Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

INGEGNERIA BIOMEDICA, ELETTRICA E DEI SISTEMI

Ciclo 34

Settore Concorsuale: 09/G1 - AUTOMATICA

Settore Scientifico Disciplinare: ING-INF/04 - AUTOMATICA

ABOUT STABILIZATION OF NON-MINIMUM PHASE SYSTEMS BY OUTPUT FEEDBACK

Presentata da: Mario Spirito

Coordinatore Dottorato

Michele Monaci

Supervisore

Lorenzo Marconi

Esame finale anno 2022

Abstract

This thesis work has been motivated by an internal benchmark dealing with the output regulation problem of a nonlinear non-minimum phase system in case of full state feedback. The system under consideration structurally suffers from finite escape time, and this condition makes the output regulation problem very hard even for very simple steady-state evolution or exosystem dynamics, such as a simple integrator.
This situation leads to the study of the approaches developed for controlling non-minimum phase systems and how they affect feedback performances. Despite a lot of frequency domain results, only a few works have been proposed for describing the performance limitations in a state space system representation. In particular, in our opinion, the most relevant research thread exploits the so-called Inner-Outer Decomposition. Such decomposition allows splitting the non-minimum phase system under consideration into a cascade of two subsystems: a minimum phase system (the outer) that contains all poles of the original system and an all-pass non-minimum phase system (the inner) that contains all the unavoidable pathologies of the unstable zero dynamics.
Such a cascade decomposition was inspiring to start working on functional observers for linear and nonlinear systems. In particular, the idea of a functional observer is to exploit only the measured signals from the system to asymptotically reconstruct a certain function of the system states, without necessarily reconstructing the whole state vector. The feature of asymptotically reconstructing a certain state functional plays an important role in the design of a feedback controller able to stabilize the non-minimum phase system.
To describe these topics we composed this thesis of mainly two parts.

Part I: non-minimum phase systems

In this first part of the thesis, we provide the reader with a general literature overview of performance limitations concerning non-minimum phase dynamics. In particular, we described the limitations in the design of the feedback controller in achieving desired closed-loop characteristics. We provide the reader with a set of proposed approaches to relax some of these limitations: a state feedback approach to remove undershoot and overshoot problems in systems with unstable zero dynamics, and a funnel control approach in which the output trajectory is driven to stay between two boundary functions for all time.
In a more theoretical framework, the non-minimum phase limitations in an ideal control case can be described by a nonzero lower bound on the output signal energy. In particular, this bound refers to the amount of energy needed by the output signal to stabilize the unstable system zero dynamics. This characteristic has been proved both for linear and for nonlinear cases. By the way, more recent results have shown that by solving a path-following problem instead of an output regulation problem, such a bound of the output signal energy can then be made arbitrarily small.
Our contributions mainly concern the output undershoot limitations and the realization of the Inner-Outer decomposition for strictly proper linear time-invariant systems. In particular, we first develop a closed-form solution to the realization of the Inner-Outer decomposition and by exploiting the cascade characteristics, we were able to upper bound the output undershoot by consequently steering the non-minimum phase output arbitrary close to the output trajectory of minimum phase systems. The only drawback of the approach is that the equivalent system closed-loop bandwidth must be reduced to a low-frequency range, which is reflected in a very slow output behaviour.

Part II: Functional observers

To achieve such a slow output behaviour, the only meaning part of the Inner-Outer cascade is the output (or the state vector) of the outer dynamics. We thus considered the problem of building a functional observer up to reconstruct from the real system output one of the outer dynamics. To pursue this idea, we first consider the case of Linear autonomous systems and then that of nonlinear autonomous systems. Our contribution to the thread of functional observers consists, for the case of linear systems, in establishing a unifying framework able to gather the most relevant literature approaches. And for the

case of nonlinear systems, it consists in extending the KKL observer approach to the case of nonlinear functional observer and finding practical applications, such as input reconstruction, unknown input observers, and observers for nonlinear controlled system.
Unfortunately, the Inner-Outer decomposition results to be a detectable cascade, due to the equivalent zero/pole cancellation, and this property does not satisfy the backward distinguishability assumptions needed in the KKL observer approach. Hence, it is not possible to reconstruct a functional of the outer system states directly from the system output but in order to do so, we necessarily have to apply a change of coordinates (or a diffeomorphism) on the original system states.

Contents

Notation

$\mathbb{C}$, $\mathbb{C}_-$ and $\mathbb{C}_+$ stand respectively for the complex plane, the closed left half complex plane and the open right half complex plane.
We consider $\mathcal{B}_R$, for positive real $R > 0$, the closed ball around the origin of radius R.
Given a matrix $A \in \mathbb{R}^{n\times n}$, $\sigma(A) \subset \mathbb{C}^n$ denotes the spectrum of A, while $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ are the minimum and the maximum values of $\sigma(A)$, respectively.

We associate with each tuple (A, B, C, D), with $A \in \mathbb{R}^{n_x\times n_x}$, $B \in \mathbb{R}^{n_x\times n_u}$, $C \in \mathbb{R}^{n_y\times n_x}$ and $D \in \mathbb{R}^{n_y\times n_u}$, the linear equations

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{1}$$

and we refer to (1) as "system (A, B, C, D)", or "system (A, B, C)" when $D = 0$. In this latter case, the system is said to be strictly proper. Moreover, the transfer matrix associated to system (1) is given by

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

For a system (A, B, C), define the vector relative degree $\bar{r} = (r_1, \dots, r_p)$ and $r = \sum_{i=1}^{n_y} r_i$, where each r_i, for $i = 1, \dots, n_y$, is defined as the smallest integer such that $C_iA^{k-1}B = 0$, for $k = 1, \dots, r_i - 1$, and $C_iA^{r-1}B \neq 0$, where C_i is the i-th row of C. There always exists, see Mueller (2009), a nonsingular matrix $T_{\text{nf}} \in \mathbb{R}^{n_x\times n_x}$, such that the system $(T_{\text{nf}}\, A\, T_{\text{nf}}^{-1},\, T_{\text{nf}}B,\, C\, T_{\text{nf}}^{-1})$ is the normal form realisation of system (A, B, C), namely

$$\begin{aligned} T_{\text{nf}}\, A\, T_{\text{nf}}^{-1} = \begin{bmatrix} F & G \\ H & \bar{A} \end{bmatrix} \quad T_{\text{nf}}B = \begin{bmatrix} 0 \\ \bar{B} \end{bmatrix} \\ C\, T_{\text{nf}}^{-1} = \begin{bmatrix} 0 & \bar{C} \end{bmatrix}, \end{aligned} \tag{2}$$

where $F \in \mathbb{R}^{(n_x-r)\times(n_x-r)}$, $\bar{A} \in \mathbb{R}^{r\times r}$, and for some matrices $G, H, \bar{B}, \bar{C}$ of suitable dimensions. In particular, for r_i being relative degree of the i-th output, the matrices $\bar{A}_i \in \mathbb{R}^{r_i\times r_i}$, $i = 1, \dots, n_y$, are in standard companion form with last row filled with the coefficients of the relative characteristic polynomial, and we have

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & \star & \cdots & \star \\ \star & \bar{A}_2 & \cdots & \star \\ & & \ddots & \\ \star & \star & & \star \\ \star & \star & \cdots & \bar{A}_p \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \\ \vdots \\ \bar{B}_p \end{bmatrix}, \quad \bar{H} = \begin{bmatrix} \bar{H}_1 \\ \bar{H}_2 \\ \vdots \\ \bar{H}_p \end{bmatrix}$$

where the terms $\star$, $\bar{B}_i$, and $\bar{H}_i$, for $i = 1, \dots, n_y$, are matrices of appropriate dimensions with all zero terms expect their last row; while $\bar{C}$ and G have the form

$$\begin{aligned} \bar{C} &= \begin{bmatrix} \bar{C}_1 & \bar{C}_2 & \dots \bar{C}_p \end{bmatrix} \\ G &= \bar{G}\bar{C} \end{aligned}$$

for some $\bar{G}$, where each $\bar{C}_i$ has all zero terms expect its i-th row which is defined as $[1, 0, \dots, 0] \in \mathbb{R}^{r_i}$, for $i = 1, \dots, n_y$.
Given $\bar{x} > 0$ and $x \in \mathbb{R}$, we define the function

$$\text{sat}_{\bar{x}}(x) = \begin{cases} x, & \text{if } |x| \leq \bar{x} \\ \text{sign}(x)\bar{x}, & \text{otherwise.} \end{cases} \tag{3}$$

Moreover, for $x = (x_1, \dots, x_{n_x}) \in \mathbb{R}^{n_x}$, we let $\text{sat}_{\bar{x}}(x) := (\text{sat}_{\bar{x}}(x_1), \dots, \text{sat}_{\bar{x}}(x_n))$.
The transmission zeros of system (A, B, C), as defined in Rosenbrock (1973), are the values $\bar{\lambda} \in \mathbb{C}$ such that

$$\text{rank} \begin{bmatrix} A - \bar{\lambda}I & B \\ C & 0 \end{bmatrix} < n + \min(n_y, n_u). \tag{4}$$

The set of such values $\bar{\lambda}$ is equal to the spectrum of the matrix F when the system (A, B, C) is in normal form realization (2), see Isidori (2017)[Ch.2]. Equivalently for a proper transfer matrix $G(s)$, its transmission zeros are the values $\bar{s} \in \mathbb{C}$ at which $G(\bar{s})$ loses rank. The transmission zeros of system (A, B, C, D), are the values $\bar{\lambda} \in \mathbb{C}$ such that

$$\text{rank} \begin{bmatrix} A - \bar{\lambda} I & B \\ C & D \end{bmatrix} < n_x + \min(n_y, n_u). \tag{5}$$

For a proper system (A, B, C, D) with D nonsingular, the transmission zeros are the eigenvalues of $A - BD^{-1}C$, i.e. the elements of $\sigma(A - BD^{-1}C)$. A system (A, B, C) (and equivalently $G(s)$) is said to be minimum phase if it has all zeros in $\mathbb{C}_-$. While it is said to be non-minimum phase if at least one of its zeros is in $\mathbb{C}_+$. A system (A, B, C) is said to be right invertible if (see Qiu and Davison (1993))

$$\text{rank} \begin{bmatrix} A - \lambda I & B \\ C & 0 \end{bmatrix} = n_x + n_y, \tag{6}$$

for at least one $\lambda \in \mathbb{C}_+$.
For a C^1 map $h : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ and a vector field $f : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$, $L_f h(x) := \frac{dh}{dx}(x) f(x)$ denotes the Lie derivative of h along the vector field f, and iteratively, $L_f^{(k)} h$ denotes the kth Lie derivative of h along f.

For a system with locally Lipschitz map f, we denote for each x in $\mathbb{R}^{n_x}$, $t \mapsto X(x, t)$ to be the unique solution with initial condition x at $t = 0$. Similarly, for a system (4.22) with inputs and locally Lipschitz map f with respect to x, we denote by $X_u(x, s, t)$ the solution subject to the input u taken at time t and initialized in x at time s, so that $X_u(x, t, t) = x$. Given an open set $\mathfrak{O}$ of $\mathbb{R}^{n_x}$ and for each x in $\mathfrak{O}$, we denote by $(\sigma_{\mathfrak{O}}^-(x), \sigma_{\mathfrak{O}}^+(x))$, or $(\sigma_{\mathfrak{O}}^-(x, s; u), \sigma_{\mathfrak{O}}^+(x, s; u))$, the maximal interval of definition of the solution $X(x, t)$, or $X_u(x, s, t)$, respectively, conditioned to take values in $\mathfrak{O}$. For a set S, we denote by $\text{cl}(S)$ its closure and with $S + \delta$ the set

$$S + \delta = \{x \in \mathbb{R}^{n_x} : \exists \chi \in S : |x - \chi| \leq \delta\} .$$

For the Kronecker product, we consider the symbol $\otimes$. Given a matrix P, with spectrum $\lambda(P)$, we indicate with $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$ the minimum and maximum eigenvalues of P, respectively.
A continuous function $\alpha : [0, +\infty) \to [0, +\infty)$ is a class-$\mathcal{K}$ map if it is increasing and $\alpha(0) = 0$.
Finally, we denote $\mathbb{W}$ a class of functions $w : D \subseteq \mathbb{R} \to \mathbb{R}^{n_w}$ and by $\mathcal{W}$ a subset of $\mathbb{R}^{n_w}$ containing their image sets $w(D)$.

Part I

Non-minimum Phase Systems

Chapter 1

Introduction: Performance limitations of Non-minimum Phase Systems

In this chapter, we describe the framework that motivated the analysis regarding the whole thesis work. In particular, we introduce what we informally called the *Astolfi Benchmark*, i.e., an output regulation problem involving a critical non-minimum phase system driven by the error signal to be regulated with the help of additional non-vanishing measurement of the system zero dynamics. We call it critical because the zero dynamics of such a system suffer from the finite escape time pathology. In the following, we better describe in more detail such a problem and then analyse the state of the art about performance limitations due to the presence of a non-minimum phase system (unstable zeros in the plant).

1.1 Motivating Benchmark

We started this Thesis work with the objective of solving the output regulation problem associated with the *Astolfi Benchmark*.
The plant under consideration is

$$\begin{aligned} \dot{z} &= z^3 + e + r(w) \\ \dot{e} &= u \\ y &= \begin{pmatrix} z \\ e \end{pmatrix} \end{aligned} \tag{1.1}$$

where z, e, u are scalar and the available measurement y is constituted of the whole state. Such a system is perturbed by an exogenous system via the term $r(w)$ where the dynamics of w is given by

$$\dot{w} = s(w) \tag{1.2}$$

Assume that there exists an invariant manifold such that $z = \Pi_0(w)$ and $e = 0$ where $\Pi(w)$ is solution of the nonlinear regulator equation Isidori and Byrnes (1990)

$$\frac{\partial \Pi_0(w)}{\partial w} s(w) = \Pi_0(w)^3 + r(w) \tag{1.3}$$

and, moreover, that Π_0 evolution in time can be described in regression form, i.e., there exists an integer number v and a function $\phi : \mathbb{R}^v \to \mathbb{R}$ such that

$$\Pi_0^{(v+1)}(w) = \phi(\Pi_0(w), \dot{\Pi}_0(w), \ldots, \Pi^{(v)}(w)). \tag{1.4}$$

Under this assumption, the output regulation problem associated with the *Astolfi Benchmark* is to find a (dynamical) control law u only exploiting the the output measurement y (partial-information case), such that system (1.1) perturbed by the exosystem (1.2) is steered on the invariant manifold $\{(z, e) \in \mathbb{R}^2 | (z, e) = (\Pi_0(w), 0)\}$ while u is kept bounded and asymptotically vanishes.
In the following, we describe general feedback limitations due to the presence of non-minimum phase zeros in the controlled plant.

1.2 Performance limitations in non-minimum phase: frequency domain

Given a plant $G(s)$ and a regulator $R(s)$ transfer functions, where both $G(s)$ and $R(s)$ are single input single output systems, we call the open loop transfer function $L(s) = R(s)G(s)$. In feedback design, very important roles are played by the sensitivity functions $S(s)$ and $T(s)$ defined as

$$S(s) = \frac{1}{1 + L(s)} = \frac{D(s)}{D(s) + N(s)}$$

$$T(s) = \frac{L(s)}{1 + L(s)} = S(s)L(s) = \frac{N(s)}{D(s) + N(s)}$$

where $D(s)$ and $N(s)$ are the denominator and the numerator of $L(s)$. To distinguish among the two sensitivity functions $S(s)$ and $T(s)$, the latter is called the *complementary* sensitivity function while $S(s)$ is simply the sensitivity function. The two functions are indeed complementary in the sense that for all s in $\mathbb{C}$ it holds

$$S(s) + T(s) = 1. \tag{1.5}$$

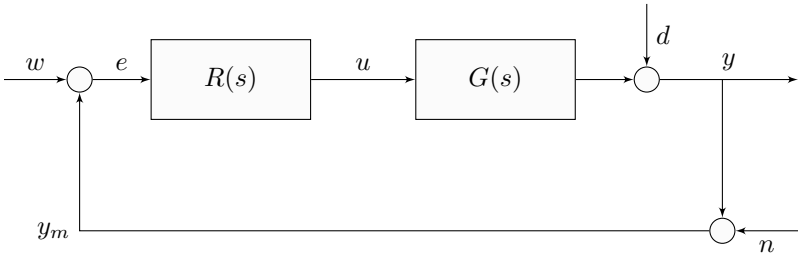

Figure 1.1: Standard control structure

According to the feedback structure reported in Fig.1.1, the output function $y(t)$ can be decomposed into three different components, i.e., y_d, y_w, and y_n, related to the three systems inputs, i.e., w, d and n

$$y(t) = y_d(t) + y_w(t) + y_n(t) = d(t)S(s) + w(t)T(s) - n(t)T(s)$$

with some abuse of notation in implicitly considering the Laplace transform of the involved signals. We can see from the output terms that the two sensitivity functions play an important role in feedback design. In particular, the sensitivity function describes the link between the output disturbance d and the measured plant output y, while the complementary function describes the whole closed loop system from the reference w to the plant output y and it is moreover the link between the measurement noise n and the output itself. We generally consider, d and n to live in the low and high-frequency ranges, respectively. Indeed, the magnitude of the sensitivity function $S(s)$ is a direct indication of the ability of the related feedback loop to perform satisfactorily despite the presence of disturbances, d, at the plant output and small plant parameter variations. Equivalently, for the complementary function $T(s)$ and the noise signal $n(t)$. Moreover, we say that the feedback system possesses sensitivity reduction at some frequency ω if $|S(j\omega)| < 1$. While, the feedback system exhibits increased sensitivity at some frequency ω if $|S(j\omega)| > 1$. In this case, the feedback increases the effect of plant parameter variations and disturbances on the output plant.
The feedback system is considered stable if $S(s)$ has no poles in the right half plane, and there is no unstable zero/pole cancellation in the open loop function, $L(s)$, see Bode et al. (1945); Freudenberg and Looze (1983, 1985).

1.2.1 Bode Integrals

Considering a stable feedback system and assume that $L(s)$ has relative degree strictly larger than 1 (so that $\lim_{s\to\infty} sL(s) = 0$) and has finitely many poles and zeros in the right half plane, and denote them as $\mathtt{P} = \{\mathrm{p}_i; i = 1, \ldots, N_\mathrm{p}\}$ and $\mathtt{Z} = \{\mathrm{z}_i; i = 1, \ldots, N_\mathrm{z}\}$, including multiplicities. When $L(s)$ has no unstable poles the first Bode integral reads as Bode et al. (1945)

$$\int_0^\infty \ln\left(|S(j\omega)|\right) d\omega = 0, \tag{1.6}$$

that is, the natural logarithm of the sensitivity function magnitude integrated over the non-negative frequencies must be zero. Hence, the negative area (associated with the feedback sensitivity reduction $|S(s)| < 1$) resulting in a certain frequency range must be compensated by a positive area (associated with the feedback sensitivity increase $|S(s)| > 1$) in the complementary frequency range.
When the open-loop function $L(s)$ has unstable poles, the constraint worsens and the Bode integral is given by Freudenberg and Looze (1983, 1985)

$$\int_0^\infty \ln\left(|S(j\omega)|\right) d\omega = \pi \sum_{i=1}^{N_p} \mathfrak{Re}\{\mathrm{p}_i\} > 0. \tag{1.7}$$

In this case, the trade-off between the sensitivity reduction and increase in some frequency range and its complementary is unbalanced towards the sensitivity increase.
Note that the two Bode integrals apply even if the system is non-minimum phase.
Exploiting the Blaschke product we can write, as in Freudenberg and Looze (1985), the open loop function

$$L(s) = L'(s)\mathfrak{B}_p^{-1}(s)\mathfrak{B}_z(s)$$

where, by denoting with the * symbol the complex conjugation, the Blaschke products read as

$$\mathfrak{B}_p(s) = \prod_{i=1}^{N_p} \frac{\mathrm{p}_i - s}{\mathrm{p}_i^\star + s},$$

$$\mathfrak{B}_z(s) = \prod_{i=1}^{N_z} \frac{z_i - s}{\mathrm{z}_i^\star + s}.$$

Hence, the $L'(s)$ term has no poles nor zeros in the right half plane. The same can be done with the two sensitivity functions

$$S(s) = S'(s)\mathfrak{B}_p(s) \tag{1.8}$$

$$T(s) = T'(s)\mathfrak{B}_z(s) \tag{1.9}$$

Defining a weighting function for a complex number s

$$\theta_s(\omega) = \arctan\left[\frac{\omega - y}{x}\right], \quad s = x + jy \tag{1.10}$$

the following theorems describe an extension of the Bode integral constraints

Theorem 1.2.1 (Th.1 in Freudenberg and Looze (1985))**.** *Let* $\mathrm{z} = x + jy$ *be an element of* $\mathtt{Z}$*, if the closed-loop system is stable the sensitivity function* $S(s)$ *must satisfy the integral constraints*

$$\int_{-\infty}^{+\infty} \ln\left(|S(j\omega)|\right) d\theta_z(\omega) = \pi \ln\left(|\mathfrak{B}_p^{-1}(\mathrm{z})|\right)$$
$$\int_{-\infty}^{+\infty} \varphi\left(S'(j\omega)\right) d\theta_z(\omega) = \pi\varphi\left(\mathfrak{B}_p^{-1}(\mathrm{z})\right).$$

Theorem 1.2.2 (Th.2 in Freudenberg and Looze (1985))**.** *Let* $\mathrm{p} = x + jy$ *be an element of* $\mathtt{P}$*, if the closed-loop system is stable the sensitivity function* $S(s)$ *must satisfy the integral constraints*

$$\int_{-\infty}^{+\infty} \ln\left(|T(j\omega)|\right) d\theta_p(\omega) = \pi \ln\left(|\mathfrak{B}_z^{-1}(\mathrm{p})|\right)$$
$$\int_{-\infty}^{+\infty} \varphi\left(T'(j\omega)\right) d\theta_p(\omega) = \pi\varphi\left(\mathfrak{B}_z^{-1}(\mathrm{p})\right).$$

In both theorems, φ stands for the phase value of its argument.
This result can be specialized Freudenberg and Looze (1985) to the case of non-minimum phase system in which we ask some sensitivity reduction in a certain frequency range $\Omega = [-\omega_2, -\omega_1] \cup [\omega_1, \omega_2]$ (i.e., Ω is a conjugate symmetric range of frequency). Let z be an open right half plane zero of $L(s)$ and let the weighted length of the frequency range Ω be denoted

$$\Theta_z\{\Omega\} = \int_\Omega d\theta_z(\omega)$$

and the weighted length of the complementary frequency range $\Omega^c = \{\omega | \omega \notin \Omega\}$ is given by

$$\Theta_z\{\Omega^c\} = \pi - \Theta_z\{\Omega\}.$$

Let the desired level of sensitivity reduction be given by

$$|S(jw)| \leq \bar{S}_\Omega < 1, \quad \forall \omega \in \Omega$$

with $\bar{S}_\Omega$ positive scalar value. The following theorem gives a lower bound on the maximum sensitivity in the complementary frequency range Ω^c due to the achievement of such a sensitivity reduction level for a non-minimum phase system.

Theorem 1.2.3. *Suppose that the closed-loop system is stable and that the level of sensitivity reduction*

$$|S(jw)| \leq \bar{S}_\Omega < 1, \quad \forall \omega \in \Omega$$

has been obtained. Then for each $z \in \mathtt{Z}$ the following bound must be satisfied

$$\|S\|_\infty \geq \left(\frac{1}{\bar{S}_\Omega}\right)^{\frac{\Theta_z(\Omega)}{\pi - \Theta_z(\Omega)}} \cdot \left|\mathfrak{B}_p^{-1}(z)\right|^{\frac{\pi}{\pi - \Theta_z(\Omega)}}$$

where $\|S\|_\infty = \sup_\omega |S(j\omega)|$.

The lower bound is necessarily greater than one (hence in the complementary frequency range the system has a sensitivity increase), indeed, $\bar{S}_\Omega < 1$, $|\mathfrak{B}_p^{-1}(z)| > 1$, and $\Theta_z(\Omega) < \pi$. To avoid this issue, one might think that a solution to this problem is to spread the sensitivity on the remaining infinite range of frequency Ω^c with an arbitrary small magnitude value. This is not usually desirable due to stability robustness. Indeed, typically uncertainties in the plant model are stronger at higher frequencies. This is why it is required to have $L(s)$ approaching 0 at high frequency, where larger relative uncertainty levels are present. Moreover, it is required to have a small loop magnitude at frequencies for which the sensor noise, $n(t)$, dominates the contributions of the plant disturbances, i.e., output disturbances and references, Freudenberg and Looze (1983). Furthermore, note that if the system is open-loop unstable and non-minimum phase with approximate right half plane pole-zero cancellations, it might have very bad sensitivity properties. Indeed, the lower bound for $\|S\|_\infty$ increases as a function of the proximity of unstable poles to the zero in question,Freudenberg and Looze (1983).

A similar result, Freudenberg and Looze (1985), can be found for the complementary sensitivity function in presence of unstable open loop poles. And lower bounds on the maximum value of $|T(j\omega)|$ can be similarly derived. In this case, since the specifications on sensor noise responds are generally imposed at high frequencies, the set $\Omega = [-\omega_2, -\omega_1] \cup [\omega_1, \omega_2]$ usually has $\omega_2 \to \infty$.

In Goodwin et al. (2001) another integral constraint has been consider relating the complementary sensitivity function $T(s)$ and the unstable zeros location $z \in \mathtt{Z}$, that is

$$\int_0^\infty \ln\left(|T(j\omega)|\right) \frac{1}{\omega^2} d\omega = \pi \sum_{i=1}^{N_z} \frac{1}{\mathrm{z}_i} - \frac{\pi}{2K_v} \tag{1.11}$$

with $K_v = \lim_{s\to 0} sL(s)$. Note, that if $L(s)$ has relative degree strictly larger then 1, or if perfect tracking/output disturbance rejection is obtain, that is $e(t) \to 0$, then $K_v = \lim_{s\to 0} sL(s) \to \infty$.
Then, a revisited version has been recently developed by Emami-Naeini and de Roover (2019). According to the author, these integral limitations/constraints forms are hiding a fundamental result, i.e., that the '*sensitivity integral constraint is related to the difference in speed (bandwidth or poles location) of the closed-loop system compared to the speed (bandwidth or poles location) of the open loop system*'. We then have the following results

Theorem 1.2.4 (Theorem 1 in Emami-Naeini and de Roover (2019))**.** *For any SISO closed-loop stable and proper rational LTI system, the Bode's integral constraint can be written as*

$$\int_0^\infty \ln\left(|S(j\omega)|\right) d\omega = \frac{\pi}{2} \sum_{i=1}^{n} \left(\mathrm{p}_{c\ell i} - \mathrm{p}_{o\ell i}\right) + \pi \sum_{i=1}^{N_p} \mathrm{p}_i \tag{1.12}$$

where $\{\mathrm{p}_{c\ell_i}\}$ and $\{\mathrm{p}_{o\ell_i}\}, i = 1, \ldots, n$ are the locations of the closed-loop and open-loop poles, while $\mathrm{p}_i \in \mathtt{P}, i = 1, \ldots, N_p$ are the open-loop unstable poles.

The fundamental relationship is that the sum of the areas underneath the $\ln(|S(s)|)$ curve is related to the difference in speeds of the closed-loop and open-loop systems. This result extends the one in Goodwin et al. (2001) that shows that if $L(s)$ is strictly proper then

$$\int_0^\infty \ln\left(|S(j\omega)|\right) d\omega = \pi \sum_{i=1}^{N_p} \mathrm{p}_i - \frac{\pi K_h}{2}$$

with $K_h = \lim_{s\to\infty} sL(s)$. And it can be shown that

$$K_h = -\sum_{i=1}^{n} \left(\mathrm{p}_{c\ell i} - \mathrm{p}_{o\ell i}\right).$$

Note that, because the imaginary part of the complex conjugate poles cancel each other, this is the same result obtained in Freudenberg and Looze (1985) for which $K_h = 0$ because the system under consideration has at least relative degree 2.
For the complementary sensitivity function $T(s)$ in Emami-Naeini and de Roover (2019) has been shown the following theorem.

Theorem 1.2.5 (Theorem 2 in Emami-Naeini and de Roover (2019))**.** *For any SISO closed-loop stable proper rational LTI system the complementary sensitivity integral constraint may be written as*

$$\int_0^\infty \ln\left(|T(j\omega)|\right) \frac{1}{\omega^2} d\omega = \pi \sum_{i=1}^{N_z} \frac{1}{\mathrm{z}_i} + \frac{\pi}{2}\left(\sum_{i=1}^{n} \frac{1}{\mathrm{p}_{c\ell i}} - \sum_{i=1}^{N_{zo\ell}} \frac{1}{\mathrm{z}_{o\ell i}}\right)$$

where $\{\mathrm{p}_{c\ell_i}\}, i = 1, \ldots, n$ *are the locations of the closed-loop poles,* $\{\mathrm{z}_{o\ell_i}\}, i = 1, \ldots, N_{zo\ell}$ *are the locations of the open-loop zeros, while* $\mathrm{z}_i \in \mathtt{Z}, i = 1, \ldots, n_z$ *are the open-loop unstable zeros.*

By exploiting the Truxal's identity we have, Emami-Naeini and de Roover (2019),

$$\sum_{i=1}^{n} \frac{1}{\mathrm{p}_{c\ell i}} - \sum_{i=1}^{N_{zo\ell}} \frac{1}{\mathrm{z}_{o\ell i}} = -\frac{1}{K_v}$$

as defined above and thus we recover Goodwin's version of the Bode's integral (1.11).
With these versions of Bode's integral, one can directly notice the relationship between the location of poles and zero in the open and closed loop system, and the sensitivity constraints in the frequency domains.

1.2.2 Output Performances in case of non-minimum phase zeros

In the following, we consider some frequency domain analysis of the output characteristics due to the presence of non-minimum phase zeros.

The first result is from Vidyasagar (1986). The author first defines an undershooting system in the locality of the initial time $t = 0$. In particular, for a strictly proper LTI SISO system with relative degree r and transfer function $G(s)$, it's well known that the unitary step response of $G(s)$, $y(t)$, at time $t = 0$, is 0 along with the its first $r - 1$ derivatives are 0. By considering, that the step response exhibits 'undershoot' if its steady-state value has a sign opposite from that of its first non-zero derivative at time $t = 0$. Thus, we define a system to have undershoot if $y^{(r)}(O)y(\infty) < 0$. Clearly, this definition only makes sense if $y(\infty) \neq 0$. This is a natural mathematical version of "the step response initially starts in the wrong direction". Then they provide the following result anywhere.

Proposition 1 (Proposition in Vidyasagar (1986))**.** *The system has initial undershoot if and only if its transfer function has an odd number of real RHP zeros.*

We modified the original proposition by adding the adjective *initial* to the word undershoot because by considering a different definition of output undershoot, such as the one in Lau et al. (2003) or Stewart and Davison (2006), the original proposition does not hold anymore. Indeed, by considering the undershoot y_{us} of the unitary step response, denoted $y(t)$, as the smallest non-negative number such that $y(t) \geq -y_{us}$, then the system output may still have points of zero crossing, see, e.g. Fig.5 in Hoagg and Bernstein (2007), or simply the output $y(t)$ may reverse its direction during its time evolution (hence, in some time interval the output derivate becomes negative before crossing the steady state value). This is why we say it is only a local (in time) result that works accordingly with the provided undershoot definition. As shown

in Hoagg and Bernstein (2007), there are two different possible output behaviour for a non-minimum phase strictly proper LTI SISO system with zero initial condition, i.e., monotonic and non-monotonic step responses. In the latter class, we can distinguish between zero crossing and non-zero crossing step response. In Table 1, in Hoagg and Bernstein (2007), they summarize the possible output characteristics (initial undershoot, zero crossing, and overshoot) relations according to zero dynamics of the plant. In particular, for strictly proper system an only sufficient condition to have zero crossing is that the system has at least one positive real zero. To have an initial undershoot the necessary and sufficient condition is that the plant has an odd number of positive real zeros. While to have a step response overshoot it is sufficient that $G(s) - G(0)$ has at least one positive real zero.
In Stewart and Davison (2006), the authors report a result from Middleton (1991),i.e., the lower bound for the output undershoot (this time the absolute value of the inf of the system step response with zero initial condition) of a strictly proper LTI SISO system with a real positive zero at $s = \mathrm{z}$

$$y_{us} \geq \frac{0.99}{e^{\mathrm{z}T_{s1\%}} - 1} y(\infty) > 0 \tag{1.13}$$

where $y(\infty) = \lim_{t\to\infty} y(t)$ and $T_{s1\%}$ is the 1% settling time of $y(t)$, i.e., the smallest time $T_{s1\%}$ that $|y(t) - y(\infty)| \leq 0.01 y(\infty)$ for all $t \geq T_{s1\%}$. Note that, for $\mathrm{z} \to 0$ (the real positive zero is close to the imaginary axis) or for $T_{s1\%} \to 0$ (thus, a very fast output response), the lower bound tends to infinity. Moreover, they show that with two distinct real positive zeros, the output response necessarily exhibits overshoot. The same two results have been similarly investigated in Lau et al. (2003)[Section B].

1.3 Performance limits in non-minimum phase: Linear Time Invariant systems (state space)

In Qiu and Davison (1993) a cheap control problem set-up has been used to describe the constraints in terms of output y energy when the system is non-minimum phase. In particular, for a linear LTI (exactly) proper non-minimum phase system in state space realization

$$\begin{aligned} \dot{x} &= Ax + Bu, \quad x(0) = x_0 \neq 0 \\ y &= Cx + Du \end{aligned} \tag{1.14}$$

the cheap control problem is to find a control action u such that the cost functional

$$J_\epsilon = \int_0^\infty y^T(t) y(t) + \epsilon u^T(t) u(t) dt \tag{1.15}$$

is minimized with $\epsilon \to 0$. In particular, the final results exploit the following properties.

- A proper transfer matrix $G(s)$ can always be factorized as $G_{\mathrm{i}}(s)G_{\mathrm{o}}(s)$ such that $G_{\mathrm{i}}(s)$ is a *inner* system and $G_{\mathrm{o}}(s)$ is a right-invertible and minimum phase system (thus, *outer*) system. Where, a stable exactly proper transfer matrix $G_{\mathrm{i}}(s)$ is called inner if $G_{\mathrm{i}}^T(-s)G_{\mathrm{i}}(s) = I$ and all the zeros of $G_{\mathrm{i}}(s)$ are in the open right half plane. Letting $(A_{\mathrm{o}}, B_{\mathrm{o}}, C_{\mathrm{o}}, D_{\mathrm{o}})$ and $(A_{\mathrm{i}}, B_{\mathrm{i}}, C_{\mathrm{i}}, D_{\mathrm{i}})$ be minimal stabilizable and detectable realization of $G_{\mathrm{o}}(s)$ and $G_{\mathrm{i}}(s)$ factors, then a stabilizable and detectable realization of $G(s)$ is given by the cascade realization

$$\begin{aligned} A &= \begin{bmatrix} A_{\mathrm{i}} & B_{\mathrm{i}}C_{\mathrm{o}} \\ 0 & A_{\mathrm{o}} \end{bmatrix}, \quad B = \begin{bmatrix} B_{\mathrm{i}}D_{\mathrm{o}} \\ B_{\mathrm{o}} \end{bmatrix} \\ C &= \begin{bmatrix} C_{\mathrm{i}} & D_{\mathrm{i}}C_{\mathrm{o}} \end{bmatrix}, \quad D = D_{\mathrm{i}}D_{\mathrm{o}}. \end{aligned} \tag{1.16}$$

- Associated with any realization (A, B, C, D) in which A is stable, we call controllability and observability grammian P_c and P_o, respectively, the solution of the following Lyapunov equations

$$\begin{aligned} AP_c + P_cA^T &= -BB^T \\ P_oA + A^TP_o &= -C^TC \end{aligned}$$

  and a minimal realization (A, B, C, D) is called *balanced realization* if such P_c and P_o are diagonal and equal, see Glover (1984). Without loss of generality we can assume that $(A_{\mathrm{o}}, B_{\mathrm{o}}, C_{\mathrm{o}}, D_{\mathrm{o}})$ and $(A_{\mathrm{i}}, B_{\mathrm{i}}, C_{\mathrm{i}}, D_{\mathrm{i}})$ exploited in (1.16) are balanced realization.

- Moreover, a balanced realization of a inner system $(A_{\mathrm{i}}, B_{\mathrm{i}}, C_{\mathrm{i}}, D_{\mathrm{i}})$ has $P_c = P_o = I$, $D_{\mathrm{i}}^T D_{\mathrm{i}} = I$ and $D^TC + B^T = 0$ (or equivalently, $DB^T + C = 0$).

As also shown afterwards in this thesis 2.2.1, in the inner-outer cascade realization, the inner factor has all the unstable zeros of $G(s)$ and all the poles of $G_{\mathrm{i}}(s)$ are equal to the negatives of zeros of $G_{\mathrm{i}}(s)$,i.e. poles and zeros of $G_{\mathrm{i}}(s)$ are mirrored with respect to the imaginary axis.
At this point, consider the cheap control problem described before in equations (1.14)-(1.15). It is well known, that the cost functional can be written as $J_\epsilon = x(0)^T P_\epsilon x(0)$, and that the optimal control that stabilize the system and achieves the optimal cost is given by $u = -(\epsilon I + D^T D)^{-1}(B^T P_\epsilon)$, where P_ϵ is the stabilizing solution of the following Algebraic Riccati Equation (ARE)

$$\begin{aligned}\left[A - B(\epsilon^2 I + D^T D)^{-1} D^T C\right]^T P_\epsilon + P_\epsilon \left[A - B(\epsilon^2 I + D^T D)^{-1} D^T C\right] + \\ C^T \left[I - D(\epsilon^2 I + D^T D)^{-1} D^T\right] C - P_\epsilon B(\epsilon^2 I + D^T D)^{-1} B^T P_\epsilon = 0. \quad (1.17)\end{aligned}$$

It is well-known that P_ϵ monotonically decreases as $\epsilon \to 0$, thus $P_0 = \lim_{\epsilon\to\infty} P_\epsilon$ exists. In particular, for a minimum phase and right invertible system such as $(A_{\mathrm{o}}, B_{\mathrm{o}}, C_{\mathrm{o}}, D_{\mathrm{o}})$, $P_0 = 0$. While for a inner system with balanced realization $(A_{\mathrm{i}}, B_{\mathrm{i}}, C_{\mathrm{i}}, D_{\mathrm{i}}$, $P_0 = I$ and $u = 0$ because $B_{\mathrm{i}}^T + D_{\mathrm{i}}^T C_{\mathrm{i}} = 0$ [1]. Thanks to this two properties and by considering a factorized realization of the system matrices (A, B, C, D) as in (1.16), we have that

$$P_0 = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

and for the cheap control problem described by (1.14)-(1.15), the cost functional for $\epsilon \to 0$, $J_0 = \lim_{\epsilon\to\infty} J_\epsilon = x^T(0) P_0 x(0) = x_{\mathrm{i}}(0)^T x_{\mathrm{i}}(0)$ is the energy used by the system output $y(t)$ in order to stabilize the unstable zero dynamics x_{i}.

1.3.1 The servomechanism problem

In the same work, Qiu and Davison (1993), the authors extend the result to the output regulation (servomechanism) framework assuming this time a zero initial condition of the plant, i.e. $x(0) = 0$. The analysed framework considers a stabilizable and detectable system (A, B, C, D) with noise-corrupted input and output

$$\begin{aligned}\dot{x} &= Ax + B(u + w_u) \\ e &= Cx + D(u + w_u) + w_e\end{aligned} \quad (1.18)$$

where w_u and w_e are exosystem-generated noise signals with some periodic behaviour, and the error signal e is generated by comparing the output to a reference signal $y^\star$, i.e., $e = y - y^\star$. To such a system we associated the following cost functional

$$J_\epsilon = \min_{\tilde{u}} \int_0^\infty e^T e + \epsilon^2 \tilde{u}^T \tilde{u} dt \quad (1.19)$$

where $\tilde{u}$ refers to the transient behaviour of the input u to its steady states u_{ss}. In both full information setup (in which we assume to measure x, w_u, and w_e) and the internal model approach, Davison (1976) Francis and Wonham (1976), the optimal cheap control problem as defined above give the same performance result, i.e. we are able to characterize $J_0 = \lim_{\epsilon\to 0} J_\epsilon$ as a function of the system unstable zeros. In particular, in Qiu and Davison (1993) the authors show that for the w_u and w_e signals oscillating with frequency ω[2] we have for some positive semi-definite M matrix

$$J_0 = W_e^T M W_e \quad (1.20)$$

where W_e refers to the vector of amplitude elements of the sinusoidal terms of w_e, i.e. for $w_e = W_{e1}\sin(\omega t) + W_{e2}\cos(\omega t)$ and $W_e^T = \begin{bmatrix} W_{e1} & W_{e2} \end{bmatrix}$, while the trace of M is

$$\mathrm{tr}M = \sum_{i=1}^{N_z} \left(\frac{1}{\mathrm{z}_i - j\omega} + \frac{1}{\mathrm{z}_i + j\omega} \right).$$

Hence, the closer to the imaginary axis the unstable zeros are the more expensive, in terms of output energy, the problem becomes.

[1] Indeed the optimal control in this case is 0 because the system is already asymptotically stable by itself. And an optimal controller would place the system poles in the open left half plane mirroring the right half plane zeros of the plant. But this is already a property of the inner system.

[2] For this scenario the *non-resonance condition*, i.e. system (A, B, C, D) has no zeros on the imaginary axis at the same frequency ω, is assumed to have necessary and sufficient conditions in order to have a solution to the output regulation problem.

1.3.2 Path-following case

The authors in Aguiar et al. (2005), extend the result of Qiu and Davison (1993), stating that a path-following approach removes the performance limitations imposed by the unstable zeros of the system only for the case of output reference tracking. For the authors, perfect tracking means that the L_2-norm of the tracking error (or its energy) can be made arbitrarily small. To achieve such perfect tracking also for the case of non-minimum phase systems, the authors consider a parametrization of the exosystem dynamics. In particular, assuming to define a timing law $\theta(t)$ afterwards, they describe the geometric path of the reference $y^\star$ to be tracked, by the system under consideration, as exosystem-generated

$$\begin{aligned} \frac{dw(\theta)}{d\theta} =& Sw(\theta) \\ y^\star(\theta) =& Qw(\theta). \end{aligned} \tag{1.21}$$

The system under consideration is the same as in Qiu and Davison (1993), i.e., eq. (1.14) with a zero initial condition $x(0) = 0$, and the path-following to be solved is as follows. For a desired path $y^\star(\theta)$, the designed controller must achieve: boundedness of the state $x(t)$, for $t > 0$, for every initial condition $(x(0), w(\theta(0))$, and convergence of the error $e(t) = y(t) - y^\star(\theta(t))$ as $t \to \infty$, for a forward motion $\dot{\theta} > c \geq 0$ with $t \geq 0$. To the path-following problem, we associate an optimal cheap control problem with cost function

$$J_0 = \int_0^\infty e^T(t)e(t)dt \tag{1.22}$$

constrained to be less than a given positive number δ, i.e. $J_0 \leq \delta$. We call these two united problems the *constrained cheap path following problem.* The main result in Aguiar et al. (2005), that follows step-by-step the approach described in Qiu and Davison (1993), is the following.

Theorem 1.3.1. *If (A, B) is stabilizable, then for any given positive constant δ there exist a timing law $\theta(t)$ and matrices K_x and K_w such that*

$$u(t) = K_x x(t) + K_w w(\theta(t))$$

solves the constrained cheap path-following problem.

And in particular, from the constructive proof, we see that the cost functional can be made arbitrarily small. Indeed, as for the case described in Qiu and Davison (1993), the optimal cost function is given by[3]

$$J_0 = W_e^T M W_e \tag{1.23}$$

where W_e is again the vector of amplitudes of the sinusoidal terms of the output reference, and the semidefinite positive matrix M has trace

$$\mathrm{tr}M = \sum_{i=1}^{N_z} \left(\frac{1}{\mathrm{z}_i - jv_d\omega} + \frac{1}{\mathrm{z}_i + jv_d\omega} \right)$$

and thus J_0 can be made arbitrarily small by choosing a sufficiently fast forward motion of the timing law (i.e., a large v_d). Obviously, the constant v_d must be selected such that the non-resonance condition in the cascade exosystem-system under consideration is fulfilled.

Thanks to this degree of freedom, the authors in Aguiar et al. (2005) were able to obtain the same kind of result even if the path-following problem must be obtained for a specified timing law, i.e., v_d is fixed, thus reconstructing the standard output tracking problem. In particular, thanks to the asymptotic convergence required property, they propose to select a decreasing finite sequence of N constants v_i, $i = 0, \ldots, N$, such that $v_N = v_d$. The values of these v_i are such that in the time interval they are applied, the related cost function J_{vi} is small enough to obtain the desired constraint on the total functional $J_0 \leq \delta$. See Aguiar et al. (2005)[Section IV] for more details.

[3] When applying a timing law

$$\dot{\theta} = v_d, \quad v_d > 0$$

, the exosystem dynamics reads as

$$\begin{aligned} \dot{w}(\theta(t)) =& v_d S w(\theta(t)) \\ y^\star(\theta(t)) =& Qw(\theta(t)). \end{aligned}$$

1.3.3 Cheap control and T-integral relationship

Lyapunov function (energy) interpretation

In Seron et al. (1999), the authors analyze the relationship between the cheap control result obtained in Qiu and Davison (1993) and the standard T-integral limitation formula resulting in Middleton (1991) in the frequency domain. The analysis in Seron et al. (1999) concerns a strictly proper square LTI system

$$\begin{aligned} \dot{x} &= Ax + Bu, \quad & x \in \mathbb{R}^{n_x},\ u \in \mathbb{R}^{n_u} \\ y &= Cx, & y \in \mathbb{R}^{n_u} \end{aligned}$$

which is stabilizable and detectable, and associate to it the cost functional J_ϵ in (1.15). They re-derive the result in Qiu and Davison (1993) for the case of relative degree one system, i.e., rank $CB = n_u$, so that the system under consideration admits a normal form

$$\begin{aligned} \dot{z} &= A_0 z + B_0 y, \quad z \in \mathbb{R}^{n_x - n_u} \\ \dot{y} &= A_1 y + A_2 z + B_1 u. \end{aligned} \tag{1.24}$$

They assume that A_0 is antistable, i.e. $-A_0$ is Hurwitz, and thus all system zeros are non-minimum phase. The Algebraic Riccati equation associated with J_ϵ in the new coordinate is given by

$$\begin{bmatrix} A_1 & A_2 \\ B_0 & A_0 \end{bmatrix}^T P(\epsilon) + P(\epsilon) \begin{bmatrix} A_1 & A_2 \\ B_0 & A_0 \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{\epsilon} P(\epsilon) \begin{bmatrix} B_1 B_1^T & 0 \\ 0 & 0 \end{bmatrix} P(\epsilon), \tag{1.25}$$

to solve it they seek a solution $P(\epsilon)$ of the form

$$P(\epsilon) = \begin{bmatrix} \epsilon P_1 + \epsilon P_2 \\ \epsilon P_2^T + P_0 + \epsilon P_3 \end{bmatrix} + O(\epsilon^2)$$

where P_0,P_1,P_2, and P_3 are independent of ϵ. Then, substituting in (1.25) yields

$$I - P_1 B_1 B_1^T P_1 + O(\epsilon) = 0 \tag{1.26a}$$

$$B_0^T P_0 - P_1 B_1 B_1^T P_2 + O(\epsilon) = 0 \tag{1.26b}$$

$$A_0^T P_0 + P_0 A_0 - P_2^T B_1 B_1^T P_2 + O(\epsilon) = 0. \tag{1.26c}$$

By setting $\epsilon = 0$, we have $P_1 = (B_1 B_1^T)^{-1/2}$, $P_2 = P_1 B_0^T P_0$. Substituting P_2 in (1.26c) and setting $\epsilon = 0$, one obtains

$$A_0^T P_0 + P_0 A_0 = P_0 B_0 B_0^T P_0 \tag{1.27}$$

and because $-A_0$ is Hurwitz, there exists a unique positive definite solution P_0.
Finally, the optimal cost functional value $J_\epsilon^\star = \begin{bmatrix} y & z \end{bmatrix} P(\epsilon) \begin{bmatrix} y & z \end{bmatrix}^T$ can be written as

$$J_\epsilon^\star = \frac{1}{2} z^T P_0 z + \frac{\epsilon}{2} (y + B_0^T P_0 z)^T (B_1 B_1^T)^{-1} (y + B_0^T P_0 z) + O(\epsilon) \tag{1.28}$$

by defining $V_0(z) = \frac{1}{2} z^T P_0 z$, and $V_1(y,z) = \frac{1}{2} (y + B_0^T P_0 z)^T (B_1 B_1^T)^{-1} (y + B_0^T P_0 z)$, the optimal cost functional value $J_\epsilon^\star$ is given by

$$J_\epsilon^\star = V_0(z) + \epsilon V_1(z,y) + O(\epsilon) \tag{1.29}$$

thus for $\epsilon \to 0$ the ideal performance is given by V_0, i.e.,

$$J_0^\star =_{\epsilon \to 0} J_\epsilon^\star = \frac{1}{2} z^T P_0 z = V_0(z). \tag{1.30}$$

Such ideal value has the interpretation of being the least amount of energy required to stabilize the zero-dynamics, indeed

$$J_0 =_{\epsilon \to 0} J_\epsilon = \frac{1}{2} \int_0^\infty y(t)^T y(t) dt. \tag{1.31}$$

Cheap Control and T-integral

We now apply the above results to the problem of regulating the output y of (3) to a constant setpoint $y^\star$. With the feedforward term $u_{ss} = -B_1^{-1}\left(A_1 - A_2A_0^{-1}B_0\right)y^\star$, where they place the equilibrium of (1.24) at $y = y^\star$. Define the error variables $e = y - y^\star$, $\tilde{z} = z + A_0^{-1}B_0y^\star$, and $\tilde{u} = u - u_{ss}$, in the new coordinates we rewrite (1.24) as

$$\begin{aligned}\dot{e} &= A_1e + A_2\tilde{z} + B_1\tilde{u}\\ \dot{\tilde{z}} &= A_0\tilde{z} + B_0e.\end{aligned} \tag{1.32}$$

Then the cheap control problem is the same as for (1.24), but with $(e, \tilde{z})$ in place of (y, z). This time the ideal performance is $V_0(\tilde{z})$. We are interested in $V_0(\tilde{z}(0)) = \tilde{z}(0)^TP_0\tilde{z}(0)$. By considering a trivial initialization for the starting system (1.24) in the origin we have that $e(0) = -y^\star$; $\tilde{z}(0) = A_0^{-1}B_0y^\star$, we can write

$$\begin{aligned}J_0^\star = V_0(\tilde{z}(0)) &= \frac{1}{2}\tilde{z}^T(0)P_0\tilde{z}(0)\\ &= \frac{1}{2}y^{\star T}B_0^T(A_0^T)^{-1}P_0A_0^{-1}B_0y^\star = \frac{1}{2}y^{\star T}\bar{P}_0y^\star,\end{aligned} \tag{1.33}$$

where the trace of $\bar{P}_0$ can be directly related to the eigenvalues of A_0^{-1}

$$\begin{aligned}\text{trace } \bar{P}_0 &= \text{trace } \left[B_0^T(A_0^T)^{-1}P_0^{1/2}P_0^{1/2}A_0^{-1}B_0\right]\\ &= \text{trace } \left[P_0^{1/2}(A_0^T)^{-1}B_0B_0^TA_0^{-1}P_0^{1/2}\right]\\ &= \text{trace } \left[P_0^{1/2}A_0^{-1}P_0^{-1/2} + P_0^{-1/2}A_0^{-T}P_0^{1/2}\right]\\ &= 2\text{trace } A_0^{-1} = 2\sum_i^{n_x-n_u}\frac{1}{\mathrm{z}_i}\end{aligned} \tag{1.34}$$

where z_i are the eigenvalues of A_0. Now, we realize the relationship with (1.11). Indeed, equation (1.11) is the Bode integral for the complementary sensitivity function T

$$\frac{1}{\pi}\int_0^\infty log|T(j\omega)|\frac{d\omega}{\omega^2} + \frac{1}{2K_v} = \sum_i^{n_x-n_u}\frac{1}{\mathrm{z}_i}. \tag{1.35}$$

For a single input single output linear system, whose controller is minimum phase and such that the closed loop is asymptotically stable, hence, the complementary sensitivity function $T(s)$ is stable we have that

$$\frac{1}{\pi}\int_0^\infty log|T(j\omega)|\frac{d\omega}{\omega^2} + \frac{1}{2K_v} = \lim_{\epsilon\to 0}\frac{1}{2}\int_0^\infty e^2(t) + \epsilon^2u(t)^Tu(t)dt \tag{1.36}$$

where, K_v is the high frequency gain of the system under consideration with $e(t) = y - y^\star$.

1.3.4 Under/overshoot of output signal

In Lau et al. (2003)[Section C] consider a the output tracking problem of a constant reference value $\bar{y}$ for strict proper LTI SISO system in normal form, i.e.,

$$\begin{aligned}\dot{z} &= Fz + Gy, \quad z(0) = 0\\ \dot{\xi} &= Hz + \bar{A}\xi + \bar{B}u, \quad \xi(0) = 0\\ y &= \bar{C}\xi\end{aligned}$$

in which the zeros of the equivalent plant $G(s)$ are the eigenvalues of F in the zero dynamics with state z. In this work, the authors consider the relationship between the output undershoot, the settling time and the location of the unstable zeros (only for the case of one and two zeros present in the system). Thus, assuming F to be nonsingular, associated to the target equilibrium $\bar{y}$ we can define the equilibrium point for $z(t)$ as $\bar{z} = -F^{-1}G\bar{y}$. Considering, the evolution of a stabilizing $y(t)$ for the zero dynamics in finite time T_s, i.e. $z(t) = \bar{z}$ for all $t \geq T_s$, such that $y(t) = \bar{y}$ for all $t \geq T_s$, we can define $y_{us} > 0$ as the minimum value[4] such that $y(t) \geq -y_{us}$ for all $t \geq 0$, and the relative undershoot as $y_{us}/\bar{y}$. For the scalar zero dynamics case with $G = 1$, we have that

$$z(t) = \int_0^t e^{F(t-\tau)}y(\tau)d\tau \tag{1.37}$$

[4]We are implicitly assuming that $\bar{y} > 0$, without loss of generality.

and because $y(t) \geq -y_{us}$ for all $t \geq 0$, we can write

$$-z(t) \leq \int_0^t e^{F(t-\tau)} d\tau \, y_{us} = F^{-1}(e^{Ft} - 1) y_{us}$$

and since for all $t \geq T_s$ we have $z(T_s) = -F^{-1}\bar{y}$ we can then write

$$F^{-1}\bar{y} \leq F^{-1}(e^{FT} - 1) y_{us}$$

that is, we thus recover the same result obtained in the frequency domain approach (1.13)

$$y_{us} \geq \frac{1}{e^{FT} - 1} \bar{y}$$

that is, the undershoot level is lower bounded by the function depending on the reference amplitude, the location of the zero and on the designed settling time of the closed loop system.

1.3.5 Non-overshooting and Non-undershooting design

In Schmid and Ntogramatzidis (2012), the authors propose a procedure to design a feedback gain matrix K to obtain a non-overshooting or non-undershooting step response. In particular, consider an (exactly) proper LTI system (A, B, C, D) in state space realization

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{1.38}$$

with $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$, and $y \in \mathbb{R}^{n_y}$, with $n_u = n_y$ (thus considering a square system), initialized at an equilibrium configuration, i.e., $u(0)$ and $x(0)$ are such that $0 = Ax(0) + Bu(0)$. The goal is to design a feedback control law for system (1.38) such that $y(t)$ asymptotically tracks a step reference $y^\star = Qw \in \mathbb{R}^{n_y}$, with $\dot{w} = 0$ being the scalar exosystem dynamics initialized at $w(0) = w_0$.

Now, assuming that system (1.38) is right invertible, stabilizable and has no zero at the origin, one is able to solve for Π and Σ the regulator equations

$$\begin{aligned} 0 &= A\Pi + B\Sigma \\ Q &= C\Pi + D\Sigma. \end{aligned} \tag{1.39}$$

Then, in this full information framework, the control law can be taken as

$$u(t) = K(x(t) - \Pi w(t)) + \Sigma w, \quad \forall t > 0 \tag{1.40}$$

and by putting the system in the error coordinates defined as $\tilde{x} = x - \Pi w$, we obtain the closed loop dynamics

$$\begin{aligned} \dot{\tilde{x}} &= (A + BK)\tilde{x} \\ y &= (C + DK)\tilde{x} + Qw \end{aligned} \tag{1.41}$$

with the closed loop matrix $(A + BK)$ asymptotically stable. As a consequence, the error converges to zero asymptotically and the output eventually reaches the value $y^\star$.
To determine if a SISO system output overshoots or undershoots, one can either analyse the output response with respect to the reference signal or directly check the tracking error behaviour, $e(t) = Qw(t) - y(t) = y^\star - y(t)$. When looking at output behaviour, in case $y(0) < y^\star$, overshooting means that the output $y(t)$, for $t > 0$, takes values less than or equal to $y^\star$, while undershooting means that $y(t)$, for $t > 0$, takes values greater than or equal to $y(0)$. On the other hand, similar considerations can be done in terms of the tracking error $e(t)$. In this case, overshooting corresponds to a zero crossing behaviour (or equivalently $e(t)$ as function of time has a real positive root), while undershooting occurs when $|e(t)| > |e(0)|$, for some $t > 0$.

Moreover, note that for (exactly) proper system at $t = 0^+$ we may have on output $y(t)$ (and consequently an error $e(t)$) discontinuity with respect to the initial equilibrium configuration $y(0)$ (and $e(0)$, respectively). We can thus introduce the notion of instantaneous overshoot and undershoot, that is having on the output $y(t)$ an overshooting or undershooting behaviour at $t \to 0^+$, rather than for $t > 0$. The authors in Schmid and Ntogramatzidis (2012) hence introduce a notation to directly characterize the instantaneous overshoot or undershoot by looking at the error behaviour around the time origin. In particular, they take $e(0^+) = \mu e(0)$ for some real scalar μ. If $\mu \leq 0$ then we have instantaneous

overshoot, if $\mu > 1$ then we have instantaneous undershoot. And for $0 < \mu \leq 1$, the overshoot occurs if there exists a positive real $\bar{t}$ such that $e(0)e(\bar{t}) < 0$ (which is again the presence of zero crossing). By the way, for strictly proper systems the discontinuity around the time origin is not possible and so $\mu = 1$.

To exploit these considerations, the authors of Schmid and Ntogramatzidis (2012) aim at finding an algorithm to construct a stabilizing feedback gain matrix K so that the error evolution $e(t)$ can be described by a particular combination of real exponential functions. This algorithm is an adaptation of Moore's one, introduced in (Moore, 1975, Proposition 1), to assign particular eigenstructure to the closed-loop system.

Lemma 1. *(Lemma 3.1 in Schmid and Ntogramatzidis (2012)) Let $\Lambda = \{\lambda_1, \ldots, \lambda_{n_x}\}$ be a self-conjugate set of n_x distinct complex numbers. Let $\mathcal{S} = \{\mathrm{s}_1, \ldots, \mathrm{s}_{n_x}\}$ be a set of n_x (not necessarily distinct) vectors in $\mathbb{R}^{n_y}$. Assume that, for each $i = \{1, \ldots, n_x\}$, the matrix equation*

$$\begin{bmatrix} A - \lambda_i I & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathrm{v}_i \\ \mathrm{w}_i \end{bmatrix} = \begin{bmatrix} 0 \\ \mathrm{s}_i \end{bmatrix} \tag{1.42}$$

has solutions sets $\mathrm{V} = \{\mathrm{v}_1, \ldots, \mathrm{v}_{n_x}\} \subset \mathbb{C}^{n_x}$ and $\mathrm{W} = \{\mathrm{w}_1, \ldots, \mathrm{w}_{n_x}\} \subset \mathbb{C}^{n_u}$. Then, provided that V has linearly independent elements, a unique real feedback matrix F exists such that, for all $i = 1, \ldots, n_x$,

$$(A + BF)\mathrm{v}_i = \lambda_i \mathrm{v}_i \tag{1.43}$$

$$(C + DF)\mathrm{v}_i = \mathrm{s}_i. \tag{1.44}$$

For square systems, i.e., $n_y = n_u$, if $\mathrm{s}_i = 0$, then (1.42) is solvable for non trivial v_i if and only if λ_i is a zero of system (1.38). For arbitrary $\mathrm{s}_i \neq 0$, then (1.42) is solvable if and only if λ_i is not a zero of system (1.38). Assuming that the solution set V and W do exist and that V has all linearly independent elements, then the feedback gain matrix $K = \hat{\mathrm{W}}\hat{\mathrm{V}}^{-1}$, where the matrices $\hat{\mathrm{V}}$ and $\hat{\mathrm{W}}$, are constructed as $\hat{\mathrm{V}} = [\mathrm{v}_1 \vdots \ldots \vdots \mathrm{v}_{n_x}]$ and $\hat{\mathrm{W}} = [\mathrm{w}_1 \vdots \ldots \vdots \mathrm{w}_{n_x}]$, for real[5] λ_i, $i = 1, \ldots, n_x$.

A particular choice of the λ_i values and of the vectors s_i, $i = 1, \ldots, n_x$, yields an exponential form of the error function vector $e(t)$, for $t > 0$, containing exactly one mode per component. Assume that system (1.38) has at least $n_x - n_y$ distinct minimum phase zeros, i.e., zeros in $\mathbb{C}_-$. Let $\{\mathrm{z}_1, \ldots, \mathrm{z}_{n_x - n_y}\}$ be $n_x - n_y$ of such zeros and let $\{\lambda_{n_x - n_y + 1}, \ldots, \lambda_{n_x}\}$ be any real distinct stable modes not coincident with invariant zeros of (1.38), construct the closed loop eigenvalues set $\Lambda = \{\mathrm{z}_1, \ldots, \mathrm{z}_{n_x - n_y}, \lambda_{n_x - n_y + 1}, \ldots, \lambda_{n_x}\}$. With $\mathbf{e}_1, \ldots, \mathbf{e}_p$ being the canonical basis of $\mathbb{R}^{n_y}$, and $\mathrm{s}_i = 0$ for $i = 1, \ldots, n_x - n_y$, and $\mathrm{s}_{n_x - n_y + j} = \mathbf{e}_j$, for $j = 1, \ldots, n_y$, assume that elements of the solution set V are linearly independent[6]. Now introduce the error coordinate $\tilde{x} = x - \Pi w$, and $\tilde{x}(0) = x(0) - \Pi w(0)$ and defied $a = \begin{bmatrix} a_1 & \ldots & a_{n_x} \end{bmatrix}^T = \hat{V}^{-1}\tilde{x}(0)$. The following Theorem describes the form of the error term.

Theorem 1.3.2 (Theorem 3.1 in Schmid and Ntogramatzidis (2012))**.** *Assume that system (1.38) is square and has at least $n_x - n_y$ distinct minimum phase zeros. Let $K = \hat{\mathrm{W}}\hat{\mathrm{V}}^{-1}$ obtained from the above steps, let $y^\star = Qw \in \mathbb{R}^{n_y}$ be any step reference, and let $(u(0), x(0), y(0))$ be the initial equilibrium configuration. Then, the error vector $e(t) = y^\star - y(t)$ obtained from the closed loop (1.38)-(1.40), has the form*

$$e(t) = \begin{bmatrix} a_{n_x - n_y + 1} e^{\lambda_{n_x - n_y + 1} t} \\ \vdots \\ a_{n_x} e^{\lambda_{n_x} t} \end{bmatrix}. \tag{1.45}$$

Consider now the case in which system (1.38) has fewer than $n_x - n_y$ minimum phase zeros and the number of such asymptotically stable zeros can be written as $n_x - ln_y + q$ for some l and q. In final goal is to obtain an error evolution with l (or less) modes per component, i.e., the number of exponential terms in the error components. For simplicity, consider the case in which $q = 0$. Then choose $\Lambda = \{\mathrm{z}_1, \ldots, \mathrm{z}_{n_x - ln_y}, \lambda_{n_x - ln_y + 1}, \ldots, \lambda_{n_x}\}$ where, z_i, for $i = 1, \ldots, n_x - ln_y$, are the minimum phase zeros of the system and λ_i, for $i = n_x - ln_y + 1, \ldots, n_x$, are freely-chosen distinct real and stable eigenvalues not coinciding with the zeros of (1.38). As above, take the s_i vectors as follows, for $i = 1, \ldots, n_x - ln_y$, let $\mathrm{s}_i = 0$, while for $i = n_x - ln_y + 1, \ldots, n_x - ln_y + l$, let $\mathrm{s}_i = \mathbf{e}_1$, for $i = n_x - ln_y + l + 1, \ldots, n_x - ln_y + 2l$, let $\mathrm{s}_i = \mathbf{e}_2$, etc.[7], till for $i = n_x - l + 1, \ldots, n_x$, let $\mathrm{s}_i = \mathbf{e}_p$, where $\{\mathbf{e}_1, \ldots, \mathbf{e}_p\}$, as before, are the canonical basis

[5] For complex conjugate pair of eigenvalues λ_i, λ_{i+1}, we refer the reader to Schmid and Ntogramatzidis (2012)(Remark 3.1)

[6] Note that for the discussion above, thanks to the choice of the s_i and the λ_i for $i = 1, \ldots, n_x$, the solution sets V and W always exist, we have no guarantee about the linear independence among the vectors in V

[7] One has to assign the j-th basis of $\mathbb{R}^{n_y}$, $\mathbf{e}_j$ to vectors s_i, for $i = n_x - ln_y + (j-1)l, \ldots, n_x - ln_y + jl$. Note that in Schmid and Ntogramatzidis (2012) equation (16) has a typo in last row.

of $\mathbb{R}^{n_y}$. With this choice of the Λ and $\mathcal{S}$ the solution sets V and W always exist, but we have to further assume that the vectors of V are all linearly independent. Now, denoting as $\mathrm{v}_{k,1}, \ldots, \mathrm{v}_{k,l}$ the eigenvectors in V associated with the canonical basis $\mathbf{e}_k$, let $\lambda_{k,1}, \ldots, \lambda_{k,l}$ be the corresponding eigenvalues, without loss of generality, ordered in increasing way. Denote, moreover, the coefficient $a = \hat{\mathrm{V}}^{-1}\tilde{x}(0)$, with elements given by

$$a = \left[a_1 \quad \cdots \quad a_{n_x - l n_y} \vdots \; a_{1,1} \quad \vdots \quad a_{1,l} \vdots \quad \cdots \quad a_{n_y,1} \quad \cdots \quad a_{n_y,l}\right]^T.$$

Then the following theorem holds.

Theorem 1.3.3 (Theorem 3.2 in Schmid and Ntogramatzidis (2012))**.** *Assume that system (1.38) is square and has exactly $n_x - l n_y$ distinct minimum phase zeros. Let $K = \hat{\mathrm{W}}\hat{\mathrm{V}}^{-1}$ obtained from the above steps, let $y^\star = Qw \in \mathbb{R}^{n_y}$ be any step reference, and let $(u(0), x(0), y(0))$ be the initial equilibrium configuration. Then, the k-th component $e_k(t)$ of the error vector $e(t) = y^\star - y(t)$ obtained from the closed loop (1.38)-(1.40), has the form*

$$e_k(t) = a_{k,1}e^{\lambda_{k,1}t} + a_{k,2}e^{\lambda_{k,2}t} \cdots + a_{k,l}e^{\lambda_{k,l}t}. \tag{1.46}$$

If $q > 0$, then we have $n_x - l + q$ minimum phase zeros, and the additional minimum phase zeros must be used in the design. This modifies the vectors s_i, so that, for $i = 1, \ldots, n_x - l n_y + q$, $\mathrm{s}_i = 0$. Then, we have to choose q of the outputs and allocate into them only $l - 1$ modes. In this case, the associated q canonical basis vectors corresponding to such outputs need to be related to only $l - 1$ eigenvalues each. The remaining $n_y - q$ will then have l modes, as shown for the case $q = 0$.

We now only need to characterize the conditions on such functions of exponential terms that imply overshoot and undershoot in the output response, or equivalently in the error behaviour. In this respective, denote by $\{\lambda_1, \ldots, \lambda_l\}$ and $\{a_1, \ldots, a_l\}$, with $\lambda_1 < \cdots < \lambda_l$. define $\beta : \mathbb{R} \to \mathbb{R}$ as

$$\beta(t) = a_1 e^{\lambda_1 t} + \cdots + a_l e^{\lambda_l t}. \tag{1.47}$$

Let $Sc\{a_1, \ldots, a_l\}$ denote the number of sign changes in the sequence of coefficient $\{a_1, \ldots, a_l\}$, and for an real interval $[\mathbf{a}, \mathbf{b}] \subseteq \mathbb{R}$, let $Z^{\beta}_{[\mathbf{a},\mathbf{b}]}$ denote the number of real roots of β in $[\mathbf{a}, \mathbf{b}]$. Let us also introduce the values $p_i = \sum_{j=1}^{i} a_j$, for $i = 1, \ldots, l$ and $q_i = q_{i-1} + p_i(\lambda_i - \lambda_{i-1})$ for $i = 1, \ldots, l-1$ with $q_0 = 0$ and $q_l = p_l$. Also, introduce $r_1 = \sum_{j=0}^{i-1} a_{l-j}$, for $i = 1, \ldots, l$, and $s_i = s_{i-1} + r_i(\lambda_{l-i+1} - \lambda_{l-i})$ for $i = 1, \ldots, l-1$ with $\mathrm{v}_0 = 0$ and $s_l = r_l$. We now need to introduce the next basic lemmas before providing the design algorithm provided in Schmid and Ntogramatzidis (2012), whose proofs can be found in their work.

Lemma 2 (Lemma 4.2 in Schmid and Ntogramatzidis (2012))**.** *Let $\lambda_1 < \lambda_2 < \lambda_3 < 0$, and, for any real constants $\{a_1, a_2, a_3\}$ with $a_3 \neq 0$, define*

$$\beta(t) = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} + a_3 e^{\lambda_3 t}.$$

There exists a positive real $\bar{t}$ such that $\beta(\bar{t}) = 0$ if and only if one of the following conditions holds:

I $Sc\{a_1, a_2, a_3\} = 1$ *and* $(a_1 + a_2 + a_3)a_3 < 0$*;*

II $Sc\{a_1, a_2, a_3\} = 2$ *and* $(a_1 + a_2 + a_3)a_3 \geq 0$*;*

III $Sc\{a_1, a_2, a_3\} = 2$ *and* $(a_1 + a_2 + a_3)a_3 < 0$*, with* $\underline{t} > 0$ *and* $|\gamma(\underline{t}s)| \geq |a_1 + a_2 + a_3|$*, where*

$$\underline{t} = \frac{1}{\lambda_3 - \lambda_1} \ln\left(\frac{a_1(\lambda_2 - \lambda_1)}{a_3(\lambda_3 - \lambda_2)}\right)$$

$$\gamma(t) = a_1\left(1 - e^{(\lambda_1 - \lambda_2)t}\right) + a_3\left(1 - e^{(\lambda_3 - \lambda_2)t}\right).$$

Lemma 3 (Lemma 4.3 in Schmid and Ntogramatzidis (2012))**.** *Let $\lambda_1 < \lambda_2 < 0$, and, for any real nonzero constants $\{a_1, a_2\}$, define*

$$\beta(t) = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t}.$$

Let $b = (a_1 + a_2)/\mu$, for some $0 < \mu \leq 1$. Then there exists positive real $\bar{t}$ such that $\beta(\bar{t}) = b$ if and only if $\underline{t}$ and $\beta(\underline{t})b \geq b^2$, where

$$\underline{t} = \frac{1}{\lambda_2 - \lambda_1} \ln\left(-\frac{a_1\lambda_1}{a_2\lambda_2}\right).$$

Lemma 4 (Lemma 4.4 in Schmid and Ntogramatzidis (2012)). *Let $\lambda_1 < \lambda_2 < \lambda_3 < 0$, and, for any nonzero constants $\{a_1, a_2, a_3\}$, define $b = (a_1 + a_2 + a_3)/\mu$, for some $0 < \mu \leq 1$. Let*

$$\beta(t) = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} + a_3 e^{\lambda_3 t}$$

and consider p_i, q_i, r_i, and s_i, for $i =, \ldots, l$, defined as above Lemma2 and with $l = 4$. Then there exists positive real $\bar{t}$ such that $\beta(\bar{t}) = b$ only if at least one of the following conditions hold:

I $Sc\{q_1, q_2, q_3, q_4\} \geq 1$;

II $Sc\{r_1, r_2, r_3, r_4\} \geq 1$.

Lemma 5 (Lemma 4.5 in Schmid and Ntogramatzidis (2012)). *For some positive integer l, let $\lambda_1 < \lambda_2 < \cdots < \lambda_l < 0$, and, for any nonzero constants $\{a_1, a_2, \ldots, a_l\}$, define*

$$\beta(t) = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} + \cdots + a_l e^{\lambda_l t}.$$

Then

(a) There exists positive real $\bar{t}$ such that $\beta(\bar{t}) = 0$ only if

$$Sc\{q_1, q_2, \ldots, q_l\} \geq 1$$

or

$$Sc\{r_1, r_2, \ldots, r_l\} \geq 1,$$

with q_i and r_i, $i = 1, \ldots, l$ defined as above Lemma 2;

(b) Let $b = \sum_{i=1}^{l} a_i/\mu$ for some $0 < \mu \leq 1$, and define $a_{l+1} = -b$ and $\lambda_{l+1} = 0$. Then there exists positive real $\bar{t}$ such that $\beta(\bar{t}) = b$ only if

$$Sc\{q_1, q_2, \ldots, q_{l+1}\} \geq 1$$

or

$$Sc\{r_1, r_2, \ldots, r_{l+1}\} \geq 1,$$

with q_i and r_i, $i = 1, \ldots, l+1$ defined as above Lemma 2, in which last element of the sequence in this case is indexed $l+1$ rather then simply l.

In the end, the proposed algorithm in Schmid and Ntogramatzidis (2012) reads as follows:

(i) Begin by determining the number of minimum phase zeros of system (1.38), i.e., $n_x - l n_y + q$, and thus solve for l and q.

(ii) For a given initial condition $x(0)$ and reference $y^\star = Qw$, determine Π and Σ from the regulator equations (1.39) and obtain also $\tilde{x}(0) = x(0) - \Pi w$.

(iii) Choose a desired interval $[\mathbf{a}, \mathbf{b}]$ of the real line (where $\mathbf{a} < \mathbf{b} < 0$), and form a candidate set $\{\lambda_1, \ldots, \lambda_{n_x}\}$ of n_x distinct closed-loop eigenvalues containing the $n_x - l + n_y$ minimum phase zeros of (1.38), and n_y sets of l eigenvalues chosen from within $[\mathbf{a}, \mathbf{b}]$.

(iv) Determine the target set $\{\mathrm{s}_1, \ldots, \mathrm{s}_{n_x}\}$ in accordance with the canonical basis of $\mathbb{R}^{n_y}$, taking into account the fact that $q \geq 0$. Then solve (1.42), for $i = 1, \ldots, n_x$ for the corresponding sets V and W and check if V has linearly independent elements. If it is not, then *return to Step (iii) and choose an alternative set of eigenvalues* within $[\mathbf{a}, \mathbf{b}]$.

(v) Obtain the coordinate vector $a = \mathrm{V}^{-1}\tilde{x}(0)$, and hence obtain the components e_k of the tracking error $e = y^\star - y$, for $k = 1, \ldots, n_y$;

(vi) For strictly proper systems, skip this step and proceed directly to Step (vii); For (exactly) proper systems, solve $e_k(0^+) = \mu e_k(0)$ for real scale μ_k, then

[a] For a step response without instantaneous overshoot, check that $\mu_k > 0$

[b] For a step response without instantaneous undershoot, check that $\mu_k < 1$ for $k = 1, \ldots, n_y$. If not, return to Step (*iii*).

(vii) If $l = 1$, proceed directly to next Step. For $l \geq 2$, for $k = 1, \ldots, n_y$ do the following :

[a] For a nonovershooting response, test each e_k for the conditions in Lemma 2 if $l \in \{2, 3\}$, or Lemma 5[(a)] if $l \geq 4$;

[b] For a nonundershooting response, test each e_k for the conditions in Lemma 3 if $l = 2$, Lemma 4 if $l = 3$, or Lemma 5[(b)] if $l \geq 4$;

[c] For a monotonic response, test each $\dot{e}_k$ for the conditions in Lemma 2 if $l \in \{2, 3\}$, or Lemma 5[(a)] if $l \geq 4$. In each case, if the conditions in the respective lemmas are satisfied for any $k \in \{1, \ldots, n_y\}$, then the sets $\{\lambda_1, \ldots, \lambda_{n_x}\}$ and V are satisfactory. If not, then *return to Step* (*iii*).

(viii) Apply Moore's algorithm to obtain the feedback matrix $K = \hat{\mathrm{W}}\hat{\mathrm{V}}^{-1}$, and define the control action u as in (1.40).

The main drawback of this approach is that, assuming that the solution sets can be found[8], the solution is not robust to parametric uncertainties of the plant, and it is thus almost impossible to obtain the same theoretical results in a real plant controller implementation. Furthermore, no conditions are provided on the choice of the eigenvalues that allow us to obtain a linearly independent set V. Indeed, the algorithm works in a trial and error fashion with no guarantee of convergence on the desired solution.

1.3.6 Funnel control approach

In Berger (2020) instead, the analysis of the performance in terms of funnel control, i.e., the system output time evolution is constrained to be in between upper and lower boundaries called funnel functions. Consider a square LTI system with strict relative degree r in normal form realization, for notation simplicity,

$$\begin{aligned} \dot{z} =& Fz + Gy + d_z(t) \\ \dot{\xi} =& Hz + \bar{A}\xi + \bar{B}(bu + \Lambda\xi) + d_r(t) \\ y =& \bar{C}\xi \end{aligned} \tag{1.48}$$

where $d_z(t)$ and $d_r(t)$ are external disturbances, and $(\bar{A}, \bar{B}, \bar{C})$ is in prime form of dimension r. Assuming that F is in block diagonal form

$$F = \begin{bmatrix} F_1 & 0 & 0 \\ 0 & F_2 & 0 \\ 0 & 0 & F_3 \end{bmatrix}, \quad G = \begin{bmatrix} G_1 \\ G_2 \\ G_3 \end{bmatrix}$$

where $\sigma(F_1) \subset \mathbb{C}_-$, $\sigma(F_2) \subset \mathbb{C}_+$, and $\sigma(F_3) \subset i\mathbb{R}$, with (F, G) stabilizable[9] (which is implied by assuming that system (1.48) is stabilizable). Consider (F, G) controllable and define $\Gamma = [0, \ldots, b^{-1}]\mathcal{C}_z^{-1}$, where $\mathcal{C}_z = [G, FG, \ldots, F^{n_x - r}G]$ is the controllability matrix associated to the (F, G) pair. In Berger (2020) the author consider a particular form of controllability for the zero dynamics for r_z m-tuples in the elements of F and G we have $[G, FG, \ldots, F^{r_z - 1}G] = \mathcal{C}_z$, being r_z the controllability index of the system zero dynamics with respect to its driving signal $y(t)$. To deal with the stabilization of the system zero dynamics, the author in Berger (2020) define a new auxiliary output $y_{new} = \Gamma z_2$, and obtain its r_z-th time derivative

$$y_{new}^{(r_z)} = \Gamma\bar{F}_2 z_2 + \Gamma^{-1} y(t) + \Gamma F^{r_z - 1} d_{z2}$$

where d_{z2} is assumed to be zero to develop the proposed approach, thus by assumption we have

$$y_{new}^{(r_z)} = \Gamma\bar{F}_2 z_2 + \Gamma^{-1} y(t), \tag{1.49}$$

from which we can write

$$y(t) = \Gamma y_{new}^{(r_z)} - \Gamma z_2.$$

Including now the time derivatives of $y(t)$ in the successive time derivative of $y_{new}^{(r_z + i)}$, for $i = 1, \ldots, r$, we have can write the complete y_{new} dynamics, whose relative degree becomes $r_z + r$:

$$\begin{aligned} y^{(r_z + r)} =& \lambda_{new}\xi_{new} + H_1 z_1 + H_3 z_3 + \Gamma^{-1} d_r(t) + u(t) \\ \dot{z}_1 =& F_{1,new} z_1 + G_{1,new}\xi_{new} + d_z 1 \\ \dot{z}_3 =& F_{3,new} z_3 + G_{3,new}\xi_{new} + d_z 1 \end{aligned}$$

[8] No guarantee on the existence of the solution sets V and W has been provided. This also implies that there is no guarantee that the algorithm will eventually converge.

[9] In order to guarantee the controllability matrix invertibility

where λ_{new} and ξ_{new} include the dynamics induced by the new output y_{new} definition so that $y_{new} = \xi_{1,new}$ and $y_{new}^{(i-1)} = \xi_{i,new}$ for $i = 2, \dots, r+r_z$. We now need to define the new output reference trajectory, $y_{new}^\star$, by solving the initial value problem associated with the steady state of the z_2 dynamics, i.e. compute $z_2^\star(0)$ such that

$$\begin{aligned} \dot{z}_2^\star &= F_2 z_2^\star + G_2 y^\star \\ y_2^\star &= \Gamma z_2^\star \end{aligned}$$

such that $z_2^\star(t)$ is bounded for all $t \geq 0$ and continuously differentiable $r + r_z$ times. Such a solution can be analytically computed as

$$z_2^\star(0) = -\int_0^\infty e^{-F_2 s} G_2 y^\star(s) ds$$

for which $y^\star(t)$ must be known for all $t \geq 0$.

Then, the proposed controller design is based on a back-stepping approach Sepulchre et al. (2012) to apply a funnel control strategy. Indeed, defining

$$\begin{aligned} e_0 &= y_{new} - y_{new}^\star \\ e_1 &= \dot{e}_0 + K_0 e_0 \\ &\dots \\ e_i &= \dot{e}_{i-1} + K_{i-1} e_{i-1}, \quad i = 2, \dots, r + r_z - 1 \end{aligned} \tag{1.50}$$

with $K_i = (1 - \phi_i(t)^2 \|e_i(t)\|^2)^{-1}$, for $i = 0, \dots, r + r_z - 1$, and control law

$$u(t) = -K_{r+r_z-1} e_{r+r_z-1} \tag{1.51}$$

where $\phi_i(t)$, $i = 0, \dots, r+r_z-1$, are the funnel functions determining the boundaries of the relative error evolutions. Such $\phi_i \in \Phi_{r+r_z-i}$, for $i = 0, \dots, r + r_z - 1$, where the class functions Φ_j are defined as

$$\Phi_j = \left\{ \phi \in C^j(\mathbb{R}^+ \to \mathbb{R}), \text{ s.t. } \phi, \dots, \phi^{(j)} \text{ are bounded, } \phi(t) > 0, \forall t \geq 0, \textit{and} \liminf_{t \to \infty} \phi(t) > 0 \right\}.$$

In Berger (2020) the following theorem has been proven.

Theorem 1.3.4 (Theorem 3.3 in Berger (2020))**.** *Consider the linear system (1.48) with strict relative degree r, with (F, G) stabilizable and $d_{z2} = 0$. Let $y^\star \in C^r(\mathbb{R}^+ \to \mathbb{R})$ and $\phi_i \in \Phi_{r+r_z-i}$, for $i = 0, \dots, r + r_z - 1$. Then the closed loop system (1.48)-(1.51), has the following properties*

- *all signals $z, \xi, z_2^\star, u, K_0, \dots, K_{r+r_z-1}$ are bounded.*
- *the error signals $e_1, \dots, e_{r+r_z-1}$ evolve uniformly within the respective performance funnels functions, that is $\forall i = 0, \dots, r + r_z - 1$, there exists $\epsilon_i > 0$ such that*

$$|e_i(t)|^2 \leq \phi_i^{-1}(t) - \epsilon_i, \quad \forall t \geq 0.$$

In Berger (2020) it is shown that the initial tracking problem, according to the error signal $e(t) = y(t) - y^\star(t)$, is achieved with some *funnel* performances, i.e.,

$$\|e(t)\| = \|y(t) - y^\star(t)\| \leq \sum_{i=1}^{r_z+1} \alpha_i (\phi_{i-1}^{-1} + \bar{\Gamma}_{i-2})$$

for some α_i and $\bar{\Gamma}_{i-2}$ for $i = 1, \dots, r_z + 1$. Then controller (1.51) achieves the prescribed performance of the original tracking error, i.e., for any $\phi \in \Phi_0$, we have $\|e(t)\| \leq \phi^{-1}(t)$ for all $t \geq 0$ such that $\phi(0)\|e(0)\| < 1$.
Unfortunately, as reported in Berger (2020), a general procedure to construct $\phi_0, \dots, \phi_{r+r_z-1}$, and $k_0, \dots, k_{r+r_z-1}$, for a given desired performance $\phi \in \Phi_0$ is not available. The author proposes to design such a function via offline simulation.
Moreover, the approach is not robust to parameter uncertainties while the knowledge of the future behaviour of the output reference signal may not be available. Thus, the approach is again very difficult to implement in a real plant controller.

1.4 Performance limits in non-minimum phase: nonlinear systems

1.4.1 Performance limitations

The same idea of section 1.3.4 can be extended to the case of nonlinear non-minimum phase systems. In Lau et al. (2003), the authors consider a SISO nonlinear system

$$\begin{aligned} \dot{\xi} &= F(\xi, z, u) \\ \dot{z} &= F_0(z, y) \\ y &= H(\xi) \end{aligned} \tag{1.52}$$

with $\xi \in \mathbb{R}^r$ and $z \in \mathbb{R}^{n_x - r}$, where r is the output relative degree uniformly defined.
Denote by $\phi(t, z_0, y)$ the solution of the zero dynamics starting at initial condition $z(0) = z_0$.
Assume that for every $\bar{y}$ in $\mathbb{R}$, $\dot{z} = F_0(z, \bar{y})$ has a unique equilibrium point $\bar{z}$ which implies $0 = F_0(\bar{z}, \bar{y})$. Moreover, assume, without loss of generality, that the state space origin is an equilibrium for the zero dynamics, i.e., $F_0(0, 0) = 0$.
Such an equilibrium point $\bar{z}$ is *unstable* if it is not (locally) asymptotically stable. It is *anti-stable* if $\dot{z} = -F_0(z, \bar{y})$ is (locally) asymptotically stable. The zero dynamics are unstable (antistable) if, for all $\bar{y}$, the corresponding equilibrium point is unstable (antistable). If $\bar{z}$ is unstable then the stable manifold $\mathcal{M}_{\bar{z}}$, corresponding to $\bar{z}$, is given by

$$\mathcal{M}_{\bar{z}} = \left\{ z_0 \in \mathbb{R}^{n_x - r} \ : \ \lim_{t \to +\infty} \phi(t, z_0, \bar{y}) = \bar{z} \right\}. \tag{1.53}$$

Recall that, for each $\gamma \geq 0$, $\mathcal{Y}_\gamma$ is the set of functions y satisfying $y(t) \geq -\alpha$ for all $t > 0$.
They thus concern with the problem of taking the system from *rest* to the equilibrium $y(t) = \bar{y} > 0$. This is equivalent to finding $y(t)$, which satisfies the following constraints:

$$\lim_{t \to +\infty} y(t) = \bar{y} \tag{1.54a}$$

$$\lim_{t \to +\infty} \phi(t, 0, y) = \bar{z}. \tag{1.54b}$$

We say that y has an *exact* finite settling time T if $y(t) = \bar{y}$ for all $t > T$.
Now, consider the system described by (1.52). For each triple (z_0, α, T) the reachable set, $\mathcal{R}_{z_0,\alpha,T}$ is the set given by

$$\mathcal{R}_{z_0,\alpha,T} = \left\{ z_{\mathbb{R}} \in \mathbb{R}^{n_x - r} : \exists y \in \mathcal{Y}_\gamma \text{ s.t. } \phi(T, z_0, y) = z_{\mathcal{R}} \right\}$$

and a set $S \subseteq \mathbb{R}^{n_x - r}$ is reachable if $S \subseteq \mathcal{R}_{z_0,\alpha,T}$. A set $S_u \subseteq \mathbb{R}^{n_x - r}$ is unreachable if $\mathcal{R}_{z_0,\alpha,T} \subseteq S_u^c$, where $S_u^c = \mathbb{R}^{n_x - r} / S_u$.

We observe that if y satisfies constraints (1.54a) and (1.54b), and $\bar{z}$ is un- stable, then y must stabilise the zero dynamics by driving z to $\mathcal{M}_{\bar{z}}$. Thus the following lemma holds.

Lemma 6. *Consider the system described by (1.52). Suppose that assumptions above are satisfied, $\bar{y} > 0$ and y_1 satisfies constraint (1.54b). Then, the following statements hold*

- *If the open set S_u is unreachable for all $y(t) \in \mathcal{Y}_0$ (and for all $t > 0$), and $\bar{z} \in S_u$, then y_1 must undershoot.*
- *If $T_{es}(y_1) = T$, and $\mathcal{M}_{\bar{z}}$ is unreachable at $t = T$ for all $y(t) \in \mathcal{Y}_\gamma$, then $r_{us}(y_1) \geq \gamma / \bar{y}$.*

For a scalar zero dynamics case, i.e.,

$$\dot{z} = F_0(z, y) = f_0(z) + g_0(z) y, \quad z(0) = 0 \tag{1.55}$$

where $z \in \mathbb{R}$, $f_0(z)$ is continuous and increasing ($df_0/dz > 0$ almost everywhere), $f_0(0)$, and $g_0(z)$ has positive sign for all z, without loss of generality. Note that the conditions on f_0 ensure that the system satisfies the above assumptions. Suppose that y is required to track a step of height $\bar{y} > 0$. Let the corresponding equilibrium point be $\bar{z}$. We have $\bar{z} > 0$ because $f(\bar{z}) = -g(\bar{z})\bar{y} < 0$. $\bar{z}$ is also anti-stable because f_0 is an increasing function. Hence $y(t)$ must drive z to $\bar{z}$. Then we have the following.

Lemma 7. *Consider the previous system. Suppose that $y(t) \in \mathcal{Y}_\gamma$ and let $z_\gamma(t)$ be the solution to initial value problem (1.55) with $y(t) = -\gamma$, $z(t) > z_\gamma$.*

Suppose that $y \in \mathcal{Y}_0$. When $\gamma = 0$, $z_\gamma(t) = 0$. Thus, have from the proposition, $z(t) \geq 0$ for all t. But then $\bar{z}$ is unreachable, and so y must undershoot.

1.4.2 Cheap control problem analysis

In Seron et al. (1999) the authors extend the results obtained in Qiu and Davison (1993) to the case of nonlinear systems. In particular, they consider a specific class of systems, i.e., square and with a global unitary relative degree so that the system dynamics can be represented in the following normal form

$$\begin{aligned} \dot{y} &= f(y,z) + g(y,z)u, \qquad && y, u \in \mathbb{R}^{n_u} \\ \dot{z} &= f_0(z) + g_0(z)y, && z \in \mathbb{R}^{n_x - n_u} \end{aligned} \tag{1.56}$$

where $f(0,0) = 0$ and $f_0(0) = 0$. In these coordinates, the system zero dynamics is given by $\dot{z} = f_0(z)$, as introduced in Isidori (2013). They assume that

Assumption 1.4.1. *There exists a $\gamma > 0$ such that the smallest singular value of $g(y,z)$ is greater than or equal to γ for all y and z.*

For such a system, the cheap optimal control problem consists of finding a feedback control u which guarantees asymptotic stability and minimizes the cost functional J_ϵ in (1.15) with $\epsilon > 0$ is small. The problem has a solution if there exists a positive semidefinite optimal value $V(y,z;\epsilon)$ satisfying the Hamilton-Jacobi-Bellman (HJB) equation

$$\frac{\partial V}{\partial y} f(y,z) + \frac{\partial V}{\partial z}[f_0(z) + g_0(z)y] + \frac{1}{2}y^T y - \frac{1}{2\epsilon}\frac{\partial V}{\partial y} g(y,z)^T g(y,z)\frac{\partial^T V}{\partial y} = 0, \quad V(0,0;\epsilon) = 0 \tag{1.57}$$

and such that the feedback control

$$u = -\frac{1}{\epsilon^2} g^T(y,z)\frac{\partial^T V}{\partial y} \tag{1.58}$$

asymptotically stabilizes (1.56). Due to the singularity that raises in the HJB equation (1.57) for $\epsilon \to 0$, they propose to solve the equation by seeking a solution in the form

$$V(y,z;\epsilon) = V_0(z) + \epsilon V_1(y,z) + O(\epsilon^2) \tag{1.59}$$

thus, obtaining

$$\frac{\partial V_0}{\partial z}[f_0(z) + g_0(z)y] + \frac{1}{2}y^T y - \frac{1}{2}\frac{\partial V_1}{\partial y} g(y,z)^T g(y,z)\frac{\partial^T V_1}{\partial y} + O(\epsilon^2) = 0. \tag{1.60}$$

Assume that

Assumption 1.4.2. *the zero dynamics is antistable, i.e., $\dot{z} = -f_0(z)$ is asymptotically stable, and there exists a positive definite function $V_0(z)$ satisfying the HJB equation*

$$\frac{\partial V_0(z)}{\partial z} f_0(z) - \frac{1}{2}\frac{\partial V_0(z)}{\partial z} g_0(z) g_0^T(z)\frac{\partial^T V_0}{\partial z} = 0, \quad V_0(0) = 0 \tag{1.61}$$

such that the feedback control

$$y^\star = -g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} \tag{1.62}$$

achieves global asymptotic stability of the system (1.56) zero dynamics.

This is associated with the optimal control problem

$$\begin{aligned} &\min_y \int_0^\infty z^T z + y^T y dt \\ &\quad \text{subj to } \dot{z} = f_0(z) + g_0(z)y. \end{aligned} \tag{1.63}$$

Considering such analysis in the cheap control problem on the initial system we can slightly modify the HJB equation in (1.57) by adding and subtracting $y^{\star T} y^\star/2$ we can write

$$\frac{1}{2}(y - y^\star)^T(y - y^\star) - \frac{1}{2}\frac{\partial V_1(z,y)}{\partial y} g(y,z) g^T(y,z)\frac{\partial^T V_1}{\partial y} = O(\epsilon). \tag{1.64}$$

Hence, by defining the output transitory behaviour as $\tilde{y} = y - y^\star$ and letting $\epsilon \to 0$, the HJB becomes

$$\frac{1}{2}\tilde{y}^T\tilde{y} - \frac{1}{2}\frac{\partial V_1(\tilde{y} + y^\star, z)}{\partial \tilde{y}} g(\tilde{y} + y^\star, z) g^T(\tilde{y} + y^\star, z)\frac{\partial^T V_1(\tilde{y} + y^\star, , z)}{\partial \tilde{y}} = 0. \tag{1.65}$$

Thus the solution of the original cheap control corresponds to

$$\begin{aligned} &\min_u \int_0^\infty \tilde{y}^T(t)\tilde{y}(t) + \epsilon^2 u^T(t)u(t)dt \\ &\text{subj to } \dot{\tilde{y}} = g(\tilde{y} + y^\star(z), z)u. \end{aligned} \tag{1.66}$$

with z being a constant. The optimal value for this problem is given by $\epsilon V_1(\tilde{y} + y^\star(z), z)$ and the optimal control law is given by

$$u = \frac{1}{\epsilon} g^T(\tilde{y} + y^\star(z), z)\frac{\partial^T V_1}{\partial \tilde{y}} = \frac{1}{\epsilon} g^T (gg^T)^{-\frac{1}{2}}\tilde{y}. \tag{1.67}$$

Via a singular perturbation analysis, we can write the system in the new coordinates $(\tilde{y}, z)$

$$\begin{aligned} \dot{\tilde{y}} &= \dot{y} - \dot{y}^\star(z) \\ &= f(\tilde{y} + y^\star, z) - \frac{1}{\epsilon}(gg^T)^{\frac{1}{2}}\tilde{y} - \frac{\partial y^\star(z)}{\partial z}\left[f_0(z) - g_0(z)g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} + g_0(z)\tilde{y}\right] \\ \dot{z} &= f_0(z) - g_0(z)g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} + g_0(z)\tilde{y} \end{aligned} \tag{1.68}$$

that can be easily put in singular perturbation form

$$\begin{aligned} \epsilon\dot{\tilde{y}} &= -(gg^T)^{\frac{1}{2}}\tilde{y} + \epsilon f(\tilde{y} + y^\star, z) - \epsilon\frac{\partial y^\star(z)}{\partial z}\left[f_0(z) - g_0(z)g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} + g_0(z)\tilde{y}\right] \\ \dot{z} &= f_0(z) - g_0(z)g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} + g_0(z)\tilde{y} \end{aligned} \tag{1.69}$$

from which we can find the boundary layer of the $\tilde{y}$ obtained by considering $\epsilon = 0$ in the first equation of (1.69), yielding $\tilde{y} = 0$, and thus the system dynamics reduces to the slow subsystem

$$\dot{z} = f_0(z) - g_0(z)g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} \tag{1.70}$$

which is globally asymptotically stable by the properties of $V_0(z)$. We can analyze the convergence properties to the boundary layer by considering a change of the time scale defining $\tau = \epsilon^{-1}t$

$$\frac{d\tilde{y}}{d\tau} = -(gg^T)^{\frac{1}{2}}\tilde{y} + \epsilon f(\tilde{y} + y^\star, z) - \epsilon\frac{\partial y^\star(z)}{\partial z}\left[f_0(z) - g_0(z)g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} + g_0(z)\tilde{y}\right] \tag{1.71}$$

that, for ϵ sufficiently small, can be approximated to

$$\frac{d\tilde{y}}{d\tau} = -(gg^T)^{\frac{1}{2}}\tilde{y} \tag{1.72}$$

which is globally exponentially stable by the properties of the function V_1.
To guarantee the asymptotic stability of the closed loop system, we need to introduce some assumptions on the interconnection with the zero dynamics in the $\tilde{y}$ one, i.e., the term

$$\phi(\tilde{y}, z) = f(\tilde{y} + y^\star, z) - \frac{\partial y^\star(z)}{\partial z}\left[f_0(z) - g_0(z)g_0^T(z)\frac{\partial^T V_0(z)}{\partial z} + g_0(z)\tilde{y}\right].$$

Lemma 8. *Assume $\|\phi(\tilde{y}, z)\| \leq k_1\|\tilde{y}\| + k_2\|y^\star(z)\|$, in $\mathcal{B}_\delta$ for some positive real k_1, k_2 and δ. Then, under Assumption 1.4.1, to each $R > 0$ there corresponds an $\epsilon_R > 0$ such that for all $\epsilon \in (0, \epsilon_R]$ the equilibrium $(\tilde{y}, z) = (0, 0)$ of (1.69) is asymptotically stable and its basin of attraction contains $\mathcal{B}_R$*

Limitations to Nonlinear Ideal Performances

The analysis above has decomposed the optimal cheap control problem into two separate subproblems: a minimum energy problem for asymptotic stabilization of the system zero dynamics and a cheap control problem for asymptotic stabilization of the boundary-layer subsystem (1.71). For $\epsilon \to 0$ the $\tilde{y}$ dynamics in (1.69) tends to zero instantaneously and the solution of (1.69) settle on the slow invariant manifold in which the z dynamics is controlled by the stabilizing action $y^\star(z)$. Due to its definition, the cost of $\tilde{y}$ stabilization, $\epsilon V_1(\tilde{y} + y^\star, z)$, decreases as $\epsilon \to 0$. What remains is the cost $V_0(z)$ of the zero dynamics stabilization. Hence, the overall stabilization cost cannot be reduced below $V_0(z)$ as described in the following Theorem.

Theorem 1.4.1. *Under the same conditions of Lemma 8, for every initial condition* $(y(0), z(0))$ *of* (1.56) *for which the cheap control problem* (1.15) *has a solution, the optimal value satisfies*

$$J_\epsilon^\star = V(y(0), z(0); \epsilon) = V_0(z(0)) + \mathcal{O}(\epsilon) \tag{1.73}$$

and thus the ideal performance is $V_0(z(0))$*, the optimal value of the minimum energy problem for the zero-dynamics of* (1.56) *controlled by the output* y*.*

Thus, the lowest attainable L_2 norm of the output of (1.56) is the least amount of energy required to stabilize the unstable zero dynamics.
It is also possible to find a nonlinear analogue of the property that, as $\epsilon \to 0$, the finite poles of the optimal linear system converge to the mirror image of the NMP zeros. An expression of this property is that the zero dynamics are 'as stable in the closed loop as they are unstable in the open loop'. Using $V_0(z)$ as a Lyapunov function we reveal an analogous property for the nonlinear zero dynamics in

Corollary 1.4.1. *For the open and closed loop zero dynamics,* $V_0(z)$ *satisfies*

$$\frac{\partial V_0}{\partial z} f_0(z) = -\frac{\partial V_0}{\partial z}\left[f_0(z) + g_0(z) y^\star(z)\right].$$

Which is immediate from the Hamilton-Jacobi-Bellman equation (1.61).

1.4.3 Path-following as an alternative to output tracking

In Aguiar et al. (2008) the authors extend their previous results Aguiar et al. (2005) for path following linear case, to nonlinear systems by developing a local analysis on the system behaviour and considering a linear exosystem dynamics.
The class of system under consideration is one of the nonlinear square systems which are locally diffeomorphic to systems in the strict-feedback form

$$\begin{aligned}
\dot{z} &= f_0(z) + g_0(z)\xi_1 \\
\dot{\xi}_1 &= f_1(z, \xi_1) + g_1(z, \xi_1)\xi_2 \\
&\cdots \\
\dot{\xi}_r &= f_r(z, \xi_1, \ldots, \xi_r) + g_r(z, \xi_1, \ldots, \xi_r)\xi_2 \\
y &= \xi_1
\end{aligned}$$

with $\xi_i \in \mathbb{R}^{n_u}$, for $i = 1, \ldots, r$, $z \in \mathbb{R}^{n_x - rm}$, $u \in \mathbb{R}^{n_u}$. $f_i(\cdot)$ and $g_i(\cdot)$ are C^k functions of their arguments (for some large k), $f_i(0, \ldots, 0) = 0$ and the matrices $g_i(\cdots)$, $i = 1, \ldots, r$, are always non-singular. And moreover we assume the system is at rest, i.e., $(z(0), \xi(0)) = (0, 0)$.
The nonlinear reference tracking problem considers a reference signal $r \in \mathbb{R}^{n_y}$ generated by a known exosystem

$$\begin{aligned}
\dot{w} &= s(w), \ s(0) = 0 \\
r &= q(w)
\end{aligned} \tag{1.74}$$

and the problem refers to finding a feedback controller such that the closed-loop system is asymptotically stable and the output y converges to r. Isidori and Byrnes (1990) show that for a system of the form

$$\dot{x} = f(x, u), \ y = h(x, u)$$

the problem is solvable if and only if there exists smooth maps $\Pi(w)$ and $c(w)$ satisfying

$$\begin{aligned}
&\frac{\partial \Pi}{\partial w} s(w) - f(w, c(w)) = 0, \ \Pi(0) = 0 \\
&h(\Pi(w), c(w)) - q(w) = 0, \ c(0) = 0.
\end{aligned} \tag{1.75}$$

The necessary and sufficient conditions, specialized for system (1.4.3), hence the reference-tracking problem is solvable, if and only if there exists maps $\Pi = \operatorname{col}(\Pi_0, \Pi_1, \ldots, \Pi_r)$, $\Pi_0 : \mathbb{R}^{n_w} \to R^{n_x - rm}$, $\Pi_i : \mathbb{R}^{n_w} \to \mathbb{R}^{n_y}$, $i = 1, \ldots, r$ and $c : \mathbb{R}^{n_y} \to \mathbb{R}^{n_u}$ that satisfy (1.75). Consider the following locally diffeomorphic change of coordinates

$$\tilde{z} = z - \Pi_0(w) \tag{1.76a}$$

$$\tilde{\xi}_i = \xi_i - \Pi_i(w), \ i = 1, \ldots, r \tag{1.76b}$$

$$\tilde{u} = u - c(w) \tag{1.76c}$$

transforms system (1.4.3) into the error system

$$\begin{aligned}
\dot{z} &= \tilde{f}_0(\tilde{z}, w) + \tilde{g}_0(\tilde{z}, w)\tilde{\xi}_1 \\
\dot{\tilde{\xi}}_1 &= \tilde{f}_1(\tilde{z}, \tilde{\xi}_1, w) + \tilde{g}_1(\tilde{z}, \tilde{\xi}_1, w)\tilde{\xi}_2 \\
&\dots \\
\dot{\tilde{\xi}}_r &= \tilde{f}_r(\tilde{z}, \tilde{\xi}_1, \dots, \tilde{\xi}_r) + \tilde{g}_r(\tilde{z}, \tilde{\xi}_1, \dots, \tilde{\xi}_r)\tilde{\xi}_2 \\
e &= \tilde{\xi}_1
\end{aligned} \tag{1.77}$$

where $\tilde{f}_i$, $\tilde{g}_i$, $i = 0, 1, \dots, r$ are appropriately defined functions satisfying $\tilde{f}(0, w) = 0$, $\tilde{g}(\tilde{z}, 0) = g_0(\tilde{z})$, $\tilde{f}_i = (0, \dots, 0, w) = 0$ and $\tilde{g}_i(\tilde{z}, \dots, \tilde{\xi}_i, 0) = g_i(\tilde{z}, \dots, \tilde{\xi}_i)$.
Also in this case, as for the linear one in section 1.3.2, we consider two optimal control problems: cheap control and minimum energy problems.
Cheap control problem: For system consisting of the error system (1.77) and exosystem (1.74) with initial condition $(\tilde{z}(0), \tilde{\xi}(0), w(0)) = (\tilde{z}_0, \tilde{\xi}_0, w_0)$, find the optimal feedback law $\tilde{u} = \kappa(\tilde{z}, \tilde{\xi}, w)$ that minimizes the cost functional

$$J_{\delta,\epsilon} = \frac{1}{2}\int_0^\infty \left(\|e(t)\|^2 + \delta\|\tilde{z}(t)\|^2 + \epsilon^{2r}\|\tilde{u}(t)\|\right) dt \tag{1.78}$$

for $\delta, \epsilon > 0$. We denote by $J^\star_{\delta,\epsilon}(\tilde{z}_0, \tilde{\xi}_0, w_0)$ the corresponding optimal value. The best attainable cheap control performance for reference tracking is then

$$J_{\mathcal{T}} = \lim_{(\delta,\epsilon)\to 0} J_{\delta,\epsilon}(\tilde{z}_0, \tilde{\xi}_0, w_0).$$

In some neighborhood of the origin and for every $\delta, \epsilon > 0$, the value $J_{\delta,\epsilon}$ is C^{k-2} under the following assumption

Assumption 1.4.3. *The linearization around $(z, \xi) = (0, 0)$ of system (1.4.3) is stabilizable and detectable and the linearization around $w = 0$ of the exosystem dynamics (1.74) is stable.*

Minimum-energy problem: For system

$$\dot{\tilde{z}} = \tilde{f}_0(\tilde{z}, w) + \tilde{g}_0(\tilde{z}, w)e, \qquad \tilde{z}(0) = \tilde{z}_0 \tag{1.79a}$$
$$\dot{w} = s(w), \qquad w(0) = w_0 \tag{1.79b}$$

with e viewed as the input, find the optimal feedback law $e = \kappa_{e,\delta}(\tilde{z}, w)$ that minimizes the cost

$$J_{e,\delta} = \frac{1}{2}\int_0^\infty \left(\delta\|\tilde{z}(t)\|^2 + \|e(t)\|^2\right) dt$$

for $\delta > 0$. We denote by $J^\star_{e,\delta}(\tilde{z}_0, w_0)$ the corresponding optimal value. Under Assumption 1.4.3, $J^\star_{e,\delta}(\tilde{z}_0, w_0)$ is C^{k-2} in some neighborhood of $(0, 0)$.
Their analysis reveals that the best-attainable cheap control performance $J_{\mathcal{T}}$ is equal to the least control effort (i.e., as $\delta \to 0$) needed to stabilize the corresponding zero dynamics (1.79a) driven the tracking error e. The following theorem summarises the analysis.

Theorem 1.4.2. *Suppose that Assumption 1.4.3 holds and that the regulator equations (1.75) has a solution in some neighborhood of $w = 0$. Then, for any $(\tilde{z}(0), \tilde{\xi}(0), w(0)) = (\tilde{z}_0, \tilde{\xi}_0, w_0)$ in some neighborhood of $(0, 0, 0)$ there exists a solution to the cheap control problem and $J_{\mathcal{T}} = \lim_{\delta\to 0} J_{e,\delta}$.*

They show that the pah-following case can be solved with arbitrarily small $\mathcal{L}_2$ norm of e. For the path following analysis, we define the corresponding cheap control problem by replacing the definition of the exosystem dynamics and, as a consequence, of the regulator equations. They focus their attention on linear exosystem dynamics, in particular, (1.74) are defined as a geometric path parameterized in *theta* along with the timing law

$$\dot{\theta}(t) = v_d, \qquad \theta(0) = \theta_0 \tag{1.80a}$$
$$\frac{\partial w}{\partial \theta}(\theta) = Sw(\theta), \qquad w(0) = w_0 \tag{1.80b}$$
$$r(t) = Qw(\theta(t)) \tag{1.80c}$$

equivalently resulting in

$$\dot{w}(t) = v_d S w(t), \quad r(t) = Qw(t) \tag{1.81}$$

where v_d will be properly defined according to the desired optimal cost. Consequently, the regulator equations are

$$\begin{aligned} v_d \frac{\partial \Pi}{\partial w} S w - f(\Pi(w), c(w)) &= 0 \\ h(\Pi(w), c(w)) - Q w &= 0. \end{aligned} \tag{1.82}$$

The result is summarized in the next Theorem.

Theorem 1.4.3. *Assume that (1.82) has a solution for almost all v_d in $(0, \infty)$. Then, for every $w(0) = w_0$ in a neighborhood around $w = 0$, there exist a timing law (1.80a) for $\theta(t)$ and a feedback law*

$$u = c(w) + \kappa_{e,\delta}(z, \xi, w) \tag{1.83}$$

which solves the geometric path following problem, i.e., the $e(t) = y(t) - r(t)$ goes to zero while the system state is bounded, by satisfying

$$\int_0^\infty \|e(t)\|^2 dt \leq \delta^\star. \tag{1.84}$$

Proof. We sketch the proof, since $J^\star_{\delta,\epsilon} = \lim_{\delta \to 0} J_{e,\delta}$, it can be shown that $J_{e,\delta}$ is bounded by

$$J_{e,\delta} \leq \frac{1}{2} \tilde{z}_0^T P_0 \tilde{z}_0$$

where P_0 is positive definite and does not depend on v_d. Observing that $\tilde{z}_0 = \Pi_0(w_0)$, since $z(0) = 0$, and that $\|\Pi_0(w_0)\|$ can be made arbitrarily small by choosing a sufficiently large v_d. Hence, $\delta^\star$ in (1.84) can be taken arbitrarily small. □

Moreover, an arbitrarily small $\mathcal{L}_2$ norm of the path-following error is attainable even when the speed v_d is specified beforehand.

Theorem 1.4.4. *Consider v_d to be specified so that (1.82) has a solution in some neighborhood of $w = 0$. Then,*

$$\int_0^\infty \|e(t)\|^2 dt \leq \delta^\star$$

can be satisfied for any $\delta^\star > 0$ with a suitable timing law $\theta(t)$ and a control law $u = c(w) + \kappa(z, \xi, w)$ with time-varying piecewise-continuous maps $c(w)$ and $\kappa(z, \xi, w)$.

Chapter 2

Stabilization of Non-minimum Phase Linear Systems via Inner-Outer Decomposition

The general idea behind the Inner-Outer decomposition is to describe a non-minimum phase system into a minimum phase system (the outer factor) driving a non-minimum phase one (the inner factor). The latter dynamics is stable but contains all the ugly non-minimum phase characteristics of the plant. Moreover, by stabilizing the outer system, the inner trajectories stay bounded and eventually stabilize by themselves. By exploiting this property, one can ideally think to feedback the output of the outer system $y_{\mathrm{o}}(t)$, or the whole outer state x_{o} as depicted in Fig.2.1, and exploit any of the available tools from the literate to design a controller for the outer system (which is minimum phase) and solve the stabilization problem of a non-minimum phase system. By exploiting this Inner-Outer decomposition, we can analyze

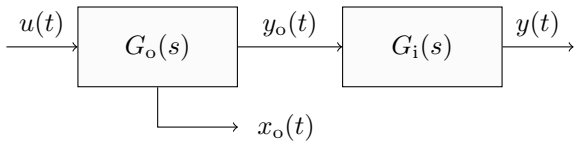

Figure 2.1: Inner-Outer decomposition Cascade system

the feedback control characteristic to make the output trajectory of a non-minimum phase system, $y(t)$, to be arbitrarily close to the output trajectory of its minimum phase 'twin 'system (the outer factor, up to a sign change).
In this chapter, we describe how to obtain an Inner-Outer decomposition for strictly linear systems and how to stabilize a controllable and observable linear non-minimum phase systems.

2.1 Preliminaries

2.1.1 Modal subspaces: stable and unstable eigenspaces

We give the definition of Modal Subspaces for a n-dimensional matrix as introduced in Francis (1987)[Ch.7].
Considering a nonsingular matrix A in $\mathbb{R}^{n\times n}$ (with no eigenvalues on the imaginary axis) and let $p_A(\lambda)$ be the characteristic polynomial of A factorized as $p_A = p_{A_-} \cdot p_{A_+}$, where the p_{A_-} (p_{A_+}) has all its zeros s with $\mathfrak{Re}\{s\} < 0$ ($\mathfrak{Re}\{s\} > 0$). We call the modal subspaces of $\mathbb{R}^n$ relative to A are

$$\begin{aligned}\mathbf{X}_-(A) &= \ker p_{A_-}(A)\\ \mathbf{X}_+(A) &= \ker p_{A_+}(A).\end{aligned}$$

It can be shown that $\mathbf{X}_-(A)$ ($\mathbf{X}_+(A)$) is also spanned by the generalized real eigenvectors v_- (v_+) of A associated to the eigenvalues λ_- (λ_+) with negative (positive) real part, i.e. $\mathfrak{Re}\{\lambda\}_- < 0$ ($\mathfrak{Re}\{\lambda\}_+ > 0$). The two modal subspaces are complementary, not generally orthogonal, that is they are independent and their direct sum is equivalent to the whole $\mathbb{R}^n$

$$\mathbb{R}^n = \mathbf{X}_-(A) \oplus \mathbf{X}_+(A).$$

The modal subspaces are the stable and unstable eigenspaces of A, i.e., the spaces spanned by the stable and unstable eigenvectors associated with the stable and unstable eigenvalues of A.

2.1.2 Modal subspaces of Hamiltonian matrix

A $2n$-dimensional Hamiltonian matrix is a matrix of the form

$$HM = \begin{bmatrix} A & -R \\ -Q & -A^T \end{bmatrix} \tag{2.1}$$

where A, Q, R are $n \times n$ matrices, in particular, Q and R are symmetric. It is also very well-known that HM and $-HM^T$ are similar matrices, i.e. for each eigenvalue of HM, λ, also $-\lambda$ is a HM eigenvalue, see also Isidori (2017)[A.6]. Clearly, the stable and unstable eigenspaces of HM live in $\mathbb{R}^{2n}$. To the Hamiltonian Matrix, HM, we associate the Algebraic Riccati Equation (ARE)

$$A^T\mathcal{P} + \mathcal{P}A - \mathcal{P}R\mathcal{P} + Q = 0. \tag{2.2}$$

In case $R = 0$ and A Hurwitz, the Algebraic Riccati Equation (2.2) is a Lyapunov equation

$$A^T\mathcal{P} + \mathcal{P}A + Q = 0 \tag{2.3}$$

has a unique, positive solution $\mathcal{P}$. We now recall some theorems from Francis (1987)[Ch.7].

Lemma 9. *Let A be Hurwitz and $R = 0$ in HM, then the stable and unstable eigenspaces relative to HM are*

$$\begin{aligned}\boldsymbol{X}_-(HM) &= span\begin{pmatrix}0\\ I\end{pmatrix}\\ \boldsymbol{X}_+(HM) &= span\begin{pmatrix}I\\ \mathcal{P}\end{pmatrix}\end{aligned}$$

where $\mathcal{P}$ is the solution of (2.3).

In case $R \neq 0$ and HM has no eigenvalues on the imaginary axis, the associated ARE (2.2), because HM is antisymmetric, it must have two n-dimensional modal subspaces. The following theorem holds

Theorem 2.1.1. *Assuming HM has no eigenvalues on the imaginary axis and (A, R) is a stabilizable pair, then the stable modal subspace $\boldsymbol{X}_-(HM)$ is complementary to the span $\begin{pmatrix}0 & I\end{pmatrix}^T$.*
Moreover, there exists a unique matrix $\mathcal{P}$ such that

$$\boldsymbol{X}_-(HM) = span\begin{pmatrix}I\\ \mathcal{P}\end{pmatrix}.$$

Such an $\mathcal{P}$ results to be symmetric, i.e. $\mathcal{P} = \mathcal{P}^T$, and a stabilizing solution of the ARE (2.2), i.e. $A - R\mathcal{P}$ is Hurwitz.

A proof of these theorems can be found in Francis (1987)[Ch.7] and Isidori (2017)[A.6].

2.1.3 Spectral factorization and Inner-Outer decomposition for proper Linear Time Invariant systems

We now introduce the standard Inner-Outer decomposition for Linear Time Invariant (LTI) proper systems as presented in Chen and Francis (1989) and exploited in Qiu and Davison (1993). This allows to connect the concepts introduced in section 2.1.1
Given a LTI proper system with minimal realization (A, B, C, D), with transfer matrix $G(s) = C\left(sI - A\right)^{-1} B + D$, by defining $G^\star(s)$ its adjoint system, i.e. $G^\star(s) = G^T(-s)$, to do the spectral factorization of $G(s)$ corresponds to find a transfer matrix $G_{\mathrm{o}}(s)$ such that $G^\star(s)G(s) = G_{\mathrm{o}}^\star(s)G_{\mathrm{o}}(s)$.
We call a matrix $G(s)$ *inner* if A transfer matrix $G(s)$ is said to be *outer* if it is minimum phase and *wide* (i.e., if it has more columns than rows), while $G(s)$ is said to be *inner* if it is stable, *tall* (i.e., its transpose is wide), $G^T(-s)G(s) = I$, $\forall s \in \mathbb{C}$, and its zeros are all in $\mathbb{C}_+$. Moreover, an inner transfer matrix is by definition an all-pass system and thus proper.

Lemma 10 (Lemma 2 in Qiu and Davison (1993))**.** *Given a system with minimal realization (A, B, C, D) and its associated transfer matrix*

$$G(s) = C(sI - A)^{-1}B + D,$$

it can always be factorized as $G(s) = G_i(s)G_o(s)$ such that $G_i(s)$ is an inner matrix and $G_o(s)$ is minimum phase and right invertible, thus outer. Moreover, all unstable poles of $G(s)$ are poles of $G_o(s)$.

In particular, finding the outer factor $G_{\mathrm{o}}(s)$ is equivalent to do the spectral factorization for $G(s)$.
The standard state space approach to find the *inner* and *outer* factors starts by considering the realization of the cascade $G^\star(s)G(s)$, i.e. $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$, where

$$\tilde{A} = \begin{bmatrix} A & 0 \\ -C^TC & -A^T \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B \\ -C^TD \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} D^TC & B^T \end{bmatrix}, \quad \tilde{D} = \begin{bmatrix} D^TD \end{bmatrix}.$$

assuming no zeros on the imaginary axis and $\tilde{D}$ nonsingular. We relate the matrices realization of the cascade $G^\star(s)G(s)$ to the Hamiltonian Matrix

$$H\tilde{M} = \tilde{A} - \tilde{B}\tilde{D}^{-1}\tilde{C} = \begin{bmatrix} A - B(D^TD)^{-1}D^TC & -B(D^TD)^{-1}B^T \\ -C^T\left[I - D(D^TD)^{-1}D^T\right]C & -A^T + C^TD(D^TD)^{-1}B^T \end{bmatrix}.$$

Lemma 11 (Lemma 1 in Chen and Francis (1989))**.** *Assuming $G(s)$ has no transmission zeros on the imaginary axis and $\tilde{D}$ nonsingular, then there exists a unique symmetric matrix $\mathcal{P}$ such that the stable eigenspace of $H\tilde{M}$, $\boldsymbol{X}_-(H\tilde{M}) = span\begin{pmatrix} I & \mathcal{P} \end{pmatrix}^T$.*
Moreover, the spectral factor $G_{\mathrm{o}}(s)$ of $G(s)$ with realization $(A, B, C_{\mathrm{o}}, D_{\mathrm{o}})$ where

$$C_{\mathrm{o}} = (D^TD)^{-\frac{1}{2}}\left(D^TC + B^T\mathcal{P}\right), \quad D_{\mathrm{o}} = (D^TD)^{\frac{1}{2}}$$

Thus we can conclude, via the respective Hamiltonian matrices, that $G^\star(s)G(s) = G_{\mathrm{o}}^\star(s)G_{\mathrm{o}}(s)$. Hence, by computing the stabilizing solution $\mathcal{P}$ of the associated ARE (2.2), we obtained the outer factor $G_{\mathrm{o}}(s)$ of the $G(s)$ decomposition and we can simple define the inner factor $G_{\mathrm{i}}(s) := G(s)G_{\mathrm{o}}^{-1}(s)$, where $G_{\mathrm{o}}^{-1}(s)$ is the right inverse of the outer factor $G_{\mathrm{o}}(s)$. The realization of the inner factor is $\left(A + BK, B\left(D^TD\right)^{-\frac{1}{2}}, C + DK, D\left(D^TD\right)^{-\frac{1}{2}}\right)$, with $K = -\left(D^TD\right)^{-1}\left(D^TC + B^T\mathcal{P}\right)$.
In Chen and Francis (1989), is presented also the case in which the $\tilde{D}$ is not invertible, we report here the main result of the work

Lemma 12 (Th.1 in Chen and Francis (1989))**.** *Assuming $G(s)$ has no transmission zeros on the imaginary axis and D^TD singular, define $E = D^T\left(DD^T\right)^{-2}D$ and the system associated ARE*

$$\left(A - BED^TC\right)^T\mathcal{P} + \mathcal{P}\left(A - BED^TC\right) - \mathcal{P}BEB^T\mathcal{P} = 0. \tag{2.4}$$

Then, the realization of outer factor $G_{\mathrm{o}}(s)$ is $(A, B, C_{\mathrm{o}}, D)$, with $C_{\mathrm{o}} = C + DEB^T\mathcal{P}$, and $G_{\mathrm{i}}(s) = G(s)G_{\mathrm{o}}^+(s)$, with $G_{\mathrm{o}}^+(s)$ being the right-inverse of $G_{\mathrm{o}}(s)$, where $\mathcal{P}$ is the stabilizing solution of (2.4), i.e. $(A - BE(D^TC + B^T\mathcal{P}), B(I - D^T(DD^T)^{-1}D))$ is stabilizable pair.

The proof of this lemma can be found in Chen and Francis (1989)[Sec.3].

Simple example

To simply get the idea of the Inner-Outer decomposition, consider the SISO non-minimum phase system

$$G(s) = \frac{s^2 - 2s}{(s+1)^3} \tag{2.5}$$

In order to obtain the outer factor we have to construct a transfer function $G_{\mathrm{o}}(s)$ that is minimum phase such that $G_{\mathrm{o}}^{\star}(s)G_{\mathrm{o}}(s) = G^{\star}(s)G(s)$, i.e.

$$G^{\star}(s)G(s) = \frac{-s(-s-2)}{(-s+1)^3}\frac{s(s-2)}{(s+1)^3}. \tag{2.6}$$

The outer transfer function has the same poles as the initial system, thus it is

$$G_{\mathrm{o}}(s) = \frac{s^2 + 2s}{(s+1)^3}. \tag{2.7}$$

We can simply obtain the inner system transfer function by considering

$$G_{\mathrm{i}}(s) = \frac{s-2}{s+2}. \tag{2.8}$$

Thus, we can write $G(s) = G_{\mathrm{i}}(s)G_{\mathrm{o}}(s)$

$$\frac{s^2 - 2s}{(s+1)^3} = \frac{s-2}{s+2} \cdot \frac{s(s+2)}{(s+1)^3} \tag{2.9}$$

2.2 The Inner-Outer decomposition for strictly proper systems

While in section 2.1.3 we studied the existence of the Inner-Outer factorization for general proper transfer matrices. We now extend the Inner-Outer decomposition in the state space to the case of strictly proper systems. And as a first result of this thesis, we propose an alternative (closed) form solution of the decomposition for strictly proper of the Inner-Outer factors systems matrices realization.
We introduce the Inner-Outer factorization by first describing the properties of inner and outer systems.

Definition 1 (Inner system, Qiu and Davison (1993))**.** *A system with minimal realization (A, B, C, D) is said to be* inner *if A is Hurwitz, $n_y \geq n_u$, $\sigma(A - BD^{-1}C) = \sigma(-A)$ (thus all zeros are in $\mathbb{C}_+$ and mirror the eigenvalues of A).*

Moreover, by defining the controllability and observability Grammian as the solution of the following Lyapunov equations

$$\begin{aligned} A\mathcal{G}_c + \mathcal{G}_c A^T &= -BB^T \\ A^T\mathcal{G}_o + \mathcal{G}_o A &= -C^T C \end{aligned} \tag{2.10}$$

for a balanced realization (as introduced in Glover (1984)) of a inner system (A, B, C, D), see Glover (1984) and Qiu and Davison (1993), we have

- $\mathcal{G}_o = \mathcal{G}_c$
- $D^T D = I$
- $D^T C + B^T = 0$ and $DB^T + C = 0$.

Definition 2 (Outer system)**.** *A system (A, B, C) is said to be* outer *if $n_y \leq n_u$, and all its zeros are in $\mathbb{C}_-$.*

For the sake of completeness, we also recall the definition of inner and outer systems introduced in Francis (1987)[Ch. 7] and Chen and Francis (1989) involving transfer matrices. A transfer matrix $G(s)$ is said to be *outer* if it is *wide* (i.e., if it has more columns than rows) and right-invertible (hence, its right inverse is analytic in $\mathbb{C}_+$) hence minimum phase, while $G(s)$ is said to be *inner* if it is stable, *tall* (i.e., its transpose is wide), its zeros are all in $\mathbb{C}_+$, and $G^T(-s)G(s) = I$, $\forall s \in \mathbb{C}$. Hence, an inner transfer matrix is by definition an all-pass system and thus proper.
The standard Inner-Outer factorization problem considered in the literature as only be described in the frequency domain case and reads as follows. Given an arbitrary transfer matrix $G(s)$, determine two transfer matrices $G_{\mathrm{i}}(s)$ and $G_{\mathrm{o}}(s)$ such that $G(s) = G_{\mathrm{i}}(s)G_{\mathrm{o}}(s)$ where $G_{\mathrm{i}}(s)$ is inner and $G_{\mathrm{o}}(s)$ is outer. $G(s) = G_{\mathrm{i}}(s)G_{\mathrm{o}}(s)$ defines an Inner-Outer factorization of $G(s)$.

Lemma 13 (Lemma 2 in Qiu and Davison (1993))**.** *Consider a system with minimal realization (A, B, C, D). Its associated transfer matrix $G(s) = C(sI - A)^{-1}B + D$ can always be factorized as $G(s) = G_i(s)G_o(s)$ such that $G_i(s)$ and $G_o(s)$ are respectively an inner and outer transfer matrix. Moreover, all unstable poles of $G(s)$ are poles of $G_o(s)$.*

Lemma 13 guarantees the existence of the Inner-Outer factorization for general proper transfer matrices. We extend this I/O result by first presenting an Inner-Outer decomposition in the state space to the case of strictly proper systems.

Remark 1. *Note that by the properties of the inner factor, we can exploit and solve the spectral factorization problem to determine the outer factor of the system and then compute the inner factor. Usually, this last term is obtained by inverting the outer factor and post-multiply the initial system by the outer right inverse, i.e., $G_{\mathrm{i}}(s) = G(s)G_{\mathrm{o}}^{+}(s)$, being $G_{\mathrm{o}}^{+}(s)$ right inverse of $G_{\mathrm{o}}(s)$.*

In this work, we provide an analysis and description of the spectral factorization and Inner-Outer decomposition problems for strictly problem systems in the state representation.
In particular, we consider a linear multi-input multi-output system (A, B, C)

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{2.11}$$

without zeros on the imaginary axis and its associated normal form realization

$$\begin{aligned} T_{\mathrm{nf}}\, A\, T_{\mathrm{nf}}^{-1} = \begin{bmatrix} F & G \\ H & \bar{A} \end{bmatrix} \quad T_{\mathrm{nf}} B = \begin{bmatrix} 0 \\ \bar{B} \end{bmatrix} \\ C\, T_{\mathrm{nf}}^{-1} = \begin{bmatrix} 0 & \bar{C} \end{bmatrix}. \end{aligned} \tag{2.12}$$

We construct a second system of the form

$$\begin{aligned} \dot{x}_{\mathrm{o}} &= A_{\mathrm{o}} x_{\mathrm{o}} + B_{\mathrm{o}} u \\ \dot{x}_{\mathrm{i}} &= A_{\mathrm{i}} x_{\mathrm{i}} + B_{\mathrm{i}} C_{\mathrm{o}} x_{\mathrm{o}} \\ y &= C_{\mathrm{i}} x_{\mathrm{i}} + D_{\mathrm{i}} C_{\mathrm{o}} x_{\mathrm{o}}. \end{aligned} \tag{2.13}$$

with $x_{\mathrm{i}} \in \mathbb{R}^{n_{\mathrm{i}}}$, $x_{\mathrm{o}} \in \mathbb{R}^{n_{\mathrm{o}}}$, $n_{\mathrm{i}}, n_{\mathrm{o}} \in \mathbb{N}$, $n_{\mathrm{i}} + n_{\mathrm{o}} > n_x$, such that

(i) System

$$\begin{aligned} \dot{x}_{\mathrm{i}} &= A_{\mathrm{i}} x_{\mathrm{i}} + B_{\mathrm{i}} u \\ y_{\mathrm{i}} &= C_{\mathrm{i}} x_{\mathrm{i}} + D_{\mathrm{i}} u; \end{aligned} \tag{2.14}$$

is inner;

(ii) System

$$\begin{aligned} \dot{x}_{\mathrm{o}} &= A_{\mathrm{o}} x_{\mathrm{o}} + B_{\mathrm{o}} u \\ y_{\mathrm{o}} &= C_{\mathrm{o}} x_{\mathrm{o}}. \end{aligned} \tag{2.15}$$

is outer.

(iii) The cascade (2.13) is input-output equivalent to (A, B, C), namely for all initial conditions $x(0)$ and for all u, the output trajectory of (2.13) with $(x_{\mathrm{i}}(0), x_{\mathrm{o}}(0)) = Tx(0)$ coincides with the output trajectory of (A, B, C).

To obtain the Inner-Outer matrices realization, let $\mathcal{P}$ be the stabilizing solution to the Algebraic Riccati Equation (ARE)

$$\mathcal{P}F + F^T\mathcal{P} + \mathcal{P}G\bar{C}^T\bar{C}G^T\mathcal{P} = 0, \tag{2.16}$$

whose existence and uniqueness are guaranteed by the absence of eigenvalues of F on the imaginary axis, and define

$$T_{\mathrm{o}} := \begin{bmatrix} I & 0 \\ -\bar{C}^T\bar{C}G^T\mathcal{P} & I \end{bmatrix}, \quad T_{\mathrm{i}} := \begin{bmatrix} I & 0 \end{bmatrix}. \tag{2.17}$$

Now, take $(A_{\mathrm{o}}, B_{\mathrm{o}}, C_{\mathrm{o}})$ in (2.13) as

$$\begin{aligned} A_{\mathrm{o}} &= \begin{bmatrix} F_{\mathrm{o}} & G_{\mathrm{o}} \\ H_{\mathrm{o}} & \bar{A}_{\mathrm{o}} \end{bmatrix} = T_{\mathrm{o}} \begin{bmatrix} F & G \\ H & \bar{A} \end{bmatrix} T_{\mathrm{o}}^{-1} \\ &= \begin{bmatrix} F + G\bar{C}^T\bar{C}G^T\mathcal{P} & G \\ H - G^T\mathcal{P}F_{\mathrm{o}} + \bar{A}\bar{C}^T\bar{C}G^T\mathcal{P} & \bar{A} - \bar{C}^T\bar{C}G^T\mathcal{P}G \end{bmatrix} \\ B_{\mathrm{o}} &= \begin{bmatrix} 0 \\ \bar{B} \end{bmatrix}, \quad C_{\mathrm{o}} := \begin{bmatrix} 0 & \bar{C} \end{bmatrix}, \end{aligned} \tag{2.18}$$

and define $(A_\mathrm{i}, B_\mathrm{i}, C_\mathrm{i}, D_\mathrm{i})$ in (2.13) as

$$\begin{aligned} A_\mathrm{i} &= F + G\bar{C}^T\bar{C}G^T\mathcal{P}, \qquad & B_\mathrm{i} &= G\bar{C}^T \\ C_\mathrm{i} &= \bar{C}G^T\mathcal{P}, & D_\mathrm{i} &= I. \end{aligned} \tag{2.19}$$

Note that $C_\mathrm{o} \neq CT_\mathrm{o}^{-1}$, that is the output $y_\mathrm{o} = C_\mathrm{o}\xi_\mathrm{o} \neq y = C\xi$. Moreover, note that, system (A, B, C) and (2.13) are related by the linear immersion map $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_\mathrm{o}+n_\mathrm{i}}$, defined as $T = \begin{bmatrix} T_\mathrm{i}^T & T_\mathrm{o}^T \end{bmatrix}^T T_\mathrm{nf}$. In particular, with $\bar{T}_\mathrm{i} = T_\mathrm{i}T_{nf}$ and $\bar{T}_\mathrm{o} = T_\mathrm{o}T_{nf}$, we have $x_\mathrm{i} = \bar{T}_\mathrm{i}x$ and $x_\mathrm{o} = \bar{T}_\mathrm{o}x$. Then, the following holds.

Theorem 2.2.1. *The realization of $(A_\mathrm{i}, B_\mathrm{i}, C_\mathrm{i}, D_\mathrm{i})$ in (2.19), and of (A_o, B_o, C_o) in (2.18) are respectively the inner and outer factors of a stabilizable realization of (A, B, C) and system (2.13) is input-output equivalent to (A, B, C).*

Proof. By the properties of the stabilizing solution $\mathcal{P}$ of (2.16), see Francis (1987), and because the pair $(F, G\bar{C}^T\bar{C}G^T)$ is stabilizable (since (A, B) is stabilizable), we have that

$$\sigma(F + G\bar{C}^T\bar{C}G^T\mathcal{P}) \subset \mathbb{C}^-. \tag{2.20}$$

In particular, $\mathcal{P}$ modifies only the eigenstructure (eigenvalues and relative eigenvectors) of the F matrix that corresponds to right half plane eigenvalues by mirroring their position into the left half plane, leaving untouched the stable part.
With these considerations, one can define an ideal output trajectory that stabilizes the zero dynamics, i.e., $\xi^\star = \bar{C}^T\bar{C}G^T\mathcal{P}z$. Thus, we define a change of coordinates for the initial system (2), $\xi_\mathrm{o} = \xi - \xi^\star$

$$\begin{aligned} \dot{\xi}_\mathrm{o} &= \dot{\xi} - \bar{C}^T\bar{C}G^T\mathcal{P}\dot{z} \\ &= Hz + \bar{A}\xi + \bar{B}u - \bar{C}^T\bar{C}G^T\mathcal{P}(Fz + G\xi) \\ &= \Big(H - \bar{C}^T\bar{C}G^T\mathcal{P}(F + G\bar{C}^T\bar{C}G^T\mathcal{P}) + \bar{A}\bar{C}^T\bar{C}G^T\mathcal{P}\Big)z + \Big(\bar{A} - \bar{C}^T\bar{C}G^T\mathcal{P}G\Big)\xi_\mathrm{o} + \bar{B}u \\ &= H_\mathrm{o}z + \bar{A}_\mathrm{o}\xi_\mathrm{o} + \bar{B}u\,. \end{aligned} \tag{2.21}$$

By this change of coordinates, with $z_\mathrm{o} = z$ and ξ_o as new states, the outer system is described as

$$\begin{aligned} \dot{z}_\mathrm{o} &= \big(F + G\bar{C}^T\bar{C}G^T\mathcal{P}\big)z_\mathrm{o} + G\xi_\mathrm{o} \\ &= F_\mathrm{o}z_\mathrm{o} + G_\mathrm{o}\xi_\mathrm{o} \\ \dot{\xi}_\mathrm{o} &= H_\mathrm{o}z_\mathrm{o} + \bar{A}_\mathrm{o}\xi_\mathrm{o} + \bar{B}_\mathrm{o}u \\ y_\mathrm{o} &= \bar{C}\xi_\mathrm{o}. \end{aligned} \tag{2.22}$$

Thus, the outer system $(A_\mathrm{o}, B_\mathrm{o}, C_\mathrm{o})$ has realization

$$\begin{aligned} A_\mathrm{o} &= \begin{bmatrix} F_\mathrm{o} & G_\mathrm{o} \\ H_\mathrm{o} & \bar{A}_\mathrm{o} \end{bmatrix} \\ &:= \begin{bmatrix} F + G\bar{C}^T\bar{C}G^T\mathcal{P} & G \\ H - \bar{C}^T\bar{C}G^T\mathcal{P}F_\mathrm{o} + \bar{A}\bar{C}^T\bar{C}G^T\mathcal{P} & \bar{A} - \bar{C}^T\bar{C}G^T\mathcal{P}G \end{bmatrix} \\ B_\mathrm{o} &= \begin{bmatrix} 0 \\ \bar{B} \end{bmatrix}, \quad C_\mathrm{o} = \begin{bmatrix} 0 & \bar{C} \end{bmatrix} \end{aligned} \tag{2.23}$$

We observe that A_o is similar to A via the linear map T_o. It is very easy to see that the system $(A_\mathrm{o}, B_\mathrm{o}, C_\mathrm{o})$ is minimum phase, and thus it is an outer factor for (A, B, C).
Now, we only need to construct the inner system

$$\begin{aligned} \dot{x}_\mathrm{i} &= A_\mathrm{i}x_\mathrm{i} + B_\mathrm{i}y_\mathrm{o} \\ y &= C_\mathrm{i}x_\mathrm{i} + D_\mathrm{i}y_\mathrm{o}\,. \end{aligned} \tag{2.24}$$

In this respect, we consider the change of coordinate $\xi_\mathrm{o} = \xi - \bar{C}^T\bar{C}G^T\mathcal{P}z_\mathrm{o}$ to obtain the real system output $y(t)$, by taking into account that $\bar{C}\bar{C}^T = I_p$,

$$\begin{aligned} y = \bar{C}\xi &= \bar{C}\xi_\mathrm{o} + \bar{C}G^T\mathcal{P}z_\mathrm{o} \\ &= \bar{C}G^T\mathcal{P}z_\mathrm{o} + y_\mathrm{o} \\ &= C_\mathrm{i}x_\mathrm{i} + D_\mathrm{i}y_\mathrm{o}\,. \end{aligned} \tag{2.25}$$

In this case, we can set $D_\mathrm{i} = I$ and $C_\mathrm{i} = \bar{C}G^T\mathcal{P}$, and $x_\mathrm{i} = z_\mathrm{o}$.
To find the inner dynamics, we consider the dynamics of z_o

$$\begin{aligned}\dot{x}_\mathrm{i} &= \dot{z}_\mathrm{o} \\ &= F_\mathrm{o} z_\mathrm{o} + G\xi_\mathrm{o}\,.\end{aligned} \tag{2.26}$$

Take B_i such that $G = B_\mathrm{i}\bar{C}$. By the property $\bar{C}\bar{C}^T = I_p$ we obtain $B_\mathrm{i} = G\bar{C}^T$ and $A_\mathrm{i} = F_\mathrm{o}$. Thus the inner dynamics read as

$$\begin{aligned}\dot{x}_\mathrm{i} &= F_\mathrm{o} z_\mathrm{o} + B_\mathrm{i}\bar{C}\xi_\mathrm{o} \\ &= F_\mathrm{o} x_\mathrm{i} + B_\mathrm{i} y_\mathrm{o} \\ y &= \bar{C}G^T\mathcal{P}x_\mathrm{i} + y_\mathrm{o},\end{aligned} \tag{2.27}$$

which is non-minimum phase and Hurwitz. Thus, system $(A_\mathrm{i}, B_\mathrm{i}, C_\mathrm{i}, D_\mathrm{i})$ is inner with

$$\begin{aligned}A_\mathrm{i} &= F + G\bar{C}^T\bar{C}G^T\mathcal{P}, \qquad & B_\mathrm{i} &= G\bar{C}^T \\ C_\mathrm{i} &= \bar{C}G^T\mathcal{P}, & D_\mathrm{i} &= I.\end{aligned} \tag{2.28}$$

By the properties of the ARE (2.16) and of its stabilizing solution $\mathcal{P}$, the stable eigenvalues of F are left unchanged in $F + G\bar{C}^T\bar{C}G^T\mathcal{P}$ while the unstable ones are mirrored to the left half plane at the same distance from the imaginary axis. This implies that the stable eigenvalues of F are both stable zeros and poles of $(A_\mathrm{i}, B_\mathrm{i}, C_\mathrm{i}, D_\mathrm{i})$, while the unstable eigenvalues of F are unstable zeros of $(A_\mathrm{i}, B_\mathrm{i}, C_\mathrm{i}, D_\mathrm{i})$ and are mirrored by the relative stable poles in $F + G\bar{C}^T\bar{C}G^T\mathcal{P}$.
By construction, the cascade of $(A_\mathrm{o}, B_\mathrm{o}, C_\mathrm{o})$ and $(A_\mathrm{i}, B_\mathrm{i}, C_\mathrm{i}, D_\mathrm{i})$ is a non-minimal realization of system (A, B, C). Indeed, we just split the output trajectory $y(t)$ into two terms and defined the two relative dynamics accordingly with the (A, B, C) dynamics. □

We observe that also Gu (2002) considers the Inner-Outer factorization for strictly proper systems. Compared to Gu (2002), Theorem 2.2.1 gives an explicit closed-form solution for the realisation of the inner and outer systems in the state space and also allows us to express the initial conditions of the inner and outer systems in terms of those of the original plant, thus guaranteeing the same input-output behaviour. The construction of Gu (2002) is instead based on an algorithmic iterative procedure.

2.2.1 Spectral factorization

Here, we describe the spectral factorization problem for analytic transfer matrices, i.e., given $G(s)$ an arbitrary transfer matrix, determine a real rational matrix $G_\mathrm{o}(s)$ such that

$$G^T(-s)G(s) = G_\mathrm{o}^T(-s)G_\mathrm{o}(s) \tag{2.29}$$

where $G_\mathrm{o}(s)$ is right-invertible (hence its pseudo-right inverse is analytic in $\mathbb{C}_+$). Then, $G_\mathrm{o}(s)$ is called a spectral factor of $G(s)$ and (2.29) defines a spectral factorization of $G(s)$. For further details, the reader can see Anderson (1967), Chen and Francis (1989), and Francis (1987). In particular, thanks to the all-pass property of the inner factor one can notice that the outer factor of the Inner-Outer decomposition has exactly the desired properties required for the spectral factor.

Corollary 2.2.1. *The outer system with realization $(A_\mathrm{o}, B_\mathrm{o}, C_\mathrm{o})$ in (2.18) is a spectral factor of the initial system (A, B, C).*

Proof. Consider, system (A, B, C) in normal form (2) and its adjoint system. The cascade of these two systems reads as

$$\begin{aligned}\dot{z} &= Fz + G\xi \\ \dot{\xi} &= Hz + \bar{A}\xi + \bar{B}u \\ \dot{z}_T &= -F^T z_T - H^T\xi_T \\ \dot{\xi}_T &= -G^T z_T - \bar{A}^T\xi_T + \bar{C}^T\bar{C}\xi \\ y_T &= -\bar{B}^T\xi_T.\end{aligned} \tag{2.30}$$

By Theorem 2.2.1, we known that system (A, B, C) is input-output equivalent to system the cascade $(A_\mathrm{o}, B_\mathrm{o}, C_\mathrm{o})$ and $(A_\mathrm{i}, B_\mathrm{i}, C_\mathrm{i}, D_\mathrm{i})$ whose realizations are given by (2.18) and (2.19). Thus, the self-adjoint

system (2.30) can be written as

$$\begin{aligned}
\dot{x}_{\mathrm{o}} &= A_{\mathrm{o}}x_{\mathrm{o}} + B_{\mathrm{o}}u \\
\dot{x}_{\mathrm{i}} &= A_{\mathrm{i}}x_{\mathrm{i}} + B_{\mathrm{i}}C_{\mathrm{o}}x_{\mathrm{o}} \\
y &= C_{\mathrm{i}}x_{\mathrm{i}} + D_{\mathrm{i}}C_{\mathrm{o}}x_{\mathrm{o}} \\
\dot{x}_{\mathrm{i}T} &= -A_{\mathrm{i}}^T x_{\mathrm{i}T} + C_{\mathrm{i}}^T y \\
\dot{x}_{\mathrm{o}T} &= -C_{\mathrm{o}}^T B_{\mathrm{i}}^T x_{\mathrm{i}T} - A_{\mathrm{o}}^T x_{\mathrm{o}T} + C_{\mathrm{o}}^T D_{\mathrm{i}}^T y \\
y_T &= -B_{\mathrm{o}}x_{\mathrm{o}T}
\end{aligned} \tag{2.31}$$

which is equivalent to the frequency domain relationship $G(-s)^T G(s) = G_{\mathrm{o}}(-s)^T G_{\mathrm{i}}(-s)^T G_{\mathrm{i}}(s) G_{\mathrm{o}}(s)$. To prove that the $(A_{\mathrm{o}}, B_{\mathrm{o}}, C_{\mathrm{o}})$ is a spectral factor of (A, B, C), is enough to show that cascade of $(A_{\mathrm{i}}, B_{\mathrm{i}}, C_{\mathrm{i}}, D_{\mathrm{i}})$ with its adjoint system is all pass (i.e., the output of such cascade is equal to its input) for all t, for a zero initial condition. Hence, we analyze the properties of the state space representation of $G_{\mathrm{i}}^T(-s)G_{\mathrm{i}}(s)$

$$\begin{aligned}
\dot{x}_{\mathrm{i}} &= A_{\mathrm{i}}x_{\mathrm{i}} + B_{\mathrm{i}}u \\
\dot{x}_{\mathrm{i}T} &= -A_{\mathrm{i}}^T x_{\mathrm{i}T} + C_{\mathrm{i}}^T(C_{\mathrm{i}}x_{\mathrm{i}} + D_{\mathrm{i}}u) \\
y_{\mathrm{i}T} &= -B_{\mathrm{i}}^T x_{\mathrm{i}T} + D_{\mathrm{i}}^T C_{\mathrm{i}}x_{\mathrm{i}} + D_{\mathrm{i}}^T D_{\mathrm{i}}u
\end{aligned} \tag{2.32}$$

in which $D_{\mathrm{i}} = I$ and we need to prove that the output term $-B_{\mathrm{i}}^T x_{\mathrm{i}T} + C_{\mathrm{i}}x_{\mathrm{i}} = 0$. Indeed, substituting $B_{\mathrm{i}} = G\bar{C}^T$ and $C_{\mathrm{i}} = \bar{C}G^T\mathcal{P}$, taking into account that for transfer matrix case the cascade (2.32) has zero initial condition ($x_{\mathrm{i}}(0) = x_{\mathrm{i}T}(0) = 0$), we show $\mathcal{P}x_{\mathrm{i}} - x_{\mathrm{i}T} = 0$ for all $t > 0$, (for $t = 0$ the equivalence is trivial satisfied). Define a change of coordinates $\chi_{\mathrm{i}} = \mathcal{P}x_{\mathrm{i}} - x_{\mathrm{i}T}$, then

$$\begin{aligned}
\dot{\chi}_{\mathrm{i}} &= \mathcal{P}\dot{x}_{\mathrm{i}} - \dot{x}_{\mathrm{i}T} \\
&= \mathcal{P}(F + G\bar{C}^T\bar{C}G^T\mathcal{P})x_{\mathrm{i}} - \mathcal{P}G\bar{C}^T u - (F + G\bar{C}^T\bar{C}G^T\mathcal{P})^T\mathcal{P}x_{\mathrm{i}} + \\
&\quad \mathcal{P}G\bar{C}^T\bar{C}G^T\mathcal{P}x_{\mathrm{i}} + \mathcal{P}G\bar{C}^T u \\
&= -(F + G\bar{C}^T\bar{C}G^T\mathcal{P})^T\chi_{\mathrm{i}} + (F^T\mathcal{P} + \mathcal{P}G\bar{C}^T\bar{C}G^T\mathcal{P} + \\
&\quad \mathcal{P}F + \mathcal{P}G\bar{C}^T\bar{C}G^T\mathcal{P} - \mathcal{P}G\bar{C}^T\bar{C}G^T\mathcal{P})x_{\mathrm{i}} \\
&= -(F + G\bar{C}^T\bar{C}G^T\mathcal{P})^T\chi_{\mathrm{i}}
\end{aligned} \tag{2.33}$$

for the property of $\mathcal{P}$. Since $\chi_{\mathrm{i}}(0) = 0$ and its dynamics is autonomous, we can conclude $\chi_{\mathrm{i}}(t) = \mathcal{P}x_{\mathrm{i}}(t) - x_{\mathrm{i}T}(t) = 0$, $\forall t \geq 0$. Thus, $y_{\mathrm{i}T} = u$ and so the cascade is all pass. We thus have the equivalence between (2.30) and the cascade of $(A_{\mathrm{o}}, B_{\mathrm{o}}, C_{\mathrm{o}})$ and its adjoint system for a zero initial condition. Hence, the outer system $(A_{\mathrm{o}}, B_{\mathrm{o}}, C_{\mathrm{o}})$ is the spectral factor of (A, B, C). □

2.2.2 Alternative Inner-Outer realization

It is well-known that Inner-Outer decomposition for transfer matrix representation is unique up to sign, Chen and Francis (1989) and Qiu and Davison (1993). In terms of state-space realizations, this means that an equivalent Inner-Outer decomposition of system (A, B, C) is given by

$$\begin{aligned}
A_{\mathrm{o}} &= \begin{bmatrix} F_{\mathrm{o}} & G_{\mathrm{o}} \\ H_{\mathrm{o}} & \bar{A}_{\mathrm{o}} \end{bmatrix} \\
&= \begin{bmatrix} F + G\bar{C}^T\bar{C}G^T\mathcal{P} & -G \\ -H + G^T\mathcal{P}F_{\mathrm{o}} - \bar{A}\bar{C}^T\bar{C}G^T\mathcal{P} & \bar{A} - \bar{C}^T\bar{C}G^T\mathcal{P}G \end{bmatrix} \\
B_{\mathrm{o}} &= \begin{bmatrix} 0 \\ -\bar{B} \end{bmatrix}, \quad C_{\mathrm{o}} = \begin{bmatrix} 0 & \bar{C} \end{bmatrix}
\end{aligned} \tag{2.34}$$

and

$$\begin{aligned}
A_{z_{\mathrm{i}}} &= F + G\bar{C}^T\bar{C}G^T\mathcal{P}, & B_{z_{\mathrm{i}}} &= -G\bar{C}^T \\
C_{z_{\mathrm{i}}} &= \bar{C}G^T\mathcal{P}, & D_{z_{\mathrm{i}}} &= -I
\end{aligned} \tag{2.35}$$

with initial conditions

$$\begin{aligned}
z_{\mathrm{i}}(0) &= z(0) = \begin{bmatrix} I & 0 \end{bmatrix} x(0) \\
x_{\mathrm{o}}(0) &= \begin{bmatrix} z(0) \\ \bar{C}^T\bar{C}G^T\mathcal{P}z(0) - \xi(0) \end{bmatrix} = \begin{bmatrix} I & 0 \\ \bar{C}^T\bar{C}G^T\mathcal{P} & -I \end{bmatrix} x(0).
\end{aligned}$$

The proof for alternative realization in (2.34) and (2.35) is based on the definition of $\xi_{\mathrm{o}}(t) = \bar{C}^T\bar{C}G^T\mathcal{P}z(t) - \xi(t)$ and does not add any value, thus it is omitted.

Remark 2. *The state space realization of the Inner-Outer Decomposition is unique (up to sign) for a particular normal form.*

Moreover, we observe that both (2.19) and (2.35) are not minimal realization of the inner factor. In (2.2.4) we show how to obtain its minimal realization.

2.2.3 Optimal control interpretation

By analysing the zero dynamics system (F, G), the change of coordinates we apply on the coordinates ξ is a backstepping approach that allows the stabilisation of the zero dynamics in an optimal way. In particular, by solving the ARE equation (2.16) we are concurrently solving the optimal expensive control problem on the zero dynamics (F, G'), where now the zero dynamics only depends on the output signal $y = \bar{C}\xi$. The latter is part of the input vector to be minimized. Indeed, by defining the cost function

$$J = \int_0^\infty z(s)^T Q z(s) + y(s)^T R y(s) ds$$

with $Q = 0$ and $R = I$, with the dynamical system

$$\dot{z} = Fz + G'y$$

and by solving the ARE (2.16), reported here for the sake of completeness,

$$\mathcal{P}F + F^T\mathcal{P} + \mathcal{P}G'G'^T\mathcal{P} = 0$$

we find the control signal $y^\star = G'^T\mathcal{P}z(t)$ is optimal in sense of the input energy J.
A similar analysis has also been treated in Gu (2002), from a different point of view.

2.2.4 Inner system minimal realization

By exploiting the Kalman decomposition for the observable/ non-observable dynamics of the system we can define the change of coordinates for (2.28). In particular we define $n_{\mathrm{i}} \leq n_x - r$, r is the sum of the elements of the vector relative degree, and n_{i} is the rank of the Observability matrix $\mathcal{O}(A_{\mathrm{i}}, C_{\mathrm{i}})$. We then define the change of coordinates

$$T_{\mathrm{i}} = \begin{bmatrix} T_{\mathrm{i}}^\star \\ \mathcal{O}_{ni} \end{bmatrix}$$

where $\mathcal{O}_{ni}$ are n_{i} linearly independent rows of the observability matrix $\mathcal{O}(A_{\mathrm{i}}, C_{\mathrm{i}})$ and $T_{\mathrm{i}}^\star$ are any rows that makes T_{i} nonsingular (and possibly $T_{\mathrm{i}}^\star B_{z_{\mathrm{i}}} = 0$). By applying the change of coordinates T_{i}, the system in (2.28) reads as

$$\begin{aligned} T_{\mathrm{i}} A_{z_{\mathrm{i}}} T_{\mathrm{i}}^{-1} &= \begin{bmatrix} A_{di} & A_{doi} \\ 0 & A_{oi} \end{bmatrix}, & T_{\mathrm{i}} B_{z_{\mathrm{i}}} &= \begin{bmatrix} B_{di} \\ B_{oi} \end{bmatrix} \\ C_{z_{\mathrm{i}}} T_{\mathrm{i}}^{-1} &= \begin{bmatrix} 0 & C_{oi} \end{bmatrix}, & D_{z_{\mathrm{i}}} &= I. \end{aligned} \tag{2.36}$$

Hence, the minimal realization of (2.28) is given by $(A_{oi}, B_{oi}, C_{oi}, D_{z_{\mathrm{i}}})$.

2.3 A simple example (continued)

Consider again the SISO non-minimum phase system as in (2.5)

$$G(s) = \frac{s^2 - 2s}{(s+1)^3}. \tag{2.37}$$

Its state space realization in controllability canonical form is given by the matrices

$$\begin{aligned} A &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -3 & -3 \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ C &= \begin{bmatrix} 0 & -2 & 1 \end{bmatrix}. \end{aligned} \tag{2.38}$$

With the change of coordinates $[z^T \; y^T] = Tx$, where

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ & C & \end{bmatrix} \tag{2.39}$$

we put the system in canonical normal form

$$\begin{aligned} TAT^{-1} &= \begin{bmatrix} F & G \\ H & \bar{A} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ \begin{bmatrix} -1 & -13 \end{bmatrix} & \begin{bmatrix} -5 \end{bmatrix} \end{bmatrix} \\ TB &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \\ CT^{-1} &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{2.40}$$

Thus, in order to find the outer factor (2.22) we need to solve the zero-associated ARE (2.16)

$$\mathcal{P}F + F^T\mathcal{P} + \mathcal{P}GG^T\mathcal{P} = 0 \tag{2.41}$$

solving element-wise we find a semi-negative definite matrix

$$\mathcal{P} = \begin{bmatrix} 0 & 0 \\ 0 & -4 \end{bmatrix} \tag{2.42}$$

for which

$$\begin{aligned} F_{\mathrm{o}} &= F + GG^T\mathcal{P} = \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} \\ G_{\mathrm{o}} &= G = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ H_{\mathrm{o}} &= H - G^T\mathcal{P}Fo + \bar{A}G^T\mathcal{P} = \begin{bmatrix} -1 & -1 \end{bmatrix} \\ \bar{A}_{\mathrm{o}} &= \bar{A} - G^T\mathcal{P}G = -1. \end{aligned} \tag{2.43}$$

Then, the outer system matrices are given by

$$\begin{aligned} A_{\mathrm{o}} &= \begin{bmatrix} F_{\mathrm{o}} & G_{\mathrm{o}} \\ H_{\mathrm{o}} & \bar{A}_{\mathrm{o}} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix} \\ B_{\mathrm{o}} &= TB = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \\ C_{\mathrm{o}} &= CT^{-1} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{2.44}$$

The outer transfer function is then

$$G_{\mathrm{o}}(s) = C_{\mathrm{o}}(sI - A_{\mathrm{o}})^{-1}B_{\mathrm{o}} = \frac{s^2 + 2s}{(s+1)^3}. \tag{2.45}$$

For the inner system, we have realization matrices

$$\begin{aligned} A_{\mathrm{i}} &= F_{\mathrm{o}} = \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} \\ B_{\mathrm{i}} &= G = \begin{bmatrix} 0 & 1 \end{bmatrix}^T \\ C_{\mathrm{i}} &= G^T\mathcal{P} = \begin{bmatrix} 0 & -4 \end{bmatrix} \\ D_{\mathrm{i}} &= 1. \end{aligned} \tag{2.46}$$

it is very easy to see that the inner system minimal realization is given by

$$(A_{oi}, B_{oi}, C_{oi}, D_{z_{\mathrm{i}}}) = (-2, 1, -4, 1)$$

and its transfer function is

$$G_{\mathrm{i}}(s) = D_{\mathrm{i}} + C_{\mathrm{i}}(sI - A_{\mathrm{i}})^{-1}B_{\mathrm{i}} = \frac{s-2}{s+2}. \tag{2.47}$$

Thus, we can write $G(s) = G_{\mathrm{i}}(s)G_{\mathrm{o}}(s)$

$$\frac{s^2 - 2s}{(s+1)^3} = \frac{s-2}{s+2} \cdot \frac{s(s+2)}{(s+1)^3} \tag{2.48}$$

2.3.1 Example of the inverted pendulum on a cart

For a more interesting example, we consider the model of the inverted pendulum on a cart as shown in (Gurumoorthy and Sanders, 1993):

$$\begin{aligned}(M+m)\ddot{p} + b\dot{p} + m\ell\left(\ddot{\theta}\cos(\theta) - \dot{\theta}^2\sin(\theta)\right) &= u \\ m\left(\ddot{p}\cos(\theta) + \ell\ddot{\theta} - g\sin(\theta)\right) &= 0\end{aligned} \tag{2.49}$$

where p is the cart position, M is the lumped mass of the cart, b is the viscous friction coefficient and u the force applied to the cart along the horizontal direction, θ is the angle of the pendulum barycenter draws with respect to the vertical axis, m is the lumped mass of the pendulum, ℓ is the distance between the centre of rotation on the cart and the pendulum barycenter.
By defining $v = \dot{p}$, $\omega = \dot{\theta}$, we write the system model in state space with coordinates $x = (p, v, \theta, \omega)$ and then by linearizing the model about the point $(1, 0, 0, 0) = (p, v, \theta, \omega)(0)$ we have (A, B, C) matrices

$$\begin{aligned}A &= \begin{bmatrix}0 & 1 & 0 & 0\\ 0 & -\frac{b}{M} & -\frac{mg}{M} & 0\\ 0 & 0 & 0 & 1\\ 0 & \frac{b}{\ell M} & \frac{(M+m)g}{\ell M} & 0\end{bmatrix},\ B = \begin{bmatrix}0\\ \frac{1}{M}\\ 0\\ -\frac{1}{\ell M}\end{bmatrix} \\ C &= \begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}.\end{aligned} \tag{2.50}$$

The linearized system has relative degree $r = 2$, thus by taking

$$T_{\text{nf}} = \begin{bmatrix}1 & 0 & 1 & 0\\ 1 & 1 & 1 & \ell\\ & C & & \\ & CA & & \end{bmatrix},\ T_{\text{nf}}^{-1} = \begin{bmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ \ell^{-1} & 0 & -\ell^{-1} & 0\\ -\ell^{-1} & \ell^{-1} & 0 & -\ell^{-1}\end{bmatrix}$$

we obtain the system normal form Isidori (2017)[Ch.2] with coordinates $(z, \xi) = T_{\text{nf}}x$

$$\begin{aligned}A_{\text{nf}} &= \begin{bmatrix}-1 & 1 & 0 & 0\\ \frac{g}{\ell} - 1 & 1 & -\frac{g}{\ell} & 0\\ 0 & 0 & 0 & 1\\ -\frac{mg}{M\ell} & 0 & \frac{mg}{M\ell} & -\frac{b}{M}\end{bmatrix},\ B_{\text{nf}} = \begin{bmatrix}0\\ 0\\ 0\\ \frac{1}{M}\end{bmatrix} \\ C_{\text{nf}} &= \begin{bmatrix}0 & 0 & 1 & 0\end{bmatrix}\end{aligned} \tag{2.51}$$

where $\bar{C} = \begin{bmatrix}1 & 0\end{bmatrix}$ and $\bar{B} = \begin{bmatrix}0 & \frac{1}{M}\end{bmatrix}^T$, and $x_{\text{nf}}(0) = (1, 1, 1, 0)^T = T_{\text{nf}}x(0)$. In particular, the normal form matrices $F, G, H, \bar{A}$ are

$$\begin{aligned}F &= \begin{bmatrix}-1 & 1\\ \frac{g}{\ell} - 1 & 1\end{bmatrix}, \qquad G = \begin{bmatrix}0 & 0\\ -\frac{g}{\ell} & 0\end{bmatrix}, \\ H &= \begin{bmatrix}0 & 0\\ -\frac{mg}{M\ell} & 0\end{bmatrix}, \qquad \bar{A} = \begin{bmatrix}0 & 1\\ \frac{mg}{M\ell} & -\frac{b}{M}\end{bmatrix}.\end{aligned} \tag{2.52}$$

With system data

$$g = 9.81\,[m/s],\ \ell = 0.325\,[m],\ m = 0.051\,[Kg],\ M = 1.378\,[Kg],\ b = 12.98\,[Ns/m]$$

we obtain the stabilizing solution $\mathcal{P}$ of $\mathcal{P}F + F^T\mathcal{P} + \mathcal{P}G\bar{C}^T\bar{C}G^T\mathcal{P} = 0$

$$\mathcal{P} = \begin{bmatrix}-0.2436 & -0.0542\\ -0.0542 & -0.0121\end{bmatrix}.$$

And the outer system reads as

$$\begin{aligned}A_{\text{o}} &= \begin{bmatrix}-1 & 1 & 0 & 0\\ -20.1965 & -9.9881 & -30.1846 & 0\\ 8.9881 & 2 & 10.9881 & 1\\ 0.7105 & 0.4067 & 1.1171 & -9.4194\end{bmatrix}, \\ B_{\text{o}} &= B_{\text{nf}} = \begin{bmatrix}0 & 0 & 0 & \frac{1}{M}\end{bmatrix}^T, \\ C_{\text{o}} &= C_{\text{nf}} = \begin{bmatrix}0 & 0 & 1 & 0\end{bmatrix}\end{aligned} \tag{2.53}$$

with initial condition $x_{\text{o}}(0) = (1, 1, -1, 0) = T_{\text{o}}T_{\text{nf}}x(0)$. The inner factor in minimal realization has matrices

$$A_{\text{i}} = -5.4941, \quad B_{\text{i}} = -10.9881, \quad C_{\text{i}} = D_{\text{i}} = 1$$

with initial condition $x_{\text{i}}(0) = [1\ 0]T_{\text{i}}z_{\text{o}}(0) = 2$.

2.4 Stabilization of non-minimum phase systems

Consider a SISO non-minimum phase linear system (A, B, C) and its Inner-Outer decomposition (2.13). Let $y_o = C_o x_o$ be the output of the outer subsystem with input u and initial state $x_o(0) = \bar{T}_o x(0)$. Furthermore, let $\bar{y}_o = G_i(0) y_o$, with $G_i(0) = D_i - C_i A_i^{-1} B_i$.
Our main objective is to start with a (state) feedback stabiliser for the outer system, resulting in an output trajectory $y_o(t)$, and to explore the design of an output feedback stabiliser for the original system so that the resulting output $y(t)$ is "practically" close to $\bar{y}_o(t)$. Of course, the behaviour of $\bar{y}_o(t)$ cannot be perfectly reproduced on the non-minimum phase system output $y(t)$ due to the intrinsic limits of performances characterising the latter. It is well-known (see Middleton (1991) and Stewart and Davison (2006)) that the output of a SISO system with one unstable zero is necessarily characterised by undershoots whose entity increases as the closed-loop setting time decreases and as the position of the zero gets closer to the imaginary axis. Such undershoot is not present in the output $y_o(t)$ because of the minimum phase behaviour of the outer system. In view of this, the practical matching between the two output behaviours can be only attained, at best, after a time $t^\star > 0$. Indeed, in the next result, we show that such a $t^\star$ can be rendered arbitrarily small provided that the outer closed-loop dynamics are sufficiently slow, with the restrictions on the outer closed-loop dynamics that depend on the position of the slowest unstable zero of the original system. The result is detailed next.
Let

$$\mathcal{X}_0 := \{x_0 \in \mathbb{R}^{n_x} : Ax_0 + Bu_0 = 0, \text{ for some real } u_0\} \tag{2.54}$$

namely $x_0 \in \mathcal{X}_0$ is a forced equilibrium for (1) when $u = u_0$.
We now assume to have a state-feedback stabiliser for the outer system $u = -K_o(\alpha_o)x_o(t)$, with α_o a positive real parameter, such that the resulting outer output trajectory originating from an initial condition $x_o(0) = \bar{T}_o x_0$ with $x_0 \in \mathcal{X}_0$ satisfies $|\dot{y}_o(t)| \le \alpha_o c_y e^{-\alpha_o t} \le \alpha_o c_y$, with real $c > 0$. Consider now the observer

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x}), \quad \hat{x}_o = \text{sat}_{\bar{x}_o}(\bar{T}_o \hat{x}) \tag{2.55}$$

with $\bar{x}_o$ and L to be designed, and consider system (1) controlled by

$$u(t) = -K_o(\alpha_o)\hat{x}_o(t)\,. \tag{2.56}$$

Then, the following result holds.

Theorem 2.4.1. *Suppose that (A, B) is controllable, (A, C) is observable and that the triplet (A, B, C) has no transmission zeros on the imaginary axis. Then, for every compact subsets $X_0 \subset \mathcal{X}_0$ and $X_{c0} \subset \mathbb{R}^{n_x}$ and $X \subset \mathbb{R}^{n_x}$ such that $x(t) \in X$, for all $t \ge 0$, there exists $\bar{x}_o > 0$ and, for every $\epsilon_y > 0$, $t^\star > 0$, there exist $\alpha_o^\star > 0$ and L such that for all $\alpha_o \le \alpha_o^\star$, the output $y(t)$ resulting from the closed-loop system (1), (2.55), (2.56) with initial conditions $\hat{x}(0) \in X_{c0}$, $x(0) \in X_0$ satisfies*

$$|y(t) - \bar{y}_o(t)| \le \epsilon_y \qquad \forall\, t \ge t^\star \tag{2.57}$$

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

Proof. Inspired by separation principle arguments, the proof is conceptually split into two parts. In the first part, we consider the case in which the state $x(t)$ is available for feedback and we show that the result is true with $t^\star = 0$ if the original system is controlled by $u(t) = -K_o x_o(t) = -K_o \bar{T}_o x(t)$. The second part follows high-gain arguments typically used in the context of output feedback stabilization of nonlinear systems (see, e.g., Isidori (2017)[Sec. 7.5]). By defining $\bar{x}_o = \sup_{x_o \in \bar{T}_o X} x_o$, the degree-of-freedom L can be chosen to quickly (i.e. in an arbitrarily amount of time $t^\star$) practically recover the state x_o. Throughout the proof, (x_i, x_o) denotes the state of the Inner-Outer realization of (A, B, C), as introduced in Section 2.2, with inner system minimal realization. *Part I:* We can explicitly write the state $x_i(t)$ as

$$x_i(t) = e^{A_i t} x_i(0) + \int_0^t e^{A_i(t-s)} B_i C_o x_o(s) ds$$

whose integral part can be written, via integration by parts, as

$$\begin{aligned}\int_0^t e^{A_i(t-s)} B_i C_o x_o(s) ds &= -\int_0^t \left[-e^{A_i(t-s)} A_i^{-1}\right] \left[B_i C_o x_o(s)\right]' ds + \left[-e^{A_i(t-s)} A_i^{-1} B_i C_o \bar{x}_o(s)\right]_0^t \\ &= \int_0^t e^{A_i(t-s)} A_i^{-1} B_i \dot{y}_o(s) ds - e^{A_i(t-t)} A_i^{-1} B_i C_o x_o(t) + e^{A_i t} A_i^{-1} B_i C_o x_o(0).\end{aligned}$$

Then, from the output system $y(t)$ we can write

$$y(t) - D_i C_o \bar{x}_o(t) + C_i A_i^{-1} B_i C_o \bar{x}_o(t) = C_i e^{A_i t}\left(x_i(0) + A_i^{-1} B_i C_o \bar{x}_o(0)\right) + C_i \int_0^t e^{A_i(t-s)} A_i^{-1} B_i \dot{y}_o(s) ds.$$

For every solution of (A, B, C) originating in $\mathcal{X}_0$, we have $x_{\rm i}(0) + A_{\rm i}^{-1} B_{\rm i} C_{\rm o} \bar{x}_{\rm o}(0) = 0$ as a consequence of the matrices definition (up to a change of coordinates, note that $\dot{x}_{\rm i}(0) = \dot{z}(0) = 0$). Then, using $G_{\rm i}(0) = D_{\rm i} - C_{\rm i} A_{\rm i}^{-1} B_{\rm i}$, we write

$$y - G_{\rm i}(0) y_{\rm o} = C_{\rm i} \int_0^t e^{A_{\rm i}(t-s)} A_{\rm i}^{-1} B_{\rm i} \dot{y}_{\rm o}(s) ds. \tag{2.58}$$

Then, by the properties of the outer output derivative in closed loop $\dot{y}_{\rm o}$, we can upper bound (2.58) as follows

$$\begin{aligned} |y - G_{\rm i}(0) y_{\rm o}| &\leq \left| C_{\rm i} \int_0^t e^{A_{\rm i}(t-s)} A_{\rm i}^{-1} B_{\rm i} \dot{y}_{\rm o}(s) ds \right| \leq \|C_{\rm i}\| \left| \int_0^t e^{A_{\rm i}(t-s)} A_{\rm i}^{-1} ds \right| \|B_{\rm i}\| \alpha_{\rm o} c_y \\ &\leq \|C_{\rm i}\| \|A_{\rm i}^{-2}\| \|(e^{A_{\rm i} t} - I)\| \|B_{\rm i}\| \alpha_{\rm o} c_y \leq \|C_{\rm i}\| \|A_{\rm i}^{-2}\| \|B_{\rm i}\| \alpha_{\rm o} c_y \end{aligned}$$

Then, for every ϵ_y there exists $\alpha_{\rm o}^\star = \alpha_{\rm o}^\star(\epsilon_y)$ such that for all $\alpha_{\rm o} \leq \alpha_{\rm o}^\star$

$$|y(t) - G_{\rm i}(0) y_{\rm o}(t)| \leq \epsilon_y, \quad \forall t \geq 0 \tag{2.59}$$

and thus (2.57) holds with $t^\star = 0$. Moreover, since $A_{\rm i}$ and $(A_{\rm o} - B_{\rm o} K_{\rm o})$ are Hurwitz, we also have $y(t) \to 0$ as $t \to \infty$.

Part II: By considering the output feedback controller (2.55) we can write the observer dynamics in error coordinates $\tilde{x} = x - \hat{x}$ and the closed loop as

$$\begin{aligned} \dot{\tilde{x}} &= (A - LC)\tilde{x} \\ \dot{x}_{\rm o} &= A_{\rm o} x_{\rm o} - B_{\rm o} K_{\rm o}(\alpha_{\rm o}) \mathrm{sat}_{\bar{x}_{\rm o}}(x_{\rm o} - \tilde{x}_{\rm o}) \\ \dot{x}_{\rm i} &= A_{\rm i} x_{\rm i} + B_{\rm i} C_{\rm o} x_{\rm o}. \end{aligned}$$

Then, by defining

$$\Delta A_{K_{\rm o}}(x_{\rm o}, \tilde{x}_{\rm o}) := B_{\rm o} K_{\rm o}(\alpha_{\rm o}) - B_{\rm o} K_{\rm o}(\alpha_{\rm o}) \mathrm{sat}_{\bar{x}_{\rm o}}(x_{\rm o} - \tilde{x}_{\rm o})$$

hence the outer dynamics reads as

$$\dot{x}_{\rm o} = (A_{\rm o} - B_{\rm o} K_{\rm o}(\alpha_{\rm o})) x_{\rm o} + \Delta A_{K_{\rm o}}(x_{\rm o}, \tilde{x}_{\rm o}).$$

Since (A, C) is observable there exists a L such that $A - LC$ is Hurwitz for any desired eigenvalues. Then, for every real positive q there exists symmetric positive definite matrix P_{LC} solution of

$$(A - LC)^T P_{LC} + P_{LC}(A - LC) = -2qI.$$

Then, defining $V(\tilde{x}) := \tilde{x}^T P_{LC} \tilde{x}$ its dynamics is given by

$$\dot{V}(\tilde{x}) = -2q|\tilde{x}|^2 \leq -\frac{2q}{\sigma_{\max}(P_{Lc})} V(\tilde{x})$$

because for all $\tilde{x}$ in $\mathbb{R}^{n_x}$

$$\sigma_{\min}(P_{LC})|\tilde{x}|^2 \leq V(\tilde{x}) \leq \sigma_{\max}(P_{LC})|\tilde{x}|^2.$$

We can explicitly write $V(t)$ from its dynamics

$$V(\tilde{x}(t)) \leq V(\tilde{x}(0)) \exp\left(-\frac{2qt}{\sigma_{\max}(P_{LC})}\right)$$

then

$$\sigma_{\min}(P_{LC})|\tilde{x}|^2 \leq \sigma_{\max}(P_{LC})|\tilde{x}(0)|^2 \exp\left(-\frac{2qt}{\sigma_{\max}(P_{LC})}\right)$$

and hence we can bind the error evolution

$$|\tilde{x}| \leq \rho \sqrt{\frac{\sigma_{\max}(P_{LC})}{\sigma_{\min}(P_{LC})}} \exp\left(-\frac{qt}{\sigma_{\max}(P_{LC})}\right) \tag{2.60}$$

with $\rho = \max_{\hat{x} \in X_{c0}} \hat{x} + \max_{x \in X_0} x$. For the observability properties of (A, C), for every $\epsilon_{\tilde{x}} > 0$ and every $t^\star > 0$ there exists L and q, (and thus P_{LC}), such that

$$|\tilde{x}| \leq \rho \sqrt{\frac{\sigma_{\max}(P_{LC})}{\sigma_{\min}(P_{LC})}} \exp\left(-\frac{qt}{\sigma_{\max}(P_{LC})}\right) \leq \epsilon_{\tilde{x}}, \quad \forall t \leq t^\star.$$

We now analysis the behavior of $\dot{y}_o$ for $t \in [0, t^\star)$ and $t \leq t^\star$. In particular, for $t \in [0, t^\star)$

$$\begin{aligned}|\dot{y}_o| = |C_o \dot{x}_o| &\leq |C_o(A_o - B_o K_o)x_o| + \|C_o B_o\||\Delta A_o K_o| \\ &\leq \alpha_o c_y + \|C_o B_o\|\delta_1\end{aligned}$$

where $C_o(A_o - B_o K_o)x_o$ is the outer output in the ideal state feedback case, and the $|\Delta A_o K_o|$ term has been upper bounded by a positive real number δ_1, i.e., $|\Delta A_o K_o| \leq \delta_1$ for all $x_o \in T_o X_0$ and $\hat{x}_o \in T_o X_{c0}$. For $t \in [t^\star, \infty)$, instead, because $\Delta A_o K_o(x_o, 0) = 0$ for $\tilde{x}_o$ sufficiently small there exists $\delta_2 > 0$ such that $|\Delta A_o K_o| \leq \delta_2 |\tilde{x}_o| = \delta_2 |T_o \tilde{x}| \leq \delta_2 \|T_o\| \epsilon_{\tilde{x}}$.

$$\begin{aligned}|\dot{y}_o| = |C_o \dot{x}_o| &\leq |C_o(A_o - B_o K_o)x_o| + \|C_o B_o\||\Delta A_o K_o| \\ &\leq \alpha_o c_y + \|C_o B_o\|\delta_2\|T_o\|\epsilon_{\tilde{x}}.\end{aligned}$$

Then, from (2.58) we can write for $t \in [0, t^\star)$

$$|y(t) - G_i)(0)y_o(t)| \leq \|C_i\|\|B_i\|\|A_i^{-2}\| \left(\alpha_o c_y + \|C_o\|\|B_o\|\delta_1\right).$$

While, for $t \in [t^\star, \infty)$ we have

$$\begin{aligned}|y(t) - G_i)(0)y_o(t)| &= \left| C_i \int_0^t e^{A_i(t-s)} A_i^{-1} B_i \dot{y}_o(s) ds \right| \\ &\leq \left| C_i \int_0^{t^\star} e^{A_i(t-s)} A_i^{-1} B_i \dot{y}_o(s) ds + C_i \int_{t^\star}^t e^{A_i(t-s)} A_i^{-1} B_i \dot{y}_o(s) ds \right| \\ &\leq \|C_i\|\|B_i\|\|A_i^{-2}\|\|e^{A_i t^\star} - I\| \left(\alpha_o c_y + \|C_o\|\|B_o\|\delta_1\right) + \\ &\quad \|C_i\|\|B_i\|\|A_i^{-2}\| + \left(\alpha_o c_y + \|C_o\|\|B_o\|\delta_2\|T_o\|\epsilon_{\tilde{x}}\right).\end{aligned}$$

For $t^\star \to 0$ the term in δ_1 vanishes and the $\epsilon_{\tilde{x}}$ term can be made arbitrary small reconstructing the same result of the state feedback case. Thus, for every ϵ_y there exists $t^\star$, $\epsilon_{\tilde{x}}^\star$ and $\alpha_o^\star$, such that for all $\epsilon_{\tilde{x}} \leq \epsilon_{\tilde{x}}^\star$ and $\alpha_o \leq \alpha_o^\star$

$$|y(t) - G_i(0)y_o(t)| \leq \epsilon_y, \quad \forall t \geq t^\star.$$

The stability is then guaranteed because the closed loop system (2.4) is a cascade of Hurwitz systems and thus $y(t) \to 0$ as $t \to \infty$. □

Moreover, notice that the upper bound (2.57) gives also an upper bound for the maximum undershoot the system output will exhibit with respect to its ideal minimum phase behaviour.

2.4.1 Comparison with other stabilizing approaches for non-minimum phase available in literature

As one of the reviewer pointed out, some approaches to stabilize non-minimum phase systems are already available in literature. We consider of particular relevance the work by Isidori (2000), Nazrulla and Khalil (2010) and Boker and Khalil (2016). The first two works are mainly describing the same approach, where Nazrulla and Khalil (2010) treated more in detail the nonlinear approach proposed in Isidori (2000), so they share the same comments. In particular, with respect to the work in Isidori (2000) (and Nazrulla and Khalil (2010) as consequence) the inner-outer decomposition approach provides an easy solution to deal with the stabilization of nonminimum phase systems and allowing us to threat them as if they were minimum phase. In other words, once an inner-outer decomposition is available, the stabilizer has to be design only for the outer part of the plant which is a minimum phase system and as a consequence we achieve the stabilization of the whole non minimum phase system. If we consider the zero dynamics of the outer factor to be (ISS) input to state stable (with respect to the input y_o), the stabilizer only has to deal with steering to zero y_o. This is true because the inner-outer realization intrinsically provides a stabilizing action for the unstable zero dynamics. This is not true for the approach proposed in Isidori (2000) because there is no guarantee that the auxiliary system therein is minimum phase and thus the problem of stabilizing such system might result more complicated then the original one, since the auxiliary system is not any more in normal form.
On a different level we find the work Boker and Khalil (2016), in which the state feedback stabilizing action is assumed to be known with some ISS properties for the zero dynamics and then they show how to realize an Extended Kalman filter plus a dirty derivative observer to apply the state feedback stabilizer in an output feedback scenario and recover, after a certain transitory, the performances of the ideal state feedback. Hence, again the problem of stability for non minimum phase is not dealt directly in a constructive way.

2.5 Example of the inverted pendulum on a cart (continued)

In section 2.3.1, we obtained the system normal form Isidori (2017)[Ch.2] with coordinates $(z, \xi) = T_{\text{nf}} x$

$$\begin{aligned} A_{\text{nf}} &= \begin{bmatrix} -1 & 1 & 0 & 0 \\ \frac{g}{\ell} - 1 & 1 & -\frac{g}{\ell} & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{mg}{M\ell} & 0 & \frac{mg}{M\ell} & -\frac{b}{M} \end{bmatrix}, \; B_{\text{nf}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{M} \end{bmatrix} \\ C_{\text{nf}} &= \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \end{aligned} \tag{2.61}$$

where $\bar{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $\bar{B} = \begin{bmatrix} 0 & \frac{1}{M} \end{bmatrix}^T$, and $x_{\text{nf}}(0) = (1, 1, 1, 0) = T_{\text{nf}} x(0)$.
With system data $g = 9.81\,[m/s]$, $\ell = 0.325\,[m]$, $m = 0.051\,[Kg]$, $M = 1.378\,[Kg]$, $b = 12.98\,[Ns/m]$ we obtain the stabilizing solution P of $PF + F^T P + PG\bar{C}^T \bar{C} G^T P = 0$

$$P = \begin{bmatrix} -0.2436 & -0.0542 \\ -0.0542 & -0.0121 \end{bmatrix}.$$

And the outer system reads as

$$\begin{aligned} A_{\text{o}} &= \begin{bmatrix} -1 & 1 & 0 & 0 \\ -20.1965 & -9.9881 & -30.1846 & 0 \\ 8.9881 & 2 & 10.9881 & 1 \\ 0.7105 & 0.4067 & 1.1171 & -9.4194 \end{bmatrix}, \\ B_{\text{o}} &= B_{\text{nf}} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{M} \end{bmatrix}^T \\ C_{\text{o}} &= C_{\text{nf}} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \end{aligned} \tag{2.62}$$

with initial condition $x_{\text{o}}(0) = (1, 1, -1, 0) = T_{\text{o}} T_{\text{nf}} x(0)$. The inner factor in minimal realization has matrices

$$A_{\text{i}} = -5.4941, \quad B_{\text{i}} = -10.9881, \quad C_{\text{i}} = D_{\text{i}} = 1$$

with initial condition $x_{\text{i}}(0) = [1\ 0] T_{\text{i}} z_{\text{o}}(0) = 2$. Note that this initial condition is contained in the $\mathcal{X}_0$ set, thus it satisfies $x_{\text{i}}(0) + A_{\text{i}}^{-1} B_{\text{i}} C_{\text{o}} \bar{x}_{\text{o}}(0) = 2 + 2 \cdot (-1) = 0$.
We then define a static state feedback gain K_{o} such that $(A_{\text{o}} - B_{\text{o}} K_{\text{o}})$ has eigenvalues $\{-1, -1.5, -2, -2.5\}$, while we define L as the observer gain such that the estimation error dynamics $(A - LC)$ has eigenvalues $\{-1, -2, -4.5, -5\} \cdot 10^4$. By saturating the input u between -5 and 5, we can control the system by dynamic output feedback. Figure 2.2 depicts the output for the case of the real system subject to output feedback along with the outer system subject to the ideal state feedback input and the outer system subject to the output feedback controller. One can notice that the observer peaking does not affect the

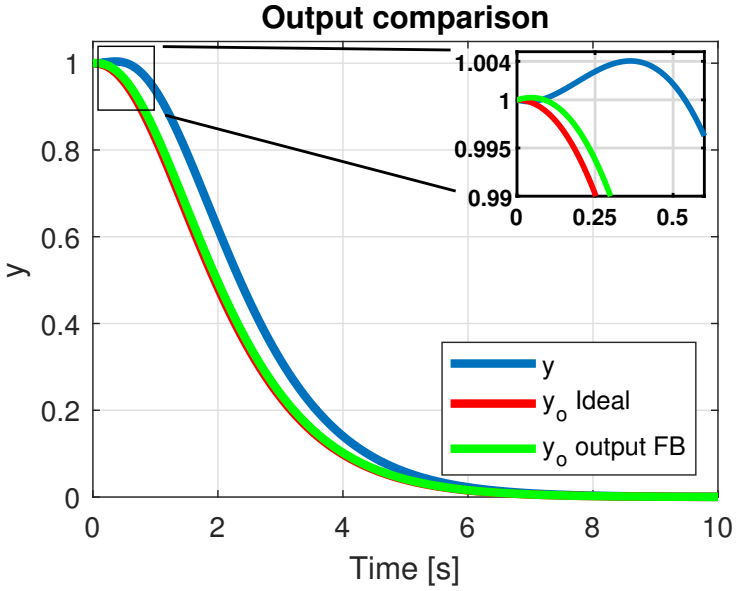

Figure 2.2: Output comparison among the system output 'y', the outer output under state feedback 'y_{o} Ideal', and the outer output subject to output FeedBack 'y_{o} output FB '.

system output thanks to the input saturation and that the output undershoot is about 0.4% of the initial condition $y(0)$. Note that, in order to provide a comparison plot we had to change sign to the outer output $y_{\text{o}}(t) = C_{\text{o}} x_{\text{o}}(t)$ because in our case $G_{\text{i}}(0) = -1$.

2.6 Conclusions

In this chapter, we presented an Inner-Outer decomposition for non-minimum phase multi-input multi-output Linear Time Invariant systems. With respect to existing results, we provide an explicit closed-form decomposition realization completely obtained in state space. The decomposition was then instrumental to present a stabilization result showing how to design an output feedback stabilizer for the original system keeping the output trajectory arbitrary close to the output trajectory of a closed-loop outer system provided that the latter is sufficiently slow.

Part II

Functional Observers

Motivations

A functional observer for a dynamical system allows to asymptotically reconstruct a target functional of the system state x, denoted $\ell(x)$, in the following. After having summarized the most relevant results in the linear system case, we present a unifying framework to gather the relevant results into a single point of view. In the successive chapter, we then specialize the standard KKL approach to the case of functional observation for nonlinear systems and provide some possible applications, such as Input reconstruction, Unknown input observers, and controlled nonlinear systems.

The world of functional observer was moreover motivated by the inner-outer decomposition framework in the sense of the input reconstruction application of the functional observer. Indeed, since the inner system of the decomposition is non-minimum phase, it does not accept a causal left inverse and reconstructing the y_{o} signal is impossible via inversion. Indeed, the sufficient and necessary condition provided in Hou and Muller (1992) to solve the input reconstruction problem via inversion is that the system must be minimum phase.
By exploiting the idea of functional observers for the nonlinear case, we can reconstruct the input of the inner system not via inversion but by exploiting the whole cascade dynamics. Indeed, the output of the outer factor y_{o} can be seen as a functional of the system cascade and thus can in principle be asymptotically reconstructed from the system output by considering a sufficiently large observer order. Unfortunately, only recently we noticed that a system with detectable dynamics does not satisfy the backward distinguishability assumption which is the building pillar of the KKL-observer approach. And due to the equivalent zero/pole cancellations of the inner-outer decomposition, the cascade becomes a detectable system. Hence, we cannot directly apply the functional observer approach to the inner-outer cascade in order to accomplish a reconstruction of the outer system output and then exploit it directly for feedback.

Chapter 3

Linear case

Given a linear autonomous dynamical system with state $x(t)$ and output measurements $y(t)$, and some functional of the system state, in the following $\ell(x(t))$, the basic idea of a functional observer is to construct a dynamical system driven by the only signal available, i.e., the output trajectory $y(t)$, to asymptotically reconstruct the functional $\ell(x(t))$, possibly, without estimating the whole state $x(t)$.

3.1 Problem statement and preliminaries

The problem under consideration that we want to analyze in this chapter is the following.

Problem 1. *Consider system*

$$\begin{aligned} \dot{x} &= Fx \\ y &= Hx \\ \ell(x) &= L_0 x \end{aligned} \tag{3.1}$$

in which $x \in \mathbb{R}^{n_x}$, $y \in \mathbb{R}^{n_y}$, *and* $\ell(x) \in \mathbb{R}^{n_\ell}$, *thus* $rank[H] = n_y < n_x$ *and* $rank[L_0] = n_\ell < n_x$. *Then, look for an observer of the form*

$$\begin{aligned} \dot{\eta} &= A\eta + By \\ \hat{\ell} &= C\eta + Dy \end{aligned} \tag{3.2}$$

with $\eta \in \mathbb{R}^{n_\eta}$, $n_\eta \geq n_\ell$, *and* A, B, C, D *such that* $\lim_{t\to+\infty} \hat{\ell} - \ell = 0$.

Inspired by Luenberger, we can look for $n_\eta \geq n_\ell$, $A \in \mathbb{R}^{n_\eta \times n_\eta}$ Hurwitz, and B, C, D such that there is T solution to

$$TF = AT + BH \tag{3.3a}$$
$$L_0 = CT + DH \tag{3.3b}$$

Indeed, in that case, since A is Hurwitz, for any initial condition in (3.2), we have $\lim_{t\to+\infty} |\eta - Tx| = 0$ from (3.3a) and therefore, $\lim_{t\to+\infty} |\hat{\ell} - \ell| = 0$ from (3.3b). On the other hand, the basic idea behind a functional observer is to construct an observer dynamics only driven by y able to reconstruct the vector ℓ without necessarily estimating the whole state x. Hence, the main challenge is to solve Problem 1 via observer dynamics with the minimum order possible. In this respect, in literature, there have been several attempts. Here, we give a short literature overview summarizing the main results dealing with establishing whether a triple (F, H, L_0) is functional observable (detectable) in the sense provided below.

The analysis of functional observers started mainly with the work of Watson Jr and Grigoriadis (1998). In this work, they take into consideration an autonomous system affected by noise as

$$\begin{aligned} \dot{x} &= Fx + G\nu \\ y &= Hx + J\nu. \end{aligned} \tag{3.4}$$

This work aims to solve a filtering problem via the estimation of a function $\ell = L_0 x$ from the noisy output y. In particular, they want to find an observer of dimensions n_ℓ whose dynamics is strictly proper, i.e.,

$$\begin{aligned} \dot{\eta} &= A\eta + By \\ \hat{\ell} &= \eta. \end{aligned} \tag{3.5}$$

The existence of an unbiased observer (3.5) is given if and only if

$$L_0 F \left(I - \begin{bmatrix} L_0 \\ H \end{bmatrix}^+ \begin{bmatrix} L_0 \\ H \end{bmatrix} \right) = 0. \tag{3.6}$$

In the end, they provide an $\mathcal{H}_\infty$ result aiming at reducing the $\mathcal{H}_\infty$ norm between the magnitude of noise ν and the norm of estimation error $e_\ell = \hat{\ell} - \ell$.
A first result taking into account as a core problem that of functional observation and the condition to characterize if a triple (F, H, L_0) is functionally observable[1] is given in Darouach (2000) and it can be summarized in the following theorem.

Theorem 3.1.1. *For system (F, H, L_0) in (3.1), there exists a functional observer (3.2) with (A, C) in observability form, if and only if the following two conditions hold*

(i) for the existence of A, B, D solving (3.3),

$$rank \begin{bmatrix} LF \\ L \\ HF \\ H \end{bmatrix} = rank \begin{bmatrix} L \\ HF \\ H \end{bmatrix}. \tag{3.7a}$$

(ii) to guarantee the stability of the error dynamics, namely A Hurwitz,

$$rank \begin{bmatrix} sL - LF \\ HF \\ H \end{bmatrix} = rank \begin{bmatrix} L \\ HF \\ H \end{bmatrix} \quad \forall s \in \mathbb{C}^+ \tag{3.7b}$$

Moreover, the spectrum of A can be assigned arbitrarily if (3.7b) is replaced by

$$rank \begin{bmatrix} sL - LF \\ HF \\ H \end{bmatrix} = rank \begin{bmatrix} L \\ HF \\ H \end{bmatrix} \quad \forall s \in \mathbb{C} \tag{3.7c}$$

In both condition (3.7a) and (3.7b), we consider a generic L instead of L_0 because in the observer design, one might need a dynamical system of order $n_\eta > n_q$. In this case, L will be of rank n_η and in general, it can be written as

$$L = \begin{bmatrix} L_0 \\ L_1 \end{bmatrix}. \tag{3.8}$$

In Moreno (2001), we have a first result that generalizes the standard definition of observability provided by the they provide an alternative result to test the Functional-Observability property of system (3.1), and this reads as follows

Theorem 3.1.2. *The triple (F, H, L_0) is Functional-Observable if and only if*

$$rank \begin{bmatrix} sI - F \\ H \\ L_0 \end{bmatrix} = rank \begin{bmatrix} sI - F \\ H \end{bmatrix}, \ \forall s \in \mathbb{C} \tag{3.9}$$

The equivalent condition for *Functional-Detectability* is given when condition (3.9) holds for any s in $\mathbb{C}^+$.
In Jennings et al. (2011) we first find a proper definition of functional-observability that we generalized into a geometric framework and extended to the case of functional detectability. In the following, for an autonomous system (M, N) we consider that the Observable space of an output matrix N is the Range $\mathcal{R}(N)$ of the observability matrix $\mathfrak{O}(M, N)$ associated to N. While the non-Observable space of N is the Kernel of the Observability matrix $\mathfrak{O}(M, N)$.

Definition 3. *[Functional-Observability] The triple (F, H, L_0) is Functional-Observable if the Observable space $\mathcal{R}(L_0)$ from L_0 is contained in the Observable space $\mathcal{R}(H)$ from H ($\mathcal{R}(L_0) \subseteq \mathcal{R}(H)$).*

In parallel, we consider the definition of Functional-Detectability.

Definition 4 (Functional-Detectability)**.** *The triple (F, H, L_0) is Functional-Detectable if the Observable sub-space $\mathcal{D}(L_0)$ from L_0 defined as $\mathcal{D}(L_0) = \mathcal{R}(L_0) \cap Ker(\mathfrak{O}(F, H))$ is contained in the region of attraction of the origin.*

[1] Here and in the following results we consider the definition of functional-observable (-detectable) according to the one provided in the relative work.

In other words, by considering the dynamical properties, in order to have Functional-Detectability of the pair (F, H, L_0) the Observable subspace from L_0, which is not in the Observable space from H, must be associated to an asymptotically stable dynamics.
Note that in these terms condition (3.7b) guarantees the properties of Functional-Detectability.
In their work, the authors provide an alternative representation of the result in Darouach (2000), as stated in the following

Theorem 3.1.3. *For system (3.1), for any Hurwitz A there exists a functional observer (3.2) with $C = [I\ 0]$, if and only if (3.7a) and*

$$rank \begin{bmatrix} sL - LF \\ HF \\ H \end{bmatrix} = rank \begin{bmatrix} L \\ HF \\ H \end{bmatrix}, \ \forall s \in \mathbb{C} \tag{3.10}$$

hold for some L_1, where L is as in (3.8).

Note that we already provide this condition in (3.7c) in Theorem 3.1.1 must hold. We only want to add that this property, which provides the possibility to have an arbitrary error convergence rate, has been only introduced in Jennings et al. (2011).
A parallel research thread has been autonomously developed in Kravaris (2016) and its main linear result can be synthesised as follows.

Theorem 3.1.4. *For a linear system (3.1) there exists a functional observer of the form (3.2) if and only if for $n_\eta > 0$ there exists a Hurwitz polynomial with coefficients α_i, $i = \{1, \dots, n_\eta\}$ such that for all $k = \{1, \dots, n_\ell\}$,*

$$L_{k,0}F^{n_\eta} + \alpha_1 L_{k,0}F^{n_\eta - 1} + \cdots + \alpha_r L_{k,0}F \in \ span\left(H_i,\ H_iF,\ \dots,\ H_iF^{n_\eta}\right)_{i=1\dots n_y}$$

where $L_{k,0}$ denotes the k-th line of L_0 and H_i is the i-th row of H.

In particular, the proof is constructive since they provide an explicit solution of the observer matrix, for the case $n_\ell = 1$, and it is interesting to notice that the coefficients β_i in $\mathbb{R}^{n_\ell \times n_y}$, $i = \{0, \dots, n_\eta\}$, defining the linear combination of the ℓ time derivatives, are exactly the coefficients of the observer zero dynamics. Indeed, if the following holds

$$L_0F^{n_\eta} + \alpha_1 L_0F^{n_\eta - 1} + \cdots + \alpha_r L_0 = \beta_0 HF^{n_\eta} + \beta_1 HF^{n_\eta - 1} + \cdots + \beta_{n_\eta - 1}HF + \beta_{n_\eta}H \tag{3.11}$$

the observer matrices are given in the standard observability canonical form, with A in companion form, C in prime form, and B and D as follows

$$\begin{aligned} A &= \begin{bmatrix} 0 & 0 & \dots & 0 & -\alpha_{n_\eta} \\ 1 & 0 & \dots & 0 & -\alpha_{n_\eta - 1} \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & 1 & -\alpha_1 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_{n_\eta} - \alpha_{n_\eta}\beta_0 \\ \beta_{n_\eta - 1} - \alpha_{n_\eta - 1}\beta_0 \\ \vdots \\ \beta_1 - \alpha_1\beta_0 \end{bmatrix} \\ C &= \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \qquad D = \beta_0. \end{aligned} \tag{3.12}$$

3.2 A unifying approach

In the following, we propose an approach to unify all the presented results into a single more intuitive framework. In particular, we first prove an equivalence among all the above-stated conditions for functional-detectability. We then, provide the equivalence among the functional-observability conditions.

Theorem 3.2.1 (Functional-Detectability)**.** *The following statements are equivalent :*

(i) There exists a matrix $L_1 \in \mathbb{R}^{(n_\eta - n_\ell) \times n}$ such that $L = [L_0^T, L_1^T]^T$ verifies conditions (3.7a) and (3.7b).

(ii)

$$rank \begin{bmatrix} sI - F \\ H \\ L_0 \end{bmatrix} = rank \begin{bmatrix} sI - F \\ H \end{bmatrix}, \ \forall s \in \mathbb{C}^+. \tag{3.13}$$

(iii) *there exists* $n_\eta \in \mathbb{N}_{>0}$ *and a Hurwitz polynomial* $\lambda^{n_\eta} + \alpha_1 \lambda^{n_\eta - 1} + \cdots + \alpha_{n_\eta - 1}\lambda + \alpha_{n_\eta}$ *such that for all* $k = \{1, \ldots, n_\ell\}$,

$$L_{k,0}F^{n_\eta} + \alpha_1 L_{k,0}F^{n_\eta - 1} + \cdots + \alpha_r L_{k,0} \in \ \mathit{span}\left(H_i,\ H_iF,\ \ldots,\ H_iF^{n_\eta}\right)_{i=\{1,\ldots,p\}}$$

where $L_{k,0}$ *denotes the* k*-th line of* L_0 *and* H_i *is the* i*-th row of* H.

(iv) *There exists* $n_\eta \in \mathbb{N}_{>0}$ *and a Hurwitz matrix* $A \in \mathbb{R}^{n_\eta \times n_\eta}$ *such that there exists a functional observer of the form* (3.2).

(v) *The triple* (F, H, L_0) *is Functional-Detectable.*

Proof. The equivalence between (i) and (iv) has been originally proved by Darouach (2000) for the case $L = L_0$ and it is easily extendable by considering $L = [L_0^T, L_1^T]^T$ and taking in (3.2) $C = [I_{n_\ell}, 0_{n_\ell \times (n_\ell - n_\eta)}]$.
Condition (ii) has been proved to be equivalent to (iv) in (Moreno, 2001, Theorem 3).
Condition (iii) has been proved to be equivalent to (iv) in Kravaris (2016).
It thus follows that (i),(ii),(iii), and (iv) are equivalent and it is enough to prove (ii) is equivalent to (v) to conclude the proof.

We define $\bar{n}_x$ the rank of the observability matrix $\mathcal{O}(F, H) = [H^T, (HF)^T, \ldots, (HF^{n_x - 1})^T]^T$ with n the dimension of the square matrix F. We define the normal form change of coordinates according to the ordered observability index $n_1, \ldots, n_p$, such that in the new coordinates, the triple (F, H, L_0) reads as

$$\begin{aligned} F &= \begin{bmatrix} F_o & 0 \\ F_\star & F_{no} \end{bmatrix} \\ H &= \begin{bmatrix} \bar{H} & 0 \end{bmatrix} \\ L_0 &= \begin{bmatrix} L_{01} & L_{02} \end{bmatrix} \end{aligned} \tag{3.14}$$

where $(F_o, \bar{H})$ is a companion form of dimension $\bar{n}$. We note that the first $\bar{n}$ components of the state represent the observable subspace from H, while the rest of the state describes the non-observable subspace from H. We then apply the same decomposition to the pair (F_{no}, L_{02}), in order to highlight the remaining observable space from L_0, thus leading to the normal form

$$\begin{aligned} F_{no} &= \begin{bmatrix} F_{\text{oL}} & 0 \\ F_{\star L} & F_{noL} \end{bmatrix} \\ L_{02} &= \begin{bmatrix} \bar{L}_{02} & 0 \end{bmatrix} \end{aligned} \tag{3.15}$$

where $(F_{oL}, \bar{L}_{02})$ is a pair in companion form and has dimension equal to the rank n_{oL} of the observability matrix $\mathcal{O}(F_{no}, L_{02})$. In those coordinates, we thus have

$$\begin{aligned} F &= \begin{bmatrix} F_o & 0 & 0 \\ F_{\star oL} & F_{oL} & 0 \\ F_{\star noL} & F_{\star L} & F_{no} \end{bmatrix} \\ H &= \begin{bmatrix} \bar{H} & 0 & 0 \end{bmatrix} \\ L_0 &= \begin{bmatrix} L_{01} & \bar{L}_{02} & 0 \end{bmatrix} \end{aligned} \tag{3.16}$$

Now, let us take into account the following matrices rank

$$\begin{aligned} \operatorname{rank} \begin{bmatrix} F - sI \\ H \\ L_0 \end{bmatrix} &= \operatorname{rank} \begin{bmatrix} F_o - sI & \vdots & 0 \\ F_\star & \vdots & F_{no} - s \\ \bar{H} & \vdots & 0 \\ L_{01} & \vdots & L_{02} \end{bmatrix} \\ &= \bar{n} + \operatorname{rank} \begin{bmatrix} F_{no} - s \\ L_{02} \end{bmatrix}, \ \forall s \in \mathbb{C} \end{aligned} \tag{3.17}$$

because the pair $(F_o, \bar{H})$ is completely observable and

$$\begin{aligned} \operatorname{rank} \begin{bmatrix} F - sI \\ H \end{bmatrix} &= \operatorname{rank} \begin{bmatrix} F_o - sI & \vdots & 0 \\ F_{\star\star} & \vdots & F_{no} - s \\ \bar{H} & \vdots & 0 \end{bmatrix} \\ &= \bar{n} + \operatorname{rank} \begin{bmatrix} F_{no} - s \end{bmatrix}, \ \forall s \in \mathbb{C} \end{aligned} \tag{3.18}$$

again because the pair $(F_o, \bar{H})$ is completely observable.
Now, to prove the sufficiency (ii) $\implies$ (v) we have

$$\text{rank}\begin{bmatrix} F_{no} - s \\ L_{02} \end{bmatrix} = \text{rank}\left[F_{no} - s\right], \ \forall s \in \mathbb{C}^+ \tag{3.19}$$

or explicitly $\forall s \in \mathbb{C}^+$

$$\text{rank}\begin{bmatrix} F_{oL} - s & 0 \\ F_{\star L} & F_{noL} - s \\ \bar{L}_{02} & 0 \end{bmatrix} = \text{rank}\begin{bmatrix} F_{oL} - s & 0 \\ F_{\star L} & F_{noL} - s \end{bmatrix} \quad .$$

Because the pair $(F_{oL}, \bar{L}_{02})$ is completely observable we have $\forall s$ that

$$\text{rank}\begin{bmatrix} F_{oL} - s \\ \bar{L}_{02} \end{bmatrix} = n_{oL}.$$

Thus if (ii) holds then

$$n_{oL} = \text{rank}\left[F_{oL} - s\right], \ \forall s \in \mathbb{C}^+. \tag{3.20}$$

which implies the Functional-Detectability property.
To prove the necessity part we go by contradiction: Not (v) $\implies$ Not (i).
We separately study the rank of the two matrices

$$\text{rank}\begin{bmatrix} F_{no} - s \\ L_{02} \end{bmatrix} = \text{rank}\begin{bmatrix} F_{oL} - s & 0 \\ F_{\star L} & F_{noL} - s \\ \bar{L}_{02} & 0 \end{bmatrix} \tag{3.21}$$

and

$$\text{rank}\left[F_{no} - s\right] = \text{rank}\begin{bmatrix} F_{oL} - s & 0 \\ F_{\star L} & F_{noL} - s \end{bmatrix} \tag{3.22}$$

We then have (3.21) is equal to $n_{oL} + \text{rank}[F_{noL} - s]$ for any s because the pair $(F_{oL}, \bar{L}_{o2})$ is completely observable. On the other hand, in (3.22) we have for any s.

$$\text{rank}\begin{bmatrix} F_{oL} - s & 0 \\ F_{\star L} & F_{noL} - s \end{bmatrix} = \text{rank}[F_{oL} - s] + \text{rank}[F_{noL} - s]. \tag{3.23}$$

If the triple (F, H, L_0) is not functional detectable it means that $\text{rank}[F_{oL} - s] < n_{oL}$ for some s in $\mathbb{C}^+$ and thus condition (ii) can not hold for any s in $\mathbb{C}^+$. This concludes the proof. □

For the sake of proof completeness, in the next subsections, we also provide the equivalences between conditions (i) and (v), and between (iii) and (v).
Note that the last part of the proof of Th.(3.2.1) is constructive, and can be exploited in building a possibly minimal order functional observer up.

Theorem 3.2.2 (Functional-Observability)**.** *The following statements are equivalent :*

(i) There exists a matrix $L_1 \in \mathbb{R}^{(n_\eta - n_\ell) \times n}$ such that $L = [L_0^T, L_1^T]^T$ verifies conditions (3.7a) and (3.7c).

(ii)

$$rank\begin{bmatrix} sI - F \\ H \\ L_0 \end{bmatrix} = rank\begin{bmatrix} sI - F \\ H \end{bmatrix}, \ \forall s \in \mathbb{C}. \tag{3.24}$$

(iii) there exists $n_\eta \in \mathbb{N}_{>0}$ such that, for any Hurwitz polynomial $\lambda^{n_\eta} + \alpha_1 \lambda^{n_\eta - 1} + \cdots + \alpha_{n_\eta - 1}\lambda + \alpha_{n_\eta}$, the following holds for $k = \{1, \ldots, n_\ell\}$,

$$L_{k,0}F^{n_\eta} + \alpha_1 L_{k,0}F^{n_\eta - 1} + \cdots + \alpha_r L_{k,0}F \in \ span\left(H_i,\ H_iF,\ \ldots,\ H_iF^{n_\eta}\right)_{i=\{1,\ldots,n_y\}} \tag{3.25}$$

where $L_{k,0}$ denotes the k-th line of L_0 and H_i is the i-th row of H.

(iv) There exists $n_\eta \in \mathbb{N}_{>0}$ such that for any Hurwitz matrix $A \in \mathbb{R}^{n_\eta \times n_\eta}$, there exists a functional observer of the form (3.2).

(v) The triple (F, H, L_0) is Functional-Observable.

Proof. The equivalence between Condition (i) and (iv) has been proved in Darouach (2000), while the equivalence between Condition (ii) and (iv) has been proved in (Moreno, 2001, Th.3). In Jennings et al. (2011) has been proved that, in this form, (ii) is equivalent to (i).
By applying the proof of Kravaris (2016) for any Hurwitz matrix A we have the equivalence between (iii) and (iv).
It thus follows that (i),(ii),(iii) and (iv) are equivalent.
To prove that (ii) is equivalent to (v), and thus conclude the proof, we can follow the steps in the proof of Theorem3.2.1, note that the functional-detectability conditions become functional-observability the term $L_{02} = 0$. □

3.2.1 Equivalence between (i) and (v) of Th.(3.2.1)

We now prove that (i) is equivalent to (v). By exploiting the coordinates in (3.14) and (3.15), we take a L_1 matrix such that

$$\begin{aligned} L = \begin{bmatrix} L_0 \\ L_1 \end{bmatrix} &= \left[\begin{array}{ccc} L_{01} & \bar{L}_{02} & 0 \\ L_{11} & \bar{L}_{12} & 0 \end{array}\right] \\ &= \left[\begin{array}{ccc} L_1' & L_2' & 0 \end{array}\right] \end{aligned} \tag{3.26}$$

where $[L_1' \; L_2']$ is any matrix that makes (3.7a) to hold. Note that the zero columns in L are mandatory because those columns are not in the observable space from L_{02} by construction. We can moreover notice that in order for (3.7a) to hold it must be that L_2' is full column rank. This is because $(F_{oL}, \bar{L}_{02})$ is an observable pair and it can be put in observability form. If L_2' is not full column rank, via manipulation of the columns we can write it of the form

$$L_2' = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

and $L_2' F_{oL}$ will add a linearly independent columns, thus not satisfying condition (3.7a). Indeed, by explicitly writing the condition we have

$$\begin{aligned} \operatorname{rank} \begin{bmatrix} LF \\ L \\ HF \\ H \end{bmatrix} = \operatorname{rank} \begin{bmatrix} L_1' F_o + L_2' F_{\star L} & L_2' F_{oL} \\ L_1' & L_2' \\ \bar{H} F_o & 0 \\ \bar{H} & 0 \end{bmatrix} = \\ \operatorname{rank} \begin{bmatrix} L \\ HF \\ H \end{bmatrix} = \operatorname{rank} \begin{bmatrix} L_1' & L_2' \\ \bar{H} F_o & 0 \\ \bar{H} & 0 \end{bmatrix} \end{aligned} \tag{3.27}$$

If L_2' is not full column rank it implies that $L_2' F_{oL}$ has a linearly independent column from L_2' and thus the rank condition can never be satisfied because

$$\operatorname{rank} \begin{bmatrix} L_2' F_{oL} \\ L_2' \end{bmatrix} > \operatorname{rank} \left[L_2' \right].$$

Hence, with loss of generality, we can consider $L_2' = [I_{n_{oL}}, 0]^T$.
Now in order to prove sufficiency ((i) $\implies$ (v)) we have that (3.7b) holds. Hence

$$\begin{aligned} \operatorname{rank} \begin{bmatrix} L(F - sI) \\ HF \\ H \end{bmatrix} = \operatorname{rank} \begin{bmatrix} L_1'(F_o - sI) + \begin{bmatrix} I \\ 0 \end{bmatrix} F_{\star oL} & \begin{bmatrix} I \\ 0 \end{bmatrix} (F_{oL} - s) & 0 \\ \bar{H} F_o & 0 & 0 \\ \bar{H} & 0 & 0 \end{bmatrix} = \\ \operatorname{rank} \begin{bmatrix} L_1'(F_o - sI) + \begin{bmatrix} I \\ 0 \end{bmatrix} F_{\star oL} \\ \bar{H} F_o \\ \bar{H} \end{bmatrix} + \operatorname{rank} \left[F_{oL} - s \right] = \\ \operatorname{rank} \begin{bmatrix} L_1' & L_2' \\ \bar{H} F_o & 0 \\ \bar{H} & 0 \end{bmatrix} = \operatorname{rank} \begin{bmatrix} L_1' \\ \bar{H} F_o \\ \bar{H} \end{bmatrix} + \operatorname{rank} \left[L_2' \right]. \end{aligned} \tag{3.28}$$

Hence, condition (i) implies that for any s in $\mathbb{C}^+$

$$\operatorname{rank}\begin{bmatrix} L_1'(F_o - sI) + \begin{bmatrix} I \\ 0 \end{bmatrix} F_{\star oL} \\ \bar{H}F_o \\ \bar{H} \end{bmatrix} = \operatorname{rank}\begin{bmatrix} L_1' \\ \bar{H}F_o \\ \bar{H} \end{bmatrix} \tag{3.29}$$

and

$$\operatorname{rank}\begin{bmatrix} F_{oL} - s \end{bmatrix} = \operatorname{rank}\begin{bmatrix} L_2' \end{bmatrix} = n_{oL}. \tag{3.30}$$

Because F_{oL} has dimension n_{oL} this implies that F_{oL} has a spectrum inside the left half plane, hence is Hurwitz.
We prove by contraction the necessary part: in particular if F_{oL} is not Hurwitz, hence, the triple (F, H, L_0) is not functional detectable we have that for some s in $\mathbb{C}^+$

$$\operatorname{rank}\begin{bmatrix} F_{oL} - s \end{bmatrix} < \operatorname{rank}\begin{bmatrix} L_2' \end{bmatrix} = n_{oL} \tag{3.31}$$

And thus, for any L_1' satisfying

$$\operatorname{rank}\begin{bmatrix} L_1'(F_o - sI) + \begin{bmatrix} I \\ 0 \end{bmatrix} F_{\star oL} \\ \bar{H}F_o \\ \bar{H} \end{bmatrix} = \operatorname{rank}\begin{bmatrix} L_1' \\ \bar{H}F_o \\ \bar{H} \end{bmatrix} \tag{3.32}$$

it is not possible to satisfy (3.7b). And this proves the necessity part.

3.2.2 Equivalence between (iii) and (v) of Th.(3.2.1)

We now want to prove that (iii) is equivalent to (v). Thus for the necessary part, assume Functional-Detectability of the triple (F, H, L_0) and consider a Hurwitz polynomial $p(\lambda)$ of degree n_η, not defined yet. Because of the lower block triangular structure of F in the new coordinates, the k-th power of F will read as

$$F^k = \begin{bmatrix} F_o^k & 0 & 0 \\ \star & F_{oL}^k & 0 \\ \star & \star & F_{noL}^k \end{bmatrix} \tag{3.33}$$

where $\star$ are elements of no interest in this proof. Then, by applying polynomial $p(\lambda)$ to F we have

$$p(F) = \begin{bmatrix} p(F_o) & 0 & 0 \\ \star & p(F_{oL}) & 0 \\ \star & \star & p(F_{noL}) \end{bmatrix} \tag{3.34}$$

by pre-multiplying by $L_0 = \begin{bmatrix} L_{01} & \bar{L}_{02} & 0 \end{bmatrix}$ leads to

$$L_0 p(F) = \begin{bmatrix} \star\star & \vdots \bar{L}_{02}\, p(F_{oL}) & \vdots 0 \end{bmatrix} \tag{3.35}$$

where the $\star\star$ term, according to the adopted coordinates, only depends on H and HF^i. In order to satisfy (iii), in order for each line of $L_0 p(F)$ to be in the span of $(H_i, H_i F, \ldots, H_i F^{n_\eta})_{i=1\ldots p}$, it must hold that the term $\bar{L}_{02} p(F_{oL}) = 0$. Now we prove necessity, because F_{oL} is Hurwitz, by F-detectability property, and we can construct the polynomial $p(\lambda) = p'(\lambda) p_{oL}(\lambda)$, with p' an arbitrary Hurwitz polynomial and p_{oL} the minimal polynomial of F_{oL}. We thus prove the existence of a Hurwitz polynomial of degree n_η, in this case, $n_\eta \geq n_{oL}$, that satisfies (iii).

To now prove sufficiency, we assume the existence of n_η and the related Hurwitz polynomial that satisfies condition (iii). We consider three non-trivial scenarios in which the condition $\bar{L}_{02} p(F_{oL}) = 0$ holds or more explicitly

$$\alpha_0 \bar{L}_{k,02} F_{oL}^{n_\eta} + \alpha_1 \bar{L}_{k,02} F_{oL}^{n_\eta - 1} + \cdots + \alpha_r \bar{L}_{k,02} = 0 \tag{3.36}$$

where $\bar{L}_{k,02}$ denotes the k-th line of $\bar{L}_{02}$ and without loss of generality, $(F_{oL}, \bar{L}_{02})$ is in observability canonical form.
The first scenario is the one in which F_{oL} has all decoupled blocks with the same characteristic polynomial and dimensions. In this case, in order to satisfy (3.36), p must necessarily be of the form $p(\lambda) = p'(\lambda) p_{oL}(\lambda)$ with p_{oL} being the characteristic polynomial of F_{oL} and p' any other arbitrary Hurwitz polynomial that brings the degree of p to be larger or equal to $n_p - 1$, with n_p the maximum observability index associated to H. This implies that F_{oL} is Hurwitz.

The second scenario is when F_{oL} has j_{oL} decouple blocks with different dimensions and characteristic polynomials. Then p necessarily must be of the form $p = p'_j p_j$, $\forall j = 1 \dots j_{oL}$, with p_j being the characteristic polynomial of the j-th block. Because p is a Hurwitz polynomial, it implies that all blocks in F_{oL} have Hurwitz components. The last scenario is the more general case in which F_{oL} has all couple blocks with different dimensions. Without loss of generality, we prove this scenario in the parametric case in which F_{oL} and $\bar{L}_{02}$ are of the form

$$\begin{aligned} F_{oL} &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\beta_0 & -\beta_1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\gamma_0 & -\gamma_1 \end{bmatrix} \\ \bar{L}_{02} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \end{aligned} \tag{3.37}$$

For $n_\eta = 2$ and $\bar{L}_{1,02}$, condition (3.36) reads as

$$\begin{aligned} \alpha_2 \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} + \alpha_1 \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} + \alpha_0 \begin{bmatrix} -\beta_0 & -\beta_1 & 1 & 0 \end{bmatrix} = \\ \begin{bmatrix} \alpha_2 - \beta_0\alpha_0 & \alpha_1 - \beta_1\alpha_0 & \alpha_0 & 0 \end{bmatrix} = 0. \end{aligned} \tag{3.38}$$

That is $\alpha_0 = 0$ which implies $\alpha_2 = \alpha_1 = 0$. We thus necessarily have $n_\eta \geq 4$. Thus, with $n_\eta = 4$ condition (3.36) reads as

$$\begin{aligned} \alpha_4 \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} -\beta_0 & -\beta_1 & 1 & 0 \end{bmatrix} + \alpha_1 \begin{bmatrix} \beta_1\beta_0 & -\beta_0 + \beta_1^2 & -\beta_1 & 1 \end{bmatrix} + \\ \alpha_0 \begin{bmatrix} (\beta_0 - \beta_1^2)\beta_0 & \beta_1\beta_0 + (\beta_0 - \beta_1^2)\beta_1 & -\beta_0 + \beta_1^2 - \alpha_0 & -\beta_1 - \gamma_1 \end{bmatrix} = 0. \end{aligned} \tag{3.39}$$

By considering, without loss of generality, $\alpha_0 = 1$ as scaling factor, the coefficient of p satisfying (3.39) must have solution

$$\begin{aligned} \alpha_1 &= \beta_1 + \gamma_1 \\ \alpha_2 &= \gamma_0 + \beta_0 + \beta_1\gamma_1 \\ \alpha_3 &= \gamma_0\beta_1 + \beta_0\gamma_1 \\ \alpha_4 &= \gamma_0\beta_0. \end{aligned} \tag{3.40}$$

The characteristic polynomial of F_{oL} reads as

$$p_{oL}(\lambda) = \lambda^4 + (\beta_1 + \gamma_1)\lambda^3 + (\gamma_0 + \beta_0 + \beta_1\gamma_1)\lambda^2 + (\gamma_0\beta_1 + \beta_0\gamma_1)\lambda + \gamma_0\beta_0\beta_0.$$

Hence in this case the polynomial p corresponds to the characteristic polynomial of F_{oL}. This implies that F_{oL} is Hurwitz. And thus that the triple (F, H, L_0) is functional detectable. The same procedure applies for any other case leading to the fact that necessarily $p(\cdot) = p'(\cdot)p_{oL}(\cdot)$ thus containing the characteristic polynomial of F_{oL}. And this implies functional detectability.
We thus proved the sufficiency.

3.3 Conclusions

We provide a unified overview of the works related to functional observers for linear systems. We also provide the geometrical definition of both functional-observability and -detectability. We proved the equivalence of these works and we provide a change of coordinates that makes the equivalence proofs constructive. Then, exploiting just a change of coordinates one can design an algorithm to construct an L_1 wide matrix for (3.8) such that the resulting observer dynamics has the smallest dimension and its converge rate can be arbitrary chosen.

Chapter 4

Nonlinear case

In this chapter, we want to extend as much as possible the linear results obtained in the previous chapter to the case of nonlinear systems.
Given a nonlinear autonomous dynamical system with state $x(t)$ and output measurements $y(t)$, and some functional of the system state $\ell(x(t))$, the basic idea of a functional observer is to construct a dynamical system driven by the only signal available, i.e., the output trajectory $y(t)$, to asymptotically reconstruct the functional $\ell(x(t))$, possibly, without estimating the whole state $x(t)$. In general, for nonlinear systems, reducing the functional-observer order is a much harder task than for a linear system and, as for the linear case, we will not address it.

4.1 Problem statement

Consider an autonomous system

$$\dot{x} = f(x) \quad , \quad y = h(x) \tag{4.1}$$

with state $x \in \mathbb{R}^{n_x}$, output $y \in \mathbb{R}^{n_y}$ and maps $f : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ and $h : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ continuously differentiable. The goal of this chapter is to investigate the possibility of reconstructing, from the measurement y, a certain continuous function $\ell : \mathbb{R}^{n_x} \to \mathbb{R}^{n_q}$ of the state x, namely design a *functional observer* processing y and providing asymptotically an estimate of $\ell(x)$, for any solution of (4.1) initialized in some set $\mathcal{X}_0 \subseteq \mathbb{R}^{n_x}$ of interest. In the following, we consider a set $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ such that any solution to (4.1) initialized in $\mathcal{X}_0$ remains in $\mathcal{X}$ for all positive times.
To this end, we build upon the literature of nonlinear Luenberger observers, also called KKL observers, and propose a functional observer of the form

$$\begin{aligned} \dot{\hat{\eta}} &= A\hat{\eta} + B(y) \\ \hat{\ell} &= \tau(\hat{\eta}) \end{aligned} \tag{4.2}$$

of appropriate dimension n_η, with $A \in \mathbb{R}^{n_\eta \times n_\eta}$ Hurwitz, $B : \mathbb{R}^{n_y} \to \mathbb{R}^{n_\eta}$ and $\tau : \mathbb{R}^{n_\eta} \to \mathbb{R}^{n_q}$ to be designed such that, for any solution $t \mapsto x(t)$ to (4.1) initialized in $\mathcal{X}_0$, any solution to (4.2) with input $y = h(x)$ verifies

$$\lim_{t\to\infty} |\hat{\ell}(t) - \ell(x(t))| = 0 \ . \tag{4.3}$$

We then say that (4.2) is a *ℓ-functional observer* for (4.1) initialized in $\mathcal{X}_0$.
The idea followed in the KKL methodology is to look for a C^1 map $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_\eta}$ transforming the dynamics (4.1) into

$$\dot{\eta} = A\eta + B(y) \quad , \quad \tau(\eta) = \ell(x) \tag{4.4}$$

namely such that

$$\frac{\partial T}{\partial x}(x) f(x) = AT(x) + B\big(h(x)\big) \tag{4.5a}$$

$$\tau \circ T(x) = \ell(x) \qquad\qquad \forall x \in \mathcal{X} \ . \tag{4.5b}$$

Indeed, if (4.5) is verified, then for any solution to (4.1) initialized in $\mathcal{X}_0$, the image $\eta = T(x)$ is solution to (4.4), and because A is Hurwitz, any solution to (4.2) verifies

$$\lim_{t\to\infty} |\hat{\eta}(t) - T(x(t))| = 0 \ . \tag{4.6}$$

It then follows by applying τ and using (4.5b), that the convergence property (4.3) holds if the map τ verifies an appropriate uniform continuity condition.

Note that if $\ell(x) = x$, namely a full-state observer is required, then we recover the paradigm of Andrieu and Praly (2006). On the other hand, for an arbitrary map ℓ, the spirit of the approach is the same as the one proposed in Kravaris (2016). However, the main difference lies in the fact that the map τ in (4.5) is allowed to be nonlinear while it is taken linear in Kravaris (2016). This allows us to prove the existence of a functional observer for a larger class of systems verifying a very general distinguishability property with respect to the map q (see Remark 3 below) The price to pay is that the obtained result deals only with existence and not constructive design.

4.2 Robust functional KKL observer

4.2.1 Main result

In Andrieu and Praly (2006), the existence of a full-state observer of the type (4.2) with $\ell = \mathrm{Id}$ is shown under a backward distinguishability assumption on the full state x. In this chapter, we relax this assumption into backward distinguishability of $\ell(x)$ only as follows.

Definition 5. *System (4.1) is backward $\mathfrak{O}$-distinguishable with respect to ℓ if there exist $\delta_d > 0$ and $\delta_\Upsilon > 0$, with $\delta_d > \delta_\Upsilon$, such that for each pair (x_a, x_b) in $(\mathfrak{O} + \delta_\Upsilon)^2$ verifying $\ell(x_a) \neq \ell(x_b)$, there exist a time $t \in (\max\{\sigma^-_{\mathfrak{O}+\delta_d}(x_a), \sigma^-_{\mathfrak{O}+\delta_d}(x_b)\}, 0]$, such that*

$$h(X(x_a, t)) \neq h(X(x_b, t)).$$

A checkable sufficient condition for backwards-distinguishability is the so-called *differential observability*, namely the fact that the map

$$x \mapsto (h(x), L_f h(x), \ldots, L_f^{(m)} h(x))$$

made of the output and its successive time derivatives, is injective on $\mathcal{O} + \delta_\Upsilon$ for some integer m. Indeed, this means that the outputs from two distinct initial conditions (x_a, x_b) in $\mathcal{O} + \delta_\Upsilon$ can instantaneously be distinguished and δ_d is any positive scalar.
We then prove our main result.

Theorem 4.2.1. *Assume $\mathcal{X}$ is compact and system (4.1) is backward $\mathfrak{O}$-distinguishable with respect to ℓ, with $\mathfrak{O}$ an open bounded set such that $\mathcal{X} \subseteq cl(\mathfrak{O})$. Then there exist $\tau : \mathbb{R}^{n_\eta} \to \mathbb{R}^{n_v}$, $A \in \mathbb{R}^{n_\eta \times n_\eta}$ Hurwitz and $B : \mathbb{R}^{n_y} \to \mathbb{R}^{n_\eta}$ with $n_\eta = 2(n+1)n_y$ such that (4.2) is a ℓ-functional observer of (4.1) initialized in $\mathcal{X}_0$. Moreover, there exists a class-$\mathcal{K}$ map α such that if (4.2) is implemented with $y = h(x) + \nu$ then*

$$\limsup_{t \to +\infty} |\hat{z}(t) - \ell(x(t))| \leq \alpha \left(\limsup_{t \to +\infty} |\nu(t)| \right). \tag{4.7}$$

More precisely, there exists a subset S of $\mathbb{C}^{n+1}$ of zero Lebesgue measure and $\ell < 0$ such that, denoting $\mathbb{C}_\ell := \{\lambda \in \mathbb{C} : \mathfrak{Re}\{\{\}\lambda\} < \ell\}$, A can be chosen as a block diagonal matrix[1] $A = \mathrm{diag}\big(I_{n_y} \otimes A_1, \ldots, I_{n_y} \otimes A_{n+1}\big)$, with

$$A_i = \begin{bmatrix} \mathfrak{Re}\{\lambda_i\} & \mathfrak{Im}\{\lambda_i\} \\ -\mathfrak{Im}\{\lambda_i\} & \mathfrak{Re}\{\lambda_i\} \end{bmatrix} \tag{4.8}$$

where $(\lambda_1, \ldots, \lambda_{n+1})$ is arbitrarily chosen in $\mathbb{C}_\ell^{n+1} \backslash S$, and

$$B(y) = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n+1} \otimes \left(y \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right). \tag{4.9}$$

Remark 3. *In Kravaris (2016), τ is imposed to be linear (and so is B), modulo a possible linear dependence of ℓ on $y = h(x)$, i.e., (4.5b) is replaced by $\ell(x) = CT(x) + Dh(x)$ for some matrices C and D to be designed. In this case, a q-functional observer exists if and only if a certain Hurwitz polynomial p of degree n_η applied to the Lie derivatives of the scalar functional q can be written as a linear combination of the outputs y_i and their n_η Lie derivatives for $i = 1, \ldots, n_y$. Besides, this equivalence is constructive, because when such a polynomial p exists, A, B, C and D can be designed as shown in Kravaris (2016), with p taken as the characteristic polynomial of A and n_η the observer's dimension. This result holds because applying p to the Lie derivatives of $q(x) = CT(x) + Dh(x)$ makes the dependence on T disappear thanks to the linearity of q with respect to T and Cayley-Hamilton theorem. In this chapter, because*

[1] The matrix A and linear map B as described here are a real realization of the complex variable dynamics $\dot{z}_{kj} = \lambda_k z_{kj} + y_j$ for $k = 1, \ldots, n+1$ and $j = 1, \ldots, n_y$.

τ is a general nonlinear function this condition is no longer necessary. In addition, relaxing it to a nonlinear dependence of the Lie derivatives of q on the Lie derivatives of h is not straightforward due to the nonlinearity of τ.

Remark 4. *Even when $n_q < n$, the dimension n_η of the observer given by Theorem 4.2.2 is the same as if we were designing a full-state KKL observer with $\ell(x) = x$. This is due to the use of Coron's Lemma Andrieu and Praly (2006) in the proof and does not allow the design of minimal/reduced order observers unlike Darouach and Fernando (2019); Fernando and Trinh (2014); Kravaris (2016).However, this observer can be used even when the full state is not observable, by extracting from η all the backwards-distinguishable states (or functions of state), through an appropriate design of τ.*

The rest of the section shows how to adapt the proof of Andrieu and Praly (2006) to prove Theorem 4.2.1.

4.2.2 Existence of T injective with respect to ℓ solving (4.5a)

The existence of a map $T : \mathcal{O} \subset \mathbb{R}^n \to \mathbb{R}^{n_\eta}$ satisfying (4.5a) has been studied in Andrieu and Praly (2006). A candidate solution is indeed

$$T(x) = \int_{-\infty}^{0} e^{-As} B\big(h(\breve{X}(x,s))\big) ds \tag{4.10}$$

with $\breve{X}(x,t)$ solution at time t of

$$\dot{x} = \chi(x) f(x) \tag{4.11}$$

passing through x at time $t = 0$, for an arbitrary locally Lipschitz function χ satisfying

$$\chi(x) = \begin{cases} 1 \text{ if } & x \in \mathfrak{O} + \delta_d \\ 0 \text{ if } & x \notin \mathfrak{O} + \delta_c \end{cases} \tag{4.12}$$

with $\delta_c > \delta_d$. The heart of the problem is rather in guaranteeing the existence of τ such that (4.5b) holds, namely the fact that $q(x)$ can be written as a function of $T(x)$. A necessary condition is that T be *injective with respect to* q on $\mathcal{X}$, namely that for any $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$T(x_a) = T(x_b) \implies q(x_a) = q(x_b) \ . \tag{4.13}$$

The following theorem generalizes (Andrieu and Praly, 2006, Theorem 3) (in the case of a bounded set $\mathcal{O}$) and gives a sufficient condition for T defined in (4.10) to be C^1 and injective with respect to ℓ.

Theorem 4.2.2. *Assume $\mathcal{X}$ is compact and system (4.1) is backward $\mathfrak{O}$-distinguishable with respect to q with corresponding δ_d in $(0, \delta_c)$. Then, there exists a subset S of $\mathbb{C}^{(n+1)}$ of zero Lebesgue measure such that the function $T : cl(\mathfrak{O}) \to \mathbb{R}^{2(n+1)\times n_y}$ defined by (4.10) is C^1 and injective with respect to q in the sense of (4.13), provided A is a block diagonal matrix $A = diag\big(I_{n_y} \otimes A_1, \dots, I_{n_y} \otimes A_{n+1}\big)$, with A_i defined in (4.8) and $(\lambda_1, \dots, \lambda_{n+1})$ arbitrarily chosen in $\mathbb{C}_\mu^{n+1} \backslash S$, with $\mathbb{C}_\mu = \{\lambda \in \mathbb{C} : \mathfrak{Re}\{\lambda\} < \mu\}$, and $B : \mathbb{R}^{n_y} \mapsto \mathbb{R}^{2(n+1)n_y}$ defined in (4.9).*

Proof. The proof of the theorem follows the same steps as in (Andrieu and Praly, 2006, Theorem 3), but with the set

$$\Upsilon := \{(x_a, x_b) \in (\mathfrak{O} + \delta_\Upsilon)^2 : x_a \neq x_b\}$$

replaced by

$$\Upsilon = \{(x_a, x_b) \in (\mathfrak{O} + \delta_\Upsilon)^2 : q(x_a) \neq q(x_b)\} \ . \tag{4.14}$$

Indeed, Υ is an open set by continuity of q. Define for $\lambda \in \mathcal{C}_\mu$

$$T_\lambda(x) = \int_{-\infty}^{0} e^{-\lambda s} h(\breve{X}(x,s)) ds \ .$$

which is shown in Andrieu and Praly (2006) to be C^1 on $\mathcal{O} + \delta_\Upsilon$ for any $\lambda \in \mathcal{C}_\mu$. Since T defined in (4.10) is built from the real and imaginary parts of $(T_{\lambda_1}, T_{\lambda_2}, \cdots, T_{\lambda_{n+1}})$, it is C^1 on $\mathrm{cl}(\mathcal{O})$. The injectivity of T with respect to q is then proved by applying Coron's lemma (Andrieu and Praly, 2006, Lemma 1) to $g : \Upsilon \times \mathcal{C}_\mu \to \mathcal{C}^{n_y}$ defined by

$$g(x_a, x_b, \lambda) = T_\lambda(x_a) - T_\lambda(x_b) \ ,$$

with Υ defined in (4.14). g is indeed holomorphic with respect to λ for all $(x_a, x_b) \in \Upsilon$ in the same way as in Andrieu and Praly (2006). Besides, for all $(x_a, x_b) \in \Upsilon$, $\lambda \mapsto g(x_a, x_b, \lambda)$ is not identically zero on $\mathcal{C}_\mu$ because for any $a < \ell$, applying Plancherel theorem,

$$\int_{-\infty}^{+\infty} |g(x_a, x_b, a + is)|^2 ds = \int_{-\infty}^{0} e^{-2\lambda s} |h(\breve{X}(x_a, s)) - h(\breve{X}(x_b, s))|^2 ds > 0$$

by backwards-distinguishability with respect to q, continuity in time and injectivity of b. It follows that (Andrieu and Praly, 2006, Lemma 1) applies and the set

$$S = \bigcup_{(x_a, x_b) \in \Upsilon} \{(\lambda_1, \cdots, \lambda_{n+1}) \in \mathcal{C}_\mu^{n+1} \ : g(x_a, x_b, \lambda_1) = \ldots = g(x_a, x_b, \lambda_{n+1}) = 0\}$$

has zero Lebesgue measure in $\mathcal{C}_\mu^{n+1}$. By definition of g and Υ, we conclude that for any $(\lambda_1, \cdots, \lambda_{n+1}) \in \mathcal{C}_\mu^{n+1} \setminus S$, the map $(T_{\lambda_1}, T_{\lambda_2}, \cdots, T_{\lambda_{n+1}})$, and therefore T, is injective with respect to q on $\mathrm{cl}(\mathfrak{O}) \subset \mathcal{O} + \delta_\Upsilon$. □

4.2.3 Existence of τ solving (4.5b)

Now that injectivity of T with respect to ℓ is guaranteed, we prove the existence of a globally defined uniformly continuous map τ verifying (4.5b). This is guaranteed by (Bernard, 2019, Lemma A.12) since $\mathcal{X}$ is compact. More precisely, there exists a map $\tau : \mathbb{R}^{n_\eta} \to \mathbb{R}^{n_q}$ and a class-$\mathcal{K}$ map γ such that (4.5b) holds and

$$|\tau(\eta_a) - \tau(\eta_b)| \leq \gamma(|\eta_a - \eta_b|) \qquad \forall (\eta_a, \eta_b) \in \mathbb{R}^{n_\eta} \times \mathbb{R}^{n_\eta}. \tag{4.15}$$

Hence, for any $x \in \mathcal{X}$ and $\hat{\eta} \in \mathbb{R}^{n_\eta}$,

$$|\tau(T(x)) - \tau(\hat{\eta})| \leq \gamma(|T(x) - \hat{\eta}|) \tag{4.16}$$

and thus according to (4.5b),

$$|q(x) - \tau(\hat{\eta})| \leq \gamma(|T(x) - \hat{\eta}|) \ . \tag{4.17}$$

Finally, using the fact that A is Hurwitz and that, according to (4.5a),

$$\frac{d}{dt}(T(x) - \hat{\eta}) = A(T(x) - \hat{\eta}) \ ,$$

we deduce that (4.3) holds and the first part of Theorem 4.2.1 is proved.

4.2.4 Robustness of the functional observer

We now study the effect of measurement additive noise on the functional observer (4.2), namely when $y = h(x) + \nu$. The error $\Delta\eta = \hat{\eta} - T(x)$ now verifies

$$\frac{d}{dt}\Delta\eta = A\Delta\eta + B(\nu) \tag{4.18}$$

and

$$|\hat{z} - \ell(x)| = |\tau(\hat{\eta}) - \tau(T(x))| \leq \gamma(|\hat{\eta} - T(x)|) = \gamma(|\Delta\eta|)$$

where $\gamma(\cdot)$ is the uniform continuity map of τ verifying (4.15). Let us define a Lyapunov function $V(\Delta\eta) = \Delta\eta^T P \Delta\eta$, with P positive definite, such that $PA + A^T P \leq -aP$ for some $a > 0$. Then

$$\begin{aligned} \dot{V} &\leq -aV + 2\Delta\eta^T P B(\nu) \\ &\leq -aV + \frac{a}{2}V + \frac{2}{a}B(\nu)^T P B(\nu) \\ &\leq -\frac{a}{2}V + \frac{2}{a}\beta|\nu|^2\lambda_{\max}(P) \end{aligned} \tag{4.19}$$

for some $\beta > 0$ depending only on n_y and n_η. It follows by standard ISS arguments that asymptotically,

$$\limsup_{t \to +\infty} |\Delta\eta(t)| \leq 2\sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}\frac{\beta}{a}\limsup_{t \to +\infty} |\nu(t)|^2 \tag{4.20}$$

and thus, we obtain the following asymptotic error property

$$\limsup_{t \to +\infty} |\hat{z} - q(x)| \leq \gamma\Big(2\sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}\frac{\beta}{a}\limsup_{t \to +\infty} |\nu|\Big). \tag{4.21}$$

4.3 Application: observer design for systems with input

Consider a system

$$\dot{x} = f(x,u) \quad , \quad y = h(x,u) \tag{4.22}$$

with state $x \in \mathbb{R}^n$, output $y \in \mathbb{R}^{n_y}$ and input $u \in \mathbb{R}^{n_u}$ in a certain class $\mathbb{U}$. We are interested in designing an observer for solutions initialized in $\mathcal{X}_0$ with input $u \in \mathbb{U}$ and remaining in a compact set $\mathcal{X}$. The goal of this observer may be for instance to reconstruct the full-state x, or a certain function of the state $\ell(x)$, or even the input u. Apart from specific structures allowing the design to be valid for any $u \in \mathbb{U}$, typically under uniform observability assumptions (linear forms Luenberger (1964), triangular forms Gauthier et al. (1992), see also (Bernard and Andrieu, 2018, Theorem 4)) or particular classes of systems Astolfi et al. (2010), there does not exist a general observer design paradigm for systems with input. Here we propose to use the functional KKL observers presented in the previous section to provide a general answer to this problem when the input is known in advance to be generated by a finite-dimensional system, or to be approximable by one, in a sense to be defined. In this latter case, the robustness proved in Theorem 4.2.2 is instrumental to allow the use of universal approximators and ensure practical estimation as detailed in Section 4.3.1. Then, we detail in particular two contexts:

- observer to reconstruct x or $\ell(x)$ when u is known: u can then be considered as an extra measurement and the observer fed with $y_{aug} = (u, y)$, see Section 4.3.2;
- functional and/or input observer to reconstruct $x, \ell(x)$, and/or u, when the latter is unknown: the observer is only fed with y, see Section 4.3.3.

4.3.1 Finite-dimensional input generator

Let us start by assuming that the input u is known to be generated, in forward time, by a finite-dimensional dynamical system of the form

$$\dot{w} = s(w) \quad , \quad u = l(w) \tag{4.23}$$

with $s : \mathbb{R}^{n_w} \to \mathbb{R}^{n_w}$ and $l : \mathbb{R}^{n_w} \to \mathbb{R}^{n_u}$ continuously differentiable, w initialized in $\mathcal{W}_0 \subset \mathbb{R}^{n_w}$ such that any solution initialized in $\mathcal{W}_0$ remains in a compact set $\mathcal{W}$ for all $t \geq 0$. This applies well in particular in the fields of electrical machines or automotive engines, where the inputs are typically periodic with a finite number of Fourier coefficients as in Chauvin et al. (2007). More precisely, defining

$$\mathbb{W} = \left\{ w : \mathbb{R}_{\geq 0} \to \mathcal{W} \text{ solution to } \dot{w} = s(w) \text{ on } \mathbb{R}_{\geq 0}, \text{ with } w(0) \in \mathcal{W}_0 \right\}$$

we assume the input u is known to belong to the class

$$\mathbb{U}_{\text{gen}} = \{ l(w) \ : \ w \in \mathbb{W} \} \ . \tag{4.24}$$

The idea is then to apply off-the-shelf the functional KKL design of the previous section on the autonomous cascade (4.23)-(4.22) for an appropriate map q_{aug}, in order to obtain an observer (4.2) for (4.22), with a certain choice of A, B, τ that works for any input $u \in \mathbb{U}_{\text{gen}}$. To appropriately define and analyse the observer, we need to consider a larger class of inputs. So, similarly as above, for $0 < \delta' < \delta$ to be defined, consider a C^1 map χ_δ satisfying

$$\chi_\delta(w) = \begin{cases} 1 \text{ if } & w \in \mathcal{W} \\ 0 \text{ if } & w \notin \mathcal{W} + \delta \end{cases} \tag{4.25}$$

and the input class

$$\begin{aligned} \mathbb{U}^\delta_{\text{gen}} = \Big\{ u : & \mathbb{R} \to \mathbb{R}^{n_u} \text{ such that } \exists w : \mathbb{R} \to \mathcal{W} + \delta, \text{ solution to } \dot{w} = \chi_\delta(w) s(w), \text{ with } w(0) \in \mathcal{W} + \delta' \ , \\ & \text{such that } u(t) = l(w(t)) \ \forall \ t \in \mathbb{R} \Big\} \ . \end{aligned} \tag{4.26}$$

Note that $\mathbb{U}_{\text{gen}} \subset \mathbb{U}^\delta_{\text{gen}}$.
In a second step, once we have designed an observer (4.2) for inputs $u \in \mathbb{U}_{\text{gen}}$, one may wonder whether the designed observer may still be used when the input $u \in \mathbb{U}$ is not in $\mathbb{U}_{\text{gen}}$. Thanks to the robustness property of the observer described in Theorem 4.2.1, the accuracy of the input generator model drives the steady-state estimation of the observer. Indeed, consider an arbitrary $u \in \mathbb{U}$, $x_0 \in \mathcal{X}_0$ and the corresponding solution x to (4.22) with output y. Then, for any $u_w = l(w) \in \mathbb{U}_{\text{gen}}$, with $w \in \mathbb{W}$, and

corresponding x_w solution to (4.22) with input u_w and output y_w, we can consider the observer (4.2) driven by y (and possibly u, see below) as driven by $y_w + \nu_w^y$ (and $u_w + \nu_w^u$) where

$$\begin{aligned}\nu_w^u &= u - u_w = u - l(w)\\ \nu_w^y &= y - y_w = y - h(x_w, l(w))\end{aligned}$$

Hence, applying Theorem 4.2.1 to the cascade (4.23)-(4.22), we deduce a result of the type

$$\limsup_{t\to+\infty} |\ell(x(t)) - \hat{z}(t)| \leq \inf_{w\in\mathbb{W}} \limsup_{t\to\infty} \{|\ell(x(t)) - \ell(x_w(t))| + \alpha\left(|(\nu_w^u(t), \nu_w^y(t))|\right)\} \tag{4.27}$$

with the quality of the estimation thus depending on how well u can be approximated by a signal $u_w \in \mathbb{U}_{\text{gen}}$ for $w \in \mathbb{W}$, and how far the corresponding $y_w, q(x_w)$ are from $y, q(x)$.
A natural idea is therefore to use in (4.23) a universal approximator of sufficiently large dimension making ν_u and ν_y sufficiently small. For instance, any sinusoidal signal of arbitrary frequency $u(t) = A\sin(\omega t + \phi)$ may be modelled via a nonlinear three-dimensional system

$$\begin{cases}\dot{w}_1 = w_2\\ \dot{w}_2 = -w_3 w_1\\ \dot{w}_3 = 0\end{cases}, \quad l(w) = w_1$$

and several such exosystems could be combined to approximate arbitrarily well any periodic signal.
In the following, we thus provide conditions ensuring the existence of asymptotic observers in the two scenarios mentioned above (known/unknown inputs) only for inputs $u \in \mathbb{U}_{\text{gen}}$. One may rely on robustness as detailed above when using the observer for an arbitrary $u \in \mathbb{U}$.

4.3.2 Observer with known input

According to (Bernard and Andrieu, 2018, Theorem 3), we already know that, under an appropriate backwards-distinguishability condition of the full state x for inputs $u \in \mathbb{U}$, a KKL observer exists for (4.22) with time-varying maps T_u depending on each individual $u \in \mathbb{U}$. Although this dependence on u is causal, it is not explicit in general, which renders the use of such an observer limited in practice unless T_u can be explicitly computed as done in some examples in Bernard and Andrieu (2018). In addition, the zero-measure set S_u outside of which the eigenvalues of A must be chosen depends on each $u \in \mathbb{U}$ and $\bigcup_{u\in\mathbb{U}} S_u$ may not be of zero-measure. Therefore, the existence of a single KKL observer working for any $u \in \mathbb{U}$ is not guaranteed. Our goal here is to use the functional observer paradigm presented in the previous section to prove the existence of a KKL functional observer relevant for at least any $u \in \mathbb{U}_{\text{gen}}$, where $\mathbb{U}_{\text{gen}}$ can be modelled by (4.24).
We start by giving a generic definition of backwards-distinguishability for systems with inputs.

Definition 6. *System (4.22) is backward $\mathfrak{O}$-distinguishable with respect to q for inputs in $\mathbb{U}$, if there exist $0 < \delta_\Upsilon < \delta_d$ such that for any pair (x_a, x_b) in $(\mathfrak{O} + \delta_\Upsilon)^2$ with $q(x_a) \neq q(x_b)$, and any $u \in \mathbb{U}$, there exists $t \in (\max\{\sigma^-_{\mathfrak{O}+\delta_d}(x_a, 0; u), \sigma^-_{\mathfrak{O}+\delta_d}(x_b, 0; u)\}, 0]$ such that*

$$h(X_u(x_a, 0, t), u(t)) \neq h(X_u(x_b, 0, t), u(t)) \ .$$

Remarking that both y and u are known, the idea is to consider u as an extra measurement and look for a functional observer of the form

$$\begin{aligned}\dot{\hat{\eta}} &= A\hat{\eta} + B\left(y, u\right)\\ \hat{z} &= \tau(\hat{\eta})\end{aligned} \tag{4.28}$$

with A, B, τ chosen such that for any solution to (4.22) initialized in $\mathcal{X}_0$ with input $u \in \mathbb{U}_{\text{gen}}$ and output y, any solution to (4.28) verifies

$$\lim_{t\to\infty} |\hat{z} - q(x)| = 0 \ .$$

Because $u \in \mathbb{U}_{\text{gen}}$, we design A, B, τ by applying the functional KKL paradigm on the extended system

$$\begin{cases}\dot{w} &= \chi_\delta(w)s(w)\\ \dot{x} &= f(x, l(w))\end{cases}, \ y_{\text{aug}} = (h(x, l(w)), l(w)) \tag{4.29}$$

in a way that makes

$$\begin{aligned}\dot{\hat{\eta}} &= A\hat{\eta} + B\left(y_{\text{aug}}\right)\\ \hat{z} &= \tau(\hat{\eta})\end{aligned} \tag{4.30}$$

a q_{aug}-functional observer for (4.29) with

$$q_{\text{aug}}(w, x) = \ell(x) \ .$$

Theorem 4.3.1. *Assume $\mathcal{X}$ is compact and that there exist $0 < \delta' < \delta$ such that system (4.22) is backward $\mathfrak{O}_x$-distinguishable with respect to a continuous map q for inputs in $\mathbb{U}_{gen}^{\delta}$ defined in (4.26), with $\mathfrak{O}_x$ a bounded open set such that $\mathcal{X} \subseteq cl(\mathfrak{O}_x)$. Then, there exist $\ell > 0$ and a set S of zero-Lebesgue measure in $\mathbb{C}^{(n_w+n+1)}$ such that there exists a map $\tau : \mathbb{R}^{2(n_w+n+1)(n_y+n_u)} \to \mathbb{R}^n$ such that (4.28) is a q-functional observer for (4.22) for any input $u \in \mathbb{U}_{gen}$ defined in (4.24), provided A is a block diagonal matrix $A = diag\big(I_{n_y+n_u} \otimes A_1, \ldots, I_{n_y+n_u} \otimes A_{n_w+n+1}\big)$, with A_i defined as in (4.8) and $(\lambda_1, \ldots, \lambda_{n_w+n+1})$ arbitrarily chosen in $\mathbb{C}_{\ell}^{n_w+n+1} \backslash S$, with $\mathbb{C}_l$ as in Theorem 4.2.2, and B defined in (4.9).*

Proof. For any solution x to (4.22) initialized in $\mathcal{X}_0$ with input $u \in \mathbb{U}_{\text{gen}}$ and output y, there exists a solution (w, x) to (4.29) initialized in $\mathbb{W}_0 \times \mathcal{X}_0$ such that $y_{\text{aug}} = (y, u)$. Therefore, it is sufficient to show that (4.30) is a q_{aug}-functional observer for (4.29) initialized in $\mathcal{W}_0 \times \mathcal{X}_0$ with respect to the continuous map $q_{\text{aug}}(w, x) = q(x)$. For that, we would like to apply Theorem 4.2.1. We know that solutions to (4.29) initialized in $\mathcal{W}_0 \times \mathcal{X}_0$ remain in forward time in the compact set $\mathcal{W} \times \mathcal{X}$, and we only need to find an open bounded set $\mathfrak{O}_w$ such that $\mathcal{W} \times \mathcal{X} \subseteq \text{cl}(\mathfrak{O}_w \times \mathfrak{O}_x)$ and (4.29) is backward $\mathfrak{O}_w \times \mathfrak{O}_x$-distinguishable with respect to q_{aug}.
Let us consider an open bounded set $\mathfrak{O}_w$, $\delta_{\Upsilon_w} > 0$ and $\delta_{d_w} > 0$ such that $\mathcal{W} \subset \mathfrak{O}_w$, $\mathfrak{O}_w + \delta_{\Upsilon_w} \subset \mathcal{W} + \delta'$ and $\mathcal{W} + \delta \subset \mathfrak{O}_w + \delta_{d_w}$.
Consider $\delta_{\Upsilon_x} < \delta_{d_x}$ given by the property of $\mathfrak{O}_x$-distinguishability. Consider a pair $(w_a, w_b) \in (\mathfrak{O}_w + \delta_{\Upsilon_w})^2$, a pair $(x_a, x_b) \in (\mathfrak{O}_x + \delta_{d_x})^2$ such that $q(x_a) \neq q(x_b)$, and the corresponding solutions $t \mapsto (W((w_i, x_i), t), X((w_i, x_i), t))$ to (4.29) for $i = a, b$. Because $t \mapsto W((w_i, x_i), t)$ does not depend on x_i, we actually write $t \mapsto W(w_i, t)$. By definition of χ_δ, $W(w_i, t) \in \mathcal{W} + \delta \subset \mathfrak{O}_w + \delta_{d_w}$ for all t and thus,

$$t_{d_x} := \max_{i=a,b} \ \sigma^-_{(\mathfrak{O}_w+\delta_{d_w})\times(\mathfrak{O}_x+\delta_{d_x})}(w_i, x_i)$$

verifies

$$t_{d_x} = \max_{i=a,b} \ \sigma^-_{\mathbb{R}^{n_w}\times(\mathfrak{O}_x+\delta_{d_x})}(w_i, x_i) \tag{4.31}$$

We have the following two cases

- either there is a time $t \in (t_{d_x}, 0]$ such that $l(W(w_a, t)) \neq l(W(w_b, t))$, in which case we trivially have
$$\begin{bmatrix} h(X((w_a, x_a), t), l(W(w_a, t)))) \\ l(W(w_a, t)) \end{bmatrix} \neq \begin{bmatrix} h(X((w_b, x_b), t), l(W(w_b, t)))) \\ l(W(w_b, t)) \end{bmatrix} \tag{4.32}$$
- or, $l(W(w_a, t)) = l(W(w_b, t))$ for all $t \in (t_{d_x}, 0]$. In this case, defining $u := l(W(w_a, \cdot))$, both $X((w_a, x_a), \cdot)$ and $X((w_b, x_b), \cdot)$ are solutions on $(t_{d_x}, 0]$ to (4.22) initialized in $\mathfrak{O}_x + \delta_{d_x}$ with $x_a \neq x_b$ and with input $u \in \mathbb{U}_{\text{gen}}^{\delta}$. Besides, according to (4.31),

  $$t_{d_x} = \max\{\sigma^-_{\mathfrak{O}_x+\delta_{d_x}}(x_a, 0; u), \sigma^-_{\mathfrak{O}_x+\delta_{d_x}}(x_b, 0; u)\} \ .$$

  By the property of backward $\mathfrak{O}_x$-distinguishability of (4.22) with respect to q, there exists $t \in (t_{d_x}, 0]$ such that
  $$h(X((w_a, x_a), t), u(t)) \neq h(X((w_b, x_b), t), u(t)) \ .$$

We conclude that in both cases there exists $t \in (t_{d_x}, 0]$ such that (4.32) holds. Therefore, (4.29) is backward $\mathfrak{O}_w \times \mathfrak{O}_x$-distinguishable with respect to q_{aug}. □

4.3.3 Observers with unknown input

Two possible applications of practical interest are the unknown input observers and the input reconstruction/observation problems.

Unknown input observer (UIO)

The goal of UIO design is to estimate the state of a system despite the presence of unknown inputs (disturbances) that are not required to be estimated. This problem is typically approached in the literature either (i) by cancelling the contribution of those inputs in the observer, through an appropriate change of coordinates and/or matrix manipulation Wang et al. (1975); Hou and Muller (1992); Chen and Saif (2006); Chakrabarty et al. (2017), or (ii) by transforming the system into a particular triangular form where robust sliding mode differentiators can be used Barbot et al. (2009) (among many others), or (iii) by using a stochastic model of the input and design a minimum-variance Kalman filter Maes et al. (2016); Azam et al. (2015).
Here, we rather propose to exploit the KKL functional paradigm of this chapter by assuming a finite-dimensional (deterministic) input generator is available, i.e., $u \in \mathbb{U}_{\text{gen}}$. In the UIO setting, the

input is unknown so the observer can only be fed with y. Therefore, we consider an observer of the form (4.2), with A, B, τ still designed so that (4.30) is a q_{aug}-functional observer for (4.29), but with

$$y_{\text{aug}} = h(x, l(w)) \quad , \quad q_{\text{aug}}(w, x) = \ell(x) \ .$$

Reproducing similar arguments as in Theorem 4.3.1, but with (4.32) replaced by

$$h(X((w_a, x_a), t), l(W(w_a, t)))) \neq h(X((w_b, x_b), t), l(W(w_b, t)))) \ ,$$

we show the existence of a ℓ-observer (4.2) for system (4.22) with $n_\eta = 2(n_w + n + 1)n_y$ for any input in $\mathbb{U}_{\text{gen}}$ if (4.22) is backward $\mathcal{O}_x$-distinguishable with respect to q for unknown inputs in $\mathbb{U}^\delta_{\text{gen}}$. In other words, Definition 6 must be adapted to the following.

Definition 7. *System (4.22) is backward $\mathfrak{O}$-distinguishable with respect to q for unknown inputs in $\mathbb{U}$, if there exist $0 < \delta_\Upsilon < \delta_d$ such that for any pair $(x_a, x_b) \in (\mathfrak{O} + \delta_\Upsilon)^2$ with $q(x_a) \neq q(x_b)$, and any pair $(u_a, u_b) \in \mathbb{U}^2$, there exists $t \in (\max\{\sigma^-_{\mathfrak{O}+\delta_d}(x_a, 0; u_a), \sigma^-_{\mathfrak{O}+\delta_d}(x_b, 0; u_b)\}, 0]$ such that*

$$h(X_{u_a}(x_a, 0, t), u_a(t)) \neq h(X_{u_b}(x_b, 0, t), u_b(t)) \ .$$

For a linear system

$$\dot{x} = Fx + Gu \ ,$$

standard linear UIO methodologies would aim at annihilating the contribution of Gu in the observer dynamics by using G^+ such that $G^+G = 0$.
In our approach instead, we relax the constraints by designing an observer able to generate asymptotically the unknown input term Gu, but only for signals u generated by a particular exosystem $\dot{w} = Sw$.
To whom is familiar with the concept of output regulation/tracking internal model, the idea is somehow to have in the observer dynamics an internal model of the part of the unknown input that is necessary to generate $\ell(x)$. In other words, the observer state η incorporates the "useful" effect of u through y, and the information about $\ell(x)$ is then extracted from η via τ thanks to distinguishability. We indeed want to highlight the similarity with this control field, in particular with Marconi et al. (2007).

Input reconstruction

Estimating the input of a system is of interest in practical applications in particular for fault diagnosis. A first approach to reconstruct the input from the output is by *inverting* the system's dynamics. This is the path taken in Szigeti et al. (2002); Edelmayer et al. (2004), where a *detector* is designed so that the cascade of the plant with its detector creates an identity map. Such an inversion requires observability of *any* input and thus that the system in question be minimum phase (i.e., its zero dynamics must be asymptotically stable) Hou and Patton (1998); Szigeti et al. (2002); Edelmayer et al. (2004). Other approaches requiring minimum-phase properties include Fridman et al. (2008), where sliding mode is applied on a particular triangular form, and Corless and Tu (1998) where a practical linear-based design is used on a system with a Lipschitz nonlinearity considered as unknown input.
Other approaches avoid the problem of non-minimum phase systems, by restricting the class of considered inputs, for instance to bounded inputs in Veluvolu and Soh (2009), or to periodic inputs with a finite number of harmonics in Chauvin et al. (2007). Our design also falls in this category since we consider the input reconstruction problem under the assumption that the input is bounded and generated by a finite-dimensional model, i.e., belongs to $\mathbb{U}_{\text{gen}}$. This does not require any minimum-phaseness property because inputs that are made indistinguishable by the unstable zero dynamics may not belong to the class of interest and may be discarded, thus recovering distinguishability. The price to pay is of course the need for an input generator model which augments the observer dimension. Anyway, still following the same arguments as before but this time with

$$y_{\text{aug}} = h(x, l(w)) \quad , \quad q_{\text{aug}}(w, x) = l(w) \ ,$$

we propose to design an input estimator of the form (4.2) such that

$$\lim_{t \to +\infty} |\hat{z}(t) - u(t)| = 0 \ ,$$

and we thus need the system (4.22) to verify the following input distinguishability property for inputs in $\mathbb{U}^\delta_{\text{gen}}$.

Definition 8. *The system (4.22) is backward input $\mathfrak{O}$-distinguishable for inputs in $\mathbb{U}$ if there exist positive real $\delta_d, \delta_\Upsilon$, with $\delta_d > \delta_\Upsilon$, such that for each pair (x_a, x_b) in $(\mathfrak{O} + \delta_\Upsilon)^2$ and (u_a, u_b) in $\mathbb{U}^2$ verifying $u_a(0) \neq u_b(0)$, there exists $t \in \max\{\sigma^-_{\mathfrak{O}+\delta_d}(x_a, 0; u_a), \sigma^-_{\mathfrak{O}+\delta_d}(x_b, 0; u_b)\}, 0]$ such that*

$$h(X_{u_a}(x_a, 0, t), u_a(t)) \neq h(X_{u_b}(x_b, 0, t), u_b(t)).$$

A similar definition (in forward time) is given in (Szigeti et al., 2002, Definition 1) under the name of input observability, but the main difference is that we only require this property for inputs in a certain class.

4.4 Conclusion

We have provided distinguishability conditions ensuring the existence of functional KKL observers for autonomous systems and for systems with inputs when the input is generated by a finite-dimensional autonomous dynamical system. Such observers consist of linear filters of the known signals (output, and input when it is known) and a nonlinear map enabling the reconstruction of the quantity of interest. The observer dimension is explicitly linked to the dimension of the system (and of the input generator). Moreover, the inherent robustness of the observer allows us to obtain practical estimation when the input is only approximately modelled by the input generator, although the observer dimension then increases with the generator dimension and thus with the required precision.
Further work includes developing numerical methods to implement such observers, extending for instance what is done in Ramos et al. (2020), as well as developing this functional KKL paradigm also in the time-varying case (extending the results of Bernard and Andrieu (2018)). Such results could indeed allow the use of more general time-varying input generators such as neural networks Salgado and Chairez (2017).

Conclusions

In our work, we presented a summary of the most relevant results in literature about the limitations in feedback design due to the presence of a non-minimum phase plant. In particular, we began with the Bode's integrals limitations appearing in the frequency domain. We then moved to the state space results in which we found lower bounds limitations in the output undershoot. Then, some approaches to construct a state feedback gain matrix that allowed no undershoot nor overshoot of the closed output trajectory, or another approach intending to stuck the output trajectory of a non-minimum phase system between two boundary functions of time, i.e., the funnel control approach. Despite the limitations we have found in the output trajectory, the most relevant theoretical work on the unstable zeros in linear system has been established via the Inner-Outer decomposition by Qiu and Davison (1993). In particular, they were able to directly link the location of such unstable zeros to a boundary of the minimum energy needed for the output trajectory in order to (unavoidably) stabilize the unstable zeros dynamics. This work has been recast into a path following scenario that allows to put into play a new design degree of freedom that brakes the limitations on the minimum energy needed for the output trajectory, leading to a arbitrary small, but not zero, output energy. These latter works has then by extended to the case of nonlinear systems. In particular, it is thanks to Seron et al. (1999) that we first we a bound between the Bode's integral results and the limitations described in Qiu and Davison (1993). They moreover extend to the case of nonlinear system the stabilization approach exploited in Qiu and Davison (1993) to stabilize square nonlinear systems with unitary vector relative degree.
Our contributions to the non-minimum phase research thread has been strongly inspired by the work of Qiu and Davison (1993) and provides an extension of the Inner-Outer decomposition to strictly proper linear systems. We then exploit this decomposition in a state feedback framework to impose an upper bound the maximum undershoot a non-minimum phase system can exhibit, by paying the price of imposing a limitation on the output time derivative. Moreover, the results we proved, in the context of dynamic output feedback, is that, by exploit a fast enough observer, we can steer the non-minimum phase system trajectory arbitrarily close to that of minimum phase one, by equivalently controlling a minimum phase 'twin' of the system under consideration enforcing a constraint on the minimum phase output time derivative.

Motivated by our results in the field of non-minimum phase systems stabilization via the Inner-Outer decomposition, we were interested in reconstructing the minimum phase (outer) output from its inner-filtered version, i.e., the real plant output y. In doing so we cannot simply invert the inner dynamics of the system because its inverse is an unstable system, thus the equivalent input reconstruction problem cannot be solved. We then wanted to exploit the outer-inner system cascade to construct a functional observer because the y_o can be viewed as functional of the cascade states. This idea led to the study of the functional observation literature and the most relevant approaches to construct such a functional observer. Apparently, these results seemed completely independent from each other, but, by exploiting a simple change of coordinates, we were able to proof the equivalence among all. Moreover, we also provide a generic definition functional-observability and -detectability for linear systems.
We then extended our work to the case of nonlinear systems by exploiting the so-called KKL observer approach. We also find interesting applications of the functional observer, such as: input reconstruction, unknown input observers, and controlled nonlinear systems. Unfortunately, due to the equivalent pole/zero cancellation present in the inner-outer decomposition its cascade becomes detectable and its state cannot satisfy the backward distinguishability property to construct the inverse map and obtain a reconstruction of the desired functional.

We are currently working on the extension of the inner-outer decomposition also to nonlinear systems admitting a normal form. Moreover, we are also focusing on different approaches to exploit the inner-outer decomposition in order to stabilize a non-minimum phase systems via dynamics output feedback.

Appendix A

Multi-Input Multi-Output Normal form

We consider the normal form of an LTI MIMO system

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{A.1}$$

with $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$, and $y \in \mathbb{R}^{n_y}$. We consider the C_i as the i-th row of matrix C, and we define the i-th relative degree r_i as the smallest integer satisfying

- $C_i B = \cdots = C_i A_i^{j-1} B = 0$ for all $j = 1, \ldots, r_i - 1$
- $C_i A_i^{r_i - 1} B \neq 0$.

We can thus construct a linear map T of the form

$$T = \begin{bmatrix} T_z \\ C_1 \\ C_1 A \\ \cdots \\ C_1 A^{r_1 - 1} \\ C_2 \\ \cdots \\ C_p A^{r_p - 1} \end{bmatrix} \tag{A.2}$$

where T_z is any wide matrix that makes T non-singular. Note that it is always possible to find such a T_z matrix with the propery that $T_z B = 0$. Map T defines a change of coordinates that puts the system in normal form

$$\begin{bmatrix} z \\ \xi \end{bmatrix} = Tx \tag{A.3}$$

with $z \in \mathbb{R}^{n_x - r}$ and $\xi \in \mathbb{R}^r$, being [1] $r = \sum_{i=1}^{n_y} r_i$. Then, system (A.1) in the normal form coordinates read as

$$\begin{aligned} \begin{bmatrix} \dot{z} \\ \dot{\xi} \end{bmatrix} &= TAT^{-1} \begin{bmatrix} z \\ \xi \end{bmatrix} + TBu \\ &= \begin{bmatrix} F & G \\ H & \bar{A} \end{bmatrix} \begin{bmatrix} z \\ \xi \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{B} \end{bmatrix} u \\ y &= CT^{-1} \begin{bmatrix} z \\ \xi \end{bmatrix} \\ &= \bar{C}\xi \end{aligned} \tag{A.4}$$

[1] In literature, we usually refer to r as the vector relative degree, see Isidori (2013). We believe that our definition of r generalizes the standard notion of relative degree given for SISO systems, as given in Isidori (2017) and we refer to $\bar{r}$ as the classical vector relative degree, i.e., $\bar{r} = \{r_1, r_2, \ldots, r_p\}$.

where the matrices $\bar{A}$, $\bar{B}$, and $\bar{C}$ are of the form

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & \star & \cdots & \star \\ \star & \bar{A}_2 & \cdots & \star \\ \cdots & \cdots & \ddots & \cdots \\ \star & \star & \cdots & \bar{A}_p \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & & & & \cdots & & & \cdots & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star & \star & \star & \star & \cdots & \star & & \star & \star & \star & \cdots & \star \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & & & & \cdots & & & \cdots & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & & 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star & \star & \star & \star & \cdots & \star & & \star & \star & \star & \cdots & \star \\ & & \cdots & & & & & \cdots & & & \ddots & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & & 0 & 0 & 1 & \cdots & 0 \\ & & \cdots & & & & & \cdots & & & \cdots & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 1 \\ \star & \star & \star & \cdots & \star & \star & \star & \star & \cdots & \star & & \star & \star & \star & \cdots & \star \end{bmatrix} \tag{A.5}$$

$$\bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \\ \vdots \\ \bar{B}_p \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star \end{bmatrix} \tag{A.6}$$

$$\bar{C} = \begin{bmatrix} \bar{C}_1 & \bar{C}_2 & \dots \bar{C}_p \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & & 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & & & & \cdots & & & \cdots & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & & 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{A.7}$$

where the matrix blocks $\bar{A}_i$, $\bar{B}_i$ and $\bar{C}_i$ are respectively in $\mathbb{R}^{r_i \times r_i}$, $\mathbb{R}^{r_i \times n_u}$ and $\mathbb{R}^{n_y \times r_i}$, with $\star$ possible nonzero elements.
Moreover, it is possible for square systems, see Mueller (2009), to find a normal form such as (A.4) where the matrices G and H have structure

$$G = \begin{bmatrix} \star & 0 & 0 & \cdots & 0 & \star & 0 & 0 & \cdots & 0 & & \star & 0 & 0 & \cdots & 0 \\ \star & 0 & 0 & \cdots & 0 & \star & 0 & 0 & \cdots & 0 & & \star & 0 & 0 & \cdots & 0 \\ & & \cdots & & & & & \cdots & & & \cdots & & & \cdots & & \\ \star & 0 & 0 & \cdots & 0 & \star & 0 & 0 & \cdots & 0 & & \star & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{A.8}$$

$$= \bar{G}\bar{C},$$

for some $\bar{G}$ and

$$H = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 \\ \star & \star & \star & \cdots & \star \end{bmatrix} \tag{A.9}$$

where the $\star$ symbols are possible nonzero elements.

Bibliography

A. P. Aguiar, J. P. Hespanha, and P. V. Kokotovic. Path-following for nonminimum phase systems removes performance limitations. *IEEE Transactions on Automatic Control*, 50(2):234–239, 2005.

A. P. Aguiar, J. P. Hespanha, and P. V. Kokotović. Performance limitations in reference tracking and path following for nonlinear systems. *Automatica*, 44(3):598–610, 2008.

B. Anderson. An algebraic solution to the spectral factorization problem. *IEEE Transactions on Automatic Control*, 12(4):410–414, 1967.

V. Andrieu and L. Praly. On the existence of a Kazantzis–Kravaris/Luenberger observer. *SIAM Journal on Control and Optimization*, 45(2):432–456, 2006.

A. Astolfi, R. Ortega, and A. Venkatraman. A globally exponentially convergent immersion and invariance speed observer for mechanical systems with non-holonomic constraints. *Automatica*, 46(1):182–189, 2010.

S. E. Azam, E. Chatzi, and C. Papadimitriou. A dual kalman filter approach for state estimation via output-only acceleration measurements. *Mechanical systems and signal processing*, 60:866–886, 2015.

J.-P. Barbot, D. Boutat, and T. Floquet. An observation algorithm for nonlinear systems with unknown inputs. *Automatica*, 45(8):1970–1974, 2009.

T. Berger. Tracking with prescribed performance for linear non-minimum phase systems. *Automatica*, 115:108909, 2020.

P. Bernard. Observer design for nonlinear systems. 2019.

P. Bernard and V. Andrieu. Luenberger observers for nonautonomous nonlinear systems. *IEEE Transactions on Automatic Control*, 64(1):270–281, 2018.

H. W. Bode et al. *Network analysis and feedback amplifier design*. van Nostrand, 1945.

A. M. Boker and H. K. Khalil. Semi-global output feedback stabilization of non-minimum phase nonlinear systems. *IEEE Transactions on Automatic Control*, 62(8):4005–4010, 2016.

A. Chakrabarty, M. J. Corless, G. T. Buzzard, S. H. Żak, and A. E. Rundell. State and unknown input observers for nonlinear systems with bounded exogenous inputs. *IEEE Transactions on Automatic Control*, 62(11):5497–5510, 2017. doi: 10.1109/TAC.2017.2681520.

J. Chauvin, G. Corde, N. Petit, and P. Rouchon. Periodic input estimation for linear periodic systems: Automotive engine applications. *Automatica*, 43(6):971–980, 2007.

T. Chen and B. A. Francis. Spectral and inner-outer factorizations of rational matrices. *SIAM Journal on Matrix Analysis and Applications*, 10(1):1–17, 1989.

W. Chen and M. Saif. Unknown input observer design for a class of nonlinear systems: an LMI approach. pages 5–pp, 2006.

M. Corless and J. Tu. State and input estimation for a class of uncertain systems. *Automatica*, 34(6): 757–764, 1998.

M. Darouach. Existence and design of functional observers for linear systems. *IEEE Transactions on Automatic Control*, 45(5):940–943, 2000.

M. Darouach and T. Fernando. On the existence and design of functional observers. *IEEE Transactions on Automatic Control*, 65(6):2751–2759, 2019.

E. Davison. The robust control of a servomechanism problem for linear time-invariant multivariable systems. *IEEE transactions on Automatic Control*, 21(1):25–34, 1976.

A. Edelmayer, J. Bokor, Z. Szabó, and F. Szigeti. Input reconstruction by means of system inversion: A geometric approach to fault detection and isolation in nonlinear systems. *International Journal of Applied Mathematics and Computer Science*, 14:189–199, 2004.

A. Emami-Naeini and D. de Roover. Bode's sensitivity integral constraints: The waterbed effect revisited. *arXiv preprint arXiv:1902.11302*, 2019.

T. Fernando and H. Trinh. A system decomposition approach to the design of functional observers. *International Journal of Control*, 87(9):1846–1860, 2014.

B. A. Francis. *A course in H_∞ control theory*, volume 88. 1987.

B. A. Francis and W. M. Wonham. The internal model principle of control theory. *Automatica*, 12(5): 457–465, 1976.

J. Freudenberg and D. Looze. Sensitivity reduction, nonminimum phase zeros, and design tradeoffs in single loop feedback systems. In *The 22nd IEEE Conference on Decision and Control*, pages 625–630. IEEE, 1983.

J. Freudenberg and D. Looze. Right half plane poles and zeros and design tradeoffs in feedback systems. *IEEE transactions on automatic control*, 30(6):555–565, 1985.

L. Fridman, Y. Shtessel, C. Edwards, and X.-G. Yan. Higher-order sliding-mode observer for state estimation and input reconstruction in nonlinear systems. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, 18(4-5):399–412, 2008.

J.-P. Gauthier, H. Hammouri, and S. Othman. A simple observer for nonlinear systems applications to bioreactors. *IEEE Transactions on automatic control*, 37(6):875–880, 1992.

K. Glover. All optimal hankel-norm approximations of linear multivariable systems and their l_∞-error bounds. *International journal of control*, 39(6):1115–1193, 1984.

G. C. Goodwin, S. F. Graebe, M. E. Salgado, et al. *Control system design*, volume 240. Prentice Hall Upper Saddle River, 2001.

G. Gu. Inner-outer factorization for strictly proper transfer matrices. *IEEE transactions on automatic control*, 47(11):1915–1919, 2002.

R. Gurumoorthy and S. Sanders. Controlling non-minimum phase nonlinear systems-the inverted pendulum on a cart example. In *1993 American Control Conference*, pages 680–685. IEEE, 1993.

J. B. Hoagg and D. S. Bernstein. Nonminimum-phase zeros-much to do about nothing-classical control-revisited part ii. *IEEE Control Systems Magazine*, 27(3):45–57, 2007.

M. Hou and P. Muller. Design of observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 37(6):871–875, 1992.

M. Hou and R. J. Patton. Input observability and input reconstruction. *Automatica*, 34(6):789–794, 1998.

A. Isidori. A tool for semi-global stabilization of uncertain non-minimum-phase nonlinear systems via output feedback. *IEEE transactions on automatic control*, 45(10):1817–1827, 2000.

A. Isidori. *Nonlinear control systems*. Springer Science & Business Media, 2013.

A. Isidori. *Lectures in feedback design for multivariable systems*. Springer, 2017.

A. Isidori and C. I. Byrnes. Output regulation of nonlinear systems. *IEEE transactions on Automatic Control*, 35(2):131–140, 1990.

L. S. Jennings, T. L. Fernando, and H. M. Trinh. Existence conditions for functional observability from an eigenspace perspective. *IEEE transactions on automatic control*, 56(12):2957–2961, 2011.

C. Kravaris. Functional observers for nonlinear systems. *IFAC-PapersOnLine*, 49(18):505–510, 2016.

K. Lau, R. H. Middleton, and J. H. Braslavsky. Undershoot and settling time tradeoffs for nonminimum phase systems. *IEEE Transactions on automatic control*, 48(8):1389–1393, 2003.

D. G. Luenberger. Observing the state of a linear system. *IEEE transactions on military electronics*, 8 (2):74–80, 1964.

K. Maes, A. Smyth, G. De Roeck, and G. Lombaert. Joint input-state estimation in structural dynamics. *Mechanical Systems and Signal Processing*, 70:445–466, 2016.

L. Marconi, L. Praly, and A. Isidori. Output stabilization via nonlinear Luenberger observers. *SIAM Journal on Control and Optimization*, 45(6):2277–2298, 2007.

R. H. Middleton. Trade-offs in linear control system design. *Automatica*, 27(2):281–292, 1991.

B. C. Moore. On the flexibility offered by state feedback in multivariable systems beyond closed loop eigenvalue assignment. In *1975 IEEE Conference on Decision and Control including the 14th Symposium on Adaptive Processes*, pages 207–214. IEEE, 1975.

J. Moreno. Quasi-unknown input observers for linear systems. In *Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01)(Cat. No. 01CH37204)*, pages 732–737. IEEE, 2001.

M. Mueller. Normal form for linear systems with respect to its vector relative degree. *Linear algebra and its applications*, 430(4):1292–1312, 2009.

S. Nazrulla and H. K. Khalil. Robust stabilization of non-minimum phase nonlinear systems using extended high-gain observers. *IEEE Transactions on Automatic Control*, 56(4):802–813, 2010.

L. Qiu and E. J. Davison. Performance limitations of non-minimum phase systems in the servomechanism problem. *Automatica*, 29(2):337–349, 1993.

L. Ramos, F. Di Meglio, V. Morgenthaler, L. Silva, and P. Bernard. Numerical design of Luenberger observers for nonlinear systems. *IEEE Conference on Decision and Control*, pages 5435–5442, 12 2020. doi: 10.1109/CDC42340.2020.9304163.

H. Rosenbrock. The zeros of a system. *International Journal of Control*, 18(2):297–299, 1973.

I. Salgado and I. Chairez. Adaptive unknown input estimation by sliding modes and differential neural network observer. *IEEE transactions on neural networks and learning systems*, 29(8):3499–3509, 2017.

R. Schmid and L. Ntogramatzidis. The design of nonovershooting and nonundershooting multivariable state feedback tracking controllers. *Systems & Control Letters*, 61(6):714–722, 2012.

R. Sepulchre, M. Jankovic, and P. V. Kokotovic. *Constructive nonlinear control*. Springer Science & Business Media, 2012.

M. M. Seron, J. H. Braslavsky, P. V. Kokotovic, and D. Q. Mayne. Feedback limitations in nonlinear systems: From bode integrals to cheap control. *IEEE Transactions on Automatic Control*, 44(4): 829–833, 1999.

J. Stewart and D. E. Davison. On overshoot and nonminimum phase zeros. *IEEE Transactions on Automatic Control*, 51(8):1378–1382, 2006.

F. Szigeti, J. Bokor, and A. Edelmayer. Input reconstruction by means of system inversion: application to fault detection and isolation. *IFAC Proceedings Volumes*, 35(1):13–18, 2002.

K. C. Veluvolu and Y. C. Soh. High-gain observers with sliding mode for state and unknown input estimations. *IEEE Transactions on Industrial Electronics*, 56(9):3386–3393, 2009.

M. Vidyasagar. On undershoot and nonminimum phase zeros. *IEEE Transactions on Automatic Control*, 31(5):440–440, 1986.

S.-H. Wang, E. Wang, and P. Dorato. Observing the states of systems with unmeasurable disturbances. *IEEE Transactions on Automatic Control*, 20(5):716–717, 1975.

J. T. Watson Jr and K. M. Grigoriadis. Optimal unbiased filtering via linear matrix inequalities. *Systems & Control Letters*, 35(2):111–118, 1998.

List of Figures