\ 0\end{bmatrix} \quad \mathbf{AV}_c\right] = \left[\mathbf{T}\begin{bmatrix}\mathbf{A}_c \\ 0\end{bmatrix} \quad \mathbf{TT}^{-1}\mathbf{AV}_c\right] = \mathbf{T}\begin{bmatrix}\mathbf{A}_c & \mathbf{T}^{-1}\mathbf{AV}_c\end{bmatrix}$$

In particular $\mathbf{T}^{-1}\mathbf{AV}_c$ is a $n\times(n-\overline{n})$ matrix that can be decomposed into two submatrices $\mathbf{A}_{12} \in \mathbb{R}^{\overline{n}\times(n-\overline{n})}$ and $\mathbf{A}_u \in \mathbb{R}^{(n-\overline{n})\times(n-\overline{n})}$ in such a manner that it allows to rewrite $\mathbf{AT}$ as

$$\mathbf{AT} = \mathbf{T}\begin{bmatrix}\mathbf{A}_c & \mathbf{A}_{12} \\ 0 & \mathbf{A}_u\end{bmatrix}$$

So finally

$$\tilde{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{AT} = \begin{bmatrix}\mathbf{A}_c & \mathbf{A}_{12} \\ 0 & \mathbf{A}_u\end{bmatrix} \tag{4.29}$$

In particular what we have is that only the last $n-\overline{n}$ rows, so the block $\begin{bmatrix}0 & \mathbf{A}_u\end{bmatrix}$, is representing the dynamic of the **non-controllable states**: knowing that $\mathbf{B}$ can be regarded as $\mathbf{VB}_c$, this also implies

$$\mathbf{B} = \begin{bmatrix}\mathbf{V} & \mathbf{V}_c\end{bmatrix}\begin{bmatrix}\mathbf{B}_c \\ 0\end{bmatrix} = \mathbf{T}\begin{bmatrix}\mathbf{B}_c \\ 0\end{bmatrix}$$

thus for the similarity transformation

$$\tilde{\mathbf{B}} = \mathbf{T}^{-1}\mathbf{B} = \cancel{\mathbf{T}^{-1}\mathbf{T}}\begin{bmatrix}\mathbf{B}_c \\ 0\end{bmatrix}$$

This shows that in the transformed matrix, only the first $\overline{n}$ states can be accessed by the input (as $\mathbf{B}_c$ is a $\overline{n}\times m$ matrix), showing that the last $n-\overline{n}$ remains untouched by the inputs $\boldsymbol{u}$, proving that they are **non-controllable**. In particular the sub-matrix $\mathbf{A}_u$ in (4.29) is the matrix ruling the **dynamics** of the **uncontrollable *part*** of the system.

Summarizing, we call **controllable decomposition** of the pair $(\mathbf{A}, \mathbf{B})$ the matrices

$$\tilde{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{AT} = \begin{bmatrix}\mathbf{A}_c & \mathbf{A}_{12} \\ 0 & \mathbf{A}_u\end{bmatrix} \qquad \tilde{\mathbf{B}} = \mathbf{T}^{-1}\mathbf{B} = \begin{bmatrix}\mathbf{B}_c \\ 0\end{bmatrix} \tag{4.30}$$

Theorem 4.6: Each LTI system characterized by a pair $(\tilde{\mathbf{B}}, \tilde{\mathbf{A}})$ in the form $\left(\begin{bmatrix}\mathbf{B}_c \\ 0\end{bmatrix}, \begin{bmatrix}\mathbf{A}_c & \mathbf{A}_{12} \\ 0 & \mathbf{A}_u\end{bmatrix}\right)$ satisfies:

i) the reachable subspace is given by

$$\mathrm{Im}\left\{\tilde{\mathbf{R}}\right\} = \mathrm{Im}\left\{\begin{bmatrix}\mathbf{I}_{\overline{n}\times\overline{n}}\\ 0\end{bmatrix}\right\}$$

ii) the pair $(\mathbf{B}_c, \mathbf{A}_c)$ is reachable/controllable.

Chapter 5

Observability and constructibility

The linear quadratic regulator (page 22), or more generally the full-state feedback (page 30), are proposed control system that, through the design of a feedback of the type $\boldsymbol{u} = -\mathbf{K}\boldsymbol{x}$, allow us to achieve desired property concerning the stability of the system of interest.

Ideally feedback of this types are the best one for linear system, however the main cons of such control system is that it requires the knowledge of all states at each time, and this is not always guaranteed. In reality in fact the states $\boldsymbol{x}$ have to be reconstructed starting from the read output $\boldsymbol{y}$ of the system. The main goal of this chapter is so to study the **observability**: as in controllability we studied the dynamic equation $\dot{\boldsymbol{x}} = \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}$ to study whenever a system was controllable, in this case we study the output equation

$$\boldsymbol{y} = \mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u}$$

to asses the values of the states $\boldsymbol{x}$ as function of the measured output $\boldsymbol{y}(t)$, starting by the idea that if $\mathbf{C}$ is invertible we can have the estimation of the states $\hat{\boldsymbol{x}}$ as $\mathbf{C}^{-1}(\boldsymbol{y} - \mathbf{D}\boldsymbol{u})$.

Intuition behind the concepts Similarly to what had been said for *controllability* and *reachability*, we define with $[t_0, t_1]$**-observability** the ability to reconstruct the initial state of a system given it's input $\boldsymbol{u}(t), \boldsymbol{y}(t)$ (with $t \in [t_0, t_1]$), while $[t_0, t_1]$**-constructibility** is instead the ability to reconstruct the final state $\boldsymbol{x}(t_1)$ given both $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$.

Unobservable subspace Given a continuous-time linear system, from the variation of constants formula (2.3) we hat that the output $\boldsymbol{y}(t)$ is defined as

$$\boldsymbol{y}(t) = \mathbf{C}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}(t_0) + \int_{t_0}^{t} \mathbf{C}(\tau)\boldsymbol{\Phi}(t, \tau)\mathbf{B}(\tau)\, d\tau + \mathbf{D}(t)\boldsymbol{u}(t)$$

The unknown quantity associated to the estimation of the state is in this case $\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}(t_0)$; assuming to have available both $\boldsymbol{u}(t), \boldsymbol{y}(t)$ (as well as the model of the system), then by simply reversing the equation we have that

$$\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}(t_0) = \boldsymbol{y}(t) - \int_{t_0}^{t} \mathbf{C}(\tau)\boldsymbol{\Phi}(t, \tau)\mathbf{B}(\tau)\, d\tau - \mathbf{D}(t)\boldsymbol{u}(t) \tag{5.1}$$

Theorem 5.1: Given two times $t_1 > t_0 \geq 0$, we write $\mathcal{N}_{\mathcal{O}}[t_0, t_1]$ the **unobservable subspace** as the set of all initial condition $\boldsymbol{x}(t_0) \in \mathbb{R}^n$ leading to zero homogeneous response:

$$\mathcal{N}_{\mathcal{O}}[t_0, t_1] = \{\boldsymbol{x}(t_0) \in \mathbb{R}^n \ : \ \mathbf{C}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}(t_0) = 0 \ \forall t \in [t_0, t_1]\} \tag{5.2}$$

The idea in this case is that if a state $\boldsymbol{x}_0$ lies in the unobservable subspace $\mathcal{N}_{\mathcal{O}}$, then it's impossible to asses if the *state-estimator* $\tilde{\boldsymbol{y}}(t)$ is zero due to initial condition $\boldsymbol{x}(t_0)$ of by the term $\mathbf{C}(t)\boldsymbol{\Phi}(t, t_0)$.

As properties of the unobservable subspace, for any 2 times $t_1 > t_0 \geq 0$ and an input pair $(\boldsymbol{u}(t), \boldsymbol{y}(t))$ defined in $[t_0, t_1]$:

i) for any initial state $\boldsymbol{x}_0$ satisfying $\tilde{y}(t) = \mathbf{C}(t)\mathbf{\Phi}(t,t_0)\boldsymbol{x}_0 \ \forall t \in [t_0,t_1]$, then for every unobservable state $\boldsymbol{x}_u \in \mathcal{N}_{\mathcal{O}}[t_0,t_1]$ it holds that

$$\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\mathbf{\Phi}(t,t_0)(\boldsymbol{x}_0 + \boldsymbol{x}_u) \qquad \forall t \in [t_0,t_1] \tag{5.3}$$

ii) when the unobservable subspace is trivial, so $\mathcal{N}_{\mathcal{O}}[t_0,t_1] = \{0\}$, then there exists at most one initial state $\boldsymbol{x}_0$ compatible with the pair $(\boldsymbol{u},\boldsymbol{y})$ satisfying (5.1).

Proof 5.1:

a) proof of *i)* is straightforward and is based on the linearity property:

$$\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\mathbf{\Phi}(t,t_0)\boldsymbol{x}_0 + \cancel{\mathbf{C}(t)\mathbf{\Phi}(t,t_0)\boldsymbol{x}_u}$$

where the second term has been cancelled due to the definition of unobservable subspace;

b) property *ii)* can be proved by absurd assuming that there exists two different initial states $\boldsymbol{x}_0 \neq \bar{\boldsymbol{x}}_0$ compliant with the given input-output pair $(\boldsymbol{u},\boldsymbol{y})$, so for which (5.1) holds. Subtracting the two equations leads to

$$0 = \mathbf{C}(t)\mathbf{\Phi}(t,t_0)\left(\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_0\right)$$

In order to be true this means that in general the non-zero vector $\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_0$ must be inside the unobservable subspace, in contradiction to the initial hypothesis of it being trivial.

Given two times $t_1 > t_0 \geq 0$, the pair $\left(\mathbf{C}(\cdot), \mathbf{A}(\cdot)\right)$ is said **$[t_0,t_1]$-observable** if it's unobservable subspace is made by the trivial set $\mathcal{N}_{\mathcal{O}}[t_0,t_1] = \{0\}$.

Note that in general matrices $\mathbf{B}, \mathbf{D}$ are not considered while dealing with observability as they do not appear in any way in the definition $\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\mathbf{\Phi}(t,t_0)\boldsymbol{x}(t_0)$.

Unconstructible subspace The variation of constant formula for a continuous-time linear system cab be also expressed as function of the final state $\boldsymbol{x}(t_1)$, thus (5.1) can be rewritten as

$$\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\mathbf{\Phi}(t,t_1)\boldsymbol{x}(t_1) = \boldsymbol{y}(t) - \int_{t_1}^{t} \mathbf{C}(\tau)\mathbf{\Phi}(t,\tau)\mathbf{B}(\tau)\,d\tau - \mathbf{D}(t)\boldsymbol{u}(t) \tag{5.4}$$

Theorem 5.2: Given two times $t_1 > t_0 \geq 0$ the **unconstructible subspace** $\mathcal{N}_{\mathcal{C}}[t_0,t_1]$ is made by the set of all the final states $\boldsymbol{x}(t_1)$ for which $\tilde{\boldsymbol{y}}(t)$ evaluates to zero, so

$$\mathcal{N}_{\mathcal{C}}[t_0,t_1] = \{\boldsymbol{x}(t_1) \in \mathbb{R}^n \ : \ \mathbf{C}(t)\mathbf{\Phi}(t,t_1)\boldsymbol{x}(t_1) = 0 \ \forall t \in [t_0,t_1]\} \tag{5.5}$$

The meaning of this theorem is similar to 5.1, however in this case the set $\mathcal{N}_{\mathcal{C}}$ contains all the final state that doesn't allow to discriminate whenever $\tilde{y}(t)$ is zero due to the state $\boldsymbol{x}(t_1)$ or the value of $\mathbf{C}(t)\mathbf{\Phi}(t,t_0)$. Also in this case we have the properties

i) if $\boldsymbol{x}_1$ satisfies $\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\mathbf{\Phi}(t,t_1)\boldsymbol{x}_1$ for all $t \in [t_0,t_1]$, for any state $\boldsymbol{x}_u \in \mathcal{N}_{\mathcal{C}}[t_0,t_1]$ we have that also $\boldsymbol{x}_1 + \boldsymbol{x}_u$ is a solution of (5.4):

$$\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\mathbf{\Phi}(t,t_1)\boldsymbol{x}_1 = \mathbf{C}(t)\mathbf{\Phi}(t,t_1)(\boldsymbol{x}_1 + \boldsymbol{x}_u)$$

ii) if the unconstructible subspace is determined by the trivial subspace $\mathcal{N}_{\mathcal{C}}[t_0,t_1] = \{0\}$, then at most 1 state $\boldsymbol{x}_r \in \mathbb{R}^n$ is compatible with the given input-output pair $\left(\boldsymbol{u}(\cdot),\boldsymbol{y}(\cdot)\right)$.

Example of unobservable system Let us consider a system that's made by a parallel connection between two SISO continuous-time sub-systems with dynamics $\dot{x}_i = A_i\boldsymbol{x}_i + B_i u$, where so the input u is common to both system and the overall output y is the sum $y_1 + y_2$ where $y_i = C_i x_i$.

Choosing $\boldsymbol{x} = (x_1, x_2)$ the states of the overall system, the associated state-space representation is

$$\begin{cases} \dot{\boldsymbol{x}} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u \\ y = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \boldsymbol{x} \end{cases}$$

In this case the *states estimator function* $\tilde{y}(t)$ can be regarded as the sum of the two $\tilde{y}_1(t) + \tilde{y}_2(t)$, expanding so to the form

$$\tilde{y}(t) = C_1 e^{A_1 t} x_1(0) + C_2 e^{A_2 t} x_2(0) = 0$$

Considering now the particular case when $A_1 = A_2 = A$ and $C_1 = C_2 = C$, this sum collapse to the form

$$\tilde{y}(t) = C e^{At} \big(x_1(0) + x_2(0)\big) = 0$$

In this case we observe that the function $\tilde{y}(t)$ can also be zero for any time t if it happens that $x_1(0) = -x_2(0)$, so we can say that the unobservable subspace is

$$\mathcal{N}_{\mathcal{O}}[t_0, t_1] = \big\{(x_1, x_2) \in \mathbb{R}^2 \text{ such that } x_1 = -x_2\big\}$$

Having a non-trivial observable subspace implies that the system is **unobservable**.

5.1 Gramians

As for reachability/controllability, given two times $t_1 > t_0 \geq 0$ we can define the $[t_0, t_1]$**-observability** and $[t_0, t_1]$**-constructibility Gramians** the symmetric matrices defiend as

$$\begin{aligned} \mathbf{W}_{\mathcal{O}}[t_0, t_1] &= \int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau, t_0) \mathbf{C}^T(\tau) \mathbf{C}(\tau) \boldsymbol{\Phi}(\tau, t_0)\, d\tau \\ \mathbf{W}_{\mathcal{C}n}[t_0, t_1] &= \int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau, t_1) \mathbf{C}^T(\tau) \mathbf{C}(\tau) \boldsymbol{\Phi}(\tau, t_1)\, d\tau \end{aligned} \tag{5.6}$$

Theorem 5.3: Given two times $t_1 > t_0 \geq 0$, then it holds that

$$\mathcal{N}_{\mathcal{O}}[t_0, t_1] = \text{Ker}\,\{\mathbf{W}_{\mathcal{O}}[t_0, t_1]\} \qquad \mathcal{N}_{\mathcal{C}}[t_0, t_1] = \text{Ker}\,\{\mathbf{W}_{\mathcal{C}n}[t_0, t_1]\} \tag{5.7}$$

Proof 5.2: To show that if a state $\boldsymbol{x}_0 \in \mathcal{N}_{\mathcal{O}}$ is unobservable we consider the quadratic form

$$\boldsymbol{x}_0^T \mathbf{W}_{\mathcal{O}} \boldsymbol{x}_0 = \int_{t_0}^{t_1} \boldsymbol{x}_0^T \boldsymbol{\Phi}^T(\tau, t_0) \mathbf{C}^T(\tau) \mathbf{C}(\tau) \boldsymbol{\Phi}(\tau, t_0) \boldsymbol{x}_0\, d\tau = \int_{t_0}^{t_1} \big|\mathbf{C}(\tau) \boldsymbol{\Phi}(\tau, t_0) \boldsymbol{x}_0\big|^2\, d\tau \geq 0$$

Having supposed that $\boldsymbol{x}_0$ is unobservable, then the product $\mathbf{C}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0$ is mutually zero, thus the quadratic form evaluates to $\boldsymbol{x}_0^T \mathbf{W}_{\mathcal{O}} \boldsymbol{x}_0 = 0$. Showing now that $\mathbf{W}_{\mathcal{O}} \boldsymbol{x}_0 = 0$ evaluates to zero for any unobservable state, then we can say that $\boldsymbol{x}_0$ belongs to the kernel of $\mathbf{W}_{\mathcal{O}}$. Considering in fact

$$\mathbf{W}_{\mathcal{O}} \boldsymbol{x}_0 = \int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau, t_0) \mathbf{C}^T(\tau) \mathbf{C}(\tau) \boldsymbol{\Phi}(\tau, t_0) \boldsymbol{x}_0\, d\tau = \int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau, t_0) \mathbf{C}^T(\tau) \tilde{\boldsymbol{y}}(\tau)\, d\tau$$

by having $\boldsymbol{x}_0 \in \mathcal{N}_{\mathcal{O}}$ ensures us that $\tilde{\boldsymbol{y}}(\tau) = 0$ for any τ implying indeed that $\mathbf{W}_{\mathcal{O}} \boldsymbol{x}_0 = 0$ for any time $t \in [t_0, t_1]$. The proof for the unconstructible subspace is dual.

As corollary of this definitions we have that the pair $\big(\mathbf{C}(\cdot), \mathbf{A}(\cdot)\big)$ is

i) $[t_0, t_1]$**-observable** if and only if the rank of the observability Gramian $\mathbf{W}_{\mathcal{O}}$ is equal to n;

ii) $[t_0, t_1]$**-constructible** if and only if the constructibility Gramian $\mathbf{W}_{\mathcal{C}n}$ is full rank.

Considering in-fact that such Gramians are made by n linearly independent columns (implying that they are full rank), then having $\text{rank}\{\mathbf{W}_{\mathcal{O}}\} = 0$ implies, from the fundamental theorem of linear algebra, that $\dim \text{Ker}\,\{\mathbf{W}_{\mathcal{O}}\} = 0$, so kernel (hence the unobservable subspace) is the trivial set $\mathcal{N}_{\mathcal{O}} = \{0\}$.

Gramians-based reconstruction

Given a continuous-time LTV system and the input-output pair $(\boldsymbol{u},\boldsymbol{y})$, pre-multiplying (5.4) by $\boldsymbol{\Phi}^T(t,t_0)\mathbf{C}^T(t)$ evaluates to

$$\boldsymbol{\Phi}^T(t,t_0)\mathbf{C}^T(t)\tilde{\boldsymbol{y}}(t) = \boldsymbol{\Phi}^T(t,t_0)\mathbf{C}^T(t)\mathbf{C}(t)\boldsymbol{\Phi}(t,t_0)\boldsymbol{x}_0$$

where $\tilde{\boldsymbol{y}}(t)$ is a quantity strictly related to both $\boldsymbol{u}$ and $\boldsymbol{y}$. Integrating this equations in time leads to

$$\begin{aligned}\int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau,t_0)\mathbf{C}^T(\tau)\tilde{\boldsymbol{y}}(\tau)\,d\tau &= \int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau,t_0)\mathbf{C}^T(\tau)\mathbf{C}(t)\boldsymbol{\Phi}(\tau,t_0)\,d\tau\,\boldsymbol{x}_0 \\ &= \mathbf{W}_{\mathcal{O}}[t_0,t_1]\boldsymbol{x}_0\end{aligned}$$

Assuming now that the system is observable (so assuming $\operatorname{rank}\{\mathbf{W}_{\mathcal{O}}\} = n$), then it means that the Gramian $\mathbf{W}_{\mathcal{O}}$ is invertible, thus the initial state $\boldsymbol{x}_0$ characterizing the given input-output pair can be obtained as

$$\boldsymbol{x}_0 = \mathbf{W}_{\mathcal{O}}^{-1}[t_0,t_1]\int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau,t_0)\mathbf{C}^T(\tau)\tilde{\boldsymbol{y}}(\tau)\,d\tau$$

Theorem 5.4: Given two times $t_1 > t_0 \geq 0$ and an input-output pair $\big(\boldsymbol{u}(t),\boldsymbol{y}(t)\big)$ defined in the time range $t \in [t_0,t_1]$, then

i) if the pair $(\mathbf{C},\mathbf{A})$ is $[t_0,t_1]$-observable, then the corresponding initial state is

$$\boldsymbol{x}_0 = \mathbf{W}_{\mathcal{O}}^{-1}[t_0,t_1]\int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau,t_0)\mathbf{C}^T(\tau)\tilde{\boldsymbol{y}}(\tau)\,d\tau \tag{5.8}$$

ii) if the pair $(\mathbf{C},\mathbf{A})$ is $[t_0,t_1]$-constructible, then the corresponding final state can be computed as

$$\boldsymbol{x}_1 = \mathbf{W}_{\mathcal{C}n}^{-1}[t_0,t_1]\int_{t_0}^{t_1} \boldsymbol{\Phi}^T(\tau,t_1)\mathbf{C}^T(\tau)\tilde{\boldsymbol{y}}(\tau)\,d\tau \tag{5.9}$$

5.2 Extension to the discrete-time case

Until now we discussed **observability/controllability** of continuous-time LTV, system, however the same concepts can be extended to discrete-time ones with some subtle details changing. In this case the variation of constants formula (2.6), page 10, is a little bit different and lead to a state estimator function

$$\tilde{\boldsymbol{y}}(t) = \mathbf{C}(t)\boldsymbol{\Phi}(t,t_0)\boldsymbol{x}_0 = \boldsymbol{y}(t) - \sum_{\tau=t_0}^{t_1-1} \mathbf{C}(t)\boldsymbol{\Phi}(t,\tau+1)\mathbf{B}(\tau)\boldsymbol{u}(\tau) - \mathbf{D}(t)\boldsymbol{u}(t)$$

The definition of unobservable subspace (5.2) is still the same, however the $[t_0,t_1]$-unconstructible subspace $\mathcal{N}_{\mathcal{C}}[t_0,t_1]$ defined as the set of $\boldsymbol{x}_1$ satisfying

$$\mathbf{C}(t)\boldsymbol{\Phi}(t,t_1)\boldsymbol{x}_1 = 0 \qquad \forall t \in [t_0,t_1]$$

requires that the state matrix $\mathbf{A}(t)$ is non-singular (invertible) for all times $t \in [t_0,t_1]$ in order to make possible the computation of the Peano-Baker series $\boldsymbol{\Phi}(t,t_1)$ backward in times.

5.3 LTI observability and constructibility

All concepts yet described simplifies if we consider a continuos-time LTI case, where so the pair $(\mathbf{C},\mathbf{A})$ is characterized by constant matrices. In this particular case the observability Gramian can be rewritten as

$$\mathbf{W}_{\mathcal{O}}[t_0,t_1] = \int_{t_0}^{t_1} e^{\mathbf{A}^T(\tau-t_0)}\mathbf{C}^T\mathbf{C}e^{\mathbf{A}(\tau-t_0)}\,d\tau$$

Regarding $\mathbf{A}^T$ as $\overline{A}$ and $\mathbf{C}^T$ as $\overline{\mathbf{B}}$ we can re-formulate this statement as

$$\mathbf{W}_{\mathcal{O}}[t_0, t_1] = \int_{t_0}^{t_1} e^{\overline{\mathbf{A}}(\tau - t_0)} \overline{\mathbf{B}}\overline{\mathbf{B}}^T e^{\overline{\mathbf{A}}^T(\tau - t_0)}\, d\tau$$

The result yet obtained is very close with the one presented in (4.12), page 28, where so we can find an analogy between observability (constructibility) and reachability (controllability), in particular we can note the following **duality** between the systems:

$$\Sigma_1 : \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right] \qquad \text{dual to} \qquad \Sigma_2 : \left[\begin{array}{c|c} \overline{\mathbf{A}} & \overline{\mathbf{B}} \\ \hline \overline{\mathbf{C}} & \overline{\mathbf{D}} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}^T & \mathbf{C}^T \\ \hline \mathbf{B}^T & \mathbf{D}^T \end{array}\right]$$

Theorem 5.5: A LTI system Σ_1, or equivalently the pair $(\mathbf{C}, \mathbf{A})$, is **observable** (**controllable**) if and only if it's **dual** system Σ_2, or equivalently the pair $(\mathbf{A}^T, \mathbf{C}^T)$ is **reachable** (**controllable**).

As in the previous chapters, the concepts of observability and controllability is equivalent for LTI system, in fact the two subspaces $\mathcal{N}_{\mathcal{O}}$ and $\mathcal{N}_{\mathcal{C}}$ are the same.

Dual to the controllability matrix $\mathbf{R}$ described at page 28, we can define the **observability matrix** $\mathbf{O}$ the one defined as

$$\mathbf{O} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} \tag{5.10}$$

5.3.1 Observability test

A pair $(\mathbf{C}, \mathbf{A})$ is **observable** (constructible) if and only if the observability matrix is full rank, so

$$\text{observability} \qquad \Leftrightarrow \qquad \text{rank}\{\mathbf{O}\} = n \tag{5.11}$$

Proof 5.3: The proof is straightforward considering the reachability matrix $\overline{\mathbf{R}}$ of the dual system that's defined as

$$\overline{\mathbf{R}} = \begin{bmatrix} \mathbf{C}^T & \mathbf{A}^T\mathbf{C}^T & \dots & \left(\mathbf{A}^T\right)^{n-1}\mathbf{C}^T \end{bmatrix} = \mathbf{O}^T$$

If $\overline{\mathbf{R}}$ is full rank (implying controllability), then it means that also $\mathbf{O}$ is full rank (implying observability).

Eigenvector test for observability *A pair $(\mathbf{C}, \mathbf{A})$ is observable if and only if no eigenvector of $\mathbf{A}$ are in the kernel of $\mathbf{C}$.*

Proof 5.4: Also in this case the straightforward proof is by duality, considering that

$$(\mathbf{C}, \mathbf{A}) \text{ observable} \qquad \Leftrightarrow \qquad \left(\mathbf{A}^T, \mathbf{C}^T\right) \text{ reachable}$$

Considering the definition of the eigenvector test for controllability applied to such system what we obtain is indeed the definition *no eigenvector of $\left(\mathbf{A}^T\right)^T = \mathbf{A}$ can lie in the kernel of $\left(\mathbf{C}^T\right)^T = \mathbf{C}$.*

PBH test for observability *A pair $(\mathbf{C}, \mathbf{A})$ is observable if and only if the matrix $\begin{bmatrix} \mathbf{A} - \lambda\mathbf{I} \\ \mathbf{C} \end{bmatrix}$ is full-rank for any complex coefficient λ:*

$$\text{observable system} \qquad \Leftrightarrow \qquad \text{rank}\left\{\begin{bmatrix} \mathbf{A} - \lambda\mathbf{I} \\ \mathbf{C} \end{bmatrix}\right\} = n \quad \forall \lambda \in \mathbb{C} \tag{5.12}$$

As in the case of reachability, this relation is true for almost any coefficient λ and the lonely *problematic values* are the eigenvalues of $\mathbf{A}$ that are making the matrix $\mathbf{A}$ rank-deficient.

Lyapunov test for observability *Given a Hurwitz (Schur) state-matrix* $\mathbf{A}$*, the pair* $(\mathbf{C}, \mathbf{A})$ *is observable if and only if exists a matrix* $\mathbf{W} = \mathbf{W}^T > 0$ *satisfying*

$$(CT): \quad \mathbf{A}^T\mathbf{W} + \mathbf{W}\mathbf{A} = -\mathbf{C}^T\mathbf{C} \qquad (DT): \quad \mathbf{A}^T\mathbf{W}\mathbf{A} - \mathbf{W} = -\mathbf{C}^T\mathbf{C} \tag{5.13}$$

Moreover the unique solution is given by

$$(CT): \quad \mathbf{W} = \lim_{t_1 - t_0 \to \infty} \mathbf{W}_{\mathcal{O}}[t_0, t_1] = \int_0^{\infty} e^{\mathbf{A}^T\tau}\mathbf{C}^T\mathbf{C}e^{\mathbf{A}\tau}\, d\tau$$
$$(DT): \quad \mathbf{W} = \lim_{t_1 - t_0 \to \infty} \mathbf{W}_{\mathcal{O}}[t_0, t_1] = \sum_{\tau=0}^{\infty} \left(\mathbf{A}^T\right)^{\tau}\mathbf{C}^T\mathbf{C}\mathbf{A}^{\tau}\, d\tau$$

5.3.2 Observable decomposition

Recalling what has been show in page 6, we can have that a system $\dot{\boldsymbol{x}}/\boldsymbol{x}^+ = \mathbf{A}\boldsymbol{x}, \boldsymbol{y} = \mathbf{C}\boldsymbol{x}$ is algebraically equivalent to $\dot{\overline{\boldsymbol{x}}}/\overline{\boldsymbol{x}}^+ = \overline{\mathbf{A}}\overline{\boldsymbol{x}}, \boldsymbol{y} = \overline{\mathbf{C}}\overline{\boldsymbol{x}}$ if exists a similarity transformation matrix $\mathbf{T}$ for which it holds

$$\begin{bmatrix} \overline{\mathbf{A}} \\ \overline{\mathbf{C}} \end{bmatrix} = \begin{bmatrix} \mathbf{T}^{-1}\mathbf{A}\mathbf{T} \\ \mathbf{C}\mathbf{T} \end{bmatrix}$$

One relevant property to observe is that the computation of the observability matrix $\overline{\mathbf{O}}$ of the algebraically equivalent system can be simply computed from $\mathbf{O}$ as

$$\overline{\mathbf{O}} = \begin{bmatrix} \overline{\mathbf{C}} \\ \overline{\mathbf{C}\mathbf{A}} \\ \vdots \\ \overline{CA}^{n-1} \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mathbf{T} \\ \mathbf{C}\mathbf{T}\mathbf{T}^{-1}\mathbf{A}\mathbf{T} \\ \vdots \\ \mathbf{C}\mathbf{T}\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\ldots\mathbf{T}^{-1}\mathbf{A}\mathbf{T} \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mathbf{T} \\ \mathbf{C}\mathbf{A}\mathbf{T} \\ \vdots \\ \mathbf{C}\mathbf{A}^{n-1}\mathbf{T} \end{bmatrix} = \mathbf{O}\mathbf{T}$$

Theorem 5.6: The pair $(\mathbf{C}, \mathbf{A})$ is observable if and only if the pair $(\overline{\mathbf{C}}, \overline{\mathbf{A}}) = (\mathbf{C}\mathbf{T}, \mathbf{T}^{-1}\mathbf{A}\mathbf{T})$ is also invertible for any non-singular $n \times n$ transformation matrix $\mathbf{T}$.

This theorem is dual to the controllable case and states that observability is an intrinsic property of the system, meaning that it's preserved for any algebraically equivalent system.

Theorem 5.7: For any unobservable pair $(\mathbf{C}, \mathbf{A})$ there exists a similarity transformation $\mathbf{T}$ that determines and algebraically equivalent system with matrices

$$\overline{\mathbf{C}} = \mathbf{C}\mathbf{T} = \begin{bmatrix} \mathbf{C}_o & 0 \end{bmatrix} \qquad \overline{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{bmatrix} \mathbf{A}_o & 0 \\ \mathbf{A}_{12} & \mathbf{A}_{\bar{o}} \end{bmatrix} \tag{5.14}$$

where $\mathbf{A}_{\bar{o}}$ is a $n_{\bar{o}} \times n_{\bar{o}}$ matrix ruling the ***unobservable part*** of the system (indeed $n_{\bar{o}} = \dim \mathrm{Ker}\,\{\mathcal{N}_{\mathcal{O}}\}$). This is the **observable decomposition**.

Important consequences are

i) the unobservable subspace of the new system is characterized by the set

$$\mathcal{N}_{\mathcal{O}} = \mathrm{Im}\left\{ \begin{bmatrix} 0 \\ \mathbf{I}_{n_{\bar{o}} \times n_{\bar{o}}} \end{bmatrix} \right\} \tag{5.15}$$

ii) the pair $(\mathbf{C}_o, \mathbf{A}_o)$ is observable.

Similarly to what has been shown for controllability, the desired transformation matrix $\mathbf{T}$ of (5.14) can be built as $\begin{bmatrix} \mathbf{V}_o & \mathbf{V}_{\bar{o}} \end{bmatrix}$, where $\mathbf{V}_{\bar{o}}$ is a base of the unobservable subspace (basis of $\mathrm{Ker}\,\{\mathbf{W}_{\mathcal{O}}\} = \mathrm{Ker}\,\{\mathbf{O}\}$) and $\mathbf{V}_0$ is it's completion.

> **Note:** From a numerical point of view, computing kernels or images of matrices is performed through the *singular value decomposition* SVD, a process that's computationally expensive and numerically unfeasible for system of higher dimensions.

Chapter 6

Kalman decomposition, stabilizability and detectability

6.1 Kalman decomposition

The **Kalman decomposition** can be seen as combination of both controllable and observable decompositions: considering in fact a generic LTI system $\dot{\boldsymbol{x}}/\boldsymbol{x}^+ = \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}$, $\boldsymbol{y} = \mathbf{C}\boldsymbol{x}$ what we can have are

- co controllable and observable states;
- $\bar{c}o$ uncontrollable but observable states;
- $c\bar{o}$ controllable but unobservable states;
- $\overline{co}$ uncontrollable and unobservable states.

Knowing that the unobservable states $\bar{o}$ are computed as $\text{Ker}\{\mathbf{O}\}$ and the controllable states as $\text{Im}\{\mathbf{R}\}$, their intersection will evaluate in the unobservable and controllable $c\bar{o}$ part of the system. With this idea in mind we can build a **similarity transformation** $\mathbf{T} \in \mathbb{R}^{n\times n}$ defined as

$$\mathbf{T} = \begin{bmatrix} \mathbf{V}_{co} & \mathbf{V}_{c\bar{o}} & \mathbf{V}_{\bar{c}o} & \mathbf{V}_{\overline{co}} \end{bmatrix} \tag{6.1}$$

where

- the vectors of $\mathbf{V}_{c\bar{o}}$ are a basis of $\text{Im}\{\mathbf{R}\} \cap \text{Ker}\{\mathbf{O}\}$;
- the vectors of $\mathbf{V}_{co}$ are the completion of $\text{Im}\{\mathbf{R}\}$;
- the vectors of $\mathbf{V}_{\overline{co}}$ are the completion of $\text{Ker}\{\mathbf{O}\}$;
- the vectors of $\mathbf{V}_{\bar{c}o}$ are the completion of $\mathbb{R}^n$;

With the similarity transformation so defined, the resulting description of the algebraically equivalent system is split accordingly to definitions of controllable and unobservable states as

$$\begin{aligned} \dot{\bar{\boldsymbol{x}}}/\bar{\boldsymbol{x}}^+ &= \begin{bmatrix} \mathbf{A}_{co} & 0 & \mathbf{A}_{13} & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{c\bar{o}} & \mathbf{A}_{23} & \mathbf{A}_{24} \\ 0 & 0 & \mathbf{A}_{\bar{c}o} & 0 \\ 0 & 0 & \mathbf{A}_{43} & \mathbf{A}_{\overline{co}} \end{bmatrix} \begin{pmatrix} \boldsymbol{x}_{co} \\ \boldsymbol{x}_{c\bar{o}} \\ \boldsymbol{x}_{\bar{c}o} \\ \boldsymbol{x}_{\overline{co}} \end{pmatrix} + \begin{bmatrix} \mathbf{B}_{co} \\ \mathbf{B}_{c\bar{o}} \\ 0 \\ 0 \end{bmatrix} \boldsymbol{u} \\ \boldsymbol{y} &= \begin{bmatrix} \mathbf{C}_{co} & 0 & \mathbf{C}_{\bar{c}o} & 0 \end{bmatrix} \bar{\boldsymbol{x}} \end{aligned} \tag{6.2}$$

Theorem 6.1: Each LTI realization can be transformed in it's **Kalman decomposition** (6.2) where

i) the pair $\left(\begin{bmatrix}\mathbf{A}_{co} & 0\\ \mathbf{A}_{21} & \mathbf{A}_{c\bar{o}}\end{bmatrix}, \begin{bmatrix}\mathbf{B}_{co}\\ \mathbf{B}_{c\bar{o}}\end{bmatrix}\right)$ is controllable;

ii) the pair $\left(\begin{bmatrix}\mathbf{C}_{co} & \mathbf{C}_{\bar{c}o}\end{bmatrix}, \begin{bmatrix}\mathbf{A}_{co} & \mathbf{A}_{13}\\ 0 & \mathbf{A}_{\bar{c}o}\end{bmatrix}\right)$ is observable;

iii) the triple $(\mathbf{C}_{co}, \mathbf{A}_{co}, \mathbf{B}_{co})$ is controllable and observable;

iv) the transfer function of the system is

$$\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} = \mathbf{C}_{co}(s\mathbf{I} - \mathbf{A}_{co})^{-1}\mathbf{B}_{co} \tag{6.3}$$

Proof 6.1: The proof of *i)*, *ii)* and *iii)* is quite straightforward as, by construction, $\mathbf{T}$ is built considering the *rules* shown for both controllable and observable decomposition. The most interesting result of this theorem is *iv)*, expressing that the transfer function of a system is associated only to the controllable and observable component of the system.
Considering now that algebraically equivalents systems are zero-state invariant, this implies

$$\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} = \overline{\mathbf{C}}(s\mathbf{I} - \overline{\mathbf{A}})^{-1}\overline{\mathbf{B}}$$

Considering now the *structure* (6.2) of the provided Kalman decomposition, we can expand the computation as

$$\begin{aligned}
\mathbf{G}(s) &= \overline{\mathbf{C}}\left(s\mathbf{I} - \begin{bmatrix}\mathbf{A}_1 & *\\ 0 & \mathbf{A}_2\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{B}_1\\ 0\end{bmatrix} = \overline{\mathbf{C}}\begin{bmatrix}(s\mathbf{I} - \mathbf{A}_1)^{-1} & *\\ 0 & (s\mathbf{I} - \mathbf{A}_2)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{B}_1\\ 0\end{bmatrix}\\
&= \overline{\mathbf{C}}\begin{bmatrix}(s\mathbf{I} - \mathbf{A}_1)^{-1}\mathbf{B}_1\\ 0\end{bmatrix} = \begin{bmatrix}\mathbf{C}_{co} & 0 & * & *\end{bmatrix}\begin{bmatrix}\left(s\mathbf{I} - \begin{bmatrix}\mathbf{A}_{co} & 0\\ * & \mathbf{A}_{c\bar{o}}\end{bmatrix}\right)^{-1}\mathbf{B}_1\\ 0\\ 0\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{C}_{co} & 0 & * & *\end{bmatrix}\begin{bmatrix}(s\mathbf{I} - \mathbf{A}_{co})^{-1} & 0 & 0 & 0\\ * & (s\mathbf{I} - \mathbf{A}_{c\bar{o}})^{-1} & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{B}_{co}\\ \mathbf{B}_{c\bar{o}}\\ 0\\ 0\end{bmatrix}
\end{aligned}$$

Computing the matrix product gives indeed (6.3), proving that the forces response of a system depends lonely on the controllable and observable states.

6.2 Stabilizability

In chapter 4 it was discussed the concept of reachability/controllability as the set of all *achievable* states of the plant and controllable systems as the ones where *all configurations are allowed*. However in reality not all LTI systems are controllable, meaning that some states can't be *accessed* with the provided inputs, leading potentially to a bad behaviour of the system.

Recalling the controllable decomposition seen at page 39 we have that each LTI system can be written in a form

$$\begin{pmatrix}\dot{\boldsymbol{x}}_c/\boldsymbol{x}_c^+\\ \dot{\boldsymbol{x}}_{\bar{c}}/\boldsymbol{x}_{\bar{c}}^+\end{pmatrix} = \begin{bmatrix}\mathbf{A}_c & \mathbf{A}_{12}\\ 0 & \mathbf{A}_{\bar{c}}\end{bmatrix}\begin{pmatrix}\boldsymbol{x}_c\\ \boldsymbol{x}_{\bar{c}}\end{pmatrix} + \begin{bmatrix}\mathbf{B}_c\\ 0\end{bmatrix}\boldsymbol{u} \tag{$\star$}$$

As we can see from this formulation, we can't act with inputs on the uncontrollable states (as expected), however those states can have a direct influence on the controllable states through the matrix $\mathbf{A}_{12}$: this means that if $\boldsymbol{x}_{\bar{c}}$ starts diverging, the dynamics of the systems can't be controlled (it would require infinite actuator effort).

We say that a pair $(\mathbf{A}, \mathbf{B})$ is **stabilizable** if it's **controllable decomposition** is such that the matrix $\mathbf{A}_{\bar{c}}$ is either **empty** (hence the system is controllable) or **Hurwitz** (or Schur for discrete-time systems).

The main idea behind stabilizability is that even if the system is uncontrollable ($\mathbf{A}_{\bar{c}}$ is non-zero), the uncontrollable states present a converging behaviour *by themselves*: with this assumption, even if we can't access those states, they will eventually not impact the controllable states (as they converge to zero) allowing the possibility to generate an input controlling them.

Eigenvector test for stabilizability The pair $(\mathbf{A}, \mathbf{B})$ is stabilizable if and only if each *bad* eigenpair $(\lambda, \boldsymbol{x})$ of the matrix $\mathbf{A}^T$ is not in the kernel of $\mathbf{B}^T$ ($\mathbf{B}^T\boldsymbol{x} = 0$).

In this case we consider *bad* eigenvalues the one that are making the LTI system not exponentially stable, so for continuous-time system where $\mathbb{R}\mathrm{e}\{\lambda\} \geq 0$ and for the discrete-time counterpart when $|\lambda| \geq 1$.

Proof 6.2:

a) firstly we show that if $(\mathbf{A}, \mathbf{B})$ is stabilizable then no bad eigenpairs of $\mathbf{A}^T$ are in the kernel of $\mathbf{B}$. By contradiction let consider a bad eigenpair $(\lambda, \boldsymbol{x})$ for which it holds $\mathbf{A}^T\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\mathbf{B}^T\boldsymbol{x} = 0$; given the similarity transformation $\mathbf{T}$ leading to the controllable decomposition $\overline{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$, then we have that $\mathbf{A} = \mathbf{T}\overline{\mathbf{A}}\mathbf{T}^{-1}$ where $\overline{\mathbf{A}}$ is in controllable canonical form ($\star$). Computing

$$\mathbf{A}^T = \mathbf{T}^{-T}\overline{\mathbf{A}}^T\mathbf{T}^T \qquad \Rightarrow \qquad \mathbf{T}^{-T}\overline{\mathbf{A}}\mathbf{T}^T\boldsymbol{x} = \lambda\mathbf{T}^{-T}\mathbf{T}^T\boldsymbol{x}$$

and also having $\overline{\mathbf{B}} = \mathbf{T}^{-1}\mathbf{B}$ leading to

$$\mathbf{B}^T = \overline{\mathbf{B}}\mathbf{T}^T$$

gives

$$\begin{cases} \mathbf{T}^{-T}\overline{\mathbf{A}}\mathbf{T}^T\boldsymbol{x} = \lambda\mathbf{T}^{-T}\mathbf{T}^T\boldsymbol{x} \\ \overline{\mathbf{B}}\mathbf{T}^T\boldsymbol{x} = 0 \end{cases}$$

Observing now that the product $\mathbf{T}^T\boldsymbol{x}$ splits the controllable and uncontrollable states giving the form $(\boldsymbol{x}_c, \boldsymbol{x}_{\bar{c}})$, then we can regard the previous system as

$$\begin{cases} \begin{bmatrix} \mathbf{A}_c^T & 0 \\ \mathbf{A}_{12}^T & \mathbf{A}_{\bar{c}}^T \end{bmatrix} \begin{pmatrix} \boldsymbol{x}_c \\ \boldsymbol{x}_{\bar{c}} \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{x}_c \\ \boldsymbol{x}_{\bar{c}} \end{pmatrix} \\ \begin{bmatrix} \mathbf{B}_c^T & 0 \end{bmatrix} \begin{pmatrix} \boldsymbol{x}_c \\ \boldsymbol{x}_{\bar{c}} \end{pmatrix} = 0 \end{cases}$$

The second equation tells us that $\boldsymbol{x}_c$ must be zero (in order to have the identity), so the non-zero part must lie into $\boldsymbol{x}_{\bar{c}}$. Implicitly this means that the chosen λ is an eigenvalue of the uncontrollable part $\mathbf{A}_{\bar{c}}$ of the system, because it must satisfy

$$\mathbf{A}_{\bar{c}}\boldsymbol{x}_{\bar{c}} = \lambda\boldsymbol{x}_{\bar{c}}$$

However if λ is a bad eigenvalue means that the uncontrollable part cannot be stabilized, contradicting the original assumption that was stating that the system was indeed stabilizable (this means that the eigenvalue doesn't pertain to $\mathbf{A}_{\bar{c}}$ but rather on $\mathbf{A}_c$).

b) We can show now that if $(\mathbf{A}, \mathbf{B})$ is not stabilizable, then there exists a bad eigenpair $(\lambda, \boldsymbol{x})$ of $\mathbf{A}^T$ for which $\mathbf{B}^T\boldsymbol{x} = 0$. Let us suppose that exists an eigenvector $\boldsymbol{x}_{\bar{c}}$ with a bad eigenvalue $\lambda \in \mathbb{C}_{bad}$, so satisfying $\mathbf{A}_{\bar{c}}^T\boldsymbol{x}_{\bar{c}} = \lambda\boldsymbol{x}_{\bar{c}}$; building the vector $\overline{\boldsymbol{x}}$ as $(0, \boldsymbol{x}_{\bar{c}})$ we can observe that

$$\overline{\mathbf{A}}^T\overline{\boldsymbol{x}} = \begin{bmatrix} \mathbf{A}_c^T & 0 \\ \mathbf{A}_{12}^T & \mathbf{A}_{\bar{c}}^T \end{bmatrix} \begin{pmatrix} 0 \\ \boldsymbol{x}_{\bar{c}} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{A}_{\bar{c}}^T\boldsymbol{x}_{\bar{c}} \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ \boldsymbol{x}_{\bar{c}} \end{pmatrix} = \lambda\overline{\boldsymbol{x}}$$

and

$$\overline{\mathbf{B}}^T\overline{\boldsymbol{x}} = \begin{bmatrix} \mathbf{B}_c^T & 0 \end{bmatrix} \begin{pmatrix} 0 \\ \boldsymbol{x}_{\bar{c}} \end{pmatrix} = 0$$

Defining $\boldsymbol{x}$ as $\mathbf{T}^{-T}\overline{\boldsymbol{x}}$, then we can regard the products $\mathbf{A}^T\boldsymbol{x}$ and $\mathbf{B}^T\boldsymbol{x}$ as

$$\mathbf{A}^T\boldsymbol{x} = \mathbf{T}^{-T}\overline{\mathbf{A}}^T\cancel{\mathbf{T}^T\mathbf{T}^{-T}}\begin{pmatrix}0\\ \boldsymbol{x}_{\bar{c}}\end{pmatrix} = \mathbf{T}^{-T}\overline{\mathbf{A}}^T\overline{\boldsymbol{x}} = \mathbf{T}^{-T}\lambda\overline{\boldsymbol{x}} = \lambda\mathbf{T}^{-T}\overline{\boldsymbol{x}} = \lambda\boldsymbol{x}$$

$$\mathbf{B}^T\boldsymbol{x} = \overline{\mathbf{B}}^T\mathbf{T}^T\mathbf{T}^{-T}\begin{pmatrix}0\\ \boldsymbol{x}_{\bar{c}}\end{pmatrix} = \begin{bmatrix}\mathbf{B}_c^T & 0\end{bmatrix}\begin{pmatrix}0\\ \boldsymbol{x}_{\bar{c}}\end{pmatrix} = 0$$

With that we showed that if $(\mathbf{A}, \mathbf{B})$ is not stabilizable then we can find a bad eigenpair $(\lambda, \boldsymbol{x})$ of $\mathbf{A}^T$ for which the eigenvector $\boldsymbol{x}$ lies in the kernel of $\mathbf{B}^T$.

PBH test for stabilizability A pair $(\mathbf{A}, \mathbf{B})$ is stabilizable if and only if the matrix $\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I} & \mathbf{B}\end{bmatrix}$ is full rank for any bad eigenvalue $\lambda \in \mathbb{C}_{bad}$.

Lyapunov test for stabilizability A pair $(\mathbf{A}, \mathbf{B})$ is stabilizable if and only if exists a symmetric positive-definite matrix $\mathbf{W}$ such that

$$(CT): \quad \mathbf{AW} + \mathbf{WA}^T - \mathbf{BB}^T < 0 \qquad\qquad (DT): \quad \mathbf{AWA}^T - \mathbf{W} - \mathbf{BB}^T < 0 \tag{6.4}$$

Note that the main difference from the Lyapunov test for controllability is that in that case we required a-priori that $\mathbf{A}$ was Hurwitz (Schur), while in this case this hypothesis isn't required.

Proof 6.3: **TO BE REWRITTEN**

Stabilization with full-state feedback Starting from the Lyapunov equality (6.4) a way to **stabilize** a stabilizable system using a **full-state feedback** control is by choosing as feedback matrix $\mathbf{K}$ the one defined as $\frac{1}{2}\mathbf{B}\mathbf{W}^{-1}$.

Theorem 6.2: A pair $(\mathbf{A}, \mathbf{B})$ is stabilizable if and only if there exists a feedback matrix $\mathbf{K}$ such that the full-state feedback has a close-loop matrix $\mathbf{A}_{cl} = \mathbf{A} - \mathbf{BK}$ that's Hurwitz (Schur).

Stabilizability is an intrinsic property of the LTI system, meaning that it doesn't depend on the realization chosen (if a system is stabilizable, also all its algebraically equivalent systems are so), and knowing that the uncontrollable states *are going to zero by their own*, then we can stabilize (with a proper feedback) the controllable ones.

As final remark, stabilizability is a weaker condition the controllability: a controllable system is always stabilizable while stabilizable system might not be controllable.

6.3 Detectability

Detectability can be regarded as the dual case of stabilizability applied to the observability of a system. In fact we say that a pair $(\mathbf{C}, \mathbf{A})$ is **detectable** if it's observable decomposition (5.14) has a matrix $\mathbf{A}_{\bar{o}}$ that's either empty (meaning that the system is observable) or Hurwitz (Schur).

What this definition tells us that if the system is not observable (so $\mathbf{A}_{\bar{o}} \neq 0$), the unobservable dynamics presents a exponential converging behaviour that leads to unobservable states that are approximatively zero: this means that, even if we can't measure those states, they eventually reach the zero value by themselves.

Tests for detectability As for stabilizability, 3 tests can be developed in order to check the detectability of the system:

- *eigenvector test:* a pair $(\mathbf{C}, \mathbf{A})$ is detectable if and only if any bad eigenvector for the matrix $\mathbf{A}$ is not in the kernel of $\mathbf{C}$;
- *PBH test:* a pair $(\mathbf{C}, \mathbf{A})$ is detectable if and only if the matrix $\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I}\\ \mathbf{C}\end{bmatrix}$ is full-rank for all bad eigenvalues $\lambda \in \mathbb{C}_{bad}$;

- *Lyapunov test:* a pair $(\mathbf{C}, \mathbf{A})$ is detectable if and only if there exists a symmetric positive-definite matrix $\mathbf{P}$ such that

$$(CT): \quad \mathbf{A}^T\mathbf{W} + \mathbf{W}\mathbf{A} - \mathbf{C}^T\mathbf{C} = 0 \qquad (DT): \quad \mathbf{A}^T\mathbf{W}\mathbf{A} - \mathbf{W} - \mathbf{C}^T\mathbf{C} = 0$$

6.4 Asymptotic estimation: Luenberger observers

The main control technique presented throughout the book has been the **full-state feedback** for which well-established design methods have been studied and is allowing us to achieve the best possible results. The main drawback of such control technique is that it requires the knowledge of all states.

In general what we can measure of the dynamical system are its outputs (not directly the states), however if the **pair** $(\mathbf{C}, \mathbf{A})$ is **detectable** it is possible to **estimate** the states of the system to control. Calling $\boldsymbol{x}$ the *true* states of the system and with $\hat{\boldsymbol{x}}$ their estimation, the simplest (and intuitive) estimator is simply a copy of the original system, so presenting a dynamic

$$\dot{\hat{\boldsymbol{x}}} = \mathbf{A}\hat{\boldsymbol{x}} + \mathbf{B}\boldsymbol{u} \tag{$\star$}$$

Calling **state estimation error** $\boldsymbol{e}$ the difference $\hat{\boldsymbol{x}} - \boldsymbol{x}$, it's dynamic can be regarded as

$$\dot{\boldsymbol{e}} = \dot{\hat{\boldsymbol{x}}} - \dot{\boldsymbol{x}} = \mathbf{A}\hat{\boldsymbol{x}} + \mathbf{B}\boldsymbol{u} - \big(\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}\big) = \mathbf{A}\big(\hat{\boldsymbol{x}} - \boldsymbol{x}\big) = \mathbf{A}\boldsymbol{e} \tag{$\circ$}$$

Supposing that $\mathbf{A}$ is a stability matrix (so is Hurwitz/Schur), then we observe that the dynamics of the error converges to zero exponentially fast, and so for any arbitrary large input $\boldsymbol{u}$ we can have a good estimation $\hat{\boldsymbol{x}}$ of the true states $\boldsymbol{x}$ (after a certain time).

Luenberger observers The main problem of ($\circ$) is that usually no system has a dynamic matrix $\mathbf{A}$ that's stable as is. The main goal now is to *stabilize* the state estimation using two other quantities that are accessible to the control plant: the output $\boldsymbol{y}$ of the system and the estimated output $\hat{\boldsymbol{y}}$ of the estimator. This second quantity can be regarded simply as

$$\hat{\boldsymbol{y}} = \mathbf{C}\hat{\boldsymbol{x}} + \mathbf{D}\boldsymbol{u} \tag{$\dagger$}$$

The dynamics of the **Luenberger observer** is as in ($\star$) but additionally we include a feedback determined by the difference $\hat{y} - \boldsymbol{y}$ that's multiplied by a ***output injection matrix gain*** $\mathbf{L} \in \mathbb{R}^{n\times m}$:

$$\dot{\hat{\boldsymbol{x}}} = \mathbf{A}\hat{\boldsymbol{x}} + \mathbf{B}\boldsymbol{u} - \mathbf{L}\big(\hat{\boldsymbol{y}} - \boldsymbol{y}\big) \tag{6.5}$$

In this case the dynamic of state estimation error evaluates to

$$\dot{\boldsymbol{e}} = \dot{\hat{\boldsymbol{x}}} - \dot{\boldsymbol{x}} = \mathbf{A}\hat{\boldsymbol{x}} + \mathbf{B}\boldsymbol{u} - \mathbf{L}\big(\hat{\boldsymbol{y}} - \boldsymbol{y}\big) - \big(\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}\big) = \mathbf{A}\big(\hat{\boldsymbol{x}} - \boldsymbol{x}\big) - \mathbf{L}\big(\hat{\boldsymbol{y}} - \boldsymbol{y}\big)$$

Considering ($\dagger$) this expression further simplifies to

$$\begin{aligned}\dot{\boldsymbol{e}} &= \mathbf{A}(\hat{\boldsymbol{x}} - \boldsymbol{x}) - \mathbf{L}\Big(\mathbf{C}\hat{\boldsymbol{x}} + \mathbf{D}\boldsymbol{u} - \big(\mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u}\big)\Big) = \big(\mathbf{A} - \mathbf{L}\mathbf{C}\big)\big(\hat{\boldsymbol{x}} - \boldsymbol{x}\big) \\ &= \big(\mathbf{A} - \mathbf{L}\mathbf{C}\big)\boldsymbol{e}\end{aligned} \tag{6.6}$$

This also allows us to rewrite (6.5) as

$$\dot{\hat{\boldsymbol{x}}} = \big(\mathbf{A} - \mathbf{L}\mathbf{C}\big)\hat{\boldsymbol{x}} + \big(\mathbf{B} - \mathbf{L}\mathbf{D}\big)\boldsymbol{u} + \mathbf{L}\boldsymbol{y} \tag{6.7}$$

From (6.6) we can see that in order to have a exponentially converging behaviour of the state estimation error we need to design a output injection matrix $\mathbf{L}$ that stabilizes the closed-loop $\mathbf{A} - \mathbf{L}\mathbf{C}$: this operation is similar to what has been presented in section 4.4, page 30, while introducing the full-state feedback and in particular the eigenvalue assignment procedure.
In that case we showed how to *tune* K in the close-loop $\mathbf{A} - \mathbf{B}\mathbf{K}$, however in this case what we have that the position of the unknown ($\mathbf{L}$ in this case) is swapped with respect to the known $\mathbf{C}$. In order to still use the same eigenvalue assignment procedure previously shown, we can consider the transpose

$$\big(\mathbf{A} - \mathbf{L}\mathbf{C}\big)^T = \mathbf{A}^T - \mathbf{C}^T\mathbf{L}^T = \overline{\mathbf{A}} - \overline{\mathbf{B}\mathbf{K}}$$

Observing that the transposition operation preserves the eigenvalues of the system, then by designing the proper feedback matrix $\overline{\mathbf{K}}$ of the close-loop $\overline{\mathbf{A}} - \overline{\mathbf{B}}\overline{\mathbf{K}}$ with the technique shown at page 30, we finally have that the output injection matrix can be computed as

$$\mathbf{L} = \overline{\mathbf{K}}^T$$

Theorem 6.3: There exists a matrix $\mathbf{L}$ such that the feedback $\mathbf{A} - \mathbf{LC}$ is a stability matrix (Hurwitz/Schur) if and only if $(\mathbf{C}, \mathbf{A})$ is a detectable pair.

This tells us, similarly to what has been said for stabilizability, the if we can find a matrix $\mathbf{L}$ that stabilizes the close loop $\mathbf{A} - \mathbf{LC}$ then we are sure the the pair $(\mathbf{C}, \mathbf{A})$ is detectable.

Theorem 6.4: Given any selection of eigenvalues $\lambda_1, \dots, \lambda_n$, there exists a matrix $\mathbf{L}$ such that $\mathbf{A} - \mathbf{LC}$ has those eigenvalues if and only if the pair $(\mathbf{C}, \mathbf{A})$ is **observable**.

Observe that this is a stronger requirement than the one presented in theorem 6.3: in fact if we prove that, by choosing a proper $\mathbf{L}$, we can assign any desired set of eigenvalues, then we ensure that the system is observable (that's a stronger condition then detectability).

6.4.1 Minimum energy estimation

Dually to the linear quadratic regulator, page 22, we can use a functional definition to chose the proper feedback matrix $\mathbf{L}$ for the Luenberger states estimator. Given a LTI system with $\mathbf{D} = 0$, process disturbance $\boldsymbol{d}$ and sensor noises $\boldsymbol{n}$, the state-space representation is

$$\begin{cases} \dot{\boldsymbol{x}} &= \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u} + \overline{\mathbf{B}}\boldsymbol{d} \\ \boldsymbol{y} &= \mathbf{C}\boldsymbol{x} + \boldsymbol{n} \end{cases}$$

The goal od the **minimum energy estimator** is so to find an estimate $\hat{\boldsymbol{x}} = \boldsymbol{x}^*$ of the states, so for which the system

$$\begin{cases} \dot{\hat{\boldsymbol{x}}} &= \mathbf{A}\hat{\boldsymbol{x}} + \mathbf{B}\boldsymbol{u} + \overline{\mathbf{B}}\hat{\boldsymbol{d}} \\ \boldsymbol{y} &= \mathbf{C}\hat{\boldsymbol{x}} + \hat{\boldsymbol{n}} \end{cases}$$

minimize the cost

$$\mathcal{J} = \min_{\hat{\boldsymbol{x}}, \hat{\boldsymbol{d}}, \hat{\boldsymbol{n}}} \int_{-\infty}^{0} \hat{\boldsymbol{n}}^T(\tau)\mathbf{Q}\hat{\boldsymbol{n}}(\tau) + \hat{\boldsymbol{d}}^T(\tau)\mathbf{R}\hat{\boldsymbol{d}}(\tau)\, d\tau$$

where $\mathbf{R}$ measures *how good we consider our model* and $\mathbf{Q}$ *how much we trust the sensors*. The solution of this minimal energy problem is obtained considering the **dual Riccati equation** defined as

$$\mathcal{J}^* = \int_{-\infty}^{t} \big(\mathbf{C}\hat{\boldsymbol{x}}(\tau) - \boldsymbol{y}(\tau)\big)^T \mathbf{Q}\big(\mathbf{C}\hat{\boldsymbol{x}}(\tau) - \boldsymbol{y}(\tau)\big) - \hat{\boldsymbol{d}}^T(\tau)\mathbf{R}\hat{\boldsymbol{d}}(\tau)\, d\tau \tag{6.8}$$

From this we can also develop the feedback invariants similar to (3.21), page 24, that are

$$\mathbf{AS} + \mathbf{SA}^T + \hat{\mathbf{B}}\mathbf{R}^{-1}\hat{\mathbf{B}}^T - \mathbf{SC}^T\mathbf{QCS} = 0 \tag{6.9}$$

Theorem 6.5: If there exists a symmetric positive-definite matrix $\mathbf{S}$ solution of (6.9) such that $\mathbf{A} - \mathbf{SC}^T\mathbf{QC}$ is Hurwitz, then the minimum energy estimation problem is solved by a Luenberger observer with feedback matrix

$$\mathbf{L} = \mathbf{SC}^T\mathbf{Q} \tag{6.10}$$

Theorem 6.6: Assume that the pair $(\mathbf{A}, \hat{\mathbf{B}})$ is stabilizable and that $(\mathbf{C}, \mathbf{A})$ is a detectable pair, then we are sure that exists $\mathbf{S} = \mathbf{S}^T > 0$ that determines a *stable* solution of (6.9).

Note that both algebraic Riccati equation (3.21) and it's dual (6.9) are containing quadratic terms, implying that generally the equation has **two solution**: one of them is **stabilizing** the system while the other one is **de-stabilizing**, so we have to carefully choose the correct solution.

Linear quadratic Gaussian LQG interpretation of the minimum energy estimation Considering now a system whose disturbance $\boldsymbol{d}$ and noise $\boldsymbol{n}$ come from uncorrelated stochastic processes normally distributed with a 0 mean, then the covariances are $\mathbb{E}\{\boldsymbol{d}(\tau)\boldsymbol{d}^T(\tau)\} = \delta(t-\tau)\mathbf{R}^{-1}$ and $\mathbb{E}\{\boldsymbol{n}(\tau)\boldsymbol{n}^T(\tau)\} = \delta(t-\tau)\mathbf{Q}^{-1}$, then the LQG (minimum energy estimator) must minimize the stochastic cost

$$\mathcal{J}_{LQG} = \mathbb{E}\{|\boldsymbol{x}(t) - \hat{\boldsymbol{x}}(t)|^2\} \tag{6.11}$$

6.5 Dynamic output feedback via state estimation

The full-state feedback design described in page 30 (sec. 4.4) was based on the assumption that all states $\boldsymbol{x}$ were accessible from the controller in order to determine the optimal control; however as was described later, this barely happens as some states could be hidden to the outside of the system: to overcome this problem state observers has been introduce in order to asymptotically estimate the states of the plant. The main idea now is to combine this powerful tools together in order to exploit the full-state feedback design on a stabilizable system if this system is also detectable (that's the requirement for the design of the observer).

Theorem 6.7: Given a **stabilizable** and **detectable plant** $\dot{\boldsymbol{x}}/\boldsymbol{x}^+ = \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}$, $\boldsymbol{y} = \mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u}$ and any selection of feedback matrices $(\mathbf{K}, \mathbf{L})$, then the **close loop system** with a **controller**

$$\begin{cases} \dot{\hat{\boldsymbol{x}}}/\hat{\boldsymbol{x}}^+ &= (\mathbf{A}-\mathbf{LC})\hat{\boldsymbol{x}} + (\mathbf{B}-\mathbf{LD})\boldsymbol{u} + \mathbf{L}\boldsymbol{y} \\ \boldsymbol{u} &= -\mathbf{K}\hat{\boldsymbol{x}} \end{cases}$$

is a linear system with **eigenvalues** corresponding to the union of the eigenvalues of $\mathbf{A}-\mathbf{LC}$ and $\mathbf{A}-\mathbf{BK}$.

The main takeaway of this theorem is that we can independently design both the close loop for stabilizing a system (design of $\mathbf{K}$) and the observer (design of $\mathbf{L}$) and still ensure that the combination of this two *sub-components* gives the desired behaviour.

Proof 6.4: In order to represent the close loop system, we can use the coordinate/states representation $(\boldsymbol{e}, \boldsymbol{x}) = (\hat{\boldsymbol{x}} - \boldsymbol{x}, \boldsymbol{x})$; the dynamics of the system is so described by

$$\begin{aligned} \dot{\boldsymbol{e}}/\boldsymbol{e}^+ &= \dot{\hat{\boldsymbol{x}}} - \dot{\boldsymbol{x}} = (\mathbf{A}-\mathbf{LC})\hat{\boldsymbol{x}} + (\mathbf{B}-\mathbf{LD})\boldsymbol{u} + \mathbf{L}(\mathbf{C}\boldsymbol{x}+\mathbf{D}\boldsymbol{u}) - \mathbf{A}\boldsymbol{x} - \mathbf{B}\boldsymbol{u} \\ &= (\mathbf{A}-\mathbf{LC})\hat{\boldsymbol{x}} + \mathbf{LC}\boldsymbol{x} = (\mathbf{A}-\mathbf{LC})\hat{x} - (\mathbf{A}-\mathbf{LC})\boldsymbol{x} = (\mathbf{A}-\mathbf{LC})\boldsymbol{e} \\ \dot{\boldsymbol{x}}/\boldsymbol{x}^+ &= \mathbf{A}\boldsymbol{x} + \mathbf{B}(\underbrace{-\mathbf{K}\hat{x}}_{=\boldsymbol{u}}) = \mathbf{A}\boldsymbol{x} - \mathbf{BK}(\underbrace{\boldsymbol{e}+\boldsymbol{x}}_{=\hat{\boldsymbol{x}}}) = (\mathbf{A}-\mathbf{BK})\boldsymbol{x} - \mathbf{BK}\boldsymbol{e} \end{aligned}$$

Exploiting the matrix representation we have

$$\begin{pmatrix} \dot{\boldsymbol{e}}/\boldsymbol{e}^+ \\ \dot{\boldsymbol{x}}/\boldsymbol{x}^+ \end{pmatrix} = \begin{bmatrix} \mathbf{A}-\mathbf{LC} & 0 \\ -\mathbf{BK} & \mathbf{A}-\mathbf{BK} \end{bmatrix} \begin{pmatrix} \boldsymbol{e} \\ \boldsymbol{x} \end{pmatrix}$$

Being the dynamic matrix lower-triangular, the eigenvalues are the one associated to the terms in the principal diagonal, thus are the union of the eigenvalues of $\mathbf{A}-\mathbf{LC}$ and $\mathbf{A}-\mathbf{BK}$.

As a rule of thumb, the estimator should be *faster* then the dynamics feedback in order to have a better behaviour of the system.

In general the dynamics of the controller is described by the state-space representation

$$\begin{cases} \dot{\hat{\boldsymbol{x}}}/\hat{\boldsymbol{x}}^+ = (\mathbf{A}-\mathbf{LC})\hat{\boldsymbol{x}} + (\mathbf{B}-\mathbf{LD})\boldsymbol{u} + \mathbf{L}\boldsymbol{y} & : \text{dynamics} \\ \qquad\qquad = (\mathbf{A}-\mathbf{LC})\hat{\boldsymbol{x}} + (\mathbf{B}-\mathbf{LD})(-\mathbf{K}\hat{\boldsymbol{x}}) + \mathbf{L}\boldsymbol{y} & \\ \qquad\qquad = (\mathbf{A}-\mathbf{LC}-\mathbf{BK}+\mathbf{LDK})\hat{\boldsymbol{x}} + \mathbf{L}\boldsymbol{y} & \\ \boldsymbol{u} = -\mathbf{K}\hat{\boldsymbol{x}} & : \text{output} \end{cases} \tag{6.12}$$

leading the the following strictly-proper controller transfer function:

$$\boldsymbol{\mathcal{K}}(s) = -\mathbf{K}\Big(s\mathbf{I} - \big(\mathbf{A} - \mathbf{LC} - \mathbf{BK} + \mathbf{LDK}\big)\Big)^{-1}\mathbf{L} \tag{6.13}$$

Note that in general the dynamics of the controller can be unstable, but the important thing is that stabilizes the close loop. We can also observe that the *tuning* of the matrix $\mathbf{L}$ can be based on a performance output $\boldsymbol{z}$ that can differ from the output $\boldsymbol{y}$ considered while designing the full-state feedback.

6.6 BIBO stability

BIBO stability is an ***external stability*** concept; in particular BIBO stands for *bounded-input bounded-output* and describes the fact that if a system is subjected to any bounded input, then also the corresponding zero-state response is bounded. Defining the infinity norm of a solution as

$$\|\boldsymbol{z}(\cdot)\|_\infty = \sup_{\tau\geq 0} |\boldsymbol{z}(\tau)| \tag{6.14}$$

so as the maximum value assumed by the vector $\boldsymbol{z}(t)$, we say that a LTV system $\dot{\boldsymbol{x}} = \mathbf{A}(t)\boldsymbol{x} + \mathbf{B}(t)\boldsymbol{u}$, $\boldsymbol{y} = \mathbf{C}(t)\boldsymbol{x} + \mathbf{D}(t)\boldsymbol{u}$ is **BIBO stable** if exists a constant $g > 0$ such that for any bounded input $\boldsymbol{u}(t)$ (with $t \geq 0$), the zero-state output response satisfies

$$\|\boldsymbol{y}(\cdot)\|_\infty \leq g\|\boldsymbol{u}(\cdot)\|_\infty \tag{6.15}$$

Theorem 6.8: For a continuous-time LTV system the following statements are equivalent:

i) (CT-LTV) is BIBO stable;

ii) *(a)* each entry of $\mathbf{D}(\cdot)$ is uniformly bounded, so

$$\exists \delta > 0 \quad \text{such that} \quad |\mathbf{D}(t)| < \delta \quad \forall t \geq 0$$

(b) each entry $g_{ij}(\cdot,\cdot)$ of the matrix $\mathbf{C}(t)\boldsymbol{\Phi}(t,\tau)\mathbf{B}(\tau)$ generates a converging integral:

$$\sup_{t\geq 0} \int_0^t \big|g_{ij}(t,\tau)\big|\, d\tau = \lim_{t\to\infty} \int_0^t |g_{ij}(t,\tau)|\, d\tau < \infty \quad \forall i,j$$

This implies

$$\exists G > 0 \quad \text{such that} \quad \int_0^\infty |g_{ij}(t,\tau)|\, d\tau < G \quad \forall t \geq 0, \forall i,j$$

Proof 6.5:

a) firstly we can show that *ii)* implies *i)*; applying the norm on the variation of constants formula (2.3) what we have is

$$\begin{aligned}
|\boldsymbol{y}(t)| &\leq \left|\int_0^t \mathbf{C}(\tau)\boldsymbol{\Phi}(t,\tau)\mathbf{B}(\tau)\boldsymbol{u}(\tau)\, d\tau\right| + \big|\mathbf{D}(t)\boldsymbol{u}(t)\big| \\
&\leq \int_0^T \big|\mathbf{C}(\tau)\boldsymbol{\Phi}(t,\tau)\mathbf{B}(\tau)\big|\boldsymbol{u}_\infty\, d\tau + \big|D(t)\big|u_\infty \\
&\leq pkGu_\infty + \delta u_\infty = \big(pkG + \delta\big)u_\infty = gu_\infty
\end{aligned}$$

where p, k are respectively the number of input and output of the system; this definition matches (6.15);

b) to show that *i)* implies *ii)* we prove by contradiction that if *ii)* is not satisfied, then also *i)* is not satisfied. Reversing the definition of BIBO stability, for each candidate g^* exists an

input $\boldsymbol{u}^*$ for which $\|\boldsymbol{y}(\cdot)\|_\infty > g^*\|\boldsymbol{u}^*(\cdot)\|_\infty$. If we consider that *ii)(a)* fails, then for any g^* exists a time t^* such that $|d_{ij}(t^*)| > g^*$ for some i, j; selecting so the control

$$\boldsymbol{u}_j^* = \begin{cases} 0 & t < t^* \\ 1 & t \geq t^* \end{cases}$$

then the corresponding output $y(t)$ evaluated at t^* disregard the BIBO stability condition (6.15):

$$\boldsymbol{y}(t^*) = 0 + \mathbf{D}(t^*)\boldsymbol{u}_j^*(t^*) = d_{ij}(t^*) > g^*$$

MISSES THE PROF WHEN *ii)(a)* FAILS

LTI case Considering now the simpler case of continuous-time LTI systems, the product $\mathbf{C}(\tau)\mathbf{\Phi}(t,\tau)\mathbf{B}(\tau)$ reduces to $\mathbf{C}e^{\mathbf{A}(t-\tau)}\mathbf{B}$ that can be regarded as $\overline{\mathbf{G}}(t-\tau)$. The condition of BIBO stability in this case collapses in the analysis of the entries $\overline{g}_{ij}(t-\tau)$ of $\overline{\mathbf{G}}$, whose norm is regarded as $\int_0^t |\overline{g}_{ij}(t-\tau)|\,d\tau$; performing a change of variable $\rho = t - \tau$ what we obtain is

$$\int_t^0 |\overline{g}_{ij}(\rho)|(-1)\,d\rho = \int_0^t |\overline{g}_{ij}(\rho)|\,d\rho$$

Theorem 6.9: For a continuous-time LTI system the following are equivalent:

i) (CT-LTI) is BIBO-stable;

ii) each entry $\overline{g}_{ij}(\cdot)$ of $\overline{\mathbf{G}}(\rho) = \mathbf{C}e^{\mathbf{A}\rho}\mathbf{B}$ satisfies

$$\lim_{t\to\infty} \int_0^t |\overline{g}_{ij}(\rho)|\,d\rho < \infty \qquad \Leftrightarrow \qquad \exists G > 0 \;:\; \int_0^t |\overline{g}_{ij}(\rho)|\,d\rho \leq G$$

iii) each entry of the transfer function matrix $\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$ has exponentially converging poles.

Proof 6.6: The relation between *i)* and *ii)* is straightforward due to theorem 6.8. We can now relate *iii)* with *i)*: computing the Laplace transform of the component $\overline{g}_{ij}(\rho)$ of the transfer function $\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$ that evaluates to the proper rational polynomial

$$\begin{aligned}\mathscr{L}\{\overline{g}_{ij}(\rho)\}(s) &= \frac{\alpha_0 s^q + \dots \alpha_{q-1}s + \alpha_q}{(s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2}\dots(s-\lambda_k)^{m_k}} \\ &= \frac{a_{11}}{s-\lambda_1} + \frac{a_{12}}{(s-\lambda_1)^2} + \dots + \frac{a_{1m_1}}{(s-\lambda_1)^{m_1}} + \frac{a_{21}}{s-\lambda_2} + \dots\end{aligned}$$

With the partial fraction expansion performed, we can anti-transform the transfer function to obtain the impulse response function of time that's of the form

$$\overline{g}_{ij}(\rho) = e^{\lambda_1\rho}p_{m_1}(\rho) + e^{\lambda_2\rho}p_{m_2}(\rho) + \dots$$

where $p_m(\rho)$ are all polynomials in ρ of order $m-1$. The asymptotic behaviour of such function is determined by the exponentials that in order to be convergent for any bounded input must have $\mathbb{R}\mathrm{e}\{\lambda_i\} < 0$ for any eigenvalue λ_i.

As a general statement while computing the impulse response of the system pole cancellations might arise while computing the transfer function: what we can so say is that if $\mathbf{A}$ is Hurwitz, then for sure the system is BIBO stable (as all poles surely have negative real part), while the converse might not always be true (*unstable* poles might have been cancelled in the computation of the transfer function).

6.7 Minimal realizations and Markov parameters

Recalling what has already said at page 6, an LTI system describe in state-space as

$$\begin{cases} \dot{\boldsymbol{x}}/\boldsymbol{x}^+ &= \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u} \\ \boldsymbol{y} &= \mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u} \end{cases} \tag{LTI}$$

is a realization of a transfer function $\hat{G}(s)$ if it holds $\hat{\mathbf{G}}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$.

Theorem 6.10: A realization of $\hat{\mathbf{G}}(s)$ is **minimal** (or ***irreducible***) if there is no other realization with lower order, so with a fewer number of states.

Theorem 6.11: Every realization is **minimal** if and only if it's both **controllable** and **observable**.

Proof 6.7: The proof of what has been just states comes from theorem 6.1: assuming that the system is not controllable and/or not observable, then the observable and controllable component $\mathbf{A}_{co}$ of the dynamics is *smaller* then the whole matrix $\mathbf{A}$; having that the transfer function of the system is associated to the observable and controllable component $\hat{\mathbf{G}}(s) = \mathbf{C}_{co}(s\mathbf{I} - \mathbf{A}_{co})^{-1}\mathbf{B}_{co} + \mathbf{D}$, then we have that some states are not being tracked by the transfer function.

This prove that a minimal realization is both controllable and observable, however the converse is much harder to show and it involves the definition of the so called **Markov parameters**.
Considering that the term $(s\mathbf{I} - \mathbf{A})^{-1}$ is the Laplace transform of the matrix exponential $e^{\mathbf{A}t}$, what we can see is that it can be rewritten (considering also the Taylor series expansion of the exponential) as

$$(s\mathbf{I} - \mathbf{A})^{-1} = \mathscr{L}\{e^{\mathbf{A}t}\} = \mathscr{L}\left\{\sum_{i=0}^{\infty} \frac{\mathbf{A}^i t^i}{i!}\right\} = \sum_{i=0} \mathbf{A}^i \mathscr{L}\left\{\frac{t^i}{i!}\right\} = \sum_{i=0}^{\infty} \frac{\mathbf{A}^i}{s^{i+1}}$$

thus the transfer function

$$\hat{\mathbf{G}}(s) = \mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B} + \mathbf{D} = \sum_{i=0}^{\infty} \frac{\mathbf{C}\mathbf{A}^i\mathbf{B}}{s^{i+1}} + \mathbf{D} \tag{6.16}$$

Having decomposed the transfer function into a series of matrices, we refer to the sequence $\mathbf{D}$, $\mathbf{CB}$, $\mathbf{CAB}$, $\mathbf{CA}^2\mathbf{B}$, $\mathbf{CA}^3\mathbf{B}$, ... as **Markov parameters** as they are able to fully describe the transfer function $\mathbf{G}(s)$. Such values cannot be computed exploiting the Cayley-Hamilton theorem as realization of different order results in different dimension of $\mathbf{A}$ (making impossible to compare them using those theorem). Markov parameters can instead be computed using the **impulse response** $\mathbf{G}(t) = \mathscr{L}^{-1}\{\hat{\mathbf{G}}(s)\} = \mathbf{C}e^{\mathbf{A}t}\mathbf{B} + \mathbf{D}\delta(t)$ of the system. In particular the first Markov parameter $\mathbf{D}$ can be simply computed as

$$\mathbf{D} = \lim_{s\to\infty} \hat{G}(s) \qquad \Rightarrow \qquad \hat{\mathbf{G}}_{sp}(s) = \mathbf{C}\big(s\mathbf{I} - \mathbf{A}\big)^{-1}\mathbf{B}$$

Having $\mathbf{G}_{sp}(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{B}$ as the impulse response of the strictly proper transfer function, evaluating it at $t = 0$ results to

$$\hat{\mathbf{G}}_{sp}(0) = \mathbf{CB}$$

Deriving in time $\mathbf{G}_{sp}(t)$ and evaluating it for the same point evaluates to

$$\left.\frac{d\mathbf{G}(t)}{dt}\right|_{t=0} = \left.\frac{d}{dt}\left(\mathbf{C}e^{\mathbf{A}t}\mathbf{B}\right)\right|_{t=0} = \left.\mathbf{C}\mathbf{A}e^{\mathbf{A}t}\mathbf{B}\right|_{t=0} = \mathbf{CAB}$$

and following

$$\left.\frac{d^2\mathbf{G}(t)}{dt^2}\right|_{t=0} = \left.\frac{d}{dt}\left(\mathbf{C}\mathbf{A}e^{\mathbf{A}t}\mathbf{B}\right)\right|_{t=0} = \left.\mathbf{C}\mathbf{A}^2e^{\mathbf{A}t}\mathbf{B}\right|_{t=0} = \mathbf{CA}^2\mathbf{B}$$

Theorem 6.12: Two different realizations are zero-state equivalent if and only if they have the same Markov parameters.

Proof 6.8: With this premise we can now show that if a system is both controllable and observable it leads to a minimal realization as stated in theorem 6.11. Let us assume that $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ with n states that's a controllable and observable realization of $\hat{G}(s)$ that's not minimal: this implies that exists the quadruple $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{D}})$ that's a minimal realization with dimension $\overline{n} < n$.
Computing the product between the observability and controllability matrix evaluates to

$$\mathbf{OR} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} \begin{bmatrix} \mathbf{B} & \mathbf{AB} & \dots & \mathbf{A}^{n-1}\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{CB} & \mathbf{CAB} & \dots & \mathbf{CA}^{n-1}\mathbf{B} \\ \mathbf{CAB} & \mathbf{CA}^2\mathbf{B} & & \\ \vdots & & \ddots & \\ \mathbf{CA}^{n-1}\mathbf{B} & & & \mathbf{CA}^{2n-2}\mathbf{B} \end{bmatrix}$$

that has rank r; computing instead $\overline{\mathbf{OR}}$ evaluates to a $\overline{n}$ rank matrix

$$\overline{\mathbf{OR}} = \begin{bmatrix} \overline{\mathbf{CB}} & \overline{\mathbf{CAB}} & \dots & \overline{\mathbf{CA}}^{\overline{n}-1}\overline{\mathbf{B}} \\ \overline{\mathbf{CAB}} & \overline{\mathbf{CA}}^2\overline{\mathbf{B}} & & \\ \vdots & & \ddots & \\ \overline{\mathbf{CA}}^{\overline{n}-1}\overline{\mathbf{B}} & & & \overline{\mathbf{CA}}^{2\overline{n}-2}\overline{\mathbf{B}} \end{bmatrix}$$

Exploiting theorem 6.12, having that the two system are realizations of the same transfer function they must share the same Markov parameters, implying so that $\mathbf{OR} = \overline{\mathbf{OR}}$, however we can clearly see that

$$\text{rank}\{\overline{\mathbf{OR}}\} \leq \text{rank}\{\overline{\mathbf{C}}\} \leq \overline{n} < n = \text{rank}\{\mathbf{OR}\}$$

contradicting so the the initial statement.

Theorem 6.13: All minimal realizations are algebraically equivalent.

To prove such statement we firstly need to recall the definitions of **Moore-Penrose pseudo inverses** matrices. Given a *fat* full column rank matrix $\mathbf{M}$ characterized so by linearly independent columns and a higher number of rows, then the matrix $\mathbf{M}^T\mathbf{M}$ is non-singular and $\mathbf{M}^l = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ is called **left-inverse** of $\mathbf{M}$ as it holds that $\mathbf{M}^l\mathbf{M} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{M} = \mathbf{I}$.
Given instead a *fat* full row rank matrix $\mathbf{N}$, then the square matrix $\mathbf{NN}^T$ is invertible and $\mathbf{N}^r = \mathbf{N}^T(\mathbf{NN}^T)^{-1}$ is called **right inverse** of $\mathbf{N}$ as $\mathbf{NN}^r = \mathbf{NN}^T(\mathbf{NN}^T)^{-1} = \mathbf{I}$.

Proof 6.9: Given two minimal realizations $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ and $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{D}})$ that are so controllable and observable and also are sharing the same Markov parameters (so $\mathbf{D} = \overline{\mathbf{D}}$), meaning that $\mathbf{OR} = \overline{\mathbf{OR}}$. We can show now that the transformation matrix

$$\mathbf{T} = \overline{\mathbf{R}}\mathbf{R}^r = \overline{\mathbf{R}}\mathbf{R}^T\left(\mathbf{R}\mathbf{R}^T\right)^{-1} \tag{6.17}$$

is the one relating the two algebraically equivalent matrices. For this reason we show that

a) $\mathbf{T}$ is invertible: considering in fact $\mathbf{IT} = \mathbf{O}^l\mathbf{OT} = \mathbf{O}^l\overline{\mathbf{OR}}\mathbf{R}^r = \mathbf{O}^l\mathbf{OR}\mathbf{R}^t = \mathbf{II} = \mathbf{I}$, implying so that $\mathbf{T}$ is invertible and in particular

$$\mathbf{T}^{-1} = \mathbf{O}^l\overline{\mathbf{O}}$$

b) we show that the transformation $\mathbf{C} = \overline{\mathbf{C}}\mathbf{T}$, $\mathbf{B} = \mathbf{T}^{-1}\overline{\mathbf{B}}$ and $\mathbf{A} = \mathbf{T}^{-1}\overline{\mathbf{A}}\mathbf{T}$ are holding. Recalling that $\mathbf{OR} = \overline{\mathbf{OR}}$, then we can rewrite $\mathbf{O} = \mathbf{ORR}^r = \overline{\mathbf{OR}}\mathbf{R}^r = \overline{\mathbf{O}}\mathbf{T}$. Computing so the

product

$$\begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{C}} \\ \overline{\mathbf{CA}} \\ \vdots \\ \overline{\mathbf{CA}}^{n-1} \end{bmatrix} \mathbf{T} = \begin{bmatrix} \overline{\mathbf{C}}\mathbf{T} \\ \overline{\mathbf{CA}}\mathbf{T} \\ \vdots \\ \overline{\mathbf{CA}}^{n-1}\mathbf{T} \end{bmatrix}$$

thus considering just the first row we showed that $\mathbf{C} = \overline{\mathbf{C}}\mathbf{T}$.

If instead we consider $\overline{\mathbf{O}}^l\mathbf{O} = \overline{\mathbf{O}}^l\overline{\mathbf{O}}\mathbf{T} = \mathbf{T}$, then we can notice that $\mathbf{TR} = \overline{\mathbf{O}}^l\mathbf{OR} = \overline{O}^l\overline{\mathbf{OR}} = \overline{\mathbf{R}}$, thus

$$\overline{\mathbf{R}} = \begin{bmatrix} \overline{\mathbf{B}} & \overline{\mathbf{AB}} & \dots & \overline{\mathbf{A}}^{n-1}\overline{\mathbf{B}} \end{bmatrix} = \begin{bmatrix} \mathbf{TB} & \mathbf{TAB} & \dots & \mathbf{TA}^{n-1}\mathbf{B} \end{bmatrix}$$

Considering so the first column we have that $\overline{\mathbf{B}} = \mathbf{TB}$, thus $\mathbf{B} = \mathbf{T}^{-1}\overline{\mathbf{B}}$.

Finally we can notice that while computing the product $\mathbf{OR}$ we have that the matrix $\mathbf{A}$ always *sticks in the middle* of matrices $\mathbf{C}$ and $\mathbf{B}$, thus if we compute $\mathbf{OAR}$ still evaluates in a matrix of Markov parameters that must be shared by $\overline{\mathbf{OAR}}$ as the system are realization of the same transfer function. Considering this equality and pre-multiplying it bu $\overline{\mathbf{O}}^l$ and post-multiplying by $\mathbf{R}^r$ results in

$$\overline{\mathbf{O}}^l\overline{\mathbf{O}}\mathbf{A}\cancel{\mathbf{R}\mathbf{R}^r} = \cancel{\overline{\mathbf{O}}^l\overline{\mathbf{O}}}\,\overline{\mathbf{A}}\overline{\mathbf{R}}\mathbf{R}^r \qquad \Rightarrow \qquad \mathbf{TA} = \overline{\mathbf{A}}\mathbf{T}$$

confirming so that $\mathbf{A} = \mathbf{T}^{-1}\overline{\mathbf{A}}\mathbf{T}$.

Chapter 7

Hybrid dynamical systems

Hybrid dynamical systems are the ones that are allowed both in continuous and discrete-time, thus it's general state-space representation is in the form

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{C} \quad : \text{continuous evolution} \\ \boldsymbol{x}^+ = \boldsymbol{g}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{D} \quad : \text{discrete evolution} \end{cases} \tag{H}$$

In this representation $\boldsymbol{f}$ is the **flow map** describing the evolution of the states in a continuous-time environment while $\boldsymbol{x}$ are in the **flow set** $\mathcal{C}$; $\mathcal{D}$ is instead the **jump set** and describe the *region* of the states where the system is allowed to *jump* with a discrete-time behaviour according to the **jump map** $\boldsymbol{g}$. Observe that in general $\mathcal{C} \cup \mathcal{D} \neq \mathbb{R}^n$ and $\mathcal{C} \cap \mathcal{D} \neq \emptyset$.

The simplest, yet effective, example is the one of a bouncing ball: given a point mass $m = 1kg$ with states (p, v) representing respectively the position and the velocity in the vertical direction (*pointing upward* with respect to ground that's at $p = 0$), then every-time the mass is above ground ($p > 0$) it's allowed to flow according to the Newton equation. Calling so $\gamma = 9.81N/kg$ the force/mass applied to the point mass, then the flow can be described as

$$\dot{\boldsymbol{x}} = \begin{pmatrix} \dot{p} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} v \\ -\gamma \end{pmatrix} = \boldsymbol{f}(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathcal{C} = \{\boldsymbol{x} = (p, v) \in \mathbb{R}^2 \ : \ p \geq 0\}$$

When the ball finally reaches ground ($p = 0$), when it's *going downward* ($v < 0$) then the bounce suddenly changes the direction of the speed (becoming positive); considering in general that in the shock some energy is lost, calling $\lambda \in [0, 1]$ the so called *restitution factor*, we can model the jumps of the systems as

$$\boldsymbol{x}^+ = \begin{pmatrix} p^+ \\ v^+ \end{pmatrix} = \begin{pmatrix} p \\ -\lambda v \end{pmatrix} = \boldsymbol{g}(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathcal{D} = \{\boldsymbol{x} = (p, v) \in \mathbb{R}^2 \ : \ p = 0 \text{ and } v \leq 0\}$$

While modelling hybrid systems, the definition of flow and jump set is of primary relevance; recalling this example, intuitively one other flow set could have been $\mathcal{C} = \{(p, v) \ : \ p > 0\}$, however this comes the first *requirement* that we won't prove (but is actually important while finding solutions of hybrid system): both **flow** and **jump** domains must be **close set**, so such that *their boundaries are included*. Direct consequence of this choice is that $\mathcal{C}$ and $\mathcal{D}$ might *overlap*, leading to a **non-unique solution** given the same initial condition.

7.1 Solutions

Hybrid time domain Given the mathematical description of an hybrid system, the associated solution must be parametrized in both continuous and discrete-time domain: for this reason we define the **hybrid time domain** $\mathbb{E}$ the subset of $\mathbb{R}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ defined as

$$\mathbb{E} = \bigcup_{j=0}^{J} I_j \times \{j\} \qquad \text{where } I_j = [t_j, t_{j+1}] \tag{7.1}$$

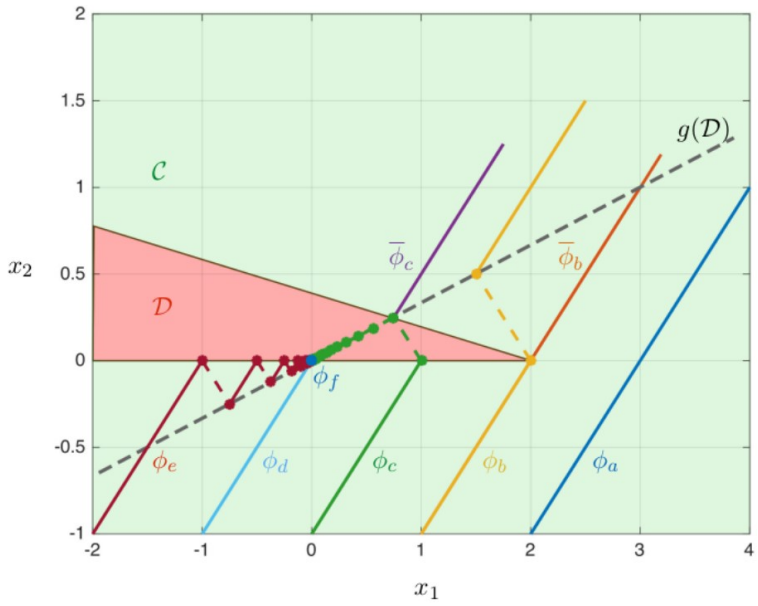

Figure 7.1: graphical representation of the solutions of (†).

where j are the discrete-time jumps and $I_j = [t_j, t_{j+1}]$ are flow intervals between them. Possibly what we want in the solutions is $J \to \infty$ and that the last interval I_j is *open on the right*. Intuitively while regarding hybrid system we say that *the time $t + j$ is always moving forward*. From this definition we call $t_0, t_1, \ldots, t_j$ as the **jump times** and must satisfy

$$0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_J$$

Solution of a hybrid system A **solution** of an hybrid system (H) is a **function** $\boldsymbol{\phi}(t, j)$ such that:

1. the domain of $\boldsymbol{\phi}(t, j)$ is an hybrid time domain as in (7.1);
2. for each intervals $(t_1, j), (t_2, j) \in \text{dom}\{\boldsymbol{\phi}\}$ with $t_2 > t_1$, then we must have

$$\frac{d}{dt}\boldsymbol{\phi}(t, j) = \boldsymbol{f}(\boldsymbol{\phi}(t, j)) \tag{F}$$

   and $\boldsymbol{\phi}(t, j) \in \mathcal{C}$ is in the flow set for almost all $t \in [t_1, t_2]$ (excluding in particular the jump times);
3. for each $(t, j), (t, j+1) \in \text{dom}\{\boldsymbol{\phi}\}$ then it must hold

$$\boldsymbol{\phi}(t, j+1) = \boldsymbol{g}(\boldsymbol{\phi}(t, j)) \tag{J}$$

   and $\boldsymbol{\phi}(t, j) \in \mathcal{D}$ lies in the jump set.

In general we regard (F) as the **flow condition** and (J) as the **jump**.

Considering now an example of hybrid system characterized by the state-space representation

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \boldsymbol{x} \in \mathcal{C} = \overline{\mathbb{R}^2 \setminus \mathcal{D}} \\ \boldsymbol{x}^+ = \boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} 3/4 \\ 1/4 \end{pmatrix} x_1 & \boldsymbol{x} \in \mathcal{D} = \{\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2 \ : \ 0 \leq 5x_2 \leq 2 - x - 1\} \end{cases} \tag{†}$$

some of its solution have been represented in figure 7.1. Starting for example from an initial condition $\boldsymbol{\phi}(0, 0) = (2, -1)$ we see that the system can only flow, thus solution $\boldsymbol{\phi}_a$ is drawn. Starting instead from $\boldsymbol{\phi}(0, 0) = (1, -1)$ the solution flows up until the point $(2, 0)$ where we have a *bifurcation*

of the solution (showing the non-uniqueness of the solution): we can either make a jump (solution $\boldsymbol{\phi}_b$) and then continue flowing or we can just flow (solution $\bar{\boldsymbol{\phi}}_b$).
Starting from $\boldsymbol{\phi}(0,0) = (0,-1)$ we flow up to $(1,0)$ where we jump at $(3/4, 1/4)$, a point that both in $\mathcal{C}$ and $\mathcal{D}$: we can so decide to flow or, more interestingly, entering a *jump loop* where we converge to the point $(0,0)$ as shown in solution $\boldsymbol{\phi}_c$.
Starting from $\boldsymbol{\phi}(0,0) = (-1,-1)$ we flow up to $(0,0)$ where we reach an *equilibrium* as the jump doesn't allowed to further move away from that point (solution $\boldsymbol{\phi}_d$). One last example is starting by $\boldsymbol{\phi}(0,0) = (-2,-1)$ (solution $\boldsymbol{\phi}_e$) where the system with a *loop* of flows and jumps converges to the point $(0,0)$.

We call **complete solutions** the ones that are evolving forever; mathematically it means that the time $t + j \to \infty$.

7.2 Stability of hybrid solutions

We say that the **origin** of an hybrid system (H) is **Lyapunov stable** if for all $\varepsilon > 0$ there exists a $\delta > 0$ (function of ε) such that

$$|\boldsymbol{\phi}(0,0)| \leq \delta \quad \Rightarrow \quad |\boldsymbol{\phi}(t,j)| \leq \varepsilon \quad \forall (t,j) \in \text{dom}\{\boldsymbol{\phi}\}$$

We say instead that the origin is **Lagrange stable** if for any $\delta > 0$ we can define a $\varepsilon > 0$ (function of δ) such that

$$|\boldsymbol{\phi}(0,0)| \leq \delta \quad \Rightarrow \quad |\boldsymbol{\phi}(t,j)| \leq \varepsilon \quad \forall (t,j) \in \text{dom}\{\boldsymbol{\phi}\}$$

As we can see the requirement for both Lyapunov and Lagrange stability is the same, but what changes is *what we fix* and *what we have to determine*: in the first case for any bound of the solution we search for a bound on the initial condition that makes the equality correct (and in general we can chose δ *very small*), while the second statement implies a more stronger condition. Given in fact any offset from the origin we have to require that the solution *doesn't diverge that much* (and stays bounded by a value ε).

The origin is also said **globally attractive** **GA** if all solutions of the system are such that

$$\lim_{t+j\to\infty} |\boldsymbol{\phi}(t,j)| = 0$$

If the origin is both Lyapunov stable and globally attractive, then it is also said **globally asymptotic stable** **GAS**.

A solution is said **uniform globally stable** **UGS** if, for any initial condition, there exists an infinitely continuous function $\alpha \in C^\infty$ bounding the solutions, so for which

$$|\boldsymbol{\phi}(t,j) \leq \alpha\big(|\boldsymbol{\phi}(0,0)|\big) \qquad \forall (t,j) \in \text{dom}\{\boldsymbol{\phi}\}$$

In practise this condition is met when the hybrid system is both Lyapunov and Lagrange stable.

Hybrid system conditions For the development of the Lyapunov theory, mostly of the time we will assume the **hybrid basic conditions** **HBC** of (H) requiring that

$$\mathcal{C}, \mathcal{D} \text{ are closed sets} \qquad \boldsymbol{f}, \boldsymbol{g} \text{ are continuous functions} \qquad \text{(HbC)}$$

If this conditions are met, by proving that the system is GAS it will automatically tells us that is also **uniformly global asymptotic stable** **UGAS**.

7.3 Lyapunov theory

The scope of the Lyapunov based theory for hybrid systems is to determine conditions that allow a system to be stable.

Theorem 7.1: The origin is globally asymptotic stable GAS (and if the hybrid basic conditions are met, then it's also UGAS) if and only if:

- it exists a continuously differentiable **Lyapunov function** $V : \boldsymbol{x} \mapsto V(\boldsymbol{x}) \in \mathbb{R}$ such that

$$\begin{cases} V(0) = 0 & \\ V(\boldsymbol{x}) > 0 & \forall \boldsymbol{x} \in \mathcal{C} \cup \mathcal{D} \setminus \{0\} \\ \lim_{|\boldsymbol{x}| \to \infty} V(\boldsymbol{x}) = \infty & \boldsymbol{x} \in \mathcal{C} \cup \mathcal{D} \end{cases} \tag{S}$$

  implying so theta V is positive definite and is radially bounded;

- it holds the **flow inequality**

$$\dot{V}(x) = \frac{\partial V}{\partial \boldsymbol{x}} \dot{\boldsymbol{x}} = \frac{\partial V}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}) < 0 \qquad \forall \boldsymbol{x} \in \mathcal{C} \setminus \{0\} \tag{F'}$$

- if holds the **jump inequality**

$$\Delta V(\boldsymbol{x}) = V\big(\boldsymbol{g}(\boldsymbol{x})\big) - V(\boldsymbol{x}) < 0 \qquad \forall \boldsymbol{x} \in \mathcal{D} \setminus \{0\} \tag{J'}$$

The underlying idea behind this theory is that when the system is flowing (F), it's flowing to zero, while when it's jumping (J), it's jumping to zero. In practise the hardest thing to do is determine the Lyapunov function $V(\boldsymbol{x})$ for the given hybrid system.

Recalling the example of the bouncing ball, a candidate Lyapunov function is the one defines as

$$V(\boldsymbol{x}) = \frac{1}{2} v^2 + \gamma p$$

Observing that thus function is positive definite for all $\boldsymbol{x} \in \mathcal{C} \cap \mathcal{D} = \{(p, v) \in \mathbb{R}^2 \ : \ p \geq 0\}$, then we can compute

$$\begin{aligned} \dot{V} &= \frac{1}{2} 2 v \dot{v} + \rho \dot{p} = v(-\gamma) + \gamma v = 0 \\ \Delta V &= \frac{1}{2} \big(v^+\big)^2 + \gamma p^+ - \frac{1}{2} v^2 - \gamma p = \frac{1}{2} (\lambda v)^2 - \frac{1}{2} v^2 = -\frac{1}{2} \big(1 - \lambda^2\big) v^2 \end{aligned}$$

As long as $\lambda < 1$ (physically, no energy is gained while bouncing but it's rather being lost), then we have that $\Delta V < 0$, proving that the system is uniformly globally asymptotic stable.