Geometric contributions to the study of LPV and Takagi-Sugeno systems

Gustave Bainier

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Contributions géométriques à l'étude des systèmes LPV et de Takagi-Sugeno

THÈSE

présentée et soutenue publiquement le 2 decembre 2024

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention génie informatique, automatique et traitement du signal)

par

Gustave Bainier

Composition du jury

Président :	Gabriela Iuliana Bara	Professeur à l'Université de Strasbourg
Rapporteurs :	Kévin Guelton	Professeur à l'Université de Reims Champagne-Ardenne
	Olivier Sename	Professeur à Grenoble INP (UGA)
Examinateurs :	Dalil Ichalal	Professeur à l'Université d'Evry (Paris Saclay)
	Jamal Daafouz	Professeur à l'Université de Lorraine (UL)
	Jean-Christophe Ponsart	Professeur à l'UL (directeur de thèse)
	Benoît Marx	Maître de conférences HDR à l'UL (co-dir. de thèse)

Centre de Recherche en Automatique de Nancy — UMR-CNRS 7039
2, avenue de la forêt de Haye 54516 Vandœuvre-lès-Nancy

Remerciements

Les travaux présentés dans ce manuscrit ont été effectués au sein du Centre de Recherche en Automatique de Nancy (CRAN), au sein du département Contrôle Identification Diagnostic (CID), sous la direction de Jean-Christophe Ponsart et de Benoît Marx.

Je tiens tout d'abord à remercier chaleureusement mes deux directeurs de thèse, Jean-Christophe Ponsart et de Benoît Marx, pour leur accompagnement, constant et éclairé, lors de cette thèse. J'ai ressenti une véritable liberté intellectuelle dans mon travail, et je les remercie vivement de m'avoir accordé leur confiance dans les chemins de traverse que j'ai pu emprunter. Je leur suis particulièrement reconnaissant pour leur ouverture d'esprit face à mes circonvolutions théoriques et mes rigidités rédactionnelles, entre autres extravagances, qui ont toujours été reçues avec humour et empathie. Merci Messieurs, mon épanouissement total durant ces trois ans est directement lié à la qualité de votre travail d'encadrement.

Je souhaite exprimer toute ma gratitude envers mes rapporteurs, Kévin Guelton et Olivier Sename, pour leur lecture approfondie de ce manuscrit, ainsi qu'au reste de mon jury : Gabriela Iuliana Bara, Dalil Ichalal et Jamal Daafouz, pour leur vif intérêt envers mes résultats. Les questions que vous m'avez posées et les remarques que vous m'avez apportées ont toutes été précieuses. Merci encore à Gabriela Iuliana Bara pour avoir assuré la présidence de mon jury.

Merci à Christian Daul, grâce à qui j'ai pu encadrer pendant presque deux ans les séances d'informatique à l'EEIGM et l'ENSGSI, ainsi qu'à tout le reste des chargés de TD qui m'ont accompagné : Sabine Richard, Christian Duclou et Hervé Lacresse. J'ai énormément appris de cette expérience d'enseignement.

Merci à tous les collègues du laboratoire que j'ai eu le plaisir de côtoyer durant ma thèse. Merci particulièrement à Romain Postoyan et Constantin Morarescu pour m'avoir épaulé, en plus de mes directeurs de thèse, dans ma recherche d'un contrat post-doctoral. Merci également à tous les services administratifs du CRAN pour avoir parfaitement pris en charge mon contrat de travail ainsi que tous mes déplacements, en France comme à l'étranger. Merci tout particulièrement à Christine Pierson et à Katia Thomas pour leur immense gentillesse.

Je souhaite également rendre hommage à José Ragot, le premier professeur que j'ai appris à connaître au CRAN après mes directeurs de thèse. Nos conversations étaient régulières, toujours empreintes d'enthousiasme, d'estime mutuelle et de sincérité. Sa disparition brutale lors de ma deuxième année de thèse a été un choc terrible, et j'adresse à tous ses proches mes plus profondes condoléances.

Merci à tous les doctorants (ou non), du CRAN (ou non), que j'ai eu le plaisir de côtoyer au cours de ces trois ans à Nancy : Alessandra, Amine, Anthony, Aurélien, Bélen, Bikas, Bouchra, Cédric, Danh, Elena, Emilie, Flora, Gau, Hassan, Hernando, Jesus, Jomphop, Jonathan, Maxime, Mira, Mohammed, Nathan, Nicolas, Olivier, Pierre, Saule, Sifeddine, Simone, Soha, Théo, Thomas, Timo, Valessa, Viviana, Xiao, Zarina, et tous ceux que j'oublie.

Merci à tous les copains de Nantes et de Paris venus assister à ma soutenance : JB, Kévin, Cam, Mr Sirletti et Léonard. Ce fut un plaisir de vous retrouver, à Nancy comme ailleurs, pendant ces trois ans.

Merci à Alexandra Elbakyan, fondatrice de Sci-Hub, grâce à qui le savoir scientifique est rendu accessible à tous. Le monde scientifique, et *a fortiori* le monde entier tout court, lui est grandement redevable tant l'utilisation de son site est devenue incontournable, toutes disciplines académiques confondues.

Merci à Gulminyam, ma complice, ma compagne et ma moitié. Tu es la tour Eiffel de ma ligne 6, la Jean Reno de mon wasabi, la Robespierre de mes catacombes. Ma vie avec toi déborde de voyages, de rires, de clins d'oeil et de fleurs.

Enfin, merci à ma maman et à mon papa. Vous avez créé les conditions nécéssaires à mon autodétermination sans faille. Je vous aime.

It is often taken for granted that a clear and distinct understanding of new ideas precedes, and should precede, their formulation and their institutional expression. Yet this is certainly not the way in which small children develop. They use words, they combine them, they play with them, until they grasp a meaning that has so far been beyond their reach. And the initial playful activity is an essential prerequisite of the final act of understanding. There is no reason why this mechanism should cease to function in the adult. The process itself is not guided by a well-defined programme, and cannot be guided by such a programme, for it contains the conditions for the realization of all possible programmes. It is guided rather by a vague urge, by a passion. The passion gives rise to specific behaviour which in turn creates the circumstances and the ideas necessary for analysing and explaining the process, for making it 'rational'.

Paul Feyerabend — Against method

Table of Contents

Notations

Logic

$\forall, \exists$	*for all, there exists*
$\vee, \wedge, \neg$	*or, and, not*

Usual sets

$\emptyset$	The empty set
$\mathbb{N}$	The positive integers ($0 \in \mathbb{N}$)
$\mathbb{Z}$	The integers
$\mathbb{R}$	The real numbers
$\mathbb{C}$	The complex numbers
$\mathbb{K}$	Usually $\mathbb{Z}$, $\mathbb{R}$, or $\mathbb{C}$
$[a, b)$	$\{x \in \mathbb{R} : a \leq x < b\}$
$[\![a, b]\!]$	$\{x \in \mathbb{Z} : a \leq x \leq b\}$
$\mathbb{K}^{n\times m}$	m-columns, n-rows matrices of $\mathbb{K}$
$\mathrm{GL}_n(\mathbb{K})$	Linear group of $\mathbb{K}^{n\times n}$
$\mathbb{S}_n(\mathbb{R})$	Symmetric matrices
$\mathbb{S}_n^+(\mathbb{R})$	Positive semidefinite matrices
$\mathbb{S}_n^{++}(\mathbb{R})$	Positive definite matrices
$\mathbb{S}_n^-(\mathbb{R})$	Negative semidefinite matrices
$\mathbb{S}_n^{--}(\mathbb{R})$	Negative definite matrices
$\mathbb{K}^n$	Vectors, identified with $\mathbb{K}^{n\times 1}$
$\mathbb{N}_m^n$	$\{(k_1, \ldots, k_n) \in \mathbb{N}^n : \sum_i k_i = m\}$
Δ_{n-1}	$\{(x_1, \ldots, x_n) \in \mathbb{R}^n_{\geq 0} : \sum_i x_i = 1\}$
<u>If $\mathbb{K} = \mathbb{Z}$ or $\mathbb{R}$:</u>	
$\tilde{\mathbb{K}}$	$\mathbb{K} \cup \{-\infty, +\infty\}, -\infty < t \in \mathbb{K} < +\infty$
<u>If $\mathbb{K} = \mathbb{Z}, \tilde{\mathbb{Z}}, \mathbb{R}$ or $\tilde{\mathbb{R}}$:</u>	
$\mathbb{K}_{\geq a}$	$\mathbb{K} \cap \{x \in \mathbb{K} : x \geq a\}$
$\mathbb{K}_{>a}$	$\mathbb{K} \cap \{x \in \mathbb{K} : x > a\}$

Set operations

$\times$	Cartesian product
$\cup, \cap$	Union, intersection
$\setminus$	Set difference
$\oplus$	Minkowski sum
$t\mathcal{S}$	$\mathcal{S}$ scaled by $t \in \mathbb{R}$ w.r.t. 0
$M\mathcal{S}$	$\mathcal{S}$ linearly transformed by M
$\mathbf{P}(\mathcal{S})$	Power set of $\mathcal{S}$

Linear algebra

$M_{(i,j)}$	(i, j)-th coefficient of $M \in \mathbb{K}^{n\times m}$
$v_{(i)}$	i-th coordinate of $v \in \mathbb{K}^n$
M^{-1}	Inverse of M
$M^\dagger$	Moore-Penrose inverse of M
$M^\top$	Transpose of M
$M^\perp$	Orthogonal complement of M
$\prod_{i=1}^n A_i$	$A_n \ldots A_1$
$\mathcal{H}(M)$	$M + M^\top$
$\mathrm{rank}(M)$	Rank of M
$\mathrm{Im}(M)$	Image of M
$\mathrm{Ker}(M)$	Kernel of M
$\det(M)$	Determinant of M
$\mathrm{Tr}(M)$	Trace of M
$\Lambda(M)$	Spectrum of M
$\lambda_{\max}(M)$	$\max_{\lambda\in\Lambda(M)} \Re(\lambda)$
$\lambda_{\min}(M)$	$\min_{\lambda\in\Lambda(M)} \Re(\lambda)$
$\langle a \vert b \rangle$	Scalar product $a^\top b$
$\langle a \vert b \rangle_P$	Weighted scalar product $a^\top P b$
$\Vert M \Vert_p$	Natural p-norm of $M \in \mathbb{K}^{n\times m}$
$\eta_P(M)$	Weighted logarithmic norm
I_n	Identity matrix of $\mathbb{K}^{n\times n}$
0_n	Null element of $\mathbb{K}^{n\times n}$ or $\mathbb{K}^n$
(the n is generally omitted)	
$\mathbf{1}_i$	Vector s.t. $(\mathbf{1}_i)_{(j)} \triangleq \delta_{ij}$
$\mathrm{span}(\mathcal{S})$	Vector space spanned by $\mathcal{S}$
$\dim(\mathcal{S})$	Intrinsic dimension of $\mathcal{S}$

Functions

$\mathcal{B}_\mathbf{k}$	Bernstein polynomial
$\mu_\mathcal{S}$	Minkowski functional of $\mathcal{S}$
$\partial f/\partial x$	Partial derivative of f
$\dot{f}$	df/dt
$\circ$	Function composition

Complex numbers

$\Re(z)$	Real part of $z \in \mathbb{C}$
$\vert z \vert$	Modulus of $z \in \mathbb{C}$

Integers

$\overline{n_{[m]} \dots n_{[1]}}^{(b)}$	Base-b notation of $n \in \mathbb{N}$
$n!$	Factorial of $n \in \mathbb{N}$
$\binom{n}{k}$	Binomial coefficient
$\binom{n}{k_1,\dots,k_m}$	Multinomial coefficient
δ_{ij}	Kronecker delta
$\#\mathcal{S}$	Cardinal of $\mathcal{S}$

Topology

$\mathcal{B}_p(c,r)$	$\{x : \|x-c\|_p < r\}$
$\mathrm{hull}(\mathcal{S})$	Convex hull of $\mathcal{S}$
$\mathrm{intr}_E(\mathcal{S})$	Interior of $\mathcal{S}$ in E
$\mathrm{cls}_E(\mathcal{S})$	Closure of $\mathcal{S}$ in E
$\partial_E\mathcal{S}$	Boundary of $\mathcal{S}$ in E

(the E in the above is often omitted)

$\mathrm{Vol}(\mathcal{S})$	Volume of $\mathcal{S}$

Control

x	State of the system
u	Input of the system
y	Output of the system
w	Disturbances of the system
f	Faults of the system
r	Residuals of the system
θ	Scheduling vector of the system
θ_i	i-th scheduling variable
$\mathbf{h}$	Activation functions of the system
h_i	i-th activation function
V	Usually a Lyapunov function

Abbreviations

i.e.	*id est*
e.g.	*exempli gratia*
s.t.	*such that*
w.r.t.	*with respect to*
resp.	*respectively*

Acronyms

System Representation

LTI	Linear Time Invariant
LTV	Linear Time Varying
LDI	Linear Differential Inclusion
LPV	Linear Parameter Varying
T-S	Takagi-Sugeno
NLSA	Nonlinear Sector Approach

Lyapunov Function

QLF	Quadratic Lyapunov Function
CQLF	Common Quadratic Lyapunov Function
nQLF	non-Quadratic Lyapunov Function
MQLF	Multiquadratic Lyapunov Function
PQLF	Piecewise Quadratic Lyapunov Function
PMQLF	Piecewise Multiquadratic Lyapunov Function

Controller Design

PDC	Parallel Distributed Compensation
nPDC	non-Parallel Distributed Compensation
BPDC	Bézier Parallel Distributed Compensation
BnPDC	Bézier non-Parallel Distributed Compensation

Algorithm, Optimization

NP	Nondeterministic Polynomial
LMI	Linear Matrix Inequality
BMI	Bilinear Matrix Inequality

Diagnosis

FDI	Fault Detection and Isolation
DIT	Direct Image Test
RDIT	Robust Direct Image Test

Geometric contributions to the study of LPV and Takagi-Sugeno systems

Chapter 1

General introduction

This opening chapter introduces the preliminary notions of control necessary to appreciate the scope of the manuscript. It is concluded by a list of contributions, a chapter-by-chapter outline of the thesis, and a list of the author's publications.

1.1 Control theory

Control theory is a field at the intersection of mathematics and engineering, focused on understanding and manipulating the behaviour of dynamical systems capable of receiving inputs, and for which output signals can be measured. This discipline has received a great amount of attention in the last century due to its numerous applications, as it plays a crucial role in electrical, mechanical and chemical engineering [45, 184, 12], robotics [182] and aeronautics [131]. It has also been applied to a large variety of other academic domains, including economics and finance [172, 232], biology [134], and more recently, research on large language models [37]. Control theory is mostly concerned with the design of generic procedures to obtain control laws, i.e. carefully constructed input signals whose purpose is to govern the behaviour of dynamical systems.

1.1.1 Dynamical systems

The systems studied by control theory are usually modeled using differential or difference equations, such as in the following state-space model:

$$\delta x(t) = f(x(t), u(t), w(t), t) \tag{1.1a}$$
$$y(t) = h_y(x(t), u(t), w(t), t) \tag{1.1b}$$
$$z(t) = h_z(x(t), u(t), w(t), t) \tag{1.1c}$$

- $x(t) \in \mathbb{R}^{n_x}$ is the (internal) state of the system;
- $u(t) \in \mathbb{R}^{n_u}$ is the input of the system for which a control law is usually designed;
- $w(t) \in \mathbb{R}^{n_w}$ is a vector of exogenous input signals, usually comprising of disturbances and reference signals;
- $y(t) \in \mathbb{R}^{n_y}$ is the measured output vector of the system;
- $z(t) \in \mathbb{R}^{n_e}$ is the tracking error of the controlled quantities (with respect to their reference), also called the regulated output vector.

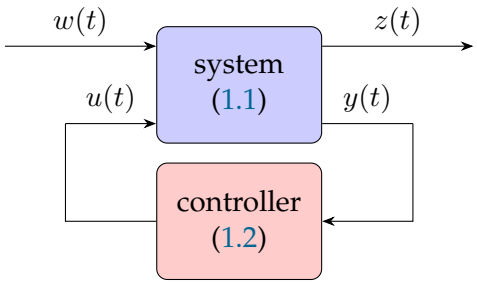


Figure 1.1: Closed-loop control architecture of the system (1.1).

Equation (1.1a) is the dynamical part of the system, with δ the shift operator, defined by:

- $\delta x(t) = \dot{x}(t)$ in the continuous-time case (and $t \in \mathbb{R}$), leading to a differential equation;
- $\delta x(t) = x(t+1)$ in the discrete-time case (and $t \in \mathbb{Z}$), leading to a difference equation.

Remark 1.1.1. *The system (1.1) is said to be time-invariant if f, h_y and h_z do not directly depend on time t.*

These systems are generally implicitly assumed to have a solution at all time, in the sense that given an initial condition $x(t_0) \in \mathbb{R}^{n_x}$, the trajectory $x(t)$ is uniquely defined on the maximal interval of existence $[t_0, +\infty)$ (eventually intersected with $\mathbb{Z}$ in the discrete-time case). Although these conditions for existence and uniqueness are guaranteed in the discrete-time case under very mild requirements, they are typically less trivial to establish in the continuous-time case. The Cauchy-Lipschitz theorem (also known as the Picard-Lindelöf theorem) provides such conditions:

Theorem 1.1.1 (Cauchy-Lipschitz theorem)**.** *Rewriting (1.1a) in the continuous-time case as $\dot{x}(t) = f(x(t), t)$, for any initial condition $x(t_0) = x_0 \in \mathbb{R}^{n_x}$, the state trajectory of the system (1.1) is uniquely defined on $\mathbb{R}$ if f is piecewise continuous with respect to t and globally Lipschitz with respect to x.*

Proof. See Theorem 3.1. of [149]. □

The control challenge usually associated with these systems consists in driving the regulated output vector $z(t)$ to 0 using a dedicated control law $u : [t_0, +\infty) \to \mathbb{R}^{n_u}$, while respecting some criteria. These criteria can take the form of physical constraints on the vectors at play, of optimality of the control law with respect to a cost function, or of robustness to disturbances and model uncertainties. A first useful notion to achieve said robustness is the *feedback* principle. It consists in utilizing a closed-loop control architecture, where the value of the input $u(t)$ is updated in real-time based on the measurement of the output $y(t)$, in order to compensate the deviation of $z(t)$ from 0 (Figure 1.1). A closed-loop controller usually takes the following form:

$$\delta x_c(t) = f_c(x_c(t), y(t), t) \tag{1.2a}$$
$$u(t) = h_c(x_c(t), y(t), t) \tag{1.2b}$$

This controller is said to be static if it has no internal state x_c (and thus, no dynamical part (1.2a)), otherwise it is said to be dynamic. If the whole state of the system (1.1) is measured, then $y(t) = x(t)$, which simplifies the control problem. Otherwise, if one is able to approximate the value of the state $x(t)$ from $u(t)$, $w(t)$ and $y(t)$, usually by means of a dynamical observer, the control law $u(t)$ can also be based on this reconstructed state, which is denoted $\hat{x}(t)$. This last strategy is generally referred to as *observer-based control*.

In order to study the feasibility of the control challenge stated above, two main structural properties are usually investigated: controllability and observability.

Definition 1.1.1 (Controllability)**.** *The system (1.1) is said to be controllable on the interval $[t_1, t_2]$ if for any state $x(t_1) \in \mathbb{R}^{n_x}$ and target $x_f \in \mathbb{R}^{n_x}$, there exists a bounded input $u : [t_1, t_2] \to \mathbb{R}^{n_u}$ such that $x(t_2) = x_f$.*

Definition 1.1.2 (Observability)**.** *The system (1.1) is said to be observable on the interval $[t_1, t_2]$ if the value of the state $x(t_1)$ can be determined from knowledge of the system inputs $u(t)$, $w(t)$ and output $y(t)$ on the interval $[t_1, t_2]$.*

Remark 1.1.2. *Assuming perfect knowledge of the model, and uniqueness of its solutions, obtaining $x(t_1)$ in the observability definition above provides the value of $x(t)$ for all $t \in [t_1, t_2]$.*

Remark 1.1.3. *For the sake of simplicity, controllability and observability are always referring in this manuscript to the definition stated above, but there exists many alternative definitions in the literature which are not not necessarily equivalent to these ones [257, 47, 156, 123].*

The state-space model (1.1) being extremely non-specific, results on this type of systems, including guarantees on their structural properties, are usually challenging to obtain, and difficult to apply. These difficulties have motivated researchers to focus on sub-classes of systems with more explicit mathematical properties, and which are therefore more practically handled. In particular, control theory has been widely developed in the setting where f, h_y and h_z are linear functions of x, u, and w. In this context, the state-space model (1.1) can be rewritten using following matrix equation:

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = \begin{pmatrix} A(t) & B_1(t) & B_2(t) \\ C_1(t) & D_{11}(t) & D_{12}(t) \\ C_2(t) & D_{21}(t) & D_{22}(t) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \tag{1.3}$$

where:

- $A(t) \in \mathbb{R}^{n_x \times n_x}$ is the state matrix;
- $B_1(t) \in \mathbb{R}^{n_x \times n_u}$ is the control input matrix;
- $B_2(t) \in \mathbb{R}^{n_x \times n_w}$ is the exogenous input matrix;
- $C_1(t) \in \mathbb{R}^{n_y \times n_x}$ is the measured output matrix;
- $D_{11}(t) \in \mathbb{R}^{n_y \times n_u}$ is the direct control input to measured output matrix;
- $D_{12}(t) \in \mathbb{R}^{n_y \times n_w}$ is the direct exogenous input to measured output matrix;
- $C_2(t) \in \mathbb{R}^{n_e \times n_x}$ is the regulated output matrix;

- $D_{21}(t) \in \mathbb{R}^{n_e \times n_u}$ is the direct control input to regulated output matrix;
- $D_{22}(t) \in \mathbb{R}^{n_e \times n_w}$ is the direct exogenous input to regulated output matrix.

If all the matrices are independent of t, then the system is said to be Linear Time Invariant (LTI), otherwise it is said to be Linear Time Varying (LTV). In a linear context, the closed-loop controller (1.2) is usually taken to be linear as well:

$$\begin{pmatrix} \delta x_c(t) \\ u(t) \end{pmatrix} = \begin{pmatrix} K_{11}(t) & K_{12}(t) \\ K_{21}(t) & K_{22}(t) \end{pmatrix} \begin{pmatrix} x_c(t) \\ y(t) \end{pmatrix} \tag{1.4}$$

A wide range of powerful results can be derived from this linear framework, such as simple controllability and observability criteria, pole assignment strategies through static state and output feedback control laws, Kalman-like filtering, the design of linear–quadratic regulator, etc. More specifically, the LTI case offers a rich *frequency domain* point of view by leveraging the Laplace transform and the z-transform respectively in the continuous-time and the discrete-time cases. This frequency-domain approach leads to an even wider range of control techniques, including loop-shaping, or $\mathcal{H}_2$ and $\mathcal{H}_\infty$ controller synthesis [283].

The Gramian-based criteria for controllability and observability of linear systems are recalled below, illustrating the practicality of the linear framework to obtain such guarantees.

Theorem 1.1.2 (Controllability and observability criteria [62])**.** *It is assumed that the exogenous input w is identically 0 ($w = 0$). Let $\Phi(t_2, t_1)$ denote the state-transition matrix of $\delta x(t) = A(t)x(t)$ between t_1 and t_2. The system (1.3) is controllable on $[t_1, t_2]$ if and only if*

$$W_c(t_2, t_1) \triangleq \int_{t_1}^{t_2} \Phi(t_2, s) B_1(s) B_1^\top(s) \Phi^\top(t_2, s) ds \in \mathrm{GL}_{n_x}(\mathbb{R}) \tag{1.5}$$

Similarly, (1.3) is observable on $[t_1, t_2]$ if and only if

$$W_o(t_2, t_1) \triangleq \int_{t_1}^{t_2} \Phi^\top(s, t_1) C_1^\top(s) C_1(s) \Phi(s, t_1) ds \in \mathrm{GL}_{n_x}(\mathbb{R}) \tag{1.6}$$

W_c and W_o are resp. called the controllability and observability Gramian of (1.3). The integral $\int_{t_1}^{t_2} ds$ should be interpreted as the sum $\sum_{s=t_1}^{t_2-1}$ in the discrete-time case.

Table 1.1: Summary of the expressions of the state transition matrix $\Phi(t_2, t_1)$.

	LTI	LTV
Discrete-time	$A^{t_2 - t_1}$	$\prod_{s=t_1}^{t_2-1} A(s)$
Continuous-time	$e^{(t_2-t_1)A}$	$\prod_{t_1}^{t_2} e^{A(s)ds}$

Remark 1.1.4. *For any initial condition $x(t_0) \in \mathbb{R}^{n_x}$ and $t_1, t_2 \in \mathbb{R}$ or $\mathbb{Z}$ such that $t_2 \geq t_1$, the state transition matrix Φ of $\delta x(t) = A(t)x(t)$ verifies $x(t_2) = \Phi(t_2, t_1)x(t_1)$. The explicit expressions of*

$\Phi(t_2, t_1)$ are summarized in Table 1.1. In particular, the expression in the continuous-time LTV case is provided under the form of a product integral. This is explained in more details in Chapter 6 of this manuscript.

Remark 1.1.5. *There also exists equivalent rank-based criteria of controllability and observability for linear systems [257], and although they are more practical to be checked by hand, this introduction sticks to the Gramian-based criteria as they will be reused in Chapter 6 of this manuscript.*

The linear state-space models are also locally useful in some nonlinear contexts, as (1.1) can often be qualitatively approximated by (1.3) using a linearization of the system dynamics around an equilibrium point (the Hartman-Grobman theorem), with (x^*, u^*, t) the selected point of linearization at time t.

$$\begin{pmatrix} A(t) & B_1(t) & B_2(t) \\ C_1(t) & D_{11}(t) & D_{12}(t) \\ C_2(t) & D_{21}(t) & D_{22}(t) \end{pmatrix} = \begin{pmatrix} \partial f/\partial x & \partial f/\partial u & \partial f/\partial w \\ \partial h_y/\partial x & \partial h_y/\partial u & \partial h_y/\partial w \\ \partial h_z/\partial x & \partial h_z/\partial u & \partial h_z/\partial w \end{pmatrix}\Bigg|_{(x^*,u^*,t)} \tag{1.7}$$

However, this approximation is only local, leading to model uncertainties in the representation of the dynamical system. In order to avoid these uncertainties, other classes of models may be needed. For example, one solution consists in using a linear differential inclusion at the level of the dynamical equation (1.1a) [48, 7]. Systems of differential inclusions are nonlinear systems whose dynamical equation takes the form of an inclusion:

$$\delta x(t) \in F(x(t), u(t), w(t), t) \tag{1.8}$$

If the shifted state vector $\delta x(t)$ belongs to the convex-hull of several LTI systems, this differential inclusion is called a *polytopic Linear Differential Inclusion (LDI)*:

$$\delta x(t) \in \text{hull}\{A_i x(t) + B_{1,i} u(t) + B_{2,i} w(t) : i = 1, \dots, m\} \tag{1.9}$$

It can be visually understood with Figure 1.2 in the input free case ($u = 0$, $w = 0$). This class of systems combines the linear framework with the capability of taking properly into account the models uncertainties. However, this robustness is obtained at the cost of the uniqueness of the solutions to (1.9).

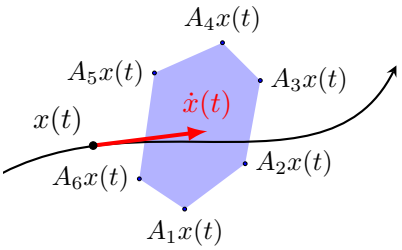


Figure 1.2: At each point $x(t)$ of a state trajectory (black line) verifying (1.9), $\dot{x}(t)$ belongs to the convex-hull $\text{hull}\{A_1, \dots, A_6\} \cdot x(t)$.

1.1.2 Stability

This section considers the nonlinear system (1.1) together with its closed-loop controller (1.2) in an exogenous input free context ($w = 0$). By concatenating the state of the system with the state

of the controller, the whole closed-loop system can be rewritten as a single dynamical system of the form:

$$\delta x(t) = f(x(t), t) \tag{1.10a}$$
$$z(t) = h(x(t), t) \tag{1.10b}$$

As previously stated, the goal of control theory consists in driving the regulated output $z(t)$ to the origin 0 for a wide range of initial conditions $x(t_0) \in \mathbb{R}^{n_x}$ and with some amount of robustness. This goal can be characterized mathematically by introducing several notions of output stability. This section formalizes these notions, and provides practical conditions to verify them using Lyapunov's method. Some preliminary definitions are however needed, namely the definitions of $\mathcal{K}$, $\mathcal{K}_\infty$ and $\mathcal{KL}$ functions, as well as the definitions of positive (semi)definite and of radially unbounded functions [149].

Definition 1.1.3 ($\mathcal{K}$, $\mathcal{K}_\infty$ function). *A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{K}$ if it is continuous, strictly increasing, and $\alpha(0) = 0$. Moreover, it is of class $\mathcal{K}_\infty$ is it is of class $\mathcal{K}$ and $\lim_{r\to\infty} \alpha(r) = +\infty$.*

Definition 1.1.4 ($\mathcal{KL}$ function). *A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{KL}$ if it is continuous, and such that for each $s \in \mathbb{R}_{\geq 0}$, $\beta(\cdot, s) \in \mathcal{K}$, and for each $r \in \mathbb{R}_{\geq 0}$, $\beta(r, \cdot)$ is decreasing on $\mathbb{R}_{\geq 0}$ with $\lim_{s\to+\infty} \beta(r, s) = 0$.*

Definition 1.1.5 (Positive (semi)definite function and matrix). *A function $V : \mathbb{R}^n \to \mathbb{R}$ is positive (semi)definite if $V(0) = 0$ and $V(x) > 0$ (resp. ≥ 0) for all $x \in \mathbb{R}^n \setminus \{0\}$. A symmetric matrix $P \in \mathbb{S}_n(\mathbb{R})$ is positive (semi)definite (i.e. $P \in \mathbb{S}_n^{++}(\mathbb{R})$, resp. $P \in \mathbb{S}_n^{+}(\mathbb{R})$) if the function $V(x) = \langle x|x\rangle_P = \langle x|Px\rangle = x^\top Px$ is positive (semi)definite.*

Definition 1.1.6 (Radially unbounded function). *A function $V : \mathbb{R}^n \to \mathbb{R}$ is radially unbounded if $\|x\| \to +\infty$ implies $V(x) \to +\infty$.*

Remark 1.1.6. *The controllability and observability Gramians $W_c(t_2, t_1)$ and $W_o(t_2, t_1)$ of Theorem 1.1.2 are positive semidefinite matrices for all $t_1, t_2 \in \mathbb{R}$ or $\mathbb{Z}$. Moreover, they are positive definite if and only if they are nonsingular.*

1.1.2.1 Stability definitions

Assuming that the origin 0 is an equilibrium point of the system (1.10) (i.e. that for all t, $f(0, t) = 0$ and $h(0, t) = 0$), and that the assumptions of Theorem 1.1.1 are verified on f, the following definitions of output stability are introduced [147, 150, 44]:

In the definitions below, $r \in \tilde{\mathbb{R}}_{>0}$, and the ball $\mathcal{B}_2(0, r)$ is now assumed to be a positively invariant set, meaning for all $x(t_0) \in \mathcal{B}_2(0, r)$, the state $x(t)$ remains bounded in $\mathcal{B}_2(0, r)$ for all $t \in [t_0, +\infty)$.

Definition 1.1.7 (Output attractivity). *The output z is attractive at the origin on $\mathcal{B}_2(0, r) \subseteq \mathbb{R}^{n_x}$ if for all $x(t_0) \in \mathcal{B}_2(0, r)$:*

$$\lim_{t\to+\infty} z(t) = 0 \tag{1.11}$$

Definition 1.1.8 (Output stability). *The output z is stable at the origin on $\mathcal{B}_2(0,r) \subseteq \mathbb{R}^{n_x}$ if there exists $\alpha \in \mathcal{K}$ such that for all $x(t_0) \in \mathcal{B}_2(0,r)$ and $t \in [t_0, +\infty)$:*

$$\|z(t)\|_2 \leq \alpha(\|x(t_0)\|_2) \tag{1.12}$$

Definition 1.1.9 (Asymptotic output stability). *The output z is asymptotically stable at the origin on $\mathcal{B}_2(0,r) \subseteq \mathbb{R}^{n_x}$ if it is both output stable and output attractive on $\mathcal{B}_2(0,r)$, i.e. if there exists $\beta \in \mathcal{KL}$ such that for all $x(t_0) \in \mathcal{B}_2(0,r)$ and $t \in [t_0, +\infty)$:*

$$\|z(t)\|_2 \leq \beta(\|x(t_0)\|_2, t - t_0) \tag{1.13}$$

Definition 1.1.10 (Exponential output stability). *The output z is exponentially stable at the origin on $\mathcal{B}_2(0,r) \subseteq \mathbb{R}^{n_x}$ if there exists $k, \lambda \in \mathbb{R}_{>0}$ such that for all $x(t_0) \in \mathcal{B}_2(0,r)$ and $t \in [t_0, +\infty)$:*

$$\|z(t)\|_2 \leq k\|x(t_0)\|_2 e^{-\lambda(t-t_0)} \tag{1.14}$$

If $r < +\infty$, then the properties above are said to hold *locally*, otherwise if $r = \infty$, then $\mathcal{B}_2(0,r) = \mathbb{R}^{n_x}$ and the properties are said to hold *globally*. If these properties hold no matter $t_0 \in \mathbb{R}$, they are said to be *uniform*.

These definitions are still not the most complete notions of stability employed in control theory, as they neglect the influence of the exogenous inputs $w(t)$. Input to output stability notions have also been introduced by the literature, but remain out of the scope of this thesis [262]. In fact, the simplifying assumption that $z(t) = x(t)$ is made in this manuscript, and *state stability* is always investigated, rather than *output stability*. These simplifications are usual in the control literature, but it is important to recognize their limitations:

- both state and output stability disregard the influence of the exogenous inputs $w(t)$;
- state stability ensures the stability of the whole internal state of the system $x(t)$, rather than simply the stability of the regulated output vector $z(t)$.

Remark 1.1.7. *For state stability ($z(t) = x(t)$), the assumption that $\mathcal{B}_2(0,r)$ is a positively invariant set can be omitted [150, 44].*

In the case of state stability (i.e. if $z(t) = x(t)$), the solutions of a globally stable continuous-time system are necessarily defined on the maximal interval of existence $[t_0, +\infty)$. However, they are not necessarily uniquely defined: state stability is not a substitute to the Cauchy-Lipschitz theorem (Theorem 1.1.1). It should be noted that the definitions given hereabove still hold when the solutions to the state-space model are non-unique by considering the set of all possible trajectories beginning at $x(t_0)$. In particular, this allows to generalize these stability notions to systems of differential inclusions, including the LDI framework (1.9) [50].

1.1.2.2 Lyapunov's method

In his seminal 1892 work on stability, Lyapunov introduced his so-called second method to demonstrate that a system of differential equations has stable solutions [181]. His methodology, known today as the Lyapunov stability criterion, or the direct method, relies on exhibiting a positive-definite function around an equilibrium point of the system, whose derivative along the system trajectories remains negative. If such a Lyapunov function exists, then the system is

stable around this equilibrium, which guarantees that any trajectory starting near to this equilibrium will always remain in its neighborhood. Moreover, a strictly negative derivative of a Lyapunov function also guarantees the asymptotic stability of the system, i.e. both stability and the asymptotic convergence of the trajectories towards this equilibrium [241, 149, 289]. In a nutshell, exhibiting a Lyapunov function provides a powerful way to learn about a system behaviour without having to explicitly compute its trajectories. This made Lyapunov functions one of the most widely used tool in the fields of dynamical systems and control theory.

The Lyapunov criterion for stability, asymptotic stability, and exponential stability are given thereafter both in the continuous-time and the discrete-time cases [150, 44].

Consider a system $\delta x(t) = f(x(t), t)$ such that the origin $x = 0$ is an equilibrium point of the system (i.e. for all t, $f(0, t) = 0$), and such that the assumptions of Theorem 1.1.1 are verified on f in the continuous-time case. In the following, $V_c : \mathbb{R}^{n_x} \times \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function and $V_d : \mathbb{R}^{n_x} \times \mathbb{Z} \to \mathbb{R}$ is a continuous function. Moreover, $W_1, W_2, W_3 : \mathbb{R}^{n_x} \to \mathbb{R}$ are three continuous positive definite functions. Finally, $D \subseteq \mathbb{R}^{n_x}$ designates a domain such that $0 \in \mathrm{intr}(D)$.

Theorem 1.1.3 (Lyapunov conditions for stability)**.** *There exists $r \in \tilde{\mathbb{R}}_{>0}$ such that the continuous-time system $\dot{x}(t) = f(x(t), t)$ is uniformly stable at the origin on $\mathcal{B}_2(0, r)$ if for all $t \in \mathbb{R}$ and $x \in D$:*

$$W_1(x) \leq V_c(x, t) \leq W_2(x) \tag{1.15a}$$

$$\frac{\partial V_c}{\partial t}(x, t) + \left\langle \frac{\partial V_c}{\partial x}(x, t) | f(x, t) \right\rangle \leq 0 \tag{1.15b}$$

Similarly, there exists $r \in \tilde{\mathbb{R}}_{>0}$ such that the discrete-time system $x(t+1) = f(x(t), t)$ is uniformly stable at the origin on $\mathcal{B}_2(0, r)$ if for all $t \in \mathbb{Z}$ and $x \in D$:

$$W_1(x) \leq V_d(x, t) \leq W_2(x) \tag{1.16a}$$

$$V_d(f(x, t), t + 1) - V_d(x, t) \leq 0 \tag{1.16b}$$

Theorem 1.1.4 (Lyapunov conditions for asymptotic stability)**.** *There exists $r \in \tilde{\mathbb{R}}_{>0}$ such that the continuous-time system $\dot{x}(t) = f(x(t), t)$ is uniformly asymptotically stable at the origin on $\mathcal{B}_2(0, r)$ if for all $t \in \mathbb{R}$ and $x \in D$:*

$$W_1(x) \leq V_c(x, t) \leq W_2(x) \tag{1.17a}$$

$$\frac{\partial V_c}{\partial t}(x, t) + \left\langle \frac{\partial V_c}{\partial x}(x, t) | f(x, t) \right\rangle \leq -W_3(x) \tag{1.17b}$$

Similarly, there exists $r \in \tilde{\mathbb{R}}_{>0}$ such that the discrete-time system $x(t+1) = f(x(t), t)$ is uniformly asymptotically stable at the origin on $\mathcal{B}_2(0, r)$ if for all $t \in \mathbb{Z}$ and $x \in D$:

$$W_1(x) \leq V_d(x, t) \leq W_2(x) \tag{1.18a}$$

$$V_d(f(x, t), t + 1) - V_d(x, t) \leq -W_3(x) \tag{1.18b}$$

Theorem 1.1.5 (Lyapunov conditions for exponential stability). *Let $a, b, c \in \mathbb{R}_{>0}$. There exists $r \in \tilde{\mathbb{R}}_{>0}$ such that the continuous-time system $\dot{x}(t) = f(x(t), t)$ is uniformly exponentially stable at the origin on $\mathcal{B}_2(0, r)$ if for all $t \in \mathbb{R}$ and $x \in D$:*

$$a\|x\|_2^2 \leq V_c(x,t) \leq b\|x\|_2^2 \tag{1.19a}$$

$$\frac{\partial V_c}{\partial t}(x,t) + \left\langle \frac{\partial V_c}{\partial x}(x,t) | f(x,t) \right\rangle \leq -c\|x\|_2^2 \tag{1.19b}$$

Similarly, there exists $r \in \tilde{\mathbb{R}}_{>0}$ such that the discrete-time system $x(t+1) = f(x(t), t)$ is uniformly exponentially stable at the origin on $\mathcal{B}_2(0, r)$ if for all $t \in \mathbb{Z}$ and $x \in D$:

$$a\|x\|_2^2 \leq V_d(x,t) \leq b\|x\|_2^2 \tag{1.20a}$$

$$V_d(f(x,t), t+1) - V_d(x,t) \leq -c\|x\|_2^2 \tag{1.20b}$$

If $D = \mathbb{R}^{n_x}$ and W_1 is radially unbounded in the theorems above, then the stability properties hold globally (i.e. $r = +\infty$).

Remark 1.1.8. *The left-hand side of* (1.15b), (1.17b) *and* (1.19b) *is the Lie derivative of $V(x,t)$ along the trajectories of $\dot{x}(t) = f(x(t), t)$. It is also denoted $\dot{V}(x(t), t)$ in this manuscript.*

Remark 1.1.9. *Extensions of these results exist for systems of differential inclusions [50].*

Despite their strength, Lyapunov stability criteria are only sufficient stability conditions, and converse results had to be established to complete Lyapunov's theory. Three well-known converse results are reported below.

- An autonomous continuous differential equation is asymptotically stable if and only if there exists a smooth Lyapunov function demonstrating its asymptotic stability [149, 289].
- A continuous-time polytopic LDI is globally exponentially stable if and only if there exists a piecewise quadratic Lyapunov function demonstrating its global asymptotic stability [193].
- An LTI system is globally exponentially stable if and only if there exists a Quadratic Lyapunov Function (QLF) demonstrating its global asymptotic stability.

This last result is generally known as the Lyapunov lemma, and it holds both for continuous-time and discrete-time LTI systems. Moreover, since an autonomous nonlinear system can often be qualitatively approximated by linearizing its dynamics around an equilibrium point (the Hartman-Grobman theorem), Lyapunov's lemma also holds locally in some nonlinear contexts as well [149, 289]. This linearizing approach to stability is sometimes called Lyapunov's indirect method [150].

Lemma 1.1.1 (Lyapunov's lemma [241, 44]). *Let $A \in \mathbb{R}^{n_x \times n_x}$. The following items are equivalent:*

1. *The LTI system $\dot{x}(t) = Ax(t)$ (resp. $x_{k+1} = Ax_k$) is globally exponentially stable;*

2. *The real part of the eigenvalues of A are strictly negative, i.e. A is Hurwitz (resp. the spectral radius of A is strictly less than 1, i.e. A is Schur);*

3. *There exists a symmetric positive definite matrix $P \in \mathbb{S}_n^{++}(\mathbb{R})$ such that $A^\top P + PA \in \mathbb{S}_{n_x}^{--}(\mathbb{R})$ (resp. $A^\top PA - P \in \mathbb{S}_{n_x}^{--}(\mathbb{R})$).*

In both cases, the matrix P of item 3 provides the QLF (1.21) defined thereafter, demonstrating the stability of the LTI system of item 1 by Theorem 1.1.5.

$$V(x) = \langle x|x\rangle_P = \langle x|Px\rangle = x^\top Px \tag{1.21}$$

The centrality of Lyapunov's lemma made QLF a well-investigated class of Lyapunov functions, despite their possible conservatism to demonstrate the stability of nonlinear systems. QLF are very simply defined, and they exhibit many interesting mathematical properties related to the symmetry of P. As an introduction, one might notice that P has real eigenvalues and orthogonal eigenvectors, that it is diagonalizable, and that the following properties hold:

Property 1.1.1. *For all $P \in \mathbb{S}_n(\mathbb{R})$, $\lambda_{\min}(P)\|x\|_2^2 \leq x^\top Px \leq \lambda_{\max}(P)\|x\|_2^2$. Equalities are obtained by considering the eigenvectors associated with $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$. This demonstrates the following statements:*

1. *$P \in \mathbb{S}_n^{+(+)}(\mathbb{R})$ if and only if $\lambda_{\min}(P) \geq 0$ (resp. > 0),*

2. *$P \in \mathbb{S}_n^{-(-)}(\mathbb{R})$ if and only if $\lambda_{\max}(P) \leq 0$ (resp. < 0),*

3. *$V(x) = x^\top Px$ is radially unbounded if and only if $P \in \mathbb{S}_n^{++}(\mathbb{R})$.*

In particular, item 1 and 2 guarantees that $\mathbb{S}_n^{+(+)}$ and $\mathbb{S}_n^{-(-)}(\mathbb{R})$ are convex cones, by convexity and positive homogeneity of $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ on $\mathbb{S}_n(\mathbb{R})$.

Remark 1.1.10. *$V(x) = -x^\top Px$ is radially unbounded if and only if $P \in \mathbb{S}_n^{--}(\mathbb{R})$.*

Moreover, QLF are easily computed numerically using semidefinite programming to solve a Linear Matrix Inequality (LMI) [48]: this is discussed in more details in the Chapter 2 of the manuscript. In the context of a polytopic LDI, the Lyapunov function (1.21) provides the following result, where P can typically be found using semidefinite programming:

Theorem 1.1.6 (Exponential stability of an LDI [48])**.** *Let $\{A_i\}_{1\leq i\leq m}$ be a set of $\mathbb{R}^{n_x\times n_x}$ matrices. The input free LDI (1.9) ($u = 0$, $w = 0$) is globally exponentially stable at the origin in the continuous-time case (resp. in the discrete-time case) if there exists a symmetric matrix $P \in \mathbb{S}_{n_x}(\mathbb{R})$ such that:*

$$P \in \mathbb{S}_{n_x}^{++}(\mathbb{R}) \tag{1.22a}$$

$$A_i^\top P + PA_i \in \mathbb{S}_{n_x}^{--}(\mathbb{R}) \text{ (resp. } A_i^\top PA_i - P \in \mathbb{S}_{n_x}^{--}(\mathbb{R})) \quad \forall i \in [\![1, m]\!] \tag{1.22b}$$

In this context, (1.21) is called a Common Quadratic Lyapunov Function (CQLF) to the set of matrices $\{A_i\}_{1\leq i\leq m}$. A simple graphical criterion to obtain P in the Theorem 1.1.6 above when $n_x = 2$ is provided in Appendix A, but in general, convex optimization is required.

1.1.3 Stabilization

The stabilization problem, which is to say the problem of designing a controller (1.2) such that the closed-loop system is output asymptotically stable at the origin, is arguably one of the most fundamental problem in control theory. This introduction will be forced to skip most results on the subject, focusing solely on *full state stabilization*. In fact, only the LTI case is tackled, and once again the exogenous inputs $w(t)$ are not considered. The linear system to be controlled now reads:

$$\delta x(t) = Ax(t) + Bu(t) \tag{1.23a}$$
$$y(t) = Cx(t) + Du(t) \tag{1.23b}$$

Under the linear framework, the linear controller (1.4) is generally considered, and taken as follows:

$$\delta\hat{x}(t) = A\hat{x}(t) + Bu(t) + L(C\hat{x}(t) + Du(t) - y(t)) \tag{1.24a}$$
$$u(t) = K\hat{x}(t) \tag{1.24b}$$

which can be rewritten as:

$$\delta\hat{x}(t) = (A + BK + LC)\hat{x}(t) - LCx(t) \tag{1.25a}$$
$$u(t) = K\hat{x}(t) \tag{1.25b}$$

The internal state of this controller is in fact an estimate of the internal state of the system ($x_c(t) = \hat{x}(t)$), obtained using a Luenberger observer of gain $L \in \mathbb{R}^{n_x\times n_y}$. The feedback law then consists of a static proportional state feedback law of gain $K \in \mathbb{R}^{n_u\times n_x}$ based on the estimated state $\hat{x}(t)$. The full closed-loop system, composed of the state $x(t)$ and of the estimation error $\hat{x}(t) - x(t)$, which both need to be stabilized at 0, is given thereafter:

$$\delta\begin{pmatrix} x(t) \\ \hat{x}(t) - x(t) \end{pmatrix} = \begin{pmatrix} A + BK & BK \\ 0 & A + LC \end{pmatrix}\begin{pmatrix} x(t) \\ \hat{x}(t) - x(t) \end{pmatrix} \tag{1.26}$$

The gains K and L can be found using a variety of techniques, including pole assignment strategies, linear–quadratic (Gaussian) optimal control, $\mathcal{H}_2$ and $\mathcal{H}_\infty$ synthesis, etc [283]. Since the full closed-loop system is described with a bloc-triangular matrix, it is easily noticed that the gains K and L can be designed independently of each other to obtain a Hurwitz matrix. Generally speaking, the *separation principle* holds for a class of systems when the observer-based stabilization problems for this class of systems can always be tackled by solving independently an observation problem and a state feedback control problem. Under the linear framework, (1.26) demonstrates that this principle holds.

Remark 1.1.11. *The observer design problem is in fact the dual of the state feedback problem: $A + BK$ is Hurwitz if and only if $A^\top + K^\top B^\top$ is Hurwitz, hence finding K such that $A + BK$ is Hurwitz is the dual of finding L such that $A + LC$ is Hurwitz.*

Remark 1.1.12. *In general, the separation principle does not correlate the performances imposed in both the observation problem and a state feedback control problem to the performances of the full closed-loop system. Only stabilization is ensured.*

Obtaining stabilizing gains K and L is closely related to the structural notions of observability and controllability introduced earlier. This is illustrated by the following theorem:

Theorem 1.1.7 (Pole assignment [297]). *Let $\Gamma \in \mathbb{C}^n$ designate a set of n complex numbers such that if there exists $z \in \Gamma$ such that $z \notin \mathbb{R}$, then $\overline{z} \in \Gamma$, with $\overline{z}$ the complex conjugate of z.*

- *For all such Γ there exists a gain matrix $K \in \mathbb{R}^{n_u \times n_x}$ such that $\Lambda(A+BK) = \Gamma$, if and only if the system (1.23) is controllable on $\tilde{\mathbb{R}}_{\geq 0} \triangleq [0, +\infty]$.*
- *For all such Γ there exists a gain matrix $L \in \mathbb{R}^{n_x \times n_y}$ such that $\Lambda(A+LC) = \Gamma$, if and only if the system (1.23) is observable on $\tilde{\mathbb{R}}_{\geq 0} \triangleq [0, +\infty]$.*

In particular, choosing K and L such that $(A+BK)$ and $(A+LC)$ are Hurwitz in the continuous time-case (resp. Schur in the discrete-time case) ensures the exponential stability of the closed-loop system (1.26).

Remark 1.1.13. *As a rule of thumb, the poles of the observer and of the controller are generally taken such that $\lambda_{\max}(A+LC) \approx 5\lambda_{\max}(A+BK) < 0$.*

One practical way to obtain these gains, although not necessarily the most refined in the LTI framework, consists in computing them numerically using semidefinite programming. Elementary stabilizing LMI conditions are provided thereafter, which guarantee the exponential stabilization of (1.26). The gains K and L can be obtained from these conditions using semidefinite programming.

Theorem 1.1.8 (Exponential stabilization of a LTI system [32]). *The closed-loop LTI system (1.26) can be globally exponentially stabilized at the origin if and only if there exists symmetric matrices $X_1, X_2 \in \mathbb{S}_{n_x}(\mathbb{R})$ and $M_1 \in \mathbb{R}^{n_u \times n_x}$, $M_2 \in \mathbb{R}^{n_x \times n_y}$ such that*

- *in the continuous-time case:*

$$X_1, X_2 \in \mathbb{S}_{n_x}^{++}(\mathbb{R}) \tag{1.27a}$$
$$X_1 A^\top + A X_1 + B M_1 + M_1^\top B^\top \in \mathbb{S}_{n_x}^{--}(\mathbb{R}) \tag{1.27b}$$
$$X_2 A + A^\top X_2 + C^\top M_2 + M_2 C \in \mathbb{S}_{n_x}^{--}(\mathbb{R}) \tag{1.27c}$$

- *in the discrete-time case:*

$$\begin{pmatrix} X_1 & X_1 A^\top + M_1^\top B^\top \\ A X_1 + B M_1 & X_1 \end{pmatrix} \in \mathbb{S}_{n_x}^{++}(\mathbb{R}) \tag{1.28a}$$
$$\begin{pmatrix} X_2 & A^\top X_2 + C^\top M_2^\top \\ X_2 A + M_2 C & X_2 \end{pmatrix} \in \mathbb{S}_{n_x}^{++}(\mathbb{R}) \tag{1.28b}$$

In both cases, the gain matrices K and L are retrieved with $K = M_1 X_1^{-1}$ and $L = X_2^{-1} M_2$. Moreover, there exists $\lambda \in \mathbb{R}_{>0}$ such that the matrices P_1 and P_2 providing the QLF $V(x, \hat{x}) = \langle x | P_1 x \rangle + \langle \hat{x} - x | P_2 (\hat{x} - x) \rangle$ can be taken as $P_1 = \lambda X_1^{-1}$, $P_2 = X_2$,

Although the methodology above is rather unrefined compared to linear–quadratic (Gaussian) optimal control, or to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ synthesis, it has the benefit of being easily deduced from the item 3 of Lyapunov's lemma (Lemma 1.1.1), and of being directly applicable to the state feedback problem in the LDI framework. Assuming this time that the state $x(t)$ is fully measured, the following results hold for polytopic LDI systems:

Theorem 1.1.9 (Exponential stabilization of an LDI [48]). *Let $\{A_i, B_i\}_{1\leq i\leq m}$ be a set of $\mathbb{R}^{n_x\times n_x} \times \mathbb{R}^{n_x\times n_u}$ matrices. The polytopic exogenous input free LDI system (1.9) ($w = 0$) can be globally exponentially stabilized at the origin by the state feedback controller $u(t) = Kx(t)$ if there exists a symmetric matrix $X \in \mathbb{S}_{n_x}(\mathbb{R})$ and $M \in \mathbb{R}^{n_u\times n_x}$ such that*

– *in the continuous-time case:*

$$X \in \mathbb{S}_{n_x}^{++}(\mathbb{R}) \tag{1.29a}$$

$$XA_i^\top + A_iX + B_iM + M^\top B_i^\top \in \mathbb{S}_{n_x}^{--}(\mathbb{R}), \quad \forall i \in [\![1, m]\!] \tag{1.29b}$$

– *in the discrete-time case:*

$$\begin{pmatrix} X & XA_i^\top + M^\top B_i^\top \\ A_iX + B_iM & X \end{pmatrix} \in \mathbb{S}_{n_x}^{++}(\mathbb{R}), \quad \forall i \in [\![1, m]\!] \tag{1.30}$$

In both cases, the matrix P of the QLF (1.21) and the gain matrix K are retrieved with $P = X^{-1}$ and $K = MP$.

Despite the uncertainties of the LDI models, some robust pole placement strategies can still be considered for these systems [67, 214, 210]. However, contrary to the usual linear framework, there is no clear duality between the controller and the observer design problem for an LDI, as the inclusion making the system uncertain leads to a much harder observation problem. This observer design problem becomes approachable using a classical Luenberger observer if real-time information is available concerning which pair $(A(t), B(t))$ is active at each time t.

The lack of real-time information under the LDI framework motivates the introduction of a *scheduling vector* $\theta(t)$, generally assumed to be measured or estimated in real-time, and which contains information regarding which pair $\big(A(\theta(t)), B(\theta(t))\big)$ is active at each time t. In particular, this scheduling vector allows to change the coefficients of the gain matrices K and L in real-time, leading to *gain-scheduled* observers and controllers [252, 253, 236]. Results surrounding the control problems associated with this methodology are the main focus of this manuscript.

1.2 Outline of the thesis

This manuscript mainly investigates the special classes of nonlinear systems known as the Takagi-Sugeno (T-S) and Linear Parameter Varying (LPV) models. In both of these frameworks, the usual nonlinear control challenges are approached by using a convex rewriting of the nonlinear system (1.1). This rewriting, which can be local or global, leads to a kind of LTV state-space model where the time dependence is not explicitly obtained, but is implicitly represented using extrinsic and intrinsic signals from the system, gathered in a *scheduling vector* $\theta \in \Theta$. In practical terms, these systems can be represented using the following state-space model:

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = \begin{pmatrix} A(\theta(t)) & B_1(\theta(t)) & B_2(\theta(t)) \\ C_1(\theta(t)) & D_{11}(\theta(t)) & D_{12}(\theta(t)) \\ C_2(\theta(t)) & D_{21}(\theta(t)) & D_{22}(\theta(t)) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \tag{1.31}$$

By leveraging the convexity properties of the model above, sometimes together with assumptions on the rate of variation of θ, the T-S and LPV frameworks manage to obtain results which are not far removed from the usual linear framework. There is a sense in which the T-S and LPV frameworks sharpen the results from the LDI framework, as ignoring the information provided by the scheduling vector θ in (1.31) leads to the following differential inclusion:

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} \in \left\{ \begin{pmatrix} A(\theta) & B_1(\theta) & B_2(\theta) \\ C_1(\theta) & D_{11}(\theta) & D_{12}(\theta) \\ C_2(\theta) & D_{21}(\theta) & D_{22}(\theta) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} : \theta \in \Theta \right\} \tag{1.32}$$

This manuscript explores a variety of results for these systems on a broad range of topics, e.g. system-modeling, gain-scheduling control or fault-diagnosis, with an emphasis on the geometric nature of the contributions.

1.2.1 Contributions

The main contributions of this thesis concern the introduction of geometric tools that had not previously been used in the context of T-S and LPV systems. In particular:

- It is shown that barycentric coordinates play a key role in the modeling of T-S systems using the nonlinear sector approach (*Chapter 3*).
- It is shown that polyhedral complexes can be used to obtain non-convex T-S models (*Chapter 4*).
- It is established that Bézier interpolations provide a geometric understanding of the multi-sums involved in the T-S framework (*Chapter 5*).
- It is demonstrated that a Lipschitz assumption on the scheduling vector of a LPV system allows to bound all its potential state transition matrices in the future, leading to useful results to characterize the near-future of these systems (*Chapter 6*).
- A set-membership approach is explored for fault detection purposes by introducing the Minkowski functional of a set (*Chapter 7*).
- Finally, the modeling of saturations and other phenomena (dead zones, hysteresis) locally affecting the actuators of a system is approached using geometric tools, such as the Minkowski functional (*Chapter 8*).

1.2.2 Chapters contents

The contents of each chapter of the thesis is summarized in further detail thereafter:

Chapter 2: Introduction to the Takagi-Sugeno framework
The T-S framework is introduced in more details. In particular, this chapter lists classical results on LMI, as well as usual LMI conditions of stability and stabilization for T-S models, both in the continuous-time and the discrete-time cases.

Chapter 3: Convex modeling of Takagi-Sugeno systems
A T-S model is usually derived from a nonlinear system through a methodology known as the nonlinear sector approach. This methodology provides T-S models which exactly represent nonlinear systems with bounded nonlinearities. Yet, this methodology assumes that these nonlinearities are bounded by a box. In this chapter, the nonlinear sector approach is revealed to rely on barycentric coordinates. Barycentric coordinates being already well-studied, it is possible to introduce flexibility in the shape of the bounding set, by allowing it to be taken in a large class of convex shapes. This flexibility has strong implications both in terms of complexity of the resulting T-S model and of its intrinsic conservatism. This chapter has been published as an article [20].

Chapter 4: Non-convex modeling of Takagi-Sugeno systems
This chapter revisits the nonlinear sector approach introduced in Chapter 3 and applies it in a piecewise manner, leading to the introduction of non-convex T-S models. These models need extra-care in order to obtain non-conservative LMI conditions from them, as it is easy to make their non-convexity worthless. LMI conditions for stability benefiting fully from the non-convexity of these models are provided.

Chapter 5: Bézier interpolations in the Takagi-Sugeno framework
Multi-sums are ubiquitous in the T-S framework. However, it is well-known that most of the terms in these multi-sums are redundant. This chapter proposes to rewrite explicitly the multi-sums in a non-redundant manner by using Bernstein polynomials. This simple idea sheds light on the underlying geometric nature of the multi-sums of the T-S framework, which turn out to be Bézier interpolation schemes. Some T-S results are revisited under this fresh perspective, including results which have already been published by the author in a conference paper [18].

Chapter 6: Anticipating the near future of an LPV system
LPV models differ from LTV models by their time-dependence which is not *explicit*, but *implicitly* obtained by the intermediate of a scheduling vector θ. This key difference is unfortunately significant enough to make most of the results from the LTV framework inapplicable in an LPV setting. However, assuming a bounded rate of variations on θ, it is possible to quantify the maximal discrepancy between an anticipated future for θ and its real future evolution, leading to LTV-like results for the near future of LPV models. After introducing two powerful tools, namely Volterra's product integral and the weighted logarithmic norm of a matrix, this chapter uses this methodology to *exactly* discretize LPV models, as well as to anticipate the evolution of some structural properties, such as controllability and observability. Some results from this chapter are adapted from two of the author's published conference papers [17, 24].

Chapter 7: Fault-isolation using a set-membership approach

Fault detection and isolation schemes consist in detecting if a system is being subject to a fault (the detection), as well as to identify which fault is active (the isolation). This is generally achieved by synthesising residuals signals from the input and output of the system. In the absence of a fault, the residuals should have their values centered at 0. However, every system is subject to mild disturbances, and a good fault detection and isolation scheme should be robust to the harmless noise affecting the residuals. The set-membership approach to fault detection and isolation consists in treating the disturbances of the system as bounded inputs, resulting in the construction of thresholds on the residual signals. Identifying a fault using this methodology is generally achieved by using specific classes of convex sets (e.g. zonotopes, ellipsoids). Rather than focusing on such a class, this chapter describes in abstract terms a set-membership fault detection and isolation scheme for uncertain linear systems. This chapter introduces the Minkowski functional of a set to obtain theoretical results which can be applied to a broad class of set-based methodologies from the literature.

Chapter 8: Saturation and dead-zone modeling
Saturations, dead-zones, dead-bands or hysteresis effects are well-known actuators and sensors faults. This chapter introduces methodologies in order to model them, and obtains theoretical guarantees on the behaviour of a system under their influence. More specifically:

- A broad class of actuator saturations is modeled under the T-S framework, with a reduced number of local models compared to existing literature. This is achieved by leveraging the Minkowski functional introduced in Chapter 7. This part of the chapter has been published as a conference paper [21].
- A unified modeling of dead-zones, dead-bands and hysteresis effects is suggested, and is investigated for LTI systems. This part of the chapter has been published as a conference paper [23].

Chapter 9: Conclusions and perspectives
This chapter concludes the thesis and offers some perspective for future works.

Appendix A: CQLF of second-order systems: a graphical criterion
A very simple graphical criterion is proposed to obtain a common quadratic Lyapunov function to a set of second order LTI systems with real coefficients. The criterion associates every Hurwitz (resp. Schur) real matrix of $\mathbb{R}^{2\times 2}$ with the interior of an ellipse on a two-dimensional plane. If the intersection of all the ellipses associated to the set of Hurwitz (resp. Schur) matrices is non-empty, it can be stated without loss of generality that there exists a common quadratic Lyapunov function to this set. All existing common quadratic Lyapunov functions can actually be retrieved from this intersection.

Appendix B: Orthogonal projection of convex sets with a $\mathcal{C}^1$ boundary
Given an Euclidean space, a topological link is obtained between the partial derivatives of the Minkowski functional associated to a $\mathcal{C}^1$ convex set and the boundary of its orthogonal projection onto the linear subspaces of the Euclidean space. A system of equations for these orthogonal projections is derived from this topological link. This appendix corresponds to the arXiv deposit [22].

Appendix C: Résumé détaillé en français
A french summary of the thesis.

1.2.3 List of publications

Journal papers

- G. Bainier, B. Marx, J.-C. Ponsart (2024). Common Quadratic Lyapunov Functions for Sets of Second-Order Linear Systems: A Simple Graphical Criterion. *IEEE Control Systems Letters*. https://doi.org/10.1109/LCSYS.2024.3418672
- G. Bainier, B. Marx, J.-C. Ponsart (2024). Generalized nonlinear sector approaches for Takagi-Sugeno models. *Fuzzy Sets and Systems, 476*. https://doi.org/10.1016/j.fss.2023.108791

Submitted papers

- G. Bainier, B. Marx, J.-C. Ponsart. A Unified Set-Membership Approach to Fault Isolation using the Minkowski functional. *International Journal of Robust and Nonlinear Control.*

Conference papers

- G. Bainier, B. Marx, J.-C. Ponsart (2024). Bézier Controllers and Observers for Takagi-Sugeno Models. *2024 American Control Conference, Toronto, Canada.* https://doi.org/10.23919/ACC60939.2024.10644414
- G. Bainier, B. Marx, J.-C. Ponsart (2024). Modeling a broad class of actuator saturations using Takagi-Sugeno models with a reduced number of local models. *2024 American Control Conference, Toronto, Canada.* https://doi.org/10.23919/ACC60939.2024.10644172
- G. Bainier, B. Marx, J.-C. Ponsart (2024). A unified modeling of dead-zone, dead-band, hysteresis, and other faulty local behaviors of actuators and sensors. *Safeprocess 2024, Ferrara, Italy.* https://doi.org/10.1016/j.ifacol.2024.07.298
- G. Bainier, J.-C. Ponsart, B. Marx (2022). Anticipating the loss of unknown input observability for sampled LPV systems. *16th European Workshop on Advanced Control and Diagnosis, ACD 2022, Nancy, France.* https://doi.org/10.1007/978-3-031-27540-1_2
- G. Bainier, B. Marx, J.-C. Ponsart (2022). Bounding the Trajectories of Continuous-Time LPV Systems with Parameters known in Real Time. *5th IFAC Workshop on Linear Parameter Varying Systems, LPVS 2022, Montréal, Canada.* https://doi.org/10.1016/j.ifacol.2022.11.292

ArXiv deposit

- G. Bainier, B. Marx, J.-C. Ponsart (2023). Orthogonal Projection of Convex Sets with a Differentiable Boundary. https://doi.org/10.48550/arXiv.2302.08937

Chapter 2

Introduction to the Takagi-Sugeno framework

This chapter provides a detailed introduction to the T-S framework, presenting classical results on LMI, as well as standard LMI conditions of stability and stabilization for continuous-time and discrete-time T-S models.

2.1 Takagi-Sugeno systems

Several models have been suggested in the literature to offer a systematic approach to the control synthesis of the nonlinear system (1.1) of Chapter 1. Among them, a recurring idea consists in approximating the shifted state vector $\delta x(t)$ and eventually the output $y(t)$ and the regulated output $z(t)$ as a time-varying weighted sum of several LTI systems of the form (1.3), the so-called local models (or submodels), to obtain the convex model given thereafter:

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = \sum_{i=1}^{n_h} h_i(\theta) \begin{pmatrix} A_i & B_{1,i} & B_{2,i} \\ C_{1,i} & D_{11,i} & D_{12,i} \\ C_{2,i} & D_{21,i} & D_{22,i} \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \tag{2.1}$$

where:

- $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$ is a vector of scheduling parameters which usually depend on $x(t)$, $u(t)$, $w(t)$ and t;
- $\mathbf{h} \triangleq (h_1, \ldots, h_{n_h})$ are the activation functions, also known as the membership functions or the weighting functions (h_i provides the weight of the i-th local model of the system).

It is easily noticed that (2.1) can be rewritten as the LPV model (1.31) introduced in Chapter 1. This model is considered convex, as the activation functions $\mathbf{h}$ satisfy the convex sum properties.

Definition 2.1.1 (Convex sum properties). *A set of functions $h_1, \ldots, h_{n_h} : \Theta \to \mathbb{R}$ is said to satisfy the convex sum properties if for all $\theta \in \Theta$, the functions verify the two following properties:*

$$h_1(\theta), \ldots, h_{n_h}(\theta) \geq 0 \qquad [\textit{positivity}] \tag{2.2a}$$

$$h_1(\theta) + \cdots + h_{n_h}(\theta) = 1 \qquad [\textit{partition of unity}] \tag{2.2b}$$

These two properties can be summarized by the inclusion of the values of $\mathbf{h} \triangleq (h_1, \ldots, h_{n_h})$ *in the* $(n_h - 1)$*-simplex* Δ_{n_h-1}*. Indeed, for all* $\theta \in \Theta$*, (2.2) is equivalent to (2.3).*

$$\mathbf{h}(\theta) \in \Delta_{n_h-1} \triangleq \left\{ (h_1, \ldots, h_{n_h}) \in \mathbb{R}^{n_h}_{\geq 0} : h_1 + \cdots + h_{n_h} = 1 \right\} \tag{2.3}$$

Several terminologies are employed in the literature to refer to this kind of model:

- the T-S (fuzzy) model, introduced by the fuzzy logic community [268];
- the polytopic (quasi-)LPV model, introduced for gain-scheduling techniques [252, 10];
- the multiple model [198];
- the linear polytopic model [8].

In particular, the quasi-LPV terminology is reserved for cases where θ depends on the state $x(t)$, whereas the linear polytopic model terminology is reserved for cases where θ does not depend on the state $x(t)$ [32].

Remark 2.1.1. *An LPV model whose scheduling vector* θ *depends on the input* $u(t)$ *generally becomes a quasi-LPV model once put in a closed-loop architecture.*

As a special case, (2.1) is also equivalent to a flattened tensor-product model [25], to a polynomial fuzzy model with only linear terms [161], or to a polytopic LDI model with coordinates known in real-time [48]. Note that many other representations are typically investigated for LPV and T-S systems, including models with a linear fractional transformation structure, with bounded matrix uncertainties, with a singular structure (also referred to as descriptor systems), with time-delays, with constraints on the signals, etc. These cases are not discussed in this chapter and the reader is referred to other works where they are considered in more details [272, 286, 52, 86, 32].

Contrary to the LTV model (1.3) of Chapter 1, which approximates the nonlinear system (1.1) locally using the linearization procedure suggested in equation (1.7), the T-S model (2.1) is able to *exactly* represent the nonlinear system (1.1), locally or globally (in case of bounded nonlinearities), using a technique called the nonlinear sector approach [206, 295]. This technique is discussed in details in the Chapter 3 of this manuscript. For now, the reader should keep in mind that the strong representation capabilities of T-S (and LPV) models are obtained at the cost of:

- an exponentially growing number of local models with respect to the size of the scheduling vector ($n_h = 2^{n_\theta}$),
- not explicitly knowing the evolution of the scheduling vector θ in advance.

Indeed, contrary to the LTV framework (1.3) discussed in the introduction, not only is the number of local models critical to the applicability of the T-S framework, but it is also not possible to know which local LTI systems in (2.1) are active at a given time t without further information on the system's initial conditions, its disturbances, etc. Despite the literature often assuming that θ is measured or estimated in real-time, this real-time knowledge is not as powerful as the all-time knowledge that comes with LTV systems. This makes most of the results from the LTV

framework not applicable in a T-S context. Typically, the Gramian-based controllability and observability criteria given in Theorem 1.1.2 of Chapter 1 are inapplicable for T-S systems.

In spite of these disadvantages, thanks to the convex nature of the T-S models, many control problems with no simple analytical solution can be still be numerically solved through convex optimization techniques, in particular by formulating them using LMI conditions [48, 272, 32]. The rest of this chapter is dedicated to the LMI formulation of said control problems. After defining and investigating usual results on LMI in Section 2.2, Section 2.3 and Section 2.4 investigate the stability and stabilization of continuous-time and discrete-time T-S models.

2.2 Linear matrix inequalities

An LMI consists of an analytical expression of the form:

$$F(z) \triangleq F_0 + z_{(1)}F_1 + \cdots + z_{(m)}F_m \in \mathbb{S}_n^+(\mathbb{R}) \tag{2.4}$$

where $F_1, \ldots, F_m \in \mathbb{S}_n(\mathbb{R})$ and $z \in \mathbb{R}^m$. The coordinates of $z \in \mathbb{R}^m$ are sometimes called the decision variables of the LMI, as they correspond to the unknown quantities to be found. In practice, the coordinates of z are often gathered in several matrices, making the equation above a developed expression which is rarely made explicit. This expression is called an inequality, in the sense that it can be rewritten using a partial ordering in the space of symmetric matrices $\mathbb{S}_n(\mathbb{R})$ known as the Loewner order (denoted $\succeq$).

$$F(z) \in \mathbb{S}_n^+(\mathbb{R}) \iff F(z) \succeq 0 \tag{2.5}$$

The inequality can be made strict (which is denoted $\succ$) by considering the cone of positive definite matrices $\mathbb{S}_n^{++}(\mathbb{R})$ rather than the cone of positive semidefinite matrices $\mathbb{S}_n^+(\mathbb{R})$. It can also be inverted by considering the cone of negative (semi)definite matrices. In any case, the matrix inequality is considered linear since, ignoring the affine term ($F_0 = 0$):

$$F(\lambda_1 z_1 + \lambda_2 z_2) = \lambda_1 F(z_1) + \lambda_2 F(z_2), \quad \forall z_1, z_2 \in \mathbb{R}^m, \forall \lambda_1, \lambda_2 \in \mathbb{R} \tag{2.6}$$

This property, taken together with the fact that $\mathbb{S}_n^+(\mathbb{R})$ and $\mathbb{S}_n^{++}(\mathbb{R})$ are convex cones, guarantees that the solution space $\{z \in \mathbb{R}^m : F(z) \in \mathbb{S}_n^{+(+)}(\mathbb{R})\}$ is also convex, and is moreover a cone if $F_0 = 0$. Solutions to an LMI are generally computed using semidefinite programming.

Example 2.2.1. *The equations* (1.22), (1.27), (1.28), (1.29) *and* (1.30) *of Chapter 1 are LMI.*

This section introduces rigorously the Loewner order on the space of symmetric matrices $\mathbb{S}_n(\mathbb{R})$ and recalls some crucial preliminary results often leveraged in the T-S literature.

2.2.1 The Loewner order

The Loewner order on the space of symmetric matrices $\mathbb{S}_n(\mathbb{R})$ is defined as follows.

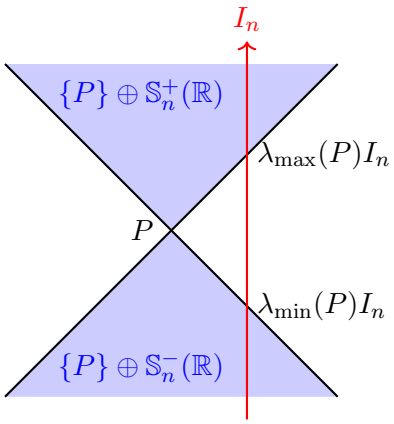


Figure 2.1: Schematic illustration of Property 2.2.1.

Definition 2.2.1 (Loewner order). *Given $P, Q \in \mathbb{S}_n(\mathbb{R})$, the Loewner order $\preceq$ is defined by the following equivalence:*

$$Q \preceq P \iff P - Q \in \mathbb{S}_n^+(\mathbb{R}) \tag{2.7}$$

Similarly, the strict Loewner order $\prec$ is defined by:

$$Q \prec P \iff P - Q \in \mathbb{S}_n^{++}(\mathbb{R}) \tag{2.8}$$

The symbols $\succeq$ and $\succ$ are defined similarly.

This definition can be geometrically understood by noticing that $Q \preceq P$ if and only if $P \in \{Q\} \oplus \mathbb{S}_n^+(\mathbb{R})$ (or similarly, if and only if $Q \in \{P\} \oplus \mathbb{S}_n^-(\mathbb{R})$). Rewriting the inequalities of Property 1.1.1 leads to the following result, which, thanks to the geometrical interpretation of the Loewner order, can be understood intuitively with Figure 2.1.

Property 2.2.1 (Classical inequality). *Let $P \in \mathbb{S}_n(\mathbb{R})$. The following inequalities hold:*

$$\lambda_{\min}(P)I_n \preceq P \preceq \lambda_{\max}(P)I_n \tag{2.9}$$

The matrices $\lambda_{\min}(P)I_n$ and $\lambda_{\max}(P)I_n$ are respectively at the boundary of $\{P\} \oplus \mathbb{S}_n^-(\mathbb{R})$ and $\{P\} \oplus \mathbb{S}_n^+(\mathbb{R})$ due to the inequality above being sharp, as discussed in Property 1.1.1. A few other elementary properties on the Loewner order are recalled below:

Property 2.2.2 (Congruence [128]). *Let $P \in \mathbb{S}_n(\mathbb{R})$ and $M \in \mathbb{R}^{m \times n}$, then*

$$P \succ 0 \Rightarrow MPM^\top \succeq 0 \tag{2.10}$$

Moreover, if M is full row rank, the following implications hold:

$$P \succ 0 \Rightarrow MPM^\top \succ 0 \tag{2.11}$$

The reciprocal holds if $m = n$, i.e. if $M \in \mathrm{GL}_n(\mathbb{R})$. The same properties hold for $(\preceq, \prec)$ by symmetry.

Property 2.2.3 (Inverse [128]). *Let $P, Q \in \mathbb{S}_n(\mathbb{R})$.*

$$P \succeq Q \succ 0 \Rightarrow Q^{-1} \succeq P^{-1} \succ 0 \tag{2.12a}$$
$$P \succ Q \succ 0 \Rightarrow Q^{-1} \succ P^{-1} \succ 0 \tag{2.12b}$$

The same properties hold for $(\preceq, \prec)$ by symmetry.

Property 2.2.4 (Perturbation). *Let $P \in \mathbb{S}_n^{++}(\mathbb{R})$ and $Q \in \mathbb{S}_n(\mathbb{R})$. There exists $r \in \mathbb{R}_{>0}$ such that for all $\varepsilon \in (0, r)$:*

$$P + \varepsilon Q \succ 0 \tag{2.13}$$

The same property holds for $\mathbb{S}_n^{--}(\mathbb{R})$ and $\prec$ by symmetry.

Property 2.2.5 (Insertion). *Let $P, Q \in \mathbb{S}_n^{++}(\mathbb{R})$. There exists $r \in \mathbb{R}_{>0}$ such that for all $\varepsilon \in (0, r)$:*

$$P \succ \varepsilon Q \succ 0 \tag{2.14}$$

The same property holds for $\mathbb{S}_n^{--}(\mathbb{R})$ and $\prec$ by symmetry.

Proof. The perturbation property is a trivial consequence of $\mathbb{S}_n^{++}(\mathbb{R})$ being an open set. The insertion property, despite being very easy, common and useful, is rarely explicitly reported in the literature, hence it is demonstrated here. Given $P, Q \in \mathbb{S}_n^{++}(\mathbb{R})$, it is clear that there exists $r \in \mathbb{R}_{>0}$ such that for all $\varepsilon \in (0, r)$, $\lambda_{\min}(P) > \varepsilon \lambda_{\max}(Q) > 0$, hence $\lambda_{\min}(P) + \lambda_{\min}(-\varepsilon Q) > 0$. Weyl's inequalities on symmetric matrices [38] yield in particular $\lambda_{\min}(P) + \lambda_{\min}(-\varepsilon Q) \leq \lambda_{\min}(P - \varepsilon Q)$, hence $\lambda_{\min}(P - \varepsilon Q) > 0$, implying $P - \varepsilon Q \succ 0$ and so $P \succ \varepsilon Q$, where $\varepsilon Q \succ 0$ since $\varepsilon > 0$. □

Remark 2.2.1. *The case $P = I_n$ in (2.11) (resp. (2.10)) guarantees that $MM^\top$ is always positive (semi)definite. Reciprocally, all positive semidefinite matrices $Q \in \mathbb{S}_n^+(\mathbb{R})$ admit such a decomposition. In particular, if $Q \in \mathbb{S}_n^{++}(\mathbb{R})$ and if M is a lower triangular matrix with positive entries on the main diagonal, then this decomposition is in fact unique and corresponds to a Cholesky factorization of Q [108].*

There are arguably three key equivalences in the literature of LMI, namely:

- Schur's complement,
- the quadratic S-procedure,
- Finsler's lemma.

These three results are recalled thereafter, beginning with Schur's complement.

Lemma 2.2.1 (Schur's complement). *Let $P \in \mathbb{S}_n(\mathbb{R})$ such that*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \tag{2.15}$$

with $P_{11} \in \mathbb{S}_k(\mathbb{R})$ a square matrix (hence P_{22} is a square matrix as well, and $P_{12} = P_{21}^\top$). The following equivalences hold:

$$P \succ 0 \Leftrightarrow \begin{cases} P_{11} \succ 0 \\ P_{22} - P_{21}P_{11}^{-1}P_{12} \succ 0 \end{cases} \Leftrightarrow \begin{cases} P_{22} \succ 0 \\ P_{11} - P_{12}P_{22}^{-1}P_{21} \succ 0 \end{cases} \tag{2.16}$$

The same property holds for $\prec$ by symmetry.

Proof. Usually, the proof is performed using a congruence argument [99]. An amusing and unusual proof of this result consists in introducing quadratic sets of the form $\mathcal{E}(Q) \triangleq \{x \in \mathbb{R}^m : x^\top Q x \leq 1\}$. It is well-established that $\mathcal{E}(Q)$ is an ellipsoid if and only if $Q \succ 0$ [249]. The proof of Schur's complement is now performed by double implication using geometric arguments.

$\boxed{\Rightarrow}$ If $P \succ 0$, then $\mathcal{E}(P)$ is an ellipsoid. In particular, the intersection of $\mathcal{E}(P)$ with the subspace $S \triangleq \operatorname{span}\begin{pmatrix} I_k \\ 0 \end{pmatrix}$ is the ellipsoid $\mathcal{E}(P_{11})$, so $P_{11} \succ 0$. Similarly, the projection of $\mathcal{E}(P)$ onto $S^\perp = \operatorname{span}\begin{pmatrix} 0 \\ I_{n-k} \end{pmatrix}$ is an ellipsoid as well [249]. Leveraging the results from Appendix B of this manuscript guarantees that this ellipsoid is in fact the set $\mathcal{E}(P_{22} - P_{21}P_{11}^{-1}P_{12})$, hence $P_{22} - P_{21}P_{11}^{-1}P_{12} \succ 0$.

$\boxed{\Leftarrow}$ Reciprocally, assuming $P_{11} \succ 0$ and $P_{22} - P_{21}P_{11}^{-1}P_{12} \succ 0$, it is deduced from the fact that both $\mathcal{E}(P_{11}) = \mathcal{E}(P) \cap S$ and the projection of $\mathcal{E}(P)$ onto $S^\perp$ (i.e. $\mathcal{E}(P_{22} - P_{21}P_{11}^{-1}P_{12})$) are two ellipsoids that $\mathcal{E}(P)$ is bounded and non empty, hence it must be an ellipsoid as well. Assuming $\mathcal{E}(P)$ is unbounded, P must have an eigenvalue in $\mathbb{R}_{\leq 0}$ (symmetric matrices have real eigenvalues). If its associated eigenvector v is such that $v \notin S$, then by projection of $\mathcal{E}(P)$ onto $S^\perp$, $\mathcal{E}(P_{22} - P_{21}P_{11}^{-1}P_{12})$ would be unbounded. Yet $\mathcal{E}(P_{22} - P_{21}P_{11}^{-1}P_{12})$ is bounded, hence $v \in S$. However, taking $v = \begin{pmatrix} v_1^\top & 0 \end{pmatrix}^\top \in S$ yields $v^\top P v = v_1^\top P_{11} v_1$, and since $P_{11} \succ 0$, P cannot have any eigenvalue in $\mathbb{R}_{\leq 0}$. In the end, $\mathcal{E}(P)$ is non-empty and bounded, which provides $P \succ 0$ and concludes the proof. □

To the author's knowledge, there exists four variations of the S-procedure in its quadratic form, depending if it is considered strict or loose, and homogeneous or not. The strict homogeneous and loose in-homogeneous cases are given below, as they are the most usual in the literature. The remarks afterward include a discussion on the two remaining cases.

Lemma 2.2.2 (Strict homogeneous S-procedure [48]). *Let $P_0, P_1, \ldots, P_m \in \mathbb{S}_n(\mathbb{R})$, $\mathcal{S}_{P_0}^- \triangleq \left\{z \in \mathbb{R}^n : z^\top P_0 z < 0\right\}$ and for all $i \in [\![1, m]\!]$, let $\mathcal{S}_{P_i}$ denote the set:*

$$\mathcal{S}_{P_i} \triangleq \left\{z \in \mathbb{R}^n : z^\top P_i z \leq 0\right\} \tag{2.17}$$

In the following, the first statement implies the second one.

1. *There exists $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_{\geq 0}$ s.t. $P_0 \prec \lambda_1 P_1 + \cdots + \lambda_m P_m$.*

2. $\mathcal{S}_{P_1} \cap \cdots \cap \mathcal{S}_{P_m} \subseteq \mathcal{S}^-_{P_0} \cup \{0\}$.

Moreover, if $m = 1$ and if there exists $z \in \mathbb{R}^n$ such that $z^\top P_1 z < 0$, then the statements are equivalent. The same property holds for $(>, \geq, \succ)$ by symmetry.

Lemma 2.2.3 (Loose in-homogeneous S-procedure [48, 49])**.** *Let $P_0, P_1, \ldots, P_m \in \mathbb{S}_{n+1}(\mathbb{R})$, and for all $i \in [\![0, m]\!]$, let $\mathcal{S}_{P_i}$ denote the set:*

$$\mathcal{S}_{P_i} \triangleq \left\{ z \in \mathbb{R}^n : \begin{pmatrix} z^\top & 1 \end{pmatrix} P_i \begin{pmatrix} z \\ 1 \end{pmatrix} \leq 0 \right\} \tag{2.18}$$

In the following, the first statement implies the second one.

1. *There exists $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_{\geq 0}$ s.t. $P_0 \preceq \lambda_1 P_1 + \cdots + \lambda_m P_m$.*
2. *$\mathcal{S}_{P_1} \cap \cdots \cap \mathcal{S}_{P_m} \subseteq \mathcal{S}_{P_0}$.*

Moreover, if $m = 1$ and if there exists $z \in \mathbb{R}^n$ such that $\begin{pmatrix} z^\top & 1 \end{pmatrix} P_1 \begin{pmatrix} z \\ 1 \end{pmatrix} < 0$, then the statements are equivalent. The same property holds for $(\geq, \succeq, >)$ by symmetry.

Remark 2.2.2. *The loose homogeneous S-procedure holds by using $P_i = \begin{pmatrix} Q_i & 0 \\ 0 & 0 \end{pmatrix}$ in the loose in-homogeneous lemma.*

Remark 2.2.3. *The strict in-homogeneous S-procedure only holds for $1 \Rightarrow 2$, and its (elementary) proof is reported below.*

Proof. $\boxed{1. \Rightarrow 2.}$ It is straightforward by introducing $\mathcal{H} \triangleq \{z \in \mathbb{R}^n : z_{(n)} = 1\}$ in the strict homogeneous lemma:

$$\mathcal{S}_{P_1} \cap \cdots \cap \mathcal{S}_{P_m} \subseteq \mathcal{S}^-_{P_0} \cup \{0\} \Rightarrow \mathcal{H} \cap \mathcal{S}_{P_1} \cap \cdots \cap \mathcal{S}_{P_m} \subseteq \mathcal{H} \cap (\mathcal{S}^-_{P_0} \cup \{0\}) \subseteq (\mathcal{H} \cap \mathcal{S}^-_{P_0}) \cup \{0\} \tag{2.19}$$

$\boxed{1. \not\Leftarrow 2.}$ The reciprocal does not hold in the strict in-homogeneous case. Indeed, considering:

$$P_0 = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 1 & 1/4 \\ 1/4 & 0 \end{pmatrix} \tag{2.20}$$

The sets of interest in the strict in-homogeneous case are as follows:

$$\left\{ z \in \mathbb{R} : \begin{pmatrix} z & 1 \end{pmatrix} P_0 \begin{pmatrix} z \\ 1 \end{pmatrix} < 0 \right\} = (-1, 0) \tag{2.21a}$$

$$\left\{ z \in \mathbb{R} : \begin{pmatrix} z & 1 \end{pmatrix} P_1 \begin{pmatrix} z \\ 1 \end{pmatrix} \leq 0 \right\} = [-1/2, 0] \tag{2.21b}$$

Clearly, item 2 is verified, as $[-1/2, 0] \subseteq (-1, 0) \cup \{0\}$, moreover $m = 1$ and for $z = -1/4$:

$$\begin{pmatrix} -1/4 & 1 \end{pmatrix} P_1 \begin{pmatrix} -1/4 \\ 1 \end{pmatrix} = -1/16 < 0 \tag{2.22}$$

All conditions should be met for item 1 to be verified. However, for all $\lambda \geq 0$:

$$\begin{pmatrix} 0 & 1 \end{pmatrix} (P_0 - \lambda P_1) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0 \tag{2.23}$$

meaning there is no $\lambda \geq 0$ such that $P_0 \prec \lambda P_1$. □

Similarly to the S-procedure, there also exists many variations of Finsler's lemma. Three of them are presented in this manuscript:

- the strict Finsler's lemma, reported here for its clear resemblance to the S-procedure;
- the positive semidefinite Finsler's lemma, which is a richer version of the strict lemma that holds for positive semidefinite matrices;
- the simultaneous Finsler's lemma, also known as the elimination lemma, or the projection lemma, reported here for its wide use in control.

Lemma 2.2.4 (Strict Finsler's lemma [313]). *Let $P_0, P_1, \ldots, P_m \in \mathbb{S}_n(\mathbb{R})$, $\mathcal{F}^-_{P_0} \triangleq \left\{z \in \mathbb{R}^n : z^\top P_0 z < 0\right\}$ and for all $i \in [\![1, m]\!]$, let $\mathcal{F}^0_{P_i}$ denote the set:*

$$\mathcal{F}^0_{P_i} \triangleq \left\{z \in \mathbb{R}^n : z^\top P_i z = 0\right\} \tag{2.24}$$

In the following, the first statement implies the second one.

1. *There exists $\mu_1, \ldots, \mu_m \in \mathbb{R}$ s.t. $P_0 \prec \mu_1 P_1 + \cdots + \mu_m P_m$.*
2. *$\mathcal{F}^0_{P_1} \cap \cdots \cap \mathcal{F}^0_{P_m} \subseteq \mathcal{F}^-_{P_0} \cup \{0\}$.*

Moreover, if $m = 1$, then the statements are equivalent. The same property holds for $(>,\succ)$ by symmetry.

Lemma 2.2.5 (Positive semidefinite Finsler's lemma [135]). *Let $M \in \mathbb{R}^{m\times n}$ be such that $\mathrm{rank}(M) < n$ and $P \in \mathbb{S}_n(\mathbb{R})$. The following statements are equivalent.*

1. *There exists $\mu \in \mathbb{R}$ s.t. $P \prec \mu M^\top M$.*
2. *$\mathrm{Ker}(M) \subseteq \left\{z \in \mathbb{R} : z^\top P z < 0\right\} \cup \{0\}$.*
3. *$(M^\perp)^\top P M^\perp \prec 0$.*
4. *There exists $X \in \mathbb{R}^{n\times m}$, s.t. $P + XM + M^\top X^\top \prec 0$*

Lemma 2.2.6 (Simultaneous Finsler's lemma [48]). *Let $M \in \mathbb{R}^{m\times n}$, $N \in \mathbb{R}^{p\times n}$ be such that $\mathrm{rank}(M) < n$ and $\mathrm{rank}(N) < n$, and $P \in \mathbb{S}_n(\mathbb{R})$. The following statements are equivalent.*

1. *There exists $\mu \in \mathbb{R}$ s.t. $P \prec \mu M^\top M$ and $P \prec \mu N^\top N$.*
2. *$\mathrm{Ker}(M) \cup \mathrm{Ker}(N) \subseteq \left\{z \in \mathbb{R} : z^\top P z < 0\right\} \cup \{0\}$.*
3. *$(M^\perp)^\top P M^\perp \prec 0$ and $(N^\perp)^\top P N^\perp \prec 0$.*
4. *There exists $X \in \mathbb{R}^{p\times m}$ s.t. $P + NXM + M^\top X^\top N^\top \prec 0$.*

Remark 2.2.4. *To the author's knowledge, the strict Finsler's lemma (Lemma 2.2.4) is not discussed in the literature when several variables $\mu_1, \dots, \mu_m$ are involved, so its (elementary) proof is reported below for $m > 1$.*

Proof. $\boxed{1. \Rightarrow 2.}$ Assuming item 1 holds, if $x \in \mathcal{F}^0_{P_1} \cap \cdots \cap \mathcal{F}^0_{P_m}$, then either $x = 0$, or item 1 directly provides $x^\top P_0 x < \mu_1 x^\top P_1 x + \cdots + \mu_m x^\top P_m x = 0$, resulting in $x \in \mathcal{F}^-_{P_0} \cup \{0\}$.

$\boxed{1. \not\Leftarrow 2.}$ The reciprocal does not hold in general for $m > 1$. Indeed, considering:

$$P_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.25}$$

It is easily found that $\mathcal{F}^-_{P_0} = \emptyset$, $\mathcal{F}^0_{P_1} = \{(x, y) : x = 0\} \cup \{(x, y) : y = 0\}$ and $\mathcal{F}^0_{P_2} = \{(x, y) : x = y\} \cup \{(x, y) : x = -y\}$. Clearly, $\mathcal{F}^0_{P_1} \cap \mathcal{F}^0_{P_2} \subseteq \mathcal{F}^-_{P_0} \cup \{0\}$, hence item 2 is verified. However, for all $\mu_1, \mu_2 \in \mathbb{R}$, $\mathrm{Tr}(P_0 - \mu_1 P_1 - \mu_2 P_2) = 2$, and the affine hyperplane $\mathrm{Tr}(\cdot) = 2$ of $\mathbb{S}_2(\mathbb{R})$ does not intersect $\mathbb{S}_2^{--}(\mathbb{R})$, making item 1 impossible to satisfy (see the ice cream Lemma A.2.1 in Appendix A for more details on the geometry of $\mathbb{S}_2^{--}(\mathbb{R})$). □

Remark 2.2.5. *The similarity between the strict Finsler's lemma (Lemma 2.2.4) and the S-procedure (Lemma 2.2.2) should be clearly visible: the variables μ_i in item 1 of Finsler's lemma play the role of the variables λ_i in item 1 of the S-procedure. In Finsler's lemma, the μ_i are however allowed to take negative values. This relaxation is explained by the fact that the inclusion which needs to be verified in item 2 is not as demanding as in the S-procedure, since:*

$$\mathcal{F}^0_P = \{z \in \mathbb{R}^n : z^\top P z = 0\} \subseteq \{z \in \mathbb{R}^n : z^\top P z \leq 0\} = \mathcal{S}_P \tag{2.26}$$

Remark 2.2.6. *In the positive semidefinite lemma (Lemma 2.2.5), the condition $\mathrm{rank}(M) < n$ ensures that $\dim \ker(M) > 0$, which in turn ensures that $M^\top M$ is positive semidefinite but not positive definite.*

Finally, a last inequality is often leveraged when dealing with LMI in the T-S framework, namely Young's relation, also known as Young's inequality, or the completion of squares property.

Lemma 2.2.7 (Young's relation)**.** *Let $M, N \in \mathbb{R}^{n \times m}$ and $P \in \mathbb{S}_n^{++}(\mathbb{R})$. The following holds:*

$$M^\top N + N^\top M \preceq M^\top P M + N^\top P^{-1} N \tag{2.27}$$

Proof. This inequality can be derived from the fact that $(M - PN)^\top P^{-1}(M - PN) \succeq 0$ by Properties 2.2.3 and 2.2.2. □

The reader is referred to [128, 307] for a more complete exposition on the properties of the Loewner order.

2.2.2 Results on double convex sums

The convex sum (2.1) of the T-S framework eventually leads to the study of LMI problems on convex sums, such as finding $z \in \mathbb{R}^m$ satisfying:

$$\sum_{i=1}^{n_h} h_i(\theta(t)) F_i(z) \prec 0, \qquad \forall t \in [t_0, +\infty) \tag{2.28}$$

where for all $i \in [\![1, n_h]\!]$, $F_i : \mathbb{R}^m \to \mathbb{S}_n(\mathbb{R})$ is an affine function of z. In practice, the scheduling vector $\theta \in \Theta$, which is only known in real-time, has to be partly neglected in order to obtain a tractable optimization problem from (2.28). The simplest procedure consists in rewriting (2.28) as follows:

$$\sum_{i=1}^{n_h} h_i F_i(z) \prec 0, \qquad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.29}$$

The solution set $\{z \in \mathbb{R}^m : (2.29)\}$ is only included in $\{z \in \mathbb{R}^m : (2.28)\}$ and no guarantees of equality hold in general: the potential size reduction between the two sets is called *conservatism*. Finding a solution to (2.29) also provides a solution to (2.28), but not finding one does not guarantee that (2.28) has no solution. Obtaining a tractable optimization problem from an inequality such as (2.28) usually requires the introduction of some amount of conservatism. The more substantial the size reduction of the solution set is, the more conservatism is introduced in the problem.

After the introduction of (2.29), the problem is easily dealt with by noticing, using a simple argument of convexity, that (2.29) holds if and only if $z \in \mathbb{R}^m$ is such that:

$$F_i(z) \prec 0, \qquad \forall i \in [\![1, n_h]\!] \tag{2.30}$$

This time, the size of the solution space $\{z \in \mathbb{R}^m : (2.30)\}$ is not reduced compared to $\{z \in \mathbb{R}^m : (2.29)\}$: no conservatism is introduced by rewriting (2.29) as an LMI. This equivalence guarantees that the only conservatism introduced in the optimization problem (2.28) is due to its rewriting as (2.29), i.e. is due to the interplay between the real-time knowledge of the scheduling vector $\theta \in \Theta$ and its relationship to the values of the activation functions $\mathbf{h}$.

It has now been demonstrated that rewriting the single convex sum problem (2.29) as an LMI is easily accomplished. However, this rewriting step becomes immediately problematic when a *double convex sum* is investigated:

$$\sum_{i=1}^{n_h} \sum_{j=1}^{n_h} h_i h_j F_{ij}(z) \prec 0, \qquad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.31}$$

Indeed, a proper LMI formulation of the double convex sum problem (2.31) generally requires the introduction of some amount of conservatism. This double convex sum problem is unfortunately extremely common under the T-S framework, as it appears naturally in the conditions of stabilization obtained for gain-scheduled controllers and observers. Several usual LMI formulations of (2.31) are presented below, all introducing some degree of conservatism in the problem. This sequence of LMI conditions are sometimes called *relaxations* of (2.31), in the sense that each new version enlarges the size of the set of feasible solutions compared to the previous one, thus gradually diminishing the amount of conservatism introduced in (2.31). However, this reduction in conservatism comes at the expense of increased complexity of the LMI conditions.

Lemma 2.2.8 (Double sum relaxation 1 [291])**.** *The inequality* (2.31) *is satisfied if the following conditions hold:*

$$F_{ij}(z) + F_{ji}(z) \prec 0, \quad \forall i, j \in [\![1, n_h]\!] \text{ s.t. } i \leq j \tag{2.32}$$

Lemma 2.2.9 (Double sum relaxation 2 [285]). *The inequality (2.31) is satisfied if the following conditions hold:*

$$\frac{2}{n_h-1}F_{ii}(z)+F_{ij}(z)+F_{ji}(z)\prec 0,\quad \forall i,j\in[\![1,n_h]\!] \text{ s.t. } i\leq j \tag{2.33}$$

Lemma 2.2.10 (Double sum relaxation 3 [301]). *The inequality (2.31) is satisfied if there exist $\{G_{ij}\}_{1\leq i,j\leq n_h}$, with $G_{ij}=G_{ji}^\top\in\mathbb{R}^{n_x\times n_x}$, and such that the following conditions hold:*

$$F_{ij}(z)+F_{ji}(z)+G_{ij}+G_{ji}\prec 0,\quad \forall i,j\in[\![1,n_h]\!] \text{ s.t. } i\leq j \tag{2.34a}$$

$$\begin{pmatrix} G_{11} & G_{12} & \dots & G_{1n_h}\\ G_{21} & G_{22} & \dots & G_{2n_h}\\ \vdots & & \ddots & \vdots\\ G_{n_h1} & G_{n_h2} & \dots & G_{n_hn_h}\end{pmatrix}\succ 0 \tag{2.34b}$$

Several procedures exist in order to obtain LMI conditions of arbitrarily large dimension from (2.31), with the guarantee that the increasing dimension of the optimization problem eventually cancels its conservatism, leading to necessary and sufficient conditions. The interested reader is referred to [238, 157] for more details. In particular, the technique of [238] is discussed in more details in the Chapter 5 of this manuscript.

2.3 Stability

This section investigates LMI conditions of state stability for input-free T-S models. These conditions are obtained by relying on the direct Lyapunov method introduced in Section 1.1.2.2 of Chapter 1.

2.3.1 Continuous-time stability

The continuous-time and input-free T-S model (2.35) is considered.

$$\dot{x}(t)=\sum_{i=1}^{n_h}h_i(\theta)A_ix(t) \tag{2.35}$$

Introducing a QLF is the simplest way to obtain an LMI condition, that, if satisfied, demonstrates the exponential stability of the T-S model (2.35). Given the QLF (1.21) of Chapter 1, i.e. $V(x)=x^\top Px$ with $P\in\mathbb{S}_{n_x}(\mathbb{R})$, this LMI condition, corresponding to Theorem 1.1.6 of Chapter 1, is stated as follows:

Theorem 2.3.1 (Quadratic stability [291]). *The T-S model (2.35) is globally exponentially stable if there exists $P\in\mathbb{S}_{n_x}(\mathbb{R})$ such that the LMI conditions (2.36) are satisfied.*

$$P\succ 0 \tag{2.36a}$$

$$\mathcal{H}(PA_j)\prec 0,\quad \forall j\in J \tag{2.36b}$$

J is a subset of $[\![1,n_h]\!]$ such that for all $i\in[\![1,n_h]\!]$, $A_i\in\text{hull}\{A_j : j\in J\}$.

Proof. Assuming (2.36) holds, three proofs of this result are given. The first proof, and the most straightforward, consists in noticing that the trajectories of (2.35) are included in the trajectories of the polytopic LDI $\dot{x}(t) \in \text{hull}\{A_j x(t) : j \in J\}$. Applying Theorem 1.1.6 of Chapter 1 concludes the first proof. The second proof consists in checking the Lyapunov conditions for exponential stability in Theorem 1.1.5 of Chapter 1: this is not detailed. The third proof consists in verifying that the exponential stability property (Definition 1.1.10 of Chapter 1) is verified. Succinctly, there exists $\varepsilon \in \mathbb{R}_{>0}$ such that $\mathcal{H}(PA_j) \prec -2\varepsilon P$ for all $j \in J$ (Property 2.2.5). By convexity, for all $A \in \text{hull}\{A_j : j \in J\}$, $\mathcal{H}(PA) \prec -2\varepsilon P$, hence for all $\theta \in \Theta$:

$$\mathcal{H}\left(P \sum_{i=1}^{n_h} h_i(\theta) A_i\right) \prec -2\varepsilon P \tag{2.37}$$

Introducing (1.21) as a QLF, the previous equation provides $\dot{V}(x(t)) \leq -2\varepsilon V(x(t))$. Grönwall's inequality [109] then yields:

$$V(x(t)) \leq e^{-2\varepsilon(t-t_0)} V(x(t_0)) \tag{2.38}$$

Moreover, since $\lambda_{\min}(P) I_{n_x} \preceq P \preceq \lambda_{\max}(P) I_{n_x}$ (Property 2.2.1), the following inequality can be established:

$$\|x(t)\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{-\varepsilon(t-t_0)} \|x(t_0)\|_2 \tag{2.39}$$

which demonstrates that the T-S model (2.35) is globally exponentially stable. □

Remark 2.3.1. *Adding $-2\mu P$ to the right-hand side of (2.36b), with $\mu \in \mathbb{R}_{>0}$, provides an easy way to check if the exponential decay rate of the trajectories is greater than μ or not.*

Remark 2.3.2. *The QLF (1.21) is sometimes called a CQLF, as it is a valid QLF for each local model $\dot{x}(t) = A_i x(t)$, $i = 1, \ldots, n_h$. A graphical criterion is discussed in Appendix A to check the existence of CQLF for second-order systems ($n_x = 2$).*

This result is rudimentary and introduces some conservatism in the stability analysis, in the sense that the conditions of Theorem 2.3.1 can fail to be verified despite (2.35) being globally exponentially stable. Less conservative approaches usually involve a non-Quadratic Lyapunov Function (nQLF) and lead to an LMI optimization problem of higher dimension. The most commonly used nQLF are

- the Piecewise Quadratic Lyapunov Function (PQLF), for which several methodologies exist to ensure the continuity of the Lyapunov function [140, 61, 30];
- the Multiquadratic Lyapunov Function (MQLF), also known as the polyquadratic Lyapunov function, or the fuzzy Lyapunov function, which interpolates between several QLF using the activation functions $\mathbf{h}$ of the T-S model [271, 196].

Both are sometimes combined into piecewise multiquadratic Lyapunov functions, also known as piecewise (fuzzy) weighting dependent Lyapunov functions [208, 207, 85]. These Lyapunov functions are introduced in the Chapter 4 of this manuscript. Other approaches include the use of line-integral Lyapunov functions [231], of polynomial Lyapunov functions [273] and of multi-polynomial Lyapunov functions [112], the last two approaches relying on sum-of-squares conditions rather than LMI conditions.

With inspiration from the necessary and sufficient stability conditions which are typically investigated for continuous-time LDI and continuous-time linear switched systems [193, 171], a necessary and sufficient stability condition leveraging the PQLF (2.40) introduced in [61] is discussed below in the context of the T-S model (2.35). To the author's knowledge, this equivalence result is typically not discussed in the T-S literature.

$$V(x) = \max_{1\leq i\leq k} \left\{ x^\top P_i x \right\} \tag{2.40}$$

Contrary to other PQLF [140, 30], the PQLF (2.40) has the benefit of being continuous *by definition*, which avoids the introduction of continuity constraints on the $\{P_i\}_{1\leq i\leq k}$ matrices.

Remark 2.3.3. *In (2.40), k stands for a positive integer. Although [61] considers that $k = n_h$, this condition can be relaxed without any trouble.*

Theorem 2.3.2 (Necessary and sufficient conditions)**.** *If the activation functions $\mathbf{h} \in \Delta_{n_h-1}$ are considered to be time-varying parametric uncertainties, then the T-S model (2.35) is globally exponentially stable if and only if there exist $k \in \mathbb{N}_{>0}$, $\varepsilon \in \mathbb{R}_{>0}$ and $\{P_i\}_{1\leq i\leq k}$, with $P_i \in \mathbb{S}^{++}_{n_x}(\mathbb{R})$ for all $i \in [\![1, k]\!]$, such that for all $x \in \mathbb{R}^{n_x}$*

$$\max_{A\in \text{hull}\{A_1,\dots,A_{n_h}\}} \left\{ x^\top \mathcal{H}(P_j A)x : j = \arg\max_{1\leq p\leq k} \left\{ x^\top P_p x \right\} \right\} \leq -\varepsilon \|x\|_2^2 \tag{2.41}$$

If so, the PQLF (2.40) demonstrates the global exponential stability of (2.35).

Proof. If the activation functions $\mathbf{h} \in \Delta_{n_h-1}$ are considered to be time-varying parametric uncertainties, then the T-S model (2.35) can be interpreted as the following polytopic LDI.

$$\dot{x}(t) \in \text{hull}\left\{ A_i x(t) : i = 1, \dots, n_h \right\} \tag{2.42}$$

For this model, global asymptotic stability is known to be equivalent to global exponential stability [162, 95, 193]. The proof of the theorem above is now performed by double implication.

$\boxed{\Leftarrow}$ Considering (2.40) as a Lyapunov function, (2.41) demonstrates the global asymptotic stability of (2.42). Succinctly, the left-hand side of (2.41) can be interpreted as the derivative of the PQLF (2.40) along the *worst* trajectories of (2.42). Conditions similar to those of Theorem 1.1.5 are then easily verified [193].

$\boxed{\Rightarrow}$ Conversely, assuming that (2.42) is globally exponentially stable, Theorem 2 of [193] ensures that there exist $k \in \mathbb{N}_{>0}$, $\varepsilon \in \mathbb{R}_{>0}$ and $\{l_i\}_{1\leq i\leq k} \in (\mathbb{R}^{n_x})^k$ such that $\text{rank}(l_1, \dots, l_k) = n_x$ and for all $x \in \mathbb{R}^{n_x}$

$$\max_{A\in \text{hull}\{A_1,\dots,A_{n_h}\}} \left\{ x^\top \mathcal{H}(l_j l_j^\top A)x : j = \arg\max_{1\leq p\leq k} \left\{ x^\top l_p l_p^\top x \right\} \right\} \leq -\varepsilon \|x\|_2^2 \tag{2.43}$$

Leveraging Property 2.2.4, there exists $\mu \in \mathbb{R}_{>0}$ such that for all $i \in [\![1, n_h]\!]$

$$-\varepsilon I_{n_x} + \mu \left(A_i + A_i^\top + I_{n_x} \right) \prec 0 \tag{2.44}$$

hence by convexity, for all $A \in \text{hull}\{A_1, \ldots, A_{n_h}\}$

$$-\varepsilon I_{n_x} \prec -\mu \left(A + A^\top + I_{n_x} \right) \tag{2.45}$$

so for all $x \in \mathbb{R}^{n_x}$ and $A \in \text{hull}\{A_1, \ldots, A_{n_h}\}$

$$-\varepsilon \|x\|_2^2 \leq -\mu x^\top \left(A + A^\top \right) x - \mu \|x\|_2^2 \tag{2.46}$$

hence, for all $x \in \mathbb{R}^{n_x}$

$$\max_{A \in \text{hull}\{A_1, \ldots, A_{n_h}\}} \left\{ x^\top \mathcal{H}(l_j l_j^\top A) x + \mu x^\top \left(A + A^\top \right) x : j = \arg\max_{1 \leq p \leq k} \left\{ x^\top l_p l_p^\top x \right\} \right\} \leq -\mu \|x\|_2^2 \tag{2.47}$$

which provides for all $x \in \mathbb{R}^{n_x}$

$$\max_{A \in \text{hull}\{A_1, \ldots, A_{n_h}\}} \left\{ x^\top \mathcal{H} \left(\left[l_j l_j^\top + \mu I_{n_x} \right] A \right) x : j = \arg\max_{1 \leq p \leq k} \left\{ x^\top l_p l_p^\top x \right\} \right\} \leq -\mu \|x\|_2^2 \tag{2.48}$$

Moreover, for all $x \in \mathbb{R}^{n_x}$ and $s \in \mathbb{R}$

$$\arg\max_{1 \leq p \leq k} \left\{ x^\top l_p l_p^\top x \right\} = \arg\max_{1 \leq p \leq k} \left\{ x^\top l_p l_p^\top x + s \right\} \tag{2.49}$$

so in particular, by setting $s = \mu x^\top x$, the following is verified

$$\arg\max_{1 \leq p \leq k} \left\{ x^\top l_p l_p^\top x \right\} = \arg\max_{1 \leq p \leq k} \left\{ x^\top \left[l_p l_p^\top + \mu I_{n_x} \right] x \right\} \tag{2.50}$$

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

$$\max_{A \in \text{hull}\{A_1, \ldots, A_{n_h}\}} \left\{ x^\top \mathcal{H} \left(\left[l_j l_j^\top + \mu I_{n_x} \right] A \right) x : j = \arg\max_{1 \leq p \leq k} \left\{ x^\top \left[l_p l_p^\top + \mu I_{n_x} \right] x \right\} \right\} \leq -\mu \|x\|_2^2 \tag{2.51}$$

The PQLF (2.40) taken with $P_i = l_i l_i^\top + \mu I_{n_x}$ for all $i \in [\![1, k]\!]$ (where $P_i \succ 0$ can be verified using Property 2.2.2 and $\mu > 0$) satisfies (2.41) (where μ plays the role of ε), which concludes the proof. □

Remark 2.3.4. *Intuitively, the proof above relies on perturbating the facets of the polytopic Lyapunov function discussed in [193] so it can be rewritten under the form of the PQLF (2.40) without affecting the proof of the exponential stability of (2.42).*

A PQLF (2.40) demonstrating the global exponential stability of the T-S model (2.35) can be computed using the following Bilinear Matrix Inequality (BMI) conditions.

Theorem 2.3.3 (Piecewise-quadratic stability [61])**.** *The T-S model (2.35) is globally exponentially stable if there exist $k \in \mathbb{N}_{>0}$, $\tau_{i,j,p} \in \mathbb{R}_{\geq 0}$ for all $(i, j, p) \in [\![1, n_h]\!] \times [\![1, k]\!]^2$, and $\{P_i\}_{1 \leq i \leq k}$, with $P_i \in \mathbb{S}_{n_x}(\mathbb{R})$ for all $i \in [\![1, k]\!]$, such that the BMI conditions (2.52) are satisfied.*

$$P_j \succ 0, \quad \forall j \in [\![1, k]\!] \tag{2.52a}$$

$$\mathcal{H}(P_j A_i) + \sum_{p=1}^{k} \tau_{i,j,p} (P_j - P_p) \prec 0, \quad \forall (i, j) \in [\![1, n_h]\!] \times [\![1, k]\!] \tag{2.52b}$$

Proof. The proof relies on applying kn_h times the strict homogeneous S-procedure (Lemma 2.2.2) where the strict sets are given for all $(i,j) \in [\![1,n_h]\!] \times [\![1,k]\!]$ by:

$$\mathcal{S}^{-}_{\mathcal{H}(P_j A_i)} \triangleq \left\{ x \in \mathbb{R}^{n_x} : x^\top \mathcal{H}(P_j A_i) x < 0 \right\} \tag{2.53}$$

and the non-strict sets are given for all $j,p \in [\![1,k]\!]$ by:

$$\mathcal{S}_{P_p - P_j} \triangleq \left\{ x \in \mathbb{R}^{n_x} : x^\top P_j x \geq x^\top P_p x \right\} \tag{2.54}$$

It is easily checked that the family $\left\{ \bigcap_{p=1}^{k} \mathcal{S}_{P_p - P_j} \right\}_{1 \leq j \leq k}$ covers the whole state-space $\mathbb{R}^{n_x}$. It can also be verified that there exists $\varepsilon \in \mathbb{R}_{>0}$ s.t. the condition (2.41) holds for the PQLF (2.40) if and only if (2.52a) holds and the following inclusions (2.55) are verified [61]:

$$\bigcap_{p=1}^{k} \mathcal{S}_{P_p - P_j} \subseteq \mathcal{S}^{-}_{\mathcal{H}(P_j A_i)}, \quad \forall (i,j) \in [\![1,n_h]\!] \times [\![1,k]\!] \tag{2.55}$$

The strict homogeneous S-procedure (Lemma 2.2.2) finally provides the BMI conditions (2.52b), concluding the proof. □

Theorem 2.3.3 is interesting for the following reasons:

- it does not rely on any assumption outside of $\mathbf{h} \in \Delta_{n_h - 1}$, in particular no bounds are put on the instantaneous variations of the activation functions $\mathbf{h}$;
- the conservatism of the BMI conditions (2.52) can be reduced by considering increasing values for $k \in \mathbb{N}_{>0}$.

However, due to the conservatism of the S-procedure, the BMI conditions (2.52) which allow to compute the PQLF (2.40) satisfying (2.41) are only sufficient. Moreover, the $\tau_{i,j,p} \in \mathbb{R}_{>0}$ have to be fixed a priori in order to obtain LMI conditions, since the conditions (2.52b) are otherwise BMI conditions.

Of course, the condition (2.41) of Theorem 2.3.2 is only a necessary and sufficient condition of stability if the activation functions $\mathbf{h}$ are viewed as *black box* signals for which there is no other information than $\mathbf{h} \in \Delta_{n_h - 1}$. Hence, despite the practicality that comes with making no assumption on the activation functions $\mathbf{h}$, the PQLF (2.40) can still be considered conservative for T-S systems where additional information on $\mathbf{h}$ are available. For example, some elementary stability results leveraging the MQLF (2.56) are reported below, which rely on some information on the rate of variation of the activation functions $\mathbf{h}$.

$$V(x,\theta) = x^\top \left(\sum_{i=1}^{n_h} h_i(\theta) P_i \right) x \tag{2.56}$$

Assumption 2.3.1. *The scheduling vector θ depends only on the state $x(t)$ of (2.35), and is continuously differentiable with respect to it.*

Assumption 2.3.2. *The activation functions $\mathbf{h}$ are continuously differentiable with respect to the scheduling vector θ.*

Assumption 2.3.3. *For all $i \in [\![1,n_h]\!]$ and $t \in \mathbb{R}$, there exists a Lipschitz constant $\phi_i \in \mathbb{R}_{>0}$ such that the Lie derivative of h_i is bounded along the trajectories of the system, i.e. $|dh_i(\theta(x(t)))/dt| \leq \phi_i$.*

Assumption 2.3.1 can be relaxed by assuming that:

- θ also depends on continuously differentiable exogenous signals;
- θ is continuously differentiable with respect to these exogenous signals;
- Assumption 2.3.3 also holds with respect to these signals.

The mention of these exogenous signals is omitted for the sake of simplicity. Moreover, Assumption 2.3.3 often holds locally, leading to local exponential stability results rather than global exponential stability results. The exponential stability is only ensured inside the outermost Lyapunov level contained in the set R where Assumption 2.3.3 is verified:

$$R \triangleq \left\{ x \in \mathbb{R}^{n_x} : \forall i \in [\![1, n_h]\!], \left| \sum_{j=1}^{n_h} h_j(\theta(x)) \left\langle \frac{\partial h_i}{\partial x} | A_j x \right\rangle \right| \leq \phi_i \right\} \tag{2.57}$$

In the following, R_V denotes the region contained inside the outermost Lyapunov level contained in the set R where Assumption 2.3.3 is verified. If Assumption 2.3.3 holds globally, $R_V = \mathbb{R}^{n_x}$.

Theorem 2.3.4 (Multiquadratic stability [137])**.** *Under the Assumptions 2.3.1, 2.3.2 and 2.3.3, the T-S model (2.35) is exponentially stable in R_V if there exist $\{P_i\}_{1 \leq i \leq n_h}$, with $P_i \in \mathbb{S}_{n_x}(\mathbb{R})$ for all $i \in [\![1, n_h]\!]$, such that the conditions (2.58) are satisfied.*

$$P_i \succ 0, \quad \forall i \in [\![1, n_h]\!] \tag{2.58a}$$

$$\sum_{i=1}^{n_h} \sum_{j=1}^{n_h} h_i h_j \left(\mathcal{H}(P_i A_j) + \sum_{k=0}^{n_h} \phi_k P_k \right) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h - 1} \tag{2.58b}$$

The inequality on the double convex sum in (2.58b) can be transformed into regular LMI conditions by leveraging the results of Section 2.2.2.

Proof. Assuming (2.58) holds, two proofs of this result are given. The first proof consists in checking the Lyapunov conditions for exponential stability in Theorem 1.1.5 of Chapter 1: this is not detailed. The second proof consists in verifying that the exponential stability property (Definition 1.1.10 of Chapter 1) is verified. Succinctly, it is easily verified that then for all $t \in \mathbb{R}$

$$\sum_{k=1}^{n_h} \dot{h}_k(\theta(x(t)))P_k + \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} h_i(\theta(x(t)))h_j(\theta(x(t)))\mathcal{H}(P_i A_j) \prec 0 \tag{2.59}$$

Leveraging Property 2.2.5, there exists $\varepsilon \in \mathbb{R}_{>0}$ such that

$$\sum_{k=1}^{n_h} \dot{h}_k(\theta(x(t)))P_k + \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} h_i(\theta(x(t)))h_j(\theta(x(t)))\mathcal{H}(P_i A_j) \prec -2\varepsilon \sum_{i=1}^{n_h} h_i(\theta(x(t)))P_i \tag{2.60}$$

The MQLF $V(x, \theta(x)) =$ (2.56) is a continuously differentiable function of $x \in \mathbb{R}^{n_x}$. Let $\underline{\lambda}, \overline{\lambda} \in \mathbb{R}_{>0}$ denote the minimum and maximum eigenvalues of all the $\{P_i\}_{1 \leq i \leq n_h}$. Since $\underline{\lambda} I_{n_x} \preceq P_i \preceq \overline{\lambda} I_{n_x}$ holds for all $i \in [\![1, n_h]\!]$ (Property 2.2.1), the following bounds are verified for all $x \in \mathbb{R}^{n_x}$:

$$\underline{\lambda} \|x\|_2^2 \leq V(x, \theta(x)) \leq \overline{\lambda} \|x\|_2^2 \tag{2.61}$$

Moreover, (2.60) provides $\dot{V}(x(t),\theta(x(t))) \leq -2\varepsilon V(x(t),\theta(x(t)))$. Grönwall's inequality [109] then yields

$$V(x(t),\theta(x(t))) \leq e^{-2\varepsilon(t-t_0)} V(x(t_0),\theta(x(t_0))) \tag{2.62}$$

hence

$$\|x(t)\|_2 \leq \sqrt{\frac{\overline{\lambda}}{\underline{\lambda}}} e^{-\varepsilon(t-t_0)} \|x(t_0)\|_2 \tag{2.63}$$

which finally demonstrates that the T-S model (2.35) is exponentially stable. □

Remark 2.3.5. *Adding $2\mu P_i$ to the sum over i in the left-hand side of (2.58b) provides an easy way to check if the exponential decay rate of the trajectories is greater than μ or not.*

The use of the Lipschitz constants $\{\phi_i\}_{1\leq i\leq n_h}$ in (2.58b) is often viewed as a drawback to the MQLF approach in the T-S literature: the LMI conditions (2.58) are obtainable *at the cost* of some Lipschitz assumptions. However, without these Lipschitz constants, no information on the activation function $\mathbf{h}$ would be included in the LMI conditions at all. In this regard, compared to the PQLF (2.40), the MQLF (2.56) is interesting *thanks to* the Lipschitz assumptions that can be made on $\mathbf{h}$, which carry properties from the activation functions $\mathbf{h}$ to the LMI conditions of stability.

Many relaxations of the LMI conditions (2.58) are found in the literature, e.g. [238, 197, 155, 88, 64]. Two versions of these LMI conditions are provided below with a reduced conservatism.

– Theorem 2.3.5 leverages the fact that

$$\sum_{i=1}^{n_h} \frac{dh_i(\theta(x(t)))}{dt} = 0 \tag{2.64}$$

in order to relax the bounds of Assumption 2.3.3 on one of the activation functions, e.g. h_p with $p \in [\![1, n_h]\!]$.

– Theorem 2.3.6 relies on the introduction of slack variables, which trades the number of LMI conditions with the number of decision variables in (2.58).

Assumption 2.3.4. *For all $i \in [\![1, n_h]\!] \setminus \{p\}$ and $t \in \mathbb{R}$, there exists a Lipschitz constant $\phi_i \in \mathbb{R}_{>0}$ such that $|dh_i(\theta(x(t)))/dt| \leq \phi_i$.*

Theorem 2.3.5 (Relaxed multiquadratic stability [271])**.** *Under the Assumptions 2.3.1, 2.3.2 and 2.3.4, the T-S model (2.35) is exponentially stable in R_V if there exist $\{P_i\}_{1\leq i\leq n_h}$, with $P_i \in \mathbb{S}_{n_x}(\mathbb{R})$ for all $i \in [\![1, n_h]\!]$, such that the conditions (2.65) are satisfied.*

$$P_i \succ 0, \quad \forall i \in [\![1, n_h]\!] \tag{2.65a}$$

$$P_i - P_p \succeq 0, \quad \forall i \in [\![1, n_h]\!] \setminus \{p\} \tag{2.65b}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \left(\mathcal{H}(P_i A_j) + \sum_{k=1}^{n_h} \phi_k (P_k - P_p) \right) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.65c}$$

The inequality on the double convex sum in (2.65c) can be transformed into regular LMI conditions by leveraging the results of Section 2.2.2.

Theorem 2.3.6 (Relaxed multiquadratic stability [197]). *Under the Assumptions 2.3.1, 2.3.2 and 2.3.3, the T-S model (2.35) is exponentially stable in R_V if there exist $\{P_i\}_{1\leq i\leq n_h}$ and $\{M_i\}_{1\leq i\leq 3}$, with $P_i \in \mathbb{S}_{n_x}(\mathbb{R})$ for all $i \in [\![1, n_h]\!]$, $M_1, M_2 \in \mathbb{R}^{n_x\times n_x}$ and $M_3 \in \mathbb{S}_{n_x}(\mathbb{R})$, such that the LMI conditions (2.66) are satisfied.*

$$P_i \succ 0, \quad \forall i \in [\![1, n_h]\!] \tag{2.66a}$$

$$P_i + M_3 \succeq 0, \quad \forall i \in [\![1, n_h]\!] \tag{2.66b}$$

$$\begin{pmatrix} \sum_{k=1}^{n_h} \phi_k(P_k + M_3) - \mathcal{H}(M_1 A_i) & P_i + M_1 - A_i^\top M_2^\top \\ P_i + M_1^\top - M_2 A_i & \mathcal{H}(M_2) \end{pmatrix} \prec 0, \quad \forall i \in [\![1, n_h]\!] \tag{2.66c}$$

Besides simply rewriting the LMI conditions (2.58) discussed previously, two key ideas exist in order to further reduce the conservatism of a stability analysis relying on a MQLF.

– The first idea consists in introducing the following generalized MQLF [66]:

$$V(x, \theta) = x^\top \left(\sum_{i_1=1}^{n_h} \cdots \sum_{i_k=1}^{n_h} h_{i_1}(\theta) \dots h_{i_k}(\theta) P_{i_1,\dots,i_k} \right) x \tag{2.67}$$

Intuitively, increasing the value of k in the generalized MQLF above provides more degree of freedom to the Lyapunov function. This is similar to increasing the value of k in the PQLF (2.40).

– The second idea consists in decomposing further the activation functions $\mathbf{h}$ into a product of intermediary functions $\mathbf{w} = (w_{i,0}, w_{i,1})_{1\leq i\leq n_\theta}$, such that:

$$h_i = w_{1,i_{[1]}} \dots w_{n_\theta, i_{[n_\theta]}}, \quad \forall i = \overline{i_{[n_\theta]} \dots i_{[1]}}^{(2)} \in [\![0, 2^{n_\theta} - 1]\!] \tag{2.68}$$

These intermediary functions are typically obtained at the modeling step of the T-S model by using the nonlinear sector approach [113], although this approach can also be generalized to generic tensor-product models [32]. Under this decomposition, the Lipschitz assumptions are put on the intermediary functions $\mathbf{w}$ rather than on the activation functions $\mathbf{h}$. Intuitively, this allows to take into account more precise information on the activation functions $\mathbf{h}$ in the LMI conditions of stability.

Unsurprisingly, these two approaches have already been combined together [29].

2.3.2 Discrete-time stability

Similarly to continuous-time stability, the stability of the discrete-time and input-free T-S model (2.69) is studied by using LMI conditions. In this section, only QLF and MQLF are considered.

$$x(t+1) = \sum_{i=1}^{n_h} h_i(\theta) A_i x(t) \tag{2.69}$$

Given the QLF (1.21), i.e. $V(x) = x^\top P x$, the LMI condition corresponding to Theorem 1.1.6 of Chapter 1 is stated as follows.

Theorem 2.3.7 (Quadratic stability [291]). *The T-S model (2.69) is globally exponentially stable if there exists $P \in \mathbb{S}_{n_x}(\mathbb{R})$ such that the LMI conditions (2.70) are satisfied.*

$$P \succ 0 \tag{2.70a}$$

$$A_j^\top P A_j - P \prec 0, \quad \forall j \in J \tag{2.70b}$$

where J is a subset of $[\![1, n_h]\!]$ such that for all $i \in [\![1, n_h]\!]$, $A_i \in \mathrm{hull}\{A_j : j \in J\}$.

Proof. Similarly to Theorem 2.3.1, assuming (2.70) holds, three proofs of this result are given. The first proof, and the most straightforward, consists in noticing that the trajectories of (2.69) are included in the trajectories of the polytopic LDI $x(t+1) \in \mathrm{hull}\{A_j x(t) : j \in J\}$. Applying Theorem 1.1.6 of Chapter 1 concludes the first proof. The second proof consists in checking the Lyapunov conditions for exponential stability in Theorem 1.1.5 of Chapter 1: this is not detailed. The third proof consists in verifying that the exponential stability property (Definition 1.1.10 of Chapter 1) is verified. Briefly, Schur's complement (Lemma 2.2.1) allows to rewrite the LMI conditions (2.70) as:

$$\begin{pmatrix} P & A_j^\top P \\ P A_j & P \end{pmatrix} \succ 0, \quad \forall j \in J \tag{2.71}$$

By convexity, the LMI condition above stands for all state matrix $A \in \mathrm{hull}\{A_j : j \in J\}$, which, after applying Schur's complement again, provides for all $\theta \in \Theta$:

$$\left(\sum_{i=1}^{n_h} h_i(\theta) A_i\right)^\top P \left(\sum_{i=1}^{n_h} h_i(\theta) A_i\right) - P \prec 0 \tag{2.72}$$

The QLF (1.21) is now introduced, yielding $V(x(t+1)) - V(x(t)) < 0$. Property 2.2.5 then guarantees that there exists $\varepsilon \in (0, 1)$ such that:

$$V(x(t+1)) - V(x(t)) < -\varepsilon V(x(t)) \tag{2.73}$$

hence $V(x(t+1)) < (1-\varepsilon) V(x(t))$, and so, for all $t \in \mathbb{Z}_{\geq t_0}$, $V(x(t)) \leq (1-\varepsilon)^{t-t_0} V(x(t_0))$. Finally, since $\lambda_{\min}(P) I_{n_x} \preceq P \preceq \lambda_{\max}(P) I_{n_x}$ (Property 2.2.1), the following inequality holds:

$$\|x(t)\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{\frac{1}{2}\ln(1-\varepsilon)(t-t_0)} \|x(t_0)\|_2 \tag{2.74}$$

which demonstrates that the T-S model (2.69) is globally exponentially stable. □

Remark 2.3.6. *Adding $-\mu P$ to the right-hand side of (2.36b), with $\mu \in (0, 1)$, provides an easy way to check if the exponential decay rate of the trajectories is greater than $-\frac{1}{2}\ln(1-\mu)$ or not.*

Remark 2.3.7. *The QLF (1.21) is sometimes called a CQLF, as it is a valid QLF for each local model $x(t+1) = A_i x(t)$, $i = 1, \ldots, n_h$. A graphical criterion is discussed in Appendix A to check the existence of CQLF for second-order systems ($n_x = 2$).*

Remark 2.3.8. *Schur's complement (Lemma 2.2.1) allows to rewrite the LMI conditions (2.70) as follows:*

$$\begin{pmatrix} P & A_j^\top P \\ PA_j & P \end{pmatrix} \succ 0, \quad \forall j \in J \tag{2.75}$$

This rewriting is practical since the positivity $P \succ 0$ is directly guaranteed, and the condition is linear in the A_j rather than quadratic. Note that the following rewriting is also valid and equivalent to the previous one:

$$\begin{pmatrix} -P & A_j^\top P \\ PA_j & -P \end{pmatrix} \prec 0, \quad \forall j \in J \tag{2.76}$$

Of course, the condition above is conservative. As for the continuous-time case, this motivates the introduction of the MQLF (2.56), i.e. $V(x,\theta) = x^\top \left(\sum_{i=1}^{n_h} h_i(\theta) P_i\right) x$, in order to reduce the conservatism of the stability analysis. The following LMI stability conditions hold for this MQLF.

Theorem 2.3.8 (Multiquadratic stability [158])**.** *The T-S model (2.69) is globally exponentially stable if there exist $k \in \mathbb{N}_{>0}$ and $\{P_i\}_{1\leq i\leq n_h}$, with $P_i \in \mathbb{S}_{n_x}(\mathbb{R})$ for all $i \in [\![1, n_h]\!]$, such that the LMI conditions (2.77) are satisfied.*

$$P_i \succ 0, \quad \forall i \in [\![1, n_h]\!] \tag{2.77a}$$

$$A_{j_0}^\top \dots A_{j_{k-1}}^\top P_{j_k} A_{j_{k-1}} \dots A_{j_0} - P_{j_0} \prec 0, \quad \forall (j_0, j_k) \in [\![1, n_h]\!]^2, (j_1, \dots, j_{k-1}) \in J \tag{2.77b}$$

J is a subset of $[\![1, n_h]\!]^{k-1}$ such that for all $(i_1, \dots, i_{k-1}) \in [\![1, n_h]\!]^{k-1}$, $\prod_{p=1}^{k-1} A_{i_p} \in \operatorname{hull}\{\prod_{p=1}^{k-1} A_{j_p} : (j_1, \dots, j_{k-1}) \in J\}$.

Proof. This result is well-established when $J = [\![1, n_h]\!]^{k-1}$ [158]. The assumption that $J = [\![1, n_h]\!]^{k-1}$ is now relaxed. By definition of the convex hull, for all $(i_1, \dots, i_{k-1}) \in [\![1, n_h]\!]^{k-1}$ there exist

$$\left(\alpha_{(j_1,\dots,j_{k-1})}^{(i_1,\dots,i_{k-1})}\right)_{(j_1,\dots,j_{k-1})\in J} \in \Delta_{\#J-1} \tag{2.78}$$

such that

$$\prod_{p=1}^{k-1} A_{i_p} = \sum_{(j_1,\dots,j_{k-1})\in J} \alpha_{(j_1,\dots,j_{k-1})}^{(i_1,\dots,i_{k-1})} \prod_{p=1}^{k-1} A_{j_p} \tag{2.79}$$

Assuming that (2.77) is verified, Schur's complement (Lemma 2.2.1) ensures for all $(j_0, j_k) \in [\![1, n_h]\!]^2$ and $(j_1, \dots, j_{k-1}) \in J$ that

$$\begin{pmatrix} P_{j_0} & A_{j_0}^\top \dots A_{j_{k-1}}^\top P_{j_k} \\ P_{j_k} A_{j_{k-1}} \dots A_{j_0} & P_{j_k} \end{pmatrix} \succ 0 \tag{2.80}$$

by convexity, the following stands for all $(j_0, j_k) \in [\![1, n_h]\!]^2$ and $(i_1, \dots, i_{k-1}) \in [\![1, n_h]\!]^{k-1}$

$$\sum_{(j_1,\dots,j_{k-1})\in J} \alpha_{(j_1,\dots,j_{k-1})}^{(i_1,\dots,i_{k-1})} \begin{pmatrix} P_{j_0} & A_{j_0}^\top \left(\prod_{p=1}^{k-1} A_{j_p}\right)^\top P_{j_k} \\ P_{j_k} \left(\prod_{p=1}^{k-1} A_{j_p}\right) A_{j_0} & P_{j_k} \end{pmatrix} \succ 0 \tag{2.81}$$

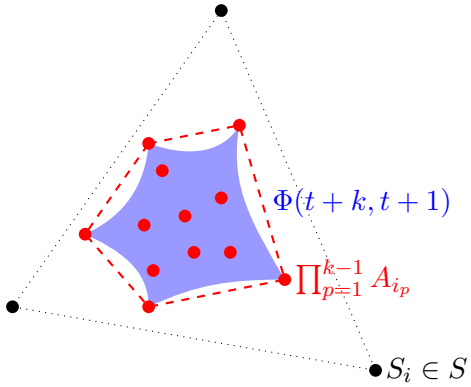


Figure 2.2: Schematic illustration of the reduction of the number of LMI in Theorem 2.3.8.

by linearity, this can be rewritten as

$$\begin{pmatrix} P_{j_0} & A_{j_0}^\top \left(\prod_{p=1}^{k-1} A_{i_p}\right)^\top P_{j_k} \\ P_{j_k}\left(\prod_{p=1}^{k-1} A_{i_p}\right) A_{j_0} & P_{j_k} \end{pmatrix} \succ 0 \tag{2.82}$$

which, after applying a Schur's complement (Lemma 2.2.1), finally provides the usual LMI conditions of the literature [158]. □

Remark 2.3.9. *As discussed in the proof above, Schur's complement (Lemma 2.2.1) allows to rewrite the LMI conditions (2.77) as follows:*

$$\begin{pmatrix} P_{j_0} & A_{j_0}^\top \dots A_{j_{k-1}}^\top P_{j_k} \\ P_{j_k} A_{j_{k-1}} \dots A_{j_0} & P_{j_k} \end{pmatrix} \succ 0, \quad \forall (j_0, j_k) \in [\![1, n_h]\!]^2, (j_1, \dots, j_{k-1}) \in J \tag{2.83}$$

This rewriting is practical since the positivity of $P_i \succ 0$ is directly guaranteed for all $i \in [\![1, n_h]\!]$, and the condition is linear in $A_{j_{k-1}} \dots A_{j_0}$ rather than quadratic. Note that the following rewriting is also valid and equivalent to the previous one:

$$\begin{pmatrix} -P_{j_0} & A_{j_0}^\top \dots A_{j_{k-1}}^\top P_{j_k} \\ P_{j_k} A_{j_{k-1}} \dots A_{j_0} & -P_{j_k} \end{pmatrix} \prec 0, \quad \forall (j_0, j_k) \in [\![1, n_h]\!]^2, (j_1, \dots, j_{k-1}) \in J \tag{2.84}$$

Theorem 2.3.8 is particularly interesting for the following reasons:

- it does not rely on any assumption outside of $\mathbf{h} \in \Delta_{n_h - 1}$, in particular no bounds are put on the instantaneous variations of the activation functions $\mathbf{h}$;
- the conservatism of the LMI conditions (2.77) can be reduced by considering increasing values for $k \in \mathbb{N}_{>0}$, and there is a sense in which for a large enough $k \in \mathbb{N}_{>0}$ they become necessary and sufficient conditions of stability (see Theorem 2.3.10);

Following the idea of the proof hereabove, the products $\prod_{p=1}^{k-1} A_{j_p}$ with $(j_1, \dots, j_{k-1}) \in J$ in (2.77) can be replaced by any set of matrices S such that for all $(i_1, \dots, i_{k-1}) \in [\![1, n_h]\!]^{k-1}$, $\prod_{p=1}^{k-1} A_{i_p} \in \mathrm{hull}(S)$. This can be leveraged to reduce the number of LMI conditions in (2.77). Intuitively, the set $\mathrm{hull}(S)$ still bounds all the possible state-transition matrices Φ of the T-S model (2.69) between $(t+1)$ and $(t+k)$, with $t \in \mathbb{Z}$. Of course, this reduction in the number of LMI is obtained at the cost of some conservatism. This idea is schematically illustrated in

Figure 2.2.

As for the continuous-time case, PQLF are also considered by the literature to demonstrate the exponential stability of discrete-time systems. This approach will not be discussed in this manuscript, and the interested reader is referred to [61, 92]. Leaving the Lyapunov framework, a necessary and sufficient exponential stability condition, which is typically investigated for discrete-time LDI and discrete-time linear switched systems [26, 171], is discussed below in the context of the T-S model (2.69).

Theorem 2.3.9 (Necessary and sufficient conditions [26]). *If the activation functions* $\mathbf{h} \in \Delta_{n_h-1}$ *are considered to be time-varying parametric uncertainties, then the T-S model* (2.69) *is globally exponentially stable if and only if there exists* $k \in \mathbb{N}_{>0}$ *such that*

$$\max_{(i_1,\dots,i_k)\in[\![1,n_h]\!]^k} \|A_{i_k}\dots A_{i_1}\|_\infty < 1 \tag{2.85}$$

Proof. Necessity is shown by contradiction in [26]. The proof of sufficiency is given below since it is not discussed in the literature. It is assumed that there exists $k \in \mathbb{N}_{>0}$ such that (2.85) holds. Let $r \in (0,1)$ be such that

$$\max_{(i_1,\dots,i_k)\in[\![1,n_h]\!]^k} \|A_{i_k}\dots A_{i_1}\|_\infty < r < 1 \tag{2.86}$$

moreover, let $s_q \in \mathbb{R}_{\geq 0}$ and $s \in \mathbb{R}_{\geq 1}$ be defined for all $q \in [\![0, k-1]\!]$ by:

$$s_q \triangleq \begin{cases} 1 & \text{if } q = 0 \\ \max_{(i_1,\dots,i_q)\in[\![1,n_h]\!]^q} \left\|A_{i_q}\dots A_{i_1}\right\|_\infty & \text{else} \end{cases} \tag{2.87a}$$

$$s \triangleq \max_{q\in[\![0,k-1]\!]} s_q \tag{2.87b}$$

For all $x(t_0) \in \mathbb{R}^{n_x}$, $p \in \mathbb{N}_{>0}$ and $q \in [\![0, k-1]\!]$, the following holds

$$x(t_0+pk+q) = \sum_{i_{pk+q}=1}^{n_h} \cdots \sum_{i_1=1}^{n_h} h_{i_{pk+q}}(\theta(t_0+pk+q-1))\dots h_{i_1}(\theta(t_0))A_{i_{pk+q}}\dots A_{i_1}x(t_0) \tag{2.88}$$

hence, by the triangle inequality, and by submultiplicativity of the infinity norm (it is assumed without loss of generality that $x(t_0) \neq 0$)

$$\begin{aligned} \frac{\|x(t_0+pk+q)\|_\infty}{\|x(t_0)\|_\infty} &\leq \sum_{i_{pk+q}=1}^{n_h} \cdots \sum_{i_1=1}^{n_h} h_{i_{pk+q}}(\theta(t_0+pk+q-1))\dots h_{i_1}(\theta(t_0))\|A_{i_{pk+q}}\dots A_{i_1}\|_\infty \\ &\leq \sum_{i_{pk+q}=1}^{n_h} \cdots \sum_{i_1=1}^{n_h} h_{i_{pk+q}}(\theta(t_0+pk+q-1))\dots h_{i_1}(\theta(t_0))sr^p \\ \frac{\|x(t_0+pk+q)\|_\infty}{\|x(t_0)\|_\infty} &\leq sr^p = s(r^{\frac{1}{k}})^{pk} \leq s\underbrace{(r^{\frac{1}{k}})^{q-k}}_{\geq 1}(r^{\frac{1}{k}})^{pk} = \frac{s}{r}(r^{\frac{1}{k}})^{pk+q} = \frac{s}{r}e^{\frac{1}{k}\ln(r)(pk+q)} \end{aligned} \tag{2.89}$$

and finally for all $t \in \mathbb{Z}_{\geq t_0}$, by the norm inequality $\frac{1}{\sqrt{n_x}}\|\cdot\|_2 \leq \|\cdot\|_\infty \leq \|\cdot\|_2$, the following inequality holds

$$\|x(t)\|_2 \leq \sqrt{n_x}\frac{s}{r}e^{\frac{1}{k}\ln(r)(t-t_0)}\|x(t_0)\|_2 \tag{2.90}$$

The exponential stability property defined in Definition 1.1.10 of Chapter 1 is verified, hence (2.69) is globally exponentially stable. □

This result is leveraged below to show that there is a sense in which the LMI conditions (2.77) of Theorem 2.3.8 are necessary and sufficient conditions of stability for a large enough $k \in \mathbb{N}_{>0}$.

Theorem 2.3.10 (Necessary and sufficient LMI conditions)**.** *If the activation functions* $\mathbf{h} \in \Delta_{n_h-1}$ *are considered to be time-varying parametric uncertainties, then the T-S model* (2.69) *is globally exponentially stable if and only if there exists* $k \in \mathbb{N}_{>0}$ *such that the LMI conditions* (2.77) *of Theorem 2.3.8 are satisfied.*

Proof. If the activation functions $\mathbf{h} \in \Delta_{n_h-1}$ are considered to be time-varying parametric uncertainties, then the T-S model (2.69) can be interpreted as the following LDI.

$$x(t+1) \in \operatorname{hull}\{A_i x(t) : i = 1, \ldots, n_h\} \tag{2.91}$$

The proof of the theorem above is now performed by double implication.

$\boxed{\Leftarrow}$ Theorem 2.3.8 demonstrates that if there exists $k \in \mathbb{N}_{>0}$ such that the LMI conditions (2.77) are satisfied, then (2.91) is globally exponentially stable.

$\boxed{\Rightarrow}$ Reciprocally, it is assumed that (2.91) is globally exponentially stable. From Theorem 2.3.9, there exists $k \in \mathbb{N}_{>0}$ such that the following holds

$$\max_{(j_1,\ldots,j_k)\in[\![1,n_h]\!]^k} \|A_{j_k}\ldots A_{j_1}\|_\infty < 1 \tag{2.92}$$

so in particular there exists $p \in \mathbb{N}_{>0}$ such that

$$\left(\max_{(j_1,\ldots,j_k)\in[\![1,n_h]\!]^k} \|A_{j_k}\ldots A_{j_1}\|_\infty\right)^p < \frac{1}{\sqrt{n_x}} \tag{2.93}$$

by submultiplicativity of the infinity norm, for all $(i_1, \ldots, i_{pk}) \in [\![1, n_h]\!]^{pk}$

$$\begin{aligned}\left\|A_{i_{pk}}\ldots A_{i_1}\right\|_\infty &\leq \left\|A_{i_{pk}}\ldots A_{i_{(p-1)k+1}}\right\|_\infty \ldots \|A_{i_k}\ldots A_{i_1}\|_\infty \\ &\leq \left(\max_{(j_1,\ldots,j_k)\in[\![1,n_h]\!]^k} \|A_{j_k}\ldots A_{j_1}\|_\infty\right)^p < \frac{1}{\sqrt{n_x}}\end{aligned} \tag{2.94}$$

and by the norm inequality $\|\cdot\|_2 \leq \sqrt{n_x}\|\cdot\|_\infty$, this provides for all $(i_1, \ldots, i_{pk}) \in [\![1, n_h]\!]^{pk}$

$$\left\|A_{i_{pk}}\ldots A_{i_1}\right\|_2 < 1 \tag{2.95}$$

now, by definition of the spectral norm, the following holds for all $(i_1, \dots, i_{pk}) \in [\![1, n_h]\!]^{pk}$

$$\sqrt{\lambda_{\max}(A_{i_1}^\top \dots A_{i_{pk}}^\top A_{i_{pk}} \dots A_{i_1})} < 1 \tag{2.96}$$

hence $\lambda_{\max}(A_{i_1}^\top \dots A_{i_{pk}}^\top A_{i_{pk}} \dots A_{i_1}) < 1$, which can finally be rewritten as

$$A_{i_1}^\top \dots A_{i_{pk}}^\top I_{n_x} A_{i_{pk}} \dots A_{i_1} \prec I_{n_x} \tag{2.97}$$

and which corresponds to the conditions (2.77) of Theorem 2.3.8, considered with $J = [\![1, n_h]\!]^{pk-1}$ and $P_i = I_{n_x}$ for all $i \in [\![1, n_h]\!]$. □

Of course, the LMI conditions (2.77) of Theorem 2.3.8 are only a necessary and sufficient conditions of stability if the activation functions $\mathbf{h}$ are viewed as *black box* signals for which there is no other information than $\mathbf{h} \in \Delta_{n_h-1}$. Hence, despite the practicality that comes with making no assumption on the activation functions $\mathbf{h}$, the MQLF (2.56) can still be considered conservative for discrete-time T-S systems where additional information on $\mathbf{h}$ are available. This idea is not developed further in this manuscript. The interested reader is referred to [164, 166] which provide LMI conditions for the local stability of (2.69) while taking into consideration the rate of variation of the activation function $\mathbf{h}$.

2.4 Stabilization

As briefly discussed in Chapter 1 of this manuscript, the stabilization problem, i.e. the design of a controller ensuring the stability of the T-S model (2.1), is arguably the most fundamental problem of the T-S framework: this section will again be forced to skip most results of the literature. More complete reviews on the subject of T-S and LPV model stabilization can be found in [272, 235, 183, 32]. In particular, this section makes three significant assumptions.

Assumption 2.4.1. *The stabilization of the full state $x(t)$ is investigated, rather than solely the stabilization of the regulated output $z(t)$.*

This assumption is a general trend of the T-S literature. The regulated output stabilization (i.e. output regulation) problem should not be confused with the output *feedback* (or output-*based*) stabilization problem. The terminological confusion is sadly not uncommon in the literature (e.g. [304]). The reader is referred to [189, 31, 124] for works dedicated to output regulation.

Assumption 2.4.2. *It is assumed that the scheduling vector $\theta \in \Theta$ is perfectly measured in real-time.*

Such T-S systems are said to have *measurable premise variables*, and they are much more easily handled than systems with *unmeasurable measurable premise variables* [133]. This assumption however limits the representation capabilities of T-S models, as the signals included in the scheduling vector have to be carefully selected in order to be measurable [195, 132].

Assumption 2.4.3. *The exogenous input $w(t)$ and regulated output $z(t)$ are not considered.*

The simplifying Assumptions 2.4.1, 2.4.2 and 2.4.3 are made until the end of this chapter. The last assumption is briefly relaxed in Section 2.4.4 in order to mention the generalized $\mathcal{H}_2$ and $\mathcal{H}_\infty$ attenuation criteria, which are often considered by the T-S literature.

2.4.1 Observer-based controller

Under the simplifying Assumptions 2.4.1, 2.4.2 and 2.4.3, the T-S system (2.1) is rewritten as follows:

$$\delta x(t) = A(\theta)x(t) + B(\theta)u(t) \tag{2.98a}$$
$$y(t) = C(\theta)x(t) + D(\theta)u(t) \tag{2.98b}$$

The literature usually investigates LPV controller of the form (2.99) by a direct analogy with the linear case discussed in Chapter 1.

$$\delta\hat{x}(t) = A(\theta)\hat{x}(t) + B(\theta)u(t) + L(\theta)(C(\theta)\hat{x}(t) + D(\theta)u(t) - y(t)) \tag{2.99a}$$
$$u(t) = K(\theta)\hat{x}(t) \tag{2.99b}$$

The internal state of this controller $\hat{x}(t)$ is in fact an estimate of the internal state of the system $x(t)$, obtained using a Luenberger observer of *scheduled* gain $L(\theta)$. The feedback law then consists of a static proportional state feedback law of *scheduled* gain $K(\theta)$, based on the estimated state $\hat{x}(t)$. The full closed-loop system, composed of the state $x(t)$ and of the estimation error $\hat{x}(t) - x(t)$, which both need to be stabilized at 0, is given thereafter:

$$\delta\begin{pmatrix} x(t) \\ \hat{x}(t) - x(t) \end{pmatrix} = \begin{pmatrix} A(\theta) + B(\theta)K(\theta) & B(\theta)K(\theta) \\ 0 & A(\theta) + L(\theta)C(\theta) \end{pmatrix}\begin{pmatrix} x(t) \\ \hat{x}(t) - x(t) \end{pmatrix} \tag{2.100}$$

Ensuring that the full closed-loop system is exponentially stable no longer works by simply imposing the diagonal matrices of (2.100) to be Hurwitz (resp. Schur) because of the time-varying scheduling vector θ. Therefore, contrary to the linear case, the block-triangular structure of (2.100) is only an *indication* that the scheduled gains $K(\theta)$ and $L(\theta)$ can be designed independently of each other, but it is *not sufficient* to establish a separation principle. Nonetheless, the separation principle holds for LPV and T-S systems with measurable premise variables, but the proof of this principle is more involved than just exhibiting the block-triangular matrix hereabove. Succinctly, it relies on designing independently the scheduled Luenberger observer and the scheduled state feedback controller, and then on demonstrating that the full observer-based controller only scales down the decay rate of the state compared to a case where the state is fully known, without affecting the overall exponential stability property [304]. In the following, the observer design and controller design problems are tackled independently of each other by relying on this separation principle.

Remark 2.4.1. *The separation principle does not hold if θ is unmeasurable, i.e. if the T-S system has unmeasurable premise variables. Indeed, if θ is not measured, it is at best obtained through an estimated scheduling vector $\hat{\theta}$, leading to a Luenberger observer replicating the dynamic of the system with an error on θ. In particular, this approximation prevents the full closed-loop system from having the block-triangular structure* (2.100) *[195].*

Other dynamical controllers which do not rely on the suggested architecture (2.99) are also sometimes investigated by the literature (e.g. [212, 111]). In particular, some controller designs skip the construction of an observer and directly obtain state stabilization using an output feedback mechanism, generally of the form $u(t) = K(\theta)y(t)$. These kinds of output-based control laws are also not investigated in this manuscript, and the reader is referred to the following works for more details [148, 60].

2.4.2 Gain-scheduling schemes

A naive approach to obtain the gains $K(\theta)$ and $L(\theta)$ in the LPV controller (2.99) consists in determining these gains on a set of operating points $\{\theta(i)\}_{1\leq i\leq k}$ using Theorem 1.1.8 of Chapter 1. The values of $K(\theta)$ and $L(\theta)$ can then be evaluated in real-time, by interpolating between the previously computed gains $\{K_i\}_{1\leq i\leq k}$ and $\{L_i\}_{1\leq i\leq k}$ depending on the value of the measured scheduling vector θ. This naive approach is however only guaranteed to stabilize (2.100) under a large enough quantity of operating points, which can be computationally heavy to obtain, and under a slow-varying assumption on θ, which is difficult to verify in practice, especially in the quasi-LPV case [76].

A less naive approach consists in utilizing the activation functions $\mathbf{h}$ of the T-S model in order to interpolate between several controller and observer gains $\{K_i\}_{1\leq i\leq k}$ and $\{L_i\}_{1\leq i\leq k}$, while rewriting properly the Lyapunov stability conditions of Theorem 1.1.4 or 1.1.5 [14]. This is the main approach of the T-S literature, as it can lead to rigorous LMI conditions to compute the gains $\{K_i\}_{1\leq i\leq k}$ and $\{L_i\}_{1\leq i\leq k}$. The stabilization of the closed-loop system (2.100) is moreover properly guaranteed, generally without a slow-varying assumption on θ. The T-S framework considers two main ways of interpolating between the gains:

– The Parallel Distributed Compensation (PDC) schemes, in which the gains are interpolated linearly with respect to the activation functions $\mathbf{h}$. The most usual PDC scheme is reported below.

$$K(\theta)=\sum_{i=1}^{n_h} h_i(\theta)K_i, \qquad L(\theta)=\sum_{i=1}^{n_h} h_i(\theta)L_i \tag{2.101}$$

This PDC scheme is generally considered together with the following QLF:

$$V(x,\hat{x})=x^\top P_1 x+(\hat{x}-x)^\top P_2(\hat{x}-x) \tag{2.102}$$

– The non-Parallel Distributed Compensation (nPDC) schemes, in which the interpolation is not linear in the activation functions $\mathbf{h}$. The two most usual nPDC schemes are reported below.

$$K(\theta)=\left(\sum_{i=1}^{n_h} h_i(\theta)K_i\right)\left(\sum_{i=1}^{n_h} h_i(\theta)P_{1,i}\right)^{-1}, \quad L(\theta)=\left(\sum_{i=1}^{n_h} h_i(\theta)P_{2,i}\right)^{-1}\left(\sum_{i=1}^{n_h} h_i(\theta)L_i\right) \tag{2.103}$$

These nPDC schemes are *specifically* considered together with the following nQLF:

$$V(x,\theta)=x^\top\left(\sum_{i=1}^{n_h} h_i(\theta)P_{1,i}\right)^{-1}x, \qquad \text{[for the controller]} \tag{2.104a}$$

$$V(x,\hat{x},\theta)=(\hat{x}-x)^\top\left(\sum_{i=1}^{n_h} h_i(\theta)P_{2,i}\right)(\hat{x}-x), \qquad \text{[for the observer]} \tag{2.104b}$$

This manuscript only investigates these PDC and nPDC schemes, but it should be noted that several variations of these interpolating techniques exist. One can mention the proportional-PDC scheme [170], as well as other augmented schemes incorporating the derivative of the scheduling vector θ (e.g. [270, 114]).

2.4.3 LMI conditions

First, the PDC scheme (2.101) is investigated using the QLF (2.102) in the continuous-time case.

Theorem 2.4.1 (Continuous quadratic PDC stabilization [32]). *The continuous-time closed-loop T-S model (2.100) is globally exponentially stabilizable using the PDC scheme (2.101) if there exist $X_1, X_2 \in \mathbb{S}_{n_x}(\mathbb{R})$ and $\{M_i\}_{1\leq i\leq n_h}$, $\{N_i\}_{1\leq i\leq n_h}$, with $M_i \in \mathbb{R}^{n_u\times n_x}$ and $N_i \in \mathbb{R}^{n_x\times n_y}$ for all $i \in [\![1, n_h]\!]$, such that the conditions (2.105) are satisfied.*

$$X_i \succ 0, \quad i = 1, 2 \tag{2.105a}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \mathcal{H}(A_i X_1 + B_i M_j) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.105b}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \mathcal{H}(X_2 A_i + N_i C_j) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.105c}$$

The inequalities on the double convex sums can be transformed into regular LMI conditions by leveraging the results of Section 2.2.2. The gain matrices K_i and L_i are retrieved with $K_i = M_i X_1^{-1}$ and $L_i = X_2^{-1} N_i$. Moreover, there exists $\lambda \in \mathbb{R}_{>0}$ such that the matrices P_1 and P_2 providing the QLF (2.102) can be taken as $P_1 = \lambda X_1^{-1}$ and $P_2 = X_2$.

Proof. See Theorem 6.11 of [32]. □

Similarly to the LMI conditions of Theorem 2.3.1, it is usual to add $-2\mu_1 X_1$ and $-2\mu_2 X_2$ respectively to the right-hand side of (2.105b) and of (2.105c). This ensures a minimum exponential decay rate of μ_2 to the observation error $\hat{x}(t) - x(t)$ and, assuming a perfectly estimated state, a minimum exponential decay rate of μ_1 to the state $x(t)$. In practice, the factor λ introduced in the theorem hereabove has to be taken into account in the choice of μ_1, as the estimation error of the observer affects the decay rate imposed by the controller. As a rule of thumb, μ_1 and μ_2 are usually selected so $\mu_2 \approx 5\mu_1$, by analogy with the linear framework (see Remark 1.1.13 of Chapter 1). The oscillations of the closed-loop dynamics can also be restrained using the concept of D-stability. D-stability provides LMI conditions ensuring that the poles of (2.100) are located in a specific region of the complex plane $\mathbb{C}$. The reader is referred to [67, 214, 210] for more details on this pole assignment strategy.

The PDC scheme is also briefly mentioned in the discrete-time case. The stabilizing LMI conditions are analogous to the continuous-time conditions described hereabove.

Theorem 2.4.2 (Discrete quadratic PDC stabilization [304]). *The discrete-time closed-loop T-S model (2.100) is globally exponentially stabilizable using the PDC scheme (2.101) if there exist $X_1, X_2 \in \mathbb{S}_{n_x}(\mathbb{R})$ and $\{M_i\}_{1\leq i\leq n_h}$, $\{N_i\}_{1\leq i\leq n_h}$, with $M_i \in \mathbb{R}^{n_u\times n_x}$ and $N_i \in \mathbb{R}^{n_x\times n_y}$ for all*

$i \in [\![1, n_h]\!]$, *such that the conditions* (2.106) *are satisfied.*

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \begin{pmatrix} X_1 & X_1 A_i^\top + M_j^\top B_i^\top \\ A_i X_1 + B_i M_j & X_1 \end{pmatrix} \succ 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.106a}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \begin{pmatrix} X_2 & A_i^\top X_2 + C_j^\top N_i^\top \\ X_2 A_i + N_i C_j & X_2 \end{pmatrix} \succ 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.106b}$$

The inequalities on the double convex sums can be transformed into regular LMI conditions by leveraging the results of Section 2.2.2. The gain matrices K_i and L_i are retrieved with $K_i = M_i X_1^{-1}$ and $L_i = X_2^{-1} N_i$. Moreover, there exists $\lambda \in \mathbb{R}_{>0}$ such that the matrices P_1 and P_2 providing the QLF (2.102) *can be taken as $P_1 = \lambda X_1^{-1}$ and $P_2 = X_2$.*

Proof. The proof follows the same outline as the proof of Theorem 6.11 in [32]. □

The nPDC scheme (2.103) is now investigated using the nQLF (2.104) in the continuous-time case. The nQLF will generally lead to a local stabilization result.

Theorem 2.4.3 (Continuous multiquadratic nPDC local stabilization [32])**.** *The continuous-time closed-loop T-S model* (2.100) *is locally exponentially stabilizable using the nPDC scheme* (2.103) *if there exist $\{X_{1,i}, X_{2,i}\}_{1\leq i\leq n_h}$ and $\{K_i\}_{1\leq i\leq n_h}$, $\{L_i\}_{1\leq i\leq n_h}$, with $X_{1,i}, X_{2,i} \in \mathbb{S}_{n_x}(\mathbb{R})$, $K_i \in \mathbb{R}^{n_u\times n_x}$ and $L_i \in \mathbb{R}^{n_x\times n_y}$ for all $i \in [\![1, n_h]\!]$, such that the conditions* (2.107) *are satisfied.*

$$X_{i,j} \succ 0, \quad (i,j) \in \{1,2\} \times [\![1, n_h]\!] \tag{2.107a}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \mathcal{H}(X_{1,i} A_j + B_j K_i) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.107b}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \mathcal{H}(X_{2,i} A_j + L_j C_i) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.107c}$$

The inequalities on the double convex sums can be transformed into regular LMI conditions by leveraging the results of Section 2.2.2.

Remark 2.4.2. *The "local" should be understood here in the sense that both the initial state $x(t_0)$ and the initial estimation error $\hat{x}(t_0) - x(t_0)$ are in a neighbourhood of 0.*

Remark 2.4.3. *As in the PDC case, the concept of D-stability is applicable to the nPDC scheme for pole assignment purposes [65].*

The local region of attraction can be characterized by including bounds on the instantaneous variations of the activation functions $\mathbf{h}$ in the LMI conditions above, as performed in the stability section of this chapter [211]. However, contrary to the stability conditions discussed earlier, these Lipschitz assumptions can only be verified after the computation of the gains K_i and L_i. Yet, these gains are computed using the LMI conditions relying on these same Lipschitz

assumptions: this is a kind of *bootstrap paradox*. An *a priori* globally Lipschitz assumption has to be made in order to compute the gains of the controller (2.99), that will eventually, *a posteriori*, prove this Lipschitz assumption to be verified locally. Therefore, as opposed to the stability LMI conditions, the stabilization LMI conditions suffer from the Lipschitz assumptions on $\mathbf{h}$, since it becomes very difficult to impose a region of attraction *a priori*. This is the main drawback of the nPDC scheme for continuous-time T-S models.

Contrary to the PDC scheme where the discrete-time case is completely analogous to the continuous-time case, the discrete-time nPDC scheme differs in some important aspects from its continuous-time equivalent. Most importantly, and similarly to the MQLF investigated in the stability section, the nQLF (2.104) is practical in the discrete-time case as it does not necessitate any Lipschitz assumptions on the activation functions $\mathbf{h}$, which avoids the paradoxical situation discussed previously. Moreover, there exists a large variety of technical ways to derive LMI stabilization conditions for nPDC in the discrete-time case. The interested reader is referred to [116, 86, 32] for more comprehensive expositions of these techniques.

2.4.4 Attenuation criteria

Exogenous input to regulated output ($w(t)$ to $z(t)$) criteria can also be imposed to the closed-loop system (2.100) considered with these exogenous input and regulated output variables. In particular, these input to output conditions usually take the form of $\mathcal{H}_2$ or $\mathcal{H}_\infty$ attenuation criteria. These criteria come from the linear framework, where they represent, in the frequency domain, some bounds on the norm of the transfer function from $w(t)$ to $z(t)$. Nonlinear systems, such as LPV and T-S systems, do not have a direct frequency domain interpretation, making it difficult to obtain a transfer function from them; nevertheless, there exists some time domain interpretations of the generalized $\mathcal{H}_2$ and of the $\mathcal{H}_\infty$ criteria which can be applied to nonlinear systems [296, 245]. In a nonlinear setting, assuming a null initial condition to the closed-loop system (i.e. $x_{c\ell}(t_0) = 0$), these criteria are given as follows.

- The generalized $\mathcal{H}_2$ criterion bounds the ratio between the maximum instantaneous 2-norm of the regulated output $z(t)$ (its $\mathcal{L}_\infty$ norm) and the square root of the energy of the exogenous input $w(t)$ (its $\mathcal{L}_2$ norm) on the interval $[t_0, +\infty)$.
- The $\mathcal{H}_\infty$ criterion bounds the ratio between the square root of the energy of the regulated output $z(t)$ (its $\mathcal{L}_2$ norm) and the square root of the energy of the exogenous input $w(t)$ (its $\mathcal{L}_2$ norm) on the interval $[t_0, +\infty)$.

These attenuation criteria can be imposed using some LMI conditions, both in the context of the gain-scheduled Luenberger observer and of the gain-scheduled state feedback controller investigated in this section. The interested reader is referred to [301, 299, 279, 116, 117] for more details. The LMI conditions to check an $\mathcal{H}_\infty$ criterion on the continuous-time closed-loop T-S system (2.108) using the QLF (1.21), i.e. $V(x_{c\ell}) = x_{c\ell}^\top P x_{c\ell}$, with $P \in \mathbb{S}^{++}_{n_{x_{c\ell}}}(\mathbb{R})$, are reported thereafter.

$$\begin{pmatrix} \dot{x}_{c\ell}(t) \\ z(t) \end{pmatrix} = \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} h_i(\theta) h_j(\theta) \begin{pmatrix} A_{c\ell,i,j} & B_{c\ell,i} \\ C_{c\ell,i} & D_{c\ell,i} \end{pmatrix} \begin{pmatrix} x_{c\ell}(t) \\ w(t) \end{pmatrix} \tag{2.108}$$

In particular, the closed-loop system (2.100) provides the augmented closed-loop state $x_{c\ell} = \begin{pmatrix} x^\top & (\hat{x} - x)^\top \end{pmatrix}^\top$.

Theorem 2.4.4 ($\mathcal{H}_\infty$ criterion [173]). *The $\mathcal{H}_\infty$ criterion*

$$\sup_{w\in\mathcal{L}_2\setminus\{0\}} \frac{\|z\|_{\mathcal{L}_2}}{\|w\|_{\mathcal{L}_2}} \triangleq \sup_{w\in\mathcal{L}_2\setminus\{0\}} \sqrt{\frac{\int_{t_0}^{+\infty}\|z(t)\|_2^2 dt}{\int_{t_0}^{+\infty}\|w(t)\|_2^2 dt}} \leq \gamma \tag{2.109}$$

is verified on the closed-loop system (2.108) for $x_{c\ell}(t_0) = 0$ if there exists $X \in \mathbb{S}_{n_{x_{c\ell}}}(\mathbb{R})$ such that the condition (2.110) is verified.

$$X \succ 0 \tag{2.110a}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \begin{pmatrix} \mathcal{H}(A_{c\ell,i,j}X) & B_{c\ell,i} & XC_{c\ell,i}^\top \\ B_{c\ell,i}^\top & -\gamma^2 I_{n_w} & D_{c\ell,i}^\top \\ C_{c\ell,i}X & D_{c\ell,i} & -I_{n_e} \end{pmatrix} \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{2.110b}$$

The inequality on the double convex sum can be transformed into regular LMI conditions by leveraging the results of Section 2.2.2. The matrix P providing the QLF (1.21) is retrieved using $P = X^{-1}$.

The condition described hereabove can no longer be easily turned into LMI conditions using the results of Section 2.2.2 once the T-S expression of the closed-loop system (2.100) with unknown gains K_i and L_i is injected into the matrices $A_{c\ell,i,j}$. Nonetheless, LMI conditions can still be obtained for this $\mathcal{H}_\infty$ criterion in the context of the observer-based controller (2.99). Again, the interested reader is referred to [301, 174, 151] for more details.

2.5 Conclusions and perspectives

In this chapter, some basics results surrounding T-S models have been introduced, exploring in particular the LMI formulation of the stability and stabilization control problems using both QLF and nQLF. After introducing generic results on LMI, it has been shown that non-conservative LMI conditions are not easily obtained in a double convex sum context. Then, the stability problem of T-S models was extensively discussed both in the continuous-time and the discrete-time cases. In particular, by analogy with the LDI framework, some necessary and sufficient stability conditions have been derived under an uncertainty assumption on the activation functions $\mathbf{h}$. This necessary and sufficient conditions are rarely discussed in the literature and remains to be investigated further. Stabilization was then investigated: the separation principle was recalled for T-S systems with measurable premise variables, and stabilization conditions were given both for PDC and nPDC controller schemes. Finally, the generalized $\mathcal{H}_2$ and $\mathcal{H}_\infty$ attenuation criteria were briefly discussed.

Chapter 3

Convex modeling of Takagi-Sugeno systems

This chapter explains how the nonlinear sector approach relies on barycentric coordinates. This observation enables the explicit derivation of exact T-S models from nonlinear systems with bounded nonlinearities, with a large flexibility in the bounding shape, impacting both the model complexity and its conservatism.

3.1 Takagi-Sugeno modeling

Two main approaches exist in the literature to obtain a T-S model from a nonlinear system:

- numerical identification using measurements of the system behaviour;
- analytical construction using an already existing nonlinear model of the system dynamic.

The first approach is out of the scope of this manuscript and the interested reader is referred to the following works [1, 138, 286, 167, 180, 284]. Common examples of the second approach are:

- the dynamic linearization at different operating points of the system [139];
- the use of unimodal basis functions [153];
- the Nonlinear Sector Approach (NLSA) [206, 295].

The latter, also called the convex polytopic transformation [35], is the most prominent in the literature of T-S models, since it provides an exact representation of the initial nonlinear system on a box-shaped set of the scheduling parameter space. Formally, for all x, u, w, and t such that the scheduling vector θ is in the box-shaped set $\Theta \subseteq \mathbb{R}^{n_\theta}$, the following equality holds between (1.1) and (2.1):

$$\begin{pmatrix} f(x,u,w,t) \\ h_y(x,u,w,t) \\ h_z(x,u,w,t) \end{pmatrix} = \sum_{i=1}^{n_h} h_i(\theta(x,u,w,t)) \begin{pmatrix} A_i & B_{1,i} & B_{2,i} \\ C_{1,i} & D_{11,i} & D_{12,i} \\ C_{2,i} & D_{21,i} & D_{22,i} \end{pmatrix} \begin{pmatrix} x \\ u \\ w \end{pmatrix} \quad \text{[exactness]} \quad (3.1)$$

As a matter of fact, this exactness is explained by a third property respected by the activation functions $\mathbf{h}$ produced by the NLSA: the *linear precision* property. This property states that given a vector θ in the convex hull of $\{\mathcal{V}_1, \ldots, \mathcal{V}_{n_h}\}$, weighting each vertex $\mathcal{V}_i$ by $h_i(\theta)$ yields back θ [294]. Linear precision makes $\mathbf{h}$ exact barycentric coordinates of the initial nonlinear system within the convex hull of the local LTI models.

Definition 3.1.1 (Barycentric coordinates [294]). *A set of functions $h_1, \dots, h_{n_h} : \Theta \to \mathbb{R}$, where Θ is a polytope with vertices $v(\Theta) = \{\mathcal{V}_i\}_{1 \leq i \leq n_h}$, is said to be barycentric coordinates of Θ if for all $\theta \in \Theta$, the functions satisfy the two convex sum properties of Definition 2.1.1 and the following linear precision property:*

$$\sum_{i=1}^{n_h} h_i(\theta)\mathcal{V}_i = \theta \qquad [\textit{linear precision}] \tag{3.2}$$

Despite its popularity, the NLSA is subject to some key issues:

- regarding the number of local models needed, which grows exponentially with the dimension of the scheduling parameter space ($n_h = 2^{n_\theta}$) [126];
- regarding the *intrinsic conservatism* of the resulting model.

Modeling a nonlinear system (1.1) with a T-S model (2.1) generally leads to an analysis where some properties of the nonlinearities of the system are neglected (such as their interdependence, periodicity, etc), which ultimately limits the number of solutions in the optimization problems obtained with a T-S model-based approach to nonlinear control questions. The choice of the T-S model can both improve or worsen this issue, which indicates that some T-S models can be *intrinsically* less conservative than others. Both the number of local models and the intrinsic conservatism of T-S models are especially problematic in the NLSA because of the imposed box-shaped geometry of Θ.

Remark 3.1.1. *Of course, a T-S model does not have "conservatism" solely by itself. The T-S framework having a limited number of commonly investigated control problems, the "intrinsic conservatism" of a T-S model can be roughly defined as the conservatism induced on all of these problems by the choice of this particular model.*

Some mitigating solutions to these modeling issues have been suggested, for example by using a polar coordinate T-S model [205], by generalizing the NLSA for polynomial fuzzy models [274], by using model reduction schemes [305, 300, 186, 84], or simply by carefully choosing the scheduling parameters of the model [160, 199]. However, most of the conservatism reduction techniques of the literature are performed *a posteriori*, that is to say after the construction of the T-S model with its $n_h = 2^{n_\theta}$ local models. As discussed in Chapter 2, the conservatism reduction is usually achieved by using nQLF in the stability analysis of the system [140, 61, 271, 231, 115], and by finding a sharp LMI formulation of the control problem at hand [238, 240, 158, 30, 196, 157]. In particular, one way of sharpening the resulting LMI conditions consists in getting rid of some useless vertices of the T-S model, e.g. by leveraging the interdependencies of the scheduling parameters [233, 146]. However, if these vertices were in fact useless, one could wonder why they were considered in the NLSA in the first place.

The author argue that the key issues of the NLSA could be addressed by acknowledging that the box-shaped bounding set Θ is ultimately incidental to the approach, and could be replaced by larger classes of convex sets. Somehow, the generalization of the NLSA to simple polytopes (see Definition 3.4.3) can already be found in the LPV literature [295], but has received very little attention yet, despite its relevance. Flexibility on the convex polytope used

to bound the scheduling vector θ is also not uncommon in the polytopic LPV model literature [9, 160, 125, 169]. However, since the generalization of the NLSA to simple polytopes is rarely leveraged in the construction of polytopic LPV models, the *explicit* expression of the activation functions is rarely given. This is still a limiting issue in the polytopic LPV framework, in particular for the PDC or nPDC schemes relying on the real-time calculation of these weights. Obtaining an explicit expression to these activation functions could largely reduce the cost time or space complexity associated with their computation, which is usually achieved by solving an optimization problem.

This chapter, published as an article in [20], presents several convex generalizations of the NLSA which are not discussed in the literature yet, and which can not only reduce the minimal number of local models needed to obtain an exact T-S model, but which can also lead to a reduction of the intrinsic conservatism of some T-S models with interdependent scheduling parameters.

This chapter is organized as follows: Section 3.2 provides a generic description of the NLSA and introduces a nonlinear system which is studied numerically in each of the following sections. Section 3.3 focuses on the NLSA for box-shaped bounding sets, which is the well-known NLSA for T-S models [206]. Section 3.4 extends the previous result to polytopic bounding sets, which includes the lesser-known generalization of the NLSA for simple polytopes [295]. Section 3.5 provides similar results for convex bounding sets with smooth boundaries, which results in a generalization of the NLSA for a new kind of T-S models which is not discussed in the literature yet. New stability results are given to study these T-S-like models using LMI. Finally, some conclusions and perspectives are discussed in Section 3.6.

3.2 The general nonlinear sector approach

The NLSA is a two-step procedure to obtain a T-S model (2.1) which exactly represents a nonlinear system (1.1) on a peculiar set. In this section, the two steps of the NLSA procedure are described succinctly, and a nonlinear system is introduced to illustrate the first step of this procedure. This nonlinear system is reused in each section of the chapter with variation on the second step of the NLSA.

3.2.1 Methodology

Step 1: The first step of the NLSA consists in rewriting (1.1) as a general LPV system of the following form:

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = E(\theta) \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \text{ with } E(\theta) \triangleq \begin{pmatrix} A(\theta) & B_1(\theta) & B_2(\theta) \\ C_1(\theta) & D_{11}(\theta) & D_{12}(\theta) \\ C_2(\theta) & D_{21}(\theta) & D_{22}(\theta) \end{pmatrix} \tag{3.3}$$

and where $\theta \in \mathbb{R}^{n_\theta}$ is a vector of scheduling parameters, which usually depends on $x(t)$, $u(t)$, $w(t)$ and t, and taken such that

$$E(\theta) = E_0 + \sum_{i=1}^{n_\theta} \theta_i E_i \tag{3.4}$$

Remark 3.2.1. *Several LPV representations are often already possible at this stage, some being more advantageous than others in terms of conservatism reduction or of structural properties (such as controllability and observability). However, this discussion is beyond the scope of this manuscript, and the interested reader is referred to [160, 199].*

Step 2: The second step consists in bounding the values of θ within a set Θ of $\mathbb{R}^{n_\theta}$ with realistic assumptions on $x(t)$, $u(t)$, $w(t)$ and t. These assumptions can sometimes lead to a T-S model exactly representing the original nonlinear system (1.1), but only locally. Assuming Θ is a polytope, from the affine relation between θ and E, barycentric coordinates of θ within Θ are also barycentric coordinates of E within its bounds. Indeed, if Θ is a polytope with vertices $v(\Theta) = \{\mathcal{V}_i\}_{1\leq i\leq n_h}$ and with barycentric coordinates $\mathbf{h}$, this *transfer* of barycentric coordinates is easily shown with the following operations:

$$\begin{aligned} E(\theta) &= E_0 + \sum_{i=1}^{n_\theta} \theta_i E_i \\ &= E_0 + \sum_{i=1}^{n_\theta} \left(\sum_{j=1}^{n_h} h_j(\theta) \left(\mathcal{V}_j\right)_{(i)} \right) E_i \\ &= \sum_{j=1}^{n_h} h_j(\theta) \left(E_0 + \sum_{i=1}^{n_\theta} \left(\mathcal{V}_j\right)_{(i)} E_i \right) \\ E(\theta) &= \sum_{j=1}^{n_h} h_j(\theta) E\left(\mathcal{V}_j\right) \end{aligned} \tag{3.5}$$

By linearity, this provides a convex-sum representation of (3.3) for all $x(t)$, $u(t)$, $w(t)$ and t such that $\theta \in \Theta$, which is an exact T-S model representation of (1.1).

This chapter focuses on the second step of the approach. In particular, based on the work of [206] and [294], the next sections provide barycentric coordinates of θ in Θ respectively when Θ is a box (Section 3.3), a polytope (Section 3.4) and a convex set with a smooth boundary (Section 3.5). An intuitive measure of the intrinsic conservativeness of the resulting T-S models is given by the size of the subset of Θ that does not contain admissible values of θ [160]. This clearly indicates that some of bounding geometries are better suited for some systems than others depending on the nonlinearities of (1.1). However, even if this intuitive measure of conservativeness is arguably accurate, some analytical difficulties or simplification can still show up during the practical manipulation of the model and affect the conservativeness of resulting stability or stabilization analysis. Rather than providing a unique process to obtain the single best NLSA to construct an exact T-S model from a nonlinear system, each section emphasizes the singularity of the presented approach. The author's goal is to underline the diversity of the possible exact T-S representations obtainable from the NLSA, a topic which is not broadly discussed in the literature yet.

3.2.2 Application

As seen in Chapter 2, the T-S models being a convex sum of linear models, it is possible to numerically conduct their stability analysis through convex optimization techniques such as semidefinite programming, in particular by leveraging results on QLF. Throughout this chapter, several T-S representations will be compared in terms of their respective capabilities in

the stability analysis of a nonlinear system, defined by the following second order nonlinear differential equation:

$$\ddot{z}(t) + (1 + \alpha \cos z(t))\dot{z}(t) + (1 + \beta \sin z(t))z(t) = 0 \tag{3.6}$$

which can be rewritten as

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} x_2(t) \\ -(1 + \beta \sin x_1(t))x_1(t) - (1 + \alpha \cos x_1(t))x_2(t) \end{pmatrix} \tag{3.7}$$

with $x_1(t) = z(t)$ and $x_2(t) = \dot{z}(t)$. The vector field associated to this differential equation is plotted on Figure 3.1 for different (α, β) values. The first step of the NLSA is applied to system (3.7), which can be rewritten as a quasi-LPV model by considering a scheduling vector $\theta = \begin{pmatrix} \theta_1 & \theta_2 \end{pmatrix}^\top$ such that

$$\dot{x}(t) = A(\theta)x(t) \tag{3.8a}$$

$$\text{with } A(\theta) \triangleq \begin{pmatrix} 0 & 1 \\ -1 - \beta\theta_1 & -1 - \alpha\theta_2 \end{pmatrix} \text{ and } \begin{cases} \theta_1 \triangleq \sin x_1(t) \\ \theta_2 \triangleq \cos x_1(t) \end{cases} \tag{3.8b}$$

Remark 3.2.2. *Of course, this LPV representation is not the only one which could be considered.*

The scheduling parameters (θ_1, θ_2) are interdependent, with $\theta_1^2 + \theta_2^2 = 1$ being always verified. This demonstrates that θ is on the unit circle centered at the origin of the scheduling space $\mathbb{R}^{n_\theta}$. Note that these interdependent nonlinearities are also found in real-world dynamical systems, e.g. in the Vertical Take-Off and Landing (VTOL) aircraft model [121], or in the kinematic model of a wheeled mobile robot [178]. Throughout this chapter, the unit circle is bounded by a set Θ whose shape varies depending on the considered NLSA. The conservatism of each NLSA is evaluated by comparing the values of α and β for which the resulting T-S representations of (3.7) are found to be globally exponentially stable. In particular, these results are compared to the following theoretical guarantee, which relies on a QLF.

Proposition 3.2.1 (Exponential stability - perturbative QLF approach)**.** *The solutions to* (3.6) *are globally exponentially stable if*

$$\max(|\alpha|, |\beta|) < 2/(5 + \sqrt{5}) \approx 0.27639... \tag{3.9}$$

Proof. Equation (3.7) is equivalent to a perturbed system of form

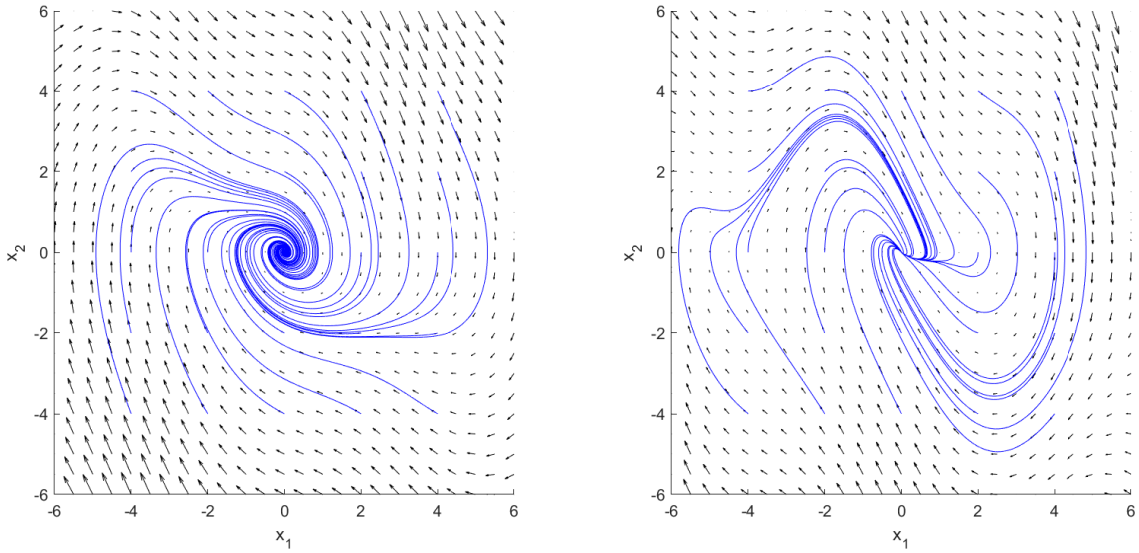
$$\dot{x}(t) = Ax(t) + g(x(t)) \tag{3.10}$$

$$\text{with } A = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} \text{ and } g(x) = \begin{pmatrix} 0 \\ -\alpha x_2 \cos x_1 - \beta x_1 \sin x_1 \end{pmatrix} \tag{3.11}$$

By Lyapunov lemma 1.1.1, if $P \in \mathbb{S}_2(\mathbb{R})$ is a positive definite matrix such that $\mathcal{H}(PA) = -I_2$, then $V(x) = x^\top Px$ is a QLF demonstrating that $\dot{x}(t) = Ax(t)$ is globally exponentially stable. It is easy to find such a P:

$$P = \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1 \end{pmatrix} \tag{3.12}$$

Moreover, if $\|g(x)\|_2 \leq \gamma \|x\|_2$ and $\gamma < \frac{1}{2\lambda_{\max}(P)}$, then a result on perturbed systems provides the global exponential stability of (3.10) as well (see Chapter 9 of [149]). Here, $\gamma = \max(|\alpha|, |\beta|)$ and $\lambda_{\max}(P) = \frac{1}{4}(5 + \sqrt{5})$, which concludes the proof. □



(a) $\alpha = -0.5, \beta = 0.3$. (b) $\alpha = 0.9, \beta = -0.8$.

Figure 3.1: Vector field associated with the differential equation (3.7) for different (α, β) values, with some trajectories (in blue) converging towards the origin for a set of evenly spread initial conditions.

Moreover, note that the following result on asymptotic stability can be stated using a nQLF, which is much more complicated to find analytically.

Proposition 3.2.2 (Asymptotic stability - nQLF approach)**.** *The solutions to* (3.6) *are globally asymptotically stable if*

$$\max(|\alpha|, |\beta|) < 1 \qquad (i.e.\ (\alpha, \beta) \in (-1, 1)^2) \tag{3.13}$$

Proof. By analogy with a nonlinear mass-spring system, the following nQLF is considered (see Figure 3.2):

$$\begin{aligned} V(x) &= \frac{1}{2}x_2^2 + \int_0^{x_1} (1 + \beta \sin\sigma)\sigma d\sigma \\ &= \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \beta(\sin(x_1) - \cos(x_1)x_1) \end{aligned} \tag{3.14}$$

where $\frac{1}{2}x_2^2$ stands for the kinetic energy of the system and $\int_0^{x_1}(1 + \beta \sin\sigma)\sigma d\sigma$ stands for its potential energy derived from the nonlinear pull-back force $-(1 + \beta \sin x_1)x_1$. V is indeed positive definite for $\beta \in (-1, 1)$ and:

$$\dot{V}(x(t)) = -(1 + \alpha \cos x_1(t))x_2^2(t) \tag{3.15}$$

hence for all $\alpha \in (-1, 1)$, the inequality $\dot{V}(x(t)) \leq 0$ is verified, and for all $\beta \in (-1, 1)$, no solution except $x(t) = 0$ stays in the set $\{x \in \mathbb{R}^{n_x} : \dot{V}(x) = 0\}$. By the invariance principle, this provides the global asymptotic stability of the origin for all $(\alpha, \beta) \in (-1, 1)^2$ [149]. □

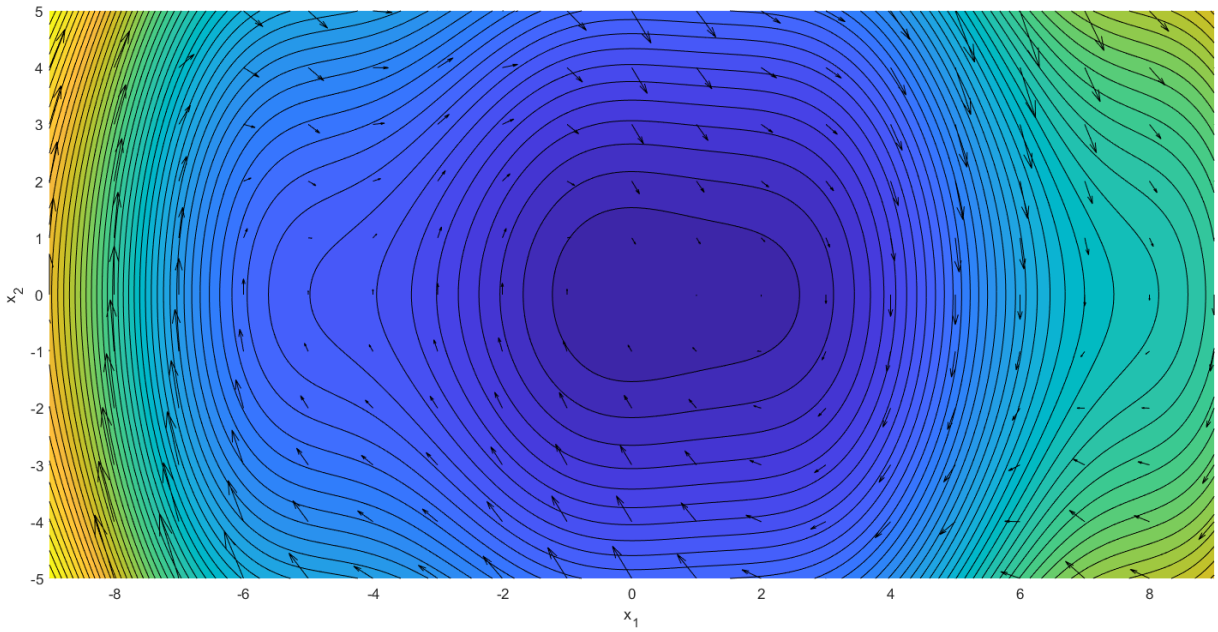


Figure 3.2: Vector field associated with the differential equation (3.7) for $(\alpha, \beta) = (0.9, -0.8)$ together with the level sets of the nQLF (3.14).

Finding this type of nQLF is in general a difficult problem. T-S modeling offers a way to rewrite the nonlinear model so the stability and stabilization analyzes can be easily and systematically handled numerically, without relying on model-specific results such as the Proposition 3.2.2 hereabove.

3.3 The nonlinear sector approach for boxes

This section summarizes the most well-known NLSA, where the scheduling vector θ is bounded within a box-shaped set Θ. Lemma 3.3.1 provides barycentric coordinates of θ inside Θ. Following the procedure of the NLSA, Theorem 3.3.1 uses these barycentric coordinates as activation functions in order to obtain a T-S model which exactly represents the nonlinear system (1.1) for all $\theta \in \Theta$. After discussing the advantages and inconveniences of the NLSA on a box-shaped set, the stability analysis of the nonlinear system (3.7) of Section 3.2.2 is performed using this NLSA.

3.3.1 Barycentric coordinates

An expression for the barycentric coordinates of a box is reported thereafter.

Lemma 3.3.1 (Barycentric coordinates of a box)**.** *Let $\theta \in \mathbb{R}^{n_\theta}$ be bounded by the box-shaped polytope $\Theta = [\underline{\theta}_1, \overline{\theta}_1] \times \cdots \times [\underline{\theta}_{n_\theta}, \overline{\theta}_{n_\theta}]$, i.e. by the convex hull of $\{\mathcal{V}_i\}_{1 \leq i \leq 2^{n_\theta}}$ with for all $i \in [\![1, 2^{n_\theta}]\!]$*

$$\mathcal{V}_i = \mathcal{V}_{\overline{(i-1)_{[n_\theta]} \ldots (i-1)_{[1]}}^{(2)}+1} \text{ with, for all } k \in [\![1, n_\theta]\!], (\mathcal{V}_i)_{(k)} = \begin{cases} \overline{\theta}_k & \text{if } (i-1)_{[k]} = 1 \\ \underline{\theta}_k & \text{if } (i-1)_{[k]} = 0 \end{cases} \tag{3.16}$$

For all $i \in [\![1, 2^{n_\theta}]\!]$, let K_i^+ and K_i^- stand resp. for the indices of 1 and 0 digits in the standard base-2 positional notation of $(i-1)$. Formally:

$$K_i^+ = \left\{k \in [\![1, n_\theta]\!] \ : \ (i-1)_{[k]} = 1\right\} \tag{3.17a}$$
$$K_i^- = \left\{k \in [\![1, n_\theta]\!] \ : \ (i-1)_{[k]} = 0\right\} \tag{3.17b}$$

Barycentric coordinates of θ in Θ associated to the vertices $v(\Theta) = \{\mathcal{V}_i\}_{1 \leq i \leq 2^{n_\theta}}$ are given by $\{h_{\mathcal{V}_i}\}_{1 \leq i \leq 2^{n_\theta}}$ with for all $i \in [\![1, 2^{n_\theta}]\!]$

$$h_{\mathcal{V}_i}(\theta) = \frac{1}{\prod_{k=1}^{n_\theta} (\overline{\theta}_k - \underline{\theta}_k)} \left(\prod_{k \in K_i^-} (\overline{\theta}_k - \theta_k) \right) \left(\prod_{k \in K_i^+} (\theta_k - \underline{\theta}_k) \right) \tag{3.18}$$

Proof. For all $n_\theta \in \mathbb{N}^*$ and $\theta \in \Theta$, the three axioms of barycentric coordinates (2.2a), (2.2b), (3.2) are easily verified with $n_h = 2^{n_\theta}$. The crucial steps of the proof of the linear precision property can be found in [206], and the two other properties are trivial by construction. This result can also be viewed as a special case of barycentric coordinates for a simple polytope: this is demonstrated in details in Section 3.4.1. □

The usual NLSA for T-S models is stated in Theorem 3.3.1.

Theorem 3.3.1 (NLSA on a box)**.** *The T-S model (2.1) with the activation functions* **h** *obtained using Lemma 3.3.1 is an exact representation of the nonlinear system (1.1) for all $x(t)$, $u(t)$, $w(t)$ and t such that $\theta \in \Theta = [\underline{\theta}_1, \overline{\theta}_1] \times \cdots \times [\underline{\theta}_{n_\theta}, \overline{\theta}_{n_\theta}]$, i.e. (3.1) holds for all $\theta \in \Theta$.*

The intrinsic conservatism of the resulting T-S model is generally fair if the initial nonlinear system (1.1) has independent and narrowly bounded scheduling parameters θ_i. Moreover, the barycentric coordinates $\{h_{\mathcal{V}_i}\}_{1 \leq i \leq 2^{n_\theta}}$ are very easy to construct, and they are polynomial in the scheduling parameters, which allows for some conservatism reduction in the stability analysis [239, 240]. However, as discussed in the introduction of this chapter, this exact representation needs a number of vertices which is exponentially growing in the number of scheduling parameters ($n_h = 2^{n_\theta}$), and it generally ceases to be the best representation as long as some of the scheduling parameters are not fully decoupled and show some interdependencies.

3.3.2 Application

Considering the LPV model (3.8), the box $\Theta = [-1, 1]^2$ is chosen to be the bounding set of θ. As illustrated by Figure 3.3, Θ has four vertices:

$$\begin{array}{ll} \mathcal{V}_1 = \begin{pmatrix} -1 & -1 \end{pmatrix}^\top & \mathcal{V}_2 = \begin{pmatrix} 1 & -1 \end{pmatrix}^\top \\ \mathcal{V}_3 = \begin{pmatrix} -1 & 1 \end{pmatrix}^\top & \mathcal{V}_4 = \begin{pmatrix} 1 & 1 \end{pmatrix}^\top \end{array} \tag{3.19}$$

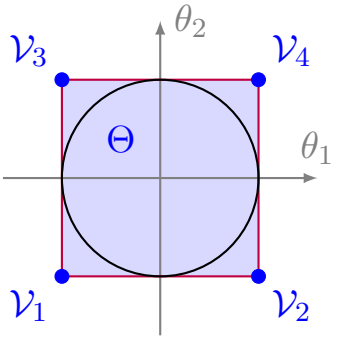


Figure 3.3: The box-shaped set Θ is bounding the scheduling vector θ of (3.8).

The barycentric coordinates of θ are given by Lemma 3.3.1:

$$h_{\mathcal{V}_1}(\theta) = \frac{1}{4}(1-\theta_1)(1-\theta_2) \tag{3.20a}$$

$$h_{\mathcal{V}_2}(\theta) = \frac{1}{4}(1+\theta_1)(1-\theta_2) \tag{3.20b}$$

$$h_{\mathcal{V}_3}(\theta) = \frac{1}{4}(1-\theta_1)(1+\theta_2) \tag{3.20c}$$

$$h_{\mathcal{V}_4}(\theta) = \frac{1}{4}(1+\theta_1)(1+\theta_2) \tag{3.20d}$$

Finally, the NLSA on the box Θ is completed and the system (3.7) is exactly represented by the following T-S model:

$$\dot{x}(t) = \sum_{i=1}^{4} h_{\mathcal{V}_i}(\theta) A(\mathcal{V}_i) x(t) \tag{3.21}$$

The stability analysis of the T-S model (3.21) is performed using Theorem 2.3.1 at several $(\alpha, \beta) \in [-1, 1]^2$ values. The (α, β)-region for which the system (3.21) is found to be exponentially stable is plotted on Figure 3.4, where it is compared to the (α, β)-region obtained with the perturbative approach (Proposition 3.2.1). The (α, β)-region of stability is visibly larger with Theorem 2.3.1 on (3.21) than with the perturbative approach, but contrary to the latter, the region is only a discrete subset of the (α, β)-plane.

3.4 The nonlinear sector approach for polytopes

This section broadens the well-known NLSA from box-shaped bounding sets to polytopic bounding sets Θ. After defining *simple polytopes* in Definition 3.4.3, Lemma 3.4.1 provides barycentric coordinates of θ inside a simple polytope Θ, and Lemma 3.4.2 extends these barycentric coordinates to general polytopes. The previous barycentric coordinates of a box-shaped set are retrieved as a special case of the proposed polytopic barycentric coordinates in Corollary 3.4.1. Following the procedure of the NLSA, Theorem 3.4.1 uses these barycentric coordinates as activation functions in order to obtain a T-S model which exactly represents the nonlinear system (1.1) for all $\theta \in \Theta$. After discussing the advantages and inconveniences of the NLSA on a polytopic set, these results are applied to the nonlinear system (3.7) of Section 3.2.2.

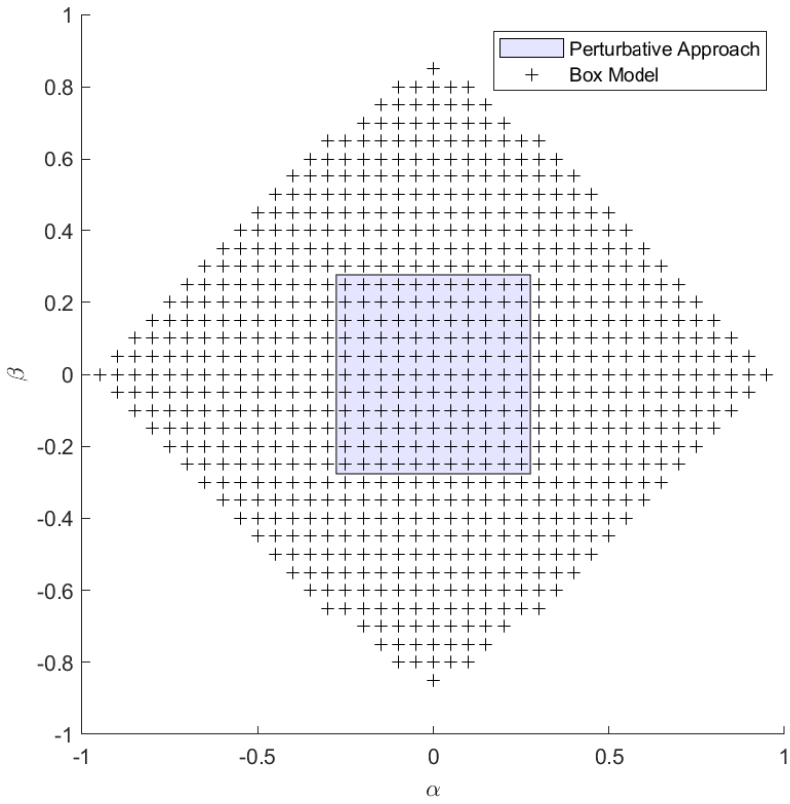


Figure 3.4: Stability (α, β)-regions of (3.6) using Theorem 2.3.1 on the T-S model (3.21) (*box model*), and using the perturbative approach (*Proposition 3.2.1*).

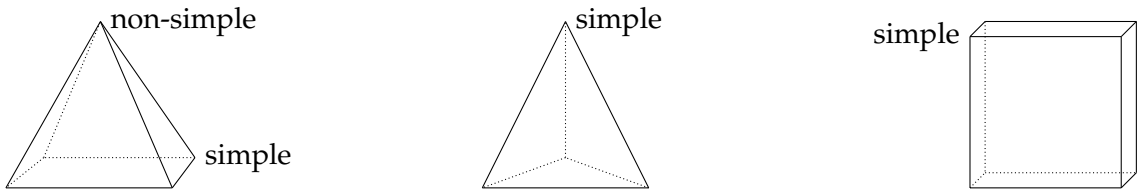
3.4.1 Barycentric coordinates

The definition of the faces of a polytope is introduced thereafter. The facets of a polytope are then used in the definitions of simple vertices and simple polytopes, which are introduced afterward. The reader is referred to [314, 110] for a more complete exposition on the theory of convex polytopes.

Definition 3.4.1 (Faces of a polytope)**.** *A n-dimensional polytope Θ is defined by the convex-hull of a finite set of vectors such that the intrinsic dimension of the convex-hull is n, i.e. $\dim(\Theta) = n$. The faces of Θ are defined as follows:*

- *the n-face of Θ is Θ itself;*
- *the $(n-1)$-faces of Θ are called its facets, and are denoted $f(\Theta)$. They are the largest (with respect to inclusion) $(n-1)$-dimensional polytopes belonging to the boundary of Θ;*
- *the k-faces of Θ are the facets of all the $(k+1)$-faces of Θ, with $k \in [\![0, n-1]\!]$;*
- *the 0-faces of Θ are called its vertices, and are denoted $v(\Theta)$.*

Definition 3.4.2 (Simple vertex [294])**.** *Given a n-dimensional polytope Θ with vertices $v(\Theta) = \{\mathcal{V}_i\}_{1 \leq i \leq n_h}$, $\mathcal{V}_i$ is a simple vertex of Θ if the set of facets of Θ containing $\mathcal{V}_i$, which is now denoted*



(a) Pyramid of $\mathbb{R}^3$. (b) Simplex Δ_3 of $\mathbb{R}^3$. (c) Box of $\mathbb{R}^3$.

Figure 3.5: Illustration of Examples 3.4.1 and 3.4.2.

$\mathrm{ind}(\mathcal{V}_i)$, *has exactly* n *elements. Formally,* $\mathcal{V}_i$ *is simple if*

$$\#\,\mathrm{ind}(\mathcal{V}_i) = \dim(\Theta) \tag{3.22}$$

Otherwise, $\mathcal{V}_i$ *is called a non-simple vertex of* Θ*.*

Definition 3.4.3 (Simple polytope [294])**.** *The subset of (non-)simple vertices of a polytope* Θ *is denoted* $v_s(\Theta) = \left\{\mathcal{V}_{s(i)}\right\}_{1\leq i\leq n_s}$ *(resp.* $v_{\bar{s}}(\Theta)$*) with* $s : [\![1, n_s]\!] \to [\![1, n_h]\!]$ *(resp.* $\bar{s}$*) an injective function.* Θ *is called a simple polytope if all of its vertices are simple. Formally,* Θ *is simple if*

$$v_s(\Theta) = v(\Theta) \quad (\text{or } v_{\bar{s}}(\Theta) = \emptyset) \tag{3.23}$$

Example 3.4.1. *The four vertices at the base of a pyramid of* $\mathbb{R}^3$ *with a square base are simple, since they are all contained in exactly three facets of the pyramid. However, its apex is not a simple vertex since it is contained in exactly four facets. Hence, the pyramid is not a simple polytope. This is illustrated in Figure 3.5a.*

Example 3.4.2. *All vertices of the* n*-simplex and of the* n*-box of* $\mathbb{R}^n$ *are simple since they are all contained in exactly* n *facets of these polytopes. Hence, the* n*-simplex and the* n*-box are simple polytopes. This is illustrated for* $\mathbb{R}^3$ *in Figures 3.5b and 3.5c.*

Example 3.4.3. *In general, the 2-dimensional polytopes of* $\mathbb{R}^2$ *are simple. Indeed, this polytopes are polygons, and every vertex of a polygon is contained in exactly two edges.*

An expression for the barycentric coordinates of a simple polytope is reported thereafter.

Lemma 3.4.1 (Barycentric coordinates of a simple polytope)**.** *Let* $\theta \in \mathbb{R}^{n_\theta}$ *be bounded by the simple polytope* Θ *with vertices* $v(\Theta) = \{\mathcal{V}_i\}_{1\leq i\leq n_h}$*. For all* $i \in [\![1, n_h]\!]$*,* $w_{\mathcal{V}_i}$ *stands for the weight function associated to* $\mathcal{V}_i$*, with*

$$w_{\mathcal{V}_i}(\theta) = \frac{\left|\det\left(\mathcal{N}_{\mathrm{ind}(\mathcal{V}_i)}\right)\right|}{\prod\limits_{\mathcal{F}\in\mathrm{ind}(\mathcal{V}_i)} \langle \mathcal{N}_{\mathcal{F}} | \mathcal{V}_i - \theta\rangle} \tag{3.24}$$

In the expression above, given $\mathcal{F}$ *a facet of* Θ*,* $\mathcal{N}_{\mathcal{F}}$ *stands for the exterior-pointing normal to* $\mathcal{F}$*. In particular,* $\mathcal{N}_{\mathrm{ind}(\mathcal{V}_i)}$ *is the matrix whose columns are formed by the exterior-pointing normals to*

each facet of $\mathrm{ind}(\mathcal{V}_i)$.

Barycentric coordinates of θ *in* Θ *associated to the vertices* $v(\Theta)$ *are given by* $\{h_{\mathcal{V}_i}\}_{1\leq i\leq n_h}$, *a normalization of the previous weights. For all* $i \in [\![1, n_h]\!]$, $h_{\mathcal{V}_i}$ *is given by*

$$h_{\mathcal{V}_i}(\theta) = \frac{w_{\mathcal{V}_i}(\theta)}{\sum_{j=1}^{n_h} w_{\mathcal{V}_j}(\theta)} \tag{3.25}$$

Proof. For all $n_h \in \mathbb{N}^*$ and $\theta \in \Theta$, the three axioms of barycentric coordinates (2.2a), (2.2b), (3.2) are verified. The proof that (3.2) stands can be found in [294], and the two other properties are trivial by construction. □

Corollary 3.4.1. *The box being a special case of a simple polytope, barycentric coordinates of a box can be recovered from the results on barycentric coordinates of a simple polytope.*

Proof. The barycentric coordinates functions constructed for boxes [206] and simple polytopes [294] are rational functions of the same minimal degree. By unicity of the barycentric coordinates of minimal degree, the two constructions are necessarily equal to each other when a box is considered [293]. To verify this claim, the formula of barycentric coordinates of a box are retrieved below from the formula of the barycentric coordinates of a simple polytope. Taking $n_h = 2^{n_\theta}$, $\Theta = [\underline{\theta}_1, \overline{\theta}_1] \times \cdots \times [\underline{\theta}_{n_\theta}, \overline{\theta}_{n_\theta}]$, the facets of the polytope Θ are given by $f(\Theta) = \left\{\mathcal{F}_k^+, \mathcal{F}_k^-\right\}_{1\leq k\leq n_\theta}$, with for all $k \in [\![1, n_\theta]\!]$

$$\mathcal{F}_k^+ = \mathrm{hull}\left\{\mathcal{V}_i : i \in [\![1, 2^{n_\theta}]\!] \,|\, k \in K_i^+\right\} \tag{3.26a}$$
$$\mathcal{F}_k^- = \mathrm{hull}\left\{\mathcal{V}_i : i \in [\![1, 2^{n_\theta}]\!] \,|\, k \in K_i^-\right\} \tag{3.26b}$$

By denoting $(e_1, \ldots, e_{n_\theta})$ the standard basis of $\mathbb{R}^{n_\theta}$, it is clear that the normals to the facets of Θ are simply given by $\mathcal{N}_{\mathcal{F}_k^+} = e_k$ and $\mathcal{N}_{\mathcal{F}_k^-} = -e_k$. Moreover, for all $i \in [\![1, 2^{n_\theta}]\!]$

$$\mathrm{ind}(\mathcal{V}_i) = \left\{\mathcal{F}_k^+ : k \in [\![1, n_\theta]\!] \,|\, k \in K_i^+\right\} \cup \left\{\mathcal{F}_k^- : k \in [\![1, n_\theta]\!] \,|\, k \in K_i^-\right\} \tag{3.27}$$

Hence $\left|\det\left(\mathcal{N}_{\mathrm{ind}(\mathcal{V}_i)}\right)\right| = 1$, and

$$\begin{aligned}\prod_{\mathcal{F}\in\mathrm{ind}(\mathcal{V}_i)} \langle \mathcal{N}_{\mathcal{F}} | \mathcal{V}_i - \theta\rangle &= \left(\prod_{k\in K_i^+} \langle \mathcal{N}_{\mathcal{F}_k^+} | \mathcal{V}_i - \theta\rangle\right)\left(\prod_{k\in K_i^-} \langle \mathcal{N}_{\mathcal{F}_k^-} | \mathcal{V}_i - \theta\rangle\right) \\ &= \left(\prod_{k\in K_i^+} (\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^-} (\theta_k - \underline{\theta}_k)\right)\end{aligned} \tag{3.28}$$

which yields

$$\sum_{i=1}^{2^{n_\theta}} w_{\mathcal{V}_i}(\theta) = \sum_{i=1}^{2^{n_\theta}} \frac{1}{\left(\prod_{k\in K_i^+}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^-}(\theta_k - \underline{\theta}_k)\right)} \tag{3.29}$$

Noticing that K_i^+ and K_i^- form a partition of $[\![1, n_\theta]\!]$ for all $i \in [\![1, 2^{n_\theta}]\!]$, multiplying the numerator and denominator of each term in the sum by $\left(\prod_{k\in K_i^-}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^+}(\theta_k - \underline{\theta}_k)\right)$ provides

$$\sum_{i=1}^{2^{n_\theta}} w_{\mathcal{V}_i}(\theta) = \frac{1}{\prod_{k=1}^{n_\theta}(\overline{\theta}_k - \theta_k)(\theta_k - \underline{\theta}_k)} \sum_{i=1}^{2^{n_\theta}} \left(\prod_{k\in K_i^-}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^+}(\theta_k - \underline{\theta}_k)\right) \tag{3.30}$$

Finally, the fact that $K^+_{2^{n_\theta}+1-i} = K_i^-$ (resp. $K^-_{2^{n_\theta}+1-i} = K_i^+$) makes it possible to re-order the terms of the sum, and obtain

$$\begin{aligned}\sum_{i=1}^{2^{n_\theta}} w_{\mathcal{V}_i}(\theta) &= \frac{1}{\prod_{k=1}^{n_\theta}(\overline{\theta}_k - \theta_k)(\theta_k - \underline{\theta}_k)} \sum_{i=1}^{2^{n_\theta}} \left(\prod_{k\in K_i^+}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^-}(\theta_k - \underline{\theta}_k)\right) \\ &= \frac{1}{\prod_{k=1}^{n_\theta}(\overline{\theta}_k - \theta_k)(\theta_k - \underline{\theta}_k)} \left(\sum_{i=1}^{2^{n_\theta}} \frac{1}{w_{\mathcal{V}_i}(\theta)}\right)\end{aligned} \tag{3.31}$$

By a succession of factorizations, one has

$$\begin{aligned}\sum_{i=1}^{2^{n_\theta}} \frac{1}{w_{\mathcal{V}_i}(\theta)} &= \sum_{i=1}^{2^{n_\theta}} \left(\prod_{k\in K_i^-}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^+}(\theta_k - \underline{\theta}_k)\right) \\ &= \left[(\overline{\theta}_{n_\theta} - \theta_{n_\theta}) + (\theta_{n_\theta} - \underline{\theta}_{n_\theta})\right] \sum_{i=1}^{2^{n_\theta-1}} \left(\prod_{k\in K_i^-\setminus\{n_\theta\}}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^+\setminus\{n_\theta\}}(\theta_k - \underline{\theta}_k)\right) \\ &= \left(\overline{\theta}_{n_\theta} - \underline{\theta}_{n_\theta}\right) \sum_{i=1}^{2^{n_\theta-1}} \left(\prod_{k\in K_i^-\setminus\{n_\theta\}}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^+\setminus\{n_\theta\}}(\theta_k - \underline{\theta}_k)\right) \\ &= \dots \\ &= \left(\prod_{k=2}^{n_\theta}\left(\overline{\theta}_k - \underline{\theta}_k\right)\right)\left((\overline{\theta}_1 - \theta_1) + (\theta_1 - \underline{\theta}_1)\right) \\ &= \prod_{k=1}^{n_\theta}\left(\overline{\theta}_k - \underline{\theta}_k\right)\end{aligned} \tag{3.32}$$

Injecting the expression of $\sum_{j=1}^{2^{n_\theta}} w_{\mathcal{V}_j}(\theta)$ and $w_{\mathcal{V}_i}(\theta)$ in $h_{\mathcal{V}_i}(\theta)$ provides

$$\begin{aligned}h_{\mathcal{V}_i}(\theta) &= \frac{w_{\mathcal{V}_i}(\theta)}{\sum_{j=1}^{n_h} w_{\mathcal{V}_j}(\theta)} \\ &= \frac{1}{\prod_{k=1}^{n_\theta}\left(\overline{\theta}_k - \underline{\theta}_k\right)} \left(\prod_{k=1}^{n_\theta}(\overline{\theta}_k - \theta_k)(\theta_k - \underline{\theta}_k)\right) w_{\mathcal{V}_i}(\theta) \\ h_{\mathcal{V}_i}(\theta) &= \frac{1}{\prod_{k=1}^{n_\theta}\left(\overline{\theta}_k - \underline{\theta}_k\right)} \cdot \frac{\prod_{k=1}^{n_\theta}(\overline{\theta}_k - \theta_k)(\theta_k - \underline{\theta}_k)}{\left(\prod_{k\in K_i^+}(\overline{\theta}_k - \theta_k)\right)\left(\prod_{k\in K_i^-}(\theta_k - \underline{\theta}_k)\right)}\end{aligned} \tag{3.33}$$

and finally, leveraging again the partition of $[\![1, n_\theta]\!]$ by K_i^+ and K_i^- gives

$$\begin{aligned} h_{\mathcal{V}_i}(\theta) &= \frac{1}{\prod_{k=1}^{n_\theta} (\overline{\theta}_k - \underline{\theta}_k)} \cdot \frac{\prod_{k=1}^{n_\theta} (\overline{\theta}_k - \theta_k)(\theta_k - \underline{\theta}_k)}{\left(\prod_{k \in K_i^+} (\overline{\theta}_k - \theta_k)\right) \left(\prod_{k \in K_i^-} (\theta_k - \underline{\theta}_k)\right)} \\ &= \frac{1}{\prod_{k=1}^{n_\theta} (\overline{\theta}_k - \underline{\theta}_k)} \left(\prod_{k \in K_i^-} (\overline{\theta}_k - \theta_k)\right) \left(\prod_{k \in K_i^+} (\theta_k - \underline{\theta}_k)\right) \end{aligned} \tag{3.34}$$

which is the expression (3.18). □

The previous barycentric coordinates can be extended to all kinds of polytopes with a perturbation trick where the non-simple vertices of the polytopes are decomposed into several simple vertices. The idea of this trick is given succinctly in the Lemma 3.4.2, and the rigorous proof of the invariance of the weight functions under the infinitesimal decomposition of non-simple vertices is found in [292].

Lemma 3.4.2 (Barycentric coordinates of a polytope)**.** *Let $\theta \in \mathbb{R}^{n_\theta}$ be bounded by the polytope Θ with vertices $v(\Theta) = \{\mathcal{V}_i\}_{1 \leq i \leq n_h}$. Each non-simple vertex $\mathcal{V}_{\overline{s}(i)} \in v_{\overline{s}}(\Theta)$ is infinitesimally disturbed into m_i distinct vertices $w_i = \{\mathcal{W}_j\}_{1 \leq j \leq m_i}$ such that the polytope Θ_s given by the convex hull of $v_s(\Theta) \cup (\bigcup_{i=1}^{n_{\overline{s}}} w_i)$ is simple. For all $\mathcal{V}_{\overline{s}(i)} \in v_{\overline{s}}(\Theta)$, the weight function of $\mathcal{V}_{\overline{s}(i)}$ in Θ is given by summing the weight functions of the simple vertices w_i in Θ_s. Moreover, for all $\mathcal{V}_{s(i)} \in v_s(\Theta)$, the weight function of $\mathcal{V}_{s(i)}$ stays unchanged between Θ and Θ_s. Formally, for all $i \in [\![1, n_h]\!]$:*

$$w^{\Theta}_{\mathcal{V}_{s(i)}}(\theta) = w^{\Theta_s}_{\mathcal{V}_{s(i)}}(\theta) \tag{3.35a}$$

$$w^{\Theta}_{\mathcal{V}_{\overline{s}(i)}}(\theta) = \sum_{j=1}^{m_i} w^{\Theta_s}_{\mathcal{W}_j}(\theta) \tag{3.35b}$$

The weight functions for the simple vertices Θ_s are defined in Lemma 3.4.1. Barycentric coordinates of θ in Θ associated to the vertices $v(\Theta)$ are finally given by normalizing the previous weights, as in Lemma 3.4.1.

The following generalization of the NLSA on a polytope can finally be obtained.

Theorem 3.4.1 (NLSA on a polytope)**.** *The T-S model (2.1) with the activation functions $\mathbf{h}$ obtained using Lemma 3.4.2 is an exact representation of the nonlinear system (1.1) for all $x(t)$, $u(t)$, $w(t)$ and t such that $\theta \in \Theta = \operatorname{hull} v(\Theta)$, i.e. (3.1) holds for all $\theta \in \Theta$.*

The extension of the NLSA to polytopic bounding sets has a lot of advantages: the number of vertices needed is flexible, and can be chosen to be linearly growing ($n_h = n_\theta + 1$) in the number of scheduling parameters by bounding θ within a simplex. It also appears to be extremely useful in order to minimize the intrinsic conservatism of the T-S model when the scheduling parameters are not fully decoupled and show some interdependencies. Indeed, moving away

from the box-shaped framework, it is now possible to get rid of some useless vertices of the model earlier than in other works [233, 146], or to move some vertices around, which can easily reduce the size of the subset of Θ that does not contain admissible values of θ, hence reducing the intrinsic conservatism of the T-S model. Finally, the barycentric coordinates $\{h_{\mathcal{V}_i}\}_{1\leq i\leq n_h}$ are rational functions of the scheduling parameters, which could possibly be leveraged for some conservatism reduction, by taking inspiration from the results found in [239, 240].

However, this extension has a few weaknesses: in high dimensions, the construction of such models can be laborious, the bounding polytope being hard to determine, and the non-simple vertices challenging to handle. Moreover, the computation of the weight functions behind the barycentric coordinates of the model also involves some divisions by zero, which do not cause any theoretical problem (by a simple argument of continuity), but which have to be taken care of numerically, for example in PDC or nPDC schemes. Note that this generalization of the NLSA was already mentioned, solely for simple polytopes, in Section 2.1.2 of [295].

3.4.2 Bounding methodology

Contrary to the usual NLSA on a box-shaped bounding set, bounding interdependent nonlinearities within a small polytope Θ is a much harder problem in high dimensions. However, this problem is not new, and several techniques to find minimal bounding polytopes are already known in the mathematical and LPV literature (see [53, 9] for some reviews) and can be leveraged to this end. Generally speaking, the numerical polytope covering techniques consist in the two following steps:

- sampling the scheduling vector θ for x, u, w and t in a range of interest, possibly with random noise to increase robustness;
- computing the convex hull of all the obtained points using a dedicated algorithm (such as the well-known *quick hull* [221]).

The resulting convex hull provides a vertex-representation (V-representation) of the bounding polytope. However, in addition to the usual LPV bounding methodology, the polytopic NLSA provides an expression for the activation functions $\{h_{\mathcal{V}_i}\}_{1\leq i\leq n_h}$. For this expression to be properly determined algorithmically, one must:

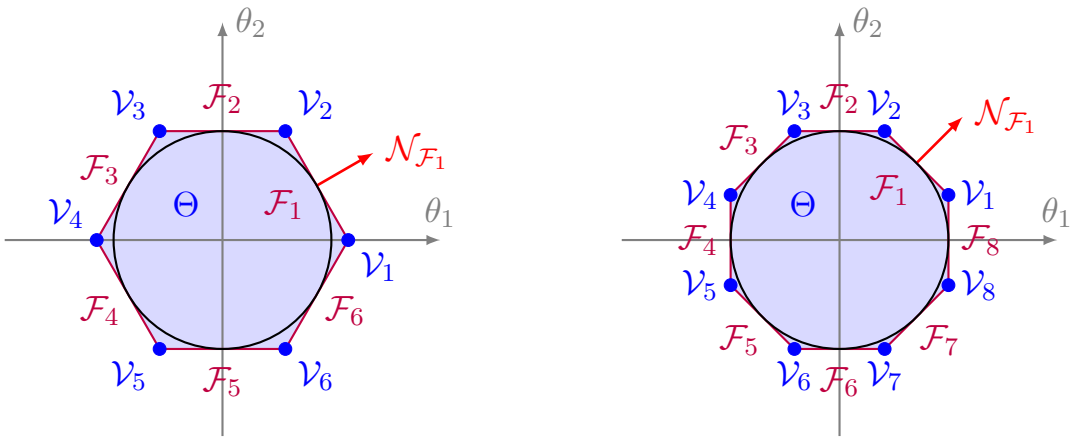
- disturb the non-simple vertices of the bounding polytope until the polytope becomes simple [145];
- compute the normals to the facets, which is equivalent to computing the half-space-representation (H-representation) of the polytope [314, 110].

Overall, the suggested bounding methodology can lead to a high number of vertices, hence polytope reduction techniques might also be needed [46, 302]. Thankfully, all these computationally heavy steps just need to be performed once. After the V- and H-representations of the bounding polytope Θ are obtained, the expressions for the activation functions $\{h_{\mathcal{V}_i}\}_{1\leq i\leq n_h}$ can be obtained analytically, and can then be evaluated in real-time, for example in order to apply PDC or nPDC control laws.

3.4.3 Application

A stability analysis is performed on the nonlinear system (3.7) of Section 3.2.2, by using two different bounding sets Θ for the scheduling vector θ of the LPV model (3.8). One of these

bounding set is chosen to be hexagonal and the other is chosen to be octagonal. As expected, it is shown that the stability results obtained using the octagonal bounding set are less conservative than those obtained using the hexagonal bounding set. Note that the hexagon is not strictly included in the square box, even if its area is lower, whereas the octagon which is strictly included in the square box.



(a) The hexagonal bounding set.

(b) The octagonal bounding set.

Figure 3.6: The sets Θ are bounding the scheduling vector θ of (3.8).

Hexagonal bounding set Considering the LPV model (3.8), the (regular) hexagon

$$\Theta = \text{hull}\,\{\mathcal{V}_1, \ldots, \mathcal{V}_6\} \tag{3.36}$$

is chosen to be the bounding set of θ. As illustrated by Figure 3.6a, Θ has six vertices:

$$\begin{aligned} \mathcal{V}_1 &= \begin{pmatrix} \frac{2}{\sqrt{3}} & 0 \end{pmatrix}^\top & \mathcal{V}_2 &= \begin{pmatrix} \frac{1}{\sqrt{3}} & 1 \end{pmatrix}^\top \\ \mathcal{V}_3 &= \begin{pmatrix} -\frac{1}{\sqrt{3}} & 1 \end{pmatrix}^\top & \mathcal{V}_4 &= \begin{pmatrix} -\frac{2}{\sqrt{3}} & 0 \end{pmatrix}^\top \\ \mathcal{V}_5 &= \begin{pmatrix} -\frac{1}{\sqrt{3}} & -1 \end{pmatrix}^\top & \mathcal{V}_6 &= \begin{pmatrix} \frac{1}{\sqrt{3}} & -1 \end{pmatrix}^\top \end{aligned} \tag{3.37}$$

This polytope has six facets which are denoted $f(\Theta) = \{\mathcal{F}_1, \ldots, \mathcal{F}_6\}$, with for all $i \in [\![1, 5]\!]$, $\mathcal{F}_i = \text{hull}\{\mathcal{V}_i, \mathcal{V}_{i+1}\}$ and $\mathcal{F}_6 = \text{hull}\{\mathcal{V}_1, \mathcal{V}_6\}$. The normals to these facets are given below.

$$\begin{aligned} \mathcal{N}_{\mathcal{F}_1} &= \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}^\top & \mathcal{N}_{\mathcal{F}_2} &= \begin{pmatrix} 0 & 1 \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_3} &= \begin{pmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}^\top & \mathcal{N}_{\mathcal{F}_4} &= \begin{pmatrix} -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_5} &= \begin{pmatrix} 0 & -1 \end{pmatrix}^\top & \mathcal{N}_{\mathcal{F}_6} &= \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}^\top \end{aligned} \tag{3.38}$$

The polytope being 2-dimensional, i.e. a polygon, it is a simple polytope. Moreover $\text{ind}(\mathcal{V}_1) = \{\mathcal{F}_6, \mathcal{F}_1\}$ and for all $i \in [\![2, 6]\!]$, $\text{ind}(\mathcal{V}_i) = \{\mathcal{F}_{i-1}, \mathcal{F}_i\}$. The weight functions, computed using

Lemma 3.4.1, are:

$$w_{\mathcal{V}_1}(\theta) = \frac{2\sqrt{3}}{3\theta_1^2 - 4\sqrt{3}\theta_1 - \theta_2^2 + 4} \tag{3.39a}$$

$$w_{\mathcal{V}_2}(\theta) = \frac{\sqrt{3}}{(\theta_2 - 1)(\sqrt{3}\theta_1 + \theta_2 - 2)} \tag{3.39b}$$

$$w_{\mathcal{V}_3}(\theta) = \frac{\sqrt{3}}{(\theta_2 - 1)(-\sqrt{3}\theta_1 + \theta_2 - 2)} \tag{3.39c}$$

$$w_{\mathcal{V}_4}(\theta) = \frac{2\sqrt{3}}{3\theta_1^2 + 4\sqrt{3}\theta_1 - \theta_2^2 + 4} \tag{3.39d}$$

$$w_{\mathcal{V}_5}(\theta) = \frac{\sqrt{3}}{(\theta_2 + 1)(\sqrt{3}\theta_1 + \theta_2 + 2)} \tag{3.39e}$$

$$w_{\mathcal{V}_6}(\theta) = \frac{\sqrt{3}}{(\theta_2 + 1)(-\sqrt{3}\theta_1 + \theta_2 + 2)} \tag{3.39f}$$

which provides the following barycentric coordinates of θ within the hexagon Θ:

$$h_{\mathcal{V}_1}(\theta) = \frac{(\theta_2^2 - 1)(3\theta_1^2 + 4\sqrt{3}\theta_1 - \theta_2^2 + 4)}{6(\theta_1^2 + \theta_2^2 - 4)} \tag{3.40a}$$

$$h_{\mathcal{V}_2}(\theta) = \frac{(\theta_2 + 1)(-\sqrt{3}\theta_1 + \theta_2 - 2)(-3\theta_1^2 + \theta_2^2 + 4\theta_2 + 4)}{12(\theta_1^2 + \theta_2^2 - 4)} \tag{3.40b}$$

$$h_{\mathcal{V}_3}(\theta) = \frac{(\theta_2 + 1)(\sqrt{3}\theta_1 + \theta_2 - 2)(-3\theta_1^2 + \theta_2^2 + 4\theta_2 + 4)}{12(\theta_1^2 + \theta_2^2 - 4)} \tag{3.40c}$$

$$h_{\mathcal{V}_4}(\theta) = \frac{(\theta_2^2 - 1)(3\theta_1^2 - 4\sqrt{3}\theta_1 - \theta_2^2 + 4)}{6(\theta_1^2 + \theta_2^2 - 4)} \tag{3.40d}$$

$$h_{\mathcal{V}_5}(\theta) = \frac{(\theta_2 - 1)(\sqrt{3}\theta_1 - \theta_2 - 2)(3\theta_1^2 - \theta_2^2 + 4\theta_2 - 4)}{12(\theta_1^2 + \theta_2^2 - 4)} \tag{3.40e}$$

$$h_{\mathcal{V}_6}(\theta) = \frac{(\theta_2 - 1)(-\sqrt{3}\theta_1 - \theta_2 - 2)(3\theta_1^2 - \theta_2^2 + 4\theta_2 - 4)}{12(\theta_1^2 + \theta_2^2 - 4)} \tag{3.40f}$$

Finally, the NLSA on the polytope Θ is completed and the system (3.7) is exactly represented by the following T-S model:

$$\dot{x}(t) = \sum_{i=1}^{6} h_{\mathcal{V}_i}(\theta) A(\mathcal{V}_i) x(t) \tag{3.41}$$

The stability analysis of the T-S model (3.41) is performed using Theorem 2.3.1 at several $(\alpha, \beta) \in [-1, 1]^2$ values. The (α, β)-region for which the system (3.41) (hexagonal model) is found to be exponentially stable is plotted on Figure 3.7, where it is compared to the (α, β)-region obtained with the T-S model (3.21) (box model), and with the perturbative approach (Proposition 3.2.1). The (α, β)-region of stability is visibly larger by using Theorem 2.3.1 on the T-S model (3.41) (hexagonal model) than on the T-S model (3.21) (box model): indeed, the hexagonal bounding set is sharper than the box-shaped bounding set, which relaxes the conservatism of the stability analysis. However, as the hexagonal bounding set is not strictly included in the box bounding set, the (α, β)-region of stability for the box model is also not strictly included in the (α, β)-region of stability for the hexagonal model.

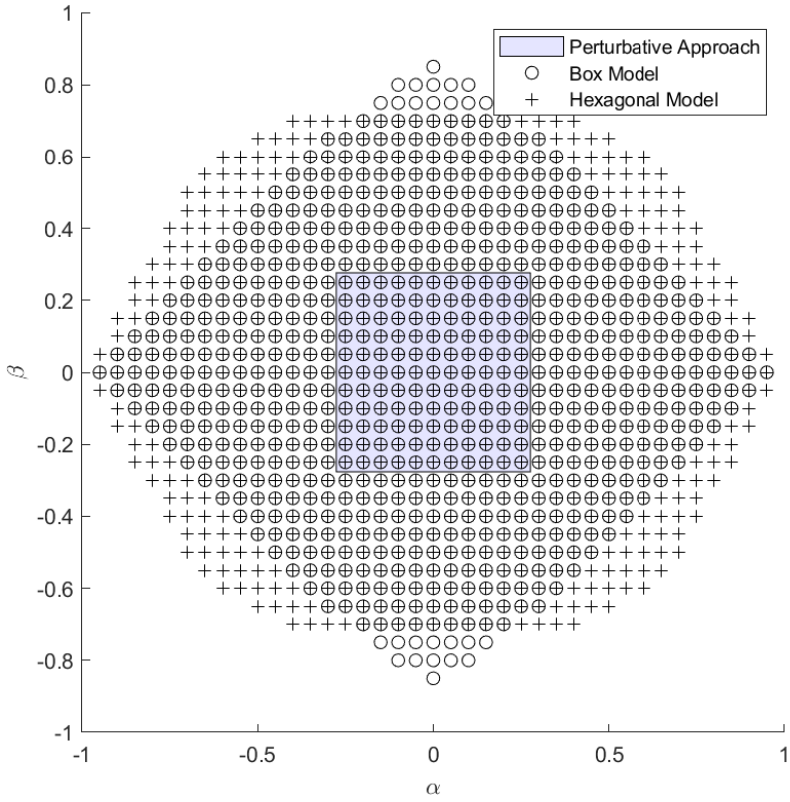


Figure 3.7: Stability (α, β)-regions of (3.6) using Theorem 2.3.1 on the T-S models (3.41) (*hexagonal model*), (3.21) (*box model*), and using the perturbative approach (*Proposition 3.2.1*).

Octagonal bounding set Similarly, the (regular) octagon $\Theta = \text{hull}\{\mathcal{V}_1, \dots, \mathcal{V}_8\}$ is chosen to be the bounding set of θ. As illustrated by Figure 3.6b, Θ has eight vertices:

$$\begin{aligned} \mathcal{V}_1 &= \begin{pmatrix} 1 & \sqrt{2}-1 \end{pmatrix}^\top & \mathcal{V}_2 &= \begin{pmatrix} \sqrt{2}-1 & 1 \end{pmatrix}^\top \\ \mathcal{V}_3 &= \begin{pmatrix} -\sqrt{2}+1 & 1 \end{pmatrix}^\top & \mathcal{V}_4 &= \begin{pmatrix} -1 & \sqrt{2}-1 \end{pmatrix}^\top \\ \mathcal{V}_5 &= \begin{pmatrix} -1 & -\sqrt{2}+1 \end{pmatrix}^\top & \mathcal{V}_6 &= \begin{pmatrix} -\sqrt{2}+1 & -1 \end{pmatrix}^\top \\ \mathcal{V}_7 &= \begin{pmatrix} \sqrt{2}-1 & -1 \end{pmatrix}^\top & \mathcal{V}_8 &= \begin{pmatrix} 1 & -\sqrt{2}+1 \end{pmatrix}^\top \end{aligned} \tag{3.42}$$

This polytope has eight facets which are denoted $f(\Theta) = \{\mathcal{F}_1, \dots, \mathcal{F}_8\}$, with for all $i \in [\![1, 7]\!]$, $\mathcal{F}_i = \text{hull}\{\mathcal{V}_i, \mathcal{V}_{i+1}\}$ and $\mathcal{F}_8 = \text{hull}\{\mathcal{V}_1, \mathcal{V}_8\}$. The normals to these facets are given below.

$$\begin{aligned} \mathcal{N}_{\mathcal{F}_1} &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^\top & \mathcal{N}_{\mathcal{F}_2} &= \begin{pmatrix} 0 & 1 \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_3} &= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^\top & \mathcal{N}_{\mathcal{F}_4} &= \begin{pmatrix} -1 & 0 \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_5} &= \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}^\top & \mathcal{N}_{\mathcal{F}_6} &= \begin{pmatrix} 0 & -1 \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_7} &= \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}^\top & \mathcal{N}_{\mathcal{F}_8} &= \begin{pmatrix} 1 & 0 \end{pmatrix}^\top \end{aligned} \tag{3.43}$$

Again, the polytope being 2-dimensional, i.e. a polygon, it is a simple polytope. Moreover $\text{ind}(\mathcal{V}_1) = \{\mathcal{F}_8, \mathcal{F}_1\}$ and for all $i \in [\![2, 8]\!]$, $\text{ind}(\mathcal{V}_i) = \{\mathcal{F}_{i-1}, \mathcal{F}_i\}$. The weight functions, computed

using Lemma 3.4.1, are:

$$\begin{aligned} w_{\mathcal{V}_1}(\theta) &= \frac{1}{(1-\theta_1)(\sqrt{2}-\theta_1-\theta_2)} & w_{\mathcal{V}_2}(\theta) &= \frac{1}{(1-\theta_2)(\sqrt{2}-\theta_1-\theta_2)} \\ w_{\mathcal{V}_3}(\theta) &= \frac{1}{(1-\theta_2)(\sqrt{2}+\theta_1-\theta_2)} & w_{\mathcal{V}_4}(\theta) &= \frac{1}{(1+\theta_1)(\sqrt{2}+\theta_1-\theta_2)} \\ w_{\mathcal{V}_5}(\theta) &= \frac{1}{(1+\theta_1)(\sqrt{2}+\theta_1+\theta_2)} & w_{\mathcal{V}_6}(\theta) &= \frac{1}{(1+\theta_2)(\sqrt{2}+\theta_1+\theta_2)} \\ w_{\mathcal{V}_7}(\theta) &= \frac{1}{(1+\theta_2)(\sqrt{2}-\theta_1+\theta_2)} & w_{\mathcal{V}_8}(\theta) &= \frac{1}{(1-\theta_1)(\sqrt{2}-\theta_1+\theta_2)} \end{aligned} \tag{3.44}$$

which provides the following barycentric coordinates of θ within the octagon Θ:

$$h_{\mathcal{V}_1}(\theta) = \frac{(\theta_2^2-1)(\theta_1+1)(\theta_1+\theta_2+\sqrt{2})(-(\theta_1-\theta_2)^2+2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45a}$$

$$h_{\mathcal{V}_2}(\theta) = \frac{(\theta_1^2-1)(\theta_2+1)(\theta_1+\theta_2+\sqrt{2})(-(\theta_1-\theta_2)^2+2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45b}$$

$$h_{\mathcal{V}_3}(\theta) = \frac{(\theta_1^2-1)(\theta_2+1)(\theta_1-\theta_2-\sqrt{2})((\theta_1+\theta_2)^2-2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45c}$$

$$h_{\mathcal{V}_4}(\theta) = \frac{(\theta_2^2-1)(\theta_1-1)(-\theta_1+\theta_2+\sqrt{2})((\theta_1+\theta_2)^2-2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45d}$$

$$h_{\mathcal{V}_5}(\theta) = \frac{(\theta_2^2-1)(\theta_1-1)(\theta_1+\theta_2-\sqrt{2})(-(\theta_1-\theta_2)^2+2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45e}$$

$$h_{\mathcal{V}_6}(\theta) = \frac{(\theta_1^2-1)(\theta_2-1)(\theta_1+\theta_2-\sqrt{2})(-(\theta_1-\theta_2)^2+2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45f}$$

$$h_{\mathcal{V}_7}(\theta) = \frac{(\theta_1^2-1)(\theta_2-1)(\theta_1-\theta_2+\sqrt{2})((\theta_1+\theta_2)^2-2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45g}$$

$$h_{\mathcal{V}_8}(\theta) = \frac{(\theta_2^2-1)(\theta_1+1)(-\theta_1+\theta_2-\sqrt{2})((\theta_1+\theta_2)^2-2)}{4(\theta_1^2+\theta_2^2-2)(2\sqrt{2}+(1-\sqrt{2})\theta_1^2+(1-\sqrt{2})\theta_2^2)} \tag{3.45h}$$

Finally, the NLSA on the polytope Θ is completed and the system (3.7) is exactly represented by the following T-S model:

$$\dot{x}(t) = \sum_{i=1}^{8} h_{\mathcal{V}_i}(\theta) A(\mathcal{V}_i) x(t) \tag{3.46}$$

The stability analysis of the T-S model (3.46) is performed using Theorem 2.3.1 at several $(\alpha, \beta) \in [-1, 1]^2$ values. The (α, β)-region for which the system (3.46) (octagonal model) is found to be exponentially stable is plotted on Figure 3.8, where it is compared to the (α, β)-region obtained with the T-S models (3.41) (hexagonal model), (3.21) (box model), and with the perturbative approach (Proposition 3.2.1). The (α, β)-region of stability is visibly larger by using Theorem 2.3.1 on the T-S model (3.46) (octagonal model) than on the two other T-S models: indeed, the octagonal bounding set is sharper than the box-shaped and hexagonal bounding sets, which relaxes the conservatism of the stability analysis. The octagonal bounding set is strictly included in the box bounding set, hence the (α, β)-region of stability for the box model is also strictly included in the (α, β)-region of stability for the octagonal model.

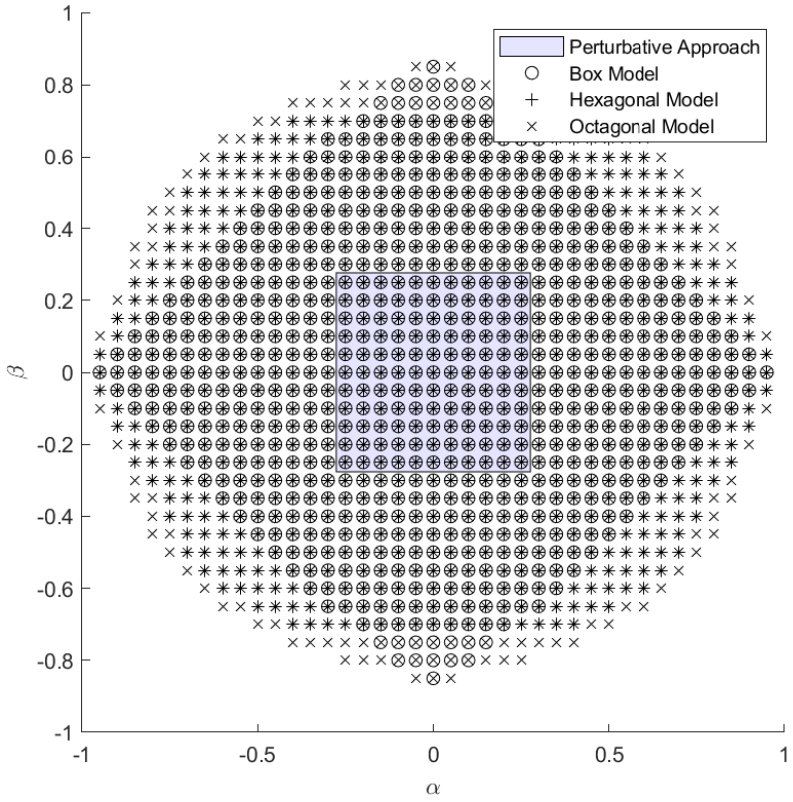


Figure 3.8: Stability (α, β)-regions of (3.6) using Theorem 2.3.1 on the T-S models (3.46) (*octagonal model*), (3.41) (*hexagonal model*), (3.21) (*box model*), and using the perturbative approach (*Proposition 3.2.1*).

In order to show that using a more sophisticated Lyapunov function does not change which model outperforms the others, the PQLF (2.40) of Chapter 2 is now used in the stability analysis of the three T-S models. Using $k = n_h$, $\tau_{i,i,k} = 0$ and $\tau_{i,j,k} = 1$ $(i \neq j)$ for all models in the LMI conditions found in Theorem 2.3.3 of Chapter 2, the (α, β)-regions for which the T-S models (3.46) (octagonal model), (3.41) (hexagonal model) and (3.21) (box model) are found to be stable are plotted on Figure 3.9. As expected, all the (α, β)-regions of stability are larger than on Figure 3.8, and the octagonal model still outperforms the hexagonal model, this latter also outperforming the box model.

So far, the conservatism reduction has only been achieved by accumulating more and more local models in the T-S representation. This is due to the nonlinearities of the toy model which are shaped like a circle, and this accumulation is not an essential feature of the polytopic NLSA. Indeed, applying the polytopic NLSA to the T-S model described in [233] after its useless vertices have been removed can reduce the number of local models from $n_h = 2^{n_\theta}$ to $n_h = n_\theta + 1$. As shown in [233], getting rid of the useless vertices of this system reduces the conservatism of the controller design problem.

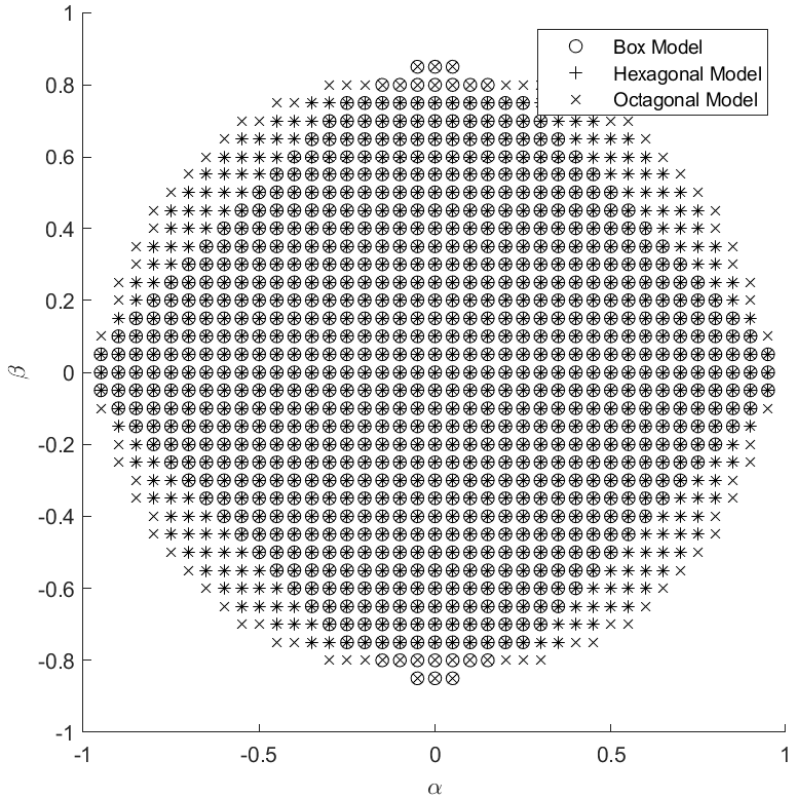


Figure 3.9: Stability (α, β)-regions of (3.6) using the PQLF of [61] (Theorem 2.3.3 of Chapter 2) on the T-S models (3.46) (*octagonal model*), (3.41) (*hexagonal model*), (3.21) (*box model*).

3.5 The nonlinear sector approach for smooth convex sets

This section introduces a new NLSA for smooth convex sets, where the scheduling vector θ is bounded within a convex set Θ with a smooth boundary $\mathcal{S}_\Theta \triangleq \partial\Theta$. Lemma 3.5.1 provides barycentric coordinates of θ inside Θ. Following the procedure of the NLSA, Theorem 3.5.1 uses these barycentric coordinates as weighting functions in order to obtain a new kind of T-S model, where the discrete sum is replaced by an integral along $\mathcal{S}_\Theta$, and which exactly represents (1.1) for all $\theta \in \Theta$. After discussing the advantages and inconveniences of the NLSA on a smooth convex set, basic results of stability for the newly introduced T-S-like models are provided, followed by an application of these results to the nonlinear system (3.7) of Section 3.2.2.

3.5.1 Barycentric coordinates

An expression for the barycentric coordinates of a smooth convex set is reported thereafter.

Lemma 3.5.1 (Barycentric coordinates of a smooth convex set)**.** *Let $\theta \in \mathbb{R}^{n_\theta}$ be bounded by the smooth and bounded convex set Θ whose boundary is a $(n_\theta - 1)$-dimensional smooth manifold $\mathcal{S}_\Theta \triangleq \partial\Theta$. For all $\mathcal{V} \in \mathcal{S}_\Theta$, $w_\mathcal{V}$ stands for the weight function associated to $\mathcal{V}$, with*

$$w_\mathcal{V}(\theta) = \frac{\kappa(\mathcal{V})}{\langle \mathcal{N}_\mathcal{V} | \mathcal{V} - \theta \rangle^{n_\theta}} \tag{3.47}$$

where $\kappa(\mathcal{V})$ represents the Gaussian curvature of $\mathcal{S}_\Theta$ at $\mathcal{V}$ and $\mathcal{N}_\mathcal{V}$ stands for the exterior-pointing normal to the supporting hyperplane of S at $\mathcal{V}$.

Barycentric coordinates of θ in Θ are given by the functions $\{h_\mathcal{V}\}_{\mathcal{V}\in\mathcal{S}_\Theta}$, obtained by a normalization of the previous weights, and defined for all $\mathcal{V} \in \mathcal{S}_\Theta$ by

$$h_\mathcal{V}(\theta) = \frac{w_\mathcal{V}(\theta)}{\int\limits_{\mathcal{W}\in\mathcal{S}_\Theta} w_\mathcal{W}(\theta) d\mathcal{S}_\Theta} \tag{3.48}$$

Proof. The three axioms for barycentric coordinates (2.2a), (2.2b), (3.2) are verified under their integral form, which can be found in [294]. The proof of the linear precision property can be found in [294] as well, and the two other properties are trivial by construction. □

The barycentric coordinates introduced hereabove rely on the expression of the Gaussian curvature of a manifold. An explicit expression of this curvature is provided below.

Lemma 3.5.2 (Expression of the Gaussian curvature [107])**.** *If there exists a smooth function $f : \mathbb{R}^{n_\theta} \to \mathbb{R}$ such that*

$$\mathcal{S}_\Theta = \{\mathcal{V} \in \mathbb{R}^{n_\theta} : f(\mathcal{V}) = 0\} \tag{3.49}$$

then the Gaussian curvature of $\mathcal{S}_\Theta$ at $\mathcal{V}$ can be expressed by

$$\kappa(\mathcal{V}) = \frac{\nabla f(\mathcal{V}) H f^\top(\mathcal{V}) \nabla f^\top(\mathcal{V})}{\|\nabla f(\mathcal{V})\|^{n_\theta+1}} \tag{3.50}$$

where $\nabla f(\mathcal{V})$ and $H f(\mathcal{V})$ stand resp. for the gradient and for the Hessian of f evaluated at $\mathcal{V}$.

Keeping the notations of Lemma 3.5.1, the following generalization of the NLSA on a smooth convex set can be stated.

Theorem 3.5.1 (NLSA on a smooth convex set)**.** *The T-S-like model (3.51) with the activation functions $\{h_\mathcal{V}\}_{\mathcal{V}\in\mathcal{S}_\theta}$ obtained using Lemma 3.5.1 is an exact representation of the nonlinear system (1.1) for all $x(t)$, $u(t)$, $w(t)$ and t such that $\theta \in \Theta = \mathrm{hull}(\mathcal{S}_\theta)$.*

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = \int\limits_{\mathcal{V}\in\mathcal{S}_\Theta} h_\mathcal{V}(\theta) \begin{pmatrix} A(\mathcal{V}) & B_1(\mathcal{V}) & B_2(\mathcal{V}) \\ C_1(\mathcal{V}) & D_{11}(\mathcal{V}) & D_{12}(\mathcal{V}) \\ C_2(\mathcal{V}) & D_{21}(\mathcal{V}) & D_{22}(\mathcal{V}) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} d\mathcal{S}_\Theta \tag{3.51}$$

This NLSA approach and its resulting T-S-like model have never been studied in the literature before, and there are still no known method to extract computable LMI conditions from the stability analysis of such systems. If this T-S-like model is geometrically one of the sharpest

convex representation of a nonlinear system presenting *smoothly interdependent* scheduling parameters, this is counterbalanced by the number of vertices $\mathcal{V} \in \mathcal{S}_\Theta$ which is infinite and uncountable, leading to tricky stability and stabilization analyzes. Some elementary results are given below in order to deal with the stability analysis of this T-S-like model using a QLF. Moreover, bounding interdependent nonlinearities within a smooth convex set Θ is a difficult problem in general with no systematic solution. However, this problem is well-studied if the class of smooth convex sets is restricted to ellipsoids, and it simply consists in finding the minimum volume ellipsoid which covers all the obtainable scheduling parameters θ (eventually after sampling them): the reader is referred to Section 8.4 of [49] for more details. Similarly to what was stated at the end of the Section 3.4.1, the computation of the continuous weight functions of the model involves some divisions by zero which have to be taken care of numerically, for example in PDC or nPDC schemes.

3.5.2 Stability

Introducing a QLF to perform the stability analysis of the continuous-time and input free T-S-like model (3.52) obtained via Theorem 3.5.1 is immediately problematic.

$$\dot{x}(t) = \int_{\mathcal{V}\in\mathcal{S}_\Theta} h_\mathcal{V}(\theta) A(\mathcal{V}) x(t) d\mathcal{S}_\Theta \tag{3.52}$$

Indeed, the result stated below, which is the straightforward generalization of the Theorem 2.3.1 of Chapter 2, contains an infinite number of LMI conditions to be satisfied.

Theorem 3.5.2 (Quadratic Stability). *The T-S-like model (3.52) is globally exponentially stable if there exists $P \in \mathbb{S}_{n_x}(\mathbb{R})$ such that the conditions (3.53) are satisfied.*

$$P \succ 0 \tag{3.53a}$$
$$\mathcal{H}(PA(\mathcal{V})) \prec 0, \quad \forall \mathcal{V} \in \mathcal{S}_\Theta \tag{3.53b}$$

Proof. By convexity, the proof of this result is similar to the proof of Theorem 2.3.1. □

In particular, if Θ is an ellipsoid, i.e. $\Theta = \mathcal{E}(Q)$ with $\mathcal{E}(Q)$ defined in (3.54), there is a reason to believe that the problem of finding such a P is Nondeterministic Polynomial (NP)-hard.

$$\mathcal{E}(Q) = \left\{ \theta \in \mathbb{R}^{n_\theta} : \theta^\top Q \theta \leq 1 \right\} \text{ with } Q \in \mathbb{S}^{++}_{n_\theta}(\mathbb{R}) \tag{3.54}$$

Indeed, given $P \in \mathbb{S}^{++}_{n_x}(\mathbb{R})$ and Θ an ellipsoid, simply checking whether the conditions (3.53b) with a non-strict constraint ($\preceq$) are satisfied or not can be turned into an NP-complete problem.

Theorem 3.5.3 (NP-hardness). *Let $\Theta = \mathcal{E}(Q)$ where $\mathcal{E}(Q)$ is the ellipsoid defined in (3.54). Finding $P \in \mathbb{S}^{++}_{n_x}(\mathbb{R})$ such that the LMI conditions (3.53b) with a non-strict constraint ($\preceq$) are satisfied is an NP-hard problem.*

Proof. $\mathcal{V} \in \mathcal{S}_\Theta$ is equivalent to the existence of $v \in \mathbb{R}^{n_\theta}$ such that

$$\begin{cases} v^\top v = 1 \\ \mathcal{V} = Q^{-\frac{1}{2}} v \end{cases} \tag{3.55}$$

By construction of the T-S model (see (3.4)), for all $\mathcal{V} \in \mathcal{S}_\Theta$

$$A(\mathcal{V}) = A_0 + \sum_{i=1}^{n_\theta} \left(Q^{-\frac{1}{2}} v \right)_{(i)} A_i \tag{3.56}$$

hence:

$$A(\mathcal{V}) = A_0 + \sum_{i=1}^{n_\theta} \left(\sum_{j=1}^{n_\theta} Q^{-\frac{1}{2}}_{(i,j)} v_{(j)} \right) A_i \tag{3.57}$$

which can be rewritten as

$$A(\mathcal{V}) = A_0 + \sum_{j=1}^{n_\theta} v_{(j)} R_j \tag{3.58}$$

with $R_j = \sum_{i=1}^{n_\theta} Q^{-\frac{1}{2}}_{(i,j)} A_i$. Hence, the LMI condition (3.53b) with a non-strict constraint ($\preceq$) is equivalent to the following condition:

$$\mathcal{H}\left(P \left(A_0 + \sum_{j=1}^{n_\theta} v_{(j)} R_j \right) \right) \preceq 0, \quad \forall v \in \mathbb{R}^{n_\theta} : v^\top v = 1 \tag{3.59}$$

Checking the conditions above for a given $P \in \mathbb{S}^{++}_{n_x}(\mathbb{R})$ can be turned into NP-complete problem, as shown in Section 3.4.1 of [27], hence, finding P satisfying these conditions is an NP-hard problem. □

Remark 3.5.1 (Erratum)**.** *In the published article from which this chapter was written, the result was stated with a strict constraint instead of a non-strict one, leading to the much more pessimistic conclusion that the LMI conditions* (3.53) *are condemned to remain intractable in general [20]. The author does not believe in this pessimistic conclusion anymore, as $\mathbb{S}^{-}_{n}(\mathbb{R})$ not being an open set seems crucial to the result above, whereas $\mathbb{S}^{--}_{n}(\mathbb{R})$ is an open set. In fact, the author now conjectures that there exists a sufficiently precise discretization of the conditions* (3.53b) *so this finite set of LMI conditions are equivalent to* (3.53b)*, in the spirit of [238].*

In spite of this pessimistic result, two tractable optimization problems are given to perform the stability analysis of the T-S-like model (3.52) when Θ is bounded by an ellipsoid:

– Theorem 3.5.4 introduces some conservatism in (3.53b), leading to tractable LMI conditions;

– Theorem 3.5.5 relies on the structure of the state matrix, where it is assumed that the scheduling vector θ acts on a single column or row of $A(\theta)$.

Theorem 3.5.4 (The universal conservative conditions). *Let $\Theta = \mathcal{E}(Q)$ where $\mathcal{E}(Q)$ is the ellipsoid defined in (3.54). By construction of the T-S-like model (3.52), for all $\theta \in \mathcal{E}(Q)$, $A(\theta)$ can be rewritten as*

$$A(\theta) = A_0 + \sum_{k=1}^{n_\theta} v_{(k)}(\theta) R_k \tag{3.60}$$

with $v(\theta) \in \mathbb{R}^{n_\theta}$ such that $v^\top(\theta)v(\theta) \leq 1$ and $A_0, R_1, \ldots, R_{n_\theta} \in \mathbb{R}^{n_x \times n_x}$. The T-S-like model (3.52) is globally exponentially stable if there exists $P \in \mathbb{S}_{n_x}(\mathbb{R})$ satisfying the LMI conditions (3.61).

$$P \succ 0 \tag{3.61a}$$

$$\begin{pmatrix} \mathcal{H}(PA_0) & \mathcal{H}(PR_1) & \mathcal{H}(PR_2) & \ldots & \mathcal{H}(PR_{n_\theta}) \\ \mathcal{H}(PR_1) & \mathcal{H}(PA_0) & 0 & \ldots & 0 \\ \mathcal{H}(PR_2) & 0 & \mathcal{H}(PA_0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \mathcal{H}(PR_{n_\theta}) & 0 & \ldots & 0 & \mathcal{H}(PA_0) \end{pmatrix} \prec 0 \tag{3.61b}$$

In particular, such a P also satisfies (3.53).

Theorem 3.5.5 (The rank 2 conditions). *It is assumed that the scheduling vector θ only acts on the l-th column of $A(\theta)$, with $l \in [\![1, n_x]\!]$. Let $\Theta = \mathcal{E}(Q)$ where $\mathcal{E}(Q)$ is the ellipsoid defined in (3.54). By construction of the T-S model, for all $\theta \in \mathcal{E}(Q)$, $A(\theta)$ can be rewritten as*

$$A(\theta) = A_0 + \sum_{k=1}^{n_\theta} v_{(k)}(\theta) r_k \tilde{e}_l^\top \tag{3.62}$$

with $\tilde{e}_l$ the l-th column of I_{n_x}, $v(\theta) \in \mathbb{R}^{n_\theta}$ such that $v^\top(\theta)v(\theta) \leq 1$, and $R \triangleq \begin{bmatrix} r_1 & \ldots & r_{n_\theta} \end{bmatrix} \in \mathbb{R}^{n_x \times n_\theta}$. The T-S-like model (3.52) is globally exponentially stable if there exists $P \in \mathbb{S}_{n_x}(\mathbb{R})$ and $\lambda \in \mathbb{R}_{\geq 0}$ satisfying the LMI conditions (3.63).

$$P \succ 0 \tag{3.63a}$$

$$\begin{pmatrix} \mathcal{H}(PA_0) + \lambda \tilde{e}_l \tilde{e}_l^\top & PR \\ R^\top P & -\lambda I_{n_x} \end{pmatrix} \prec 0 \tag{3.63b}$$

These LMI conditions can be verified if and only if there exists $P \in \mathbb{S}_{n_x}(\mathbb{R})$ satisfying (3.53).

Proof. The universal conditions are obtained by applying Theorem 3.4 of [27]. The rank 2 conditions are demonstrated as follows. For all $\theta \in \mathcal{E}(Q)$, there exists $v(\theta) \in \mathbb{R}^{n_\theta}$ such that $\|v(\theta)\|_2 \leq 1$ and

$$A(\theta) = A_0 + Rv(\theta)\tilde{e}_l^\top \tag{3.64}$$

hence, when the scheduling vector θ acts on a single column of $A(\theta)$, (3.52) can be interpreted as a particular norm bound LDI system, and the LMI conditions demonstrating the stability of such systems are given in Section 5.1 of [48]. In this peculiar case, the equivalence between the LMI conditions (3.63) and (3.53) is found in Proposition 3.1 of [27]. Unsurprisingly, this equivalence is a result of the strict homogeneous S-procedure (Lemma 2.2.2 of Chapter 2). □

Corollary 3.5.1. *The LMI conditions (3.53) also demonstrate the stability of the T-S-like model*

$$\dot{x}(t) = \int_{\mathcal{V}\in\mathcal{S}_\Theta} h_{\mathcal{V}}(\theta) A^\top(\mathcal{V}) x(t) d\mathcal{S}_\Theta \tag{3.65}$$

hence, it is also possible to use the LMI conditions (3.63) for T-S-like models where the scheduling vector θ acts on a single row of $A(\theta)$, instead of on a single column.

Proof. Succinctly, if there exists P such that the LMI conditions (3.53) hold, then, multiplying these conditions left and right by P^{-1}, the LMI conditions (3.66) also hold by congruence (Property 2.2.2 of Chapter 2).

$$P^{-1} \succ 0 \tag{3.66a}$$

$$\forall \mathcal{V} \in \mathcal{S}_\Theta,\ \mathcal{H}(A(\mathcal{V})P^{-1}) \prec 0 \tag{3.66b}$$

Hence there exists $P^{-1} \succ 0$ such that for all $\mathcal{V} \in \mathcal{S}_\Theta$, $\mathcal{H}(P^{-1}A^\top(\mathcal{V})) \prec 0$, which are the conditions of exponential stability (3.53) applied to (3.65). □

Remark 3.5.2. *The name of Theorems 3.5.4 and 3.5.5 are taken from [27].*

Leaving the quadratic framework, the natural generalization of the MQLF for T-S-like systems is given by

$$V(x) = x^\top \left(\int_{\mathcal{V}\in\mathcal{S}_\Theta} h_{\mathcal{V}}(\theta) P(\mathcal{V}) d\mathcal{S}_\Theta \right) x \tag{3.67}$$

with for all $\mathcal{V} \in \mathcal{S}_\Theta$, $P(\mathcal{V}) \in \mathbb{S}^{++}_{n_x}(\mathbb{R})$. Similarly, the natural generalizations of the PDC and nPDC control laws are given by

$$u(t) = \left(\int_{\mathcal{V}\in\mathcal{S}_\Theta} h_{\mathcal{V}}(\theta) K(\mathcal{V}) d\mathcal{S}_\Theta \right) x(t) \tag{3.68a}$$

$$u(t) = \left(\int_{\mathcal{V}\in\mathcal{S}_\Theta} h_{\mathcal{V}}(\theta) K(\mathcal{V}) d\mathcal{S}_\Theta \right) \left(\int_{\mathcal{V}\in\mathcal{S}_\Theta} h_{\mathcal{V}}(\theta) P(\mathcal{V}) d\mathcal{S}_\Theta \right)^{-1} x(t) \tag{3.68b}$$

The question of whether there exist tractable optimization problems to check the stability and the stabilization of T-S-like models by leveraging these expressions remains open for future works.

3.5.3 Application

Considering the LPV model (3.8), the disc $\Theta = \left\{\theta \in \mathbb{R}^2 : \theta_1^2 + \theta_2^2 \leq 1\right\}$ is chosen to be the bounding set of θ, as illustrated by Figure 3.10. The Gaussian curvature of a circle is constant and given by the inverse of its radius, and for all $\mathcal{V} \in \mathcal{S}_\Theta$, $\mathcal{N}_\mathcal{V} = \mathcal{V}$. The weight functions, computed using Lemma 3.5.1, are:

$$\begin{aligned} w_{\mathcal{V}}(\theta) =& \frac{1}{\left(\mathcal{V}_{(1)}(\mathcal{V}_{(1)} - \theta_1) + \mathcal{V}_{(2)}(\mathcal{V}_{(2)} - \theta_2)\right)^2} \\ =& \frac{1}{\left(1 - \mathcal{V}_{(1)}\theta_1 - \mathcal{V}_{(2)}\theta_2\right)^2} \end{aligned} \tag{3.69}$$

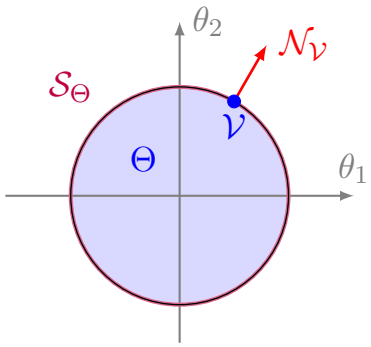


Figure 3.10: The disc Θ is bounding the scheduling vector θ of (3.8).

moreover, the following equality is verified [294]:

$$\int_{\mathcal{V}\in\mathcal{S}_\Theta} w_\mathcal{V}(\theta) d\mathcal{S}_\Theta = \frac{2\pi}{(1-\theta_1^2-\theta_2^2)^{3/2}} \tag{3.70}$$

which provides the following barycentric coordinates of θ within the disc Θ:

$$h_\mathcal{V}(\theta) = \frac{(1-\theta_1^2-\theta_2^2)^{3/2}}{2\pi\left(1-\mathcal{V}_{(1)}\theta_1-\mathcal{V}_{(2)}\theta_2\right)^2} \tag{3.71}$$

Finally, the NLSA on the disc Θ is completed and the system (3.7) is exactly represented by the following T-S-like model:

$$\dot{x}(t) = \int_{\mathcal{V}\in\mathcal{S}_\Theta} h_\mathcal{V}(\theta) A(\mathcal{V}) x(t) d\mathcal{S}_\Theta \tag{3.72}$$

The stability analysis of the T-S-like model (3.72) is performed using Theorem 3.5.4 and Theorem 3.5.5 (which is applicable since θ acts on a single row of $A(\theta)$ in (3.8)) at several $(\alpha,\beta) \in [-1,1]^2$ values. The (α,β)-regions for which the system (3.72) (disc model) is found to be exponentially stable are plotted on Figure 3.11, where they are compared to the (α,β)-regions obtained with the hexagonal T-S model (3.41) and the perturbative approach (Proposition 3.2.1). The (α,β)-region of stability is visibly larger by using Theorem 3.5.5 (rank 2 conditions) instead of Theorem 3.5.4 (universal conditions) on the T-S-like model (3.72), which confirms that the rank 2 LMI conditions are less conservative than the universal LMI conditions. Moreover, the (α,β)-region of stability for the hexagonal T-S model (3.41) is about the same size as the region computed on the T-S-like model (3.72) with Theorem 3.5.4, which indicates that, to some degree, the sharpness of the smooth convex shape has compensated the conservatism introduced by the universal LMI conditions.

However, when these universal LMI conditions are compared to the stability results obtained using the T-S model (3.46) (octagonal model) on Figure 3.12, the conservatism introduced in the stability analysis is not compensated by the sharpness of the smooth convex shape anymore. Nevertheless, in all cases, the rank 2 conditions outperform the others.

3.6 Conclusions and perspectives

The NLSA is a way to construct T-S models which exactly represent nonlinear systems whose nonlinearities are bounded by a box-shaped set. This chapter has generalized the NLSA for

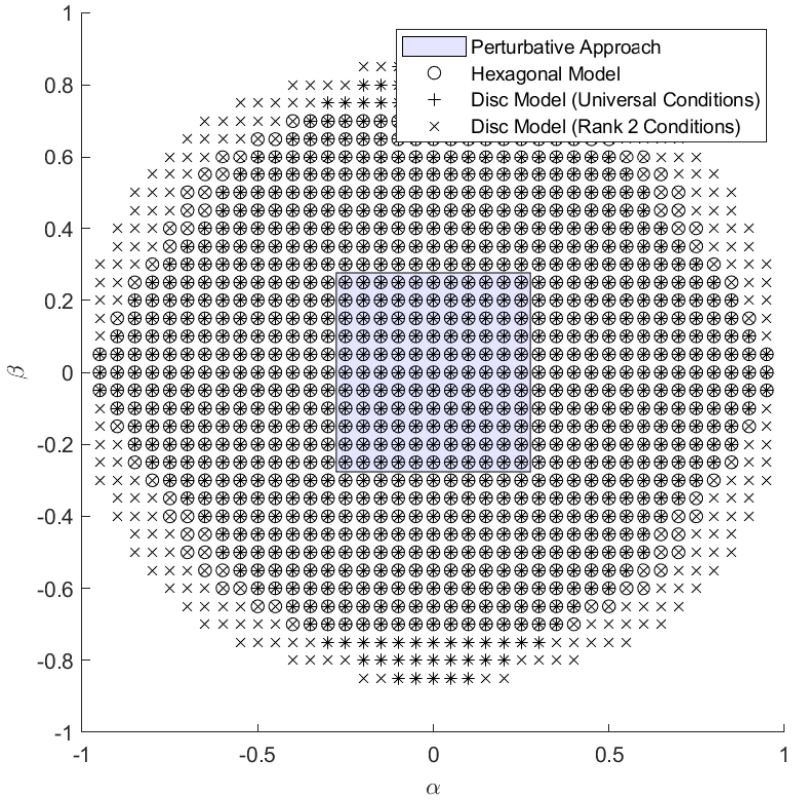


Figure 3.11: Stability (α, β)-regions of (3.6) using Theorem 3.5.4 (*universal conditions*) and Theorem 3.5.5 (*rank 2 conditions*) on the T-S-like model (3.72) (*disc model*), using Theorem 2.3.1 on the T-S model (3.41) (*hexagonal model*), and using the perturbative approach (*Proposition 3.2.1*).

larger classes of convex bounding sets: extending it for polytopic and smooth convex sets. These generalizations provide new ways of reducing the intrinsic conservatism of T-S representations with interdependent scheduling parameters, which has been illustrated numerically through the study of simple LMI criteria for stability analysis of the resulting models. In particular, it has been shown that the LMI criteria for the newly introduced T-S-like models are still conservative, and the problem of stability and stabilization for these models is left open for further investigations.

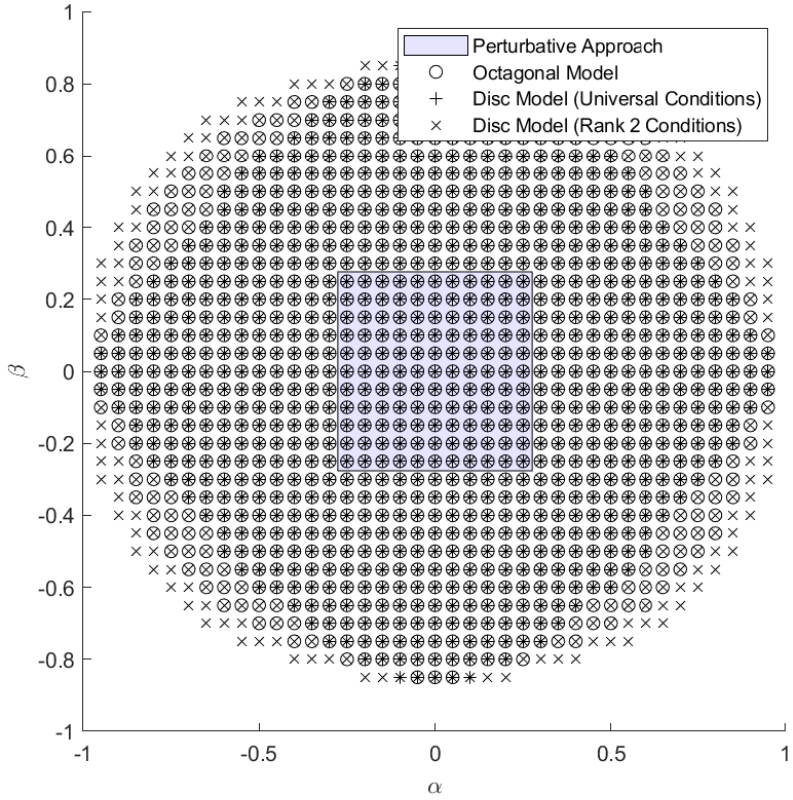


Figure 3.12: Stability (α, β)-regions of (3.6) using Theorem 3.5.4 (*universal conditions*) and Theorem 3.5.5 (*rank 2 conditions*) on the T-S-like model (3.72) (*disc model*), using Theorem 2.3.1 on the T-S model (3.46) (*octagonal model*), and using the perturbative approach (*Proposition 3.2.1*).

Chapter 4

Non-convex modeling of Takagi-Sugeno systems

This chapter revisits the nonlinear sector approach from the previous chapter, applying it piecewise to develop non-convex T-S models. Some LMI conditions for stability that effectively leverage the non-convexity of these models are provided.

4.1 The curse of convexity

Even if this is rarely *explicitly* discussed in the literature, it is well-known that the convex nature of the T-S models is in fact double-edged. Rewriting a nonlinear system (1.1) as a T-S model (2.1), e.g. by relying on the NLSA of Chapter 3, leads to the embedding of the original nonlinear system (1.1) in the convex-hull of a set of LTI systems. This embedding is extremely practical, as it gives access to all the LMI conditions of stability and stabilization discussed in Chapter 3. However, this convex-hull of LTI models can sometimes include a problematic model, preventing most of the usual LMI conditions of stability and stabilization from admitting a solution. Typically, an asymptotically stable, controllable or observable nonlinear system (1.1) can be embedded via (2.1) in a convex set of LTI systems, which may contain an unstable, non-controllable or non-observable system. This curse of convexity can be described more rigorously as follows.

Theorem 4.1.1 (The curse of convexity)**.** *Any method aiming at demonstrating the asymptotic stability of a nonlinear system $\dot{x}(t) = f(x(t), t)$ and relying on an LPV representation $\dot{x}(t) = A(\theta)x(t)$, with $\theta \in \Theta$, is condemned to never admit a solution, if*

1. *the method implies that for all $\theta \in \text{hull}(\Theta)$, there exists a positive-definite matrix $P_\theta \in \mathbb{S}_{n_x}^{++}(\mathbb{R})$ such that $A^\top(\theta)P_\theta + P_\theta A(\theta) \prec 0$;*
2. *there exists $\theta \in \text{hull}(\Theta)$ such that $A(\theta)$ is a non-Hurwitz matrix.*

Similarly, any method aiming at stabilizing a nonlinear system $\dot{x}(t) = f(x(t), u(t), t)$ and relying on an LPV representation $\dot{x}(t) = A(\theta)x(t) + B(\theta)u(t)$, with $\theta \in \Theta$, is condemned to never admit a solution, if

1. *the method implies that for all* $\theta \in \mathrm{hull}(\Theta)$, *there exists a positive-definite matrix* $P_\theta \in \mathbb{S}^{++}_{n_x}(\mathbb{R})$ *and a matrix* $K_\theta \in \mathbb{R}^{n_u \times n_x}$ *such that* $(A(\theta)+B(\theta)K_\theta)^\top P_\theta + P_\theta(A(\theta)+B(\theta)K_\theta) \prec 0$;
2. *there exists* $\theta \in \mathrm{hull}(\Theta)$ *such that* $(A(\theta), B(\theta))$ *is non-stabilizable pair.*

Similar issues occur for observability, as well as in a discrete-time setting.

Proof. In both cases, item 1 and item 2 are mutually exclusive [283]. □

Remark 4.1.1. *The phrase "curse of convexity" is somewhat contentious, as convexity is typically regarded as a beneficial property. Alternative terms that might better capture this concept include "curse of convexification" or "curse of convex embedding".*

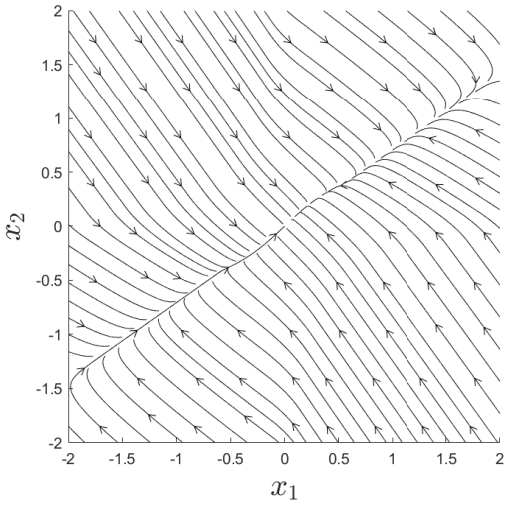
In the T-S framework, this convexity curse affects all the stability and stabilization LMI conditions discussed in Chapter 2.

Example 4.1.1. *Let us consider the following nonlinear system:*

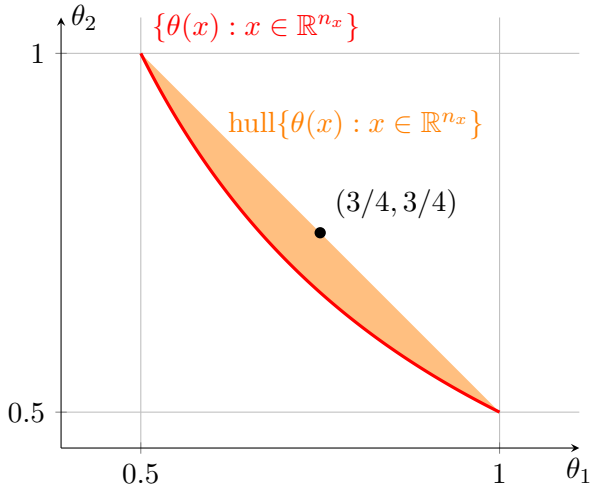
$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} -rx_1(t) + x_2(t)(3 + \mathrm{sat}(4x_1(t)x_2(t)))/4 \\ -rx_2(t) + 2x_1(t)/(3 + \mathrm{sat}(4x_1(t)x_2(t))) \end{pmatrix} \tag{4.1}$$

with r *a constant scalar value taken in the interval* $\left(\sqrt{2}/2, 3/4\right]$, *and where* sat *denotes the unitary saturation function:*

$$\mathrm{sat}(x) \triangleq \begin{cases} x & \text{if } x \in [-1, 1] \\ x/|x| & \text{else} \end{cases} \tag{4.2}$$



(a) Vector field associated with (4.1) for $r = 3/4$.



(b) Convex-hull of $\theta \in [0.5, 1]^2$ verifying $\theta_2 = 1/2\theta_1$.

Figure 4.1: Despite (4.1) being clearly asymptotically stable (left figure), the convex-hull of the state matrices $A(\theta)$ of the LPV model (4.3) for $\theta \in [0.5, 1]^2$ verifying $\theta_2 = 1/2\theta_1$ contains the matrix $A(3/4, 3/4)$, which is not Hurwitz (right figure).

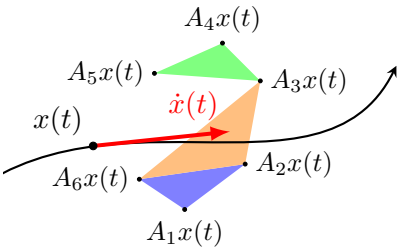


Figure 4.2: At each point $x(t)$ of the state trajectory (black line), $\dot{x}(t)$ belongs to $(\text{hull}\{A_1, A_2, A_6\} \cup \text{hull}\{A_2, A_3, A_6\} \cup \text{hull}\{A_3, A_4, A_5\}) \cdot x(t)$.

The vector field associated to this differential equation is plotted on Figure 4.1a. The system clearly seems asymptotically stable. Now (4.1) is rewritten as the LPV model (4.3) by considering a scheduling vector $\theta = \begin{pmatrix} \theta_1 & \theta_2 \end{pmatrix}^\top$.

$$\dot{x}(t) = A(\theta)x(t) \tag{4.3a}$$

$$\text{with } A(\theta) \triangleq \begin{pmatrix} -r & \theta_1 \\ \theta_2 & -r \end{pmatrix} \text{ and } \begin{cases} \theta_1 \triangleq (3 + \text{sat}(4x_1x_2))/4 \\ \theta_2 \triangleq 2/(3 + \text{sat}(4x_1x_2)) \end{cases} \tag{4.3b}$$

The scheduling parameters (θ_1, θ_2) *are bounded in* $[0.5, 1]^2$ *and are interdependent, with* $\theta_2 = 1/2\theta_1$ *being always verified. This model is typically problematic in a T-S setting, since the convex-hull of* $\{A(\theta(x)) : x \in \mathbb{R}^{n_x}\}$ *always contains the matrix* $A(3/4, 3/4)$*, which is never Hurwitz for* $r \in \left(\sqrt{2}/2, 3/4\right]$ *(see Figure 4.1b).*

Mitigating solutions to this curse of convexity are known, typically by managing to exclude portions of the simplex Δ_{n_h-1} which are never reached by the activation functions $\mathbf{h}$, and for which the usual LMI conditions do not need to be verified [240, 30, 312]. However, this requires knowledge on which subsets of the simplex Δ_{n_h-1} can be reached by the activation functions $\mathbf{h}$, and these approaches tend to generate a large number of generally unintuitive LMI conditions.

This chapter aims at bypassing the curse of convexity at the modeling step itself, by introducing non-convex T-S systems. Intuitively, these models consist of a *juxtaposition of convex linear embedding* of the system (1.1). More precisely, the shifted state vector $\delta x(t)$ and the output vectors $y(t)$ and $z(t)$ of (1.1) are rewritten using several convex combination of several LTI local models partitioning a non-convex set (Figure 4.2). Formally, these systems take form of a switched T-S model, where the switching is performed between several T-S model, while maintaining continuity with respect to θ between the active local models. Such models can be written as follows:

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = \sum_{i=1}^{n_v} v_i(\theta) \sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) \begin{pmatrix} A_{i,j} & B_{1,i,j} & B_{2,i,j} \\ C_{1,i,j} & D_{11,i,j} & D_{12,i,j} \\ C_{2,i,j} & D_{21,i,j} & D_{22,i,j} \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \tag{4.4}$$

where:

- the family $\{v_i\}_{1 \leq i \leq n_v}$, with $v_i(\theta) \in \{0, 1\}$ and $\sum_{i=1}^{n_v} v_i(\theta) = 1$, can be interpreted as the switching signal between the local T-S models;
- for all $i \in [\![1, n_v]\!]$, $\mathbf{h}_i \triangleq (h_{i,1}, \ldots, h_{i,n_{h_i}}) \in \Delta_{n_{h_i}-1}$ are the activation functions of the i-th local T-S model.

The challenges associated with these non-convex T-S models are two-folded. First, the continuity with respect to θ of these models should be ensured at the modeling step. Moreover, the results obtained from these models should not be equivalent to those that could be obtained by considering the convex-hull of the whole juxtaposition of polytopes (in other words, the curse of convexity should be conjured). Although these kinds of partitioning techniques have already been investigated both by the T-S [155] and the LPV [15, 244, 13] frameworks for straightforward conservatism reduction purposes, to the author's knowledge, these two challenges have generally been overlooked by the literature.

This chapter is organized as follows: the derivation of exact non-convex T-S models such as (4.4) is formalized in Section 4.2 using the NLSA of Chapter 3. Section 4.3 provides local stability LMI conditions for the previously introduced non-convex T-S models, by adapting well-known results to this non-convex context. A numerical illustration of this result is provided in Section 4.4, demonstrating that the curse of convexity has been conjured. Finally, some conclusions and perspectives are discussed in Section 4.5

4.2 A non-convex nonlinear sector approach

This section expands the NLSA on polytopic bounding sets described in Chapter 3 to all the bounding sets that can be defined using the facets $\{\Theta_i\}_{1\leq i\leq n_v}$ of a *polyhedral complex* $\mathcal{S}$. After defining a polyhedral complex and its facets in Definition 4.2.1 and Definition 4.2.2, following the procedure of the NLSA, Theorem 4.2.1 provides a switched T-S model which exactly represents the nonlinear system (1.1) for all $\theta \in \bigcup_{i=1}^{n_v} \Theta_i$. This switched T-S model is continuous with respect to the scheduling vector θ.

Definition 4.2.1 (Polyhedral complex [314]). *Let $\mathcal{S}$ stand for a family of polytopes. This family is called a polyhedral complex if:*

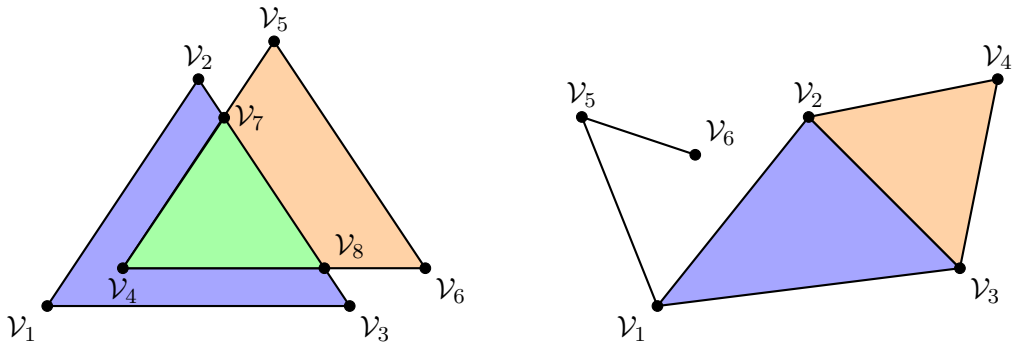
1. *for all $\Theta \in \mathcal{S}$ such that $\dim(\Theta) > 0$, the facets of Θ are also in $\mathcal{S}$, i.e. $f(\Theta) \subset \mathcal{S}$;*
2. *for all $\Theta_1, \Theta_2 \in \mathcal{S}$, the intersection between these two polytopes is either empty, i.e. $\Theta_1 \cap \Theta_2 = \emptyset$, or is a face $\Theta_3 \in \mathcal{S}$ which is common to both Θ_1 and Θ_2 (with $\dim(\Theta_3) \leq \min(\dim(\Theta_1), \dim(\Theta_2))$).*

Note that the first condition implies that if $\Theta \in \mathcal{S}$, then all its faces belong to $\mathcal{S}$ as well.

Definition 4.2.2 (Facets of polyhedral complex). *Let $\mathcal{S}$ be a polyhedral complex. The facets of $\mathcal{S}$, denoted $f(\mathcal{S})$, are the polytopes of $\mathcal{S}$ which are not a face of a larger polytope in $\mathcal{S}$.*

Definition 4.2.3 (Vertices of polyhedral complex). *Let $\mathcal{S}$ be a polyhedral complex. The vertices of $\mathcal{S}$, denoted $v(\mathcal{S})$, are the polytopes of $\mathcal{S}$ of dimension 0.*

Remark 4.2.1. *A polyhedral complex is not necessarily path-connected, although only path-connected polyhedral complexes are investigated in this chapter.*



(a) The family of polytopes $\mathcal{S}_1$ in Example 4.2.1 is not a polyhedral complex.

(b) The family of polytopes $\mathcal{S}_2$ in Example 4.2.2 is a polyhedral complex.

Figure 4.3: Illustration of Examples 4.2.1 and 4.2.2.

Example 4.2.1. *The family of polytopes $\mathcal{S}_1$ illustrated in Figure 4.3a is given by*

$$\begin{aligned}\mathcal{S}_1 = \{&(123), (456), (478), (12), (23), (31), (45), (56), (64), \dots \\ &(47), (78), (84), (1), (2), (3), (4), (5), (6), (7), (8)\}\end{aligned} \tag{4.5}$$

where $(k_1, \dots, k_m)$ stands for $\text{hull}\{\mathcal{V}_{k_1}, \dots, \mathcal{V}_{k_m}\}$. This family is not a polyhedral complex. Indeed, $(123) \cap (456) = (478)$. Although (478) is in $\mathcal{S}$, it is neither a face of (123) nor of (456). It is recalled that according to the Definition 3.4.1 of Chapter 3, the only face of a polytope which has the same dimension as this polytope is the polytope itself.

Example 4.2.2. *The family of polytopes $\mathcal{S}_2$ illustrated in Figure 4.3b is given by*

$$\mathcal{S}_2 = \{(123), (234), (12), (23), (34), (42), (31), (15), (56), (1), (2), (3), (4), (5), (6)\} \tag{4.6}$$

where $(k_1, \dots, k_m)$ stands for $\text{hull}\{\mathcal{V}_{k_1}, \dots, \mathcal{V}_{k_m}\}$. This family is a polyhedral complex. Its facets and vertices are given by:

$$f(\mathcal{S}_2) = \{(123), (234), (15), (56)\} \tag{4.7a}$$

$$v(\mathcal{S}_2) = \{(1), (2), (3), (4), (5), (6)\} \tag{4.7b}$$

In order to demonstrate the continuity of the exact switched T-S model which is presented at the end of this section, it is necessary to introduce the following lemma.

Lemma 4.2.1 (Continuity on a polyhedral complex)**.** *Let $\mathcal{S}$ be a polyhedral complex with facets $f(\mathcal{S}) = \{\Theta_i\}_{1 \leq i \leq n_v}$, and let $\{\Gamma_{\mathcal{V}}\}_{\mathcal{V} \in v(\mathcal{S})}$ be a set of matrices. The following function is continuous on $\bigcup_{\Theta \in f(\mathcal{S})} \Theta$:*

$$g : \theta \mapsto \sum_{i=1}^{n_v} v_i(\theta) \sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) \Gamma_{\mathcal{V}_{i,j}} \tag{4.8}$$

where:

- *$v_i(\theta) \triangleq \delta_{i,n^*(\theta)}$, with $n^*(\theta)$ the index of a unique facet of $\mathcal{S}$ such that $\theta \in \Theta_{n^*(\theta)}$;*
- *$\mathbf{h}_i \triangleq (h_{i,1}, \dots, h_{i,n_{h_i}}) \in \Delta_{n_{h_i}-1}$ are the barycentric coordinates of θ in Θ_i;*
- *$\mathcal{V}_{ij}$ is j-th vertex of the facet Θ_i of $\mathcal{S}$.*

Proof. By continuity of the barycentric coordinates with respect to $\theta \in \mathbb{R}^{n_\theta}$, for all $i \in [\![1, n_v]\!]$, the following sum is continuous with respect to $\theta \in \Theta_i$:

$$\sum_{j=1}^{n_{h_i}} h_{i,j}(\theta)\Gamma_{\mathcal{V}_{i,j}} \tag{4.9}$$

Hence, g is a continuous function of θ if the values of the active sums (4.9) in (4.8) coincide at the boundary of the facets of $\mathcal{S}$. Let $i \in [\![1, n_v]\!]$. For all $\theta \in \partial\Theta_i$, the boundary of Θ_i, either θ only belongs to the facet Θ_i of $\mathcal{S}$, and there is nothing to demonstrate, or there exists $k_1, \ldots, k_r \in [\![1, n_v]\!]$ such that $\theta \in \bigcup_{i=1}^{r} \Theta_{k_i}$. In the latter case, for all $j \in \{k_1, \ldots, k_r\}$ (and in particular for j^* such that $n^*(\theta) = j^*$), by definition of a polyhedral complex, the intersection between Θ_i and Θ_j is a face $\mathcal{F}$ common to both Θ_i and Θ_j. Let $n_\mathcal{F}$ stand for the number of vertices $v(\mathcal{F})$ associated with this common face $\mathcal{F}$. There exists two injective functions $\varphi_1 : [\![1, n_\mathcal{F}]\!] \to [\![1, n_{h_i}]\!]$ and $\varphi_2 : [\![1, n_\mathcal{F}]\!] \to [\![1, n_{h_j}]\!]$ such that $v(\mathcal{F}) = \{\mathcal{V}_{i,\varphi_1(k)}\}_{1\leq k\leq n_\mathcal{F}} = \{\mathcal{V}_{j,\varphi_2(k)}\}_{1\leq k\leq n_\mathcal{F}}$, with for all $k \in [\![1, n_\mathcal{F}]\!]$, $\mathcal{V}_{i,\varphi_1(k)} = \mathcal{V}_{j,\varphi_2(k)}$, hence $\Gamma_{\mathcal{V}_{i,\varphi_1(k)}} = \Gamma_{\mathcal{V}_{j,\varphi_2(k)}}$. Moreover, by definition of the barycentric coordinates, for all $\mathcal{V}_{i,j}$ of $v(\Theta_i)$ and $v(\Theta_j)$ which are not vertices of $\mathcal{F}$, $h_{i,j}(\theta) = 0$, which also implies $h_{i,\varphi_1(k)} = h_{j,\varphi_2(k)}$ for all $k \in [\![1, n_\mathcal{F}]\!]$. In the end, the following stands:

$$g(\theta) = \sum_{k=1}^{n_{h_i}} h_{i,k}(\theta)\Gamma_{\mathcal{V}_{i,k}} = \sum_{k=1}^{n_{h_j}} h_{j,k}(\theta)\Gamma_{\mathcal{V}_{j,k}} \tag{4.10}$$

which ensures that for all $i \in [\![1, n_v]\!]$, there is no discontinuity of $g(\theta)$ at any point of Θ_i, providing the overall continuity of g on $\bigcup_{\Theta\in f(\mathcal{S})} \Theta$. □

The following generalization of the NLSA on the facets of a polyhedral complex can be found.

Theorem 4.2.1 (NLSA on the facets of a polyhedral complex)**.** *Let $\mathcal{S}$ be a polyhedral complex with facets $f(\mathcal{S}) = \{\Theta_i\}_{1\leq i\leq n_v}$. The switched T-S model (4.11) is an exact representation of the nonlinear system (1.1) for all $x(t)$, $u(t)$, $w(t)$ and t such that $\theta \in \bigcup_{i=1}^{n_v} \Theta_i$.*

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = \sum_{i=1}^{n_v} v_i(\theta) \sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) \begin{pmatrix} A(\mathcal{V}_{ij}) & B_1(\mathcal{V}_{ij}) & B_2(\mathcal{V}_{ij}) \\ C_1(\mathcal{V}_{ij}) & D_{11}(\mathcal{V}_{ij}) & D_{12}(\mathcal{V}_{ij}) \\ C_2(\mathcal{V}_{ij}) & D_{21}(\mathcal{V}_{ij}) & D_{22}(\mathcal{V}_{ij}) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \tag{4.11}$$

where:

- $v_i(\theta) \triangleq \delta_{i,n^*(\theta)}$*, with $n^*(\theta)$ the index of a unique facet of $\mathcal{S}$ such that $\theta \in \Theta_{n^*(\theta)}$;*
- $\mathbf{h}_i \triangleq (h_{i,1}, \ldots, h_{i,n_{h_i}}) \in \Delta_{n_{h_i}-1}$ *are the barycentric coordinates of θ in Θ_i obtained with Theorem 3.4.1 of Chapter 3;*
- $\mathcal{V}_{ij}$ *is j-th vertex of the facet Θ_i.*

This representation is continuous with respect to $\theta \in \bigcup_{i=1}^{n_v} \Theta_i$.

Proof. For all $i \in [\![1, n_v]\!]$, Theorem 3.4.1 states that for all $x(t)$, $u(t)$, $w(t)$ and t such that $\theta \in \Theta_i$, (4.11) can be rewritten as the nonlinear system (1.1), hence, (3.1) also stands for all $x(t)$, $u(t)$,

$w(t)$ and t such that $\theta \in \bigcup_{i=1}^{n_v} \Theta_i$. Moreover, Lemma 4.2.1 ensures that this representation is continuous with respect to $\theta \in \bigcup_{i=1}^{n_v} \Theta_i$. □

Remark 4.2.2. *The common faces to the facets $f(\mathcal{S}) = \{\Theta_i\}_{1 \leq i \leq n_v}$ of $\mathcal{S}$ correspond to the switch regions (also called the switch surfaces) of the switched system (4.11) [215, 216]. Thanks to the continuity of (4.11) with respect to θ, the transition between the local T-S models in these switch regions is seamless.*

4.3 Non-convex stability conditions

Contrary to the usual switched T-S models, there is no use in studying the stability of the input-free switched T-S model (4.12) obtained via Theorem 4.2.1 with the results of Chapter 2.

$$\dot{x}(t) = \sum_{i=1}^{n_v} v_i(\theta) \left(\sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) A(\mathcal{V}_{i,j}) x(t) \right) \tag{4.12}$$

Indeed, by the curse of convexity (Theorem 4.1.1), demonstrating the stability of (4.12) on $\bigcup_{i=1}^{n_v} \Theta_i$ through the results of Chapter 2 directly provides the stability of the T-S model obtained by applying the polytopic NLSA (Theorem 3.4.1 of Chapter 3) to the convex-hull of $\bigcup_{i=1}^{n_v} \Theta_i$, and there is no point in considering a non-convex bounding set in the first place.

In order to benefit from the non-convexity of the bounding set $\bigcup_{i=1}^{n_v} \Theta_i$, PQLF and MQLF are not sufficient, and the following Piecewise Multiquadratic Lyapunov Function (PMQLF) is introduced:

$$V(x, \theta) = \sum_{i=1}^{n_v} v_i(\theta) x^\top \left(\sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) P_{i,j} \right) x \tag{4.13}$$

This type of Lyapunov functions is also called piecewise (fuzzy) weighting dependent Lyapunov functions [208, 207]. Contrary to the existing literature on switched T-S system, where continuity conditions have to be imposed on the matrices $\{P_{i,j}\}$ to obtain a continuous Lyapunov function (4.13) with respect to θ, it is straightforward to achieve this continuity by leveraging Lemma 4.2.1. It is in fact sufficient to consider the polyhedral complex $\mathcal{S}$ underlying the T-S model (4.12) obtained via Theorem 4.2.1, and to associate a single matrix $P_\mathcal{V}$ to each vertex $\mathcal{V} \in v(\mathcal{S})$ of this polyhedral complex. Simply put, the set of matrices $\{P_{i,j}\}$ corresponding to a same vertex should be equal.

The LMI conditions to perform the stability analysis of (4.12) can now be very easily derived from Theorem 2.3.4 and Theorem 2.3.6 of Chapter 2, by leveraging the continuity property of (4.13) imposed via Lemma 4.2.1. Note that the natural extension of these theorems relies heavily on the construction of the switched T-S model (4.12) via Theorem 4.2.1, and is not directly applicable to the usual switched T-S models.

Assumption 4.3.1. *The scheduling vector θ depends only on the state $x(t)$ of (4.12), and is continuously differentiable with respect to it.*

Assumption 4.3.2. *For all $i \in [\![1, n_v]\!]$, the activation functions $\mathbf{h}_i$ are continuously differentiable with respect to the scheduling vector θ.*

Assumption 4.3.3. *For all $i \in [\![1, n_v]\!]$, $j \in [\![1, n_{h_i}]\!]$ and $t \in \mathbb{R}$, there exists a Lipschitz constant $\phi_{i,j} \in \mathbb{R}_{>0}$ such that the Lie derivative of $h_{i,j}$ is bounded along the trajectories of the system, i.e. $|dh_{i,j}(\theta(x(t)))/dt| \leq \phi_{i,j}$.*

Assumption 4.3.4. *For all initial conditions $x(t_0) \in \mathbb{R}^{n_x}$, there only exists a finite number of $t \in [t_0, +\infty)$ such that $\theta(x(t)) \in \bigcup_{i=1}^{n_v} \partial\Theta_i$.*

Assumption 4.3.1 can be relaxed by assuming that:

- θ also depends on continuously differentiable exogenous signals;
- θ is continuously differentiable with respect to these exogenous signals;
- Assumption 4.3.3 also holds with respect to these signals.

The mention of these exogenous signals is omitted for the sake of simplicity. Moreover, Assumption 4.3.3 often holds locally, leading to local exponential stability results rather than global exponential stability results. The exponential stability is only ensured inside the outermost Lyapunov level contained in the set $\bigcap_{i=1}^{n_v} R_i$ where Assumption 4.3.3 is verified, with:

$$R_i \triangleq \left\{ x \in \mathbb{R}^{n_x} : \theta(x) \in \Theta_i \text{ and } \forall j \in [\![1, n_{h_i}]\!], \left| \sum_{k=1}^{n_{h_i}} h_{i,k}(\theta(x)) \left\langle \frac{\partial h_{i,j}}{\partial x} | A_{i,k} x \right\rangle \right| \leq \phi_{i,j} \right\} \tag{4.14}$$

In the following, R_V denotes the region contained inside the outermost Lyapunov level contained in the set $\bigcap_{i=1}^{n_v} R_i$ where Assumption 4.3.3 is verified. If Assumption 4.3.3 holds globally, $R_V = \mathbb{R}^{n_x}$. Finally, Assumption 4.3.4 is a technicality which has to be introduced to compensate for the loss of differentiability of the Lyapunov function (4.13) along the switching surface of the system, and it can be difficult to verify *a priori*. This last assumption allows the author to apply the Theorem 1 of [216] to obtain simple stability conditions. However, this assumption is probably not fundamentally useful in this chapter, and the author believes that the proposed conditions impose enough constraints on the upper right Dini derivative of the Lyapunov function (4.13) along the system trajectories to obtain exponential stability, despite the loss of differentiability [69]. This remains to be properly demonstrated.

Theorem 4.3.1 (Non-convex multiquadratic stability)**.** *Under the Assumptions 4.3.1, 4.3.2, 4.3.3 and 4.3.4, the T-S model (4.12) obtained via Theorem 4.2.1 is exponentially stable in R_V if there exists $\{P_{i,j}\}_{1 \leq j \leq n_{h_i}}^{1 \leq i \leq n_v}$ with $P_{i,j} \in \mathbb{S}_{n_x}(\mathbb{R})$ such that (4.13) is continuous with respect to θ on $\bigcup_{i=1}^{n_v} \Theta_i$, and such that the conditions (4.15) are satisfied for all $i \in [\![1, n_v]\!]$.*

$$P_{i,j} \succ 0, \quad \forall j \in [\![1, n_{h_i}]\!] \tag{4.15a}$$

$$\sum_{j_1=1}^{n_{h_i}} \sum_{j_2=1}^{n_{h_i}} h_{i,j_1} h_{i,j_2} \left(\mathcal{H}(P_{i,j_1} A_{i,j_2}) + \sum_{k=0}^{n_{h_i}} \phi_{i,k} P_{i,k} \right) \prec 0, \quad \forall \mathbf{h}_i \in \Delta_{n_{h_i}-1} \tag{4.15b}$$

The inequality on the double convex sum in (4.15b) can be transformed into regular LMI conditions by leveraging the results of Section 2.2.2.

Theorem 4.3.2 (Relaxed non-convex multiquadratic stability)**.** *Under the Assumptions 4.3.1, 4.3.2, 4.3.3 and 4.3.4, the T-S model (4.12) obtained via Theorem 4.2.1 is exponentially stable in*

R_V if there exist $\{P_{i,j}\}_{1\leq j\leq n_{h_i}}^{1\leq i\leq n_v}$ and $\{M_{i,j}\}_{1\leq j\leq 3}^{1\leq i\leq n_v}$, with $M_{i,1}, M_{i,2} \in \mathbb{R}^{n_x\times n_x}$ and $M_{i,3} \in \mathbb{S}_{n_x}(\mathbb{R})$, $P_{i,j} \in \mathbb{S}_{n_x}(\mathbb{R})$ such that (4.13) is continuous with respect to θ on $\bigcup_{i=1}^{n_v}\Theta_i$, and such that the LMI conditions (4.16) are satisfied for all $i \in [\![1, n_v]\!]$ and $j \in [\![1, n_{h_i}]\!]$.

$$P_{i,j} \succ 0 \tag{4.16a}$$

$$P_{i,j} + M_{3,j} \succeq 0 \tag{4.16b}$$

$$\begin{pmatrix} \sum_{k=1}^{n_{h_i}} \phi_{i,k}(P_{i,k} + M_{i,3}) - \mathcal{H}(M_{i,1}A_j) & P_{i,j} + M_{i,1} - A_{i,j}^\top M_{i,2}^\top \\ P_{i,j} + M_{i,1}^\top - M_{i,2}A_{i,j} & \mathcal{H}(M_{i,2}) \end{pmatrix} \prec 0 \tag{4.16c}$$

Proof. The continuity of the PMQLF $V(x,\theta)$ (4.13) with respect to θ can be imposed by taking all the matrices $\{P_{i,j}\}$ corresponding to a same vertex equal to each other (Lemma 4.2.1). θ is a continuous function of $x \in \mathbb{R}^{n_x}$, and the PMQLF $V(x,\theta)$ (4.13) is a continuous function of $x \in \mathbb{R}^{n_x}$ and $\theta \in \Theta$, hence $V(x,\theta(x))$ is a continuous function of $x \in \mathbb{R}^{n_x}$. Moreover $V(x(t),\theta(x(t)))$ is continuously differentiable with respect to $t \in \mathbb{R}$ for all $x(t) \in \mathbb{R}^{n_x}$ such that $\theta(x(t)) \in \mathrm{intr}(\Theta_i)$, and for all initial condition $x(t_0)$, there only exists a finite number of t such that $\theta(x(t)) \in \bigcup_{i=1}^{n_v}\partial\Theta_i$, hence $V(x(t),\theta(x(t)))$ is continuous and piecewise differentiable with respect to $t \in \mathbb{R}$. Let $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}_{>0}$ denote the minimum and maximum eigenvalues of $\{P_{i,j}\}_{1\leq j\leq n_{h_i}}^{1\leq i\leq n_v}$. From $\lambda_{\min}I_{n_x} \prec P_{i,j} \prec \lambda_{\max}I_{n_x}$, for all $x \in \mathbb{R}^{n_x}$, the following bounds are verified

$$\lambda_{\min}\|x\|_2^2 \leq V(x,\theta(x)) \leq \lambda_{\max}\|x\|_2^2 \tag{4.17}$$

Moreover, if the conditions (4.15) or (4.16) are satisfied, by Theorem 2.3.4 and Theorem 2.3.6 of Chapter 2, for all $i \in [\![1, n_{h_i}]\!]$ there exists $\varepsilon_i \in \mathbb{R}_{>0}$ such that if $\theta(x(t)) \in \mathrm{intr}(\Theta_i)$, then $\dot{V}(x(t),\theta(x(t))) \leq -\varepsilon_i\|x(t)\|^2$. Finally, the conditions of Theorem 1 of [216] are satisfied, and the T-S model (4.12) is globally exponentially stable. □

Remark 4.3.1. *Although these LMI conditions rely on a polyhedral complex which is not necessarily convex, they remain LMI conditions, and as such, define a convex set of solutions.*

4.4 Application

A stability analysis is performed on the nonlinear system (4.1) of Example 4.1.1. The two polytopes $\Theta_1 = \mathrm{hull}\{\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3\}$, $\Theta_2 = \mathrm{hull}\{\mathcal{V}_3, \mathcal{V}_4, \mathcal{V}_5\}$ are chosen to be the bounding set of θ in the LPV representation (4.3) of (4.1). As illustrated by Figure 4.4, the polytopes Θ_1 and Θ_2 are the facets of the following polyhedral complex:

$$\mathcal{S} = \{(123), (345), (12), (23), (31), (34), (45), (53), (1), (2), (3), (4), (5)\} \tag{4.18}$$

where $(k_1, \ldots, k_m)$ stands for $\mathrm{hull}\{\mathcal{V}_{k_1}, \ldots, \mathcal{V}_{k_m}\}$, and where the vertices are given by:

$$\mathcal{V}_1 = \begin{pmatrix} 1/2 & 1 \end{pmatrix}^\top \tag{4.19a}$$

$$\mathcal{V}_2 = \begin{pmatrix} 2-\sqrt{2} & 2(\sqrt{2}-1) \end{pmatrix}^\top \tag{4.19b}$$

$$\mathcal{V}_3 = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}^\top \tag{4.19c}$$

$$\mathcal{V}_4 = \begin{pmatrix} 2(\sqrt{2}-1) & 2-\sqrt{2} \end{pmatrix}^\top \tag{4.19d}$$

$$\mathcal{V}_5 = \begin{pmatrix} 1 & 1/2 \end{pmatrix}^\top \tag{4.19e}$$

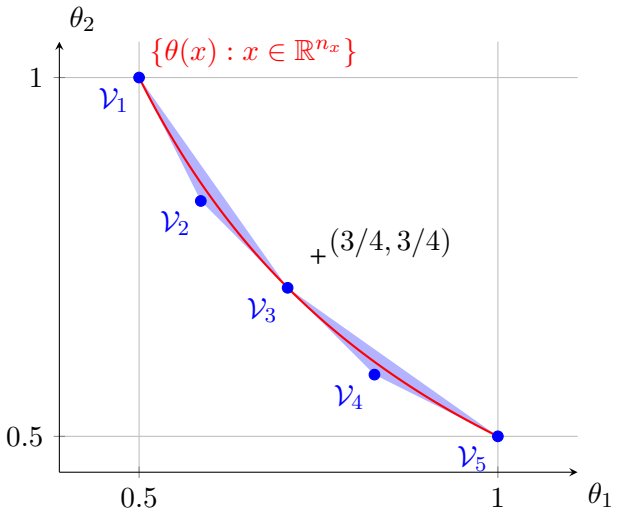


Figure 4.4: The polytopes Θ_1 and Θ_2 are bounding the scheduling vector θ.

The polytopes Θ_1 and Θ_2 have the following facets:

$$\text{facets of } \Theta_1 : \begin{cases} \mathcal{F}_{1,1} = \text{hull}\{\mathcal{V}_1, \mathcal{V}_2\} \\ \mathcal{F}_{1,2} = \text{hull}\{\mathcal{V}_2, \mathcal{V}_3\} \\ \mathcal{F}_{1,3} = \text{hull}\{\mathcal{V}_3, \mathcal{V}_1\} \end{cases} \qquad \text{facets of } \Theta_2 : \begin{cases} \mathcal{F}_{2,1} = \text{hull}\{\mathcal{V}_3, \mathcal{V}_4\} \\ \mathcal{F}_{2,2} = \text{hull}\{\mathcal{V}_4, \mathcal{V}_5\} \\ \mathcal{F}_{2,3} = \text{hull}\{\mathcal{V}_5, \mathcal{V}_3\} \end{cases} \tag{4.20}$$

The normals to these facets are given below.

$$\text{normals of } \Theta_1 : \begin{cases} \mathcal{N}_{\mathcal{F}_{1,1}} = -\begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_{1,2}} = -\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_{1,3}} = \begin{pmatrix} \sqrt{2/3} & 1/\sqrt{3} \end{pmatrix}^\top \end{cases} \tag{4.21a}$$

$$\text{normals of } \Theta_2 : \begin{cases} \mathcal{N}_{\mathcal{F}_{2,1}} = -\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_{2,2}} = -\begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}^\top \\ \mathcal{N}_{\mathcal{F}_{2,3}} = \begin{pmatrix} 1/\sqrt{3} & \sqrt{2/3} \end{pmatrix}^\top \end{cases} \tag{4.21b}$$

The two polytopes being 2-dimensional, i.e. polygons, they are simple polytopes. The weight functions, computed using Lemma 3.4.1, are:

$$\text{weights of } \Theta_1 : \begin{cases} w_{\mathcal{V}_1}^{\Theta_1}(\theta) = \frac{2(\sqrt{2}-2)}{(2\theta_1+\theta_2-2)(2\sqrt{2}\theta_1+2\theta_2-\sqrt{2}-2)} \\ w_{\mathcal{V}_2}^{\Theta_1}(\theta) = \frac{1}{(\theta_1+\theta_2-\sqrt{2})(2\theta_1+\theta_2-2)} \\ w_{\mathcal{V}_3}^{\Theta_1}(\theta) = \frac{\sqrt{2}(\sqrt{2}-2)}{(\theta_1+\theta_2-\sqrt{2})(2\sqrt{2}\theta_1+2\theta_2-\sqrt{2}-2)} \end{cases} \tag{4.22a}$$

$$\text{weights of } \Theta_2 : \begin{cases} w_{\mathcal{V}_3}^{\Theta_2}(\theta) = \frac{\sqrt{2}(\sqrt{2}-2)}{(\theta_1+\theta_2-\sqrt{2})(2\sqrt{2}\theta_2+2\theta_1-\sqrt{2}-2)} \\ w_{\mathcal{V}_4}^{\Theta_2}(\theta) = \frac{1}{(\theta_1+\theta_2-\sqrt{2})(2\theta_2+\theta_1-2)} \\ w_{\mathcal{V}_5}^{\Theta_2}(\theta) = \frac{2(\sqrt{2}-2)}{(2\theta_2+\theta_1-2)(2\sqrt{2}\theta_2+2\theta_1-\sqrt{2}-2)} \end{cases} \tag{4.22b}$$

which provides the following barycentric coordinates of θ within the triangle Θ_1:

$$h_{1,1}(\theta) = \frac{2(\sqrt{2}-2)(\theta_1+\theta_2-\sqrt{2})}{7\sqrt{2}-10} \qquad \text{[associated with } \mathcal{V}_1 \in \Theta_1\text{]} \tag{4.23a}$$

$$h_{1,2}(\theta) = \frac{2\sqrt{2}\theta_1+2\theta_2-\sqrt{2}-2}{7\sqrt{2}-10} \qquad \text{[associated with } \mathcal{V}_2 \in \Theta_1\text{]} \tag{4.23b}$$

$$h_{1,3}(\theta) = \frac{\sqrt{2}(\sqrt{2}-2)(2\theta_1+\theta_2-2)}{7\sqrt{2}-10} \qquad \text{[associated with } \mathcal{V}_3 \in \Theta_1\text{]} \tag{4.23c}$$

and the following barycentric coordinates of θ within the triangle Θ_2:

$$h_{2,1}(\theta) = \frac{\sqrt{2}(\sqrt{2}-2)(2\theta_2+\theta_1-2)}{7\sqrt{2}-10} \qquad \text{[associated with } \mathcal{V}_3 \in \Theta_2\text{]} \tag{4.24a}$$

$$h_{2,2}(\theta) = \frac{2\sqrt{2}\theta_2+2\theta_1-\sqrt{2}-2}{7\sqrt{2}-10} \qquad \text{[associated with } \mathcal{V}_4 \in \Theta_2\text{]} \tag{4.24b}$$

$$h_{2,3}(\theta) = \frac{2(\sqrt{2}-2)(\theta_1+\theta_2-\sqrt{2})}{7\sqrt{2}-10} \qquad \text{[associated with } \mathcal{V}_5 \in \Theta_2\text{]} \tag{4.24c}$$

Moreover, the following functions are introduced:

$$v_1(\theta) = \begin{cases} 1 & \text{if } \theta \in \Theta_1 \\ 0 & \text{else} \end{cases} \tag{4.25a}$$

$$v_2(\theta) = \begin{cases} 1 & \text{if } \theta \in \Theta_2 \setminus \{\mathcal{V}_3\} \\ 0 & \text{else} \end{cases} \tag{4.25b}$$

Finally, the NLSA on the polyhedral complex $\mathcal{S}$ is completed (Theorem 4.2.1) and the system (4.1) is exactly represented by the following T-S model:

$$\dot{x}(t) = \sum_{i=1}^{2} v_i(\theta) \sum_{j=1}^{3} h_{i,j}(\theta) A(\mathcal{V}_j) x(t) \tag{4.26}$$

The stability analysis of the T-S model (4.26) is performed using Theorem 4.3.2 at several (r, ϕ) values, where r is a parameter of (4.1) and ϕ is the value of the Lipschitz constant systematically used in Assumption 4.3.3, i.e. $\phi_{i,j} = \phi$ for all $i \in \{1, 2\}$ and $j \in \{1, 2, 3\}$. It should be highlighted that the Assumption 4.3.3 makes these results only locally valid. Despite the curse of convexity (Theorem 4.1.1) which should affect the system, several PMQLF of the form (4.27) demonstrating the local exponential stability of (4.1) can be found for $r \in [0.7178, 3/4] \subset \left(\sqrt{2}/2, 3/4\right]$.

$$V(x, \theta(x)) = \sum_{i=1}^{2} v_i(\theta(x)) x^\top \left(\sum_{j=1}^{3} h_{i,j}(\theta(x)) P_{i,j} \right) x \tag{4.27}$$

The (r, ϕ)-region for which the system (4.26) is found to be locally exponentially stable is plotted on Figure 4.5. It should be noted that none of the results from Chapter 2 would have been able

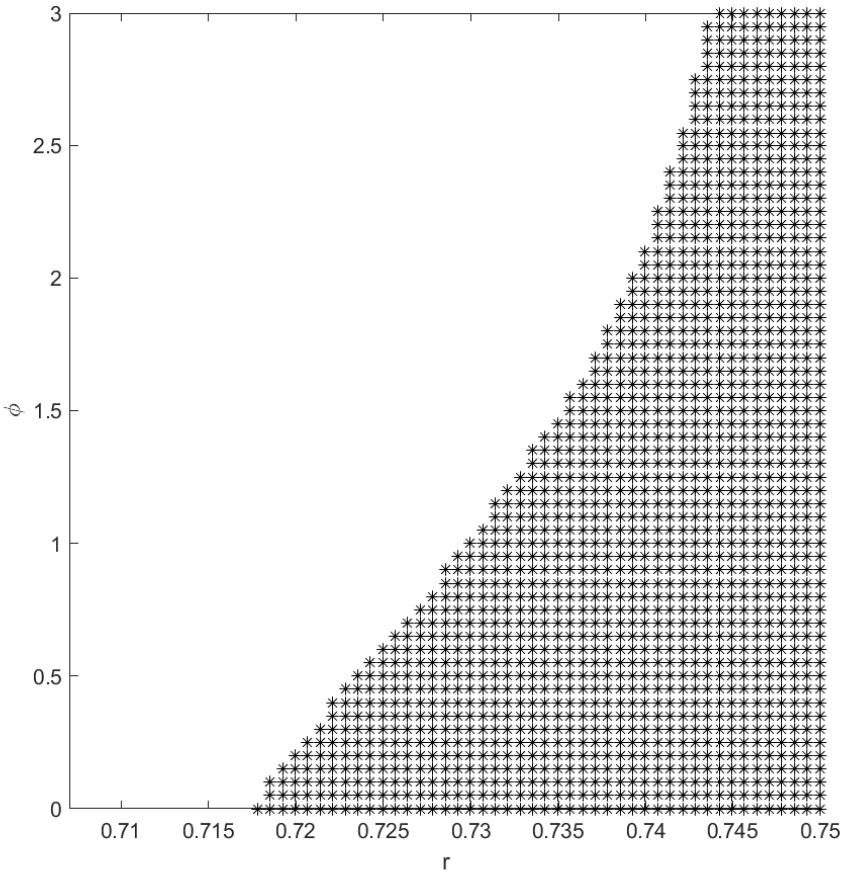


Figure 4.5: Stability (r, ϕ)-region of (4.1) using Theorem 4.3.2 on the T-S model (4.26).

to provide these results. For $r = 0.73$ and $\phi = 0.5$, the following values of $\{P_{i,j}\}$ are obtained:

$$P_{11} = \begin{pmatrix} 1.9047 & -0.0692 \\ -0.0692 & 1.4119 \end{pmatrix}, \qquad P_{12} = \begin{pmatrix} 1.5815 & 0.1094 \\ 0.1094 & 1.3404 \end{pmatrix} \tag{4.28a}$$

$$P_{13} = P_{21} = \begin{pmatrix} 1.2994 & 0.2684 \\ 0.2684 & 1.2994 \end{pmatrix} \tag{4.28b}$$

$$P_{22} = \begin{pmatrix} 1.3404 & 0.1094 \\ 0.1094 & 1.5815 \end{pmatrix}, \qquad P_{23} = \begin{pmatrix} 1.4119 & -0.0692 \\ -0.0692 & 1.9047 \end{pmatrix} \tag{4.28c}$$

The equality between P_{13} and P_{21} in (4.28b) ensures the continuity of the Lyapunov function (4.27), as detailed in Lemma 4.2.1.

Figure 4.6 illustrates the vector field associated with (4.1), together with the PMQLF (4.27) taken with the values obtained hereabove. Clearly, the continuity of the Lyapunov function is verified. Although it is not illustrated by this figure, the exponential stability is only ensured inside the outermost Lyapunov level contained in the set $\bigcap_{i=1}^{n_v} R_i$ where Assumption 4.3.3 is verified, where the R_i are defined by (4.14).

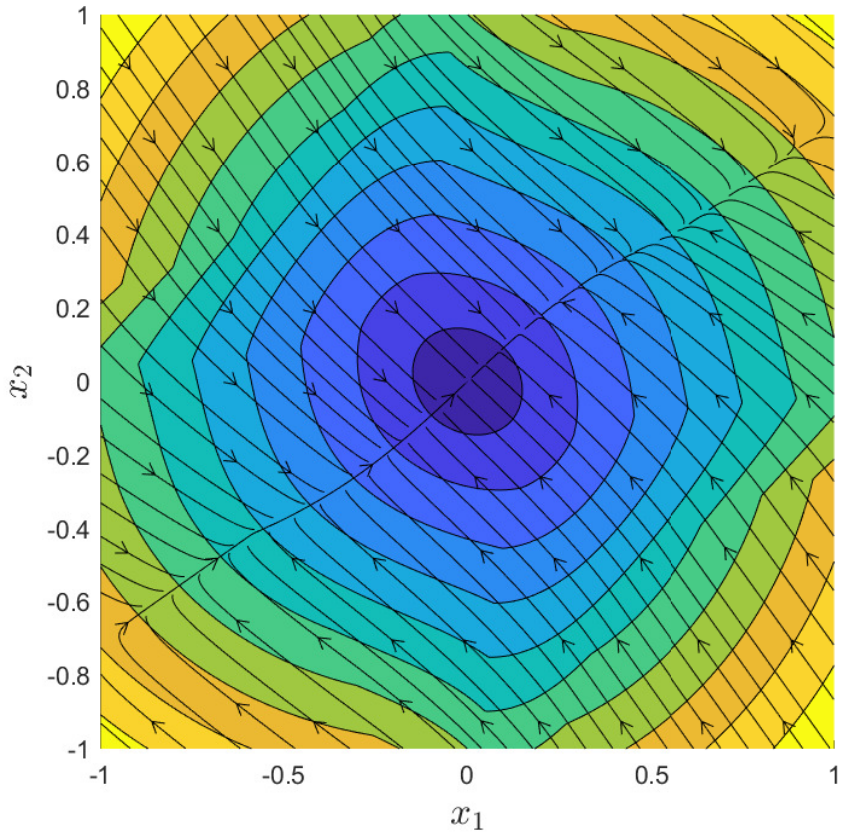


Figure 4.6: Vector field associated with (4.1) for $r = 0.73$, plotted together with the level sets of the PMQLF (4.27) obtained with (4.28).

4.5 Conclusions and perspectives

The convex nature of the T-S models being subject to a convexity curse (Theorem 4.1.1), non-convex T-S representations were introduced and investigated in this chapter. In order to derive these representations, the NLSA of Chapter 3 has been generalized to larger classes of non-convex bounding sets, by extending it to polyhedral complexes, a notion introduced in this chapter (Definition 4.2.1). LMI conditions of stability were then obtained for this new class of T-S models. It was demonstrated on a numerical example that the curse of convexity can be bypassed by leveraging the newly introduced non-convex representation. The results remain to be extended to LMI conditions of stabilization.

Chapter 5

Bézier interpolations in the Takagi-Sugeno framework

This chapter proposes to rewrite the multi-sums of the T-S framework using Bernstein polynomials to eliminate their redundancy, revealing their geometric nature as Bézier interpolation schemes. Some T-S results are revisited with this new perspective.

5.1 Multi-sums in the Takagi-Sugeno framework

As discussed in the Chapter 2 of this manuscript, multi convex sums of the form (5.1), such as the double convex sums introduced in Section 2.2.2 (where $m = 2$), are ubiquitous in the T-S framework .

$$\sum_{i_1=1}^{n_h} \cdots \sum_{i_m=1}^{n_h} h_{i_1}(\theta) \dots h_{i_m}(\theta) \Gamma_{i_1,\dots,i_m} \tag{5.1}$$

However, they suffer from a major inconvenience: the amount of matrices Γ is exponentially growing with m, the number of nested sums. Formally:

$$\#\left\{\Gamma_{i_1,\dots,i_m}\right\}_{1\leq i_1,\dots,i_m\leq n_h} = n_h^m \tag{5.2}$$

The combinatorial explosion due to multi-sums impacts a significant portion of the results of the T-S literature, especially those relying on generalized MQLF [66, 29], as well as generalized PDC or nPDC schemes [11, 165]. These generalizations have the major inconvenience of vastly increasing the number of decision variables in the LMI conditions for stability and stabilization.

However, it is actually well-known that a lot of the matrices $\Gamma_{i_1,\dots,i_m}$ in (5.1) are superfluous. For example, by considering (5.1) with $n_h = m = 2$, it is easily noticed that $\Gamma_{1,2}$ and $\Gamma_{2,1}$ play a similar role in the multi-sum and could be replaced by a single matrix of the form $G = (\Gamma_{1,2} + \Gamma_{2,1})/2$.

$$\sum_{i_1=1}^{2} \sum_{i_2=1}^{2} h_{i_1}(\theta) h_{i_2}(\theta) \Gamma_{i_1,i_2} = h_1^2(\theta)\Gamma_{1,1} + 2h_1(\theta)h_2(\theta)G + h_2^2(\theta)\Gamma_{2,2} \tag{5.3}$$

Moreover, the resulting products of activation functions h_1^2, $2h_1h_2$, and h_2^2 can be considered as new activation functions following the usual convex sum properties (Definition 2.1.1 of Chap-

Table 5.1: Comparison between the number of matrices Γ in (5.5) and in the multi-sum (5.1)

(n_h, m)	terms in (5.5)	terms in (5.1)
$(2,2)$	3	4
$(2,3)$	4	8
$(2,4)$	5	16
$(2,5)$	6	32
$(3,2)$	6	9
$(3,3)$	10	27
$(3,4)$	15	81
$(3,5)$	21	243

(n_h, m)	terms in (5.5)	terms in (5.1)
$(4,2)$	10	16
$(4,3)$	20	64
$(4,4)$	35	256
$(4,5)$	56	1024
$(5,2)$	15	25
$(5,3)$	35	125
$(5,4)$	70	625
$(5,5)$	126	3125

ter 2). Formally, for all $\theta \in \Theta$:

$$h_1^2(\theta),\ 2h_1(\theta)h_2(\theta),\ h_2^2(\theta) \geq 0 \qquad \text{[positivity]} \tag{5.4a}$$

$$h_1^2(\theta) + 2h_1(\theta)h_2(\theta) + h_2^2(\theta) = (h_1(\theta) + h_2(\theta))^2 = 1 \qquad \text{[partition of unity]} \tag{5.4b}$$

Of course, the T-S literature has already addressed the redundancy of multi-sums, particularly in order to obtain relaxed LMI conditions of stability and stabilization [11, 238]. However, to the author's surprise, no *explicit formula* is used in the T-S literature to handle the multi-sum (5.1) directly as a non-redundant convex sum, leaving a rewriting process to be worked out either algorithmically or manually (as performed in (5.3)). In the end, the non-redundant expressions provided by the existing T-S literature are defined implicitly, and are not necessarily convenient to employ during practical manipulations.

This chapter is concerned with replacing the redundant multi-sums (5.1) of the T-S framework with their *explicit* non-redundant expression (5.5), skipping the usual regrouping and rewriting step.

$$\sum_{i_1+\cdots+i_{n_h}=m} \frac{m!}{i_1!\dots i_{n_h}!} h_1^{i_1}(\theta)\dots h_{n_h}^{i_{n_h}}(\theta)\Gamma_{i_1,\dots,i_{n_h}} \tag{5.5}$$

The non-redundant expression (5.5) brings to light the geometric nature of the multi-sum (5.1), which turns out to be a simple Bézier interpolation scheme, a fact which is not discussed in the T-S literature yet. The explicit Bézier rewriting (5.5) of (5.1) drastically reduces the number of matrices Γ which are manipulated, without hindering the capabilities of the multi-sum, as it simply consists in regrouping its redundant terms.

$$\#\left\{\Gamma_{i_1,\dots,i_{n_h}}\right\}_{i_1+\cdots+i_{n_h}=m} = \frac{(m+n_h-1)!}{m!(n_h-1)!} \ll n_h^m \tag{5.6}$$

The number of matrices $\{\Gamma_{i_1,\dots,i_{n_h}}\}_{i_1+\cdots+i_{n_h}=m}$ in the sum (5.5) is compared in Table 5.1 to the number of matrices $\{\Gamma_{i_1,\dots,i_m}\}_{1\leq i_1,\dots,i_m\leq n_h}$ in (5.1) for several values of (n_h, m), providing the amount of useless matrices economized using the Bézier interpolation scheme (5.5) rather than the naive multi-sum (5.1). It should be noted that this combinatorial result has already been worked out by the T-S literature [238].

The geometric idea behind the Bézier interpolation (5.5) introduced in this chapter is the following. For $m = 1$, the matrices $\{\Gamma_i\}_{1\leq i\leq n_h}$ can be interpreted as the vertices of a polytope.

The activation functions $\mathbf{h} \in \Delta_{n_h-1}$ then define all the points inside this polytope through a convex sum. For $m \geq 2$, the Bézier multi-sum (5.5) consists in adding control points to this initial polytope, deforming the shape of the points obtained for $\mathbf{h} \in \Delta_{n_h-1}$ through a Bézier interpolation scheme. The shape of the polytope becomes *malleable* with a degree of flexibility increasing with the number of control points considered, which is itself increasing with m.

The chapter is organized as follows: in Section 5.2, the Bernstein polynomials are introduced in order to properly define the Bézier interpolation (5.5). Several useful results surrounding both the Bernstein polynomials and the Bézier interpolations schemes are introduced. Section 5.3 applies the introduced formalism to a series of results from the T-S framework, focusing particularly on the computation of the gain matrices of Bézier controllers and observers. This latter work has been published as a conference paper [18]. Finally, some conclusions and perspectives are discussed in Section 5.4.

5.2 Bernstein polynomials and Bézier interpolations

At this stage, it is not necessarily clear why (5.5) is a Bézier interpolation scheme. This section introduces the Bernstein polynomials in order to properly define Bézier interpolations, while providing useful properties surrounding both of these tools.

5.2.1 Multivariate Bernstein polynomials

The multivariate Bernstein polynomials are defined thereafter.

Definition 5.2.1 (The multivariate Bernstein polynomials [96])**.** *The multivariate Bernstein polynomials of degree $m \in \mathbb{N}_{>0}$ with $n \in \mathbb{N}_{>0}$ variables are defined for all $\mathbf{k} \in \mathbb{N}_m^n \triangleq \{(k_1, \dots, k_n) \in \mathbb{N}^n : \sum_{i=1}^n k_i = m\}$ as follows:*

$$\begin{aligned} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) &\triangleq \binom{m}{\mathbf{k}} \mathbf{X}^{\mathbf{k}} \\ &= \binom{m}{k_1, \dots, k_n} X_1^{k_1} \dots X_n^{k_n} \\ &= \frac{m!}{k_1! \dots k_n!} X_1^{k_1} \dots X_n^{k_n} \end{aligned} \tag{5.7}$$

Usually, $\mathbf{X} \triangleq (X_1, \dots, X_n)$ belongs to the $(n-1)$-simplex Δ_{n-1}.

Remark 5.2.1. *The literature often defines the multivariate Bernstein polynomials as follows:*

$$\mathrm{B}_{\mathbf{k}}^m(\mathbf{X}) = \frac{m!}{k_1! \dots k_{n-1}!(m - k_1 \cdots - k_{n-1})!} X_1^{k_1} \dots X_{n-1}^{k_{n-1}} (1 - X_1 \cdots - X_{n-1})^{m-k_1 \cdots - k_{n-1}} \tag{5.8}$$

This definition is equivalent to (5.7), but it has the advantage of taking its argument $\mathbf{X}$ in the $(n-1)$-box $[0,1]^{n-1}$ rather than in the $(n-1)$-simplex Δ_{n-1} [96].

Remark 5.2.2. *The usual univariate Bernstein polynomials are retrieved as follows:*

$$B_k^m(X) \triangleq \mathcal{B}_{k,m-k}(X, 1-X) = \binom{m}{k, m-k} X^k (1-X)^{m-k} = \binom{m}{k} X^k (1-X)^{m-k} \tag{5.9}$$

Remark 5.2.3. *The literature sometimes defines the multivariate Bernstein polynomials as a simple product of several univariate Bernstein polynomials, with* $\mathbf{X} \in [0,1]^p$ *[96].*

$$B_{\mathbf{k}}^{\mathbf{m}}(\mathbf{X}) \triangleq B_{k_1}^{m_1}(X_1) \dots B_{k_p}^{m_p}(X_p) \tag{5.10}$$

This product-based generalization is however ill-suited to the study of T-S systems which are not obtained from a box-based NLSA.

Remark 5.2.4. *The product-based generalization discussed in the previous remark can also be applied to the multivariate Bernstein polynomials as they are defined in this manuscript, with* $(\mathbf{X}_1, \dots, \mathbf{X}_p) \in \Delta_{n_1-1} \times \dots \times \Delta_{n_p-1}$.

$$\mathscr{B}_{\mathbf{k}_1,\dots,\mathbf{k}_p}(\mathbf{X}_1, \dots, \mathbf{X}_p) \triangleq \mathcal{B}_{\mathbf{k}_1}(\mathbf{X}_1) \dots \mathcal{B}_{\mathbf{k}_p}(\mathbf{X}_p) \tag{5.11}$$

This multi-multivariate generalization is particularly well-suited to the study of tensor product models, although this is not discussed in this manuscript.

The Bernstein polynomials of degree m appear naturally in the development of the polynomial $(X_1 + \dots + X_n)^m$. This result is also known as the multinomial theorem.

Theorem 5.2.1 (Multinomial theorem [264])**.** *For all* $\mathbf{X} \in \mathbb{R}^n$*:*

$$\left(\sum_{k=1}^{n} X_k \right)^m = \sum_{\mathbf{k} \in \mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \tag{5.12}$$

Proof. The proof is done by induction on n, using $(X_1 + \dots + X_{n-1} + X_n)^m = (X_1 + \dots + (X_{n-1} + X_n))^n$ and applying the usual binomial theorem on the $(X_{n-1} + X_n)^k$ terms at the induction step. □

They are as many multivariate Bernstein polynomial of degree m as they are elements in $\mathbb{N}_m^n$, a number given by counting the weak n-composition of m, i.e. by counting all the ways of adding $k_1, \dots, k_n \in \mathbb{N}$ such that $k_1 + \dots + k_n = m$. The following property justifies the combinatorial expression (5.6) in the introduction to this chapter.

Property 5.2.1 (Cardinal of $\mathbb{N}_m^n$)**.** *For all* $n, m \in \mathbb{N}_{>0}$*:*

$$\#\mathbb{N}_m^n = \binom{m+n-1}{n-1} = \binom{m+n-1}{m} = \frac{(m+n-1)!}{m!(n-1)!} \tag{5.13}$$

Proof. See the *stars and bars* proof in [263] (page 26). □

Moreover, it is now demonstrated that the collection of Bernstein polynomials of degree m satisfies the convex sum properties on a simplex (Definition 2.1.1 of Chapter 2). This guarantees (5.5) to be a convex sum.

Property 5.2.2 (Convex sum properties). *For all* $\mathbf{X} \in \Delta_{n-1}$*:*

$$\forall \mathbf{k} \in \mathbb{N}_m^n : \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \geq 0 \qquad [\textit{positivity}] \tag{5.14a}$$

$$\sum_{\mathbf{k} \in \mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) = 1 \qquad [\textit{partition of unity}] \tag{5.14b}$$

Proof. The positivity follows from the definitions of Δ_{n-1} and of the multivariate Bernstein polynomials. The multinomial theorem (Theorem 5.2.1) provides the partition of unity, as for all $\mathbf{X} \in \Delta_{n-1}$:

$$\sum_{\mathbf{k} \in \mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) = (X_1 + \cdots + X_n)^m = 1 \tag{5.15}$$

□

It is also noticed that the Bernstein polynomials are homogeneous polynomials. In fact, the collection of Bernstein polynomials of degree m contains exactly all the monomials of degree m. These monomials form a basis to the vector-space of homogeneous polynomial of degree m, and they are scaled so the multinomial theorem (Theorem 5.2.1) holds, guaranteeing that the convex sum properties discussed above are verified.

Property 5.2.3 (Homogeneity). *For all* $\mathbf{k} \in \mathbb{N}_m^n$, $\mathbf{X} \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$*:*

$$\mathcal{B}_{\mathbf{k}}(\lambda \mathbf{X}) = \lambda^m \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \tag{5.16}$$

Proof. The proof is straightforward, as $\mathcal{B}_{\mathbf{k}}(\lambda \mathbf{X}) = \frac{m!}{k_1! \ldots k_n!} \lambda^{k_1 + \cdots + k_n} X_1^{k_1} \ldots X_n^{k_n}$. □

Finally, the maximum and minimum values of the Bernstein polynomials with n variables are explicitly provided on the $(n-1)$-simplex Δ_{n-1}.

Property 5.2.4 (Bounds on Δ_{n-1}). *For all* $\mathbf{k} \in \mathbb{N}_m^n \setminus \{0\}$*:*

$$\min_{\mathbf{X} \in \Delta_{n-1}} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) = 0 \tag{5.17a}$$

$$\max_{\mathbf{X} \in \Delta_{n-1}} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) = \mathcal{B}_{\mathbf{k}}\left(\frac{k_1}{m}, \ldots, \frac{k_n}{m}\right) = \binom{m}{\mathbf{k}} \frac{k_1^{k_1} \ldots k_n^{k_n}}{m^m} \tag{5.17b}$$

Proof. $\boxed{\min}$ Taking $i \in [\![1, n]\!]$ such that $k_i \neq 0$ yields $\mathcal{B}(\dots, X_{i-1}, 0, X_{i+1}, \dots) = 0$. Together with the positivity of the Bernstein polynomials for $\mathbf{X} \in \Delta_{n-1}$, this demonstrates the minimum value.

$\boxed{\max}$ Now if there exists $i \in [\![1, n]\!]$ such that $k_i = 0$, it is clear that for all $\mathbf{X} \in \Delta_{n-1}$:

$$\mathcal{B}_{\mathbf{k}}(\mathbf{X}) = \mathcal{B}_{\mathbf{k}}(\dots, X_{i-1}, 0, X_{i+1}, \dots) = \mathcal{B}_{\dots,k_{i-1},k_{i+1},\dots}(\dots, X_{i-1}, X_{i+1}, \dots) \tag{5.18}$$

which implies $\max_{\mathbf{X}\in\Delta_{n-1}} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) = \max_{\mathbf{Y}\in\Delta_{n-2}} \mathcal{B}_{\dots,k_{i-1},k_{i+1},\dots}(\mathbf{Y})$. Without loss of generality, one can therefore consider that for all $i \in [\![1, n]\!]$, $k_i \neq 0$. At each point of the boundary $\mathbf{X} \in \partial\Delta_{n-1}$, there exists a $i \in [\![1, n]\!]$ such that $X_i = 0$, hence for all $\mathbf{X} \in \partial\Delta_{n-1}$, $\mathcal{B}_{\mathbf{k}}(\mathbf{X}) = 0$. The Bernstein polynomials being smooth, positive and not identically zero in the convex set Δ_{n-1}, their maximum in Δ_{n-1} must be obtained in the interior of the simplex, i.e. in $\mathrm{intr}(\Delta_{n-1})$. The following Lagrangian is introduced:

$$\mathcal{L}(\mathbf{X}, \lambda) = \mathcal{B}_{\mathbf{k}}(\mathbf{X}) + \lambda(1 - X_1 - \cdots - X_n) \tag{5.19}$$

The first order Karush–Kuhn–Tucker conditions [103] provide for all $\mathbf{X} \in \mathrm{intr}(\Delta_{n-1})$:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial(\mathbf{X}, \lambda)}(\mathbf{X}, \lambda) = 0 &\Leftrightarrow \begin{cases} \sum_{i=1}^n X_i = 1 \\ \frac{\partial}{\partial \mathbf{X}} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) = (\lambda, \dots, \lambda)^\top \end{cases} \\
&\Leftrightarrow \begin{cases} \sum_{i=1}^n X_i = 1 \\ \lambda = k_1 X_1^{k_1-1} X_2^{k_2} \dots X_n^{k_n} = \cdots = k_n X_1^{k_1} \dots X_{n-1}^{k_{n-1}} X_n^{k_n-1} \end{cases} \\
&\Leftrightarrow \begin{cases} \sum_{i=1}^n X_i = 1 \\ \frac{k_1}{X_1} = \cdots = \frac{k_n}{X_n} \end{cases} \\
&\Leftrightarrow (X_1, \dots, X_n) = \left(\frac{k_1}{m}, \dots, \frac{k_n}{m}\right)
\end{aligned} \tag{5.20}$$

Since there exists a maximum to $\mathcal{B}_{\mathbf{k}}$ inside of $\mathrm{intr}(\Delta_{n-1})$, and since any maximum of $\mathcal{B}_{\mathbf{k}}$ inside of $\mathrm{intr}(\Delta_{n-1})$ should respect the first order Karush–Kuhn–Tucker conditions, the maximum of $\mathcal{B}_{\mathbf{k}}$ in Δ_{n-1} is necessarily obtained at $\left(\frac{k_1}{m}, \dots, \frac{k_n}{m}\right)$. □

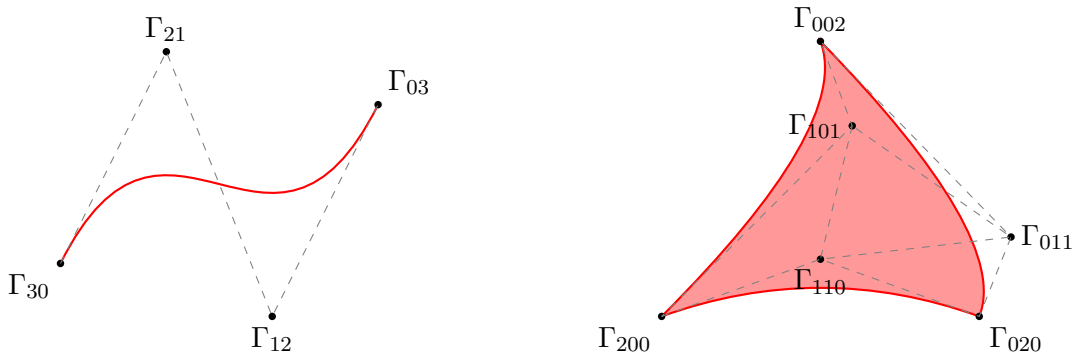
5.2.2 Bézier interpolation schemes

A Bézier interpolation scheme relies on weighting *control points* $\{\Gamma_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{N}^n_m}$ of a given space by the multivariate Bernstein polynomials of degree m, with parameters $\mathbf{X} \triangleq (X_1, \dots, X_n)$ belonging to the $(n-1)$-simplex Δ_{n-1}. The definition of a Bézier interpolation, provided thereafter, should be reminiscent of the convex sum (5.5) introduced at the beginning of this chapter: it is in fact strictly the same expression once the variables $\mathbf{X}$ are replaced with the activation functions $\mathbf{h}$ of a T-S model.

Definition 5.2.2 (Bézier interpolation [90])**.** *Let $n, m \in \mathbb{N}_{>0}$. The Bézier interpolation of $(\Gamma_{\mathbf{k}})_{\mathbf{k}\in\mathbb{N}^n_m}$ at $\mathbf{X} \in \Delta_{n-1}$ is given by:*

$$\Gamma(\mathbf{X}) \triangleq \sum_{\mathbf{k}\in\mathbb{N}^n_m} \mathcal{B}_{\mathbf{k}}(\mathbf{X})\Gamma_{\mathbf{k}} \tag{5.21}$$

and the set $\Gamma(\Delta_{n-1})$ is called a Bézier simplex.



(a) A cubic Bézier curve.

(b) A quadratic Bézier triangle.

Figure 5.1: Illustration of two Bézier simplices.

Example 5.2.1 (Bézier curve). *For $n = 2$, the multivariate Bernstein polynomials in the definition above can be replaced with the usual univariate Bernstein polynomials, and $\Gamma([0,1])$ is called a Bézier curve. In particular, for $n = 2$ and $m = 1$, the Bézier interpolation is simply a linear interpolation between two points:*

$$\Gamma(X) = X\Gamma_{10} + (1-X)\Gamma_{01} \tag{5.22}$$

Likewise, for $n = 2$ and $m = 2$, a quadratic Bézier curve is obtained:

$$\Gamma(X) = X^2\Gamma_{20} + 2X(1-X)\Gamma_{11} + (1-X)^2\Gamma_{02} \tag{5.23}$$

and for $n = 2$ and $m = 3$, a cubic Bézier curve is obtained:

$$\Gamma(X) = X^3\Gamma_{30} + 3X^2(1-X)\Gamma_{21} + 3X(1-X)^2\Gamma_{12} + (1-X)^3\Gamma_{03} \tag{5.24}$$

This latter case is illustrated in Figure 5.1a.

Example 5.2.2 (Bézier triangle). *For $n = 3$, $\Gamma(\Delta_2)$ is called a Bézier triangle. In particular, for $n = 3$ and $m = 1$, the Bézier interpolation defines a regular flat triangle:*

$$\Gamma(X_1, X_2, X_3) = X_1\Gamma_{100} + X_2\Gamma_{010} + X_3\Gamma_{001} \tag{5.25}$$

Likewise, for $n = 3$ and $m = 2$, a quadratic Bézier triangle is obtained:

$$\Gamma(X_1, X_2, X_3) = X_1^2\Gamma_{200} + 2X_1X_2\Gamma_{110} + X_2^2\Gamma_{020} + 2X_2X_3\Gamma_{011} + X_3^2\Gamma_{002} + 2X_3X_1\Gamma_{101} \tag{5.26}$$

This latter case is illustrated in Figure 5.1b.

Remark 5.2.5. *These Bézier interpolations are sometimes directly called Bernstein polynomials by the mathematical literature [91, 96]. However, this disregards how they are referred to in more applied fields (e.g. computer graphics, robotics...) [89, 90].*

Remark 5.2.6. *Bézier surfaces are often defined using the product-based generalization of Bézier polynomials $B^{\mathbf{m}}_{\mathbf{k}}(\mathbf{X})$ (Remark 5.2.3) rather than the multivariate Bernstein polynomials as they are defined in this manuscript [91, 96]. Again, this is ill-suited to the study of T-S systems which are not obtained from a box-based NLSA. However, the multi-multivariate polynomials $\mathscr{B}_{\mathbf{k}_1,\dots,\mathbf{k}_p}(\mathbf{X}_1,\dots,\mathbf{X}_p)$ (Remark 5.2.4) are well-suited to the study of tensor product models, although this is again not discussed in this manuscript.*

The Bézier interpolations schemes are also homogeneous expressions, by linearity with respect to the Bernstein polynomials.

Property 5.2.5 (Homogeneity). *For all* $\lambda \in \mathbb{R}$ *and* $\mathbf{X} \in \mathbb{R}^n$*:*

$$\Gamma(\lambda \mathbf{X}) = \sum_{\mathbf{k} \in \mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\lambda \mathbf{X}) \Gamma_{\mathbf{k}} = \sum_{\mathbf{k} \in \mathbb{N}_m^n} \lambda^m \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \Gamma_{\mathbf{k}} = \lambda^m \Gamma(\mathbf{X}) \tag{5.27}$$

Although this is not discussed in this manuscript, it has already been showed that these Bézier interpolation schemes can approximate any continuous function f defined on Δ_{n-1} along with its partial derivatives (provided that they are continuously defined). The Bézier interpolation schemes are converging in a pointwise fashion as $m \to +\infty$ [91, 96]. This result could be useful in the future, for example in order to demonstrate the asymptotic capabilities in terms of stability and stabilization of the results relying on generalized MQLF, as well as on generalized PDC and nPDC control schemes. The pointwise convergence however leads to pessimism, and a uniform convergence result would be more encouraging. This is left as an open question in this manuscript.

Any Bézier interpolation of order $m_1 \in \mathbb{N}_{>0}$ can also be obtained from a Bézier interpolation of order $m_2 \in \mathbb{N}_{>m_1}$. This is sometimes called the degree elevation property of the Bézier interpolation schemes [89]. The trick simply relies on pre-multiplying the interpolating sum (5.21) by $1 = (X_1 + \cdots + X_n)^{m_2 - m_1}$.

Lemma 5.2.1 (Degree elevation). *For all* $\mathbf{X} \in \Delta_{n-1}$*:*

$$\sum_{\mathbf{k} \in \mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \Gamma_{\mathbf{k}} = \left(\sum_{j=1}^n X_j \right) \left(\sum_{\mathbf{k} \in \mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \Gamma_{\mathbf{k}} \right) = \sum_{\mathbf{k} \in \mathbb{N}_{m+1}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{j=1}^n \frac{k_j}{m+1} \Gamma_{\mathbf{k} - \mathbf{1}_j} \tag{5.28}$$

with $\mathbf{1}_j \triangleq (0, \ldots, 0, \underbrace{1}_{j\text{-th coordinate}}, 0, \ldots, 0)$.

Proof. See Section 1.4 of [89]. This is also a special case of the interpolation elevation demonstrated later in Lemma 5.2.2. □

Theorem 5.2.2 (Generalized degree elevation). *Let $p \in \mathbb{N}_{\geq 2}$. For all $\mathbf{X} \in \Delta_{n-1}$:*

$$\begin{aligned}\sum_{\mathbf{k}\in\mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X})\Gamma_{\mathbf{k}} &= \left(\sum_{j=1}^{n} X_j\right)^p \left(\sum_{\mathbf{k}\in\mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X})\Gamma_{\mathbf{k}}\right) \\ &= \sum_{\mathbf{k}\in\mathbb{N}_{m+p}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \frac{m!}{(m+p)!} \sum_{\mathbf{j}\in[\![1,n]\!]^p} k_{j_p} \left(\prod_{i=1}^{p-1} \left(k_{j_i} - \sum_{r=i+1}^{p} \delta_{j_i,j_r}\right)\right) \Gamma_{\mathbf{k}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_p}}\end{aligned} \tag{5.29}$$

with $\delta_{i,j}$ the Kronecker delta, defined by:

$$\delta_{ij} \triangleq \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{5.30}$$

Proof. The proof consists in applying p times Lemma 5.2.1:

$$\begin{aligned}&\left(\sum_{i=1}^{n} X_i\right)^p \left(\sum_{\mathbf{k}\in\mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X})\Gamma_{\mathbf{k}}\right) = \left(\sum_{i=1}^{n} X_i\right)^{p-1} \left(\sum_{\mathbf{k}\in\mathbb{N}_{m+1}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{j_1=1}^{n} \frac{k_{j_1}}{m+1}\Gamma_{\mathbf{k}-\mathbf{1}_{j_1}}\right) \\ &= \left(\sum_{i=1}^{n} X_i\right)^{p-2} \left(\sum_{\mathbf{k}\in\mathbb{N}_{m+2}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} \frac{(k_{j_1} - \delta_{j_1,j_2})k_{j_2}}{(m+1)(m+2)}\Gamma_{\mathbf{k}-\mathbf{1}_{j_1}-\mathbf{1}_{j_2}}\right) \\ &= \left(\sum_{i=1}^{n} X_i\right)^{p-3} \left(\sum_{\mathbf{k}\in\mathbb{N}_{m+2}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{\mathbf{j}\in[\![1,n]\!]^3} \frac{(k_{j_1} - \delta_{j_1,j_2} - \delta_{j_1,j_3})(k_{j_2} - \delta_{j_2,j_3})k_{j_3}}{(m+1)(m+2)(m+3)}\Gamma_{\mathbf{k}-\mathbf{1}_{j_1}-\mathbf{1}_{j_2}-\mathbf{1}_{j_3}}\right) \\ &= \dots \\ &= \left(\sum_{i=1}^{n} X_i\right)^{2} \sum_{\mathbf{k}\in\mathbb{N}_{m+p-2}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{\mathbf{j}\in[\![1,n]\!]^{p-2}} \frac{(k_{j_1} - \delta_{j_1,j_2}\cdots - \delta_{j_1,j_{p-2}})\dots k_{j_{p-2}}}{(m+1)\dots(m+p-2)}\Gamma_{\mathbf{k}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_{p-2}}} \\ &= \left(\sum_{i=1}^{n} X_i\right) \sum_{\mathbf{k}\in\mathbb{N}_{m+p-1}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{\mathbf{j}\in[\![1,n]\!]^{p-1}} \frac{(k_{j_1} - \delta_{j_1,j_2}\cdots - \delta_{j_1,j_{p-1}})\dots k_{j_{p-1}}}{(m+1)\dots(m+p-1)}\Gamma_{\mathbf{k}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_{p-1}}} \\ &= \sum_{\mathbf{k}\in\mathbb{N}_{m+p}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{\mathbf{j}\in[\![1,n]\!]^{p}} \frac{(k_{j_1} - \delta_{j_1,j_2}\cdots - \delta_{j_1,j_p})(k_{j_2} - \delta_{j_2,j_3}\cdots - \delta_{j_2,j_p})\dots k_{j_p}}{(m+1)\dots(m+p)}\Gamma_{\mathbf{k}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_p}} \\ &= \sum_{\mathbf{k}\in\mathbb{N}_{m+p}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \frac{m!}{(m+p)!} \sum_{\mathbf{j}\in[\![1,n]\!]^p} k_{j_p} \left(\prod_{i=1}^{p-1} \left(k_{j_i} - \sum_{r=i+1}^{p} \delta_{j_i,j_r}\right)\right) \Gamma_{\mathbf{k}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_p}}\end{aligned} \tag{5.31}$$

□

The degree elevation property described above is now considered with the addition of new control points in the Bézier interpolation scheme. The name *interpolation elevation* is suggested for these new identities, as this time the number of independent control points increases with the degree of the interpolation.

Lemma 5.2.2 (Interpolation elevation). *Given the sets of control points* $(\Gamma_{\mathbf{k},j})$ *with* $\mathbf{k} \in \mathbb{N}_m^n$ *and* $j \in [\![1, n]\!]$*, for all* $\mathbf{X} \in \Delta_{n-1}$*:*

$$\sum_{\mathbf{k}\in\mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \left(\sum_{j=1}^{n} X_j \Gamma_{\mathbf{k},j} \right) = \sum_{\mathbf{k}\in\mathbb{N}_{m+1}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{j=1}^{n} \frac{k_j}{m+1} \Gamma_{\mathbf{k}-\mathbf{1}_j,j} \tag{5.32}$$

Proof. Given $\mathbf{k} \in \mathbb{N}^n$, if there exists $i \in [\![1, n]\!]$ such that $k_i < 0$, then it is considered that $\binom{m}{\mathbf{k}} = 0$. The following equalities stand:

$$\begin{aligned}
\sum_{\mathbf{k}\in\mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \left(\sum_{j=1}^{n} X_j \Gamma_{\mathbf{k},j} \right) &= \sum_{\mathbf{k}\in\mathbb{N}_m^n} \binom{m}{\mathbf{k}} \mathbf{X}^{\mathbf{k}} \left(\sum_{j=1}^{n} X_j \Gamma_{\mathbf{k},j} \right) \\
&= \sum_{\mathbf{k}\in\mathbb{N}_m^n} \sum_{j=1}^{n} \binom{m}{\mathbf{k}} \mathbf{X}^{\mathbf{k}+\mathbf{1}_j} \Gamma_{\mathbf{k},j} \\
&= \sum_{\mathbf{k}\in\mathbb{N}_{m+1}^n} \sum_{j=1}^{n} \binom{m}{\mathbf{k}-\mathbf{1}_j} \mathbf{X}^{\mathbf{k}} \Gamma_{\mathbf{k}-\mathbf{1}_j,j} \\
&= \sum_{\mathbf{k}\in\mathbb{N}_{m+1}^n} \binom{m+1}{\mathbf{k}} \mathbf{X}^{\mathbf{k}} \sum_{j=1}^{n} \binom{m}{\mathbf{k}-\mathbf{1}_j} \binom{m+1}{\mathbf{k}}^{-1} \Gamma_{\mathbf{k}-\mathbf{1}_j,j} \\
&= \sum_{\mathbf{k}\in\mathbb{N}_{m+1}^n} \binom{m+1}{\mathbf{k}} \mathbf{X}^{\mathbf{k}} \sum_{j=1}^{n} \frac{k_j}{m+1} \Gamma_{\mathbf{k}-\mathbf{1}_j,j} \\
&= \sum_{\mathbf{k}\in\mathbb{N}_{m+1}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \sum_{j=1}^{n} \frac{k_j}{m+1} \Gamma_{\mathbf{k}-\mathbf{1}_j,j}
\end{aligned} \tag{5.33}$$

□

Contrary to the degree elevation property, finding a practical closed-form expression to the generalization of the interpolation elevation is not straightforward and is left as an open question of this manuscript.

$$\sum_{\mathbf{k}\in\mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \left(\sum_{\mathbf{j}\in\mathbb{N}_p^n} \mathcal{B}_{\mathbf{j}}(\mathbf{X}) \Gamma_{\mathbf{k},\mathbf{j}} \right) = \sum_{\mathbf{k}\in\mathbb{N}_{m+p}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \ldots ? \ldots \tag{5.34}$$

This expression is however easily found by using a multi-sum rather than a Bézier interpolation in one of the two sums. However, this has the disadvantage of introducing many redundant terms in the multi-sum.

Theorem 5.2.3 ((Almost) generalized interpolation elevation). *Let $p \in \mathbb{N}_{\geq 2}$. Given the sets of control points $(\Gamma_{\mathbf{k},\mathbf{j}})$ with $\mathbf{k} \in \mathbb{N}_m^n$ and $\mathbf{j} \in [\![1,n]\!]^p$, for all $\mathbf{X} \in \Delta_{n-1}$:*

$$\begin{aligned}&\sum_{\mathbf{k}\in\mathbb{N}_m^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \left(\sum_{\mathbf{j}\in[\![1,n]\!]^p} \left(\prod_{i=1}^{p} X_{j_i} \right) \Gamma_{\mathbf{k},\mathbf{j}} \right) = \\ &\sum_{\mathbf{k}\in\mathbb{N}_{m+p}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \frac{m!}{(m+p)!} \sum_{\mathbf{j}\in[\![1,n]\!]^p} k_{j_p} \left(\prod_{i=1}^{p-1} \left(k_{j_i} - \sum_{r=i+1}^{p} \delta_{j_i,j_r} \right) \right) \Gamma_{\mathbf{k}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_p},\mathbf{j}}\end{aligned} \tag{5.35}$$

Proof. The proof follows the same steps as Theorem 5.2.2, by applying p times Lemma 5.2.2. □

5.3 Applications in the Takagi-Sugeno framework

The Bézier interpolation scheme introduces *explicit* expressions to the non-redundant multi-sums of the T-S framework, at places where a rewriting step would otherwise be necessary to ignore the redundant terms. In order to illustrate the benefits of a Bézier sum compared to the usual naive multi-sum, the explicit expression of the LMI conditions for multi-sums found in [238] is provided in Section 5.3.1. Moreover, the idea of Bézier-T-S models is presented in Section 5.3.2. Finally, the generalized PDC and nPDC schemes [11, 165] are rewritten as Bézier-PDC and Bézier-nPDC schemes in Section 5.3.3. Overall, the accumulation of these results should be seen as arguments in favor of the adoption of a Bézier formalism in the T-S framework.

5.3.1 LMI results on multiple convex sums

As previously discussed in Chapter 2, the multi-sums, ubiquitous in the T-S framework, often lead to the study of matrix inequality problems such as finding $z \in \mathbb{R}^m$ satisfying:

$$\sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) F_{\mathbf{i}}(z) \prec 0, \qquad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{5.36}$$

where the case $m = 2$ corresponds to the double convex sum LMI problem discussed in Section 2.2.2. Based on the work of [238], a procedure is suggested hereafter to obtain a set of tractable LMI conditions from the inequality (5.36). This procedure consists in increasing the number of terms in the convex sum (5.36) using the generalized degree elevation of a Bézier interpolation (Theorem 5.2.2). This eventually provides necessary and sufficient LMI conditions.

Theorem 5.3.1 (Multiple sum relaxation). *Let $p \in \mathbb{N}_{\geq 2}$. The inequality (5.36) is satisfied if the following conditions hold:*

$$\sum_{\mathbf{j}\in[\![1,n_h]\!]^p} i_{j_p} \left(\prod_{k=1}^{p-1} \left(i_{j_k} - \sum_{r=k+1}^{p} \delta_{j_k,j_r} \right) \right) F_{\mathbf{i}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_p}}(z) \prec 0, \quad \forall \mathbf{i} \in \mathbb{N}_{m+p}^{n_h} \tag{5.37}$$

If p is sufficiently large, the conditions (5.37) are equivalent to (5.36).

Proof. The degree elevation trick (Theorem 5.2.2) provides:

$$\sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h})F_{\mathbf{i}} = \sum_{\mathbf{i}\in\mathbb{N}_{m+p}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h})\frac{m!}{(m+p)!}\sum_{\mathbf{j}\in[\![1,n_h]\!]^p} i_{j_p}\left(\prod_{k=1}^{p-1}\left(i_{j_k}-\sum_{r=k+1}^{p}\delta_{j_k,j_r}\right)\right)F_{\mathbf{i}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_p}} \tag{5.38}$$

Moreover, as the Bernstein polynomials satisfy the convex sum properties (Property 5.2.2), (5.36) is satisfied if the following conditions hold:

$$\frac{m!}{(m+p)!}\sum_{\mathbf{j}\in[\![1,n_h]\!]^p} i_{j_p}\left(\prod_{k=1}^{p-1}\left(i_{j_k}-\sum_{r=k+1}^{p}\delta_{j_k,j_r}\right)\right)F_{\mathbf{i}-\mathbf{1}_{j_1}\cdots-\mathbf{1}_{j_p}}(z) \prec 0, \quad \forall \mathbf{i}\in\mathbb{N}_{m+p}^{n_h} \tag{5.39}$$

Since $\mathbb{S}_{n_x}^{--}(\mathbb{R})$ is a cone, the scaling by $\frac{m!}{(m+p)!}$ is superfluous in the conditions above, leading to the LMI conditions (5.37). Finally, the proof that these conditions become necessary and sufficient for a sufficiently large p is found in [238]. □

Remark 5.3.1. *The case $p=0$ and $p=1$ in the conditions (5.37) are respectively given by:*

$$F_{\mathbf{i}}(z) \prec 0, \qquad \forall \mathbf{i}\in\mathbb{N}_m^{n_h} \quad (\textit{for } p=0) \tag{5.40a}$$

$$\sum_{j=1}^{n_h} i_j F_{\mathbf{i}-\mathbf{1}_j}(z) \prec 0, \qquad \forall \mathbf{i}\in\mathbb{N}_{m+1}^{n_h} \quad (\textit{for } p=1) \tag{5.40b}$$

The relative novelty of Theorem (5.3.1) compared to [238] is two-folded:

- it is now clear that this relaxation relies on the generalized degree elevation of a Bézier interpolation scheme (Theorem 5.2.2);
- an *explicit* expression of the relaxed LMI conditions is obtained.

Example 5.3.1. *Considering the matrix inequality (5.36) for $n_h=2$ and $m\in\mathbb{N}_{>0}$. Theorem 5.3.1 provides the following tractable LMI conditions for $p\in\{0,1,2,3\}$.*
For $p=0$, the LMI conditions are given by:

$$F_{i,m-i} \prec 0, \qquad i=0,\dots,m \tag{5.41}$$

For $p=1$, the LMI conditions are given by:

$$F_{0,m} \prec 0 \tag{5.42a}$$

$$iF_{i-1,m+1-i} + (m+1-i)F_{i,m-i} \prec 0, \qquad i=1,\dots,m \tag{5.42b}$$

$$F_{m,0} \prec 0 \tag{5.42c}$$

For $p=2$, the LMI conditions are given by:

$$F_{0,m} \prec 0 \tag{5.43a}$$

$$2(m+1)F_{0,m} + (m+1)mF_{1,m-1} \prec 0 \tag{5.43b}$$

$$i(i-1)F_{i-2,m+2-i} + 2i(m+2-i)F_{i-1,m+1-i} + (m+2-i)(m+1-i)F_{i,m-i} \prec 0 \tag{5.43c}$$

$$(m+1)mF_{m-1,1} + 2(m+1)F_{m,0} \prec 0 \tag{5.43d}$$

$$F_{m,0} \prec 0 \tag{5.43e}$$

with $i = 2, \ldots, m$ *in* (5.43c).
For $p = 3$, *the LMI conditions are given by:*

$$F_{0,m} \prec 0 \tag{5.44a}$$
$$3(m+2)(m+1)F_{0,m} + (m+2)(m+1)mF_{1,m-1} \prec 0 \tag{5.44b}$$
$$6(m+1)F_{0,m} + 6(m+1)mF_{1,m-1} + (m+1)m(m-1)F_{2,m-2} \prec 0 \tag{5.44c}$$
$$\begin{aligned} &i(i-1)(i-2)F_{i-3,m+3-i} + 3i(i-1)(m+3-i)F_{i-2,m+2-i} + \ldots \\ &3i(m+3-i)(m+2-i)F_{i-1,m+1-i} + (m+3-i)(m+2-i)(m+1-i)F_{i,m-i} \prec 0 \end{aligned} \tag{5.44d}$$
$$(m+1)m(m-1)F_{m-2,2} + 6(m+1)mF_{m-1,1} + 6(m+1)F_{m,0} \prec 0 \tag{5.44e}$$
$$(m+2)(m+1)mF_{m-1,1} + 3(m+2)(m+1)F_{m,0} \prec 0 \tag{5.44f}$$
$$F_{m,0} \prec 0 \tag{5.44g}$$

with $i = 3, \ldots, m$ *in* (5.44d).

5.3.2 Bézier modeling

The *Bézier-T-S model* (5.45) is proposed in this section.

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ e(t) \end{pmatrix} = \sum_{\mathbf{i} \in \mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) \begin{pmatrix} A_{\mathbf{i}} & B_{1,\mathbf{i}} & B_{2,\mathbf{i}} \\ C_{1,\mathbf{i}} & D_{11,\mathbf{i}} & D_{12,\mathbf{i}} \\ C_{2,\mathbf{i}} & D_{21,\mathbf{i}} & D_{22,\mathbf{i}} \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \tag{5.45}$$

Contrary to a typical T-S model which is essentially a *polytopic* (hence convex) LDI equipped with a scheduling vector θ; the Bézier-T-S model suggested hereabove is essentially a *Bézier-simplex* LDI equipped with a scheduling vector θ. The Bézier-simplex LDI associated with the Bézier-T-S model (5.45) is introduced in (5.46), where the set on the right-hand side is a Bézier-simplex, which is not guaranteed to be convex.

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ e(t) \end{pmatrix} \in \left\{ \sum_{\mathbf{i} \in \mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \begin{pmatrix} A_{\mathbf{i}} & B_{1,\mathbf{i}} & B_{2,\mathbf{i}} \\ C_{1,\mathbf{i}} & D_{11,\mathbf{i}} & D_{12,\mathbf{i}} \\ C_{2,\mathbf{i}} & D_{21,\mathbf{i}} & D_{22,\mathbf{i}} \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} : \mathbf{h} \in \Delta_{n_h - 1} \right\} \tag{5.46}$$

To the author's knowledge, both the Bézier-T-S model (5.45) and the Bézier-simplex LDI (5.46) are not yet studied by the literature. However, it could be interesting to investigate them for the following reasons:

- the intrinsic conservatism of the model could be reduced compared to the use of a standard T-S model, by precisely taking into account the interdependent nonlinearities of the modeled system;
- smoothness and non-convexity are introduced in the representation while still relying on typical convex tools and a finite number of local models;
- for all $m \in \mathbb{N}_{>0}$, only n_h activation functions are required to study a system with $\frac{(m+n_h-1)!}{m!(n_h-1)!}$ local models.

Moreover, most stability and stabilization results of the T-S framework should be easily extrapolated to a Bézier-T-S framework. For example, the simplest quadratic stability result for T-S model (Theorem 2.3.1 of Chapter 2) is generalized thereafter.

Theorem 5.3.2 (Quadratic stability). *The continuous-time Bézier-T-S model*

$$\dot{x}(t) = \sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta))A_{\mathbf{i}}x(t) \tag{5.47}$$

is globally exponentially stable if there exists $P \in \mathbb{S}_{n_x}(\mathbb{R})$ such that the conditions (5.48) are satisfied.

$$P \succ 0 \tag{5.48a}$$

$$\sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h})\mathcal{H}(PA_{\mathbf{i}}) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{5.48b}$$

The inequality on the Bézier sum can be transformed into regular LMI conditions by leveraging Theorem 5.3.1.

Proof. It follows the same steps as the two last proofs of Theorem 2.3.1 in Chapter 2. □

Remark 5.3.2. *The same result holds for the Bézier-simplex LDI associated with (5.47).*

Remark 5.3.3. *This result benefits from the smoothness of the set at the right-hand side of the Bézier-simplex LDI associated with (5.47), but not from its non-convexity. Given a P satisfying (5.48), it also satisfies $\mathcal{H}(PA) \prec 0$ for all $A \in \text{hull}\left\{\sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h})A_{\mathbf{i}} : \mathbf{h} \in \Delta_{n_h-1}\right\}$, hence the curse of convexity applies (Theorem 4.1.1 of Chapter 4).*

The Bézier-T-S representation could be particularly well-suited to exactly represent nonlinear systems with interdependent *polynomials* nonlinearities, as demonstrated by the following example.

Example 5.3.2. *Consider the following nonlinear system:*

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} (x_1^2(t)-1)x_1(t) \\ (x_1(t)-1)x_2(t) \end{pmatrix} \tag{5.49}$$

A Bézier-T-S system which exactly represents (5.49) can be found using the following activation functions, where $\mathbf{h} \in \Delta_1$ for all $\theta \triangleq x_1 \in [-1+\varepsilon, 1-\varepsilon]$, with $\varepsilon \in (0,1)$.

$$h_1(x_1) = \frac{x_1 + 1 - \varepsilon}{2(1-\varepsilon)} \tag{5.50a}$$

$$h_2(x_1) = 1 - h_1(x_1) = \frac{-x_1 + 1 - \varepsilon}{2(1-\varepsilon)} \tag{5.50b}$$

The following equalities stand for all $x_1 \in \mathbb{R}$:

$$x_1^2 - 1 = h_1^2(x_1)\varepsilon(\varepsilon-2) + 2h_2(x_1)h_1(x_1)\left(\varepsilon(\varepsilon-2) - 2(1-\varepsilon)^2\right) + h_2^2(x_1)\varepsilon(\varepsilon-2) \tag{5.51a}$$

$$x_1 - 1 = -h_1^2(x_1)\varepsilon - 2h_2(x_1)h_1(x_1) + h_2^2(x_1)(\varepsilon-2) \tag{5.51b}$$

Hence, the following Bézier-T-S model is an exact representation of (5.49)

$$\dot{x}(t) = \sum_{\mathbf{i}\in\mathbb{N}_2^2} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta))A_{\mathbf{i}}x(t) \tag{5.52}$$

where

$$A_{20} = \begin{pmatrix} \varepsilon(\varepsilon-2) & 0 \\ 0 & -\varepsilon \end{pmatrix} \tag{5.53a}$$

$$A_{11} = \begin{pmatrix} \varepsilon(\varepsilon-2)-2(1-\varepsilon)^2 & 0 \\ 0 & -1 \end{pmatrix} \tag{5.53b}$$

$$A_{02} = \begin{pmatrix} \varepsilon(\varepsilon-2) & 0 \\ 0 & \varepsilon-2 \end{pmatrix} \tag{5.53c}$$

It is easily verified that for all $\lambda \in \mathbb{R}_{>0}$, $V(x) = x_1^2 + \lambda x_2^2$ *is a CQLF to all the local models given hereabove. In particular,* $\lambda = 1$ *demonstrates the local exponential stability of* (5.52) *on the ball* $\mathcal{B}_2(0, 1-\varepsilon)$. *This example is illustrated in Figure 5.2.*

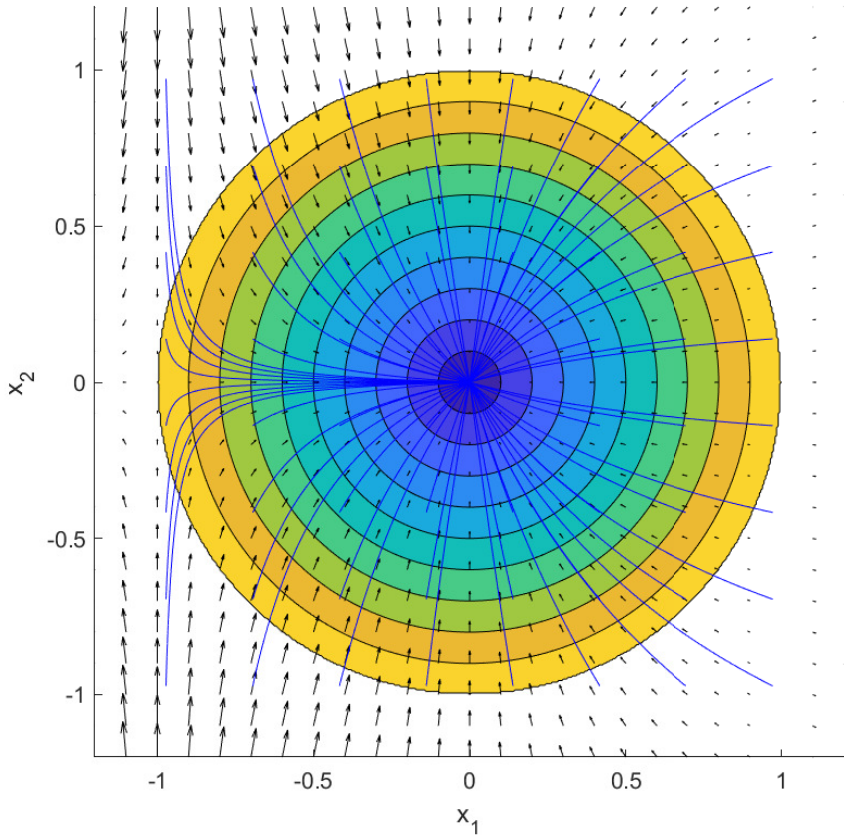


Figure 5.2: Vector field associated with the differential equation (5.49) together with the level sets of the local QLF $V(x) = \|x\|_2^2$, and with some trajectories (in blue) converging towards the origin for a set of evenly spread initial conditions.

Applying a typical polytopic NLSA (as in Chapter 3) to the nonlinear system (5.49) of the example hereabove would yield a T-S model with at least three local models and three activation functions.

$$\dot{x}(t) = \sum_{i=0}^{2} h_{\mathcal{V}_i}(\theta) A_{i,2-i} \tag{5.54}$$

Yet, the volume covered by the polytopic LDI associated with (5.54) would be far too large compared to the set of *truly reachable* local models. Indeed, given the considered interdependent

nonlinearities for $x_1 \in [-1+\varepsilon, 1-\varepsilon]$, the set of truly reachable local models is the following quadratic Bézier curve, a *non-convex and smooth* set of intrinsic dimension 1:

$$\left\{h^2 A_{20} + 2h(1-h)A_{11} + (1-h)^2 A_{02} : h \in [0,1]\right\} \tag{5.55}$$

The Bézier model (5.52) still uses three local models, however, it uses solely two activation functions. Moreover, the right-hand side of its associated Bézier-simplex LDI covers exactly the set (5.55) of truly reachable local models for $x_1 \in [-1+\varepsilon, 1-\varepsilon]$, which is indicative of the reduced intrinsic conservatism of this Bézier modeling technique.

Obtaining a generalized procedure to rewrite a polynomial nonlinear system using an exact Bézier-T-S representation, while utilizing as few activation functions as possible remains an open problem of this manuscript.

5.3.3 Bézier controllers and observers

The simplifying Assumptions 2.4.1, 2.4.2 and 2.4.3 of Chapter 2 are once again considered in all subsequent matters of this chapter. The following continuous-time T-S model is studied in this section:

$$\dot{x}(t) = \sum_{i=1}^{n_h} h_i(\theta)(A_i x(t) + B_i u(t)) \tag{5.56a}$$

$$y(t) = \sum_{i=1}^{n_h} h_i(\theta)(C_i x(t) + D_i u(t)) \tag{5.56b}$$

The objective is to design the gains of the following gain-scheduled state-feedback controller:

$$u(t) = K(\theta)x(t) \tag{5.57}$$

or of the following the gain-scheduled Luenberger observer:

$$\dot{\hat{x}}(t) = \sum_{i=1}^{n_h} h_i(\theta)\left(A_i \hat{x}(t) + B_i u(t)\right) + L(\theta)\left(\hat{y} - y(t)\right) \tag{5.58a}$$

$$\hat{y}(t) = \sum_{i=1}^{n_h} h_i(\theta)\left(C_i \hat{x}(t) + D_i u(t)\right) \tag{5.58b}$$

Bézier interpolation schemes are now applied to the gains $K(\theta)$ and $L(\theta)$. Indeed, the PDC and nPDC gain-scheduling schemes investigated in Section 2.4.2 of Chapter 2 can be generalized into Bézier Parallel Distributed Compensation (BPDC) and Bézier non-Parallel Distributed Compensation (BnPDC) schemes as follows:

– The BPDC scheme consists in a Bézier interpolation of the gain matrices:

$$K(\theta) = \sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta))K_{\mathbf{i}}, \qquad L(\theta) = \sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta))L_{\mathbf{i}} \tag{5.59}$$

– The BnPDC scheme consists in the following interpolations:

$$K(\theta) = \left(\sum_{\mathbf{i} \in \mathbb{N}_{m-1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) K_{\mathbf{i}} \right) \left(\sum_{i=1}^{n_h} h_i(\theta) P_{1,i} \right)^{-1} \tag{5.60a}$$

$$L(\theta) = \left(\sum_{i=1}^{n_h} h_i(\theta) P_{2,i} \right)^{-1} \left(\sum_{\mathbf{i} \in \mathbb{N}_{m-1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) L_{\mathbf{i}} \right) \tag{5.60b}$$

where the matrices $\{P_{1,i}, P_{2,i}\}_{1 \leq i \leq n_h}$ are used in the expression of a nQLF and of a MQLF, depending on the closed-loop state or on the state estimation error.

This section provides the stabilizability conditions allowing for the computation of the sets of gains $\{K_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{N}_m^n}$ or $\{L_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{N}_m^n}$. In the case of the design of an observer-based controller, according to the separation principle, as discussed in Section 2.4 of Chapter 2, the controller (5.57) and the observer (5.58) can be designed independently of each other to obtain the observer-based controller (2.99) combining the two.

Bézier-PDC The following results hold for the BPDC control law:

$$u(t) = \sum_{\mathbf{i} \in \mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) K_{\mathbf{i}} x(t) \tag{5.61}$$

and for its BPDC observer counterpart:

$$\dot{\hat{x}}(t) = \sum_{j=1}^{n_h} h_j(\theta) \left(A_j \hat{x}(t) + B_j u(t) \right) + \sum_{\mathbf{i} \in \mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) L_{\mathbf{i}} (\hat{y}(t) - y(t)) \tag{5.62a}$$

$$\hat{y}(t) = \sum_{j=1}^{n_h} h_j(\theta) \left(C_j \hat{x}(t) + D_j u(t) \right) \tag{5.62b}$$

considered with the usual QLF (1.21), i.e. $V(x) = x^\top P x$.

Theorem 5.3.3 (BPDC controller). *Given $m \in \mathbb{N}$, the system (5.56) is globally exponentially stabilizable using the BPDC control law (5.61) if there exists $X \in \mathbb{S}_{n_x}(\mathbb{R})$ and $\{M_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{N}_m^{n_h}}$, with $M_{\mathbf{i}} \in \mathbb{R}^{n_u \times n_x}$ for all $\mathbf{i} \in \mathbb{N}_m^{n_h}$, such that the conditions (5.63) are satisfied.*

$$X \succ 0 \tag{5.63a}$$

$$\sum_{\mathbf{i} \in \mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \sum_{j=1}^{n_h} \frac{i_j}{m+1} \mathcal{H}(A_j X + B_j M_{\mathbf{i}-\mathbf{1}_j}) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h - 1} \tag{5.63b}$$

The inequality on the Bézier sum can be transformed into regular LMI conditions by leveraging Theorem 5.3.1. The gain matrices are retrieved with $K_{\mathbf{i}} = M_{\mathbf{i}} X^{-1}$, and the matrix P providing the QLF (1.21) is obtained with $P = X^{-1}$.

Proof. The dynamic of the closed-loop system is given by:

$$\begin{aligned}\dot{x}(t) &= \sum_{j=1}^{n_h} h_j(\theta) \left(A_j + \sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) B_j K_{\mathbf{i}} \right) x(t) \\ &= \sum_{j=1}^{n_h} h_j(\theta) \sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) \left(A_j + B_j K_{\mathbf{i}} \right) x(t)\end{aligned} \tag{5.64}$$

The interpolation elevation trick (Lemma 5.2.2) provides:

$$\dot{x}(t) = \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) \sum_{j=1}^{n_h} \frac{i_j}{m+1} (A_j + B_j K_{\mathbf{i}-\mathbf{1}_j}) x(t) \tag{5.65}$$

Theorem 5.3.2 can then be applied to the closed-loop system above, and a final left and right multiplication by P^{-1} concludes the proof by congruence (Property 2.2.2). □

Theorem 5.3.4 (BPDC observer). *Given $m \in \mathbb{N}$, the observation error between the state of (5.56) and the state of the observer (5.62) is globally exponentially stabilizable if there exists $P \in \mathbb{S}_{n_x}(\mathbb{R})$ and $\{N_{\mathbf{i}}\}_{\mathbf{i}\in\mathbb{N}_m^{n_h}}$, with $N_{\mathbf{i}} \in \mathbb{R}^{n_x \times n_y}$ for all $\mathbf{i} \in \mathbb{N}_m^{n_h}$, such that the conditions (5.66) are satisfied.*

$$P \succ 0 \tag{5.66a}$$

$$\sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \sum_{j=1}^{n_h} \frac{i_j}{m+1} \mathcal{H}(P A_j + N_{\mathbf{i}-\mathbf{1}_j} C_j) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{5.66b}$$

The inequality on the Bézier sum can be transformed into regular LMI conditions by leveraging Theorem 5.3.1. The gain matrices are retrieved with $L_{\mathbf{i}} = P^{-1} N_{\mathbf{i}}$, and the matrix P provides the QLF $V(e) = e^\top P e$.

Proof. The dynamic of the estimation error $e(t) = \hat{x}(t) - x(t)$ is given by:

$$\dot{e}(t) = \sum_{j=1}^{n_h} h_j(\theta) \left(A_j + \sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta)) L_{\mathbf{i}} C_j \right) e(t) \tag{5.67}$$

From here, the proof follows the same steps as for Theorem 5.3.3, without the final left and right multiplication by P^{-1}. □

Example 5.3.3. *To illustrate the conservatism reduction brought by the BPDC controller design, the following T-S model (taken from [87, 239]) is considered:*

$$\mathcal{S}_{(a,b)} : \quad \dot{x}(t) = \sum_{i=1}^{3} h_i(\theta)(A_i(a)x(t) + B_i(b)u(t)) \tag{5.68}$$

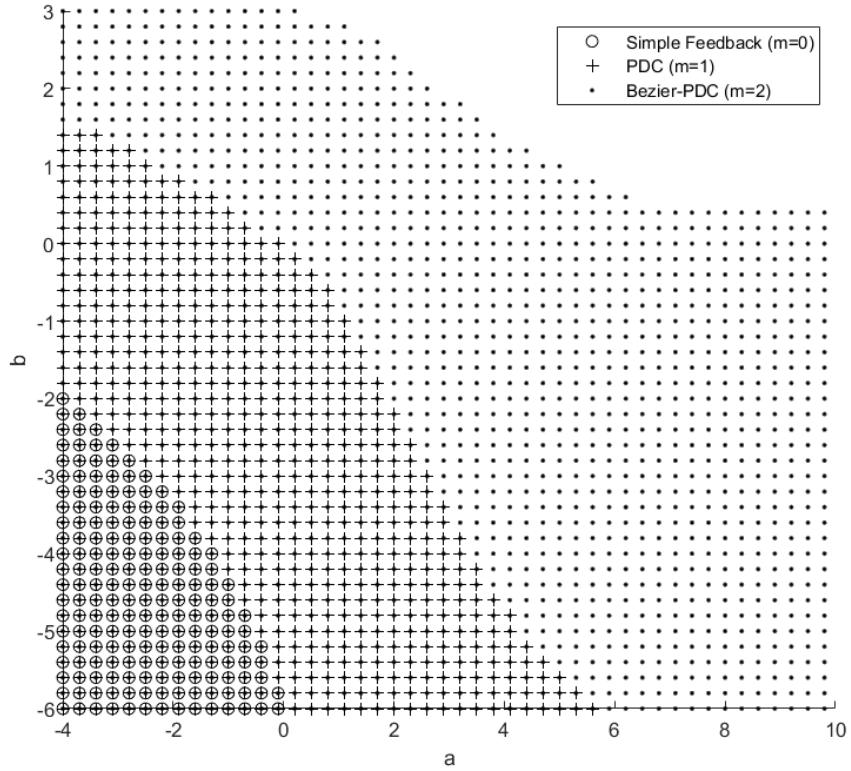


Figure 5.3: Stabilizability (a, b)-regions of $\mathcal{S}_{(a,b)}$ with the BPDC control laws computed with the LMI conditions of Theorem 5.3.3 and Theorem 5.3.1 considered with $p = 0$.

with

$$
\begin{aligned}
A_1 &= \begin{pmatrix} 1.59 & -7.29 \\ 0.01 & 0 \end{pmatrix} \qquad & B_1 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
A_2 &= \begin{pmatrix} 0.02 & -4.64 \\ 0.35 & 0.21 \end{pmatrix} \qquad & B_2 &= \begin{pmatrix} 8 \\ 0 \end{pmatrix} \\
A_3 &= \begin{pmatrix} -a & -4.33 \\ 0 & 0.05 \end{pmatrix} \qquad & B_3 &= \begin{pmatrix} 6-b \\ -1 \end{pmatrix}
\end{aligned}
\tag{5.69}
$$

together with the QLF: $V(x) = x^\top X^{-1} x$*, with a symmetric* $X \in \mathbb{S}_{n_x}^{++}(\mathbb{R})$*. The stabilization problem of* $\mathcal{S}_{(a,b)}$ *is considered at several values of* $(a, b) \in \mathbb{R}^2$ *for the BPDC control laws* $u(t) = K_m(\mathbf{h}(\theta))x(t)$*, with* $m \in [\![0, 2]\!]$ *and*

$$K_0(\mathbf{h}) = K_{000} \tag{5.70a}$$
$$K_1(\mathbf{h}) = h_1 K_{100} + h_2 K_{010} + h_3 K_{001} \tag{5.70b}$$
$$K_2(\mathbf{h}) = h_1^2 K_{200} + h_2^2 K_{020} + h_3^2 K_{002} + 2h_1 h_2 K_{110} + 2h_1 h_3 K_{101} + 2h_2 h_3 K_{011} \tag{5.70c}$$

For $m = 2$*,* $n^m - (m + n - 1)!/m!(n - 1)! = 3$*, hence* 3 *redundant gain matrices have been economized compared to the usual multi-sum approach, which is a total of* 6 *useless decision variables. Theorem 5.3.3 and Theorem 5.3.1 considered with* $p = 0$ *provide the following LMI conditions to com-*

pute the BPDC feedback $K_2(\mathbf{h})$:

$$\forall i \in [\![1,3]\!] : \ \mathcal{H}(A_iX + B_iM_{000+2\cdot\mathbf{1}_i}) \prec 0 \tag{5.71a}$$

$$\frac{1}{3}\mathcal{H}(2[A_1X + B_1M_{110}] + [A_2X + B_2M_{200}]) \prec 0 \tag{5.71b}$$

$$\frac{1}{3}\mathcal{H}(2[A_1X + B_1M_{101}] + [A_3X + B_3M_{200}]) \prec 0 \tag{5.71c}$$

$$\frac{1}{3}\mathcal{H}([A_1X + B_1M_{020}] + 2[A_2X + B_2M_{110}]) \prec 0 \tag{5.71d}$$

$$\frac{1}{3}\mathcal{H}(2[A_2X + B_2M_{011}] + [A_3X + B_3M_{020}]) \prec 0 \tag{5.71e}$$

$$\frac{1}{3}\mathcal{H}([A_1X + B_1M_{002}] + 2[A_3X + B_3M_{101}]) \prec 0 \tag{5.71f}$$

$$\frac{1}{3}\mathcal{H}([A_2X + B_2M_{002}] + 2[A_3X + B_3M_{011}]) \prec 0 \tag{5.71g}$$

$$\frac{1}{3}\mathcal{H}\left(\sum_{i=1}^{3}[A_iX + B_iM_{111-\mathbf{1}_i}]\right) \prec 0 \tag{5.71h}$$

The proportional feedback $K_0(\mathbf{h})$ and the PDC feedback $K_1(\mathbf{h})$ are computed with the LMI conditions given by Theorem 5.3.3 and Theorem 5.3.1 for $p = 0$ as well. Figure 5.3 illustrates the (a, b)-regions for which a solution is found to these LMI conditions. The (a, b)-region gets larger as m increases. This demonstrates the increased capabilities of the BPDC control law compared to the usual PDC approach.

Bézier-nPDC Under the Assumptions 2.3.1, 2.3.2 and 2.3.3 of Chapter 2, and with θ still considered to be perfectly measured (Assumption 2.4.2), the previous results, i.e. Theorem 5.3.3 and Theorem 5.3.4, can be extended to the BnPDC control law:

$$u(t) = \left(\sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta))K_{\mathbf{i}}\right) Q(\mathbf{h}(\theta))x(t) \tag{5.72}$$

considered with the nQLF $V(x,\theta) = x^\top Q(\mathbf{h}(\theta))x$, as well as to its BnPDC observer counterpart:

$$\dot{\hat{x}}(t) = \sum_{j=1}^{n_h} h_j(\theta)\left(A_j\hat{x}(t) + B_ju(t)\right) + Q(\mathbf{h}(\theta)) \sum_{\mathbf{i}\in\mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta))L_{\mathbf{i}}(\hat{y}(t) - y(t)) \tag{5.73a}$$

$$\hat{y}(t) = \sum_{j=1}^{n_h} h_j(\theta)(C_j\hat{x}(t) + D_ju(t)) \tag{5.73b}$$

considered with the MQLF $V(x,\theta) = x^\top Q^{-1}(\mathbf{h}(\theta))x$, where $Q^{-1}(\mathbf{h}) = \sum_{k=1}^{n_h} h_kP_k$ and $P_k \in \mathbb{S}_{n_x}(\mathbb{R})$ for all $k \in [\![1, n_h]\!]$.

Theorem 5.3.5 (BnPDC controller). *Under the Assumptions 2.3.1, 2.3.2 and 2.3.3 of Chapter 2, and given $m \in \mathbb{N}$, the system (5.56) is globally exponentially stabilizable using the BnPDC control*

law (5.72) if there exists $\{P_k\}_{1\leq k\leq n_h}$ *and* $\{K_{\mathbf{i}}\}_{\mathbf{i}\in\mathbb{N}_m^{n_h}}$*, with* $P_k \in \mathbb{S}_{n_x}(\mathbb{R})$ *for all* $k \in [\![1, n_h]\!]$ *and* $K_{\mathbf{i}} \in \mathbb{R}^{n_u\times n_x}$ *for all* $\mathbf{i} \in \mathbb{N}_m^{n_h}$*, such that the conditions (5.74) are satisfied.*

$$P_i \succ 0, \quad i = 1, \dots, n_h \tag{5.74a}$$

$$\sum_{\mathbf{i}\in\mathbb{N}_{m+2}^{n}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \sum_{j=1}^{n} \left(\phi_j P_j + \sum_{k=1}^{n} \frac{i_k(i_j - \delta_{j,k})}{(m+2)(m+1)} \mathcal{H}(T_{\mathbf{i}-\mathbf{1}_j-\mathbf{1}_k,j,k}) \right) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{5.74b}$$

where $T_{\mathbf{i},j,k} \triangleq [A_j P_k + B_j K_{\mathbf{i}}]$*. The inequality on the Bézier sum can be transformed into regular LMI conditions by leveraging Theorem 5.3.1.*

Proof. For clarity, the mention of the scheduling vector $\theta(x)$ is omitted in this proof. Thanks to the interpolation elevation trick (Lemma 5.2.2), the closed-loop dynamic of (5.56) with (5.72) is given by:

$$\dot{x}(t) = \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \tilde{A}_{\mathbf{i}}(\mathbf{h}) x(t) \tag{5.75}$$

where $\tilde{A}_{\mathbf{i}}(\mathbf{h}) \triangleq \sum_{j=1}^{n_h} \frac{i_j}{m+1}[A_j + B_j K_{\mathbf{i}-\mathbf{1}_j} Q(\mathbf{h})]$. Following the proof of Theorem 2.3.4 in Chapter 2, the exponential stability of (5.75) is guaranteed if $R(\mathbf{h}) \prec 0$ for all $\theta \in \Theta$, where:

$$R(\mathbf{h}) = \dot{Q}(\mathbf{h}) + \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \mathcal{H}(Q(\mathbf{h})\tilde{A}_{\mathbf{i}}(\mathbf{h})) \tag{5.76}$$

By congruence (Property 2.2.2), $R(\mathbf{h}) \prec 0$ holds if and only if:

$$S(\mathbf{h}) \triangleq Q^{-1}(\mathbf{h}) R(\mathbf{h}) Q^{-1}(\mathbf{h}) \prec 0 \tag{5.77}$$

Developing the expression of S provides:

$$S(\mathbf{h}) = -[\dot{Q^{-1}}](\mathbf{h}) + \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \mathcal{H}(\tilde{A}_{\mathbf{i}}(\mathbf{h}) Q^{-1}(\mathbf{h})) \tag{5.78}$$

where $[\dot{Q^{-1}}](\mathbf{h}) = -Q^{-1}(\mathbf{h})\dot{Q}(\mathbf{h})Q^{-1}(\mathbf{h}) = \sum_{k=1}^{n_h} \dot{h}_k P_k$. Moreover, the following equalities hold:

$$\begin{aligned} \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \tilde{A}_{\mathbf{i}}(\mathbf{h}) Q^{-1}(\mathbf{h}) &= \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \sum_{j=1}^{n_h} \frac{i_j}{m+1} (A_j Q^{-1}(\mathbf{h}) + B_j K_{\mathbf{i}-\mathbf{1}_j}) \\ &= \sum_{k=1}^{n_h} h_k \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \Gamma_{\mathbf{i},k} \end{aligned} \tag{5.79}$$

with $\Gamma_{\mathbf{i},k} \triangleq \sum_{j=1}^{n_h} \frac{i_j}{m+1}[A_j P_k + B_j K_{\mathbf{i}-\mathbf{1}_j}]$. Thanks to the interpolation elevation trick (Lemma 5.2.2) once again, this provides:

$$\begin{aligned} \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \tilde{A}_{\mathbf{i}}(\mathbf{h}) Q^{-1}(\mathbf{h}) &= \sum_{\mathbf{i}\in\mathbb{N}_{m+2}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \sum_{k=1}^{n_h} \frac{i_k}{m+2} \Gamma_{\mathbf{i}-\mathbf{1}_k,k} \\ &= \sum_{\mathbf{i}\in\mathbb{N}_{m+2}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}) \sum_{j=1}^{n_h} \sum_{k=1}^{n_h} \frac{i_k(i_j - \delta_{j,k})}{(m+2)(m+1)} T_{\mathbf{i}-\mathbf{1}_j-\mathbf{1}_k,j,k} \end{aligned} \tag{5.80}$$

with $T_{\mathbf{i},j,k} = [A_j P_k + B_j K_{\mathbf{i}}]$. Finally, the Lipschitz assumptions $|\dot{h}_k| \leq \phi_k$ guarantee that the conditions (5.74) of the theorem imply $S(\mathbf{h}) \prec 0$, which concludes the proof. □

Theorem 5.3.6 (BnPDC observer). *Under the Assumptions 2.3.1, 2.3.2 and 2.3.3 of Chapter 2, and given $m \in \mathbb{N}$, the observation error between the state of (5.56) and the state of the observer (5.73) is globally exponentially stabilizable if there exists $\{P_k\}_{1\leq k\leq n_h}$ and $\{L_{\mathbf{i}}\}_{\mathbf{i}\in\mathbb{N}_m^{n_h}}$, with $P_k \in \mathbb{S}_{n_x}(\mathbb{R})$ for all $k \in [\![1, n_h]\!]$ and $L_{\mathbf{i}} \in \mathbb{R}^{n_x\times n_y}$ for all $\mathbf{i} \in \mathbb{N}_m^{n_h}$, such that the conditions (5.74) are satisfied with $T_{\mathbf{i},j,k} = [P_k A_j + L_{\mathbf{i}} C_j]$.*

Proof. The dynamic of the estimation error $e(t) \triangleq \hat{x}(t) - x(t)$ is given by

$$\dot{e}(t) = \sum_{\mathbf{i}\in\mathbb{N}_{m+1}^{n_h}} \mathcal{B}_{\mathbf{i}}(\mathbf{h}(\theta))\tilde{A}_{\mathbf{i}}(\mathbf{h}(\theta))e(t) \tag{5.81}$$

where $\tilde{A}_{\mathbf{i}}(\mathbf{h}) = \sum_{j=1}^{n} \frac{i_j}{m+1}\left(A_j + Q(\mathbf{h})L_{\mathbf{i}-\mathbf{1}_j}C_j\right)$. From here, the proof follows the same steps as for Theorem 5.3.5, without the left and right multiplications by $Q^{-1}(\mathbf{h})$. □

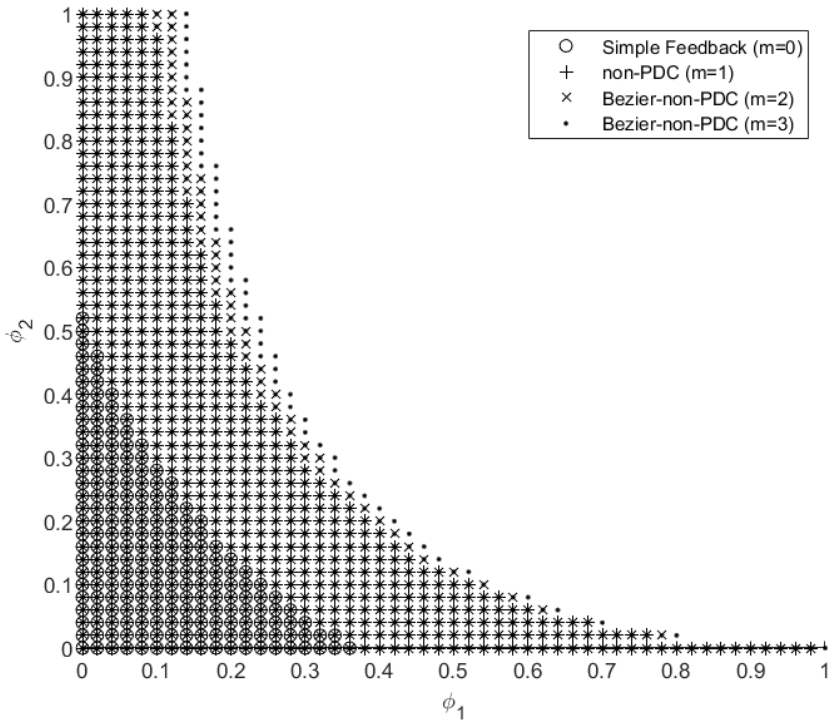


Figure 5.4: Stabilizability (ϕ_1, ϕ_2)-regions of $\mathcal{T}_{(\phi_1,\phi_2)}$ with a BnPDC control laws computed with the LMI conditions of Theorem 5.3.5 and Theorem 5.3.1 considered with $p = 0$.

Example 5.3.4. *To illustrate the conservatism reduction brought by the BnPDC controller design, the following T-S model is considered*

$$\mathcal{T}_{(\phi_1,\phi_2)}: \quad \dot{x}(t) = \sum_{i=1}^{2} h_i(\theta)\left(A_i x(t) + B_i u(t)\right) \tag{5.82}$$

with

$$
\begin{aligned}
A_1 &= \begin{pmatrix} 2 & -10 & 3 & 1 & 5 \\ 2 & 0 & 1 & 2 & 4 \\ -1 & 0 & -5 & 0 & -2 \\ 1 & 0 & 5 & 0 & -1 \\ -1 & 5 & 4 & 3 & 1 \end{pmatrix} \quad B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ -2 & 0 \end{pmatrix} \\
A_2 &= \begin{pmatrix} 0 & 5 & 2 & -1 & 1 \\ 1 & 2 & 1 & -2 & -1 \\ -1 & 0 & -10 & -1 & -1 \\ 1 & 0 & -10 & 1 & -1 \\ 4 & 5 & -1 & -2 & 5 \end{pmatrix} \quad B_2 = \begin{pmatrix} 0 & 1 \\ 0 & 2 \\ 1 & -1 \\ 1 & -2 \\ 1 & 1 \end{pmatrix}
\end{aligned} \tag{5.83}
$$

under the Assumptions 2.3.1, 2.3.2 and 2.3.3 of Chapter 2, with $|\dot{h}_1| \leq \phi_1$ *and* $|\dot{h}_2| \leq \phi_2$*, together with the nQLF* $V(x) = x^\top [h_1 P_1 + h_2 P_2]^{-1} x$*, with symmetric* $P_1, P_2 \in \mathbb{S}_{n_x}^{++}(\mathbb{R})$*. The stabilization problem of* $\mathcal{T}_{(\phi_1,\phi_2)}$ *is considered at several values of* $(\phi_1, \phi_2) \in \mathbb{R}^2_{\geq 0}$ *for the control laws* $u(t) = K_m(\mathbf{h})[h_1 P_1 + h_2 P_2]^{-1} x(t)$*, with* $m \in [\![0, 3]\!]$ *and*

$$
\begin{aligned}
K_0(\mathbf{h}) &= K_{00} && \text{(5.84a)} \\
K_1(\mathbf{h}) &= h_1 K_{10} + h_2 K_{01} && \text{(5.84b)} \\
K_2(\mathbf{h}) &= h_1^2 K_{20} + 2h_1 h_2 K_{11} + h_2^2 K_{02} && \text{(5.84c)} \\
K_3(\mathbf{h}) &= h_1^3 K_{30} + 3h_1^2 h_2 K_{21} + 3h_1 h_2^2 K_{12} + h_2^3 K_{03} && \text{(5.84d)}
\end{aligned}
$$

For $m = 3$*,* $n^m - (m+n-1)!/m!(n-1)! = 4$ *redundant gain matrices have been economized compared to the usual multi-sum approach, i.e.* 40 *useless decision variables. Let* $R(\phi) = -\phi_1 P_1 - \phi_2 P_2$*. Theorem 5.3.3 and Theorem 5.3.1 considered with* $p = 0$ *provide the following LMI conditions to compute the simple feedback* $K_0(\mathbf{h})[h_1 P_1 + h_2 P_2]^{-1}$*:*

$$
\begin{aligned}
\mathcal{H}(A_1 P_1 + B_1 K_{00}) &\prec R(\phi) && \text{(5.85a)} \\
\frac{1}{2}\mathcal{H}([A_1 P_2 + B_1 K_{00}] + [A_2 P_2 + B_2 K_{00}]) &\prec R(\phi) && \text{(5.85b)} \\
\mathcal{H}(A_2 P_2 + B_2 K_{00}) &\prec R(\phi) && \text{(5.85c)}
\end{aligned}
$$

Similarly, the LMI conditions to compute the nPDC feedback $K_1(\mathbf{h})[h_1 P_1 + h_2 P_2]^{-1}$ *are given by:*

$$
\begin{aligned}
\mathcal{H}(A_1 P_1 + B_1 K_{10}) &\prec R(\phi) && \text{(5.86a)} \\
\frac{1}{3}\mathcal{H}([A_1 P_1 + B_1 K_{01}] + [A_1 P_2 + B_1 K_{10}] + [A_2 P_1 + B_2 K_{10}]) &\prec R(\phi) && \text{(5.86b)} \\
\frac{1}{3}\mathcal{H}([A_1 P_2 + B_1 K_{01}] + [A_2 P_1 + B_2 K_{01}] + [A_2 P_2 + B_2 K_{10}]) &\prec R(\phi) && \text{(5.86c)} \\
\mathcal{H}(A_2 P_2 + B_2 K_{01}) &\prec R(\phi) && \text{(5.86d)}
\end{aligned}
$$

the LMI conditions to compute the BnPDC feedback $K_2(\mathbf{h})[h_1P_1 + h_2P_2]^{-1}$ *are given by:*

$$\mathcal{H}(A_1P_1 + B_1K_{20}) \prec R(\phi) \tag{5.87a}$$

$$\frac{1}{4}\mathcal{H}(2[A_1P_1 + B_1K_{11}] + [A_1P_2 + B_1K_{20}] + [A_2P_1 + B_2K_{20}]) \prec R(\phi) \tag{5.87b}$$

$$\frac{1}{6}\mathcal{H}([A_1P_1 + B_1K_{02}] + 2[A_1P_2 + B_1K_{11}] + \dots \\ 2[A_2P_1 + B_2K_{11}] + [A_2P_2 + B_2K_{20}]) \prec R(\phi) \tag{5.87c}$$

$$\frac{1}{4}\mathcal{H}([A_1P_2 + B_1K_{02}] + [A_2P_1 + B_2K_{02}] + 2[A_2P_2 + B_2K_{11}]) \prec R(\phi) \tag{5.87d}$$

$$\mathcal{H}(A_2P_2 + B_2K_{02}) \prec R(\phi) \tag{5.87e}$$

and finally the LMI conditions to compute the BnPDC feedback $K_3(\mathbf{h})[h_1P_1 + h_2P_2]^{-1}$ *are given by:*

$$\forall i \in [\![1, 2]\!] : \ \mathcal{H}(A_iP_i + B_iK_{00+3\cdot\mathbf{1}_i}) \prec R(\phi) \tag{5.88a}$$

$$\frac{1}{5}\mathcal{H}(3[A_1P_1 + B_1K_{21}] + [A_1P_2 + B_1K_{30}] + [A_2P_1 + B_2K_{30}]) \prec R(\phi) \tag{5.88b}$$

$$\frac{1}{5}\mathcal{H}([A_1P_2 + B_1K_{03}] + [A_2P_1 + B_2K_{03}] + 3[A_2P_2 + B_2K_{12}]) \prec R(\phi) \tag{5.88c}$$

$$\frac{1}{10}\mathcal{H}(3[A_1P_1 + B_1K_{12}] + 3[A_1P_2 + B_1K_{21}] + \dots \\ 3[A_2P_1 + B_2K_{21}] + [A_2P_2 + B_2K_{30}]) \prec R(\phi) \tag{5.88d}$$

$$\frac{1}{10}\mathcal{H}([A_1P_1 + B_1K_{03}] + 3[A_1P_2 + B_1K_{12}] + \dots \\ 3[A_2P_1 + B_2K_{12}] + 3[A_2P_2 + B_2K_{21}]) \prec R(\phi) \tag{5.88e}$$

Figure 5.4 illustrates the (ϕ_1, ϕ_2)*-regions for which a solution is found to the LMI given above. This* (ϕ_1, ϕ_2)*-region gets larger as* m *increases, which demonstrates the increased capabilities of the BnPDC control law compared to the usual nPDC approach.*

5.4 Conclusions and perspectives

In this chapter, a reformulation of the multi-sums found in the T-S framework was introduced, using Bernstein polynomials. This rewriting provides a more efficient representation of multi-sums by eliminating their redundant terms, and reveals the multi-sums to be, in fact, Bézier interpolations in disguise. After properly introducing both the Bernstein polynomials and the Bézier interpolations, together with some important properties, this formalism was applied to the T-S framework. An explicit rewriting of the LMI conditions found in [238] was introduced. Moreover, the introduction of Bézier-T-S models was proposed, and has yet to be fully investigated. Finally, the generalized PDC and nPDC schemes have been reformulated into BPDC and BnPDC versions. Simple LMI formulations of the resulting stabilization problems have been provided for QLF and nQLF. Collectively, these results advocate for the adoption of the Bézier formalism in the T-S framework. It should be noted that the LMI conditions given in this chapter could be relaxed using other relaxation schemes than [238]. The extension of this Bézier approach also remains to be explored for generalized MQLF, for discrete-time T-S models, as well as for T-S models with an unmeasurable scheduling vector.

Chapter 6

Anticipating the near future of an LPV system

This chapter assumes a bounded variation rates on the scheduling vector θ of an LPV model, and introduces tools to quantify the discrepancies between the true future of the system and one where the value of θ remains constant. This methodology enables LTV-like results for LPV models, such as an exact discretization technique, and predictions on some structural properties like controllability and observability.

6.1 From real-time knowledge to near future knowledge

LPV models, including T-S models, act at each instant t as a specific LTI model. This behaviour is also found both in LDI and in LTV models. However, LDI models do not benefit from any information regarding which LTI model is active at a given time t, while LTV models explicitly provide the relationship between the time and the active LTI models. LPV are half-way between these two frameworks, as the relation between the active local LTI models and the time t is implicitly given through a scheduling vector θ. More precisely, the active LTI model at a time t is assumed to only be known in real-time, based on the measurement or on the estimation of the scheduling vector θ. This real-time knowledge being not as powerful as the all-time knowledge that comes with LTV systems, the analysis and the controller and observer design for LPV models cannot rely easily on the results dedicated to LTV models, for which the future

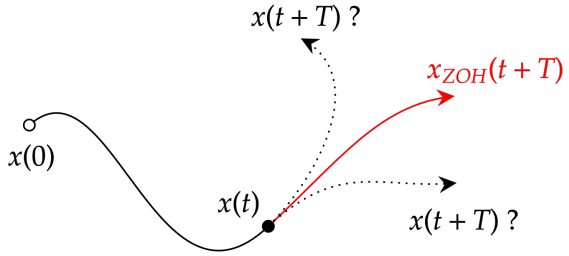


Figure 6.1: Potential discrepancy between the future trajectory of an LPV system and its trajectory under a zero-order hold assumption on θ.

needs to be known in advance.

Typically, the LPV models are subject to some issues regarding their discretization, which necessitates to know the scheduling vector behaviour between two samples. Because of the lack of all-time knowledge, a zero-order hold assumption is generally made on the scheduling vector θ in order to discretize LPV systems. This zero-order hold assumption consists in considering that the scheduling vector remains constant between two samples, by holding its last known value [281, 280]. This zero-order hold assumption made on θ is typically crucial in the following situations.

- Discrete-time LPV representations are often derived from continuous-time LPV models using a zero-order hold assumption on θ [281, 280].
- If a controller is synthesized for a continuous-time LPV system [209, 246], its practical implementation is generally sampled (assuming that a numerical controller is used), and the dynamic of θ is neglected during the sampling period. This sampled-data issue has already been discussed for example in [269, 228].

In both of these context, the dynamic of θ is neglected during the sampling period, and, to the author's knowledge, no rigorous bounds are derived to quantify the potential discrepancy between the trajectory of an LPV system, and its trajectory under a zero-order hold assumption on θ (see Figure 6.1).

To overcome this issue, the present chapter suggests to rely on a common assumption of the LPV framework, namely: a Lipschitz assumption on the scheduling vector θ. This assumption was already leveraged for T-S models in order to establish both stability and stabilization results in Chapters 3, 4, and 5, where it was directly formulated as Lipschitz assumptions on the activation functions $\mathbf{h}$ (Assumptions 2.3.3 and 4.3.3). This Lipschitz assumption is also common in the LPV framework, e.g. in model predictive control schemes [267, 34, 168, 144]. Combining a Lipschitz assumption on θ with its real-time knowledge leads to an *uncertain* knowledge of the value of θ in a near future. This uncertainty being quantified by the Lipschitz constant, bounds on the future state trajectory of a continuous-time LPV system can be obtained in terms of norm-bounded uncertain matrices. The uncertain matrices are introduced in the system to represent its divergence compared to an estimated future, typically the zero-order hold future. The further into the future the system is investigated, the greater its uncertainties become, which makes this methodology mainly useful to obtain results in a near future. In particular, bounds are obtained on the future controllability and observability Gramians (Theorem 1.1.2 of Chapter 1) of an LPV system, leading to near future controllability and observability results similar to those of LTV systems.

Two crucial tools are introduced by this chapter, namely:

- Volterra's product integration, which corresponds in this context to the state-transition matrix of an LTV system;
- the weighted logarithmic norm of a matrix, which can effectively bound the uncertain matrices representing the divergence between the future of an LPV system and its estimated future.

The results of this chapter are partially adapted from some of the author's previous results, published as conference papers in [17, 24].

The chapter is organised as follows: Volterra's product integration is introduced in details in Section 6.2, then the weighted logarithmic norm of a matrix is introduced in Section 6.3. In Section 6.4, these two tools are leveraged in order to obtain results on the near future of an LPV system, including a methodology to exactly discretize these systems as well as a way to anticipate their loss of controllability and observability. Finally, some conclusions and perspectives are discussed in Section 6.5.

6.2 Product integration

Remark 6.2.1. *This section is dense and may be skipped; readers are encouraged to proceed directly to Section 6.3 after reviewing this preliminary material.*

The intuitive idea behind product integration consist in defining a usual integral, but in a multiplicative setting. The usual Riemann integral is introduced thereafter, from right to left (although by commutativity of the addition, the order of operation is not important here), with $t_1 < t_2$ and $\delta_s > 0$ an infinitesimally small quantity:

$$\int_{t_1}^{t_2} A(s)ds \approx A(t_2)\delta_s + A(t_2 - \delta_s)\delta_s + \cdots + A(t_1 + \delta_s)\delta_s + A(t_1)\delta_s \tag{6.1}$$

The *right-to-left* multiplicative counterpart to the equation above can be intuitively taken by substituting matrix addition with matrix multiplication, and scalar multiplication with fractional matrix power:

$$\prod_{t_1}^{t_2} A(s)^{ds} \approx A(t_2)^{\delta_s} A(t_2 - \delta_s)^{\delta_s} \dots A(t_1 + \delta_s)^{\delta_s} A(t_1)^{\delta_s} \tag{6.2}$$

This expression introduces fractional matrix powers of the form $A^\alpha = e^{\alpha \log(A)}$, with $\alpha \in \mathbb{R}_{>0}$. The logarithm used in the fractional matrix power definition is however only defined for non-singular matrices, and it is not unique. To remove any ambiguity from this construction, this *right-to-left multiplicative Riemann integral* is directly written using the exponential expression of the fractional matrix power, with the logarithm removed:

$$\prod_{t_1}^{t_2} e^{A(s)ds} \approx e^{A(t_2)\delta_s} e^{A(t_2-\delta_s)\delta_s} \dots e^{A(t_1+\delta_s)\delta_s} e^{A(t_1)\delta_s} \tag{6.3}$$

which can also be noted as follows, after applying a first-order Taylor expansion:

$$\prod_{t_1}^{t_2} (I + A(s)ds) \approx (I + A(t_2)\delta_s)\,(I + A(t_2 - \delta_s)\delta_s) \dots (I + A(t_1 + \delta_s)\delta_s)\,(I + A(t_1)\delta_s) \tag{6.4}$$

As the reader can expect, this product integral can be rigorously defined using the multiplicative equivalent to a Riemann integral, as introduced hereabove. It should be noted that a multiplicative Lebesgue integration theory also exists [80, 260]. Historically, product integration has been introduced by Vito Volterra in order to study the solutions to differential equation of the form $\dot{x}(t) = A(t)x(t)$ [290]. Indeed, since the solutions to $\dot{x}(t) = Ax(t)$ are given by

$x(t) = e^{(t-t_0)A}x(t_0)$, it is rather intuitive that the solutions to the time-varying case are going to be obtained with:

$$x(t) = \left(\prod_{t_0}^{t} e^{A(s)ds}\right) x(t_0) \approx e^{A(t)\delta_s} e^{A(t-\delta_s)\delta_s} \dots e^{A(t_0+\delta_s)\delta_s} e^{A(t_0)\delta_s} x(t_0) \tag{6.5}$$

This manuscript actually defines the product integral using the solutions of the differential equation $\dot{x}(t) = A(t)x(t)$ directly, essentially reducing the product integral to a mere eccentricity of notation to denote the usual state-transition matrix $\Phi(t_2, t_1)$. However, the author strongly believes this notation to be the most natural and well-suited one for the applications that follow, and it will be used in the rest of this chapter.

6.2.1 Peano-Baker series

The Peano-Baker series is essentially the Picard iteration leveraged in order to demonstrate the existence and uniqueness of the solutions to the differential equation $\dot{x}(t) = A(t)x(t)$ [16]. In this chapter, the product integral of A is defined using this Peano-Baker series.

Definition 6.2.1 (Peano-Baker series). *Given $A : \mathbb{R} \to \mathbb{K}^{n\times n}$ a piecewise continuous function (with $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$), the product integral of A between t_1 and $t_2 \in \mathbb{R}$ is defined using the Peano-Baker series:*

$$\prod_{t_1}^{t_2} e^{A(s)ds} \triangleq \sum_{k=0}^{+\infty} J_k(t_2, t_1) \tag{6.6}$$

where the J_k are defined recursively using:

$$\begin{cases} J_{k+1}(t_2, t_1) = \int_{t_1}^{t_2} A(s)J_k(s, t_1)ds \\ J_0(t_2, t_1) = I_n \end{cases} \qquad \forall t_1, t_2 \in \mathbb{R} \tag{6.7}$$

Remark 6.2.2. *Once the expression of each J_k is fully developed, the Peano-Baker series reads:*

$$\begin{aligned} \prod_{t_1}^{t_2} e^{A(s)ds} &\triangleq I_n + \int_{t_1}^{t_2} A(s)ds + \sum_{k=1}^{+\infty} \int_{t_1}^{t_2} \int_{t_1}^{s_1} \cdots \int_{t_1}^{s_k} A(s_1) \dots A(s_k) ds_k \dots ds_1 \\ &= I_n + \int_{t_1}^{t_2} A(s)ds + \int_{t_1}^{t_2} \int_{t_1}^{s_1} A(s_1)A(s_2)ds_2 ds_1 + \dots \end{aligned} \tag{6.8}$$

Property 6.2.1 (State transition matrix). *The product integral of the piecewise continuous matrix $A : \mathbb{R} \to \mathbb{K}^{n\times n}$ is the state-transition matrix of the system:*

$$\dot{x}(t) = A(t)x(t) \tag{6.9}$$

i.e. for all solution $x : \mathbb{R} \to \mathbb{R}^n$ to the equation (6.9), the following holds for all $t_1, t_2 \in \mathbb{R}$:

$$x(t_2) = \left(\prod_{t_1}^{t_2} e^{A(s)ds}\right) x(t_1) \tag{6.10}$$

Proof. The terms of the Peano-Baker series can be retrieved from the Picard iteration demonstrating the existence and unicity of the solutions to (6.9). See [16] for more details. □

This product integral can also be denoted using a time-ordering operator. This time-ordered notation is usual for the product integral in physics, particularly in quantum field theory [55, 93].

Property 6.2.2 (Time-ordered exponential)**.** *Introducing the linear time-ordering (or chronological) operator $\mathcal{T}$ such that for all $\sigma \in \mathfrak{S}_n$ (where $\mathfrak{S}_n$ stands for the symmetric group of order n) and $s_1 \leq \cdots \leq s_n$:*

$$\mathcal{T}\big(A(s_{\sigma(1)}) \dots A(s_{\sigma(n)})\big) = A(s_1) \dots A(s_n) \tag{6.11}$$

then the following holds:

$$\prod_{t_1}^{t_2} e^{A(s)ds} = \mathcal{T} e^{\int_{t_1}^{t_2} A(s)ds} \tag{6.12}$$

Proof. The steps of the proof are succinctly given below:

$$\begin{aligned}
\prod_{t_1}^{t_2} e^{A(s)ds} &= I_n + \sum_{k=1}^{+\infty} \int_{t_1}^{t_2} \int_{t_1}^{s_1} \cdots \int_{t_1}^{s_k} A(s_1) \dots A(s_k) ds_k \dots ds_1 \\
&= \mathcal{T}(I_n) + \sum_{k=1}^{+\infty} \frac{1}{\#\mathfrak{S}_k} \int_{t_1}^{t_2} \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} \mathcal{T}\big(A(s_1) \dots A(s_k)\big) ds_k \dots ds_1 \\
&= \mathcal{T}\left(I_n + \sum_{k=1}^{+\infty} \frac{1}{k!} \int_{t_1}^{t_2} \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} A(s_1) \dots A(s_k) ds_k \dots ds_1 \right) \\
&= \mathcal{T}\left(\sum_{k=0}^{+\infty} \frac{1}{k!} \left(\int_{t_1}^{t_2} A(s) ds \right)^k \right) \\
&= \mathcal{T}\left(e^{\int_{t_1}^{t_2} A(s)ds} \right)
\end{aligned} \tag{6.13}$$

□

Finally, the following bounds can be established:

Property 6.2.3. *The following inequality holds for all $t_1, t_2 \in \mathbb{R}$:*

$$\left\| \prod_{t_1}^{t_2} e^{A(s)ds} \right\|_2 \leq e^{\left| \int_{t_1}^{t_2} \|A(s)\|_2 ds \right|} \tag{6.14}$$

Proof. See Theorem 4.1 of [80]. It is not difficult to obtain this inequality using the triangle inequality and the submultiplicativity of the 2-norm in the time-ordered formulation of the product integral. □

Remark 6.2.3. *This inequality may not be very precise, and it will be improved later in this chapter using the weighted logarithmic norm of $A(t)$ (Property 6.3.3).*

Property 6.2.4 (Bounds on the Peano-Baker series remainder)**.** *The following bounds hold on the remainder of the Peano-Baker series, for all $m \in \mathbb{N}$ and $t_1, t_2 \in \mathbb{R}$:*

$$\begin{aligned}\left\| \sum_{k=m+1}^{+\infty} J_k(t_2, t_1) \right\|_2 &\leq e^{\left|\int_{t_1}^{t_2} \|A(s)\|_2 ds\right|} - \sum_{k=0}^{m} \frac{1}{k!} \left| \int_{t_1}^{t_2} \|A(s)\|_2 ds \right|^k \\ &\leq \frac{1}{(m+1)!} \left| \int_{t_1}^{t_2} \|A(s)\|_2 ds \right|^{m+1} e^{\left|\int_{t_1}^{t_2} \|A(s)\|_2 ds\right|}\end{aligned} \tag{6.15}$$

Proof. See Theorem 4.2 of [80]. □

6.2.2 State transition matrix

The results from this section are usually discussed when the product integral is considered as the state transition matrix of the LTV system $\dot{x}(t) = A(t)x(t)$. The following results are therefore well-known in control.

Property 6.2.5 (State transition matrix properties)**.** *The usual properties of a state transition matrix are recalled below:*

- *Identity:*

$$\prod_{t}^{t} e^{A(s)ds} = I_n, \quad \forall t \in \mathbb{R} \tag{6.16}$$

- *Inverse:*

$$\left(\prod_{t_1}^{t_2} e^{A(s)ds} \right)^{-1} = \prod_{t_2}^{t_1} e^{A(s)ds}, \quad \forall t_1, t_2 \in \mathbb{R} \tag{6.17}$$

- *Chasles relation:*

$$\prod_{t_1}^{t_3} e^{A(s)ds} = \left(\prod_{t_2}^{t_3} e^{A(s)ds} \right) \left(\prod_{t_1}^{t_2} e^{A(s)ds} \right), \quad \forall t_1, t_2, t_3 \in \mathbb{R} \tag{6.18}$$

- *Partial derivative:*

$$\begin{aligned}\frac{\partial}{\partial t_2} \left(\prod_{t_1}^{t_2} e^{A(s)ds} \right) &= A(t_2) \left(\prod_{t_1}^{t_2} e^{A(s)ds} \right) \\ \frac{\partial}{\partial t_1} \left(\prod_{t_1}^{t_2} e^{A(s)ds} \right) &= - \left(\prod_{t_1}^{t_2} e^{A(s)ds} \right) A(t_1)\end{aligned} \tag{6.19}$$

Proof. All properties directly derive from the equation (6.10) of Property 6.2.1. □

Theorem 6.2.1 (Similarity). *Let $A : \mathbb{R} \to \mathbb{K}^{n\times n}$ be a piecewise continuous matrix and $P \in \mathrm{GL}_n(\mathbb{K})$. The following identity holds for all $t_1, t_2 \in \mathbb{R}$:*

$$P\left(\prod_{t_1}^{t_2} e^{A(s)ds}\right)P^{-1} = \prod_{t_1}^{t_2} e^{PA(s)P^{-1}ds} \tag{6.20}$$

Proof. See Theorem 2.5.12 of [260]. □

Remark 6.2.4. *This identity corresponds to a change of coordinates of the LTV system $\dot{x}(t) = A(t)x(t)$. Not using the product integral notation hides the fact that this change of coordinates simply generalizes the identity $Pe^AP^{-1} = e^{PAP^{-1}}$.*

Theorem 6.2.2 (Commutative case). *If the family of matrices $\{A(t)\}_{t\in[t_1,t_2]}$ commutes (i.e. $A(s_1)A(s_2) = A(s_2)A(s_1)$ for all $s_1, s_2 \in [t_1, t_2]$), then:*

$$\prod_{t_1}^{t_2} e^{A(s)ds} = e^{\int_{t_1}^{t_2} A(s)ds} \tag{6.21}$$

Proof. See Theorem 1.3 of [80] or Lemma 2.4.2 of [260]. □

Remark 6.2.5. *Although this result is well-known, not using the product integral notation hides the fact that it simply generalizes the identity $e^Ae^B = e^{A+B}$ for two commuting matrices A and B.*

Theorem 6.2.3 (Liouville's formula). *Let $A : \mathbb{R} \to \mathbb{K}^{n\times n}$ be a piecewise continuous matrix. For all $t_1, t_2 \in \mathbb{R}$:*

$$\det\left(\prod_{t_1}^{t_2} e^{A(s)ds}\right) = e^{\int_{t_1}^{t_2} \mathrm{Tr}\left(A(s)\right)ds} \tag{6.22}$$

Proof. See Theorem 1.4 of [80] or Theorem 2.5.11 of [260] □

Remark 6.2.6. *Although Liouville's formula is well-known, not using the product integral notation hides the fact that it simply generalizes the identity $\det(e^A) = e^{\mathrm{Tr}(A)}$.*

Remark 6.2.7. *Given a set $S(0) \subseteq \mathbb{R}^n$, if each $x(0) \in S(0)$ follows the dynamic $\dot{x}(t) = A(t)x(t)$ with $A : \mathbb{R} \to \mathbb{R}^{n\times n}$ piecewise continuous, then the set at time t is $S(t) = \left(\prod_0^t e^{A(s)ds}\right) S(0)$, and its volume is given by:*

$$\mathrm{Vol}(S(t)) = \left|\det\left(\prod_0^t e^{A(s)ds}\right)\right| \mathrm{Vol}(S(0)) \tag{6.23}$$

Liouville's formula (6.22) *allows to rewrite this expression in a practical way [298], and find that the dynamic* $\dot{x}(t) = A(t)x(t)$ *is* n*-contractive if and only if*

$$\lim_{t\to+\infty} \int_0^t \mathrm{Tr}\big(A(s)\big)ds = -\infty \tag{6.24}$$

Theorem 6.2.4 (Solutions to a differential Sylvester equation)**.** *Given* $A, B, C : \mathbb{R} \to \mathbb{K}^{n\times n}$ *three piecewise continuous matrix, the solutions to the differential Sylvester equation:*

$$\dot{X}(t) = A(t)X(t) + X(t)B(t) + C(t) \tag{6.25}$$

are given by:

$$X(t) = \left(\prod_{t_0}^{t} e^{A(s)ds}\right) X(t_0) \left(\prod_{t}^{t_0} e^{-B(s)ds}\right) + \int_{t_0}^{t} \left(\prod_{\tau}^{t} e^{A(s)ds}\right) C(\tau) \left(\prod_{t}^{\tau} e^{-B(s)ds}\right) d\tau \tag{6.26}$$

Proof. This is easily checked that (6.26) is solution to (6.25) by rewriting $\dot{X}(t)$ using Property 6.2.5. Uniqueness of the solutions with respect to the initial condition $X(t_0)$ is obtained by noticing that (6.26) is simply a linear differential equation, which is well-known to have a unique trajectory associated to each initial condition. □

Remark 6.2.8. *Alternative expressions to* $X(t)$ *can be found using the following identities in* (6.26)*:*

$$\prod_{t}^{t_0} e^{-B(s)ds} = \left(\prod_{t_0}^{t} e^{-B(s)ds}\right)^{-1} = \left(\prod_{t_0}^{t} e^{B^\top(s)ds}\right)^{\top} \tag{6.27}$$

which are essentially ways to represent the left-to-right product integral [260].

Corollary 6.2.1 (Trajectories of an LTV system)**.** *The trajectories of the LTV system* $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ *are given by:*

$$\begin{aligned} x(t) &= \left(\prod_{t_0}^{t} e^{A(s)ds}\right) x(t_0) + \int_{t_0}^{t} \left(\prod_{\tau}^{t} e^{A(s)ds}\right) B(\tau)u(\tau)d\tau \\ &= \left(\prod_{t_0}^{t} e^{A(s)ds}\right) \left(x(t_0) + \int_{t_0}^{t} \left(\prod_{t_0}^{\tau} e^{A(s)ds}\right)^{-1} B(\tau)u(\tau)d\tau\right) \end{aligned} \tag{6.28}$$

Theorem 6.2.5 (Floquet's theorem)**.** *If the piecewise continuous matrix* $A : \mathbb{R} \to \mathbb{K}^{n\times n}$ *is* T*-periodic (i.e.* $A(t+T) = A(t)$ *for all* $t \in \mathbb{R}$*), then there exists a* T*-periodic matrix* $F : \mathbb{R} \to \mathbb{C}^{n\times n}$

and a constant matrix $K \in \mathbb{C}^{n\times n}$ such that for all $t \in \mathbb{R}_{\geq 0}$:

$$\prod_0^t e^{A(s)ds} = F(t)e^{tK} \tag{6.29}$$

Closed-form expressions for F and K are provided below:

$$F(t) = \left(\prod_0^{t-\lfloor t/T\rfloor T} e^{A(s)ds}\right) e^{(\lfloor t/T\rfloor T - t)K}, \quad K = \frac{1}{T}\log\left(\prod_0^T e^{A(s)ds}\right) \tag{6.30}$$

where $\lfloor\cdot\rfloor$ stands for the floor function.

Proof. Since A is T-periodic, given $k \in \mathbb{N}$ and $\delta \in [0, T)$ such that $t = kT + \delta$, the following equalities hold:

$$\begin{aligned}
\prod_0^{kT+\delta} e^{A(s)ds} &= \left(\prod_{kT}^{kT+\delta} e^{A(s)ds}\right)\left(\prod_0^{kT} e^{A(s)ds}\right) \\
&= \left(\prod_0^{\delta} e^{A(s+kT)ds}\right)\left(\prod_{(k-1)T}^{kT} e^{A(s)ds}\right)\cdots\left(\prod_T^{2T} e^{A(s)ds}\right)\left(\prod_0^{T} e^{A(s)ds}\right) \\
&= \left(\prod_0^{\delta} e^{A(s+kT)ds}\right)\left(\prod_0^{T} e^{A(s+(k-1)T)ds}\right)\cdots\left(\prod_0^{T} e^{A(s+T)ds}\right)\left(\prod_0^{T} e^{A(s)ds}\right) \\
&= \left(\prod_0^{\delta} e^{A(s)ds}\right)\left(\prod_0^{T} e^{A(s)ds}\right)^k
\end{aligned} \tag{6.31}$$

The product integral $\left(\prod_0^T e^{A(s)ds}\right)$ being nonsingular, it is possible to take its logarithm in $\mathbb{C}^{n\times n}$:

$$\begin{aligned}
\prod_0^{kT+\delta} e^{A(s)ds} &= \left(\prod_0^{\delta} e^{A(s)ds}\right) e^{k\log\left(\prod_0^T e^{A(s)ds}\right)} \\
&= \left(\prod_0^{\delta} e^{A(s)ds}\right) e^{(t-\delta)\log\left(\prod_0^T e^{A(s)ds}\right)/T} \\
&= \left(\prod_0^{t-\lfloor t/T\rfloor T} e^{A(s)ds}\right) e^{(\lfloor t/T\rfloor T - t)K} e^{tK}
\end{aligned} \tag{6.32}$$

□

Remark 6.2.9. *The product integral notation is once again eye-opening: e^K in Floquet's theorem (Theorem 6.2.5) appears to be the product average of A over a period T. This product average can be understood by analogy with the usual average:*

$$(\textit{product average}) \quad \left(\prod_0^T e^{A(s)ds}\right)^{\frac{1}{T}} \longleftrightarrow \frac{1}{T}\int_0^T A(s)ds \quad (\textit{usual average}) \tag{6.33}$$

This product average contains all necessary information to conclude on the stability of $\dot{x}(t) = A(t)x(t)$, as its eigenvalues are related to the characteristic multipliers (or the Floquet multipliers, or Poincaré multipliers) of the periodic system [194].

6.2.3 Multiplicative calculus

The results from this section are usually discussed when the product integral is considered as a calculus tool, by analogy with the usual integral $\int$, rather than as a state transition matrix. To the author's knowledge, these results are typically never discussed in control theory.

Theorem 6.2.6 (Change of variable). *Let $A : \mathbb{R} \to \mathbb{K}^{n\times n}$ be a piecewise continuous matrix and $\varphi \in \mathcal{C}^1([t_1, t_2], \mathbb{R})$, with φ' the derivative of φ. The following holds:*

$$\prod_{t_1}^{t_2} e^{A(s)ds} = \prod_{\varphi(t_1)}^{\varphi(t_2)} e^{A(\varphi(s))\varphi'(s)ds} \tag{6.34}$$

Proof. See Theorem 3.5 of [80] or Theorem 2.5.10 of [260]. □

Remark 6.2.10. *There is a clear analogy with the usual change of variable:*

$$\int_{t_1}^{t_2} A(s)ds = \int_{\varphi(t_1)}^{\varphi(t_2)} A(\varphi(s))\varphi'(s)ds \tag{6.35}$$

Remark 6.2.11. *This change of variable corresponds to a scaling of time of the LTV system $\dot{x}(t) = A(t)x(t)$.*

Theorem 6.2.7 (Duhamel's formula). *Let $A, B : \mathbb{R} \to \mathbb{K}^{n\times n}$ be two piecewise continuous matrices. For all $t_1, t_2 \in \mathbb{R}$:*

$$\prod_{t_1}^{t_2} e^{A(s)ds} - \prod_{t_1}^{t_2} e^{B(s)ds} = \int_{t_1}^{t_2} \left(\prod_{\tau}^{t_2} e^{B(s)ds}\right) (A(\tau) - B(\tau)) \left(\prod_{t_1}^{\tau} e^{A(s)ds}\right) d\tau \tag{6.36}$$

Proof. See Theorem 5.1 of [80]. □

Remark 6.2.12. *This result is essentially a broad generalization of the following identity:*

$$A^m - B^m = \sum_{k=0}^{m-1} A^{m-1-k}(A - B)B^k \tag{6.37}$$

which, as will be seen later, is particularly useful in order to upper-bound $\|A^m - B^m\|_2$. Similarly, Duhamel's formula is practical in order to upper-bound $\|\prod_{t_1}^{t_2} e^{A(s)ds} - \prod_{t_1}^{t_2} e^{B(s)ds}\|_2$, which is crucial to the results of Section 6.4.

Theorem 6.2.8 (Fundamental theorem of multiplicative calculus). *Let $A \in \mathcal{C}^1(\mathbb{R}, \mathrm{GL}_n(\mathbb{K}))$ be a continuously differentiable matrix. The following holds for all $t_1, t_2 \in \mathbb{R}$:*

$$\prod_{t_1}^{t_2} e^{A'(s)A^{-1}(s)ds} = A(t_2)A^{-1}(t_1) \tag{6.38}$$

Proof. See Theorem 3.1 of [80]. □

Remark 6.2.13. *In Volterra's matrix calculus, the quantity $A'(t)A^{-1}(t)$ is called the left derivative of the matrix function A [260]. There is a clear analogy with the usual fundamental theorem of calculus:*

$$\int_{t_1}^{t_2} A'(s)ds = A(t_2) - A(t_1) \tag{6.39}$$

Theorem 6.2.9 (Contour integration). *Taking a continuously differentiable curve $\varphi \in \mathcal{C}^1([t_1, t_2], \mathbb{C})$ and a continuous $A \in \mathcal{C}^0(\mathbb{C}, \mathbb{C}^{n\times n})$, the contour product integral of A along φ is defined as:*

$$\prod_{\varphi} e^{A(z)dz} \triangleq \prod_{t_1}^{t_2} e^{A(\varphi(s))\varphi'(s)ds} \tag{6.40}$$

This expression only depends on $\varphi(t_1)$ and $\varphi(t_2)$ for $\varphi([t_1, t_2]) \subseteq G$, provided that G is a simply connected domain of $\mathbb{C}$ and that A is holomorphic (or analytic) on $G \subseteq \mathbb{C}$, i.e. that its matrix entries are complex differentiable at every $z \in G$.

If A is holomorphic (or analytic) on G, a simply connected domain of $\mathbb{C}$, the contour product integral is comparable to a state transition matrix for the differential equation $\frac{dx}{dz}(z) = A(z)x(z)$ with $z \in \mathbb{C}$. For all $z_1, z_2 \in G$, any continuously differentiable curve φ such that $\varphi([t_1, t_2]) \subseteq G$, with $\varphi(t_1) = z_1$ and $\varphi(t_2) = z_2$, provides:

$$x(z_2) = \left(\prod_{\varphi} e^{A(z)dz}\right) x(z_1) \tag{6.41}$$

In particular, this implies Cauchy's product integral theorem. If $\varphi(t_1) = \varphi(t_2)$:

$$\prod_{\varphi} e^{A(z)dz} = I_n \tag{6.42}$$

Proof. See Sections 2.6 and 2.7 of [260], and Chapter 2 of [80] for a more complete exposition of product integration in the complex domain. □

Remark 6.2.14. *There is a clear analogy with the usual Cauchy's integral theorem. If A is holomorphic (or analytic) on a simply connected domain of $\mathbb{C}$ with φ a smooth closed curve of this domain, then:*

$$\int_{\varphi} A(z)dz = 0_n \tag{6.43}$$

6.2.4 Generalized Lie-product formula

This section generalizes the Lie-product formula (or Trotter-Kato formula) stating that:

$$\lim_{m\to+\infty}\left(e^{\frac{1}{m}A}e^{\frac{1}{m}B}\right)^m = e^{A+B} \tag{6.44}$$

This result has already been generalized to a large variety of abstract settings [56]. Yet, to the author's knowledge, this result is not so well-known in product integration theory, nor in control theory. The generalization discussed below might find some application in averaging theory for linear time-periodic systems, or to obtain minimum dwell time guarantees of stability for some classes of switched linear systems, although it will probably not lead to fundamentally new results.

Theorem 6.2.10 (Generalized Lie-product formula)**.** *Let $A:\mathbb{R}\to\mathbb{K}^{n\times n}$ be a piecewise continuous matrix. The following limit holds:*

$$\lim_{m\to+\infty}\left(\prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds}\right)^m = e^{\int_{t_1}^{t_2}A(s)ds} \tag{6.45}$$

The convergence rate is in $O\left(\frac{1}{m}\right)$.

Proof. The proof is similar, with some adjustments, to the proof of the non-generic case (6.44). First, the time-ordering operator is leveraged in order to obtain the following result:

$$\begin{aligned}
&\left(\prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds}\right) - e^{\frac{1}{m}\int_{t_1}^{t_2}A(s)ds} = \mathcal{T}\left(\sum_{k=0}^{+\infty}\frac{1}{k!m^k}\left(\int_{t_1}^{t_2}A(s)ds\right)^k\right) - \sum_{k=0}^{+\infty}\frac{1}{k!m^k}\left(\int_{t_1}^{t_2}A(s)ds\right)^k\\
&=\sum_{k=2}^{+\infty}\frac{1}{k!m^k}\int_{t_1}^{t_2}\cdots\int_{t_1}^{t_2}\left(\mathcal{T}\big(A(s_1)\dots A(s_k)\big) - A(s_1)\dots A(s_k)\right)ds_k\dots ds_1
\end{aligned} \tag{6.46}$$

Hence, the following inequalities hold:

$$\begin{aligned}
&\left\|\left(\prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds}\right) - e^{\frac{1}{m}\int_{t_1}^{t_2}A(s)ds}\right\|_2\\
&\le\sum_{k=2}^{+\infty}\frac{1}{k!m^k}\int_{t_1}^{t_2}\cdots\int_{t_1}^{t_2}\left\|\mathcal{T}\big(A(s_1)\dots A(s_k)\big)\right\|_2 + \left\|A(s_1)\dots A(s_k)\right\|_2 ds_k\dots ds_1\\
&\le\sum_{k=2}^{+\infty}\frac{1}{k!m^k}\int_{t_1}^{t_2}\cdots\int_{t_1}^{t_2}2\left\|A(s_1)\right\|_2\dots\left\|A(s_k)\right\|_2 ds_k\dots ds_1\\
&\le\sum_{k=2}^{+\infty}\frac{2}{k!m^k}\int_{t_1}^{t_2}\cdots\int_{t_1}^{t_2}\left(\sup_{s\in[t_1,t_2]}\|A(s)\|_2\right)^k ds_k\dots ds_1\\
&\le 2\sum_{k=2}^{+\infty}\frac{(t_2-t_1)^k}{k!m^k}\left(\sup_{s\in[t_1,t_2]}\|A(s)\|_2\right)^k = O\left(\frac{1}{m^2}\right)
\end{aligned} \tag{6.47}$$

Moreover, leveraging the identity $X^m - Y^m = \sum_{k=0}^{m-1} X^k(X-Y)Y^{m-k-1}$ and Property 6.2.3, the following inequalities are obtained:

$$\begin{aligned}
&\left\| \left(\prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds} \right)^m - e^{\int_{t_1}^{t_2} A(s)ds} \right\|_2 = \left\| \left(\prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds} \right)^m - \left(e^{\frac{1}{m}\int_{t_1}^{t_2} A(s)ds} \right)^m \right\|_2 \\
&\leq \left\| \sum_{k=0}^{m-1} \left(\prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds} \right)^k \left(\prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds} - e^{\frac{1}{m}\int_{t_1}^{t_2} A(s)ds} \right) e^{\frac{m-k-1}{m}\int_{t_1}^{t_2} A(s)ds} \right\|_2 \\
&\leq \left\| \prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds} - e^{\frac{1}{m}\int_{t_1}^{t_2} A(s)ds} \right\|_2 \cdot \sum_{k=0}^{m-1} \left\| \prod_{t_1}^{t_2} e^{\frac{1}{m}A(s)ds} \right\|_2^k \left\| e^{\frac{1}{m}\int_{t_1}^{t_2} A(s)ds} \right\|_2^{m-k-1} \\
&\leq O\left(\frac{1}{m^2}\right) \cdot m \cdot e^{\frac{m-1}{m}\left|\int_{t_1}^{t_2} \|A(s)\|_2\right| ds} = O\left(\frac{1}{m}\right)
\end{aligned} \tag{6.48}$$

Since $\lim_{m\to+\infty} \frac{1}{m} = 0$, this concludes the proof. □

6.3 Logarithmic norm

The logarithmic norm of a matrix is introduced in this section. Despite its name, it is not a norm, as it can take negative values. It is also called the matrix measure or the Lozinskii measure of a matrix. It consists of a function $\eta : \mathbb{R}^{n\times n} \to \mathbb{R}$ such that for all $t \in \mathbb{R}_{\geq 0}$, the quantity $e^{t\eta(A)}$ is a close upper-bound to $\|e^{tA}\|_2$:

$$\|e^{tA}\|_2 \leq e^{t\eta(A)} \tag{6.49}$$

This inequality can already be verified by simply taking $\eta(A) = \|A\|_2$. However, as the 2-norm is necessarily a positive number, the quantity $e^{t\|A\|_2}$ always explodes to infinity as $t \to +\infty$, which may not be satisfying. More specifically, the underlying goal of the logarithmic norm is to find a measure $\eta(A)$ of A such that if $\lim_{t\to+\infty} \|e^{tA}\|_2 = 0$, then $\lim_{t\to+\infty} e^{t\eta(A)} = 0$ as well. If verified, this powerful property would provide upper-bounds preserving (at least to some extend) the asymptotical behaviour of the quantity which they dominate. In particular, for a Hurwitz matrix A, $\eta(A)$ should ideally be taking a negative value.

The definition of the logarithmic norm of a matrix is introduced in this section, and it is then generalized to obtain the *weighted* logarithmic norm of a matrix. This latter tool is then mixed with product integration in order to obtain useful bounds on the state transition matrix of an LTV system.

6.3.1 Definitions and properties

Some equivalent definitions of the logarithmic norm are given hereafter.

Definition 6.3.1 (Logarithmic norm). *Given $A \in \mathbb{R}^{n\times n}$, the logarithmic norm of A is defined by [72]:*

$$\eta(A) \triangleq \lim_{h\to 0^+} \frac{\|I + hA\|_2 - 1}{h} \tag{6.50}$$

It can also be defined using a ratio of scalar products [130]:

$$\eta(A) = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\langle Ax|x\rangle}{\|x\|_2^2} \tag{6.51}$$

or using the following expression [77, 130]:

$$\eta(A) = \frac{1}{2}\lambda_{\max}(A + A^\top) \tag{6.52}$$

Although it can be demonstrated that the upper-bound (6.49) holds [130], with in general $\eta(A) \leq \|A\|_2$ [177], the following example shows that this notion of logarithmic norm is not completely satisfying yet, as for a Hurwitz matrix A, $\eta(A)$ can sometimes take a positive value.

Example 6.3.1. *Considering the following matrix A:*

$$A = \begin{pmatrix} 0 & 1/2 \\ -1 & -1 \end{pmatrix} \tag{6.53}$$

Since $\lambda_{\max}(A) = -1/2$, A is Hurwitz, hence by the Lyapunov lemma (Lemma 1.1.1 of Chapter 1), $\lim_{t\to+\infty}\|e^{tA}\|_2 = 0$. However, $\eta(A) = \frac{1}{4}(-2+\sqrt{5}) > 0$, hence $\lim_{t\to+\infty} e^{t\eta(A)} = +\infty$. This is at odds with the underlying goal of the logarithmic norm stated in the introduction of the section, since ideally, $\eta(A)$ should yield a negative value.

By taking inspiration from the Lyapunov lemma (Lemma 1.1.1 of Chapter 1), one can get the idea of introducing a positive definite matrix $P \in \mathbb{S}_n^{++}(\mathbb{R})$ in the definition of the logarithmic norm, such that if $V(x) = x^\top Px$ is a QLF demonstrating the exponential stability of $\dot{x}(t) = Ax(t)$, then $\eta(A) < 0$, so that $\lim_{t\to+\infty} e^{t\eta(A)} = 0$. This motivates the introduction of the *weighted* logarithmic norm of a matrix, now denoted η_P, and defined thereafter.

Definition 6.3.2 (Weighted logarithmic norm [130])**.** *Given $A \in \mathbb{R}^{n\times n}$ and $P \in \mathbb{S}_n^{++}(\mathbb{R})$, the P-weighted logarithmic norm of A is defined by:*

$$\eta_P(A) \triangleq \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle Ax|x\rangle_P}{\langle x|x\rangle_P} = \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{x^\top(A^\top P + PA)x}{2x^\top Px} \tag{6.54}$$

It can also be defined by:

$$\eta_P(A) = \frac{1}{2}\lambda_{\max}(P^{\frac{1}{2}}AP^{-\frac{1}{2}} + P^{-\frac{1}{2}}A^\top P^{\frac{1}{2}}) \tag{6.55}$$

It is first verified that there exists at least one QLF $V(x) = x^\top Px$ demonstrating the exponential stability of the system $\dot{x}(t) = Ax(t)$, and providing a logarithmic norm such that $\eta_P(A) < 0$. This weak property is detailed below.

Property 6.3.1. *If $A \in \mathbb{R}^{n\times n}$ is Hurwitz, then $P \in \mathbb{S}_n^{++}(\mathbb{R})$ satisfying the Lyapunov equation:*

$$A^\top P + PA = -2I_n \tag{6.56}$$

provides:

$$\eta_P(A) = -\frac{1}{\lambda_{\max}(P)} < 0 \tag{6.57}$$

Proof. See Theorem 2.3. of [129] □

However, since the 2-norm in the quantity $\|e^{tA}\|_2$ is not itself weighted by P, this current definition allows for some situations where:

$$\|e^A\|_2 \geq e^{\eta_P(A)} \tag{6.58}$$

This is the case with the matrix A defined in (6.53) of Example 6.3.1, as demonstrated in the following example.

Example 6.3.2. *Considering the following matrix P:*

$$P = \begin{pmatrix} 5 & 1 \\ 1 & 3/2 \end{pmatrix} \tag{6.59}$$

this P satisfies $A^\top P + PA = -2I_n$ with A defined in (6.53), yet $\eta_P(A) = -0.1899$, hence $e^{\eta_P(A)} \approx 0.8270$, while $\|e^A\|_2 \approx 1.0133$.

This time, the solution is not to change the definition of the logarithmic norm, but rather to restate its purpose in a wiser way. If there exists a constant k_P term, only depending on P, and such that for all $t \in \mathbb{R}_{\geq 0}$:

$$\|e^{tA}\|_2 \leq k_P e^{t\eta_P(A)} \tag{6.60}$$

then, the inequality (6.49) is still obtained, up to a constant. This constant does not fundamentally change the asymptotic behaviour of the quantity $k_P e^{t\eta_P(A)}$ compared to $e^{t\eta_P(A)}$ as $t \to +\infty$. The next result provides an explicit expression for k_P, and states that if P is such that the QLF $V(x) = x^\top Px$ demonstrates the exponential stability of $\dot{x}(t) = Ax(t)$, then $\eta_P(A) < 0$.

Property 6.3.2. *Given $A \in \mathbb{R}^{n\times n}$ and $P \in \mathbb{S}_n^{++}(\mathbb{R})$, the following inequality holds for all $t \in \mathbb{R}_{\geq 0}$:*

$$\|e^{tA}\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{t\eta_P(A)} \tag{6.61}$$

Moreover if $A^\top P + PA \prec 0$ (implying that A is Hurwitz), then $\eta_P(A) < 0$.

Proof. This is Property 6.3.3, demonstrated later, in the case of a constant matrix A and $t_2 = t_1 + t$. If $A^\top P + PA \prec 0$, the negativity of $\eta_P(A)$ follows directly from $P \in \mathbb{S}_n^{++}(\mathbb{R})$, $(A^\top P + PA) \in \mathbb{S}_n^{--}(\mathbb{R})$, and the definition (6.54) of η_P (Definition 6.3.2). □

6.3.2 Applications to product integration

Equation (6.61) of Property 6.3.2 can be generalized to a setting where the matrix exponential is substituted with an expression leveraging Volterra's product integral.

Property 6.3.3. *Let $A : \mathbb{R} \to \mathbb{R}^{n\times n}$ be a piecewise continuous matrix and $P \in \mathbb{S}_n^{++}(\mathbb{R})$. For all $t_1, t_2 \in \mathbb{R}$ such that $t_1 \leq t_2$, the following holds:*

$$\left\| \prod_{t_1}^{t_2} e^{A(s)ds} \right\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{\int_{t_1}^{t_2} \eta_P(A(s))ds} \tag{6.62}$$

Proof. The function $V(x) = x^\top Px$ is introduced. Its Lie derivative along the solutions of $\dot{x}(t) = A(t)x(t)$ is given by:

$$\dot{V}(x(t)) = x^\top(t)(A^\top(t)P + PA(t))x(t) \tag{6.63}$$

By definition of the logarithmic norm, V verifies for all solution x to $\dot{x}(t) = A(t)x(t)$ and $t \in \mathbb{R}$:

$$\dot{V}(x(t)) \leq 2\eta_P(A(t))V(x(t)) \tag{6.64}$$

which, by Grönwall's inequality [109], yields for all $t \geq t_0$:

$$V(x(t)) \leq e^{2\int_{t_0}^{t} \eta_P(A(s))ds} V(x(t_0)) \tag{6.65}$$

The Property 2.2.1 of Chapter 2 then provides:

$$\|x(t)\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{\int_{t_0}^{t} \eta_P(A(s))ds} \|x(t_0)\|_2 \tag{6.66}$$

and since $x(t) = \left(\prod_{t_0}^{t} e^{A(s)ds}\right) x(t_0)$, the following stands for all $x(t_0) \in \mathbb{R}^n \setminus \{0\}$:

$$\frac{\left\| \left(\prod_{t_0}^{t} e^{A(s)ds}\right) x(t_0) \right\|_2}{\|x(t_0)\|_2} \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{\int_{t_0}^{t} \eta_P(A(s))ds} \tag{6.67}$$

Which concludes the proof, by definition of the spectral norm $\|\cdot\|_2$. □

Moreover, it is now demonstrated that a Lipschitz assumption on the matrix A implies a similar Lipschitz property for the weighted logarithmic norm of this matrix.

Lemma 6.3.1. *Let $P \in \mathbb{S}_n^{++}(\mathbb{R})$. If $A : \mathbb{R} \to \mathbb{R}^{n\times n}$ is a L_A-Lipschitz function, then $\eta_P(A(\cdot))$ is also a Lipschitz function, with for all $t_1, t_2 \in \mathbb{R}$, at least:*

$$|\eta_P(A(t_1)) - \eta_P(A(t_2))| \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} L_A |t_1 - t_2| \tag{6.68}$$

Proof. For all $B_1, B_2 \in \mathbb{R}^{n\times n}$, the following stands:

$$\begin{aligned}\eta_P(B_1+B_2) &= \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle (B_1+B_2)x|x\rangle_P}{\langle x|x\rangle_P} \\ &= \max_{x\in\mathbb{R}^n\setminus\{0\}} \left(\frac{\langle B_1x|x\rangle_P}{\langle x|x\rangle_P} + \frac{\langle B_2x|x\rangle_P}{\langle x|x\rangle_P}\right) \\ &\leq \left(\max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle B_1x|x\rangle_P}{\langle x|x\rangle_P}\right) + \left(\max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle B_1x|x\rangle_P}{\langle x|x\rangle_P}\right) \\ \eta_P(B_1+B_2) &\leq \eta_P(B_1) + \eta_P(B_2)\end{aligned} \tag{6.69}$$

Now, for all $t_1, t_2 \in \mathbb{R}$, the following holds:

$$\begin{aligned}\eta_P(A(t_1)) &= \eta_P(A(t_2) + A(t_1) - A(t_2)) \\ &\leq \eta_P(A(t_2)) + \eta_P(A(t_1) - A(t_2)) \\ \Rightarrow \eta_P(A(t_1)) - \eta_P(A(t_2)) &\leq \eta_P(A(t_1) - A(t_2))\end{aligned} \tag{6.70}$$

Moreover, by the Property 2.2.1 of Chapter 2:

$$\begin{aligned}\eta_P(A(t_1) - A(t_2)) &\leq \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle (A(t_1)-A(t_2))x|x\rangle_P}{\lambda_{\min}(P)\|x\|_2^2} \\ &\leq \frac{1}{\lambda_{\min}(P)} \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle P(A(t_1)-A(t_2))x|x\rangle}{\|x\|_2^2} \\ &\leq \frac{1}{\lambda_{\min}(P)} \eta\left(P(A(t_1)-A(t_2))\right)\end{aligned} \tag{6.71}$$

Leveraging the inequality $\eta(A) \leq \|A\|_2$ [177], the following is obtained:

$$\eta_P(A(t_1) - A(t_2)) \leq \frac{1}{\lambda_{\min}(P)} \|P(A(t_1) - A(t_2))\|_2 \tag{6.72}$$

which, by submultiplicativity of the spectral norm, provides:

$$\eta_P(A(t_1)) - \eta_P(A(t_2)) \leq \frac{\|P\|_2}{\lambda_{\min}(P)} \|A(t_1) - A(t_2)\|_2 \tag{6.73}$$

and since $P \in \mathbb{S}_n^{++}(\mathbb{R})$, $\|P\|_2 = \lambda_{\max}(P)$. Interverting the role of t_1 and t_2 in the previous equations yields the same upper-bound. Since A is L_A-Lipschitz, combining (6.73) with its interverted counterpart finally provides:

$$|\eta_P(A(t_1)) - \eta_P(A(t_2))| \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \|A(t_1) - A(t_2)\|_2 \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} L_A |t_1 - t_2| \tag{6.74}$$

□

Assuming that the values of A are only known up to a time t_0, it is possible to still find an equivalent to Property 6.3.3 for $t_1, t_2 \geq t_0$, by assuming that A remains Lipschitz, or that its logarithmic norm remains bounded.

Lemma 6.3.2. *Let $P \in \mathbb{S}_n^{++}(\mathbb{R})$ and let $A : \mathbb{R} \to \mathbb{R}^{n\times n}$ be a L_A-Lipschitz function with for all $t \in \mathbb{R}$, $\eta_P(A(t)) \leq \sigma$. Let τ be defined by $\tau \triangleq t_0 + \frac{\sigma - \eta_P(A(t_0))}{L_\eta}$ where L_η is the Lipschitz constant of $\eta_P(A(\cdot))$, with $L_\eta \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} L_A$. For all $t_0, t_1, t_2 \in \mathbb{R}$ such that $t_0 \leq t_1 \leq t_2$:*

$$\left\| \prod_{t_1}^{t_2} e^{A(s)ds} \right\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{K(\tau)} \tag{6.75}$$

where

$$K(\tau) \triangleq \begin{cases} (t_2 - t_1)\sigma & \text{if } \tau \leq t_1 \\ t_2\sigma - t_1(\eta_P(A(t_0)) - \frac{L_\eta}{2}(2t_0 - t_1)) - \frac{1}{2L_\eta}(\sigma - \eta_P(A(t_0)) + L_\eta t_0)^2 & \text{if } \tau \in (t_1, t_2) \\ (t_2 - t_1)\eta_P(A(t_0)) + \frac{1}{2}L_\eta \left(t_2^2 - t_1^2 + 2t_0(t_1 - t_2)\right) & \text{if } \tau \geq t_2 \end{cases} \tag{6.76}$$

Proof. Lemma 6.3.1 provides the inequality $L_\eta \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} L_A$. Moreover:

$$\int_{t_1}^{t_2} \eta_P(A(s))ds \leq \int_{t_1}^{t_2} \min(\eta_P(A(t_0) + L_\eta(s - t_0), \sigma)ds \tag{6.77}$$

and since $s \mapsto \eta_P(A(t_0)) + L_\eta(s - t_0)$ is an increasing map, the following inequality follows

$$\begin{aligned} \int_{t_1}^{t_2} \eta_P(A(s))ds &\leq \min_{s\in(t_1,t_2)} (s - t_1)\eta_P(A(t_0)) + \frac{1}{2}L_\eta(s^2 - 2t_0 s - t_1^2 + 2t_0 t_1) + (t_2 - s)\sigma \\ &= \min_{s\in(t_1,t_2)} as^2 + bs + c \end{aligned} \tag{6.78}$$

where:

$$a = \frac{1}{2}L_\eta > 0 \tag{6.79a}$$

$$b = \eta_P(A(t_0)) - L_\eta t_0 - \sigma \tag{6.79b}$$

$$c = L_\eta(t_0 - t_1/2)t_1 + t_2\sigma - t_1\eta_P(A(t_0)) \tag{6.79c}$$

The minimum of $as^2 + bs + c$ being reached for $\tau = -b/2a$, the sharpest upper-bound of (6.77) depends on whether τ belongs to (t_1, t_2) or not.

$$\int_{t_1}^{t_2} \eta_P(A(s))ds \leq \begin{cases} (t_2 - t_1)\sigma & \text{if } \tau \leq t_1 \\ c - b^2/(4a) & \text{if } \tau \in (t_1, t_2) \\ (t_2 - t_1)\eta_P(A(t_0)) + \frac{1}{2}L_\eta \left(t_2^2 - t_1^2 + 2t_0(t_1 - t_2)\right) & \text{if } \tau \geq t_2 \end{cases} \tag{6.80}$$

The results are then applied to Property 6.3.3, providing (6.75) and thus concluding the proof. □

The following observations can be made about the inequalities of Lemma 6.3.2.

Observation 6.3.1. *The value of τ can be seen as the characteristic time above which the upper-bound $\eta_P(A(t)) \leq \sigma$ becomes advantageous in the inequalities compared to the L_η-Lipschitzness of $\eta_P(A(\cdot))$. In particular, the first inequality of (6.75) only relies on the upper-bound $\eta_P(A(t)) \leq \sigma$, whereas the last inequality only relies on the Lipschitz assumption made on A. The second inequality benefits from both assumptions.*

Observation 6.3.2. *The first and last upper-bounds of (6.75) remain true for all values of τ, but are less sharp than the second one for $\tau \in (t_1, t_2)$. This second upper-bound only holds for $\tau \in (t_1, t_2)$.*

The value of the state transition matrix of $\dot{x}(t) = A(t)x(t)$ between t_1 and t_2 is now compared to the value of the state transition matrix of the system $\dot{x}(t) = A(t_0)x(t)$ between t_1 and t_2, where $A(t_0)$ is assumed to be the last known value of A.

Lemma 6.3.3. *Let $P \in \mathbb{S}_n^{++}(\mathbb{R})$ and let $A : \mathbb{R} \to \mathbb{R}^{n\times n}$ be a L_A-Lipschitz function with for all $t \in \mathbb{R}$, $\eta_P(A(t)) \leq \sigma$. Again, L_η is the Lipschitz constant of $\eta_P(A(\cdot))$, with $L_\eta \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} L_A$. For all $t_0, t_1, t_2 \in \mathbb{R}$ such that $t_0 \leq t_1 \leq t_2$, the following upper-bounds hold:*

$$\left\| \prod_{t_1}^{t_2} e^{A(s)ds} - e^{(t_2-t_1)A(t_0)} \right\|_2 \leq \frac{L_A}{2} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(t_2^2 - t_1^2 + 2t_0(t_1 - t_2)\right) e^{(t_2-t_1)\sigma} \tag{6.81a}$$

$$\left\| \prod_{t_1}^{t_2} e^{A(s)ds} - e^{(t_2-t_1)A(t_0)} \right\|_2 \leq \frac{L_A}{L_\eta} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(e^{\frac{1}{2}L_\eta\left(t_2^2 - t_1^2 + 2t_0(t_1-t_2)\right)} - 1\right) e^{(t_2-t_1)\eta_P(A(t_0))} \tag{6.81b}$$

Proof. Given $t_0, t_1, t_2 \in \mathbb{R}$ such that $t_0 \leq t_1 \leq t_2$, Duhamel's formula (Theorem 6.2.7) provides:

$$\begin{aligned} \prod_{t_1}^{t_2} e^{A(s)ds} - e^{(t_2-t_1)A(t_0)} &= \prod_{t_1}^{t_2} e^{A(s)ds} - \prod_{t_1}^{t_2} e^{A(t_0)ds} \\ &= \int_{t_1}^{t_2} \left(\prod_{v}^{t_2} e^{A(t_0)ds}\right) (A(v) - A(t_0)) \left(\prod_{t_1}^{v} e^{A(s)ds}\right) dv \\ &= \int_{t_1}^{t_2} e^{(t_2-v)A(t_0)} (A(v) - A(t_0)) \left(\prod_{t_1}^{v} e^{A(s)ds}\right) dv \end{aligned} \tag{6.82}$$

then, by submultiplicativity of the spectral norm and the Lipschitz assumption on A, the following is obtained:

$$\begin{aligned} \left\| \prod_{t_1}^{t_2} e^{A(s)ds} - e^{(t_2-t_1)A(t_0)} \right\|_2 &\leq \int_{t_1}^{t_2} \left\| e^{(t_2-v)A(t_0)} \right\|_2 \|A(v) - A(t_0)\|_2 \left\| \prod_{t_1}^{v} e^{A(s)ds} \right\|_2 dv \\ &\leq L_A \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \int_{t_1}^{t_2} e^{(t_2-v)\eta_P(A(t_0))} (v - t_0) \left\| \prod_{t_1}^{v} e^{A(s)ds} \right\|_2 dv \end{aligned} \tag{6.83}$$

Applying the first upper-bound of Lemma 6.3.2 to (6.83) provides the following inequalities:

$$\begin{aligned}\left\|\prod_{t_1}^{t_2} e^{A(s)ds} - e^{(t_2-t_1)A(t_0)}\right\|_2 &\leq L_A \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(\int_{t_1}^{t_2} (v-t_0)dv\right) e^{(t_2-t_1)\sigma} \\ &\leq \frac{L_A}{2}\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(t_2^2 - t_1^2 + 2t_0(t_1-t_2)\right) e^{(t_2-t_1)\sigma}\end{aligned} \tag{6.84}$$

Similarly, applying the third upper-bound of Lemma 6.3.2 to (6.83) provides the following inequalities:

$$\begin{aligned}&\left\|\prod_{t_1}^{t_2} e^{A(s)ds} - e^{(t_2-t_1)A(t_0)}\right\|_2 \\ &\leq L_A \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \int_{t_1}^{t_2} e^{(t_2-v)\eta_P(A(t_0))}(v-t_0)e^{(v-t_1)\eta_P(A(t_0))+\frac{1}{2}L_\eta\left(v^2-t_1^2+2t_0(t_1-v)\right)}dv \\ &\leq L_A \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(\int_{t_1}^{t_2} (u-t_0)e^{\frac{1}{2}L_\eta(u^2-2t_0u)}du\right) e^{-\frac{1}{2}L_\eta(t_1^2-2t_0t_1)+(t_2-t_1)\eta_P(A(t_0))} \\ &\leq \frac{L_A}{L_\eta}\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(e^{\frac{1}{2}L_\eta\left(t_2^2-t_1^2+2t_0(t_1-t_2)\right)} - 1\right) e^{(t_2-t_1)\eta_P(A(t_0))}\end{aligned} \tag{6.85}$$

It is noticed that this second inequality is obtained without relying on the assumption that $\eta_P(A(t)) \leq \sigma$. □

The following observation can be made about the inequalities of Lemma 6.3.3.

Observation 6.3.3. *The second inequality* (6.81b) *only relies on the Lipschitz assumption made on* A*, whereas the first one* (6.81a) *also benefits from the upper-bound* $\eta_P(A(t)) \leq \sigma$*.*

Observation 6.3.4. *Depending on the context,* (6.81a) *or* (6.81b) *can be advantageous.*

6.4 Applications to LPV systems

This section quantifies the greatest possible discrepancy between the real future of a continuous-time LPV system with a Lipschitz assumption on the scheduling vector θ, and an artificially constructed prediction of the future for which the scheduling vector θ and the input are being held constant. Quantifying this discrepancy allows to obtain an exact discretization of this LPV system, as well as an estimation of the value of its Gramians in the future, and thus an evaluation of their possible loss of controllability or observability.

6.4.1 Exact discretization

The expression of the continuous-time LPV system investigated in this section is given as follows.

$$\dot{x}(t) = A(\theta(t))x(t) + B(\theta(t))u(t) \tag{6.86a}$$
$$y(t) = C(\theta(t))x(t) \tag{6.86b}$$

where A, B and C depend on the scheduling vector θ, which is exclusively known in real-time. This expression has, up to the real-time knowledge assumption made on θ, the same nature as

its LTV counterpart, hence, in the following, the LTV notations are kept in the LPV context, due to their concision.

$$A(t) \equiv A(\theta(t)),\ B(t) \equiv B(\theta(t)),\ C(t) \equiv C(\theta(t)) \tag{6.87}$$

Throughout this section, a $P \in \mathbb{S}_n^{++}(\mathbb{R})$ is fixed, and the following assumptions are made:

Assumption 6.4.1. *A is a L_A-Lipschitz function. Moreover, L_η is the Lipschitz constant of $\eta_P(A(\cdot))$, with $L_\eta \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} L_A$ (Lemma 6.3.1).*

Assumption 6.4.2. *B is a L_B-Lipschitz function.*

Assumption 6.4.3. *$\sigma \in \mathbb{R}$ upper-bounds the P-weighted logarithmic norm of A, i.e.*

$$\sup_{t \in \mathbb{R}} \eta_P(A(t)) \leq \sigma \tag{6.88}$$

The Assumptions 6.4.1 and 6.4.2 are similar to the usual assumption that the scheduling vector θ has a bounded rate of instantaneous variations, and the values for L_A and L_B can usually be derived from (6.87) using the bounds on the first derivative of θ. Similarly, the Assumption 6.4.3 can be viewed as a generalization of the usual assumption that the scheduling vector θ remains in a bounded set $\Theta \subset \mathbb{R}^{n_\theta}$. The value of σ can easily be deduced from the bounds of A thanks to the Lemma 1 of [143].

Knowing the values of A and B at a time t and assuming the control u is being held constant between t and $t+T$, an estimation of the discretized state-space representation of the system (6.86) at a given time t and with a sampling period T can generally be obtained using the following zero-order hold assumption on θ.

Assumption 6.4.4. *(Zero-order hold) For all $s \in [t, t+T]$, $A(s) = A(t)$, $B(s) = B(t)$, $C(s) = C(t)$.*

The zero-order hold discretization of (6.86) is given as follows:

$$x_{ZOH}(t+T) = e^{TA(t)} x(t) + \left(\int_0^T e^{sA(t)} ds \right) B(t) u(t) \tag{6.89a}$$

$$y(t) = C(t) x(t) \tag{6.89b}$$

The idea of this section consists in estimating the error made under this zero-order hold assumption, thus upper-bounding the spectral norm of $\Delta_A(t)$ and $\Delta_B(t)$, two matrices modeling the discretization error in the representation and taken such that the following equality is satisfied:

$$x(t+T) = \left(e^{TA(t)} + \Delta_A(t) \right) x(t) + \left(\left(\int_0^T e^{sA(t)} ds \right) B(t) + \Delta_B(t) \right) u(t) \tag{6.90}$$

The system (6.90) can be viewed as the *exact discretization* of the LPV system (6.86) at a sampling period T. This discretization is exact in the sense that it takes into account the uncertainties introduced by the zero-order hold discretization of the parameter θ, which, to the author's knowledge, were previously not taken into account in the literature [281, 280]. This discretization can therefore be used in order to synthesize a sampled controller or observer for the continuous-time LPV system. The bounds on $\Delta_A(t)$ and $\Delta_B(t)$ are provided by the theorem hereafter.

Theorem 6.4.1 (Discretization error). *Under Assumptions 6.4.1, 6.4.2 and 6.4.3, the following inequalities hold on the spectral norm of $\Delta_A(t)$:*

$$\|\Delta_A(t)\|_2 \leq \frac{L_A}{2}\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}T^2 e^{\sigma T} \tag{6.91a}$$

$$\|\Delta_A(t)\|_2 \leq \frac{L_A}{L_\eta}\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}\left(e^{\frac{1}{2}L_\eta T^2}-1\right)e^{\eta_P(A(t))T} \tag{6.91b}$$

Similarly, by introducing $\Delta_{B,1}(t)$ and $\Delta_{B,2}(t)$ such that:

$$\|\Delta_B(t)\|_2 \leq \|\Delta_{B,1}(t)\|_2 + \|\Delta_{B,2}(t)\|_2\|B(t)\|_2 \tag{6.92}$$

the following inequalities hold on the spectral norm of $\Delta_{B,1}(t)$:

$$\|\Delta_{B,1}(t)\|_2 \leq L_B\sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}\int_0^T re^{(T-r)\sigma}dr \tag{6.93a}$$

$$\|\Delta_{B,1}(t)\|_2 \leq L_B\sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}\int_0^T re^{(T-r)\eta_P(A(t))+\frac{1}{2}L_\eta(T^2-r^2)}dr \tag{6.93b}$$

and the following inequalities hold on the spectral norm of $\Delta_{B,2}(t)$:

$$\|\Delta_{B,2}(t)\|_2 \leq \frac{L_A}{2}\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}\int_0^T \left(T^2-r^2\right)e^{\sigma(T-r)}dr \tag{6.94a}$$

$$\|\Delta_{B,2}(t)\|_2 \leq \frac{L_A}{L_\eta}\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}\int_0^T \left(e^{\frac{1}{2}L_\eta(T^2-r^2)}-1\right)e^{(T-r)\eta_P(A(t))}dr \tag{6.94b}$$

Proof. Equations (6.90) and (6.28) of Corollary 6.2.1 provide:

$$\Delta_A(t) = \prod_t^{t+T} e^{A(s)ds} - e^{TA(t)} \tag{6.95a}$$

$$\Delta_B(t) = \int_t^{t+T}\left(\prod_r^{t+T} e^{A(s)ds}B(r) - e^{(t+T-r)A(t)}B(t)\right)dr \tag{6.95b}$$

Hence, the bounds on $\Delta_A(t)$ are directly obtained by applying Lemma 6.3.3 with $t_0 = t_1 = t$ and $t_2 = t+T$. The bounds on $\Delta_B(t)$ are more difficult to obtain, and $\Delta_B(t)$ is rewritten as follows:

$$\begin{aligned}\Delta_B(t) &= \int_t^{t+T}\prod_r^{t+T} e^{A(s)ds}\left(B(r)-B(t)\right)dr + \int_t^{t+T}\left(\prod_r^{t+T}e^{A(s)ds} - \prod_r^{t+T}e^{A(t)ds}\right)dr\, B(t) \\ &= \Delta_{B,1}(t) + \Delta_{B,2}(t)B(t)\end{aligned} \tag{6.96}$$

In particular, by applying Lemma 6.3.2 with $t_0 = t$, $t_1 = t+r$ and $t_2 = t+T$, the following can

be found:

$$\begin{aligned}\|\Delta_{B,1}(t)\|_2 &\leq L_B \int_0^T \left\| \prod_{t+r}^{t+T} e^{A(s)ds} \right\|_2 r dr \\ &\leq L_B \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \begin{cases} \int_0^T r e^{(T-r)\sigma} dr \\ \int_0^T r e^{(T-r)\eta_P(A(t)) + \frac{1}{2}L_\eta (T^2-r^2)} dr \end{cases}\end{aligned} \tag{6.97}$$

Finally, applying Lemma 6.3.3 with $t_0 = t$, $t_1 = t + r$ and $t_2 = t + T$ provides:

$$\|\Delta_{B,2}(t)\|_2 \leq \frac{L_A}{2} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \int_0^T \left(T^2 - r^2\right) e^{\sigma(T-r)} dr \tag{6.98a}$$

$$\|\Delta_{B,2}(t)\|_2 \leq \frac{L_A}{L_\eta} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \int_0^T \left(e^{\frac{1}{2}L_\eta (T^2-r^2)} - 1\right) e^{(T-r)\eta_P(A(t))} dr \tag{6.98b}$$

□

Several observations should be made on these inequalities.

Observation 6.4.1. *The inequalities* (6.91b), (6.93b) *and* (6.94b) *only rely on the Lipschitz constants of Assumptions 6.4.1 and 6.4.2, whereas the others also benefit from the value of the upper-bound σ of Assumption 6.4.3.*

Observation 6.4.2. *The upper-bounds* (6.91b), (6.93b) *and* (6.94b) *are time-varying, since they depend on the value of $\eta_P(A(t))$. However, they can easily be made time-invariant by leveraging Assumption 6.4.3.*

Depending on the context, some inequalities can be more advantageous than others. The previous inequalities are now illustrated through the study of two examples.

Example 6.4.1. *Consider the following one-dimensional LPV system:*

$$\dot{x}(t) = \theta(t)x(t) + u(t) \tag{6.99}$$

where $\theta : \mathbb{R} \to \mathbb{R}$ is assumed to be a 1-Lipschitz scheduling parameter, only known in real-time, and such that $\theta(0) = 0$. Let $T \in \mathbb{R}_{>0}$. It is assumed that the control input u is held constant on $[0, T]$. The inequalities (6.91b), (6.93b) *and* (6.94b) *of Theorem 6.4.1, considered with:*

$$P = \lambda_{\max}(P) = \lambda_{\min}(P) = L_A = L_\eta = 1, \quad \eta_P(A(0)) = L_B = 0 \tag{6.100}$$

provide:

$$x(T) = (1 + \Delta_A)x(0) + (T + \Delta_B)u(0) \tag{6.101}$$

where:

$$|\Delta_A| \leq e^{\frac{1}{2}T^2} - 1 \tag{6.102a}$$

$$|\Delta_B| \leq |\Delta_{B,1}| + |\Delta_{B,2}| \leq 0 + \left(\int_0^T e^{\frac{1}{2}(T^2-r^2)} dr \right) - T \tag{6.102b}$$

After observing the scheduling vector θ during the time span $[0, T]$, suppose its evolution was given by:

$$\theta(t) = t \tag{6.103}$$

The true value of $x(T)$ can now be obtained as follows:

$$x(T) = e^{\frac{1}{2}T^2}x(0) + \int_0^T \left(e^{\frac{1}{2}(T^2-r^2)}dr\right) u(0) \tag{6.104}$$

Hence, $\Delta_A = e^{\frac{1}{2}T^2} - 1$ and $\Delta_B = \left(\int_0^T e^{\frac{1}{2}(T^2-r^2)}dr\right) - T$. The upper-bounds (6.102a) and (6.102b) are obtained, which illustrates the sharpness of the inequalities (6.91b), (6.93b) and (6.94b) in this context.

Example 6.4.2. *Consider the following LPV system:*

$$\dot{x}(t) = A(\theta(t))x(t) + B(\theta(t))u(t) \tag{6.105a}$$

$$\text{with } A(\theta(t)) = \begin{pmatrix} \theta_1(t) & 1/2 \\ -1 & -1 \end{pmatrix} \text{ and } B(\theta(t)) = \begin{pmatrix} \theta_2(t) \\ 1 \end{pmatrix} \tag{6.105b}$$

where $\theta : \mathbb{R} \to \mathbb{R}^2$ is assumed to be a 1-Lipschitz scheduling parameter, only known in real-time, and such that $\theta(0) = 0$. It is easily deduced that $A(\theta(\cdot))$ and $B(\theta(\cdot))$ are 1-Lipschitz functions as well. It is also assumed that $\eta_P(A(t)) \leq 1$, and that the control input u is held constant on $[0, T]$. This provides:

$$x(T) = \left(\left(e^{TA(\theta(0))} + \Delta_A\right)x(0) + \left(\left(\int_0^T e^{sA(0)}ds\right)B(\theta(0)) + \Delta_B\right)u(0)\right) \tag{6.106}$$

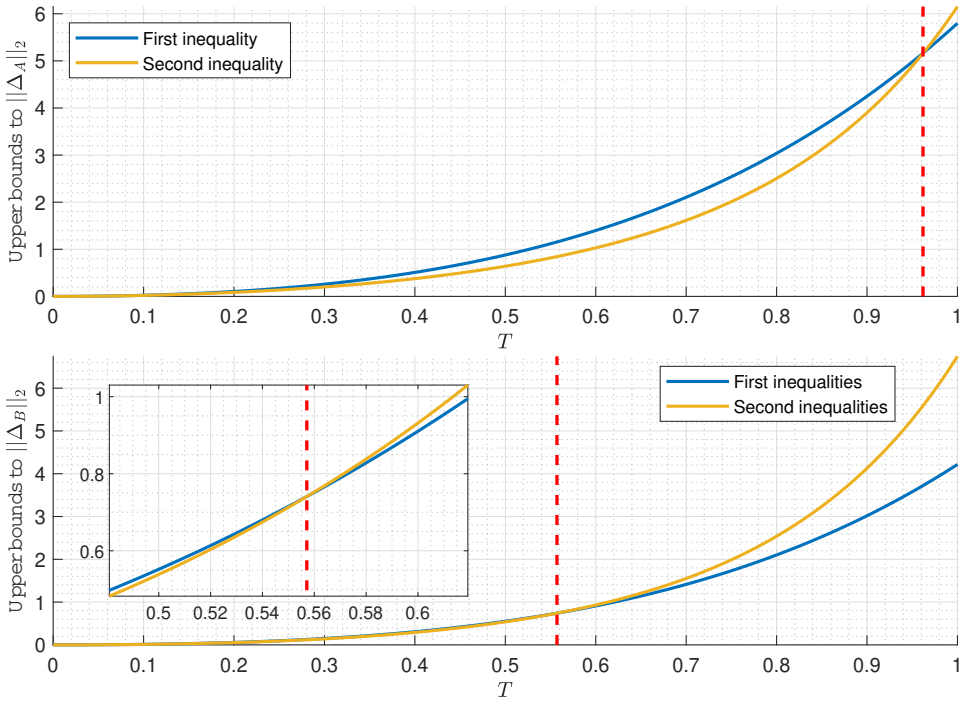


Figure 6.2: Upper-bounds to $\|\Delta_A\|_2$ and $\|\Delta_B\|_2 \leq \|\Delta_{B,1}\|_2 + \|\Delta_{B,2}\|_2\|B(\theta(0))\|_2$ of Example 6.4.2 obtained with the inequalities of Theorem 6.4.1, for $T \in [0, 1]$. In the first plot, the red dotted line represents the threshold above which the inequality (6.91a) becomes advantageous compared to (6.91b) to upper-bound $\|\Delta_A\|_2$. Similarly, in the second plot, the red dotted line represents the threshold above which the inequalities (6.93a) and (6.94a) on $\|\Delta_{B,1}\|_2$ and $\|\Delta_{B,2}\|_2$ become advantageous compared to (6.93b) and (6.94b) to upper-bound $\|\Delta_B\|_2$.

with $\|\Delta_B\|_2 \leq \|\Delta_{B,1}\|_2 + \|\Delta_{B,2}\|_2\|B(\theta(0))\|_2$. *All the inequalities of Theorem 6.4.1 are considered with:*

$$P = \begin{pmatrix} 5 & 1 \\ 1 & 3/2 \end{pmatrix}, \tag{6.107a}$$

$$\lambda_{\min}(P) = 1.2344\ldots, \qquad \lambda_{\max}(P) = 5.2656\ldots, \tag{6.107b}$$

$$L_A = 1, \qquad L_\eta = 4.2656\ldots, \tag{6.107c}$$

$$L_B = 1, \qquad \|B(\theta(0))\|_2 = 1, \tag{6.107d}$$

$$\eta_P(A(0)) = -0.1899\ldots, \qquad \sigma = 1 \tag{6.107e}$$

These inequalities provide upper-bounds to $\|\Delta_A\|_2$ *and* $\|\Delta_B\|_2 \leq \|\Delta_{B,1}\|_2 + \|\Delta_{B,2}\|_2\|B(\theta(0))\|_2$, *which are plotted in Figure 6.2 for* $T \in [0, 1]$. *It can be noticed that the inequalities* (6.91b), (6.93b) *and* (6.94b) *provide sharper bounds to* $\|\Delta_A\|_2$ *and* $\|\Delta_B\|_2$ *in the short term compared to the inequalities* (6.91a), (6.93a) *and* (6.94a), *which become advantageous for larger values of* T.

6.4.2 Near future controllability and observability

In this section, sufficient conditions to guarantee controllability and observability of an LPV system in the near future are obtained. Such conditions are based on a zero-order hold estimation of the controllability and observability Gramians, where the estimation error is upper-bounded with the help of the previous results of this chapter. As an introduction to this section, the Gramian-based conditions of controllability and observability for the continuous-time LTV system (6.108) are recalled.

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \tag{6.108a}$$

$$y(t) = C(t)x(t) \tag{6.108b}$$

Theorem 6.4.2 (Usual controllability and observability criteria)**.** *The LTV system* (6.108) *is controllable (resp. observable) on* $[t_1, t_2]$ *if and only if*

$$W_c(t_2, t_1) \triangleq \int_{t_1}^{t_2} \left(\prod_{s}^{t_2} e^{A(r)dr}\right) B(s)B^\top(s) \left(\prod_{s}^{t_2} e^{A(r)dr}\right)^\top ds \succ 0 \tag{6.109a}$$

$$\text{resp. } W_o(t_2, t_1) \triangleq \int_{t_1}^{t_2} \left(\prod_{t_1}^{s} e^{A(r)dr}\right)^\top C^\top(s)C(s) \left(\prod_{t_1}^{s} e^{A(r)dr}\right) ds \succ 0 \tag{6.109b}$$

Proof. This is a rewriting of the Theorem 1.1.2 of Chapter 1 in the continuous-time LTV case. □

This section considers the following LPV system.

$$\dot{x}(t) = A(\theta(t))x(t) + B(t)u(t) \tag{6.110a}$$

$$y(t) = C(t)x(t) \tag{6.110b}$$

The matrices B and C are now assumed to be continuous functions whose values are known in advance at all time $t \in \mathbb{R}$. The scheduling parameter θ providing the values of A in real-time is

still denoted implicitly. Since A cannot be known in advance, assuming that the last measured value of θ is obtained at a time t, a zero-order hold estimation $A(s) = A(t)$ is considered for $s \in [t_1, t_2]$. An estimation of the controllability and observability Gramians (6.109) of the system (6.110) between t_1 and t_2, with $t \leq t_1 < t_2$, can be carried out using this zero-order hold assumption on θ.

$$\hat{W}_c(t_2, t_1) = \int_{t_1}^{t_2} e^{(t_2-s)A(t)} B(s) B^\top(s) e^{(t_2-s)A^\top(t)} ds \tag{6.111a}$$

$$\hat{W}_o(t_2, t_1) = \int_{t_1}^{t_2} e^{(s-t_1)A^\top(t)} C^\top(s) C(s) e^{(s-t_1)A(t)} ds \tag{6.111b}$$

The idea of this section consists in obtaining upper-bounds to $\|\Delta_{W_c}(t_2, t_1)\|_2$ and $\|\Delta_{W_o}(t_2, t_1)\|_2$ in (6.112), where the two matrices $\Delta_{W_c}(t_2, t_1)$ and $\Delta_{W_o}(t_2, t_1)$ define the error made in the Gramian calculations.

$$\Delta_{W_c}(t_2, t_1) = W_c(t_2, t_1) - \hat{W}_c(t_2, t_1) \tag{6.112a}$$

$$\Delta_{W_o}(t_2, t_1) = W_o(t_2, t_1) - \hat{W}_o(t_2, t_1) \tag{6.112b}$$

Theorem 6.4.3 (Gramian inequalities). *Let τ be defined by $\tau \triangleq t + \frac{\sigma - \eta_P(A(t))}{L_\eta}$. Under Assumptions 6.4.1 and 6.4.3, the following inequalities hold on the spectral norm of $\Delta_{W_c}(t_2, t_1)$ and $\Delta_{W_o}(t_2, t_1)$ for all $t_1 \in [t, t_2)$:*

$$\|\Delta_{W_c}(t_2, t_1)\|_2 \leq \int_{t_1}^{t_2} \|\Delta_{c,1}(s)\|_2 \left(\|\Delta_{c,2}(s)\|_2 + \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{(t_2-s)\eta_P(A(t))} \right) \|B(s)\|_2^2 ds \tag{6.113a}$$

$$\|\Delta_{W_o}(t_2, t_1)\|_2 \leq \int_{t_1}^{t_2} \|\Delta_{o,1}(s)\|_2 \left(\|\Delta_{o,2}(s)\|_2 + \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{(s-t_1)\eta_P(A(t))} \right) \|C(s)\|_2^2 ds \tag{6.113b}$$

where:

$$\|\Delta_{c,1}(s)\|_2 \leq \frac{L_A}{2} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(t_2^2 - s^2 + 2t(s - t_2) \right) e^{(t_2-s)\sigma} \tag{6.114a}$$

$$\|\Delta_{c,1}(s)\|_2 \leq \frac{L_A}{L_\eta} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(e^{\frac{1}{2}L_\eta \left(t_2^2 - s^2 + 2t(s-t_2) \right)} - 1 \right) e^{(t_2-s)\eta_P(A(t))} \tag{6.114b}$$

$$\|\Delta_{o,1}(s)\|_2 \leq \frac{L_A}{2} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(s^2 - t_1^2 + 2t(t_1 - s) \right) e^{(s-t_1)\sigma} \tag{6.114c}$$

$$\|\Delta_{o,1}(s)\|_2 \leq \frac{L_A}{L_\eta} \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(e^{\frac{1}{2}L_\eta \left(s^2 - t_1^2 + 2t(t_1-s) \right)} - 1 \right) e^{(s-t_1)\eta_P(A(t))} \tag{6.114d}$$

and

$$\|\Delta_{c,2}(s)\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \begin{cases} e^{(t_2-s)\sigma} & \text{if } \tau \leq s \\ e^{t_2\sigma - s(\eta_P(A(t)) - \frac{L_\eta}{2}(2t-s)) - \frac{1}{2L_\eta}(\sigma - \eta_P(A(t)) + L_\eta t)^2} & \text{if } \tau \in (s, t_2) \\ e^{(t_2-s)\eta_P(A(t)) + \frac{1}{2}L_\eta\left(t_2^2 - s^2 + 2t(s-t_2)\right)} & \text{if } \tau \geq t_2 \end{cases} \tag{6.115a}$$

$$\|\Delta_{o,2}(s)\|_2 \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \begin{cases} e^{(s-t_1)\sigma} & \text{if } \tau \leq t_1 \\ e^{s\sigma - t_1(\eta_P(A(t)) - \frac{L_\eta}{2}(2t-t_1)) - \frac{1}{2L_\eta}(\sigma - \eta_P(A(t)) + L_\eta t)^2} & \text{if } \tau \in (t_1, t_2) \\ e^{(s-t_1)\eta_P(A(t)) + \frac{1}{2}L_\eta\left(s^2 - t_1^2 + 2t(t_1-s)\right)} & \text{if } \tau \geq s \end{cases} \tag{6.115b}$$

Proof. First, for any matrices $E, F \in \mathbb{R}^{p\times q}$, the following identity holds:

$$2\left(EE^\top - FF^\top\right) = (E+F)(E-F)^\top + (E-F)(E+F)^\top \tag{6.116}$$

hence:

$$\|EE^\top - FF^\top\|_2 \leq \|E-F\|_2\|E+F\|_2 \tag{6.117}$$

Now, applying this inequality to (6.112), with:

$$E = \left(\prod_s^{t_2} e^{A(r)dr}\right) B(s), \qquad \text{resp. } E = \left(\prod_{t_1}^{s} e^{A(r)dr}\right)^\top C^\top(s) \tag{6.118a}$$

$$F = e^{(t_2-s)A(t)}B(s), \qquad \text{resp. } F = e^{(s-t_1)A^\top(t)}C^\top(s) \tag{6.118b}$$

provides:

$$\|\Delta_{W_c}(t_2,t_1)\|_2 \leq \int_{t_1}^{t_2} \|\Delta_{c,1}(s)\|_2\|\tilde{\Delta}_{c,2}(s)\|_2\|B(s)\|_2^2 ds \tag{6.119a}$$

$$\|\Delta_{W_o}(t_2,t_1)\|_2 \leq \int_{t_1}^{t_2} \|\Delta_{o,1}(s)\|_2\|\tilde{\Delta}_{o,2}(s)\|_2\|C(s)\|_2^2 ds \tag{6.119b}$$

with:

$$\|\Delta_{c,1}(s)\|_2 = \left\|\prod_s^{t_2} e^{A(r)dr} - e^{(t_2-s)A(t)}\right\|_2 \tag{6.120a}$$

$$\|\tilde{\Delta}_{c,2}(s)\|_2 = \left\|\prod_s^{t_2} e^{A(r)dr} + e^{(t_2-s)A(t)}\right\|_2 \leq \left\|\prod_s^{t_2} e^{A(r)dr}\right\|_2 + \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{(t_2-s)\eta_P(A(t))} \tag{6.120b}$$

$$\|\Delta_{o,1}(s)\|_2 = \left\|\prod_{t_1}^{s} e^{A(r)dr} - e^{(s-t_1)A(t)}\right\|_2 \tag{6.120c}$$

$$\|\tilde{\Delta}_{o,2}(s)\|_2 = \left\|\prod_{t_1}^{s} e^{A(r)dr} + e^{(s-t_1)A(t)}\right\|_2 \leq \left\|\prod_{t_1}^{s} e^{A(r)dr}\right\|_2 + \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} e^{(s-t_1)\eta_P(A(t))} \tag{6.120d}$$

The proof is concluded by applying Lemma 6.3.2 to (6.120b) and (6.120d), and Lemma 6.3.3 to (6.120a) and (6.120c). □

Thanks to the previous upper-bounds on the Gramian estimation error, sufficient conditions of near future controllability and observability of the LPV system (6.86) can be exhibited.

Theorem 6.4.4 (near future controllability and observability)**.** *Despite not having access to the value of A in the system (6.110) during the interval $[t_1, t_2]$, the controllability (resp. observability) Gramian defined in (6.109) (Theorem 6.4.2) is guaranteed to be positive definite if the zero-order hold controllability (resp. observability) Gramian defined in (6.111) satisfies:*

$$\hat{W}_c(t_2, t_1) - m_c I_n \succ 0 \tag{6.121a}$$

$$\text{resp. } \hat{W}_o(t_2, t_1) - m_o I_n \succ 0 \tag{6.121b}$$

where m_c and m_o are upper-bounds verifying respectively:

$$\|\Delta_{W_c}(t_2, t_1)\|_2 \leq m_c \tag{6.122a}$$

$$\|\Delta_{W_o}(t_2, t_1)\|_2 \leq m_o \tag{6.122b}$$

If (6.121) is verified, the system (6.110) is said to be near future controllable (resp. observable) on $[t_1, t_2]$.

Proof. The proof only focuses on the controllability Gramian, the reasoning being similar for the observability Gramian. Thanks to (6.112), $W_c(t_2, t_1) \succ 0$ holds if and only if:

$$\hat{W}_c(t_2, t_1) + \Delta_{W_c}(t_2, t_1) \succ 0 \tag{6.123}$$

Since $\Delta_{W_c}(t_2, t_1) \in \mathbb{S}_n(\mathbb{R})$, $\|\Delta_{W_c}(t_2, t_1)\|_2 = \max\{|\lambda_{\min}(\Delta_{W_c}(t_2, t_1))|, |\lambda_{\max}(\Delta_{W_c}(t_2, t_1))|\}$. Combined with Property 2.2.1, this provides:

$$\hat{W}_c(t_2, t_1) - \|\Delta_{W_c}(t_2, t_1)\|_2 I_n \preceq \hat{W}_c(t_2, t_1) + \Delta_{W_c}(t_2, t_1) \tag{6.124}$$

hence:

$$\hat{W}_c(t_2, t_1) - m_c I_n \preceq W_c(t_2, t_1) \tag{6.125}$$

So if $\hat{W}_c(t_2, t_1) - m_c I_n$ is positive definite, then $W_c(t_2, t_1)$ is positive definite as well. □

Depending on the context, some inequalities of Theorem 6.4.3 can be more advantageous than others to use in Theorem 6.4.4. The previous inequalities are now illustrated through the study of two examples.

Example 6.4.3. *Consider once again the one-dimensional LPV system of Example 6.4.1:*

$$\dot{x}(t) = \theta(t)x(t) + u(t) \tag{6.126}$$

where $\theta : \mathbb{R} \to \mathbb{R}$ is assumed to be a L_A-Lipschitz scheduling parameter only known in real-time, such that for all $t \in \mathbb{R}$, $\mu(\theta(t)) \leq \sigma$, with $L_A = 1$, $\sigma = 1/2$. Of course, no matter the behavior of θ, it is easily seen that for all $t_1 \in \mathbb{R}_{\geq 0}$, there exists $t_2 \in [t_1, +\infty)$ such that the system is controllable on $[t_1, t_2]$, provided that the behavior of θ is known during this interval. The point of this example is to show that the condition of Theorem 6.4.4, despite being conservative, is sufficient to get to this conclusion.

The zero-order hold assumption is used to estimate the controllability Gramian between t_1 and t_2 from $t = 0$ and $\theta(0) = 0$:

$$\hat{W}_c(t_2, t_1) = \int_{t_1}^{t_2} 1 ds = t_2 - t_1 \tag{6.127}$$

The inequality (6.114a) and the first inequality of (6.115a) in Theorem 6.4.3 are considered with $t = 0$ and $P = 1$ to upper-bound the estimation error $\Delta_{W_c}(t_2, t_1)$.

$$\begin{aligned} \|\Delta_{W_c}(t_2, t_1)\|_2 &\leq \int_{t_1}^{t_2} (t_2 - s)(t_2 + s) e^{t_2 - s} ds \\ &= 2(t_2 + 1) + \left(t_2^2 - t_1^2 - 2(t_1 + 1)\right) e^{t_2 - t_1} \end{aligned} \tag{6.128}$$

Theorem 6.4.4 states that the system is near future controllable on $[t_1, t_2]$ if $\alpha(t_1, t_2) > 0$ where $\alpha(t_1, t_2)$ is defined by:

$$\begin{aligned} \alpha(t_1, t_2) &\triangleq \hat{W}_c(t_2, t_1) - \|\Delta_{W_c}(t_2, t_1)\|_2 \\ &= \left(2(t_1 + 1) + t_1^2 - t_2^2\right) e^{t_2 - t_1} - t_2 - t_1 - 2 \end{aligned} \tag{6.129}$$

It is easy to check that for all $t_1 \in \mathbb{R}_{>0}$, $\alpha(t_1, t_1) = 0$. Moreover, taking the derivative of α with respect to t_2 and evaluating this derivative for $t_2 = t_1$ provides:

$$\frac{\partial \alpha}{\partial t_2}(t_1, t_1) = 1 \tag{6.130}$$

Since α is a smooth function, the equation above guarantees that for all $t_1 \in \mathbb{R}_{\geq 0}$, there exists $\varepsilon \in \mathbb{R}_{>0}$ such that $\alpha(t_1, t_1 + \varepsilon) > 0$, which was the intended conclusion.

Example 6.4.4. *Consider the following LPV system:*

$$\dot{x}(t) = A(\theta(t))x(t) \tag{6.131a}$$

$$y(t) = Cx(t) \tag{6.131b}$$

$$\text{with } A(\theta(t)) = \begin{pmatrix} \theta_{11}(t) & \theta_{12}(t) \\ \theta_{21}(t) & \theta_{22}(t) \end{pmatrix} \text{ and } C = \begin{pmatrix} 0 & 1 \end{pmatrix} \tag{6.131c}$$

where $\theta : \mathbb{R} \to \mathbb{R}^4$ is assumed to be a Lipschitz scheduling parameter, only known in real-time. It is easily deduced that $A(\theta(\cdot))$ is also a Lipschitz function, and its Lipschitz constant is denoted L_A. It is assumed that:

$$A(\theta(0)) = \begin{pmatrix} -5 & 0 \\ a & -2 \end{pmatrix} \tag{6.132}$$

The inequality (6.114d) and the third inequality of (6.115b) in Theorem 6.4.3 are considered with $t = t_1 = 0$ and $P = I_2$ to upper-bound the spectral norm of the estimation error $\Delta_{W_c}(t_2, 0)$ between $W_c(t_2, 0)$ and $\hat{W}_c(t_2, 0)$ using a quantity $m_o(t_2, 0)$. Now, let t_2^* denote the time above which the observability of (6.131) on the interval $[0, t_2^*]$ can no longer be guaranteed. Formally:

$$t_2^* = \sup \left\{ t_2 \in \mathbb{R}_{\geq 0} : \hat{W}_c(t_2, 0) - m_o(t_2, 0) I_2 \succ 0 \right\} \tag{6.133}$$

The values of t_2^* are plotted as a function of $a \in [-5, 0]$ and $L_A \in [0.15, 0.4]$ in Figure 6.3. For example, for $a = -1$ and $L_A = 0.15$, the system (6.131) is guaranteed to be observable on $[0, t_2]$ for all $t_2 < t_2^* \approx 24.5$. Without any surprise, t_2^* decreases for larger values of the Lipschitz constant L_A. However, the correlation between t_2^* and a is less straightforward. For $a = 0$, the pair $(A(\theta(0)), C)$ is unobservable, which explains the smaller values of t_2^* obtained for $a \approx 0$. Moreover, for larger values of $-a$, the logarithmic norm $\eta(A(\theta(0)))$ takes positive values rather negative ones, which decreases the quality of the bound $m_o(t_2, 0)$. This trade-off might explain why the largest values of t_2^* are obtained for $a \approx -1$.

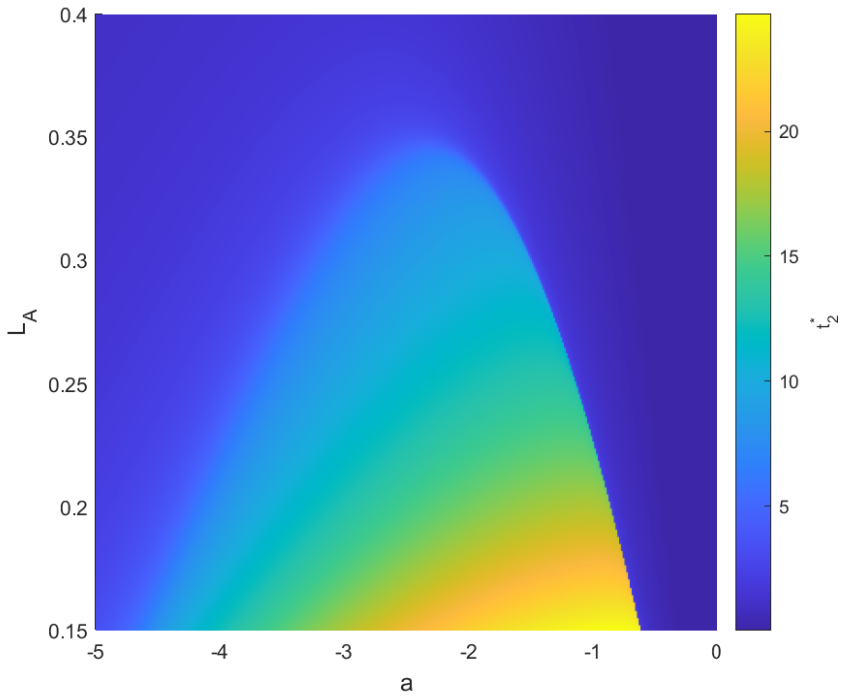


Figure 6.3: The values of t_2^* in Example 6.4.4 as a function of $a \in [-5, 0]$ and $L_A \in [0.15, 0.4]$.

6.5 Conclusions and perspectives

In this chapter, uncertain matrices are introduced in order to bound the future trajectory of a continuous-time LPV system with its parameter exclusively known in real-time. These matrices of uncertainties can be upper-bounded by quantifying the greatest possible discrepancy between the real future of a continuous-time LPV system with a Lipschitz assumption on the scheduling vector θ, and an artificially constructed prediction of the future for which the scheduling vector θ and the input are being held constant. Volterra's product integration and the weighted logarithmic norm of a matrix are introduced by this chapter to achieve this goal. The obtained bounds are leveraged to exactly discretize an LPV system, as well as to demonstrate controllability and observability of such a system in a near future. These near future structural guarantees, namely near future observability and near future controllability, remain to be investigated in the context of practical gain-scheduled observers and controllers. Other applications of these bounds also remain to be investigated. In particular, the extension to a continuous-time setting of the discrete-time model-predictive control techniques for LPV systems with a bounded rate of variation [267, 34, 168, 144] may also rely on these results.

Chapter 7

Fault-isolation using a set-membership approach

This chapter presents an abstract set-membership approach to fault detection and isolation for uncertain linear systems, using the Minkowski functional to derive theoretical results that apply broadly to set-based methodologies, enabling robust detection despite system disturbances.

7.1 Set-membership approach to fault detection and isolation

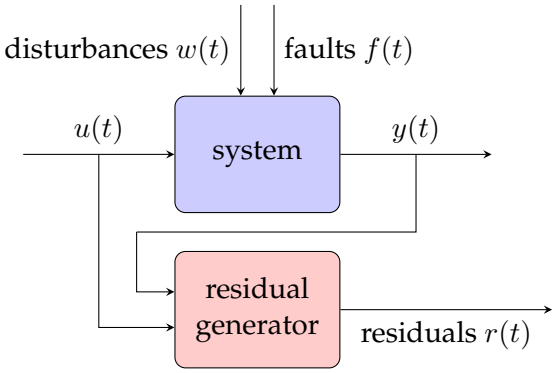


Figure 7.1: Residuals generation architecture.

Given a monitored dynamical system for which a mathematical model is known, fault diagnosis consists in identifying the faults occurring in the system, its sensors, or its actuators, by analyzing - for example - the discrepancies between the measured inputs and outputs of the system and their expected or estimated behaviour. This task has been approached both by the control theory community using Fault Detection and Isolation (FDI) techniques, and the diagnosis (DX) community, by relying on artificial intelligence techniques. Although this work focuses on techniques from the FDI community, it is worth mentioning the pioneer works of the DX community [229, 218, 74], as well as more recent works bridging the gap between the two communities [100, 51, 282]. In order to achieve fault diagnosis, residual signals are generally generated by numerically checking the consistency between the inputs and the outputs of the system onto a single vector, measurable in real-time, nominally centered at the origin, and which diverges from it in case of discrepancies (Figure 7.1) [98, 213, 101, 63, 79]. From

these residuals, the fault diagnosis process is generally handled in three steps: fault detection (detect if at least one fault is active), fault isolation (pinpoint which set of faults is active) and fault identification (evaluate the magnitude of each occurring fault). One obvious way of performing fault diagnosis consists in leveraging physical redundancy in the sensors and actuators monitoring the system (e.g. using triplex sensors), but these solutions usually require costly extra equipment, hence analytical redundancy methods are often preferred. After some pioneer works in aerospace engineering [75] and chemical engineering [185, 201], several analytical residual generating methods have been developed in the literature for both linear and nonlinear systems. The most well-known schemes include the parity space approach [68, 213, 102] and the use of diagnostic observers [248, 33, 97, 213]. Of course, the residual signals obtained through these methods are affected by the system's parametric uncertainties (such as modeling errors) and exogenous disturbances (such as measurement noise), and one of the challenges in performing fault diagnosis is to distinguish the unavoidable noise in the residual signals from the signature of an actual fault affecting the system, while taking into account its parametric uncertainties.

Usually, the presence of noise and parametric uncertainties are handled in two ways [98, 101, 79]:

- firstly, by minimizing or even canceling their influence on the residual generation through some robustness techniques (the *active* approach);
- secondly, by processing the residual signals with consideration for the statistical influence of the noise and of the uncertainties (the *passive* approach).

However, identifying a probabilistic model of the noise and using it with consideration for the parametric uncertainties of the model is generally a difficult task which limits the applicability of the statistical methods of the passive approach. This obstacle has led to the development of so-called *set-membership* (or *bounding*) approaches, where the passive step is achieved through the construction of an adaptive threshold on the residuals which only relies on the knowledge of the noise and uncertainty bounds [190, 71].

The difficulty common to all set-membership approaches - which also includes reachability analysis, robust Model Predictive Control (MPC) and state estimation - concerns the computation of the bounding sets, which are generally hard to compute exactly, especially in real-time. In the literature, geometric results are generally leveraged to characterize their inner- and outer-approximations. The usual method consists in approximating all the sets of the problem by convex sets of a convenient class: orthotopes in the case of interval analysis [2], including interval observer techniques [188, 226, 82]; ellipsoids [159, 230]; parallelotopes [224]; zonotopes [217, 42, 223, 43]; polytopes [41], sometimes considered as constrained zonotopes [250, 225]; or more generic convex shapes, represented using hybrid zonotopes [39] or constrained convex generators [258, 259]. Dedicated algorithms are then used to compute the exact representation of the sets of interest, or at least some inner- or outer-approximations when the exact sets cannot be easily computed. The classes of convex sets and their representations are chosen depending on the system and the task at hand, as a trade-off between their geometric accuracy and the efficiency of the algorithms associated with their representation. These set-memberships techniques have sometimes been merged with results from the probabilistic paradigm mentioned earlier, to enhance their efficiency [204, 28, 70].

Although less common in the literature, there already exist generalized set-membership approaches, which usually rely on the support functions associated with the convex sets of interest [163, 227]. Under these generalized approaches, very little structural properties are assumed on the sets of interests, leading to interesting unified results applicable to all other techniques. This present work falls under the category of these more generalized frameworks.

This chapter is an attempt at providing a unified set-theoretic thresholding perspective to FDI, and at establishing results that can be applied to a large class of set-membership approaches, with a possibly higher degree of accuracy than, for example, by only using ellipsoids. The idea is to establish a flexible framework to the manipulation of convex (or even star-convex) sets, allowing for easier combinations of the already existing set-membership approaches to FDI, while unifying them to a new degree of generality. This chapter does not aim at competing with these already existing methodologies, but rather complements them by offering an alternative and unified analytical point of view to represent the sets of interest, similarly to support functions. All algorithm dedicated to special classes of convex sets remain applicable under the proposed methodology, and can be enhanced by leveraging the generic results found in this chapter.

To this end, this chapter introduces a tool from order theory: a (upper semi-)lattice structure; as well as a tool from convex analysis: the Minkowski functional (or Minkowski function, or gauge function). The Minkowski functional has already gained popularity for the control of dynamical systems due to its practicality, in particular to find so-called Minkowski-Lyapunov functions demonstrating the stability of ordinary differential equations [118, 40, 227]. The author also believes that these functions are convenient in the context of set-membership approaches, and in particular for FDI, since the Minkowski functional of a set often provides a way to describe this set using a single inequality, which can very easily serve as a threshold that can be scaled up or down. More generally, the Minkowski functional of a set $\mathcal{S}$ possesses algebraic properties linked with the geometric properties of $\mathcal{S}$ which are interesting to know in the context of a set-membership approach. However, as far as the author is aware, its use has not been adopted in a FDI context yet, and could be very well-suited to complement already existing set-membership approaches, in particular the generic approaches relying on support functions.

The main contributions of the chapter are listed hereafter:

- The (upper semi-)lattice is introduced as a tool from order theory to isolate the fault occurrences of a system.
- The Minkowski functional of a set is introduced as a thresholding tool, which can represent a set implicitly through a simple inequality. The evaluation of the Minkowksi functional gives an intuitive measurement of the extend to which a set-membership relation is verified. This evaluation is introduced later in this document as the Minkowski signal associated with the set-membership relation.
- The usual operations of union, intersection, Minkowski sum and linear transformation of convex sets are investigated under the lens of the Minkowksi functional.
- The expression of the Minkowski functional is leveraged to obtain analytical criteria on the minimal fault magnitude guaranteeing its isolation.

Although the union and intersection operation are easily dealt with, the Minkowksi sum

and the linear transformation of sets represented by their Minkowski functional remain challenging to obtain without relying on pre-existing algorithms dedicated to special classes of convex sets. Some inner- and outer-approximations are nevertheless obtained to approximate the Minkowski sum of convex sets, and a tight result is obtained for ellipsoids. Similarly, two methodologies to analytically characterize the linear transformation of a convex set are derived. The opposite properties are found for support functions associated with convex sets, where the union and intersection set operations are difficult to handle, contrary to the Minkowski sum and the linear transformation which are easily dealt with. These properties probably explain the prevalence of the support functions in the set-membership literature. The support function of a set is however not well-adapted for thresholding purposes. Both tools are complementary in this regard. As a matter of fact, the Minkowski functional of a compact and convex set is the support function associated with its polar set [247, 227].

The chapter is organized as follows. Section 7.2 presents a highly generic set-membership approach to FDI for residuals following an uncertain linear internal structure (a notion introduced in Section 7.2.1). Section 7.3 introduces the Minkowski functional as a practical tool to compute inner- and outer-approximations of the sets manipulated in the set-membership approach described earlier, and to generate *Minkowski signals* from the residuals, which are easily thresholded. In particular, Section 7.3.3 provides two analytical characterizations of the linear transformation of smooth convex sets. Section 7.4 illustrates the described methodology with an academic example. Finally, some conclusions and perspectives are discussed in Section 7.5.

7.2 Set-membership diagnosis

After introducing in Section 7.2.1 the internal structure of the residual that will be studied throughout this chapter (Definition 7.2.1), it is recalled in Section 7.2.2 how the set-membership approach to diagnosis deals with the question of fault detection using the Direct Image Test (Theorem 7.2.1), and then a generic extension of this approach to fault isolation is suggested in Section 7.2.3 by leveraging some order theory results (Theorem 7.2.2). For now, these results are stated using an all-encompassing set-theoretic framework.

7.2.1 Residuals internal structure

The residuals synthesized for system diagnosis can be written under two forms [101]:

- their computational form, giving their relationship to the measured signals used in their calculation;
- their internal form, giving their relationship to the noises and faults affecting the system.

It is assumed that the computational form of the residuals provides a vector of signals $r(\theta, t)$, with θ denoting the time varying parametric uncertainties of the system, and where $r(\theta, t) = 0$ is verified in a noise- and fault-free context.

This chapter focuses on the set-membership approach to diagnosis schemes where the residuals follow an *uncertain linear internal structure*. This structure is introduced in Definition 7.2.1 below.

Definition 7.2.1 (Uncertain linear internal structure)**.** *The residual signal $r(\theta,t) \in \mathbb{R}^{n_r}$ is said to follow an uncertain linear internal structure if its internal form can be written as:*

$$r(\theta,t) = \sum_{i=1}^{n_w} W_i(\theta,t) w_i(t) + \sum_{i=1}^{n_f} F_i(\theta,t) f_i(t) \tag{7.1}$$

with $w_i(t) \in \mathcal{W}_i \subset \mathbb{R}^{n_{w_i}}$ the bounded noises affecting the residual signals, $f_i(t) \in \mathcal{F}_i \subset \mathbb{R}^{n_{f_i}}$ the potential faults to be detected, $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$ the time-varying uncertainties of the system, and $W_i \in \mathbb{R}^{n_r \times n_{w_i}}$, $F_i \in \mathbb{R}^{n_r \times n_{f_i}}$ uncertain time-varying matrices. Let $\Omega(t) \subseteq \Theta$ denote the bounding set for the uncertainties θ of the system at time t. The union over all the uncertainties $\theta \in \Omega(t)$ of the residual signals is referred to as the residual set and is denoted $R(t)$:

$$R(t) \triangleq \bigcup_{\theta \in \Omega(t)} \{r(\theta,t)\} \tag{7.2}$$

This structure is common among residuals obtained for Linear Time Invariant, Time Varying, and Parameter Varying (LTI, LTV, LPV) discrete-time systems. In particular, residual signals following this internal structure can be obtained with parity space approaches or some simple diagnostic observer designs [101, 213, 217, 43]. Note that this structure can also be leveraged for residuals with a more complex internal structure by means of linearization, or by considering the residuals' internal nonlinearities as new parametric uncertainties.

Remark 7.2.1. *Despite (7.1) being an instantaneous expression, delays on the disturbances, on the faults and on the time-varying components of the matrices are not difficult to handle by considering $\tilde{w}_i(t) = w_i(t-\tau_i)$, $\tilde{f}_i(t) = f_i(t-\tau_i)$, $\tilde{W}_i(\theta,t) = W_i(\theta,t-\tau_i)$ and $\tilde{F}_i(\theta,t) = F_i(\theta,t-\tau_i)$, with $\tau_i \geq 0$. However, if the parametric uncertainties $\theta \in \Omega(t)$ are also time-varying and subject to delays, the residuals analysis can become more challenging, and some assumptions on the variation rate of θ may become extremely useful to approximate the bounding set $\Omega(t)$.*

Remark 7.2.2. *The choice of writing the residuals in the form (7.1) might be surprising, as they can be simply rewritten as the uncertain linear transformation of a single fault and noise vector. The sums are introduced in order to distinguish between several types of additive faults and noises (actuator, sensor, or parametric faults and noises), at potentially different instant of the past. Moreover, the faults and noises are assumed to be vectors instead of scalars in order to limit the number of terms in the sums, hence reducing the complexity and conservatism of the FDI methods described in this chapter. The model described above contains in fact m_w scalar noises and m_f scalar faults, with:*

$$m_w \triangleq \sum_{i=1}^{n_w} n_{w_i}, \qquad m_f \triangleq \sum_{i=1}^{n_f} n_{f_i} \tag{7.3}$$

Example 7.2.1 (Parity Relations [217])**.** *Considering the following uncertain discrete-time linear system:*

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t + G_1(\theta)f_t + V_1(\theta)w_t \tag{7.4a}$$
$$y_t = C(\theta)x_t + D(\theta)u_t + G_2(\theta)f_t + V_2(\theta)w_t \tag{7.4b}$$

where $x_t \in \mathbb{R}^{n_x}$ is the state of the system, $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$ is a vector of parametric uncertainties, $u_t \in \mathbb{R}^{n_u}$ is the known input vector, $y_t \in \mathbb{R}^{n_y}$ is the measured output vector, $f_t \in \mathcal{F} \subset \mathbb{R}^{n_x}$ are potential faults to detect, and $w_t \in \mathcal{W} \subset \mathbb{R}^{n_x}$ are the bounded noises affecting the system. In the following parity space

approach, θ is assumed to be a constant and unknown uncertainty at a horizon of h time steps. It is moreover assumed that an estimation of θ, denoted $\hat{\theta}$, is known. The residual $r([\theta,\hat{\theta}],t)$ is a vector signal with the following computational and internal form resp.:

$$r([\theta,\hat{\theta}],t) \triangleq N(\hat{\theta})\left(Y_t - \Gamma_u(\hat{\theta})U_t\right) \tag{7.5a}$$

$$r([\theta,\hat{\theta}],t) = N(\hat{\theta})\left(\mathcal{O}(\theta)x_t + \Gamma_w(\theta)W_t + \Gamma_f(\theta)F_t + \left(\Gamma_u(\theta) - \Gamma_u(\hat{\theta})\right)U_t\right) \tag{7.5b}$$

with for all $(Z,z) \in \{(U,u),(Y,y),(W,w),(F,f)\}$, $Z_t = \begin{pmatrix} z_{t-h}^\top & \dots & z_t^\top \end{pmatrix}^\top$, and where the matrix N is synthesized such that $N(\hat{\theta})\mathcal{O}(\theta)x_t \approx 0$, with

$$\Gamma_u = \begin{bmatrix} D & & & 0 \\ CB & \ddots & & \\ \vdots & \ddots & & \\ CA^{h-1}B & \dots & CB & D \end{bmatrix}, \quad \Gamma_w = \begin{bmatrix} V_2 & & & 0 \\ CV_1 & \ddots & & \\ \vdots & \ddots & & \\ CA^{h-1}V_1 & \dots & CV_1 & V_2 \end{bmatrix},$$
$$\Gamma_f = \begin{bmatrix} G_2 & & & 0 \\ CG_1 & \ddots & & \\ \vdots & \ddots & & \\ CA^{h-1}G_1 & \dots & CG_1 & G_2 \end{bmatrix}, \quad \mathcal{O} = \begin{bmatrix} C^\top & [CA]^\top & \dots & [CA^h]^\top \end{bmatrix}^\top \tag{7.6}$$

where the dependence on θ or $\hat{\theta}$ has been omitted for concision. The residual r follows the uncertain linear internal structure introduced in Definition 7.2.1 if $N(\hat{\theta})\mathcal{O}(\theta)x_t$ and $(\Gamma_u(\theta) - \Gamma_u(\hat{\theta}))U_t$ are assumed to be bounded noise signals.

Example 7.2.2 (Luenberger observer [97])**.** *The system (7.4) is considered once again, but θ is now assumed to be an uncertainty measured in real-time. The Luenberger observer (7.7) is designed with a matrix $L \in \mathbb{R}^{n_x \times n_y}$ chosen such that the system generating the state estimation error $e_{t+1} = (A(\theta_t) + LC(\theta_t))e_t$ is globally asymptotically stable.*

$$\hat{x}_{t+1} = A(\theta_t)\hat{x}_t + B(\theta_t)u_t + L(y_t - C(\theta_t)\hat{x}_t - D(\theta_t)u_t) \tag{7.7}$$

A residual $r([\theta_t,\theta_{t-1}],t)$ can be obtained with the following computational and internal form resp.:

$$r([\theta_t,\theta_{t-1}],t) \triangleq y_t - C(\theta_t)\hat{x}_t - D(\theta_t)u_t \tag{7.8a}$$

$$\begin{aligned} r([\theta_t,\theta_{t-1}],t) =& C(\theta_t)[A(\theta_{t-1})(x_{t-1} - \hat{x}_{t-1}) - Lr([\theta_{t-1},\theta_{t-2}],t-1) + G_1(\theta_{t-1})f_{t-1} + V_1(\theta_{t-1})w_{t-1}] \\ & \cdots + G_2(\theta_t)f_t + V_2(\theta_t)w_t \end{aligned} \tag{7.8b}$$

The residual r follows the uncertain linear internal structure introduced in Definition 7.2.1 if the state estimation error $(x_{t-1} - \hat{x}_{t-1})$ and the residual at the previous time-step $r([\theta_{t-1},\theta_{t-2}],t-1)$ are assumed to be noise signals, with bounds that can be evaluated recursively, e.g. using set-membership state estimation [248, 33].

Remark 7.2.3. *From now on, the time-dependence in $r(\theta,t)$, $R(t)$, $W(\theta,t)$, $w_i(t)$, $F(\theta,t)$, and $f_i(t)$ is omitted, resp. $r(\theta)$, R, $W(\theta)$, w_i, $F(\theta)$ and f_i.*

7.2.2 Fault detection

As discussed in the introduction, fault detection consists in performing a consistency test to a monitored dynamical system in order to detect if the system is subject to at least one occurring fault. In a fault-free context, a residual $r(\theta)$ following the structure (7.1) is only affected by noises, and belongs to $\mathcal{R}_\theta$, a finite Minkowski sum of uncertain linear transformations of the sets $(\mathcal{W}_i)_{1\leq i\leq n_w}$. Knowing precisely these bounding sets, the set-membership approach provides $r(\theta) \in \mathcal{R}_\theta$ as an elementary consistency test across all uncertainties $\theta \in \Theta$, where $\mathcal{R}_\theta$ is the set of all the residual values possibly due to the noises, defined by:

$$\mathcal{R}_\theta \triangleq \bigoplus_{i=1}^{n_w} W_i(\theta)\mathcal{W}_i, \text{ and } \mathcal{R}_\Theta \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_\theta \tag{7.9}$$

The boundary of $\mathcal{R}_\theta$ can be considered as an *exact* or *clear-cut* threshold for fault detection, since it provides the sharpest criterion to detect a fault which also avoids any false detection. Checking if the residuals verify $\cup_{\theta\in\Theta}(r(\theta) \setminus \mathcal{R}_\theta) \neq \emptyset$ in order to detect a fault is usually called the Direct Image Test (DIT) [223, 309], and it relies on the following property:

Theorem 7.2.1 (Direct Image Test)**.** *Considering that the following statement holds:*

$$\forall i \in [\![1, n_f]\!], f_i = 0 \Rightarrow \forall\theta \in \Theta, r(\theta) \in \mathcal{R}_\theta \tag{7.10}$$

the contrapositive statement provides:

$$\exists\theta \in \Theta, r(\theta) \notin \mathcal{R}_\theta \Rightarrow \exists i \in [\![1, n_f]\!], f_i \neq 0 \tag{7.11}$$

which is to say: if there exists an uncertainty $\theta \in \Theta$ for which the residual $r(\theta)$ does not belong to the set $\mathcal{R}_\theta$ (or more succinctly, if $\cup_{\theta\in\Theta}(r(\theta) \setminus \mathcal{R}_\theta) \neq \emptyset$), then at least one fault has occurred in the system.

Remark 7.2.4. *Note that no internal model of the faults influence on the residuals is leveraged at this point, hence this fault detection scheme can be applied to any residuals signals for which the influence of the noises on the residuals reduces to the internal structure (7.1) in a fault-free context.*

Remark 7.2.5 (Robust Direct Image Test)**.** *This test is not equivalent to checking if $(\cup_{\theta\in\Theta} r(\theta)) \setminus (\cup_{\theta\in\Theta}\mathcal{R}_\theta)) \triangleq R \setminus \mathcal{R}_\Theta \neq \emptyset$ (with R defined in (7.2)). This latter test, although not as sharp as (7.11), is still interesting as it does not require to know which uncertainty θ is responsible for each residual in R, which is often the case in fault diagnosis schemes. This chapter refers to this latter test as the Robust Direct Image Test (RDIT).*

The DIT still works using inner-approximations of Θ or outer-approximations of $\mathcal{R}_\Theta$. In particular, the literature sometimes evaluates $\hat{\Omega}(t)$, an estimated set of values for θ at time t (such that $\Omega(t) \subseteq \hat{\Omega}(t) \subseteq \Theta$ in (7.2)) using some additional assumptions on θ (e.g. known dynamic, bounded rate of variation, partial estimation, etc). The DIT described above is then performed with this set estimate in order to detect a fault. This approach is sometimes called the *inverse image test*, where the term *inverse* refers to the dynamical construction of $\hat{\Omega}(t)$, which can involve the computation of the pre-image set of some output signals [223, 309]. This chapter will mainly focus on DIT, but its extension to inverse image tests is possible as long as residuals of the form (7.1) can be constructed. In that case, instead of Θ, the set $\hat{\Omega}(t)$ should be considered.

7.2.3 Fault isolation

Following the ideas previously introduced, checking what are the occurring faults in the system can again be achieved by means of consistency tests. Moving forward, $\mathbf{P}_{n_f}$ stands for the power set of $[\![1, n_f]\!]$, i.e. the set of all the subsets of $[\![1, n_f]\!]$, and for all $\mathcal{I} \subseteq [\![1, n_f]\!]$ (i.e. $\mathcal{I} \in \mathbf{P}_{n_f}$), $\overline{\mathcal{I}} \triangleq [\![1, n_f]\!] \setminus \mathcal{I}$. Assuming $\mathcal{I} \in \mathbf{P}_{n_f}$ a list of potentially active faults indices, the obtainable residuals are given by the set $\mathcal{R}_{\theta,\mathcal{I}}$ defined hereafter.

Definition 7.2.2 (Feasible set)**.** *The feasible set $\mathcal{R}_{\theta,\mathcal{I}}$ is the set of residuals obtainable with $\mathcal{I} \in \mathbf{P}_{n_f}$ a list of potentially active faults indices. The residuals structure (7.1) provides:*

$$\mathcal{R}_{\theta,\mathcal{I}} \triangleq \mathcal{R}_\theta \oplus \left(\bigoplus_{i \in \mathcal{I}} F_i(\theta) \mathcal{F}_i \right), \text{ and } \mathcal{R}_{\Theta,\mathcal{I}} \triangleq \bigcup_{\theta \in \Theta} \mathcal{R}_{\theta,\mathcal{I}} \tag{7.12}$$

where it is considered that $\mathcal{R}_{\theta,\emptyset} = \mathcal{R}_\theta$ in a fault-free context. The collections of all the feasible sets are denoted FS_θ and FS_Θ, with:

$$FS_\theta \triangleq \{\mathcal{R}_{\theta,\mathcal{I}} : \mathcal{I} \in \mathbf{P}_{n_f}\}, \qquad FS_\Theta \triangleq \{\mathcal{R}_{\Theta,\mathcal{I}} : \mathcal{I} \in \mathbf{P}_{n_f}\} \tag{7.13}$$

It is easily verified that there are $\#\mathrm{FS}_\theta = 2^{n_f}$ feasible sets, hence their calculation can rapidly become computationally heavy, especially over all uncertainties θ. Moreover, in a faulty situation, the fault isolation process has to discriminate between all the feasible sets containing $r(\theta)$ in order to estimate the list of possible active faults. This task is difficult, and to the author's knowledge, the literature has only tackled this problem for specific fault isolation schemes. In particular, fault isolation has motivated the design of Dedicated and Generalized Observer Schemes (the DOS and GOS architectures) [97] and the use of structured residuals and fault matrices [102]. Leveraging some tools from order theory, this section provides a straightforward and all-encompassing procedure to fault isolation, which generalizes the DIT discussed previously.

As a first observation, one can notice that the inclusion hierarchy of the feasible sets (7.13) contains *at least* the same lattice structure as the power set $\mathbf{P}_{n_f}$ taken with the inclusion $\subseteq$ as a lattice ordering (Figures 7.2 and 7.3). This is formalized in the following property.

Property 7.2.1 (Order structure of the feasible sets)**.** *Let FS_θ be considered with the partial order $\preceq$ defined by $\mathcal{R}_{\theta,\mathcal{I}_1} \preceq \mathcal{R}_{\theta,\mathcal{I}_2}$ if and only if $\mathcal{I}_1 \subseteq \mathcal{I}_2$. $(FS_\theta, \preceq)$ and $(\mathbf{P}_{n_f}, \subseteq)$ have an isomorphic lattice structure. Moreover if $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$ (i.e., each fault can be null), then:*

$$\forall \mathcal{I}_1, \mathcal{I}_2 \in \mathbf{P}_{n_f},\ \mathcal{R}_{\theta,\mathcal{I}_1} \preceq \mathcal{R}_{\theta,\mathcal{I}_2} \Rightarrow \mathcal{R}_{\theta,\mathcal{I}_1} \subseteq \mathcal{R}_{\theta,\mathcal{I}_2} \tag{7.14}$$

meaning $\preceq$ is only a restriction of the partial order $\subseteq$ on FS_θ.

Proof. The map $\varphi : \mathbf{P}_{n_f} \to \mathrm{FS}_\theta$ defined such that for all $\mathcal{I} \in \mathbf{P}_{n_f}$, $\varphi(\mathcal{I}) = \mathcal{R}_{\theta,\mathcal{I}}$, is by definition an order isomorphism, and order isomorphisms preserve lattice structures, hence, since $(\mathbf{P}_{n_f}, \subseteq)$ is a lattice, then $(\mathrm{FS}_\theta, \preceq)$ has an isomorphic lattice structure. Moreover if $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$, from equation (7.2) and by definition of the Minkowski sum, (7.14) holds, and φ is an order homomorphism between $(\mathbf{P}_{n_f}, \subseteq)$ and $(\mathrm{FS}_\theta, \subseteq)$. □

Remark 7.2.6. *This property is easily extended to FS_Θ (Figure 7.2).*

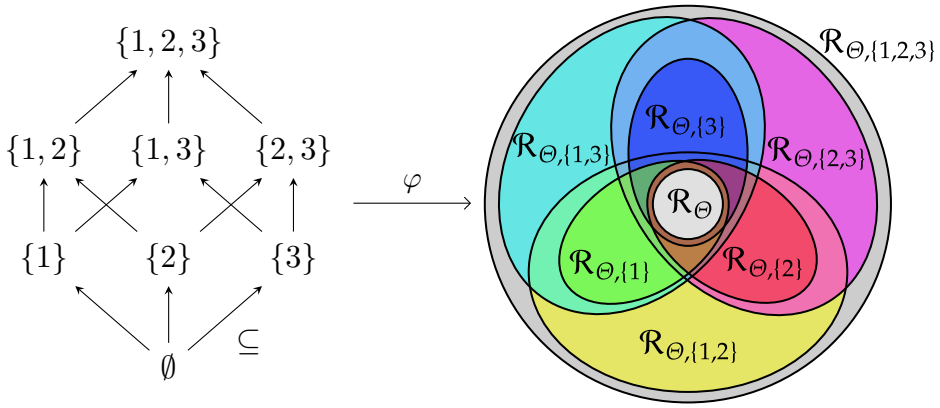


Figure 7.2: Hasse diagram of the lattice $(\mathbf{P}_3, \subseteq)$ in relation to the inclusion hierarchy of the feasible sets FS_Θ.

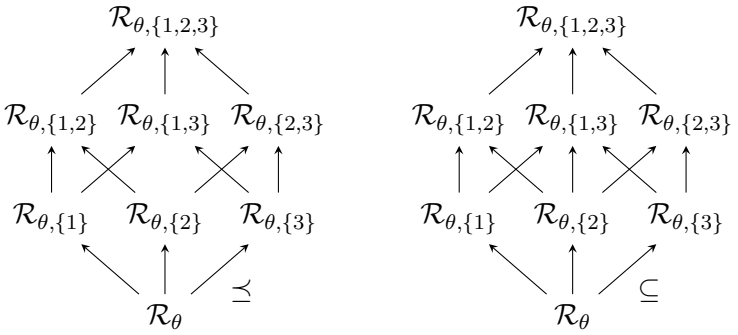


Figure 7.3: Hasse diagram of the lattice $(\mathrm{FS}_\theta, \preceq)$ (on the left) and of the partially ordered set $(\mathrm{FS}_\theta, \subseteq)$ (on the right) with $n_f = 3$, where the non-trivial inclusion $\mathcal{R}_{\theta,\{2\}} \subseteq \mathcal{R}_{\theta,\{1,3\}}$ is verified.

Remark 7.2.7. *Notice that the reciprocal to (7.14) is not verified in general, since non-trivial inclusions such as $\mathcal{R}_{\theta,\{2\}} \subseteq \mathcal{R}_{\theta,\{1,3\}}$ can hold true in practice. Hence, even if $(FS_\theta, \subseteq)$ contains at least the lattice structure of $(\mathbf{P}_{n_f}, \subseteq)$ by restriction to $(FS_\theta, \preceq)$, it is not necessarily a lattice itself (Figure 7.3).*

As discussed previously, the DIT described in Theorem 7.2.1 can be generalized to the feasible sets (7.13). In that case, in order to prove that at least one fault of index $i \in \mathcal{I}$ has occurred in the system, one must show that a residual $r(\theta)$ does not belong to the feasible set which has the complement of $\mathcal{I}$ (denoted $\overline{\mathcal{I}}$) as a list of potentially active faults. Succinctly, there must exists a $\theta \in \Theta$ such that $r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}}$ to demonstrate that one of the fault of index $i \in \mathcal{I}$ is active. Contrary to the DIT to fault detection, all sets of potentially active faults $\mathcal{I} \in \mathbf{P}_{n_f}$ have to be considered, and the question of which elements of $\mathbf{P}_{n_f}$ verifying the DIT provide the most meaningful results must be raised. It is shown below that the sets of FS_θ satisfying the DIT follow an upper semi-lattice structure [73], where the strongest statements about the system's faults are found at the bottom of the semi-lattice (Example 7.2.3 illustrated by Figure 7.4).

Theorem 7.2.2 (Generalized Direct Image Test)**.** *If $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$ (i.e., each fault can be null), then, the DIT set defined below is an upper semi-lattice with inclusion taken as partial order.*

$$DIT = \left\{ \mathcal{I} \in \mathbf{P}_{n_f} \ : \ \exists \theta \in \Theta \ s.t. \ r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}} \right\} \tag{7.15}$$

Denoting P_i with $i \in, [\![1, n_f]\!]$ and $Q_{\mathcal{I}}$ with $\mathcal{I} \triangleq \{i_1, \ldots, i_p\} \in \mathbf{P}_{n_f}$ the following statements:

$$P_i \triangleq [f_i \neq 0] : \text{“ The } i\text{-th vector fault is active”} \tag{7.16a}$$

$$Q_{\mathcal{I}} \triangleq P_{i_1} \vee \cdots \vee P_{i_p} : \text{“ At least one of the } i_1\text{-th, } \ldots, i_p\text{-th vector fault is active”} \tag{7.16b}$$

the DIT set demonstrates that the statement $\bigwedge_{\mathcal{I} \in DIT} Q_{\mathcal{I}}$ holds true. Moreover, the following equivalence holds:

$$\left(\bigwedge_{\mathcal{I} \in DIT} Q_{\mathcal{I}} \right) \Leftrightarrow \left(\bigwedge_{\mathcal{I} \in m_e(DIT)} Q_{\mathcal{I}} \right) \tag{7.17}$$

where $m_e(DIT)$ stands for the sets of minimal elements of $(DIT, \subseteq)$. These elements belong to $\mathbf{P}_{n_f}$ and summarize all the information about the faults' activity that can be deduced from DIT.

Proof. This proof is split in three parts. First, $(\text{DIT}, \subseteq)$ is shown to be an upper semi-lattice. Then $\bigwedge_{\mathcal{I} \in \text{DIT}} Q_{\mathcal{I}}$ is shown to hold. Finally, the equivalence (7.17) is proved.

To prove that $(\text{DIT}, \subseteq)$ is an upper semi-lattice, as $\text{DIT} \subseteq \mathbf{P}_{n_f}$, it is sufficient to demonstrate the stability of DIT with respect to inclusion [73]. Let $\mathcal{I}_1 \in \text{DIT}$ and $\mathcal{I}_2 \in \mathbf{P}_{n_f}$ such that $\mathcal{I}_1 \subseteq \mathcal{I}_2$. By definition there exits $\theta \in \Theta$ for which $r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}_1}}$. Since $\overline{\mathcal{I}_2} \subseteq \overline{\mathcal{I}}_1$, by an application of the lattice isomorphism φ, $\mathcal{R}_{\theta,\overline{\mathcal{I}_2}} \preceq \mathcal{R}_{\theta,\overline{\mathcal{I}}_1}$, and finally if $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$, (7.14) provides $\mathcal{R}_{\theta,\overline{\mathcal{I}_2}} \subseteq \mathcal{R}_{\theta,\overline{\mathcal{I}}_1}$, hence $r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}_2}}$, meaning $\mathcal{I}_2 \in \text{DIT}$.

The statement $\bigwedge_{\mathcal{I} \in \text{DIT}} Q_{\mathcal{I}}$ holds by a direct and repeated application of the usual DIT to the sets $\mathcal{R}_{\Theta,\overline{\mathcal{I}}}$ with $\mathcal{I} \in \mathbf{P}_{n_f}$.

Finally the equivalence (7.17) is shown by double implication.

$\boxed{\Rightarrow}$ This implication is an easy consequence of $m_e(\text{DIT}) \subseteq \text{DIT}$.

$\boxed{\Leftarrow}$ On one hand, for all $\mathcal{I}_1, \mathcal{I}_2 \in \text{DIT}$, $(\mathcal{I}_1 \subseteq \mathcal{I}_2) \Rightarrow (Q_{\mathcal{I}_1} \Rightarrow Q_{\mathcal{I}_2})$ is trivial considering that $Q_{\mathcal{I}_j}$ holds if and only if at least one of the $(P_i)_{i \in \mathcal{I}_j}$ holds, with $j \in \{1, 2\}$. On the other hand, considering $\{\mathcal{I}_1 \in \text{DIT} : \mathcal{I}_1 \subseteq \mathcal{I}_2\}$ a subset of DIT, since this set is finite, it must contain at least one minimal element, which will also be a minimal element of DIT by definition. Hence for all $\mathcal{I}_2 \in \text{DIT}$, there exists $\mathcal{I}_1 \in m_e(\text{DIT})$ such that $\mathcal{I}_1 \subseteq \mathcal{I}_2$, which implies $(Q_{\mathcal{I}_1} \Rightarrow Q_{\mathcal{I}_2})$. Repeating the process for all $Q_{\mathcal{I}}$ in $\left(\bigwedge_{\mathcal{I} \in \text{DIT}} Q_{\mathcal{I}}\right)$ demonstrates the implication. □

Corollary 7.2.1 (Generalized Robust Direct Image Test)**.** *The following inclusion is verified (with R defined in (7.2)):*

$$RDIT \triangleq \left\{ \mathcal{I} \in \mathbf{P}_{n_f} \ : \ R \setminus \mathcal{R}_{\Theta,\overline{\mathcal{I}}} \neq \emptyset \right\} \subseteq DIT \tag{7.18}$$

The set RDIT (Robust Direct Image Test) on the left also follows an upper semi-lattice structure with respect to inclusion, where the minimal elements summarize all the information about the faults' activity that can be deduced from the elements in the set.

Remark 7.2.8. *As noted in Remark 7.2.5, the tests performed to obtain the elements of RDIT do not require to know which uncertainty θ is responsible for each residual $r(\theta)$ in R.*

Corollary 7.2.2 (Fault isolation)**.** *If there exists $\mathcal{I} \in DIT$ (or $\mathcal{I} \in RDIT$) such that $\#\mathcal{I} = 1$, i.e. $\mathcal{I} = \{i\}$ with $i \in [\![1, n_f]\!]$, then the i-th fault has been isolated and $f_i \neq 0$ is guaranteed.*

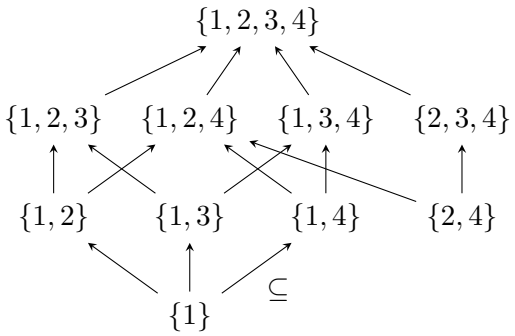


Figure 7.4: Hasse diagram of upper semi-lattice $(\mathrm{DIT}, \subseteq)$ defined in Example 7.2.3.

Remark 7.2.9. *No matter how many faults are active in the system, $r(\theta) \in \mathcal{R}_{\theta,[\![1,n_f]\!]}$ and $R \subseteq \mathcal{R}_{\Theta,[\![1,n_f]\!]}$ are always verified by definition, and the edge cases $\emptyset \in DIT$ and $\emptyset \in RDIT$ do not need to be considered. If these edge cases happen anyway, it means that the model used for the residual is incorrect or incomplete.*

Remark 7.2.10. *As for the DIT and the RDIT, these generalized DIT and RDIT still work using inner-approximations of Θ or outer-approximations of $\mathcal{R}_{\theta,\overline{\mathcal{I}}}$.*

Example 7.2.3. *Consider a system with $n_f = 4$ potentially active faults, and a residual R such that the DIT results in:*

$$DIT = \{\{1\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \{1,2,3,4\}\} \tag{7.19}$$

The upper semi-lattice structure of DIT is illustrated by Figure 7.4. The minimal elements of DIT are given below:

$$m_e(DIT) = \{\{1\}, \{2,4\}\} \tag{7.20}$$

Hence, the generalized DIT provides the truth statement $P_1 \wedge (P_2 \vee P_4)$, i.e. $[f_1 \neq 0] \wedge ([f_2 \neq 0] \vee [f_4 \neq 0])$ which is to say: the first fault of the system is active, and the second or the fourth (or both) faults are active.

7.3 The Minkowski functional for set-membership fault isolation

In practice, the feasible sets (7.13) introduced in Definition 7.2.2 can be difficult to compute, and when their direct calculation is not possible, inner- and outer-approximation of these sets are often used as fault detection thresholds. In this section, the author suggests a unifying approach where the geometric problem of finding inner- and outer-approximations to the feasible sets of the previous section is translated into the problem of upper- and lower-bounding their Minkowski functionals.

Section 7.3.1 begins with the introduction of the general properties of the Minkowski functional, and, in particular, how to use it for the usual set operations (union, intersection, Minkowski sum and left-invertible linear transformations). In Section 7.3.3, two new characterizations of the (non-invertible) linear transformation of smooth convex sets are provided. Finally, these results are leveraged in Section 7.3.4 to obtain a generic solution to the inner- and outer-approximation problem of the feasible sets (7.13), providing upper-bounds on the minimal faults magnitude guaranteeing fault isolation through the generalized image test.

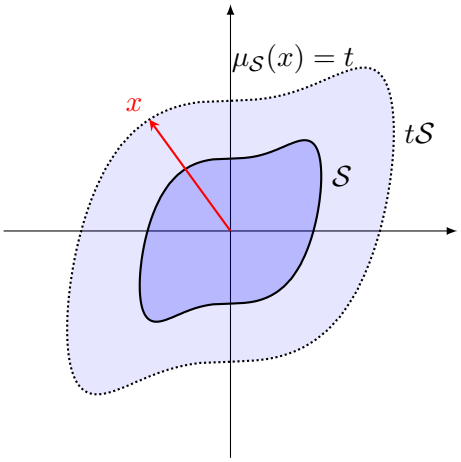


Figure 7.5: Minkowski functional of a set $\mathcal{S}$ of $\mathbb{R}^2$, star-convex at 0.

7.3.1 The Minkowski functional

Intuitively, the Minkowski functional $\mu_{\mathcal{S}}$ associated with the set $\mathcal{S}$ is defined such that for all x, $\mu_{\mathcal{S}}(x)$ provides the smallest scaling t of the set $\mathcal{S}$ with respect to the origin such that the scaled set $t\mathcal{S}$ reaches x (Figure 7.5). This definition is formalized below, and the expression of the Minkowski functional is specified for the usual classes of convex sets.

Definition 7.3.1 (Minkowski functional [247])**.** *Given $\mathcal{S}$ a non-empty set of $\mathbb{R}^n$, the Minkowski functional associated to $\mathcal{S}$ is the map $\mu_{\mathcal{S}} : \mathbb{R}^n \to \tilde{\mathbb{R}}_{\geq 0}$ defined by:*

$$\mu_{\mathcal{S}}(x) \triangleq \inf \left\{ t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{S} \right\} \tag{7.21}$$

Remark 7.3.1. *The set $\mathcal{S}$ is said to be absorbing if $\mu_{\mathcal{S}} < +\infty$.*

Remark 7.3.2. *Alternative functionals can be used for off-centered sets $\mathcal{S} \subset \mathbb{R}^n$, where the scaling is performed with respect to a point $x_0 \in \mathbb{R}^n$ instead of the origin, such that $\mu_{\mathcal{S}}^{x_0} \triangleq \mu_{\mathcal{S} \oplus \{-x_0\}}$.*

Example 7.3.1 (p-Ball [247])**.** *The Minkowski functional of $\mathcal{B}_p \subset \mathbb{R}^n$ the unit ball of norm $p \in \tilde{\mathbb{R}}_{\geq 1}$ centered at the origin is given by:*

$$\mu_{\mathcal{B}_p}(x) = \|x\|_p \tag{7.22}$$

where $\|x\|_p$ stands for the p-norm of x.

Example 7.3.2 (Ellipsoid [40])**.** *The Minkowski functional of an ellipsoid $\mathcal{E} \subset \mathbb{R}^n$ centered at the origin is given by:*

$$\mu_{\mathcal{E}}(x) = \mu_{M\mathcal{B}_2}(x) = \sqrt{x^\top Q x} \tag{7.23}$$

with $M \in \mathbb{R}^{n \times m}$ a full row rank matrix, and $Q \in \mathbb{S}_n^{++}(\mathbb{R})$, the positive definite matrix given by $Q = (MM^\top)^{-1}$.

Example 7.3.3 (Zonotope [6])**.** *The Minkowski functional of a zonotope $\mathcal{Z} \subset \mathbb{R}^n$ centered at the origin is given by:*

$$\mu_{\mathcal{Z}}(x) = \mu_{G\mathcal{B}_\infty}(x) = \|Lx\|_\infty \tag{7.24}$$

with $G \in \mathbb{R}^{n\times m}$ a full row rank matrix, and $L \in \mathbb{R}^{p\times n}$ the matrix whose $p = \binom{m}{n-1}$ rows are given by $l_{\mathcal{I}}/d_{\mathcal{I}}$ for all $\mathcal{I} \subseteq [\![1,m]\!]$ with $\#\mathcal{I} = m-n+1$, and such that

$$l_{\mathcal{I}} = \frac{\times^n(G^{\langle\overline{\mathcal{I}}\rangle})^\top}{\|\times^n(G^{\langle\overline{\mathcal{I}}\rangle})\|_2} \tag{7.25a}$$

$$d_{\mathcal{I}} = \|l_{\mathcal{I}}G\|_1 \tag{7.25b}$$

with $G^{\langle\overline{\mathcal{I}}\rangle}$ the matrix G taken without its columns of index in $\mathcal{I}$, and $\times^n$ the generalized vector product defined by:

$$\forall H \in \mathbb{R}^{n\times(n-1)},\ \times^n(H) \triangleq \begin{pmatrix} \dots & (-1)^{i+1}\det H^{[i]} & \dots \end{pmatrix}^\top \tag{7.26}$$

where $H^{[i]}$ stands for the sub-matrix defined by H without its i-th row.

Remark 7.3.3. *A parallelotope is a zonotope where $m = n$. An orthotope is a zonotope where $G = I_n$.*

Example 7.3.4 (Polytope [94])**.** *Given $\mathcal{P} \subset \mathbb{R}^n$ a compact and convex polytope whose interior contains the origin and whose halfspace-representation is given by:*

$$\mathcal{P} = \{x \in \mathbb{R}^n : \forall k \in [\![1,m]\!],\ \langle h_k|x\rangle \leq 1\} \tag{7.27}$$

with $(h_k)_{1\leq k\leq m}$ as set of m vectors of $\mathbb{R}^n$, then, the Minkowski functional of $\mathcal{P}$ is given by:

$$\mu_{\mathcal{P}}(x) = \max_{k\in[\![1,m]\!]} \langle h_k|x\rangle \tag{7.28}$$

Remark 7.3.4. *As discussed in [250], it is possible to extract the polytopic halfspace-representation of constrained zonotopes, and it is therefore also possible to obtain their Minkowski functional.*

Remark 7.3.5. *Hybrid zonotopes being the union of a finite number of constrained zonotopes, it is also possible to obtain an explicit expression of their Minkowksi functional [39]. It is the minimum of all the Minkowski functionals associated with the constrained zonotopes of this union (see Property 7.3.3 and Remark 7.3.6).*

Example 7.3.5 (Radially parameterized set)**.** *Let $\mathcal{O} \subset \mathbb{R}^n$ be a star-convex set at 0 whose boundary can be parameterized radially by the map $\rho : [0,\pi)^{n-2}\times[0,2\pi) \to \mathbb{R}_{>0}$ such that*

$$\partial\mathcal{O} = \left\{ \rho(\varphi) \begin{pmatrix} \cos(\varphi_1) \\ \sin(\varphi_1)\cos(\varphi_2) \\ \vdots \\ \sin(\varphi_1)\dots\sin(\varphi_{n-2})\cos(\varphi_{n-1}) \\ \sin(\varphi_1)\dots\sin(\varphi_{n-2})\sin(\varphi_{n-1}) \end{pmatrix} : \varphi \in [0,\pi)^{n-2}\times[0,2\pi) \right\} \tag{7.29}$$

with $\varphi \triangleq (\varphi_1,\dots,\varphi_{n-1})$. The Minkowski functional of $\mathcal{O}$ is given by:

$$\mu_{\mathcal{O}}(x) = \begin{cases} \|x\|_2/\rho(\varphi(x)) & \text{if } x \neq 0 \\ 0 & \text{else} \end{cases} \tag{7.30}$$

where $\varphi(x) = (\varphi_1(x),\dots,\varphi_{n-1}(x))$ stands for the following angular coordinates of x:

$$\varphi_k(x) = \arccos\frac{x_k}{\sqrt{\sum_{i=k}^n x_i^2}}, \qquad \forall k \in [\![1,n-2]\!] \tag{7.31a}$$

$$\varphi_{n-1}(x) = 2\,\mathrm{arccot}\,\frac{x_{n-1}+\sqrt{x_n^2+x_{n-1}^2}}{x_n} \tag{7.31b}$$

Proof. It is easily verified that $\mu_{\mathcal{O}}(0) = 0$. Moreover, for all $x \in \mathbb{R}^n \setminus \{0\}$, the following equalities stand:

$$\begin{aligned}\mu_{\mathcal{O}}(x) &= \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{O}\} \\ &= \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\rho(\varphi(x))\mathcal{B}_2\} \\ &= \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : \frac{1}{t\rho(\varphi(x))}x \in \mathcal{B}_2\} \\ &= \mu_{\mathcal{B}_2}(x/\rho(\varphi(x))) = \|x\|_2/\rho(\varphi(x))\end{aligned} \tag{7.32}$$

This concludes the proof. □

7.3.2 Properties

First, the classical properties on the Minkowski functional associated with star-convex and convex sets are recalled. In particular, when $\mathcal{S}$ is a star-convex set at 0 (i.e. $t\mathcal{S} \subseteq \mathcal{S}$ for all $t \in [0, 1]$ [247]) of $\mathbb{R}^n$ whose interior contains 0, checking if $x \in \mathbb{R}^n$ belongs - in a loose sense - to $t\mathcal{S}$, can simply be achieved by verifying the inequality $\mu_{\mathcal{S}}(x) \leq t$. This fact is formalized, among others, in the two properties below.

Property 7.3.1 (Star-convex properties [243, 242])**.** *Let $\mathcal{S}$ be a bounded star-convex set at 0 of $\mathbb{R}^n$ whose interior contains 0. The Minkowski functional $\mu_{\mathcal{S}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ associated with $\mathcal{S}$ satisfies:*

1. *For all $x \in \mathbb{R}^n$, $0 \leq \mu_{\mathcal{S}}(x) < +\infty$,*
2. *For all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}_{\geq 0}$, $\mu_{\mathcal{S}}(tx) = t\mu_{\mathcal{S}}(x)$, and $\mu_{\mathcal{S}}(tx) = \mu_{\frac{1}{t}\mathcal{S}}(x)$ if $t \neq 0$,*
3. *$\mu_{\mathcal{S}}^{-1}([0, 1)) \subseteq \mathcal{S} \subseteq \mu_{\mathcal{S}}^{-1}([0, 1])$.*

Property 7.3.2 (Convex properties [179])**.** *Let $\mathcal{S}$ be a convex set whose interior contains 0. The Minkowski functional $\mu_{\mathcal{S}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ associated with $\mathcal{S}$ satisfies:*

1. *For all $x_1, x_2 \in \mathbb{R}^n$, $\mu_{\mathcal{S}}(x_1 + x_2) \leq \mu_{\mathcal{S}}(x_1) + \mu_{\mathcal{S}}(x_2)$,*
2. *$\mu_{\mathcal{S}} \in \mathcal{C}^0(\mathbb{R}^n, \mathbb{R}_{\geq 0})$,*
3. *$\mu_{\mathcal{S}}^{-1}([0, 1)) = \mathrm{intr}(\mathcal{S})$, $\mu_{\mathcal{S}}^{-1}([0, 1]) = \mathrm{cls}(\mathcal{S})$, $\mu_{\mathcal{S}}^{-1}(\{1\}) = \partial\mathcal{S}$.*

Item 1 hereabove combined with item 2 of Property 7.3.1 makes $\mu_{\mathcal{S}}$ a convex function.

For all non-empty sets $\mathcal{S}_1$ and $\mathcal{S}_2$, it is easily established from Definition 7.3.1 that $\mathcal{S}_1 \subseteq \mathcal{S}_2$ provides $\mu_{\mathcal{S}_2} \leq \mu_{\mathcal{S}_1}$, however $\mathcal{S}_1$ and $\mathcal{S}_2$ need to be star-convex at 0 for the reciprocal to hold (this is a particular case of (7.33b) stated below). Moreover, the usual set operations of union, intersection and Minkowski sum can also be expressed in terms of Minkowski functionals as follows.

Property 7.3.3 (Usual set operations)**.** *Given M a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ in matrix form, and $\mathcal{S}_1$, $\mathcal{S}_2$ two bounded star-convex sets at 0 of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and whose interior contains 0. For all $x \in \mathbb{R}^n$, the following identities hold:*

$$\mu_{\mathcal{S}_1 \cup_M \mathcal{S}_2}(x) = \min\{\mu_{\mathcal{S}_1}(x), \mu_{\mathcal{S}_2}(Mx)\} \tag{7.33a}$$

$$\mu_{\mathcal{S}_1 \cap_M \mathcal{S}_2}(x) = \max\{\mu_{\mathcal{S}_1}(x), \mu_{\mathcal{S}_2}(Mx)\} \tag{7.33b}$$

$$\mu_{\mathcal{S}_1 \oplus_M \mathcal{S}_2}(x) = \inf_{x_1 + x_2 = x} \{\max\{\mu_{\mathcal{S}_1}(x_1), \mu_{\mathcal{S}_2}(Mx_2)\}\} \tag{7.33c}$$

Given $\star \in \{\cup, \cap, \oplus\}$, then the set operation $\mathcal{S}_1 \star_M \mathcal{S}_2$ stands for $\mathcal{S}_1 \star \{x \in \mathbb{R}^n : Mx \in \mathcal{S}_2\}$.

Proof. Given any set operation $\star \in \{\cup, \cap, \oplus\}$, by definition of $\mathcal{S}_1 \star_M \mathcal{S}_2$, and since $\mu_{\{x \in \mathbb{R}^n : Mx \in \mathcal{S}_2\}} = \mu_{\mathcal{S}_2} \circ M$ holds, then, the proof reduces to showing that (7.33) holds when $m = n$ and $M = I_n$. Note that (7.33a) holds even if $\mathcal{S}_1$ and $\mathcal{S}_2$ are simply non-empty sets without further assumptions: this follows from Definition 7.3.1 and from $\inf \cup = \min \inf$ in this context (see Exercise 1.5 from [266], which can be solved by contradiction):

$$\inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t(\mathcal{S}_1 \cup \mathcal{S}_2)\} = \inf \cup_{i \in \{1,2\}} \{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{S}_i\} = \min_{i \in \{1,2\}} \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{S}_i\} \tag{7.34}$$

The proof of (7.33b) can be found in Lemma 5.49, page 192 of [5]. Finally, the proof of (7.33c) can be found in Lemma 1, page 9 of [127] in the convex and balanced case, but it is easily noticed that the proof only leverages Property 7.3.1, hence Lemma 1 at page 9 of [127] also holds in the star-convex case. □

Remark 7.3.6. *The identities* (7.33a), (7.33b) *and* (7.33c) *can easily be generalized to finite collections* $(\mathcal{S}_i)_{1 \leq i \leq m}$ *of bounded star-convex sets at* 0 *of* $\mathbb{R}^n$ *and whose interior contains* 0*:*

$$\mu_{\cup_{i=1}^m \mathcal{S}_i}(x) = \min_{1 \leq i \leq m} \mu_{\mathcal{S}_i}(x) \tag{7.35a}$$

$$\mu_{\cap_{i=1}^m \mathcal{S}_i}(x) = \max_{1 \leq i \leq m} \mu_{\mathcal{S}_i}(x) \tag{7.35b}$$

$$\mu_{\oplus_{i=1}^m \mathcal{S}_i}(x) = \inf_{x_1 + \cdots + x_m = x} \left\{ \max_{1 \leq i \leq m} \mu_{\mathcal{S}_i}(x_i) \right\} \tag{7.35c}$$

The identity (7.33a) *can be extended to uncountable collections* $(\mathcal{S}_j)_{j \in [0,1]}$ *simply by replacing the* $\min_{1 \leq i \leq m}$ *by an* $\inf_{j \in [0,1]}$*. The existence of this* inf *is easily deduced by the fact that the Minkowski functional is bounded from below by* 0*, and the deduction of this identity follows the same steps as* (7.34)*, by leveraging* $\inf \cup = \inf \inf$ *in this context (using a similar proof by contradiction).*

Among the previous set operations, the Minkowski functional associated with the Minkowski sum (7.33c) is impractical for direct computation since it combines both an $\inf$ and a $\max$ in its expression. However, elementary upper- and lower-bounds can be retrieved respectively by evaluating the $\max$ at specific values of x_i (hence upper-bounding the $\inf$), or by determining the $\inf$ of the average of all of the $\mu_{\mathcal{S}_i}(x_i)$ (hence lower-bounding the $\max$). The author also suggests in Property 7.3.4 some less straightforward inequalities to bound the value of (7.33c).

Property 7.3.4 (Minkowski sum of convex sets)**.** *Let* $(\mathcal{S}_i)_{1 \leq i \leq m}$ *be a collection of bounded star-convex sets at* 0 *of* $\mathbb{R}^n$ *whose interior contains* 0*. For all* $x \in \mathbb{R}^n \setminus \{0\}$*, the following holds:*

$$\frac{1}{m} \cdot \mu_{\text{hull}(\cup_{i=1}^m \mathcal{S}_i)}(x) \leq \mu_{\oplus_{i=1}^m \mathcal{S}_i}(x) \leq \frac{1}{\sum_{i=1}^m \frac{1}{\mu_{\mathcal{S}_i}(x)}} \tag{7.36}$$

which can be lower-bounded again using $\max_{1 \leq i \leq m}\{s_i \mu_{\mathcal{S}_i}(x)\} \leq \mu_{\text{hull}(\cup_{i=1}^m \mathcal{S}_i)}(x)$*, where the* $(s_i)_{1 \leq i \leq m} \in \mathbb{R}^m_{>0}$ *are taken such that* $\text{hull}(\cup_{i=1}^m \mathcal{S}_i) \subseteq \cap_{i=1}^m \frac{1}{s_i} \mathcal{S}_i$*.*

Proof. $\boxed{\mu_{\oplus} \leq \dots}$ Let $x \in \mathbb{R}^n \setminus \{0\}$. The $\mathcal{S}_i$ being star-convex at 0, their Minkowski sum is also star-convex at 0. Hence given $(t_i)_{1 \leq i \leq m} \in \mathbb{R}^m_{>0}$ taken such that $t_i x \in \mathcal{S}_i$, then for all $0 \leq \lambda \leq (\sum_{i=1}^m t_i)$, $\lambda x \in \oplus_{i=1}^m \mathcal{S}_i$ is verified by definition of the Minkowski sum. This provides the following lower-bound:

$$\sup \oplus_{i=1}^m \{t \in \mathbb{R}_{>0}, tx \in \mathcal{S}_i\} \leq \sup\{t \in \mathbb{R}_{>0}, tx \in \oplus_{i=1}^m \mathcal{S}_i\} \tag{7.37}$$

Note that $\sup \oplus = \sum \sup$ in this context [222], hence the expression above can be rewritten as [200]:

$$\sum_{i=1}^{m} \frac{1}{\inf\{t \in \mathbb{R}_{>0}, x \in t\mathcal{S}_i\}} \leq \frac{1}{\inf\{t \in \mathbb{R}_{>0}, x \in t \oplus_{i=1}^{m} \mathcal{S}_i\}} \tag{7.38}$$

Recognizing the Definition 7.3.1 of the Minkowski functional in the expression above provides the upper-bound in (7.36).

$\boxed{\cdots \leq \mu_{\oplus}}$ Since for all i, $\mathcal{S}_i \subseteq \text{hull}(\bigcup_{j=1}^{m} \mathcal{S}_j)$, the following upper-bound holds:

$$\sup\{t \in \mathbb{R}_{>0}, tx \in \oplus_{i=1}^{m}\mathcal{S}_i\} \leq \sup\{t \in \mathbb{R}_{>0}, tx \in \oplus_{i=1}^{m} \text{hull}(\cup_{i=1}^{m}\mathcal{S}_i)\} \tag{7.39}$$

Moreover, since $\oplus_{i=1}^{m} \text{hull}(\bigcup_{i=1}^{m} \mathcal{S}_i) = m \cdot \text{hull}(\bigcup_{i=1}^{m} \mathcal{S}_i)$, the following holds:

$$\frac{1}{\inf\{t \in \mathbb{R}_{>0}, x \in t \oplus_{i=1}^{m} \mathcal{S}_i\}} \leq \frac{m}{\inf\{t \in \mathbb{R}_{>0}, x \in t\,\text{hull}(\cup_{i=1}^{m}\mathcal{S}_i)\}} \tag{7.40}$$

Recognizing the Definition 7.3.1 of the Minkowski functional provides the lower-bound in (7.36). Finally, since all $\mathcal{S}_i$ contain 0 in their interior, for all i there exists $\varepsilon_i > 0$ such that $\varepsilon_i\mathcal{B}_2 \subseteq \mathcal{S}_i$, meaning there exists $s_i \in \mathbb{R}_{>0}$ such that $\text{hull}(\cup_{i=1}^{m}\mathcal{S}_i) \subseteq \frac{1}{s_i}\varepsilon_i\mathcal{B}_2 \subseteq \frac{1}{s_i}\mathcal{S}_i$. Given $\text{hull}(\cup_{i=1}^{m}\mathcal{S}_i) \subseteq \cap_{i=1}^{m}\frac{1}{s_i}\mathcal{S}_i$, the straightforward generalization of (7.33b) to m intersections combined with item 2 of Property 7.3.1 provides $\max_{1\leq i\leq m}\{s_i\mu_{\mathcal{S}_i}(x)\} \leq \mu_{\text{hull}(\cup_{i=1}^{m}\mathcal{S}_i)}(x)$. □

Although these inner- and outer-approximations may not always be practically handled, the Minkowski functional of a Minkowski sum can still be obtained for special classes of convex sets for which dedicated algorithms already exist (e.g. constrained zonotopes [250, 225]). Moreover, if the Minkowski sum is difficult to handle, the usage of the Minkowski functional can also lead to advantageous results for special classes of convex set, as demonstrated hereafter on ellipsoids.

Property 7.3.5 (Minkowski sum of two ellipsoids). *Let $\mathcal{E}(Q_1) \triangleq \{x \in \mathbb{R}^n : x^\top Q_1 x \leq 1\}$ and $\mathcal{E}(Q_2)$ be defined similarly, with $Q_1, Q_2 \in \mathbb{S}_n^{++}(\mathbb{R})$. The eigenvalues of the matrix pencil (Q_2^{-1}, Q_1^{-1}), solutions of the equation*

$$\det(Q_1^{-1} - \lambda Q_2^{-1}) = 0 \tag{7.41}$$

are denoted as:

$$\underline{\lambda} = \lambda_1 \leq \cdots \leq \lambda_n = \overline{\lambda} \tag{7.42}$$

Let $\mathcal{I}$ be a finite subset of $[\underline{\lambda}, \overline{\lambda}]$. The following inequality holds for all $x \in \mathbb{R}^n$:

$$\max_{p\in\mathcal{I}} \mu_{\mathcal{E}(Q(p))}(x) \leq \mu_{\mathcal{E}(Q_1)\oplus\mathcal{E}(Q_2)}(x) \tag{7.43}$$

where

$$Q(p) \triangleq \left((1+p^{-1})Q_1^{-1} + (1+p)Q_2^{-1}\right)^{-1} \tag{7.44}$$

This outer-approximation can be made arbitrarily precise by increasing the number of points in $\mathcal{I}$.

Proof. Theorem 2.2.3 at page 118 of [159] states that:

$$\mathcal{E}(Q_1) \oplus \mathcal{E}(Q_2) = \bigcap_{p\in[\underline{\lambda},\overline{\lambda}]} \mathcal{E}(Q(p)) \tag{7.45}$$

hence:

$$\mathcal{E}(Q_1) \oplus \mathcal{E}(Q_2) \subseteq \bigcap_{p \in \mathcal{I}} \mathcal{E}(Q(p)) \tag{7.46}$$

The result is then obtained by applying (7.35b) to the right-hand side of the inclusion (7.46). □

Finally, the Minkowski functional of a linearly transformed set is given below.

Property 7.3.6 (Linear transformation)**.** *Given M a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ in matrix form and $\mathcal{S}$ a non-empty set of $\mathbb{R}^n$, if M has a left inverse, then for all $x \in \mathrm{Im}(M)$:*

$$\mu_{M\mathcal{S}}(x) = \mu_{\mathcal{S}}(M^\dagger x) \tag{7.47}$$

with $M^\dagger$ the Moore–Penrose inverse of M, such that $M^\dagger M = I_n$.

Proof. The linear transformation M is injective, hence $M^\dagger$ is the inverse of M on $\mathrm{Im}(M)$. Moreover, for all $t \in \mathbb{R}_{\geq 0}$, $tM\mathcal{S} \subseteq \mathrm{Im}(M)$, hence (7.47) holds for all $x \in \mathrm{Im}(M)$. □

Other practical characterizations of the linear transformation of a convex set are obtained thereafter. This manuscript focuses on the non-invertible linear case since, to the author's knowledge, practical characterizations of $\mu_{M\mathcal{S}}$ are not found in the literature of the Minkowski functional.

7.3.3 Linear transformation of smooth convex sets

In order to generalize Property 7.3.6, this chapter provides two characterizations of the Minkowski functional of a linearly transformed smooth compact and convex set $\mathcal{S}$ of $\mathbb{R}^n$ whose interior contains the origin. The first characterization is obtained for strictly convex sets with the help of the Legendre transform, which is a common tool in convex analysis. The second characterization is obtained using a result on the orthogonal projection of convex set onto the linear subspaces of $\mathbb{R}^n$ [22]. From now on, M is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, k is defined by $k \triangleq \dim(\mathrm{Im}(M)) = \dim(\mathrm{Ker}(M)^\perp)$ with the assumption that $k \geq 1$ (indeed, $M\mathcal{S} = \{0\}$ if $k = 0$, which is not an interesting case to consider).

Definition 7.3.2 (Legendre transform [247])**.** *Given a closed and convex function $f : \mathbb{R}^n \to \tilde{\mathbb{R}}$, the Legendre transform of f is given by the convex function $f^* : \mathbb{R}^n \to \tilde{\mathbb{R}}$ such that:*

$$f^*(y) \triangleq \sup_{x \in \mathbb{R}^n} \left(\langle y | x \rangle - f(x) \right) \tag{7.48}$$

Theorem 7.3.1 (First characterization)**.** *Let $\mathcal{S}$ be a smooth compact and strictly convex set of $\mathbb{R}^n$ whose interior contains 0, and $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a bijective strictly increasing and strictly convex function such that $g \circ \mu_{\mathcal{S}}$ is differentiable over $\mathbb{R}^n$. Let $f_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $f_2 : \mathbb{R}^m \to \mathbb{R}^m$ be two functions defined by*

$$f_1 = \frac{\partial}{\partial x}[g \circ \mu_{\mathcal{S}}] \tag{7.49a}$$

$$f_2 = M \circ \frac{\partial}{\partial x}[(g \circ \mu_{\mathcal{S}})^*] \circ M^\top \tag{7.49b}$$

Then, f_1 is invertible on $f_1(\mathbb{R}^n)$, and for all $y \in \frac{\partial}{\partial x} f_1(\mathbb{R}^n)$

$$(g \circ \mu_{\mathcal{S}})^*(y) = \left\langle f_1^{-1}(y) | y \right\rangle - g \circ \mu_{\mathcal{S}} \circ f_1^{-1}(y) \tag{7.50}$$

From here, if f_2 is invertible on $f_2(\mathbb{R}^m)$, it is possible to retrieve the expression of $\mu_{M\mathcal{S}}$ by computing

$$\mu_{M\mathcal{S}}(y) = g^{-1}(\langle f_2^{-1}(y)|y\rangle - (g\circ\mu_{\mathcal{S}})^*\circ M^\top \circ f_2^{-1}(y)) \tag{7.51}$$

which offers a characterization of the linear transformation of $\mathcal{S}$ by M that - in spite of its apparent complexity - can sometimes be treated algebraically. The problem is reduced to finding a suitable g and an inverse for f_1 and f_2.

Proof. First it is proven that $g\circ\mu_{\mathcal{S}}$ is a strictly convex function. Let $x_1, x_2 \in \mathbb{R}^n$ such that $x_1 \neq x_2$ and $t \in (0,1)$. On one hand, if $\mu_{\mathcal{S}}(x_1) = \mu_{\mathcal{S}}(x_2) = \lambda$, then necessarily $x_1, x_2 \neq 0$, and by strict convexity of $\mathcal{S}$:

$$\begin{aligned} tx_1 + (1-t)x_2 \in \text{intr}(\lambda\mathcal{S}) &\Leftrightarrow \mu_{\lambda\mathcal{S}}(tx_1 + (1-t)x_2) < 1 \\ &\Leftrightarrow \mu_{\mathcal{S}}(tx_1 + (1-t)x_2) < \lambda \end{aligned} \tag{7.52}$$

Since g is a strictly increasing function, $(g\circ\mu_{\mathcal{S}})(tx_1 + (1-t)x_2) < g(\lambda) = tg(\lambda) + (1-t)g(\lambda)$, hence

$$(g\circ\mu_{\mathcal{S}})(tx_1 + (1-t)x_2) < t(g\circ\mu_{\mathcal{S}})(x_1) + (1-t)(g\circ\mu_{\mathcal{S}})(x_2) \tag{7.53}$$

On the other hand, if $\mu_{\mathcal{S}}(x_1) \neq \mu_{\mathcal{S}}(x_2)$, then by convexity of $\mu_{\mathcal{S}}$ and from the fact that g is increasing, the following holds: $(g\circ\mu_{\mathcal{S}})(tx_1 + (1-t)x_2) \leq g(t\mu_{\mathcal{S}}(x_1) + (1-t)\mu_{\mathcal{S}}(x_2))$. Finally by strict convexity of g, it follows:

$$(g\circ\mu_{\mathcal{S}})(tx_1 + (1-t)x_2) < t(g\circ\mu_{\mathcal{S}})(x_1) + (1-t)(g\circ\mu_{\mathcal{S}})(x_2) \tag{7.54}$$

Hence $g\circ\mu_{\mathcal{S}}$ is a strictly convex function. The rest of the proof is based on explicitly calculating the right-hand side of the following identity (the Legendre transform of the identity found at page 142-143 of [234]):

$$g\circ\mu_{M\mathcal{S}} = ((g\circ\mu_{\mathcal{S}})^*\circ M^\top)^* \tag{7.55}$$

Given $f : \mathbb{R}^n \to \mathbb{R}$ a strictly convex and differentiable function, $\frac{\partial}{\partial x}f$ is invertible on $\frac{\partial}{\partial x}f(\mathbb{R}^n)$, and for all $y \in \frac{\partial}{\partial x}f(\mathbb{R}^n)$, the Legendre transform of f at y is given by (see Remark 1.6.18 of [247]):

$$f^*(y) = \left\langle (\frac{\partial}{\partial x}f)^{-1}(y)|y \right\rangle - f\circ(\frac{\partial}{\partial x}f)^{-1}(y) \tag{7.56}$$

Applying the previous result to $g\circ\mu_{\mathcal{S}}$ provides:

$$(g\circ\mu_{\mathcal{S}})^*(y) = \langle f_1^{-1}(y)|y\rangle - g\circ\mu_{\mathcal{S}}\circ f_1^{-1}(y) \tag{7.57}$$

The result can be applied again, this time to $(g\circ\mu_{\mathcal{S}})^*\circ M^\top$. Under the assumption that $(g\circ\mu_{\mathcal{S}})^*\circ M^\top$ is strictly convex and differentiable, then $f_2 = \frac{\partial}{\partial x}((g\circ\mu_{\mathcal{S}})^*\circ M^\top) = M\circ\frac{\partial}{\partial x}[(g\circ\mu_{\mathcal{S}})^*]\circ M^\top$ is invertible on $f_2(\mathbb{R}^m)$, and the identity (7.55) can be written:

$$\begin{aligned} g\circ\mu_{M\mathcal{S}} &= ((g\circ\mu_{\mathcal{S}})^*\circ M^\top)^* \\ &= \left\langle (\frac{\partial}{\partial x}[(g\circ\mu_{\mathcal{S}})^*\circ M^\top])^{-1}(y)|y \right\rangle - (g\circ\mu_{\mathcal{S}})^*\circ M^\top\circ(\frac{\partial}{\partial x}[(g\circ\mu_{\mathcal{S}})^*\circ M^\top])^{-1}(y) \\ &= \left\langle (M\circ\frac{\partial}{\partial x}[g\circ\mu_{\mathcal{S}}]^*\circ M^\top)^{-1}(y)|y \right\rangle - (g\circ\mu_{\mathcal{S}})^*\circ M^\top\circ(M\circ\frac{\partial}{\partial x}[g\circ\mu_{\mathcal{S}}]^*\circ M^\top)^{-1}(y) \end{aligned} \tag{7.58}$$

which concludes the proof. □

Remark 7.3.7. *This first result is a broad generalization of the well-known result on ellipsoids which states that $M\mathcal{E}(P) = \mathcal{E}((MP^{-1}M^\top)^{-1})$, where $\mathcal{E}(P) \triangleq \{x \in \mathbb{R}^n : x^\top P x \leq 1\}$ and M is a full row rank matrix.*

Example 7.3.6. *Consider the Minkowski functional associated to $\mathcal{B}_{4/3}$, the unit ball of norm $4/3$ of $\mathbb{R}^3$:*

$$\mu_{\mathcal{B}_{4/3}}(x,y,z) \triangleq \left(x^{4/3} + y^{4/3} + z^{4/3}\right)^{3/4} \tag{7.59}$$

and let us compute the Minkowski functional associated to $M\mathcal{B}_{4/3}$ where M is defined by:

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \tag{7.60}$$

Since $\mathcal{B}_{4/3}$ is a smooth compact and strictly convex set, Theorem 7.3.1 is applied.

The function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $g(s) = (3/4)s^{4/3}$ is introduced. It is easily verified that g is bijective (with $g^{-1}(s) = \left(2/\sqrt{3}\right)^{3/2} s^{3/4}$), strictly increasing, strictly convex, and that $g \circ \mu_{\mathcal{B}_{4/3}}$ is smooth over $\mathbb{R}^3$. Following Theorem 7.3.1, $f_1 = \frac{\partial}{\partial x}[g \circ \mu_{\mathcal{B}_{4/3}}]$ and its inverse are given by:

$$f_1(x,y,z) = \begin{pmatrix} x^{1/3} & y^{1/3} & z^{1/3} \end{pmatrix}^\top \tag{7.61a}$$

$$f_1^{-1}(x,y,z) = \begin{pmatrix} x^3 & y^3 & z^3 \end{pmatrix}^\top \tag{7.61b}$$

This provides the Legendre transform of $g \circ \mu_{\mathcal{B}_{4/3}}$:

$$(g \circ \mu_{\mathcal{B}_{4/3}})^*(x,y,z) = \frac{1}{4}(x^4 + y^4 + z^4) \tag{7.62}$$

Again, following Theorem 7.3.1, $f_2 = M \circ \frac{\partial}{\partial x}[(g \circ \mu_{\mathcal{B}_{4/3}})^*] \circ M^\top$ and its inverse are given by:

$$f_2(x,y) = \begin{pmatrix} 2(x+y)^3 + (x-y)^3 \\ 2(x+y)^3 - (x-y)^3 \end{pmatrix} \tag{7.63a}$$

$$f_2^{-1}(x,y) = \begin{pmatrix} 2^{-4/3}\left(2^{-1/3}(x+y)^{1/3} + (x-y)^{1/3}\right) \\ 2^{-4/3}\left(2^{-1/3}(x+y)^{1/3} - (x-y)^{1/3}\right) \end{pmatrix} \tag{7.63b}$$

Finally, the Minkowski functional associated to $M\mathcal{B}_{4/3}$ is given by:

$$\mu_{M\mathcal{B}_{4/3}}(x,y) = \left(\frac{2}{\sqrt{3}}\right)^{3/2} \left(\frac{x}{2^{4/3}}\left(\frac{(x+y)^{1/3}}{2^{1/3}} + (x-y)^{1/3}\right) + \frac{y}{2^{4/3}}\left(\frac{(x+y)^{1/3}}{2^{1/3}} - (x-y)^{1/3}\right) - \frac{1}{2^{10/3}}\left(\frac{(x+y)^{4/3}}{2^{1/3}} + (x-y)^{4/3}\right)\right)^{3/4} \tag{7.64}$$

Let $(v_1, \ldots, v_k)$ denotes a basis of $\mathrm{Ker}(M)^\perp$ and $(v_{k+1}, \ldots, v_n)$ denotes a basis of $\mathrm{Ker}(M)$. $V_{\mathrm{Ker}^\perp}$ is the matrix whose columns are formed by $(v_1, \ldots, v_k)$ and V_{Ker} is the matrix whose columns are formed by $(v_{k+1}, \ldots, v_n)$. The invertible matrix $V \in \mathbb{R}^{n \times n}$ is also defined by the matrix whose columns are formed by $(v_1, \ldots, v_n)$, i.e. such that:

$$V = \begin{pmatrix} V_{\mathrm{Ker}^\perp} & V_{\mathrm{Ker}} \end{pmatrix} \tag{7.65}$$

Theorem 7.3.2 (Second characterization). *Let $\mathcal{S}$ be a smooth compact and convex set of $\mathbb{R}^n$ whose interior contains 0. Then the following identity holds:*

$$\mu_{M\mathcal{S}}(x) = \begin{cases} \mu_{p_{\mathrm{Ker}^\perp}(\mathcal{S})}((MV_{\mathrm{Ker}^\perp})^\dagger x) & \text{if } x \in \mathrm{Im}(M) \\ +\infty & \text{else} \end{cases} \tag{7.66}$$

where $p_{\mathrm{Ker}^\perp}(\mathcal{S})$ denotes the orthogonal projection of $\mathcal{S}$ onto $\mathrm{Ker}(M)^\perp$. The Minkowski functional on the right-hand side can be computed using the following system:

$$\mu_{p_{\mathrm{Ker}^\perp}(\mathcal{S})}(y_{\mathrm{Ker}^\perp}) = \inf \left\{ t \in \mathbb{R}_+^* \,:\, \exists y_{\mathrm{Ker}} \in \mathbb{R}^{n-k} \mid \begin{cases} \eta_{\mathcal{S}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) \leq t \\ y_{\mathrm{Ker}^\perp} + y_{\mathrm{Ker}} \neq 0 \Rightarrow \frac{\partial \eta_{\mathcal{S}}}{\partial y_{\mathrm{Ker}}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) = 0 \end{cases} \right\} \tag{7.67}$$

where $y_{\mathrm{Ker}^\perp} \in \mathbb{R}^k$ and $y_{\mathrm{Ker}} \in \mathbb{R}^{n-k}$ are expressed in the $(v_1, \ldots, v_k)$ and $(v_{k+1}, \ldots, v_n)$ basis respectively, with:

$$\eta_{\mathcal{S}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) \triangleq \mu_{\mathcal{S}}\left(V \begin{pmatrix} y_{\mathrm{Ker}^\perp} \\ y_{\mathrm{Ker}} \end{pmatrix}\right) \tag{7.68}$$

The right-hand side of (7.67) can sometimes be treated algebraically, providing a characterization of the linear transformation of $\mathcal{S}$ by M. The problem is reduced to solving the equation $\frac{\partial \eta_{\mathcal{S}}}{\partial y_{\mathrm{Ker}}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) = 0$ with y_{Ker} unknown.

Proof. By definition of $\mathrm{Ker}(M)$ and $\mathrm{Ker}(M)^\perp$, the following equality holds:

$$M = MV_{\mathrm{Ker}^\perp} \begin{pmatrix} I_k & 0 \end{pmatrix} V^{-1} \tag{7.69}$$

where moreover, $MV_{\mathrm{Ker}^\perp}$ is left-invertible, hence by Property 7.3.6, the following holds:

$$\mu_{M\mathcal{S}}(x) = \mu_{\begin{pmatrix} I_k & 0 \end{pmatrix} V^{-1}\mathcal{S}}((MV_{\mathrm{Ker}^\perp})^\dagger x) \tag{7.70}$$

From here, it can be noticed that $\begin{pmatrix} I_k & 0 \end{pmatrix} V^{-1}\mathcal{S}$ is the orthogonal projection of $\mathcal{S}$ onto $\mathrm{Ker}(M)^\perp$ expressed in the $(v_1, \ldots, v_k)$ basis. Finally, the results from [22] provides the expression (7.66) and (7.67) when $\mathcal{S}$ is a smooth compact and convex set of $\mathbb{R}^n$ whose interior contains 0, concluding the proof. □

Example 7.3.7. *Consider the Minkowski functional associated to $\mathcal{B}_4$, the unit ball of norm 4 of $\mathbb{R}^2$:*

$$\mu_{\mathcal{B}_4}(x, y) \triangleq \left(x^4 + y^4\right)^{1/4} \tag{7.71}$$

and let us compute the Minkowski functional associated to $M\mathcal{B}_4$ where M is defined by:

$$M = \begin{pmatrix} 1 & 2 \end{pmatrix} \tag{7.72}$$

Since $\mathcal{B}_4$ is a smooth compact and convex set, Theorem 7.3.2 is applied.

The matrix V associated with M can be taken to be $V = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$. Hence:

$$\eta_{\mathcal{B}_4}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) \triangleq \left((y_{\mathrm{Ker}^\perp} + 2y_{\mathrm{Ker}})^4 + (2y_{\mathrm{Ker}^\perp} - y_{\mathrm{Ker}})^4\right)^{1/4} \tag{7.73}$$

and:

$$\frac{\partial \eta_{\mathcal{B}_4}}{\partial y_{\mathrm{Ker}}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) = 0 \Leftrightarrow 17y_{\mathrm{Ker}}^3 + 18y_{\mathrm{Ker}}^2 y_{\mathrm{Ker}^\perp} + 24y_{\mathrm{Ker}} y_{\mathrm{Ker}^\perp}^2 - 6y_{\mathrm{Ker}^\perp}^3 = 0 \tag{7.74}$$

the equation on the right possesses a single real solution given by $y_{\mathrm{Ker}} = \frac{1}{17}(-6 - 5\cdot 2^{1/3} + 10\cdot 2^{2/3}) y_{\mathrm{Ker}^\perp}$, moreover $(MV_{\mathrm{Ker}^\perp})^\dagger = 1/5$. Finally, the Minkowski functional associated to $M\mathcal{B}_4$ is given by:

$$\mu_{M\mathcal{B}_4}(x) = \mu_{p_{\mathrm{Ker}^\perp}(\mathcal{B}_4)}(x/5) = \eta_{\mathcal{B}_4}\left(x/5, x(-6 - 5\cdot 2^{1/3} + 10\cdot 2^{2/3})/85\right) \tag{7.75}$$

7.3.4 Analytic fault isolation

The flexibility and generality of the Minkowski functional for set-membership approaches is illustrated here by applying its properties to the fault isolation scheme of Theorem 7.2.2. In Theorem 7.3.3, the problem of calculating the DIT set (7.15) is reduced to determining analytical thresholds under the form of scalar inequalities. In Theorem 7.3.4, a condition on the minimal fault magnitude is given to guarantee isolability by the DIT. In both theorems, for all i, the sets $\mathcal{W}_i$ and $\mathcal{F}_i$ of Definition 7.2.1 are assumed to be bounded star-convex sets at 0 whose interior contains 0.

Theorem 7.3.3 (Analytic Direct Image Test)**.** *For all $\mathcal{I} \in \mathbf{P}_{n_f}$, the Minkowski functional associated with the feasible sets $\mathcal{R}_{\theta,\mathcal{I}}$ (7.13) introduced in Definition 7.2.2 can be computed by:*

$$\mu_{\theta,\mathcal{I}}(r) = \inf_{\substack{\Sigma r_i = r \\ i \in \mathcal{J}}} \max \left\{ \max_{1 \leq i \leq n_w} \mu_{W_i(\theta)\mathcal{W}_i}(r_i), \max_{i \in \mathcal{I}} \mu_{F_i(\theta)\mathcal{F}_i}(r_{n_w+i}) \right\} \tag{7.76}$$

with $\mathcal{J} = [\![1, n_w]\!] \cup (\{n_w\} \oplus \mathcal{I})$. Given $\underline{\mu}_{\theta,\overline{\mathcal{I}}}$ and $\overline{\mu}_{\theta,\overline{\mathcal{I}}}$ resp. a lower-bound and an upper-bound to $\mu_{\theta,\overline{\mathcal{I}}}$, the DIT set (7.15) can be approximated by:

$$\left\{ \mathcal{I} \in \mathbf{P}_{n_f} \, : \, \sup_{\theta \in \Theta} \underline{\mu}_{\theta,\overline{\mathcal{I}}}(r(\theta)) > 1 \right\} \subseteq DIT \subseteq \left\{ \mathcal{I} \in \mathbf{P}_{n_f} \, : \, \sup_{\theta \in \Theta} \overline{\mu}_{\theta,\overline{\mathcal{I}}}(r(\theta)) > 1 \right\} \tag{7.77}$$

Proof. This is simply a rewriting of Theorem 7.2.2 obtained by leveraging the Minkowski functional identities of Property 7.3.1 and Remark 7.3.6. □

Corollary 7.3.1 (Analytic Robust Direct Image Test)**.** *Similarly, the robust DIT set (7.18) can be approximated by:*

$$\left\{ \mathcal{I} \in \mathbf{P}_{n_f} \, : \, \sup_{\theta_2 \in \Theta} \inf_{\theta_1 \in \Theta} \underline{\mu}_{\theta_1,\overline{\mathcal{I}}}(r(\theta_2)) > 1 \right\} \subseteq RDIT \subseteq \left\{ \mathcal{I} \in \mathbf{P}_{n_f} \, : \, \sup_{\theta_2 \in \Theta} \inf_{\theta_1 \in \Theta} \overline{\mu}_{\theta_1,\overline{\mathcal{I}}}(r(\theta_2)) > 1 \right\} \tag{7.78}$$

As noted in Remark 7.2.5 and Corollary 7.2.1, the tests performed to obtain these two sets do not require to know which uncertainty θ is responsible for each residual in R.

Remark 7.3.8. *By definition, the sets on the left-hand side of (7.77) and (7.78) avoid false fault detection. Similarly, the sets on the right-hand side can guarantee the absence of faults.*

The $\inf$ in (7.76) comes from Minkowski sums, and can be handled either by leveraging the inequalities described before and in Property 7.3.4, or by using dedicated algorithms which compute Minkowski sums for special classes of convex shapes. Similarly, computing the linear transformations of the sets $\mathcal{W}_i$ and $\mathcal{F}_i$ can be achieved either by using dedicated algorithms, or by leveraging Theorem 7.3.1 or 7.3.2. All of these approaches allow to find upper- and lower-bounds to $\mu_{\theta,\mathcal{I}}$, here denoted $\overline{\mu}_{\theta,\mathcal{I}}$ and $\underline{\mu}_{\theta,\mathcal{I}}$, and which respectively correspond to an inner- and outer-approximation of the feasible set $\mathcal{R}_{\theta,\mathcal{I}}$. These upper- and lower-bounds provide closed-form inequalities for residual thresholding, and their real-time evaluation can be viewed as the computation of *meta-residuals* synthesizing *how close is the system from the boundary of $\mathcal{R}_{\theta,\mathcal{I}}$*. The author believes that the name *Minkowski signals* is appropriate to refer the real-time evaluation of these functions in the context of system diagnosis.

Theorem 7.3.4 (Minimal fault magnitude for isolation). *Let $Q_{\mathcal{I}}$ with $\mathcal{I} \triangleq \{i_1, \ldots, i_p\} \in \mathbf{P}_{n_f}$ denote the following statement:*

$$Q_{\mathcal{I}} : \text{“ At least one of the } i_1\text{-th, } \ldots, i_p\text{-th vector fault is active”} \tag{7.79}$$

This statement is guaranteed to be verified (or isolated) using the DIT set (7.15) if the faults $(f_i)_{i\in\mathcal{I}}$ satisfy the following inequality:

$$\sup_{\theta\in\Theta} \mu_{\mathcal{S}_\theta}\left(\sum_{i\in\mathcal{I}} F_i(\theta) f_i\right) > 1 \text{ with } \mathcal{S}_\theta \triangleq \mathcal{R}_{\theta,\overline{\mathcal{I}}} \oplus \left(-\mathcal{R}_{\theta,\overline{\mathcal{I}}}\right) \tag{7.80}$$

Similarly, this statement is guaranteed to be verified (or isolated) using the RDIT set (7.18) if the faults $(f_i)_{i\in\mathcal{I}}$ satisfy the following inequality:

$$\inf_{\theta\in\Theta} \mu_{\mathcal{S}_\Theta}\left(\sum_{i\in\mathcal{I}} F_i(\theta) f_i\right) > 1 \text{ with } \mathcal{S}_\Theta \triangleq \mathcal{R}_{\Theta,\overline{\mathcal{I}}} \oplus \left(-\mathcal{R}_{\Theta,\overline{\mathcal{I}}}\right) \tag{7.81}$$

Proof. Since the operations of linear transformation, union and Minkowski sum preserve star-convexity at 0, $\mathcal{S}_\theta$ is star-convex at 0. Moreover, the inequality (7.80) implies that there exists $\theta \in \Theta$:

$$\sum_{i\in\mathcal{I}} F_i(\theta) f_i \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}} \oplus \left(-\mathcal{R}_{\theta,\overline{\mathcal{I}}}\right) \tag{7.82}$$

meaning there exists $\theta \in \Theta$, such that for all $w_i \in \mathcal{W}_i$ (with $i \in [\![1, n_w]\!]$) and $f_i \in \mathcal{F}_i$ (with $i \in \overline{\mathcal{I}}$),

$$\sum_{i\in\mathcal{I}} F_i(\theta) f_i + \sum_{i=1}^{n_w} W_i(\theta) w_i + \sum_{i\in\overline{\mathcal{I}}} F_i(\theta) f_i \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}} \tag{7.83}$$

hence, by Theorem 7.2.2, (7.83) guarantees that $\mathcal{I} \in$ DIT, which in turn provides the statement $Q_{\mathcal{I}}$. The proof follows the same principle in the robust case. □

7.4 Application

7.4.1 Example 1

Consider the following academic example of residuals $r(t)$ following the uncertain linear internal structure of Definition 7.2.1:

$$\begin{pmatrix} r_1(t) \\ r_2(t) \end{pmatrix} = \begin{pmatrix} 1 & \theta_1(t) & 0 \\ 0 & 1 & \theta_2(t) \end{pmatrix} \begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \end{pmatrix} + \begin{pmatrix} 1 & \theta_1(t) \\ 0 & \varepsilon \end{pmatrix} \begin{pmatrix} f_{1,1}(t) \\ f_{1,2}(t) \end{pmatrix} + \begin{pmatrix} \varepsilon \\ 1 \end{pmatrix} f_2(t) \tag{7.84a}$$

$$\text{i.e. } r(t) = W(\theta(t))w(t) + F_1(\theta(t))f_1(t) + F_2(\theta(t))f_2(t) \tag{7.84b}$$

where $r(t)$ are the residuals of the system, $\theta(t) \in \Theta$ are time-varying parametric uncertainties, $f_1(t) \in \mathcal{F}_1$ and $f_2(t) \in \mathcal{F}_2$ are potential faults to detect, and $w(t) \in \mathcal{W}$ are bounded noises affecting the residuals. This example is chosen with matrices in $\mathbb{R}^{2\times3}$, $\mathbb{R}^{2\times2}$ and $\mathbb{R}^{2\times1}$ to cover as many cases of matrix multiplication as possible. According to the set-membership approach

developed in Section 7.2, the three following sets should be determined or approximated in order to threshold the residuals of the system:

$$\mathcal{R}_{\theta,\emptyset} \triangleq W(\theta)\mathcal{W} \tag{7.85a}$$
$$\mathcal{R}_{\theta,\{1\}} \triangleq W(\theta)\mathcal{W} \oplus F_1(\theta)\mathcal{F}_1 \tag{7.85b}$$
$$\mathcal{R}_{\theta,\{2\}} \triangleq W(\theta)\mathcal{W} \oplus F_2(\theta)\mathcal{F}_2 \tag{7.85c}$$

These sets can be approximated using any sort of set-membership methodologies already available in the literature, e.g. by leveraging results on orthotopes, ellipsoids, parallelotopes, zonotopes, constrained zonotopes, hybrid zonotopes, etc; or (not exclusively) by using the analytical results exposed in this chapter. From here, following the analytic RDIT (Corollary 7.3.1) methodology of this chapter, the problem consists in obtaining or approximating the following Minkowski signals:

$$\mu_{\Theta,\emptyset}(r) \triangleq \mu_{\mathcal{R}_{\Theta,\emptyset}}(r) = \inf_{\theta\in\Theta} \mu_{\mathcal{R}_{\theta,\emptyset}}(r) \qquad \text{with } \mathcal{R}_{\Theta,\emptyset} \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_{\theta,\emptyset} \tag{7.86a}$$
$$\mu_{\Theta,\{1\}}(r) \triangleq \mu_{\mathcal{R}_{\Theta,\{2\}}}(r) = \inf_{\theta\in\Theta} \mu_{\mathcal{R}_{\theta,\{2\}}}(r) \qquad \text{with } \mathcal{R}_{\Theta,\{2\}} \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_{\theta,\{2\}} \tag{7.86b}$$
$$\mu_{\Theta,\{2\}}(r) \triangleq \mu_{\mathcal{R}_{\Theta,\{1\}}}(r) = \inf_{\theta\in\Theta} \mu_{\mathcal{R}_{\theta,\{1\}}}(r) \qquad \text{with } \mathcal{R}_{\Theta,\{1\}} \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_{\theta,\{1\}} \tag{7.86c}$$

where:

- $\mu_{\Theta,\emptyset}(r) > 1$ guarantees that at least one of the vector faults f_1 or f_2 is active (detection);
- $\mu_{\Theta,\{1\}}(r) > 1$ guarantees that the vector fault f_1 is active (isolation);
- $\mu_{\Theta,\{2\}}(r) > 1$ guarantees that the fault f_2 is active (isolation).

The sets $\mathcal{W}$, $\mathcal{F}_1$ and $\mathcal{F}_2$ are modeled by:

$$\mathcal{W} \triangleq \mathcal{E}(Q_{w,1}) \cup G_{w,2}\mathcal{B}_\infty \tag{7.87a}$$
$$\mathcal{F}_1 \triangleq \mathcal{E}(Q_{f_1}) \tag{7.87b}$$
$$\mathcal{F}_2 \triangleq \mathcal{I}_{f_2} = [-1/\varepsilon, 1/\varepsilon] \tag{7.87c}$$

In order to be as illustrative as possible, these sets are selected to exhibit many of the properties developed in this chapter. However, any other choice could have been made at this point. A more involved modeling of these sets would lead to better detection capabilities, but also to more elaborate expressions for the Minkowski signals, whereas simpler sets would be more easily handled, but less accurate in their detection capabilities. The expression of $\mu_{\Theta,\emptyset}(r)$ is explicitly obtained as:

$$\mu_{\Theta,\emptyset}(r) = \inf_{\theta\in\Theta} \min \left\{ \sqrt{r^\top \left(W(\theta)Q_{w,1}^{-1}W^\top(\theta)\right)^{-1} r}, \left\| L_{w,2}(\theta) r^\top \right\|_\infty \right\} \tag{7.88}$$

where Theorem 7.3.1 is leveraged to handle the linear transformation of the ellipsoid, and $L_{w,2}$ is obtained from $W(\theta)G_{w,2}$ using the results of Example 7.3.3.

In order to obtain an expression for $\mu_{\Theta,\{1\}}(r)$, $\mathcal{E}(Q_{w,1})$ is approximated by the zonotope $G_{w,1}\mathcal{B}_\infty$, leading to an hybrid zonotope defined by:

$$\begin{aligned} W(\theta)\mathcal{W} \oplus F_2(\theta)\mathcal{F}_2 &= W(\theta)\left(G_{w,1}\mathcal{B}_\infty \cup G_{w,2}\mathcal{B}_\infty\right) \oplus F_2(\theta)\mathcal{I}_{f_2} \\ &= \left(W(\theta)G_{w,1}\mathcal{B}_\infty \oplus F_2(\theta)\mathcal{I}_{f_2}\right) \cup \left(W(\theta)G_{w,2}\mathcal{B}_\infty \oplus F_2(\theta)\mathcal{I}_{f_2}\right) \\ &= \begin{pmatrix} W(\theta)G_{w,1} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix}\mathcal{B}_\infty \cup \begin{pmatrix} W(\theta)G_{w,2} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix}\mathcal{B}_\infty \end{aligned} \tag{7.89}$$

The Minkowski functional associated with this hybrid zonotope is given by:

$$\mu_{\Theta,\{1\}}(r) = \inf_{\theta\in\Theta} \min\left\{\left\|\mathcal{L}_1(\theta)r\right\|_\infty, \left\|\mathcal{L}_2(\theta)r\right\|_\infty\right\} \tag{7.90}$$

where $\mathcal{L}_1(\theta)$ and $\mathcal{L}_2(\theta)$ are resp. obtained from $\begin{pmatrix} W(\theta)G_{w,1} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix}$ and $\begin{pmatrix} W(\theta)G_{w,2} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix}$ using the results of Example 7.3.3. By leveraging the central symmetry of the set $\mathcal{R}_{\Theta,\{2\}}$, it can be noted that Theorem 7.3.4 guarantees that the magnitude of the vector fault f_1 is sufficient to ensure a fault detection if the following inequality holds:

$$\inf_{\theta\in\Theta} \mu_{\Theta,\{1\}}(F_1(\theta)f_1) > 2 \tag{7.91}$$

Finally, in order to obtain an analytical expression to $\mu_{\Theta,\{2\}}(r)$, $G_{w,2}\mathcal{B}_\infty$ is approximated by the ellipsoid $\mathcal{E}(Q_{w,2})$, hence:

$$\begin{aligned} W(\theta)\mathcal{W} \oplus F_1(\theta)\mathcal{F}_1 &= W(\theta)\left(\mathcal{E}(Q_{w,1}) \cup \mathcal{E}(Q_{w,2})\right) \oplus F_1(\theta)\mathcal{E}(Q_{f_1}) \\ &= \left(W(\theta)\mathcal{E}(Q_{w,1}) \oplus F_1(\theta)\mathcal{E}(Q_{f_1})\right) \cup \left(W(\theta)\mathcal{E}(Q_{w,2}) \oplus F_1(\theta)\mathcal{E}(Q_{f_1})\right) \end{aligned} \tag{7.92}$$

Although only ellipsoids are leveraged in the representation of the initial sets, $\mathcal{R}_{\Theta,\{1\}}$ is obtained here as a generic convex shape whose Minkowski functional can be approximated arbitrarily closely by:

$$\mu_{\Theta,\{2\}}(r) = \inf_{\theta\in\Theta} \min\left\{\max_{p\in\mathcal{I}_1} \sqrt{r^\top \mathcal{Q}_1(\theta,p)r}, \max_{p\in\mathcal{I}_2} \sqrt{r^\top \mathcal{Q}_2(\theta,p)r}\right\} \tag{7.93}$$

where $\mathcal{Q}_1(\theta,p)$, $\mathcal{Q}_2(\theta,p)$ and $\mathcal{I}_1$ and $\mathcal{I}_2$ are obtained with Property 7.3.5 on the Minkowski sum of two ellipsoids.

$$\mathcal{Q}_1(\theta,p) = \left((1+p^{-1})W(\theta)Q_{w,1}^{-1}W^\top(\theta) + (1+p)F_1(\theta)Q_{f_1}^{-1}F_1(\theta)^\top\right)^{-1} \tag{7.94a}$$

$$\mathcal{Q}_2(\theta,p) = \left((1+p^{-1})W(\theta)Q_{w,2}^{-1}W^\top(\theta) + (1+p)F_1(\theta)Q_{f_1}^{-1}F_1(\theta)^\top\right)^{-1} \tag{7.94b}$$

Again, by leveraging the central symmetry of the set $\mathcal{R}_{\Theta,\{1\}}$, it can be noted that Theorem 7.3.4 guarantees that the magnitude of the fault f_2 is sufficient to ensure a fault detection if the following inequality holds:

$$\inf_{\theta\in\Theta} \mu_{\Theta,\{2\}}(F_2(\theta)f_2) > 2 \tag{7.95}$$

The simulations have been carried out with the following numerical values:

$$\begin{aligned} &\Theta = [-1,1]^2, && \theta(t) = \begin{pmatrix} \cos(t) & \sin(t) \end{pmatrix}^\top, \\ &Q_{w,1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.3 \end{pmatrix}, && Q_{w,2} = \begin{pmatrix} 0.4 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ &G_{w,1} = Q_{w,1}^{-1/2}, && G_{w,2} = Q_{w,2}^{-1/2}, \\ &Q_{f_1} = \begin{pmatrix} 0.025 & 0.01 \\ 0.01 & 1 \end{pmatrix}, && \varepsilon = 0.2, \end{aligned} \tag{7.96}$$

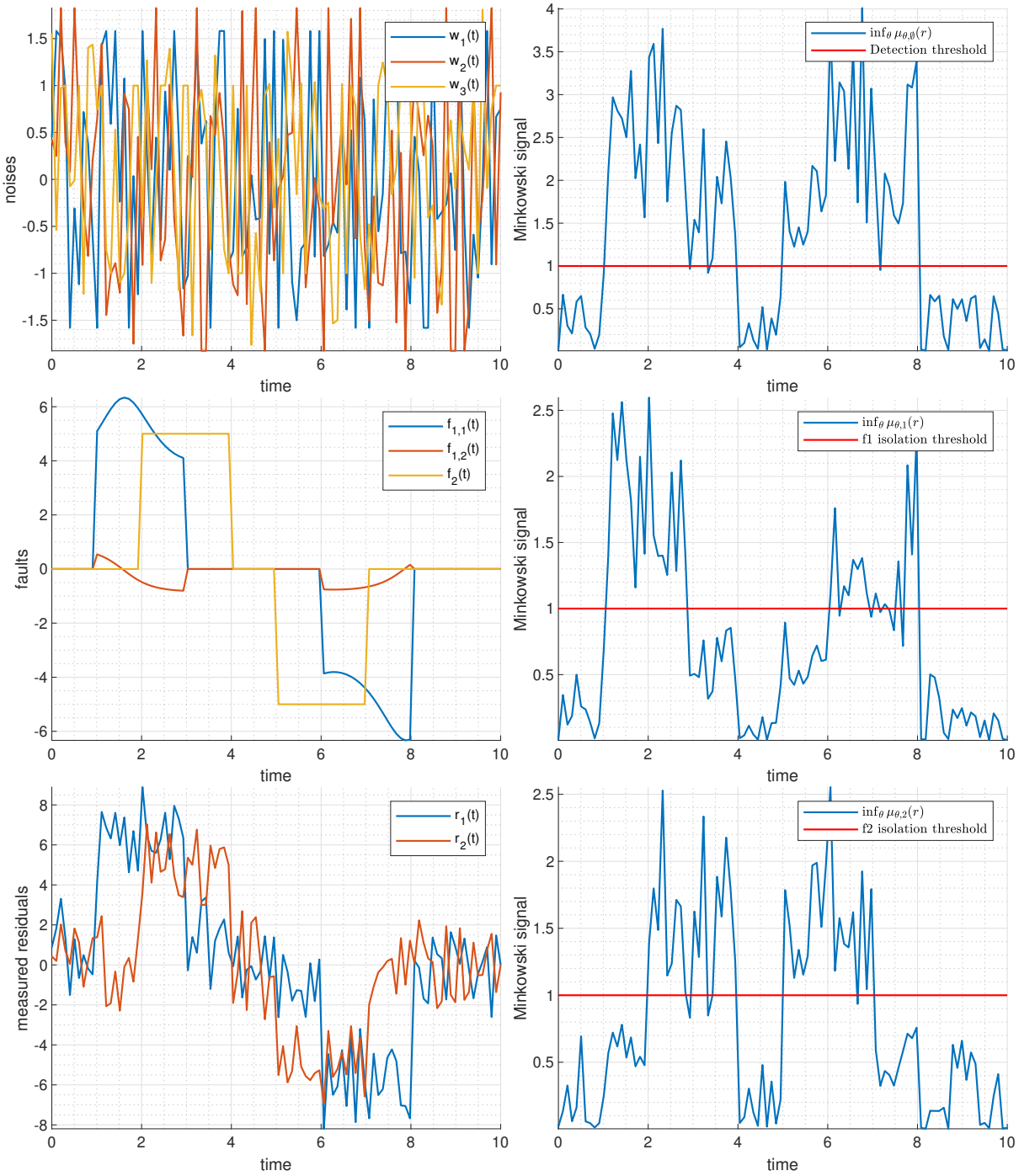


Figure 7.6: Computation of the Minkowski signals to perform the robust direct image.

The results of the simulation can be found in Figure 7.6 of this document. The vector fault f_1 is active for $t \in [1,3] \cup [6,8]$, and the fault f_2 is active for $t \in [2,4] \cup [5,7]$. The Minkowski signal $\mu_{\Theta,\emptyset}(r)$ detects faults for $t \in [1,4] \cup [5,8]$ almost perfectly (with the exception of a few points). Moreover, the Minkowski signal $\mu_{\Theta,\{1\}}(r)$ perfectly isolates the fault f_1 for $t \in [1,3]$, and isolates, although less precisely, its activity in the interval $[6,8]$. Finally, the Minkowski signal $\mu_{\Theta,\{2\}}(r)$ almost perfectly isolates the fault f_2 for $t \in [2,4] \cup [5,7]$, again with the exception of a few points that could be easily dealt with by classical filtering techniques. The quality of the detection demonstrates the quality of the feasible sets underlying the Minkowski signals, which were computed by combining both the properties of this document and some existing results from the set-membership literature. Overall, for each $\theta \in \Theta$, the Minkowski signals are computed explicitly, while:

- $\mathcal{R}_{\theta,\emptyset}$ is considered to be the union of an ellipse and a zonotope;
- $\mathcal{R}_{\theta,\{1\}}$ is considered to be the union of two zonotopes (or an hybrid zonotope);
- $\mathcal{R}_{\theta,\{2\}}$ is considered to be the Minkowski sum of two ellipsoids.

This ultimately demonstrates the unifying nature of the proposed thresholding framework.

7.4.2 Example 2

Consider the following academic example of uncertain linear discrete-time system subject to faults and noises:

$$x_{t+1} = \begin{pmatrix} -1 & 0 & \theta_t & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix} x_t + \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} f_t + w_t \tag{7.97a}$$

$$y_t = \left(\begin{array}{c|c} I_3 & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \end{array}\right) x_t + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} g_t + v_t \tag{7.97b}$$

where $x_t \in \mathbb{R}^4$ is the state of the system, $\theta_t \in \Theta$ stands for a parametric uncertainty, $y_t \in \mathbb{R}^3$ is the output vector, $f_t,\ g_t \in \mathbb{R}$ are potential faults to detect, and $w_t \in \mathbb{R}^4$, $v_t \in \mathbb{R}^3$ are the bounded noises affecting the system. The uncertainties are assumed to be slow-varying, meaning the residuals can be generated using the parity space approach, as described in Example 7.2.1, with

$$r(\theta,t) \triangleq N(\theta) Y_t \tag{7.98a}$$

$$r(\theta,t) = N(\theta)(V_t + G_t) + N(\theta)\Gamma(W_t + F_t) \tag{7.98b}$$

with for all $(Z,z) \in \{(Y,y),(W,w),(V,v),(F,f),(G,g)\}$, $Z_t = \begin{pmatrix} z_{t-1}^\top & z_t^\top \end{pmatrix}$, and:

$$\mathcal{O}(\theta) \triangleq \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & \theta & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \quad \Gamma \triangleq \left(\begin{array}{c|c} \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \hline \begin{matrix} I_3 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \end{array}\right), \quad N(\theta) = \begin{pmatrix} 1 & 1 & 0 & 1 & \theta & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \tag{7.99}$$

For all $\theta_1, \theta_2 \in \Theta$, $N(\theta_1)\mathcal{O}(\theta_2)x = \begin{pmatrix} (\theta_2 - \theta_1)x_3 & 0 \end{pmatrix}^\top \approx 0$ since x_3 is exponentially stable in the absence of faults and disturbances. In a rigorous setting, ultimate bounds should be determined for x_3 in the presence of the bounded noises w_t and v_t, with or without activity on the faults f_t and g_t. These ultimate bounds would allow $(\theta_2 - \theta_1)x_3$ to be considered as a new bounded noise affecting the first residual. However, to keep the illustrative example as uncomplicated as possible, this approximation error is simply ignored here. The noises are assumed to be bounded by $v_t \in \mathcal{E}_v$ and $w_t \in \mathcal{E}_w$, where $\mathcal{E}_v$ and $\mathcal{E}_w$ are the ellipsoids defined by the Minkowski functionals $\mu_{\mathcal{E}_v}(v_t) = \sqrt{v_t^\top P_v v_t}$ and $\mu_{\mathcal{E}_w}(w_t) = \sqrt{w_t^\top P_w w_t}$. The noise terms V_t and W_t are therefore contained in the convex bounded sets $\mathcal{V}$ and $\mathcal{W}$ defined by $\mu_{\mathcal{V}}(V_t) = \max\{\mu_{\mathcal{E}_v}(v_{t-1}), \mu_{\mathcal{E}_v}(v_t)\}$ and $\mu_{\mathcal{W}}(W_t) = \max\{\mu_{\mathcal{E}_w}(w_{t-1}), \mu_{\mathcal{E}_w}(w_t)\}$.

In order to detect the faults f_t and g_t, leveraging Theorem 7.3.1 (which, in that case, reduces to the usual expression for the linear transformation of an ellipsoid), the following upper- and lower-bounds are used as ellipsoidal inner- and outer-approximations respectively:

$$\frac{1}{\sqrt{2}}\sqrt{r^\top Q_v^{-1}(\theta) r} \leq \mu_{N(\theta)\mathcal{V}}(r) \leq \sqrt{r^\top Q_v^{-1}(\theta) r} \tag{7.100a}$$

$$\frac{1}{\sqrt{2}}\sqrt{r^\top Q_w^{-1}(\theta) r} \leq \mu_{N(\theta)\Gamma\mathcal{W}}(r) \leq \sqrt{r^\top Q_w^{-1}(\theta) r} \tag{7.100b}$$

with:

$$Q_v(\theta) = N(\theta) \begin{pmatrix} P_v^{-1} & 0 \\ 0 & P_v^{-1} \end{pmatrix} N^\top(\theta), \qquad Q_w(\theta) = N(\theta)\Gamma \begin{pmatrix} P_w^{-1} & 0 \\ 0 & P_w^{-1} \end{pmatrix} \Gamma^\top N^\top(\theta) \tag{7.101}$$

Finally, using the upper-bound of Property 7.3.4 and the minimum trace outer ellipsoidal approximation (which is a special case of Property 7.3.5 [119]) of the ellipsoids found above, the following Minkowski signals associated with the feasible set $\mathcal{R}_{\theta,\emptyset} \triangleq (N(\theta)\mathcal{V}) \oplus (N(\theta)\Gamma\mathcal{W})$ are obtained

$$\underline{\mu}_{\theta,\emptyset}(r) = \sqrt{r^\top Q_\emptyset^{-1} r} \tag{7.102a}$$

$$\overline{\mu}_{\theta,\emptyset}(r) = \left(1/\sqrt{r^\top Q_v^{-1}(\theta) r} + 1/\sqrt{r^\top Q_w^{-1}(\theta) r}\right)^{-1} \tag{7.102b}$$

with:

$$Q_\emptyset = 2\left(1 + \sqrt{\frac{\mathrm{Tr}(Q_w(\theta))}{\mathrm{Tr}(Q_v(\theta))}}\right) Q_v(\theta) + 2\left(1 + \sqrt{\frac{\mathrm{Tr}(Q_v(\theta))}{\mathrm{Tr}(Q_w(\theta))}}\right) Q_w(\theta) \tag{7.103}$$

and where the DIT to fault detection can be achieved by following Theorem 7.3.3 using the inequalities below.

$$\underline{\mu}_{\theta,\emptyset}(r) \leq \mu_{\theta,\emptyset}(r) \leq \overline{\mu}_{\theta,\emptyset}(r) \tag{7.104}$$

Similarly, in order to isolate the fault g_t from the set of faults $\{f, g\}$, it is assumed that the fault f_t is bounded in an interval $\mathcal{I}_f = [-\alpha_f, \alpha_f]$ whose Minkowski functional is given by $\mu_{\mathcal{I}_f}(f_t) = |f_t|/\alpha_f$. The fault term F_t is therefore contained in the bounded set $\mathcal{F}$ defined by its Minkowski functional $\mu_{\mathcal{F}}(F_t) = \max\{\mu_{\mathcal{I}_f}(f_{t-1}), \mu_{\mathcal{I}_f}(f_t)\}$. The following upper- and lower-bounds are obtained:

$$\frac{1}{\sqrt{2}}\sqrt{r^\top Q_f^{-1}(\theta) r} \leq \mu_{N(\theta)\Gamma\mathcal{F}}(r) \leq \sqrt{r^\top Q_f^{-1}(\theta) r} \tag{7.105}$$

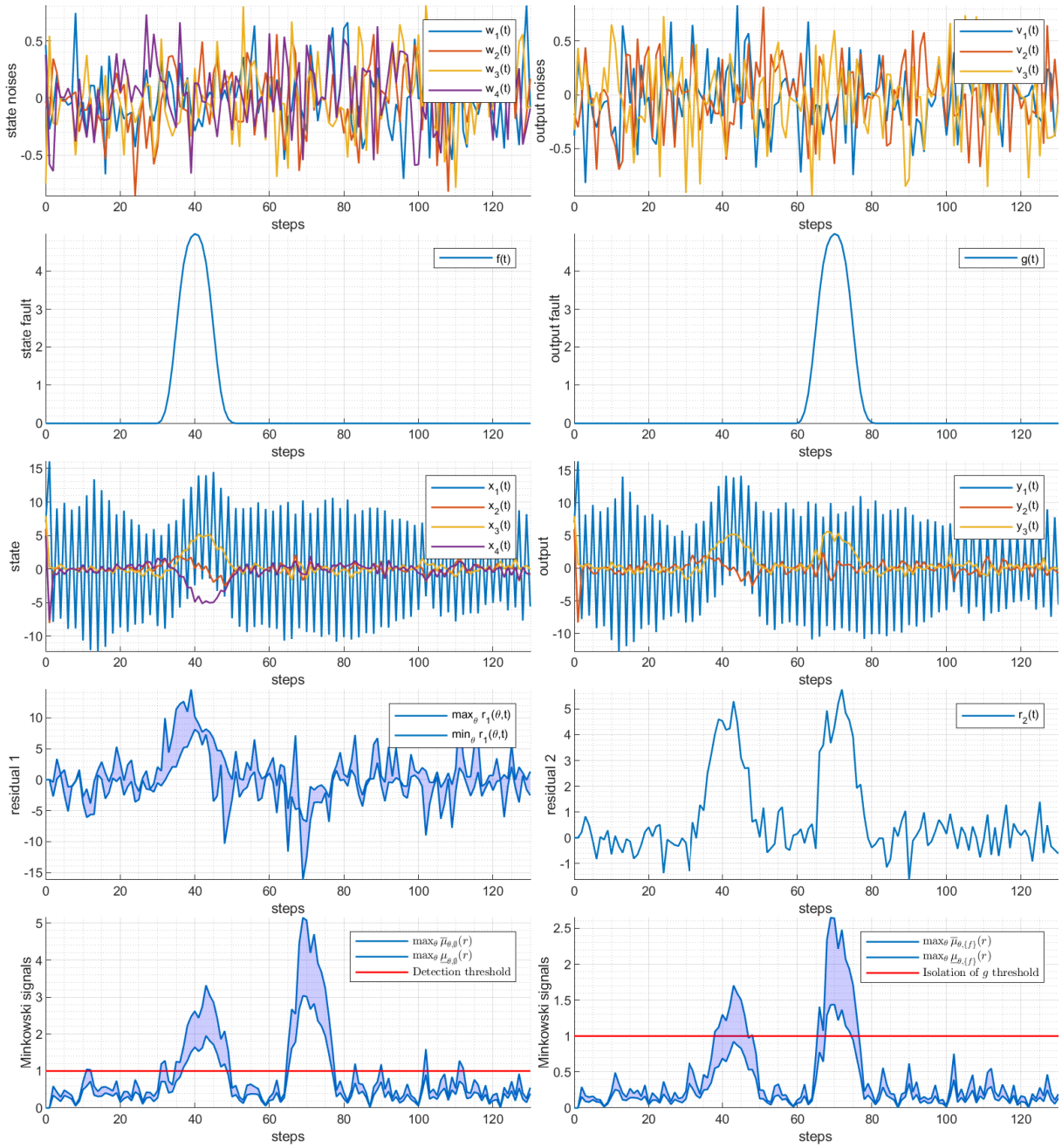


Figure 7.7: Application of the parity space approach described in Example 7.2.1 to the system (7.97), and computation of the Minkowski signals to perform the direct image described in Theorem 7.3.3.

with $Q_f(\theta) = \alpha_f^2 N(\theta)\Gamma\Gamma^\top N^\top(\theta)$. Using once more the upper-bound of Property 7.3.4 and the minimum trace outer ellipsoidal approximation (which is a special case of Property 7.3.5 [119]) of the ellipsoids found above, the feasible set $\mathcal{R}_{\theta,\{f\}} = (N(\theta)\mathcal{V}) \oplus (N(\theta)\Gamma\mathcal{W}) \oplus (N(\theta)\Gamma\mathcal{F})$ is approximated by:

$$\underline{\mu}_{\theta,\{f\}}(r) = \sqrt{r^\top \left(\left(1 + \sqrt{\frac{2\,\mathrm{Tr}(Q_f(\theta))}{\mathrm{Tr}(Q_\emptyset(\theta))}} \right) Q_\emptyset(\theta) + 2\left(1 + \sqrt{\frac{\mathrm{Tr}(Q_\emptyset(\theta))}{2\,\mathrm{Tr}(Q_f(\theta))}} \right) Q_f(\theta) \right)^{-1} r} \tag{7.106a}$$

$$\overline{\mu}_{\theta,\{f\}}(r) = \left(1/\sqrt{r^\top Q_v^{-1}(\theta) r} + 1/\sqrt{r^\top Q_w^{-1}(\theta) r} + 1/\sqrt{r^\top Q_f^{-1}(\theta) r} \right)^{-1} \tag{7.106b}$$

The DIT to isolate g_t can be achieved by following Theorem 7.3.3 using the inequalities below.

$$\underline{\mu}_{\theta,\{f\}}(r) \leq \mu_{\theta,\{f\}}(r) \leq \overline{\mu}_{\theta,\{f\}}(r) \tag{7.107}$$

The system is now simulated numerically with the initial conditions $x_0 = [7; 6; 8; 5]^\top$, using $\theta(t) = 3.5 - 2.5\sin(t/50)$ to model the uncertainties (hence $\Theta = [1, 6]$), and $P_w = I_4$, $P_v = I_3$ to bound the noises. The fault f_t is made active for $t \in [\![30, 50]\!]$, and the fault g_t is made active for $t \in [\![60, 80]\!]$. Moreover, $\alpha_f = 5$ is set as the maximum fault magnitude for f_t. For simplicity sake, the Minkowski signals are evaluated for θ taken in an equally-spaced grid of singletons $\tilde{\Theta}$ covering the uncertainty interval Θ. Results are presented on Figure 7.7. It is clear that both faults f_t and g_t are detected by the Minkowski signal $\max_{\theta\in\tilde{\Theta}} \underline{\mu}_{\theta,\emptyset}(r)$ for $t \in [\![37, 48]\!] \cup [\![66, 76]\!]$, and the fault g_t is also correctly isolated by the Minkowski signal $\max_{\theta\in\tilde{\Theta}} \underline{\mu}_{\theta,\{f\}}(r)$ for $t \in [\![68, 74]\!]$. These statements can be respectively observed at the bottom left and right of Figure 7.7. As stated in Remark 7.3.8 of Theorem 7.3.3, only the lower-bounding signals $\underline{\mu}_{\theta,\emptyset}(r)$ and $\underline{\mu}_{\theta,\{f\}}(r)$ are able to avoid false detection. The upper-bounding signals $\overline{\mu}_{\theta,\emptyset}(r)$ and $\overline{\mu}_{\theta,\{f\}}(r)$ are not to be seen as fault detection or isolation indicators, but they provide a guarantee that no fault is occurring when their amplitude is below 1. In particular, no conclusion can be stated on the faults' activity when $\max_{\theta\in\tilde{\Theta}} \underline{\mu}_{\theta,\emptyset}(r) < 1 < \max_{\theta\in\tilde{\Theta}} \overline{\mu}_{\theta,\emptyset}(r)$. Likewise, f_t is neither guaranteed to be isolated nor active when $\max_{\theta\in\tilde{\Theta}} \underline{\mu}_{\theta,\{f\}}(r) < 1 < \max_{\theta\in\tilde{\Theta}} \overline{\mu}_{\theta,\{f\}}(r)$, which is the case for $t \in [\![38, 46]\!]$. Again, this can be observed at the bottom right of Figure 7.7.

7.5 Conclusions and perspectives

This chapter first stated a generic set-theoretic FDI scheme to evaluate residuals following an uncertain linear internal structure. This scheme is obtained using results from order theory, which remain to be generalized to fault-trees for multi-component systems. Then, the Minkowski functional has been introduced as a novel and unified analytical thresholding tool for model-based diagnosis. Using the properties of the Minkowksi functional, including two characterizations of linear transformations of smooth convex sets, the threshold computation problem of the previously introduced fault isolation scheme has been stated in an analytical way. This led to the introduction of Minkowski signals, an intuitive measurement of the extend to which a set-membership relation is verified. An analytical expression of the minimal fault magnitude guaranteeing fault isolation has also been provided. The fault isolation scheme described in the chapter has finally been illustrated by an academic example. Simplifying the expression of the Minkowski functional of a convex set, particularly after it has been subject to several sets operation, remains an open challenge which needs to be tackled in future works.

Chapter 8

Saturation and dead-zone modeling

This chapter introduces methods for modeling actuator and sensor faults like saturations, dead-zones, dead-bands, and hysteresis, providing theoretical guarantees on the system behavior. Actuator saturations are considered with a T-S framework approach, by using fewer local models than current literature with the help of the Minkowski functional. A unified model for dead-zones, dead-bands, and hysteresis is also suggested and studied for LTI systems.

8.1 Faulty behaviours of actuators

Faulty behaviours of actuators (and sensors), which includes saturations, dead-zones, dead-bands or hysteresis effects, are common constraints that significantly affect system performances and even asymptotic stability. This chapter suggests new modeling methodologies for these behaviours, thus leading to new results in the subsequent stability and stabilization analyzes of the faulty system.

Actuator saturation arises due to material limitations or safety requirements, making it impossible to apply unbounded control signals. This phenomenon introduces a nonlinearity in the system which can significantly reduce its closed-loop performance and potentially cause instability. Hence, the saturation has to be accounted for in the design of a controller. Two main strategies exist in the literature in order to design a controller subject to saturation [278].

- The first strategy involves a two-step process. First, a nominal controller is developed without considering the actuator saturation; then, an anti-windup compensator is designed to handle the saturation constraint. This compensator computes the discrepancy between the unsaturated and the saturated control signals and retroactively uses it to modify the pre-designed controller [306]. This approach is ordinarily adopted in linear settings, and only a few extensions exists for T-S systems (e.g. [154, 202, 176]).
- In the second strategy, the saturation constraints are incorporated right from the start of the controller design process. Various methods have been developed using this approach, including the set invariance framework, which ensures that any state trajectory starting within an invariant set remains bounded inside it, preventing states from exceeding known bounds [278]. However, guaranteeing a large domain of attraction with this approach is often obtained by imposing constraints on the feedback gains, which hinders the performances of the controllers designed to prevent saturation (e.g. [272, 58, 308]).

The T-S community usually deals with saturation using the set invariance framework with a polytopic representation of the nonlinearity introduced by the saturation, and merges it with gain scheduling (e.g. [58, 237, 311, 36, 308]). Unfortunately, in a typical T-S fashion, this leads to a large number of local models and generates optimization problems with a huge number of LMI conditions, limiting their applicability. In particular, given a T-S system with n_h local models and an input vector $u(t)$ of dimension n_u, the number of local models needed to represent $u(t)$ saturated outside of an orthotope is typically $2^{n_u} n_h$ ([58, 237, 311, 308] etc) or $3^{n_u} n_h$ ([36]). Moreover, these existing T-S techniques are not only difficult to grasp intuitively and difficult to compute, but also limited in their intrinsic modeling capabilities, as the methodology does not extend to cases where $u(t)$ is saturated outside of a non-orthotopic polytope.

Section 8.2 of this chapter, published as a conference paper [21], is dedicated to the mitigation of the disadvantages of the T-S approach to actuator saturation. This is achieved by obtaining a flexible representation which captures a broad class of actuator saturations. The representation discussed in this chapter leverages the Minkowski functional associated with the saturating set, and only demands $2n_h$ local models, no matter the geometry of the saturating set or the dimension of $u(t)$. This leads to a reduced number of LMI in the local stabilization conditions, hence simplifying and reducing the conservatism of the approach. Moreover, a heuristic method is given to increase the chances of obtaining a large guaranteed domain of attraction, while explicitly allowing saturation to happen (as in [36]). Despite these advantages, this method comes with a few drawbacks: there is no simple and generic solution to obtain a minimal size for the domain of attraction, and, in the orthotopic case, the lack of *component-wise decomposition* of the saturated input vector may lead to a smaller domain of attraction than previous approaches. Numerical simulations are given in order to point out the contributions of the proposed method, as well as to quantify its shortcomings.

In addition to actuator saturation, *local* faulty behaviours such as dead-zones, dead-bands, and hysteresis are also frequently observed in physical systems, including hydraulic servo-valves, electric servomotors, among other applications. Although this is less discussed in the literature, these faulty local behaviours can also be observed on sensors, such as on relative pressure sensors. These faults tend to deteriorate the controller and observer performances, sometimes even leading to instability of the whole system. These types of local behaviours have been well investigated in the control literature, with an emphasis on the dead-zone of actuators [276, 277]. In particular, two primary approaches are commonly used to mitigate the impact of dead-zones [104].

- The first method is to implement an active compensation control strategy on the actuators based on a dead-zone inverse [275]: this method is however limited since the inverse is discontinuous, and it does not exist if the dead-zone is in fact a dead-band, or if it affects a sensor.
- The second approach consists in modeling the dead-zone as a combination of a linear control input with a constant or time-varying gain and a bounded disturbance-like term. This disturbance-like term is typically treated as an uncertain nonlinearity and is handled using robust feedback mechanisms [310].

Similarly, the impact of hysteresis effects has been investigated by the control literature from a theoretical point of view. In particular, a lot of models of hysteresis have been developed over the years for control purposes, including, among others, the Preisach model [220], the Duhem

model [81] and its variations (e.g. the Bouc-Wen, LuGre, Dahl models...). The reader is referred to the review [120] for more details.

Section 8.3 of this chapter, published as a conference paper [23], extends the underlying idea of the second approach to dead-zone modeling, while allowing it to take into account a broader range of nonlinear local behaviours, including some of the previously mentioned hysteresis models. It does so by introducing a new unifying representation which encapsulates the nonlinearities discussed above, leading to ultimate bound guarantees both in the cases of nonlinear and linear systems. Even if the suggested framework is more conservative compared to pre-existing specialized methods, it nevertheless requires minimal knowledge about the local nonlinearities, and moreover establishes a duality between the actuators and sensors local uncertainties.

8.2 Modeling actuator saturations using T-S systems

This section presents a flexible T-S representation capturing a wide range of actuator saturations, utilizing the Minkowski functional (Definition 7.3.1 of Chapter 7) to limit the required local models to $2n_h$ regardless of the saturating set's geometry or input dimension. In Section 8.2.1 a simple and exact T-S representation of a saturating actuator is obtained. Section 8.2.2 provides a local stabilization technique for the previously introduced T-S model by employing a saturated PDC controller, and details a heuristic to maximize the guaranteed domain of attraction. In Section 8.2.3, practical insights on the proposed method are presented through numerical simulations. The simplifying assumptions 2.4.1, 2.4.2 and 2.4.3 of Chapter 2 are once again considered in all subsequent matters.

8.2.1 T-S modeling of a saturation

The T-S model (8.1) is presumed to represent a nonlinear system subject to actuator saturation outside of $\mathcal{S} \subset \mathbb{R}^{n_u}$, a convex set containing the origin in its interior ($0 \in \text{intr}(\mathcal{S})$).

$$\dot{x}(t) = \sum_{i=1}^{n_h} h_i(\theta) \left(A_i x(t) + B_i u_{\mathcal{S}}(t)\right) \tag{8.1}$$

In this model, $u_{\mathcal{S}}(t) \in \mathcal{S} \subset \mathbb{R}^{n_u}$ stands for the saturated input vector. The T-S model (8.1) can easily be rewritten as (8.2) with an unrestricted input vector $u(t) \in \mathbb{R}^{n_u}$, using the map $u \mapsto \text{act}(u)u$, guaranteeing a continuous mapping from $\mathbb{R}^{n_u}$ to $\mathcal{S}$:

$$\dot{x}(t) = \sum_{i=1}^{n_h} h_i(\theta) \left(A_i x(t) + \text{act}(u(t)) B_i u(t)\right) \tag{8.2}$$

with act : $\mathbb{R}^{n_u} \to (0, 1]$ the continuous scalar map defined using the Minkowski functional associated with the set $\mathcal{S}$ (Definition 7.3.1 of Chapter 7):

$$\text{act}(u) \triangleq \begin{cases} 1 & \text{if } \mu_{\mathcal{S}}(u) \leq 1 \\ 1/\mu_{\mathcal{S}}(u) & \text{if } \mu_{\mathcal{S}}(u) > 1 \end{cases} \tag{8.3}$$

Continuity is indeed obtained from the item 2 of Property 7.3.2. This rewriting is powerful, since it reduces all the nonlinearities induced by the saturation of $u(t)$ to a bounded scalar term pre-multiplying the input matrices B_i. From here, an exact representation of (8.2) is obtained.

Lemma 8.2.1 (Open-loop T-S rewriting). *Let $\tau \in [0,1)$ and $\mathcal{U}_\tau \triangleq \{u \in \mathbb{R}^{n_u} : act(u) \geq \tau\}$. For all $u \in \mathcal{U}_\tau$, the T-S model (8.4) is an exact representation of (8.2).*

$$\dot{x}(t) = \sum_{i=1}^{n_h} \sum_{k=1}^{2} h_i(\theta) h_k^\tau(u(t)) \left(A_i x(t) + B_{i,k} u(t)\right) \tag{8.4}$$

with $B_{i,1} = B_i$, $B_{i,2} = \tau B_i$ and:

$$h_k^\tau(u) \triangleq \begin{cases} (act(u) - \tau)/(1-\tau) & \text{if } k = 1 \\ (1 - act(u))/(1-\tau) & \text{if } k = 2 \end{cases} \tag{8.5}$$

Proof. Since act$(u) \in [\tau, 1]$, (8.5) follows immediately from the NLSA (see Chapter 3). □

Remark 8.2.1. *For all $\tau \in (0,1)$, (8.4) only represents (8.2) locally ($\mathcal{U}_\tau \subset \mathbb{R}^{n_u}$). For $\tau = 0$, (8.4) is a global representation of (8.2) ($\mathcal{U}_0 = \mathbb{R}^{n_u}$). However, if $\tau = 0$, then $B_{i,2} = \tau B_i = 0$, which can lead to a loss of controllability [36].*

8.2.2 Saturated PDC state feedback

This section investigates the local stabilization of the saturated T-S model (8.1) controlled by (8.6), a PDC state feedback employing the same activation functions as in (8.1).

$$\begin{aligned} u_{\mathcal{S}}(t) &= \text{act}(u(t))u(t) \\ \text{with } u(t) &= \sum_{i=1}^{n_h} h_i(\theta) K_i x(t) \end{aligned} \tag{8.6}$$

Remark 8.2.2. *In a similar context, the T-S literature has already considered a PDC feedback law which involves the activation functions $\{h_k^\tau\}_{k=1,2}$ [308]. However, this gives rise to a self-referential effect in which the control signal $u(t)$ depends on its own current value. Indeed, the simulations of [308] do not employ these activation functions.*

As a direct consequence of Lemma 8.2.1, injecting (8.6) within the T-S system (8.4) provides an exact representation of the closed-loop system [20].

Lemma 8.2.2 (Closed-loop T-S rewriting). *Let $\tau \in [0,1)$ and*

$$\mathcal{X}_\tau \triangleq \left\{ x \in \mathbb{R}^{n_x} : \mu_{\mathcal{S}} \left(\sum_{j=1}^{n_h} h_j K_j x \right) \leq \frac{1}{\tau} \right\} \tag{8.7}$$

(if $\tau = 0$, $\mathcal{X}_0 = \mathbb{R}^{n_x}$). For all $x \in \mathcal{X}_\tau$, the T-S model (8.8) is an exact representation of (8.1) taken with the control law (8.6).

$$\dot{x}(t) = \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} \sum_{k=1}^{2} h_i(\theta) h_j(\theta) h_k^\tau(u(t)) \left(A_i + B_{i,k} K_j\right) x(t) \tag{8.8}$$

Proof. It is a straightforward rewriting of the open-loop system found in Lemma 8.2.1. □

From this representation, the following local stabilization conditions can be obtained.

Theorem 8.2.1 (Local PDC stabilization conditions). *Let $\mathcal{E}_{\lambda^*}(P) \triangleq \{x \in \mathbb{R}^{n_x} : x^\top P x \leq \lambda^*\}$ denote the largest ellipsoid contained within the set $\mathcal{X}_\tau$. The T-S model (8.1) taken with the control law (8.6) is guaranteed to be exponentially stable on $\mathcal{E}_{\lambda^*}(P)$ if there exists $X \in \mathbb{S}_{n_x}(\mathbb{R})$ and $\{M_i\}_{1\leq i\leq n_h}$ with $M_i \in \mathbb{R}^{n_u\times n_x}$ for all $i \in [\![1, n_h]\!]$, such that the conditions (8.9) are satisfied.*

$$X \succ 0 \tag{8.9a}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \mathcal{H}\left(A_i X + B_i M_j\right) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{8.9b}$$

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \mathcal{H}\left(A_i X + \tau B_i M_j\right) \prec 0, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{8.9c}$$

The inequalities on the double convex sums can be transformed into regular LMI conditions for a fixed τ by leveraging the results of Section 2.2.2. The gain matrices K_i are retrieved with $K_i = M_i X^{-1}$. Moreover, the matrix P providing the QLF $V(x) = x^\top P x$ demonstrating the local exponential stability of the closed-loop can be obtained with $P = X^{-1}$.

Proof. Introducing the quadratic Lyapunov function $V(x) = x^\top P x$, the LMI (8.9b) and (8.9c) are obtained by applying the results of Theorem 2.4.1 of Chapter 2 to the closed loop system (8.8) resp. for $k = 1, 2$. □

Remark 8.2.3. *If $\tau = 0$, then $\mathcal{E}_{\lambda^*}(P) = \mathbb{R}^{n_x}$, additionally (8.9c) can have a solution for $\tau = 0$ only if the input-free system is already globally exponentially stable.*

Remark 8.2.4. *It is easy to adapt the LMI (8.9b) to impose a minimum decay rate or an $\mathcal{H}_\infty$ attenuation criterion when the controller is not saturating. Identical adaptations can be made to both the LMI (8.9b) and (8.9c) to impose similar (but typically less demanding) guarantees up to the saturation level $act(u) \geq \tau$.*

Finding the largest ellipsoid $\mathcal{E}_{\lambda^*}(P)$ contained inside $\mathcal{X}_\tau$ is a hard problem in general, hence optimizing the size of $\mathcal{E}_{\lambda^*}(P)$ through the conditions (8.9) is a difficult task which could be tackled using computationally heavy derivative free optimization methods. Moreover, the actual domain of attraction for a given set of gains $\{K_i\}_{1\leq i\leq n_h}$ can be much larger than $\mathcal{E}_{\lambda^*}(P)$ itself, hence this optimization problem would only maximize the *guaranteed* domain of attraction, and not necessarily the domain of attraction itself. However, our approach has provided a set of conditions (8.9) which are limited in their number, size, and which are extremely simple to grasp. This leads to a heuristic solution to the problem of maximizing the guaranteed domain of attraction. Three key observations are given below and schematically illustrated in Figure 8.1.

Observation 8.2.1. *Let $\mathfrak{T}_{(P,\{K_i\}_{1\leq i\leq n_h})}$ denote the interval of all scalar values of $\tau \in [0, 1)$ solving (8.9) for fixed Lyapunov function and set of gains $(P, \{K_i\}_{1\leq i\leq n_h})$. Of course, (8.7) guarantees that the minimal value τ of $\mathfrak{T}_{(P,\{K_i\}_{1\leq i\leq n_h})}$ provides the largest $\mathcal{X}_\tau$ set, which in turns maximizes the guaranteed domain of attraction $\mathcal{E}_{\lambda^*}(P)$.*

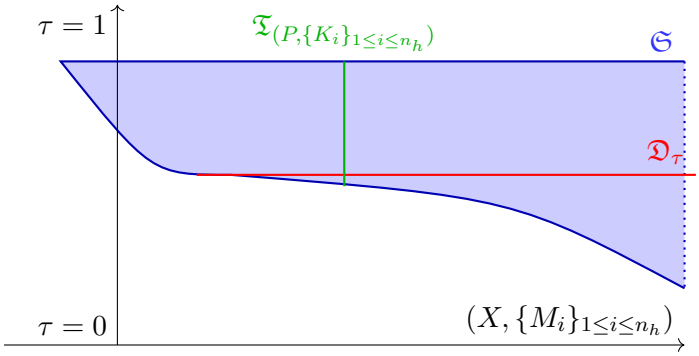


Figure 8.1: Schematic illustration of the sets $\mathfrak{T}_{(P,\{K_i\}_{1\leq i\leq n_h})}$, $\mathfrak{D}_\tau$ and $\mathfrak{S}$ described in the Observations 8.2.1, 8.2.2 and 8.2.3.

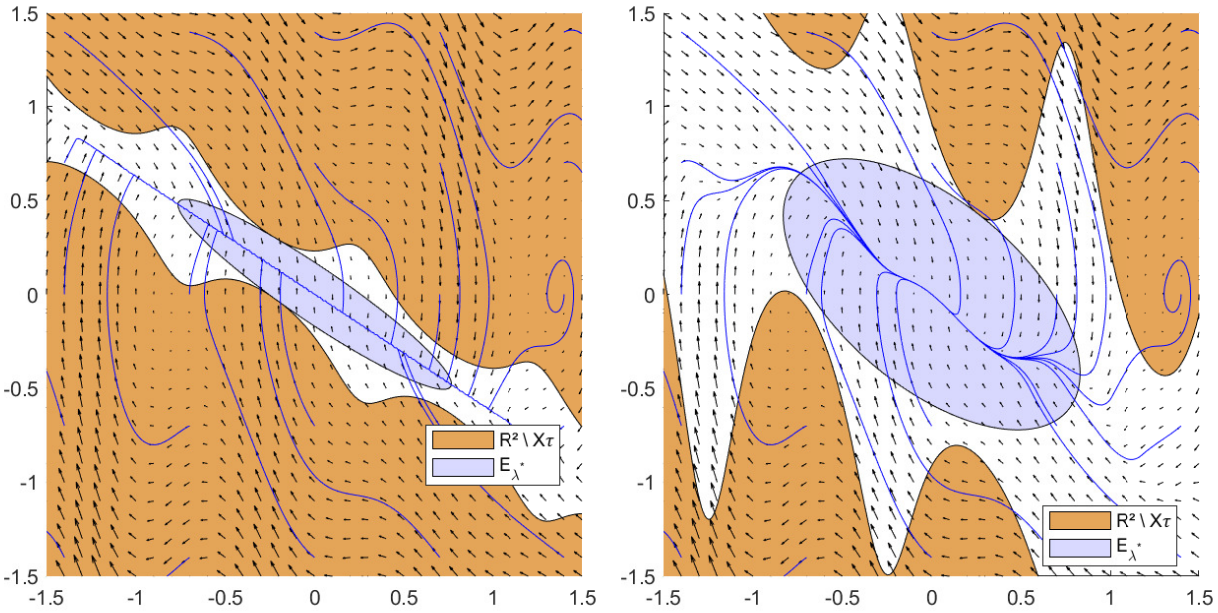
Observation 8.2.2. *Let $\mathfrak{D}_\tau$ denote the convex set of all decision variables $(X, \{M_i\}_{1\leq i\leq n_h})$ solving (8.9) for a fixed $\tau \in [0, 1)$. Well chosen weighted sums of (8.9c) and (8.9b) provide that for all $\tau_1, \tau_2 \in [0, 1)$ such that $\tau_1 \leq \tau_2$, then $\mathfrak{D}_{\tau_1} \subseteq \mathfrak{D}_{\tau_2}$, with the empty set being a subset of every possible set, including itself. Therefore, trying to minimize τ in (8.9) when $(P, \{K_i\}_{1\leq i\leq n_h})$ are not fixed does not guarantee a larger domain of attraction $\mathcal{E}_{\lambda^*}(P)$, but rather limits the size of the Lyapunov function and feedback gain set from which a solution is picked by the solver.*

Observation 8.2.3. *Let $\mathfrak{S}$ denote the set of all decision variables $(\tau, X, \{M_i\}_{1\leq i\leq n_h})$ solving (8.9). For all $(\tau, X, \{M_i\}_{1\leq i\leq n_h}) \in \mathfrak{S}$ and $\alpha \in \mathbb{R}_{>0}$ it is easily verified that $(\tau, \alpha X, \{\alpha M_i\}_{1\leq i\leq n_h}) \in \mathfrak{S}$. Hence, as long as $\mathfrak{S}$ is not empty, it is unbounded. Moreover, if there also exists $\beta \in (0, 1]$ such that $(\tau, X, \{M_i/\beta\}_{1\leq i\leq n_h})$ is a solution to (8.9b), then $(\beta\tau, X, \{M_i/\beta\}_{1\leq i\leq n_h}) \in \mathfrak{S}$. Intuitively, as long as the stabilizing gains can be scaled up in the unsaturated system, τ can be scaled down accordingly. It can be deduced from (8.7) and the positive homogeneity of the Minkowski functional (item 2 of Property 7.3.1) that this scaling has no effect on the size of $\mathcal{X}_\tau$, hence on the guaranteed domain of attraction $\mathcal{E}_{\lambda^*}(P)$.*

Observation 8.2.1 indicates that keeping P and $\{K_i\}_{1\leq i\leq n_h}$ relatively fixed gives relevance to the pursue of the minimization of τ as a means to maximize the guaranteed domain of attraction $\mathcal{E}_{\lambda^*}(P)$. However, Observation 8.2.2 underscores that providing excessive flexibility to these variables diminishes this relevance. This is partly explained by Observation 8.2.3, since scaling up the feedback gains can result in a reduced value of τ without any substantial impact on the guaranteed domain of attraction $\mathcal{E}_{\lambda^*}(P)$. Heuristically, it can therefore be conjectured that constraining further the optimization problem to limit this meaningless scaling effect should enhance the results obtained by minimizing τ. To this end, it is suggested to impose a *maximum* decay rate $\alpha/2 > 0$ to the unsaturated system, resulting in the following additional condition, which is added to (8.9):

$$\sum_{i=1}^{n_h}\sum_{j=1}^{n_h} h_i h_j \mathcal{H}\left(A_i X + B_i M_j\right) \succ -\alpha X, \quad \forall \mathbf{h} \in \Delta_{n_h-1} \tag{8.10}$$

The purpose of this heuristic is not to enforce a low decay rate: the objective is to ensure that the solver does not seek large gains *in order* to minimize τ. Large values of α can therefore be taken in order to not compromise the performances of the controller. In case of redundant inputs, this heuristic might need to be repeated on a selection of columns of B_i. The effectiveness of this



(a) Result with $\tau = 0.004$ and no α.

(b) Result with $\tau = 0.18$ and $\alpha = 15$.

Figure 8.2: Vector field associated with the closed-loop state-space, with the complement (in orange) of the exact region $\mathcal{X}_\tau$ (in white) and the guaranteed domain of attraction $\mathcal{E}_{\lambda^*}(P)$ (in light blue), with some trajectories (in blue) converging towards the origin for a set of evenly spread initial conditions.

heuristic is investigated numerically in the next section through the measure of the influence of τ on the volume of the guaranteed domain of attraction (8.11) (Figures 8.3 and 8.5)

$$\mathrm{Vol}(\mathcal{E}_{\lambda^*}(P)) = \frac{2\pi^{n_x/2}}{n_x \Gamma\left(n_x/2\right)} \sqrt{\det((P/\lambda^*)^{-1})} \tag{8.11}$$

where Γ stands for the usual gamma function [141].

8.2.3 Application

The LMI problems (8.9) and (8.10) are investigated using the relaxation scheme provided in Theorem 5.3.1 of Chapter 5 with $p = 3$.

Investigating the maximum decay rate heuristic. Consider the following nonlinear second-order system:

$$\ddot{y}(t) = \sin(2\pi y(t))(y(t) + \dot{y}(t)) + u_{\mathcal{S}}(t) \tag{8.12}$$

where the input $u_{\mathcal{S}}(t)$ is saturating outside the set $\mathcal{S} = [-1, 1]$ (which gives $\mu_{\mathcal{S}}(u) = |u|$). This system is rewritten as the exact T-S model (8.13) using Lemma 8.2.1:

$$\dot{x}(t) = \sum_{i=1}^{2} \sum_{k=1}^{2} h_i(\theta) h_k^\tau(u(t))(A_i x(t) + B_{i,k} u(t)) \tag{8.13}$$

with $x = \begin{pmatrix} x_1 & x_2 \end{pmatrix}^\top = \begin{pmatrix} y & \dot{y} \end{pmatrix}^\top$, $\theta = x_1$, $u_{\mathcal{S}} = \mathrm{act}(u)u$, and

$$\begin{aligned} A_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & B_{1,1} = B_{2,1} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ A_2 &= \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} & B_{1,2} = B_{2,2} &= \tau B_{1,1} \end{aligned} \tag{8.14}$$

and with the following activation functions:

$$h_1(x_1) = (1 + \sin(2\pi x_1))/2, \qquad h_2(x_1) = 1 - h_1(x_1) \tag{8.15}$$

In order to stabilize the system, a PDC state feedback law of the form (8.6) is calculated at several values of τ through the LMI (8.9) of Theorem 8.2.1, both with and without imposing a maximum decay rate $\alpha/2$ through (8.10). The closed loop state-space of (8.12) is plotted on Figures 8.2a and 8.2b with the *exact region* $\mathcal{X}_\tau$ and guaranteed region of attraction $\mathcal{E}_{\lambda^*}(P)$ of (8.13). In both cases, the blue trajectories show effective regions of attraction much larger that the guaranteed ones.

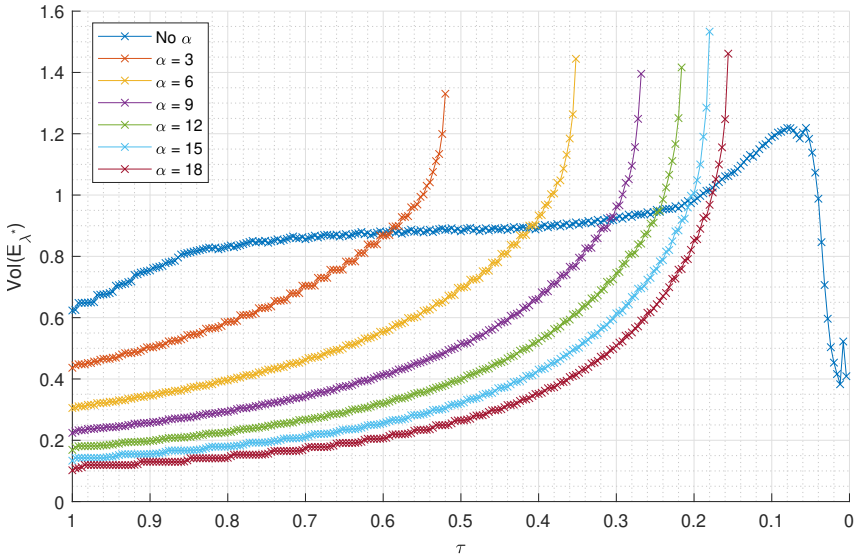


Figure 8.3: Volume of the guaranteed domain of attraction $\text{Vol}(\mathcal{E}_{\lambda^*}(P))$ depending on τ at several values of α.

Numerically, as long as no maximum decay rate is imposed through the LMI (8.10), it seems that τ can be chosen arbitrarily in $(0, 1)$ with an unclear effect on the volume of the guaranteed domain of attraction $\text{Vol}(\mathcal{E}_{\lambda^*}(P))$. However, adding a maximum decay rate unambiguously makes the minimization problem of τ relevant to obtain a large volume for the guaranteed domain of attraction, with no clear correlation between the choice of α and the largest value of $\text{Vol}(\mathcal{E}_{\lambda^*}(P))$ (Figure 8.3).

Investigating the number of local models. The T-S model (8.13) is modified so $u(t) \in \mathbb{R}^2$, $\mathcal{S} = [-1, 1]^2$, and

$$\begin{aligned} A_1 &= \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix} & B_{1,1} = B_{2,1} &= I_2 \\ A_2 &= \begin{pmatrix} 0 & 1 \\ -a & -b \end{pmatrix} & B_{1,2} = B_{2,2} &= \tau I_2 \end{aligned} \tag{8.16}$$

with $a, b \in \mathbb{R}$. Using the same LMI as previously, but imposing (8.10) on both columns of $B_{i,1}$ because of the input redundancy, the conservatism of computing a saturated PDC controller is investigated using:

- this section representation of $\mathcal{S}$ ($2r = 4$ local models),
- the usual literature representation ($2^2 r = 8$ local models).

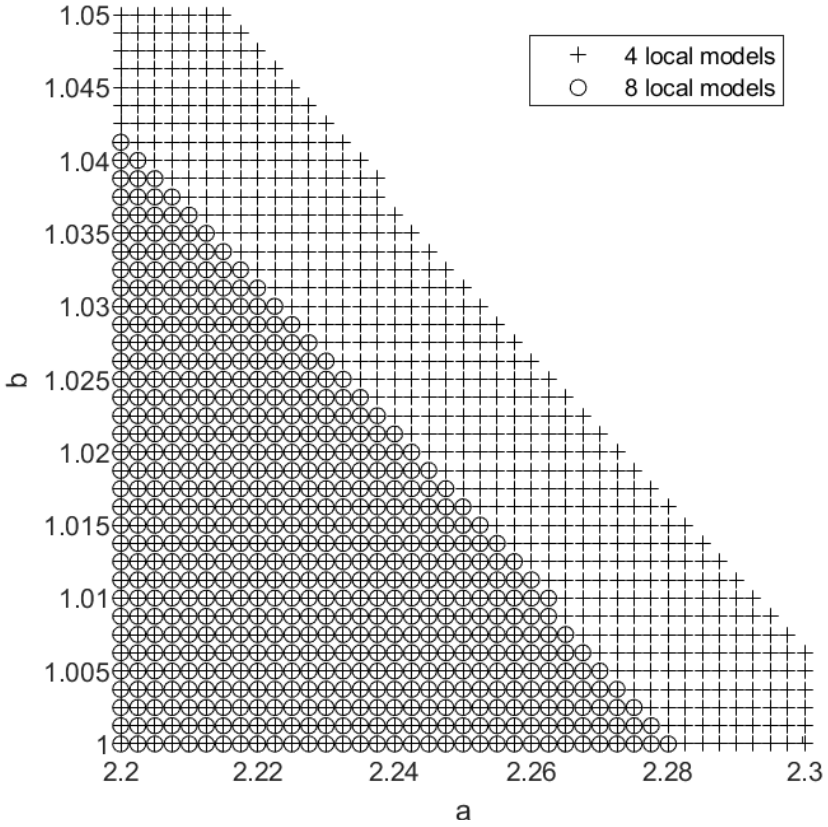


Figure 8.4: Stabilizability (a, b)-regions for $\tau = 0.11$ and $\alpha = 15$.

Fixing $(\tau, \alpha) = (0.11, 15)$, Figure 8.4 compares for several values of a, b the conservatism of both representations through the feasibility space of the LMI computing a saturated PDC controller. Unsurprisingly, the feasibility space is larger using the proposed representation with 4 local models than with 8. This was easily anticipated since the LMI problem with 4 local models is included in the LMI problem with 8 local models.

Fixing $(a, b) = (2.2, 1)$, Figure 8.5 investigates the heuristic of imposing (8.10) to obtain a large domain of attraction. Again, this heuristic tends to make the minimization of τ relevant to obtain a large volume for the guaranteed domain of attraction. However, the improvement is smaller for the T-S system with 8 local models than with 4 local models. It can be conjectured that the multiplication of local models naturally constraints the optimization problem, making the minimization of τ more effective on its own, without having to impose new LMI conditions. It is worth highlighting that the guaranteed domain of attraction is also smaller by using 4 models than 8 models, but this disadvantage needs to be put into perspective: this is

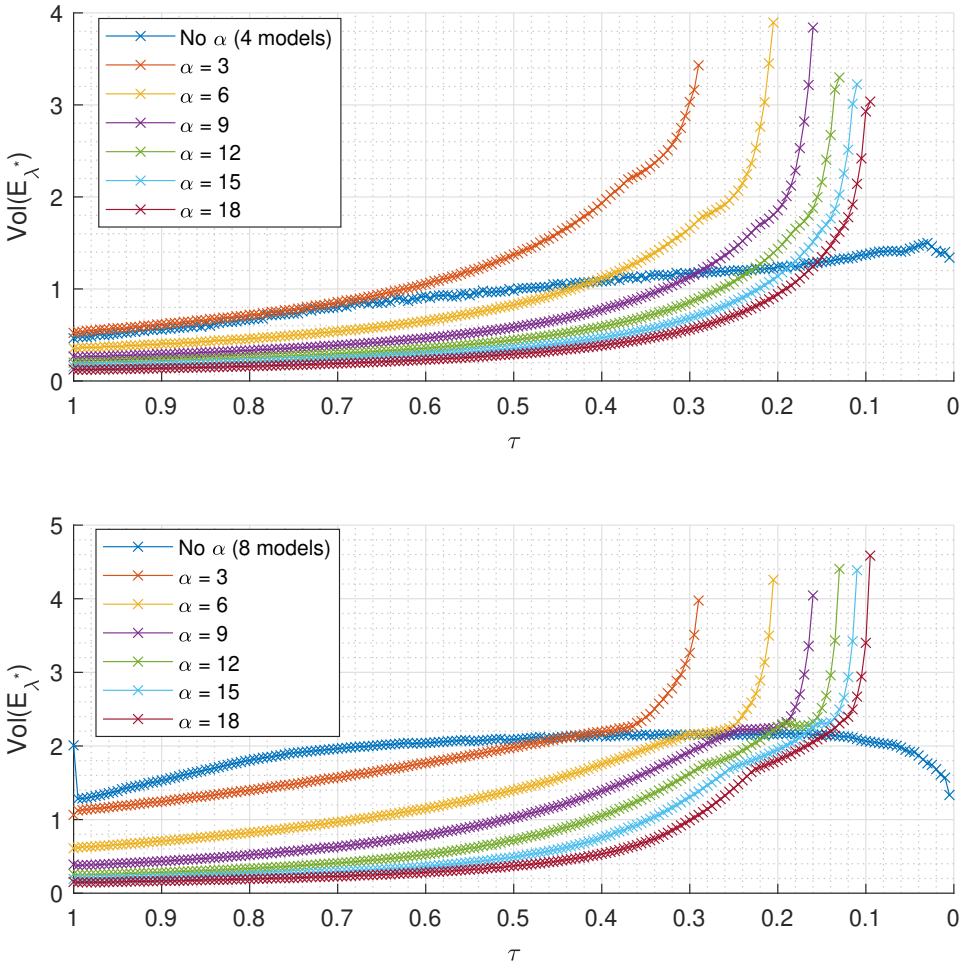


Figure 8.5: Volume of the guaranteed domain of attraction $\text{Vol}(\mathcal{E}_{\lambda^*}(P))$ depending on τ at several values of α, with $(a, b) = (2.2, 1)$.

not a comparison of the actual domain of attraction, but a comparison of an easily computed theoretical guarantee for the PDC controller.

8.3 A generalized model of actuators dead-zone, dead-band and hysteresis

In this section, a unifying representation which encapsulates local nonlinearities such as dead-zone, dead-band and hysteresis effects on actuators or sensors is presented. Ultimate bound guarantees for both nonlinear and linear systems are provided under this unified framework. While the suggested representation is more conservative than existing specialized methods, it requires minimal knowledge of the local nonlinearities. In Section 8.3.1, the unified modeling of faulty local nonlinear behaviours on actuators and sensors is introduced. Early ultimate bounds results for LTI systems are provided in Section 8.3.2, and these results are applied on a simple example in Section 8.3.3.

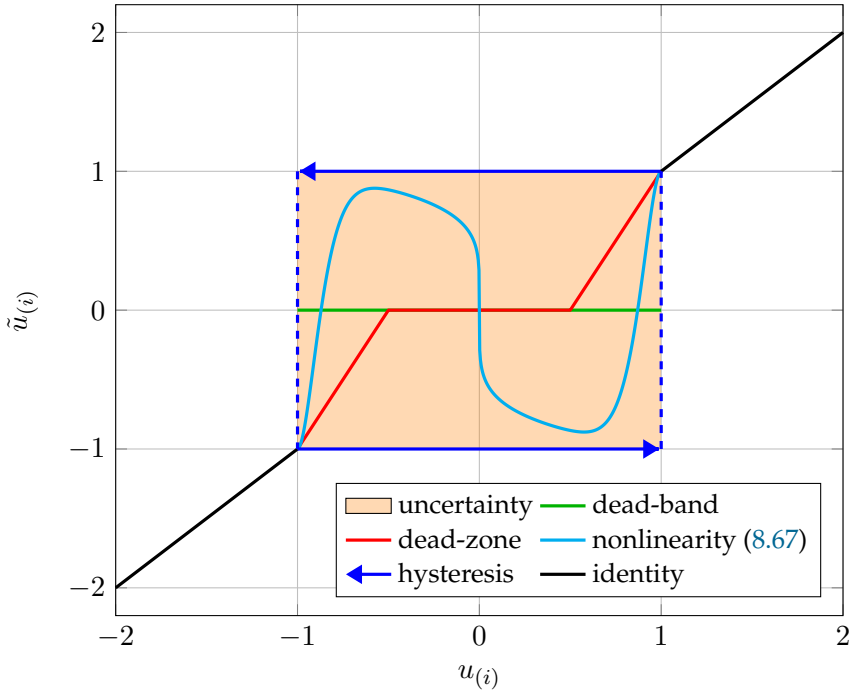


Figure 8.6: Nonlinear local behaviours of actuators and their suggested unified uncertain modeling, with $\overline{d} = -\underline{d} = 1$.

8.3.1 Unified modeling of actuators local nonlinearities

Consider the nonlinear system

$$\dot{x}(t) = f\left(x(t), \tilde{u}(t)\right) \tag{8.17a}$$
$$\tilde{u}(t) = h(u(t)) \tag{8.17b}$$

with $\tilde{u}(t) \in \mathbb{R}^{n_u}$ the effective control input of the system and $u(t) \in \mathbb{R}^{n_u}$ the reference signal given to the actuators. The function f is continuously differentiable and globally Lipschitz in $(x, \tilde{u})$, and h is a nonlinear function modeling the nonlinear faults affecting the actuators, such as their dead-zone, dead-band or hysteresis. In this section, instead of specifying exactly h, the author suggest a generic expression for h of the form

$$\tilde{u}(t) = h(u(t)) = u(t) + \delta(t) \tag{8.18}$$

where $\delta(t) \in \mathbb{R}^{n_u}$ is considered to be an unknown piecewise continuous signal. This allows h to model a wide range of nonlinear actuator faults without precise knowledge on them. The assumptions made on $\delta(t)$ differ depending on whether the actuator nonlinear faults affect $\tilde{u}$ component-wise or not. In both cases, two vectors $\overline{d}$ and $\underline{d}$ of $\mathbb{R}^{n_u}$ are defined such that $\underline{d}_{(i)} < 0 < \overline{d}_{(i)}$ for $i = 1, \ldots, n_u$.

8.3.1.1 Local component-wise nonlinearities

If each coordinate of the control input vector $\tilde{u}_{(i)}(t)$ is subject to a nonlinear distortion with respect to the nominal signal $u_{(i)}(t)$ for $u_{(i)}(t) \in [\underline{d}_{(i)}, \overline{d}_{(i)}]$, then h is nonlinear in the following

domain:

$$\mathcal{D}_u \triangleq \left\{ u \in \mathbb{R}^{n_u} : \exists i \in [\![1, n_u]\!] \text{ s.t. } u_{(i)} \in [\underline{d}_{(i)}, \overline{d}_{(i)}] \right\} \tag{8.19}$$

Outside of $\mathcal{D}_u$, h is considered to be the identity function. It is moreover assumed that this input distortion remains bounded according to the following property:

$$u_{(i)}(t) \in [\underline{d}_{(i)}, \overline{d}_{(i)}] \Rightarrow \tilde{u}_{(i)}(t) \in [\underline{d}_{(i)}, \overline{d}_{(i)}] \tag{8.20}$$

Usual candidates for h satisfying these properties include the dead-band:

$$[h(u)]_{(i)} = 0 \quad \text{if } u_{(i)} \in [\underline{d}_{(i)}, \overline{d}_{(i)}] \tag{8.21}$$

the continuous dead-zone:

$$[h(u)]_{(i)} = \begin{cases} \frac{\underline{d}_{(i)}}{\underline{d}_{(i)} - \underline{\varepsilon}_{(i)}} (u_{(i)} - \underline{\varepsilon}_{(i)}) & \text{if } \underline{d}_{(i)} \leq u_{(i)} < \underline{\varepsilon}_{(i)} \\ 0 & \text{if } \underline{\varepsilon}_{(i)} \leq u_{(i)} \leq \overline{\varepsilon}_{(i)} \\ \frac{\overline{d}_{(i)}}{\overline{d}_{(i)} - \overline{\varepsilon}_{(i)}} (u_{(i)} - \underline{\varepsilon}_{(i)}) & \text{if } \overline{\varepsilon}_{(i)} < u_{(i)} \leq \overline{d}_{(i)} \end{cases} \tag{8.22}$$

with $\underline{d}_{(i)} < \underline{\varepsilon}_{(i)} < 0 < \overline{\varepsilon}_{(i)} < \overline{d}_{(i)}$ for $i = 1, \ldots, n_u$, or even the Preisach hysteresis [220]:

$$[h(u)]_{(i)} = \text{“last value of } u_{(i)} \text{ outside of } [\underline{d}_{(i)}, \overline{d}_{(i)}]\text{”} \tag{8.23}$$

Instead of choosing one of the representation above, (8.18) is preferred, with the unknown signal δ satisfying:

$$\begin{cases} \delta_{(i)} \in [\underline{d}_{(i)} - u_{(i)}, \overline{d}_{(i)} - u_{(i)}] & \text{if } u_{(i)} \in [\underline{d}_{(i)}, \overline{d}_{(i)}] \\ \delta_{(i)} = 0 & \text{else} \end{cases} \tag{8.24}$$

It is easily verified that the dead-zone, dead-band and hysteresis described above can all be embedded in this representation (Figure 8.6). Moreover, it is also easily verified that δ is bounded inside a set D defined by:

$$D \triangleq [\underline{d}_{(1)} - \overline{d}_{(1)}, \overline{d}_{(1)} - \underline{d}_{(1)}] \times \cdots \times [\underline{d}_{(n_u)} - \overline{d}_{(n_u)}, \overline{d}_{(n_u)} - \underline{d}_{(n_u)}] \tag{8.25}$$

hence $\|\delta\|_2 \leq \left\|\overline{d} - \underline{d}\right\|_2$. Note that in the case of a dead-zone, this modeling can be leveraged after applying a first approximate smooth dead-zone inverse to the reference input signal, hence reducing the conservatism of the assumptions of this section.

8.3.1.2 Local nonlinearities of $\mathbb{R}^{n_u}$

Although very rarely discussed in the dead-zone literature, some local actuators faults may lead to a distortion of $\tilde{u}$ with respect to its nominal signal u which is restricted to a bounded set surrounding the origin of the input space $\mathbb{R}^{n_u}$. In that case, h can be considered nonlinear in an orthotopic domain $\mathcal{D}_u^{\square}$ defined by:

$$\mathcal{D}_u^{\square} \triangleq \left\{ u \in \mathbb{R}^{n_u} : \forall i \in [\![1, n_u]\!],\ u_{(i)} \in [\underline{d}_{(i)}, \overline{d}_{(i)}] \right\} \tag{8.26}$$

Outside of $\mathcal{D}_u^{\square}$, h is considered to be the identity function. Again, it is assumed that this input distortion remains bounded according to the following property:

$$u \in \mathcal{D}_u^{\square} \Rightarrow \tilde{u} \in \mathcal{D}_u^{\square} \tag{8.27}$$

These kind of local nonlinearities are typically found on actuators with several degree of freedom, e.g. radial dead-zones of joysticks [106, 78]. They are usually purposefully imposed to physical systems in order to avoid over-sensitivity of the system to small input values, but it is nonetheless useful to study their effect [54]. Moreover, even if the nonlinearities of the actuators are actually component-wise, the assumptions of this section can be leveraged as an intermediary step in order to diminish the overall conservatism of the study (see Corollary 8.3.1).

It is easily verified that $\delta \in D$ still holds true here, with D defined in (8.25). Moreover, a practical upper-bounding to $\|\delta\|_2$ in that case is provided by leveraging the following lemma.

Lemma 8.3.1. *Let $\mathbf{1}_S$ stand for the indicator function, defined by:*

$$\mathbf{1}_S(x) \triangleq \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{else} \end{cases} \tag{8.28}$$

For all $\beta, \eta > 0$, the following inequality holds:

$$\mathbf{1}_{\mathcal{B}_2(0,1)} \leq \exp\left(\beta(1 - \|\cdot\|_2^\eta)\right) \tag{8.29}$$

moreover the following limit is verified

$$\lim_{m\to\infty} \exp\left(\frac{1}{m}(1 - \|\cdot\|_2^m)\right) = \mathbf{1}_{\mathcal{B}_2(0,1)} \text{ pointwise} \tag{8.30}$$

but this convergence is not uniform.

Proof. Let $x \in \mathbb{R}^n$. If $x \notin \mathcal{B}_2(0,1)$, then $\exp\left(\beta(1 - \|x\|_2^\eta)\right) \geq 0$ is obvious. Moreover, $\|x\|_2 > 1$, hence $\lim_{m\to\infty} \frac{1}{m}(1 - \|x\|_2^m) = -\infty$, providing (8.30) for such x. If $x \in \mathcal{B}_2(0,1)$, then $\|x\|_2 \leq 1$ and $\exp\left(\beta(1 - \|x\|_2^\eta)\right) \geq e^0 = 1$. Moreover, $\lim_{m\to\infty} \frac{1}{m}(1 - \|x\|_2^m) = 0$, again providing (8.30) for such x. Finally, the convergence (8.30) cannot be uniform since $\exp\left(\frac{1}{m}(1 - \|\cdot\|_2^m)\right)$ is continuous on $\mathbb{R}^n$ for all m, whereas $\mathbf{1}_{\mathcal{B}_2(0,1)}$ is discontinuous. □

Indeed, the previous lemma can be applied in the context of a local nonlinearity as follows.

Lemma 8.3.2. *Given $\mathcal{E}(M^\top M)$ an ellipsoid centered at the origin and covering $\mathcal{D}_u^\square$, then for all $\beta, \eta > 0$, the following upper-bound holds:*

$$\|\delta(u)\|_2 \leq \left\|\overline{d} - \underline{d}\right\|_2 \exp\left(\beta\left(1 - \|Mu\|_2^\eta\right)\right) \tag{8.31}$$

Proof. For all $u \notin \mathcal{D}_u^\square$, $\|\delta(u)\|_2 = 0$, so (8.31) holds. For all $u \in \mathcal{D}_u^\square$,

$$\|\delta(u)\|_2 \leq \left\|\overline{d} - \underline{d}\right\|_2$$

The minimum-volume ellipsoid centered at the origin covering $\mathcal{D}_u^\square$ provides:

$$\|\delta(u)\|_2 \leq \left\|\overline{d} - \underline{d}\right\|_2 \mathbf{1}_{\mathcal{B}_2(0,1)}(Mu) \tag{8.32}$$

and (8.31) is obtained with Lemma 8.3.1. □

The ellipsoid $\mathcal{E}(M^\top M)$ covering $\mathcal{D}_u^\square$ can be found by solving the following optimization problem.

Lemma 8.3.3 (Minimum-volume covering ellipsoid [265]). *Given a set $\mathcal{S} \subset \mathbb{R}^n$, the minimum volume ellipsoid $\mathcal{E}(M^\top M)$ centered at the origin and covering $\mathcal{S}$ is found with the optimization problem (8.33).*

$$\begin{aligned} M = &\arg\min_{M \succ 0} (-\log\det M) \\ &\text{s.t. } (Mx)^\top Mx \leq 1, \qquad \text{for all } x \in \mathcal{S} \end{aligned} \tag{8.33}$$

In particular, this problem is convex when $\mathcal{S}$ is the convex hull of a finite set, as $\mathcal{S}$ can be reduced to this finite set.

8.3.1.3 A generic ultimate bound guarantee

The nonlinear system (8.17) can be rewritten by considering δ as a new unknown input

$$\dot{x}(t) = f_a(x(t), u(t), \delta(t)) \quad (= f\left(x(t), u(t) + \delta(t)\right)) \tag{8.34}$$

with f_a a continuously differentiable and globally Lipschitz function in (x, u, δ). Given a state feedback reference signal, the next result provides a simple condition to guarantee the ultimate boundedness of the closed-loop system, no matter if the nonlinearities of the actuators act on the control input component-wise or not.

Theorem 8.3.1 (Ultimate boundedness). *If there exists a state feedback law $u(t) = g(x(t))$ with g continuously differentiable and globally Lipschitz in x such that*

$$\dot{x}(t) = f_a(x(t), g(x(t)), 0) \tag{8.35}$$

has a globally exponentially stable equilibrium at $x = 0$, then for all bounded signal $\delta \in D$ (with D defined in (8.25)), there exists $r > 0$ such that for all initial state $x_0 \in \mathbb{R}^{n_x}$, the state trajectory x of (8.34) taken with $u(t) = g(x(t))$ is ultimately bounded by $\mathcal{B}_2(0, r)$.

Proof. From Lemma 4.5 page 108 of [150], the system (8.34) with $u(t) = g(x(t))$ is input-to-state stable in δ. In particular, by an application of Definition 4.4 page 107 of [150] for all initial state $x_0 \in \mathbb{R}^{n_x}$, the state trajectory x is ultimately bounded by

$$\gamma\left(\sup_{t_0 \leq \tau \leq t} \|\delta(\tau)\|_2\right) \leq \gamma\left(\sup_{\delta \in D} \|\delta\|_2\right) \leq \gamma\left(\left\|\overline{d} - \underline{d}\right\|_2\right) \tag{8.36}$$

with γ a class $\mathcal{K}$ function. In other words, there exists $r > 0$ such that for all x_0, there exists a $T > 0$ providing $x(t) \in \mathcal{B}_2(0, r)$ for all $t \geq t_0 + T$. □

8.3.1.4 Duality regarding sensors

Contrary to actuators, dead-zone of sensors are not theoretically invertible. Instead, the output $\tilde{y} \in \mathbb{R}^{n_y}$ of sensors subject to faults like dead-zone, dead-band or hysteresis becomes uncertain when measured in a specific range of values. By making similar assumptions as previously, a system subject to local nonlinear sensor faults may be written using an unknown signal δ:

$$\dot{x}(t) = f(x(t)) \tag{8.37a}$$
$$\tilde{y}(t) = h(y(t)) = y(t) + \delta(t), \quad \text{with } \delta \in D \tag{8.37b}$$

with $y \in \mathbb{R}^{n_y}$ the theoretical value that the sensors should measure. This symmetry of assumptions between the nonlinear actuator and sensor faults allowed by our simple modeling is not usual in the dead-zone literature, although it should provide ways to generalize the ultimate bounds discussed in the next section to the observer design problem, by duality between the linear state feedback control laws and the Luenberger observers.

8.3.2 Ultimate bounds for LTI systems

This section establishes ultimate bound guarantees for a stabilizable LTI system subject to local nonlinear actuator faults (8.38), taken with a linear state-feedback control $u(t) = Kx(t)$ imposing a decay-rate of $\alpha > 0$.

$$\dot{x}(t) = Ax(t) + B\tilde{u}(t) \tag{8.38a}$$
$$\tilde{u}(t) = h(u(t)) = u(t) + \delta(t), \quad \text{with } \delta \in D \tag{8.38b}$$

It is assumed that there exists a gain matrix K and a symmetric positive definite matrix $P \in \mathbb{S}_{n_x}^{++}(\mathbb{R})$ such that the following holds:

$$(A + BK)^\top P + P(A + BK) \preceq -2\alpha P \tag{8.39}$$

meaning the control law $u(t) = Kx(t)$ imposes a minimum decay rate $\alpha > 0$ to the nominal closed-loop system $\dot{x}(t) = (A + BK)x(t)$.

Theorem 8.3.2 (Generic ultimate bound). *System* (8.38) *associated with the control law* $u(t) = Kx(t)$ *such that* (8.39) *holds is ultimately bounded by the ellipsoid* $\mathcal{E}_r(P)$ *defined by*

$$\mathcal{E}_r(P) \triangleq \{x \in \mathbb{R}^n : x^\top P x \leq r\} \tag{8.40}$$

with

$$r = \frac{\lambda_{\max}(P)}{\alpha^2 \lambda_{\min}^2(P)} \|PB\|_2^2 \left\|\overline{d} - \underline{d}\right\|_2^2 \tag{8.41}$$

Proof. It is easily obtained from (8.39) that the derivative of the Lyapunov function $V(x) = x^\top P x$ along the trajectories of (8.38) with $u(t) = Kx(t)$ respects:

$$\begin{aligned} \dot{V}(x(t)) &\leq -2\alpha x^\top(t) P x(t) + 2x^\top(t) PB\delta(t) \\ &\leq -2\alpha\lambda_{\min}(P)\|x(t)\|_2^2 + 2\|x(t)\|_2 \|PB\|_2 \|\delta(t)\|_2 \end{aligned} \tag{8.42}$$

Thus, $\dot{V}(x) < 0$ is verified when

$$\|x\|_2 > \frac{\|PB\|_2\|\delta\|_2}{\alpha\lambda_{\min}(P)} \tag{8.43}$$

hence, the ultimate bound of the system can be computed as the following level-set:

$$V(x) = \lambda_{\max}(P)\left(\frac{\|PB\|_2\|\delta\|_2}{\alpha\lambda_{\min}(P)}\right)^2 \tag{8.44}$$

and the inequality $\|\delta\|_2 \leq \left\|\overline{d} - \underline{d}\right\|_2$ concludes the proof. □

The result discussed above holds whether the distortion of the input is component-wise or not. However, because it does not take into account the dependency between the uncertainties $\delta(t)$ and the system state $x(t)$, this ultimate bound is rather conservative. Other ultimate bounds are suggested below, both when the nonlinear fault affecting $\tilde{u}(t)$ is component-wise or not.

In order to deal with the component-wise case of Section 8.3.1.1, where h is nonlinear on $\mathcal{D}_u$ (with $\mathcal{D}_u$ defined in (8.19)), the set of indices (8.45) is introduced to list which coordinates of δ are active (i.e. not equal to zero) at a given state x.

$$\mathcal{I}(x) \triangleq \left\{ i \in [\![1, n_u]\!] : \underline{d}_{(i)} \leq (Kx)_{(i)} \leq \overline{d}_{(i)} \right\} \tag{8.45}$$

Moreover, for each non-empty subset $\mathcal{J} \subseteq [\![1, n_u]\!]$, the polytope (8.46) is defined under a half-space representation as follows:

$$\begin{aligned} \mathcal{P}_{\mathcal{J}} \triangleq \left\{ x \in \mathbb{R}^{n_x} : \begin{cases} \mathcal{J} \subseteq \mathcal{I}(x) \\ \|x\|_\infty \leq a\sqrt{\sum_{j\in\mathcal{J}}\left(\overline{d}_{(j)} - \underline{d}_{(j)}\right)^2} \end{cases} \right\} \\ \text{with } a = \frac{\|PB\|_2}{\alpha\lambda_{\min}(P)} \end{aligned} \tag{8.46}$$

note that $\mathcal{J} \subseteq \mathcal{I}(x)$ is verified for all $x \in \mathbb{R}^{n_x}$ such that:

$$\forall j \in \mathcal{J},\ \underline{d}_{(j)} \leq (Kx)_{(j)} \leq \overline{d}_{(j)} \tag{8.47}$$

Theorem 8.3.3 (Component-wise ultimate bound). *If $\delta \in D$ follows the assumptions of Section 8.3.1.1, with D defined in (8.25), system (8.38) associated with the control law $u(t) = Kx(t)$ such that (8.39) holds is ultimately bounded by $\mathcal{E}_r(P)$ with*

$$\begin{aligned} r = \max_{x\in\mathbb{R}^n}\ & x^\top Px \\ \text{s.t. } x \in & \bigcup_{\mathcal{J}\subseteq[\![1,n_u]\!]} \mathcal{P}_{\mathcal{J}} \end{aligned} \tag{8.48}$$

which can be computed by enumerating the vertices of all the polytopes $\mathcal{P}_{\mathcal{J}}$ defined in (8.46).

Proof. Similarly to the proof of Theorem 8.3.2, for $V(x) = x^\top Px$, $\dot{V}(x) < 0$ is verified when

$$\|x\|_2 > \frac{\|PB\|_2\|\delta\|_2}{\alpha\lambda_{\min}(P)} \tag{8.49}$$

in particular, if $\|x\|_\infty > a\|\delta\|_2$ then $\dot{V}(x) < 0$ holds, with a defined in (8.46). Since for all $i \notin \mathcal{I}(x)$, $\delta_{(i)} = 0$, it follows that if $\mathcal{I}(x) = \emptyset$ then $\|\delta\|_2 = 0$, and if $\mathcal{I}(x) \neq \emptyset$ then

$$\|\delta\|_2^2 = \sum_{i\in\mathcal{I}(x)} \delta_{(i)}^2 \leq \sum_{i\in\mathcal{I}(x)} \left(\overline{d}_{(i)} - \underline{d}_{(i)}\right)^2 \tag{8.50}$$

Reciprocally, if $\dot{V}(x) \geq 0$, then $\|x\|_\infty \leq a\|\delta\|_2$. It is shown by contradiction that $\mathcal{I}(x) \neq \emptyset$. If $\mathcal{I}(x) = \emptyset$, then $\|\delta\|_2 = 0$, hence $x = 0$, which also ensures $\mathcal{I}(x) \neq \emptyset$ by (8.45). So there exists a non-empty $\mathcal{J} \subseteq [\![1, n_u]\!]$ such that $\mathcal{J} \subseteq \mathcal{I}(x)$ and so $\|x\|_\infty \leq a\sqrt{\sum_{i\in\mathcal{J}}(\overline{d}_{(i)} - \underline{d}_{(i)})^2}$. This provides that $\dot{V}(x)$ can only be positive for x inside a polytope $\mathcal{P}_\mathcal{J}$, which is to say $\dot{V}(x) < 0$ for all $x \notin \bigcup_{\mathcal{J}\subseteq[\![1,n_u]\!]} \mathcal{P}_\mathcal{J}$. In the end, the ultimate bound of the system can be computed as the smallest level-set of V containing $\bigcup_{\mathcal{J}\subseteq[\![1,n_u]\!]} \mathcal{P}_\mathcal{J}$, i.e. through the optimization problem (8.48). □

A similar reasoning can be carried out when δ follows the assumptions of Section 8.3.1.2, where h is nonlinear on $\mathcal{D}_u^\square$ (8.26), by enumerating the vertices of the polytope $\mathcal{P}_{[\![1,n_u]\!]}$ directly.

Keeping the assumptions of Section 8.3.1.2, the next result leverages the upper-bound of Lemma 8.3.2 in order to obtain an ultimate bound through a simple semidefinite programming optimization problem. This optimization problem is a particular instance of the problem of finding an ellipsoid covering the intersection of several ellipsoids. This problem is however known not to be convex, and conservatism needs to be introduced in the analysis, as demonstrated in the following lemma.

Lemma 8.3.4 (Radius of the intersection of ellipsoids [122])**.** *Given $m+1$ (eventually degenerated) ellipsoids $\{\mathcal{E}_{r_i}(Q_i)\}_{0\leq i\leq m}$ with $Q_0 \in \mathbb{S}_n^{++}(\mathbb{R})$ and $Q_i \in \mathbb{S}_n^{+}(\mathbb{R})$ for all $i = 1, \ldots, m$, if a minimal radius $r_0 > 0$ exists such that*

$$\bigcap_{i=1}^{m} \mathcal{E}_{r_i}(Q_i) \subseteq \mathcal{E}_{r_0}(Q_0) \tag{8.51}$$

then r_0 is obtained by solving

$$\begin{aligned} r_0 = \max_{x\in\mathbb{R}^n}\ & x^\top x \\ \text{s.t. } & x^\top T_i x \leq r_i, \text{ for all } i = 1, \ldots, m \end{aligned} \tag{8.52}$$

with $T_i \triangleq Q_0^{-\frac{1}{2}\top} Q_i Q_0^{-\frac{1}{2}}$. However, this problem is not convex, and there are no efficient numerical procedures to solve it exactly. The following rank 1 dropping relaxation is used in order to upper-bound the minimal radius:

$$\begin{aligned} r_0 \leq \max_{X\in\mathbb{S}_n^+(\mathbb{R})}\ & \operatorname{Tr} X \\ \text{s.t. } & \operatorname{Tr}(T_i X) \leq r_i, \text{ for all } i = 1, \ldots, m \end{aligned} \tag{8.53}$$

Proof. The reader is referred to [122] for a review on the subject, including the proof of (8.53), among other convex relaxations of (8.52). □

The ultimate bound relying on the previous optimization problem is provided thereafter.

Theorem 8.3.4 (Non component-wise ultimate bound). *If $\delta \in D$ follows the assumptions of Section 8.3.1.2, with D defined in (8.25), system (8.38) associated with the control law $u(t) = Kx(t)$ such that (8.39) holds is ultimately bounded by $\mathcal{E}_r(P)$ with*

$$\begin{aligned} r = \max_{X \in \mathbb{S}^+_{n_x}(\mathbb{R})} & \operatorname{Tr} X \\ \text{s.t. } & \operatorname{Tr}(T_i X) \leq r_i, \text{ for } i = 1, 2 \end{aligned} \tag{8.54}$$

where

$$T_1 \triangleq P^{-\frac{1}{2}\top} R \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} R^\top P^{-\frac{1}{2}} \tag{8.55a}$$

$$T_2 \triangleq P^{-\frac{1}{2}\top} R \begin{pmatrix} 0 & 0 \\ 0 & I_{n_x - p} \end{pmatrix} R^\top P^{-\frac{1}{2}} \tag{8.55b}$$

$$r_1 \triangleq \frac{\left(W_0\left(\eta\beta\lambda_{\min}^{\eta/2}(\Lambda) r_2^{\eta/2}\right)/\eta\beta\right)^{\frac{2}{\eta}}}{\lambda_{\min}(\Lambda)} \tag{8.55c}$$

$$r_2 \triangleq \frac{\|PB\|_2^2}{\alpha^2 \lambda_{\min}^2(P)} \left\|\overline{d} - \underline{d}\right\|_2^2 e^{2\beta} \tag{8.55d}$$

where W_0 denotes the principal branch of the Lambert W function, R is a unitary matrix such that $\lambda_{\min}(\Lambda) > 0$ in

$$K^\top Q K = R\, diag(\lambda_1, \ldots, \lambda_p, 0, \ldots, 0) R^\top = R \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} R^\top \tag{8.56}$$

and $Q \triangleq M^\top M$, with M provided by Lemma 8.3.2.

Proof. Similarly to the proof of Theorem 8.3.2, for $V(x) = x^\top P x$, $\dot{V}(x) < 0$ is verified when

$$\|x\|_2 > \frac{\|PB\|_2 \|\delta\|_2}{\alpha \lambda_{\min}(P)} \tag{8.57}$$

leveraging the upper-bound (8.31) of Lemma 8.3.2, $\dot{V}(x) < 0$ stands for all x such that:

$$\|x\|_2 > \frac{\|PB\|_2}{\alpha \lambda_{\min}(P)} \left\|\overline{d} - \underline{d}\right\|_2 \exp\left(\beta\left(1 - (u^\top Q u)^{\frac{\eta}{2}}\right)\right) \tag{8.58}$$

i.e.

$$\|x\|_2 \exp\left(\beta (x^\top K^\top Q K x)^{\frac{\eta}{2}}\right) > \sqrt{r_2} \tag{8.59}$$

$K^\top Q K$ is real-symmetric hence unitary diagonalizable. We take $x = Rz = R \begin{pmatrix} z_1^\top & z_2^\top \end{pmatrix}^\top$, with R a unitary matrix and such that (8.56) holds. This provides $\dot{V}(x) < 0$ for all z such that:

$$\sqrt{(Rz)^\top Rz} \exp\left(\beta (z_1^\top \Lambda z_1)^{\frac{\eta}{2}}\right) > \sqrt{r_2} \tag{8.60}$$

note that since R is unitary, $R^\top R = I_{n_u}$ follows, hence $\dot{V}(x) < 0$ holds if:

$$\sqrt{(z_1^\top z_1 + z_2^\top z_2)\exp\left(2\beta(z_1^\top \Lambda z_1)^{\frac{\eta}{2}}\right)} > \sqrt{r_2} \tag{8.61}$$

which also holds, since $2\beta(z_1^\top \Lambda z_1)^{\frac{\eta}{2}} \geq 0$, if:

$$\sqrt{z_1^\top z_1 \exp\left(2\beta(z_1^\top \Lambda z_1)^{\frac{\eta}{2}}\right) + z_2^\top z_2} > \sqrt{r_2} \tag{8.62}$$

Finally $\dot{V}(x) < 0$ holds if:

$$\max\left\{\|z_1\|_2 \exp\left(\beta\lambda_{\min}^{\eta/2}(\Lambda)\|z_1\|_2^\eta\right), \|z_2\|_2\right\} > \sqrt{r_2} \tag{8.63}$$

In particular for z_1:

$$\begin{aligned} &\|z_1\|_2 \exp\left(\beta\lambda_{\min}^{\eta/2}(\Lambda)\|z_1\|_2^\eta\right) > \sqrt{r_2} \\ \Leftrightarrow\ &\|z_1\|_2 > \frac{\left(W_0\left(\eta\beta\lambda_{\min}^{\eta/2}(\Lambda) r_2^{\eta/2}\right)/\eta\beta\right)^{\frac{1}{\eta}}}{\lambda_{\min}^{1/2}(\Lambda)} = \sqrt{r_1} \end{aligned} \tag{8.64}$$

In the end, the ultimate bound of the system can be computed as the smallest level-set of V containing the intersection of the two degenerate ellipsoids $\|z_1\|_2^2 < r_1$ and $\|z_2\|_2^2 < r_2$, and Lemma 8.3.4 provides the tractable optimization problem (8.54) to obtain this level-set. □

This ultimate bound can actually be leveraged with a $\delta \in D$ following the assumptions of Section 8.3.1.1 if a sufficiently small ultimate bound is already known.

Corollary 8.3.1. *If $\delta \in D$ follows the assumptions of Section 8.3.1.1, system (8.38) associated with the control law $u(t) = Kx(t)$ such that (8.39) holds is ultimately bounded by $\mathcal{E}_{r_1}(P)$ with r_1 defined by (8.54) if there exists $r_2 > r_1$ such that $\mathcal{E}_{r_2}(P) \subset \mathcal{X}$ with $\mathcal{E}_{r_2}(P)$ an ultimate bound for the system and $\mathcal{X}$ defined by*

$$\mathcal{X} \triangleq \left\{x \in \mathbb{R}^{n_x} : Kx \in \mathcal{D}_u^{\square}\right\} \tag{8.65}$$

Proof. For all $x \in \mathcal{X}$, the assumptions on δ of Section 8.3.1.1 and of Section 8.3.1.2 are equivalent. Hence $\dot{V}(x) < 0$ holds for all $x \notin \mathcal{E}_{r_2}(P)$ and for all $x \in \mathcal{X} \setminus \mathcal{E}_{r_1}(P)$. Since $\mathcal{E}_{r_2}(P) \subset \mathcal{X}$, overall, $\dot{V}(x) < 0$ for all $x \notin \mathcal{E}_{r_1}(P)$, which concludes the proof. □

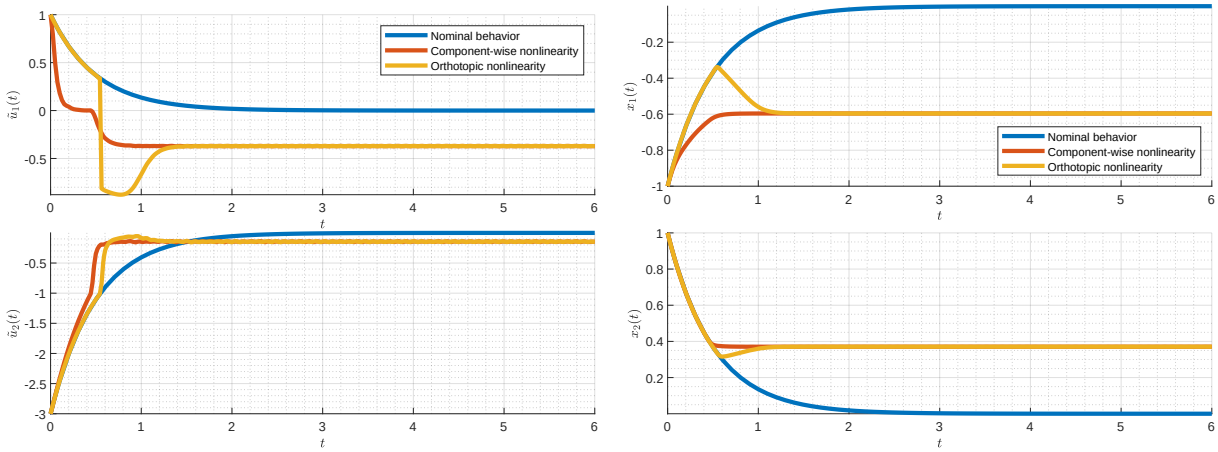
8.3.3 Application

The results of the previous section are applied to (8.38) taken with

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, \qquad B = I_2, \qquad K = \begin{pmatrix} -2 & -1 \\ -1 & -4 \end{pmatrix} \tag{8.66}$$

It is easily verified that (8.39) holds with $\alpha = 2$, $P = I_2$. Moreover, the vectors $\overline{d}, \underline{d} \in \mathbb{R}^2$ are defined by $\overline{d}_{(i)} = -\underline{d}_{(i)} = 1, i = 1, 2$. In simulation (Figures 8.7a, 8.7b), the nonlinear function h is taken such that (8.67) holds on $\mathcal{D}_u$ (defined in (8.19))and $\mathcal{D}_u^{\square}$ (defined in (8.26)) respectively (Figure 8.6).

$$\tilde{u}_{(i)} = h(u_{(i)}) = -\text{sign}(u_{(i)})|u_{(i)}|^{\frac{1}{5}}\cos(\pi u_{(i)}^5) \tag{8.67}$$



(a) Closed-loop input $\tilde{u}(t)$.

(b) Closed-loop state $x(t)$.

Figure 8.7: Input and state trajectories of the closed-loop system without input distortion (in blue), with component-wise input distortion (in orange), and with a local orthotopic input distortion (in yellow).

Component-wise nonlinearity A simple application of Theorem 8.3.2 provides

$$\|x\|_2 \leq \sqrt{2} = \text{UB}_1 \tag{8.68}$$

as a first ultimate bound of the system. Moreover, applying Theorem 8.3.3 with

$$\mathcal{P}_{\{1\}} = \left\{x \in \mathbb{R}^2 : \max\left\{|2x_1 + x_2|, \|x\|_\infty\right\} \leq 1\right\} \tag{8.69a}$$

$$\mathcal{P}_{\{2\}} = \left\{x \in \mathbb{R}^2 : \max\left\{|x_1 + 4x_2|, \|x\|_\infty\right\} \leq 1\right\} \tag{8.69b}$$

$$\mathcal{P}_{\{1,2\}} = \left\{ x \in \mathbb{R}^2 : \max \left\{ \begin{matrix} |2x_1 + x_2|, \\ |x_1 + 4x_2|, \\ \|x\|_\infty/\sqrt{2} \end{matrix} \right\} \leq 1 \right\} \tag{8.69c}$$

provides the norm of the vertex $\mathcal{V} = \begin{pmatrix} -1 & 1 \end{pmatrix}^\top$ as a valid ultimate bound of the system, hence $\|x\|_2 \leq \sqrt{2} = \text{UB}_2$. Despite taking into account the link between δ and x, the conservatism introduced by the infinity norm makes $\text{UB}_2 = \text{UB}_1$ for this particular system.

Orthotopic nonlinearity Assuming the system is ultimately bounded by $\mathcal{X}$, the set defined in (8.65), (which is verified when h is given by (8.67)), the ultimate bounds given below also hold in the component-wise case. Theorem 8.3.3 now provides the norm of the vertex $\mathcal{V} = \begin{pmatrix} 5/7 & -3/7 \end{pmatrix}^\top$ as a valid ultimate bound, hence $\|x\|_2 \leq \frac{6}{7} = \text{UB}_3$. Finally, applying Theorem 8.3.4 with

$$T_1 = I_2 \qquad T_2 = \emptyset \tag{8.70a}$$

$$r_1 = \frac{\left(W_0\left(\eta\beta(11/2 - 3\sqrt{2})^{\eta/2} r_2^\eta\right)/\eta\beta\right)^{\frac{2}{\eta}}}{11/2 - 3\sqrt{2}} \qquad r_2 = 2e^{2\beta} \tag{8.70b}$$

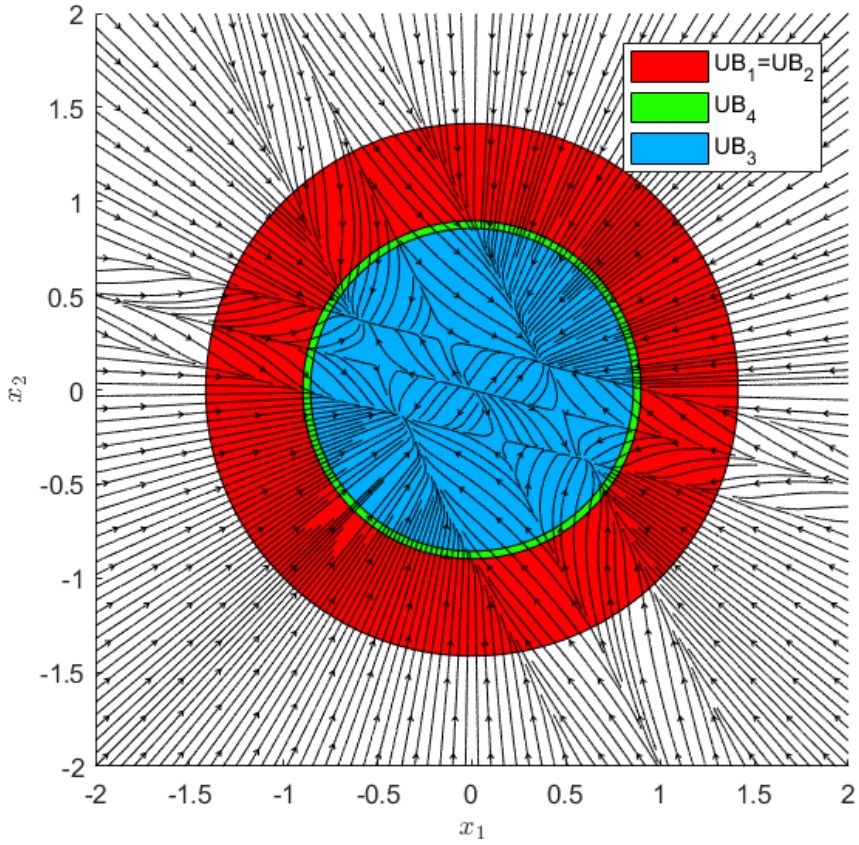


Figure 8.8: Ultimate-bounds over the phase space of the closed-loop taken with component-wise nonlinearities.

and

$$Q = \text{diag}\,(1/2, 1/2) \tag{8.71a}$$

$$K^\top Q K = \begin{pmatrix} 5/2 & 3 \\ 3 & 17/2 \end{pmatrix} \tag{8.71b}$$

$$\Gamma = \text{diag}\left(11/2 - 3\sqrt{2}, 11/2 + 3\sqrt{2}\right) \tag{8.71c}$$

$$R = \begin{pmatrix} -\frac{1+\sqrt{2}}{\sqrt{4+2\sqrt{2}}} & -\frac{1-\sqrt{2}}{\sqrt{4-2\sqrt{2}}} \\ \frac{1}{\sqrt{4+2\sqrt{2}}} & \frac{1}{\sqrt{4-2\sqrt{2}}} \end{pmatrix} \tag{8.71d}$$

provides the following ultimate bound for all $\beta, \eta > 0$:

$$\|x\|_2 \leq \frac{\left(W_0\left(\eta\beta(11/2 - 3\sqrt{2})^{\eta/2} 2^{\frac{\eta}{2}} e^{\eta\beta}\right)/\eta\beta\right)^{\frac{1}{\eta}}}{\sqrt{11/2 - 3\sqrt{2}}}$$

which, evaluated at $\beta = 10^{-3}$ and $\eta = 10^3$ yields $\|x\|_2 \leq 0.8973... = \text{UB}_4$. The ultimate bounds UB_1, UB_2, UB_3 and UB_4 can be visually compared over the phase space of the closed-looped system in Figure 8.8.

8.4 Conclusions and perspectives

In this chapter, some challenges posed by faulty behaviours of actuators were addressed, including saturations, dead-zones, dead-bands, and hysteresis effects. The proposed modeling methodologies offer innovative ways to capture and represent these nonlinearities. In particular, this chapter has presented a novel T-S representation of a large class of saturated actuators by leveraging the Minkowski functional associated with the saturating set. Contrary to previous works which solely considers orthotopic or ellipsoidal saturations, this representation is valid for all convex saturations. Moreover, very few local models are needed, which drastically reduces the number of LMI in the usual local stabilization conditions for a PDC state feedback law. Furthermore, a heuristic method to enhance the chances of achieving a large guaranteed domain of attraction is provided and numerically examined on two examples. This work can be generalized to other control law, such as the nPDC or the output feedback approaches. In the end, guaranteeing a large domain of attraction in the design of a saturated control for a nonlinear system remains a complex issue. A novel and unified approach was introduced in this chapter for addressing dead-zone, dead-band, hysteresis and others nonlinear local faults on actuators and sensors of physical system which requires minimal prior knowledge on these faults. Practical ultimate bounds are given for LTI systems affected by these nonlinear actuator faults. Similarly to the usual dead-zone literature, more robust bounds should be obtainable through adaptive control by leveraging a real-time estimate of the uncertain term introduced in the chapter, or by finding sufficient conditions to reduce the assumptions on δ of Section 8.3.1.1 to those of Section 8.3.1.2. These questions remain open for further investigations.

Chapter 9

Conclusions and perspectives

In this thesis, a large variety of control-theoretic results on nonlinear systems have been explored through the lens of T-S and LPV models, leveraging their (usually) convex properties in order to simplify the inherent challenges of nonlinear control. The T-S and LPV frameworks can generally be seen as refinements of the LDI framework, where information concerning which linear model is active at a time t is obtainable through a scheduling vector θ. This scheduling vector is generally considered measured or estimated in real-time. This approach allows for the generalization of linear techniques to systematically ensure the control or the observation of nonlinear systems, which would otherwise require a more difficult analysis. These techniques rely in particular on LMI conditions ensuring the stability or the stabilization of the system, or of the observation error.

The core contributions of this work is the introduction of geometric tools which were not previously employed in these contexts. For T-S systems, barycentric coordinates were shown to play a key role in the NLSA (*Chapters 3 and 4*), while polyhedral complexes were shown to be useful in order to approach non-convex T-S models (*Chapters 4*). Moreover, Bézier interpolations were found to be fundamental to the geometrical understanding of multi-sums (*Chapter 5*). For LPV systems, a Lipschitz assumption on the scheduling vector was shown, using Volterra's product integration and the weighted logarithmic norm, to bound all the obtainable state transition matrices in the future, leading to usable results in a near-future (*Chapter 6*). Additionally, a set-membership approach was explored for fault detection purposes, incorporating the Minkowski functional of a set (*Chapter 7*). Finally, the modeling of actuators saturations as well as others local nonlinearities (including dead-zones) was also approached using geometrical tools, including the previously introduced Minkowski functional (*Chapter 8*).

Of course, while this thesis has laid a variety of theoretical foundations that were illustrated on several academic examples, the implications of the all the results and tools introduced in this thesis can still be further investigated. These perspectives are now presented on a chapter-by-chapter basis. In the following, a one diamond symbol (♢) indicates the perspectives where the author is confident that results can be easily achieved, while two diamonds (♢♢) denote the perspectives where the author has some preliminary ideas but is uncertain about their difficulty. Prospects considered particularly challenging, and for which an in-depth analysis is probably required, are marked with three diamonds (♢♢♢).

Chapter 2: Introduction to the Takagi-Sugeno framework

Assuming that the activation functions $\mathbf{h} \in \Delta_{n_h-1}$ are time-varying uncertainties, necessary and sufficient conditions of stability are obtainable for T-S models, but they are rarely discussed in the literature, hence:

♢♢♢ The question of obtaining an LMI formulation of the necessary and sufficient exponential stability conditions provided in the continuous-time case (Theorem 2.3.2) using a PQLF remains open.

♢♢♢ This question can be extended to the case where some Lipschitz assumptions on the activation functions $\mathbf{h}$ are added to the previous assumption.

Chapter 3: Convex modeling of Takagi-Sugeno systems

♢♢ Stability and stabilization results for T-S-like models where the discrete-sum is replaced with a continuous integral remain to be investigated. In particular, tractable LMI conditions have to be extracted from the following problem:

$$\int_{\mathcal{V}_1 \in \mathcal{S}_\Theta} \cdots \int_{\mathcal{V}_m \in \mathcal{S}_\Theta} h_{\mathcal{V}_1} \dots h_{\mathcal{V}_m} \Gamma_{\mathcal{V}_1,\dots,\mathcal{V}_m}(z) d\mathcal{S}_\Theta \dots d\mathcal{S}_\Theta \prec 0 \tag{9.1}$$

It should be possible to demonstrate that there exists a sufficiently precise discretization of the conditions (9.1) so this finite set of LMI conditions is equivalent to (9.1), in the spirit of [238].

♢♢♢ Efficient methodologies remain to be investigated in order to systematically obtain a bounding set to the nonlinearities of a system. This set should ideally be a simple polytope of relatively small size, with as few vertices as possible, leading to a T-S model with low intrinsic conservatism.

♢♢♢ In particular, obtaining a simple measurement of the intrinsic conservatism of a T-S model would help to discriminate all the possible bounding polytopes: such a metric has yet to be introduced, and needs to be both easy to compute, and more involved than a simple volume or diameter measurement of the bounding polytope. To obtain such a metric, it might be interesting to examine the geometric properties of the following Gramian polytopes

$$\mathcal{P}_Q \triangleq \text{hull}\left\{ \int_0^{+\infty} e^{tA_i^\top} Q e^{tA_i} dt : i = 1, \dots, n_h \right\} \subseteq \mathbb{S}_n^+(\mathbb{R}) \tag{9.2}$$

where $Q \in \mathbb{S}_n^+(\mathbb{R})$ in order to investigate the conservatism of the stability analysis, and where $Q \in \left\{B_i B_i^\top : i = 1, \dots, n_h\right\}$ or $Q \in \left\{C_i^\top C_i : i = 1, \dots, n_h\right\}$ in order to investigate the conservatism of controllers or observers design. If theoretical guarantees were difficult to obtain, a large set of examples should be investigated in order to validate the metric, at least *experimentally*.

Chapter 4: Non-convex modeling of Takagi-Sugeno systems

♢ The Assumption 4.3.4 should be replaced with an assumption that can easily be verified *a priori*. Conditions on the upper right Dini derivative of the Lyapunov function along the system trajectories should be investigated, as they might be sufficient in this context [69].

♢ The results of this chapter remain to be extended to other LMI stability and stabilization conditions from the T-S literature, e.g. by relying both on the piecewise-nQLF

$$V(x,\theta) = \sum_{i=1}^{n_v} v_i(\theta) x^\top \left(\sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) P_{i,j} \right)^{-1} x \tag{9.3}$$

and on the following piecewise-nPDC scheme:

$$K(\theta) = \sum_{i=1}^{n_v} v_i(\theta) \left(\sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) K_{i,j} \right) \left(\sum_{j=1}^{n_{h_i}} h_{i,j}(\theta) P_{i,j} \right)^{-1} \tag{9.4}$$

♢♢♢ Similarly to Chapter 3, efficient bounding methodologies remain to be investigated in order to obtain a sharp juxtaposition of simple polytopes bounding the nonlinearities of a nonlinear system.

♢♢♢ It is unclear how to quantify the intrinsic conservatism of this approach, as the trade-off between the number of polytopes and their relative size has an unclear effect on conservatism, and as the number of vertices shared by several polytopes also seems to influence conservatism.

Chapter 5: Bézier interpolations in the Takagi-Sugeno framework

♢ This Bézier interpolation framework can also be applied to most of the usual results of the T-S framework, especially those relying on the following generalized MQLF:

$$V(x,\theta) = x^\top \left(\sum_{\mathbf{k} \in \mathbb{N}_m^{n_h}} \mathcal{B}_{\mathbf{k}}(\mathbf{h}(\theta)) P_{\mathbf{k}} \right) x \tag{9.5}$$

♢ All results from this chapter can be extended to tensor-product models by using multi-multivariate Bernstein polynomials in the Bézier interpolation scheme:

$$\sum_{\mathbf{k}_1,\ldots,\mathbf{k}_p \in \mathbb{N}_{m_1}^{n_1} \times \cdots \times \mathbb{N}_{m_p}^{n_p}} \mathscr{B}_{\mathbf{k}_1,\ldots,\mathbf{k}_p}(\mathbf{h}_1,\ldots,\mathbf{h}_p) \Gamma_{\mathbf{k}_1,\ldots,\mathbf{k}_p} \tag{9.6}$$

where $\mathscr{B}_{\mathbf{k}_1,\ldots,\mathbf{k}_p}(\mathbf{h}_1,\ldots,\mathbf{h}_p) \triangleq \mathcal{B}_{\mathbf{k}_1}(\mathbf{h}_1) \ldots \mathcal{B}_{\mathbf{k}_p}(\mathbf{h}_p)$.

♢♢ A closed-form expression remains to be found in order to explicitly rewrite the Bézier interpolation of order m_1 of a Bézier interpolation of order m_2 as a Bézier interpolation of order $m_1 + m_2$.

$$\sum_{\mathbf{k}_1 \in \mathbb{N}_{m_1}^n} \mathcal{B}_{\mathbf{k}_1}(\mathbf{X}) \left(\sum_{\mathbf{k}_2 \in \mathbb{N}_{m_2}^n} \mathcal{B}_{\mathbf{k}_2}(\mathbf{X}) \Gamma_{\mathbf{k}_1,\mathbf{k}_2} \right) = \sum_{\mathbf{k} \in \mathbb{N}_{m_1+m_2}^n} \mathcal{B}_{\mathbf{k}}(\mathbf{X}) \ldots ? \ldots \tag{9.7}$$

♢♢♢ A systematic methodology could be developed in order to exactly rewrite a polynomial nonlinear system under the form of a Bézier-T-S model. This methodology should also be adapted to at least approximate any sufficiently smooth nonlinear system.

◇◇◇ Assuming that the activation functions $\mathbf{h} \in \Delta_{n_h-1}$ are time-varying uncertainties, necessary and sufficient LMI stability conditions of remain to be found for the Bézier-T-S models.

◇◇◇ The asymptotic capabilities of the BPDC and BnPDC schemes need to be investigated with regard to the approximations capabilities of Bézier interpolations, possibly leading to necessary and sufficient controllability conditions with respect to the initial nonlinear system.

Chapter 6: Anticipating the near future of an LPV system

◇ Applications of the suggested exact discretization remain to be investigated, e.g. in a sampled-data setting.

◇ A logarithmic norm with time-varying weights remains to be introduced and studied, such as:

$$\eta_{P,t}(A) \triangleq \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^\top \left(A^\top P(t) + P(t)A + \dot{P}(t) \right) x}{2x^\top P(t) x} \tag{9.8}$$

◇◇ The guaranteed bounds between the future state-transition matrix and an estimation of this matrix might be useful in model predictive control schemes.

◇◇◇ Given a piecewise continuous real matrix $A(t) \in \mathbb{R}^{n \times n}$, it would be interesting to find practical bounds to the following quantity:

$$\ldots ? \ldots \leq \mathrm{Tr} \left(\prod_{t_1}^{t_2} e^{A(t)dt} \right) \leq \ldots ? \ldots \tag{9.9}$$

Such bounds would be helpful in order to approximate the eigenvalues of $\prod_{t_1}^{t_2} e^{A(t)dt}$ for $A(t) \in \mathbb{R}^{2\times 2}$.

◇◇◇ It would also be interesting to find a practical series expansion to the expression $\prod_{t_1}^{t_2} e^{A(t)dt}$ for $A(t) \in \mathbb{R}^{2\times 2}$ without relying on nested integrals or on a time-ordering operator. If such an expression exists, the example found in [105] gives us reasons to think that it might involve hypergeometric functions. By triangularization, the problem reduces to obtaining an explicit solution to a scalar Riccati differential equation [287]. This equation is already well-studied, but does not seem to always admit an explicit solution.

◇◇◇ The near-future structural guarantees, namely near-future observability and near-future controllability, remain to be investigated in the context of practical gain-scheduled observers and controllers. For example, one can wonder under which circumstances a near-future controllability guarantee implies the existence of a gain scheduled state feedback ensuring the decrease of a certain Lyapunov function on a specific time interval. Also, can this be related to the robust controllability and observability criteria which are found in the literature ? e.g. [251, 261]...

Chapter 7: Fault-isolation using a set-membership approach

◇◇◇ The generalized direct image test suggested in this manuscript rely on a lattice structure which remains to be generalized in a context where a fault-tree / a fault-matrix is available for the system.

◇◇◇ Sharper bounds, especially practical *lower-bounds*, remain to be found for the Minkowski functional of a Minkowski sum of convex sets.

$$\ldots ? \ldots \leq \mu_{\oplus_{i=1}^{m} \mathcal{S}_i} \leq \ldots ? \ldots \tag{9.10}$$

◇◇◇ Obtaining more practical conditions to handle the linear transformation of generic convex sets remains an open problem.

◇◇◇ Set reduction techniques remain to be investigated under the Minkowski framework to avoid nested analytical expressions to evaluate Minkowski signals.

Chapter 8: Saturation and dead-zone modeling

Concerning saturations:

◇ The proposed approach on saturations can be extended to other control laws, such as the nPDC or output feedback approaches.

◇ Instead of rewriting the saturated input using a single scalar nonlinearity, it might be useful to only regroup some coordinates of the saturated input together, making it straightforward to obtain a T-S model with $2^m n_h$ local models, with m ranging from 1 (all the input coordinates are grouped together: this is the approach of the thesis) to n_u (none of the input coordinates are grouped together: this is the usual approach of the literature). Taking $\{P_1, \ldots, P_m\}$ a partition of $[\![1, n_u]\!]$, such decomposition can be written as:

$$\sum_{k=1}^{m} \text{act}_{\mathcal{S}_k}(u_{P_k}(t)) u_{P_k}(t) = \sum_{k_1=1}^{2} \cdots \sum_{k_m=1}^{2} h_{k_1}^{\tau_1}(u_{P_1}(t)) \ldots h_{k_m}^{\tau_m}(u_{P_m}(t)) u(t) \tag{9.11}$$

◇◇ The proposed approach on saturations remains to be investigated while imposing an *a priori* guarantee on the attraction domain. This might be non-trivial due to the proposed rewriting being only valid locally.

◇◇ Generally speaking the Minkowski functional associated with the saturating set remains to be investigated in the context of the usual anti-windup schemes of the literature.

◇◇◇ The heuristic solution to obtain a large guaranteed domain of attraction remains to be investigated formally. This might be difficult, as the relationship between the value of τ and the volume of the guaranteed domain of attraction does not seem, at least *experimentally* (a single solution was investigated for each value of τ), to be trivially monotonic.

Concerning the unified modeling of dead-zones, dead-bands and hysteresis:

◇◇ The ultimate bounds remain to be investigated for observers, as well as for T-S systems.

◇◇◇ Obtaining stabilization conditions guaranteeing that a small enough ultimate bound is respected remains an open problem of this thesis.

◇◇◇ Similarly to the usual dead-zone literature, more robust ultimate bounds might be obtainable through adaptive control by leveraging a real-time estimate of the uncertain term introduced in the dead-zone modeling.

◇◇◇ Sufficient conditions to reduce the assumptions on δ of Section 8.3.1.1 to those of Section 8.3.1.2 remain to be found.

Appendices

Appendix A

CQLF of second-order systems: a graphical criterion

A.1 Introduction

A very simple graphical criterion is proposed in this appendix, published as an article in [19], in order to obtain a CQLF $V(x) = x^\top P x$ with $P \in \mathbb{S}_2^{++}(\mathbb{R})$ to a set of second order LTI systems $\{A_i\}_{1\leq i\leq m}$ with real coefficients. Simply put, the criterion associates every Hurwitz (resp. Schur) 2×2 real matrix with the interior of an ellipse on a two-dimensional plane. If the intersection of all the ellipses associated to the set of Hurwitz (resp. Schur) matrices $\{A_i\}_{1\leq i\leq m}$ is non-empty, it can be stated without loss of generality that there exists a CQLF to this set. All existing CQLF to $\{A_i\}_{1\leq i\leq m}$ can actually be retrieved from this intersection.

CQLF have been widely studied in the context of continuous-time and discrete-time LDI, which are nonlinear systems whose trajectories are included at each instant in the convex-hull of a finite set of LTI systems, respectively defined by:

$$\dot{x}(t) \in \text{conv}\{A_i x(t) : i = 1, \dots, m\} \tag{A.1a}$$
$$x_{k+1} \in \text{conv}\{A_i x_k : i = 1, \dots, m\} \tag{A.1b}$$

We recall Theorem 1.1.6 of Chapter 1 for $n_x = 2$:

Theorem A.1.1 (Exponential stability of an LDI). *Let $\{A_i\}_{1\leq i\leq m}$ be a set of $\mathbb{R}^{2\times 2}$ matrices. The LDI (A.1a) (resp. (A.1b)) is globally exponentially stable if there exists a symmetric positive-definite matrix $P \in \mathbb{S}_2^{++}(\mathbb{R})$ such that for all i, the $A_i^\top P + P A_i \in \mathbb{S}_2^{--}(\mathbb{R})$ (resp. $A_i^\top P A_i - P \in \mathbb{S}_2^{--}(\mathbb{R})$).*

LDI systems (A.1a) and (A.1b) encapsulate many other classes of nonlinear systems, including T-S systems. Their ever-presence makes Theorem A.1.1 a simple yet extremely common result of the modern nonlinear control literature.

As it is usually the case for Lyapunov criterion restricted to QLF, Theorem A.1.1 only offers a sufficient condition to exponential stability, and it is well-known not to be a necessary condition. However, despite its apparent simplicity and its clear similarity to the Lyapunov lemma,

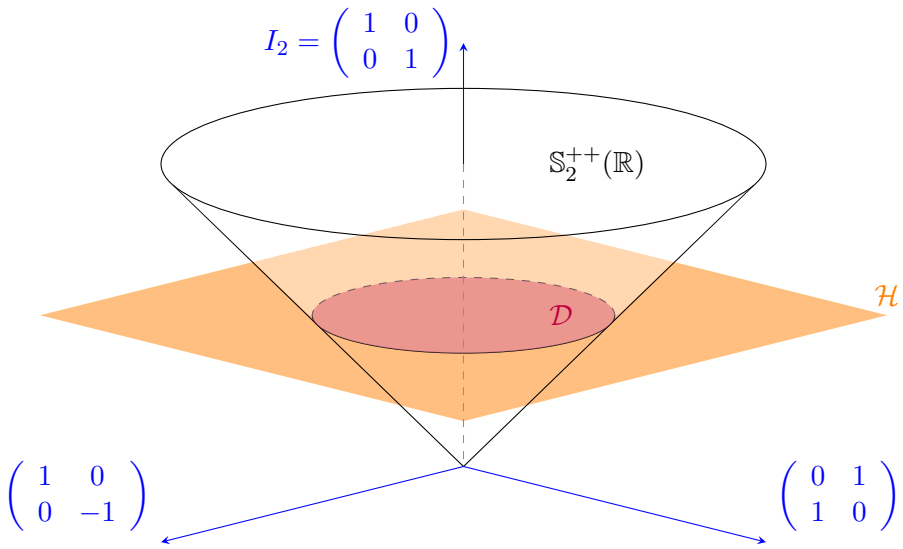


Figure A.1: Cone of positive-definite matrices $\mathbb{S}_2^{++}(\mathbb{R})$ in $\mathbb{S}_2(\mathbb{R})$, and its intersection with the affine hyperplane $\mathcal{H}$, the subspace of matrices with a trace of 2.

this result lacks a straightforward converse. Simply put, there are no elementary criterion to know if a given set of matrices $\{A_i\}_{1\leq i\leq m}$ is going to admit a CQLF or not. The problem is actually so hard, in both the continuous-time and the discrete-time cases, that existing results in the literature are limited to two-dimensional systems ($n = 2$) [255, 256, 203, 4], to sets made of two matrices ($m = 2$) [187], and to other restrictive conditions such as $m = 2$ and a rank one difference between the matrices [152], or both $n = 2$ and $m = 2$ [303, 254]. The necessary conditions proposed are moreover very often difficult to grasp intuitively.

To the author's knowledge, the strikingly uncomplicated graphical criterion discussed in this appendix has not yet been stated in such a simple manner in the literature so far, despite its practical interest. It should be highlighted that the graphical criterion is somewhat similar to the plots leading to the results found in Theorem 4.1 of [256]. However, the plots in [256] are constructed in a more convoluted fashion, and contrary to [256], every geometrical shape defined in this appendix is obtained through the exact same inequality, unifying the graphical criteria without requiring to check some preliminary assumptions on the matrices $\{A_i\}_{1\leq i\leq m}$. Typically, the suggested criterion also works on non-Hurwitz (resp. non-Schur) matrices, by associating them to an empty set or to an unbounded set which does not intersect any other ellipse given by a Hurwitz (resp. Schur) matrix. This suggested unification facilitates the numerical implementation of the graphical representation.

A.2 The ice cream cone

As illustrated by Figure A.1, Equation (A.3) of Lemma A.2.1 demonstrates that the set of positive-semidefinite matrices, denoted $\mathbb{S}_2^{+}(\mathbb{R})$, is a quadratic cone (also called a Lorentz cone, or an ice cream cone) in the space of symmetric 2×2 real matrices, oriented in the identity matrix I_2 direction [49]. Moreover, a strict inequality in (A.3) defines the set of positive-definite

matrices $\mathbb{S}_2^{++}(\mathbb{R})$ as the interior of $\mathbb{S}_2^{+}(\mathbb{R})$.

Lemma A.2.1 (The ice cream lemma). *Let $P \in \mathbb{S}_2(\mathbb{R})$ and $z_1, z_2, z_3 \in \mathbb{R}$ be such that:*

$$P = z_1 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + z_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + z_3 I_2 \tag{A.2}$$

The matrix $P \in \mathbb{S}_2^{+}(\mathbb{R})$ if and only if

$$\sqrt{z_1^2 + z_2^2} - z_3 \leq 0 \tag{A.3}$$

Moreover, $P \in \mathbb{S}_2^{++}(\mathbb{R})$ if and only if the inequality (A.3) is strict.

Proof. The matrix P is real-symmetric, hence both of its eigenvalues are real. Its smallest eigenvalue is:

$$\lambda_{\min}(P) = \mathrm{Tr}(P)/2 - \sqrt{\mathrm{Tr}^2(P)/4 - \det(P)} \tag{A.4}$$

It is easily verified that $\mathrm{Tr}(P) = 2z_3$ and $\det(P) = z_3^2 - z_1^2 - z_2^2$. Moreover, P is positive (semi)definite if and only if $\lambda_{\min}(P) > 0$ (resp. ≥ 0). Rewriting this condition with z_1, z_2 and z_3 directly yields (A.3). □

Remark A.2.1. *The cones $\mathbb{S}_2^{-}(\mathbb{R})$ and $\mathbb{S}_2^{--}(\mathbb{R})$ are symmetric to $\mathbb{S}_2^{+}(\mathbb{R})$ and $\mathbb{S}_2^{++}(\mathbb{R})$ with respect to the hyperplane $\mathrm{Tr}(\cdot) = 0$. Formally, $P \in \mathbb{S}_2^{-}$ (resp. $\mathbb{S}_2^{--}(\mathbb{R})$) if and only if*

$$\sqrt{z_1^2 + z_2^2} + z_3 \leq 0 \ \ (\textit{resp.} \ < 0) \tag{A.5}$$

Now let $\mathcal{H}$ denote the affine hyperplane of symmetric matrices with a trace of 2 (i.e. $z_3 = 1$):

$$\mathcal{H} \triangleq \{P \in \mathbb{S}_2(\mathbb{R}) : \mathrm{Tr}(P) = 2\} \tag{A.6}$$

The intersection between $\mathbb{S}_2^{++}(\mathbb{R})$ and $\mathcal{H}$ can be found using (A.3) to be an open disk of radius 1. In the $\mathcal{H}$ plane, this disk is denoted $\mathcal{D}$, and defined by:

$$\mathcal{D} \triangleq \left\{z \in \mathbb{R}^2 : z_1^2 + z_2^2 < 1\right\} \tag{A.7}$$

A.3 Graphical criterion

The main idea of the suggested graphical criterion consists in restricting the set of symmetric matrices $P \in \mathbb{S}_2(\mathbb{R})$ for which $A^\top P + PA$ (resp. $A^\top PA - P$) is negative-definite to the affine hyperplane $\mathcal{H}$. It will be noticed that this restriction is made without loss of generality from $\mathbb{S}_2^{++}(\mathbb{R})$ to $\mathcal{D}$, as a simple scaling of P allows to set its trace to 2 while preserving, by homogeneity, the negativeness of $A^\top P + PA$ (resp. $A^\top PA - P$). This ultimately reduces the problem of finding a CQLF to a set of matrices $\{A_i\}_{1 \leq i \leq m}$ to the geometrical problem of finding this common P in the open disk $\mathcal{D}$ at the intersection of $\mathbb{S}_2^{++}(\mathbb{R})$ and $\mathcal{H}$.

Given a matrix $A \in \mathbb{R}^{2\times 2}$ such that:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{A.8}$$

given the following linear operators:

$$\mathcal{L}_A^c(P) = A^\top P + PA, \quad \mathcal{L}_A^d(P) = A^\top PA - P \tag{A.9}$$

and given $P_z \in \mathcal{H}$ linearly defined for $z \in \mathbb{R}^2$ by:

$$P_z \triangleq z_1 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + z_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + I_2 \tag{A.10}$$

the following sets are now introduced, in the continuous-time case:

$$\mathcal{Q}_A \triangleq \{z \in \mathbb{R}^2 : \mathcal{L}_A^c(P_z) \in \mathbb{S}_2^{--}(\mathbb{R})\} \tag{A.11a}$$
$$= \{z \in \mathbb{R}^2 : \lambda_{\max}(\mathcal{L}_A^c(P_z)) < 0\} \tag{A.11b}$$
$$= \{z \in \mathbb{R}^2 : \sqrt{f_1^2(z) + f_2^2(z)} + f_3(z) < 0\} \tag{A.11c}$$

and in the discrete-time case:

$$\mathcal{R}_A \triangleq \{z \in \mathbb{R}^2 : \mathcal{L}_A^d(P_z) \in \mathbb{S}_2^{--}(\mathbb{R})\} \tag{A.12a}$$
$$= \{z \in \mathbb{R}^2 : \lambda_{\max}(\mathcal{L}_A^d(P_z)) < 0\} \tag{A.12b}$$
$$= \{z \in \mathbb{R}^2 : \sqrt{g_1^2(z) + g_2^2(z)} + g_3(z) < 0\} \tag{A.12c}$$

with:

$$f_1(z) \triangleq z_1(a_{12} - a_{21}) + z_2(a_{11} + a_{22}) + a_{12} + a_{21}$$
$$f_2(z) \triangleq z_1(a_{11} + a_{22)} + z_2(a_{21} - a_{12}) + a_{11} - a_{22}$$
$$f_3(z) \triangleq z_1(a_{11} - a_{22}) + z_2(a_{12} + a_{21}) + a_{11} + a_{22}$$

and:

$$g_1(z) \triangleq \frac{z_1}{2}(a_{21}^2 + a_{12}^2 - a_{11}^2 - a_{22}^2 + 2) + z_2(a_{22}a_{12} - a_{11}a_{21}) + (a_{22}^2 + a_{12}^2 - a_{21}^2 - a_{11}^2)/2$$
$$g_2(z) \triangleq z_1(a_{11}a_{12} - a_{21}a_{22}) + z_2(a_{11}a_{22} + a_{12}a_{21} - 1) + a_{11}a_{12} + a_{21}a_{22}$$
$$g_3(z) \triangleq \frac{z_1}{2}(a_{11}^2 + a_{12}^2 - a_{21}^2 - a_{22}^2) + z_2(a_{11}a_{21} + a_{12}a_{22}) + (a_{11}^2 + a_{12}^2 + a_{21}^2 + a_{22}^2)/2 - 1$$

Hereafter, $\mathcal{L}_A^{c,d}(P)$ stands for $\mathcal{L}_A^c(P)$ (resp. $\mathcal{L}_A^d(P)$).

Remark A.3.1. *By linearity of $\mathcal{L}_A^c(P_z)$ and $\mathcal{L}_A^d(P_z)$ with respect to $z \in \mathbb{R}^2$, and by convexity of $\mathbb{S}_2^{--}(\mathbb{R})$, $\mathcal{Q}_A$ and $\mathcal{R}_A$ are convex.*

Theorem A.3.1 (Graphical criterion)**.** *Given $\{A_i\}_{1\leq i\leq m}$ a set of $\mathbb{R}^{2\times 2}$ matrices, there exists a matrix $P \in \mathbb{S}_2^{++}(\mathbb{R})$ such that for all i, $\mathcal{L}_{A_i}^c(P) \in \mathbb{S}_2^{--}(\mathbb{R})$ if and only if*

$$\mathcal{Q}_{A_1} \cap \cdots \cap \mathcal{Q}_{A_m} \cap \mathcal{D} \neq \emptyset \tag{A.15}$$

Similarly, there exists a matrix $P \in \mathbb{S}_2^{++}(\mathbb{R})$ such that for all i, $\mathcal{L}_{A_i}^d(P) \in \mathbb{S}_2^{--}(\mathbb{R})$ if and only if

$$\mathcal{R}_{A_1} \cap \cdots \cap \mathcal{R}_{A_m} \cap \mathcal{D} \neq \emptyset \tag{A.16}$$

Proof. $\boxed{\Rightarrow}$ If there exists a $P \in \mathbb{S}_2^{++}(\mathbb{R})$ such that $\mathcal{L}_{A_i}^{c,d}(P) \in \mathbb{S}_2^{--}(\mathbb{R})$ for all i, then by homogeneity [48], $P' = 2P/\operatorname{Tr}(P) \in \mathcal{H} \cap \mathbb{S}_2^{++}(\mathbb{R})$, and $\mathcal{L}_{A_i}^{c,d}(P') \in \mathbb{S}_2^{--}(\mathbb{R})$ for all i as well. Moreover as $P' \in \mathcal{H}$ there exists $z \in \mathbb{R}^2$ such that $P_z = P'$. By applying Lemma A.2.1, since $P_z \in \mathbb{S}_2^{++}(\mathbb{R})$, then $z \in \mathcal{D}$, and since $\mathcal{L}_{A_i}^{c,d}(P_z) \in \mathbb{S}_2^{--}(\mathbb{R})$, then $z \in \mathcal{Q}_{A_i}$ (resp. $z \in \mathcal{R}_{A_i}$) for all i.

$\boxed{\Leftarrow}$ If there exists $z \in \mathcal{Q}_{A_1} \cap \cdots \cap \mathcal{D}$ (resp. $z \in \mathcal{R}_{A_1} \cap \cdots \cap \mathcal{D}$), then $P_z \in \mathbb{S}_2^{++}(\mathbb{R})$ such that $\mathcal{L}_{A_i}^{c,d}(P_z) \in \mathbb{S}_2^{--}(\mathbb{R})$ for all i. □

Remark A.3.2. *Helly's theorem states that given a finite collection of convex sets of $\mathbb{R}^d$, the intersection of this collection is non-empty if the intersection of each sub-collection with $d+1$ convex sets is non-empty. The sets of Theorem A.3.1 being convex and subsets of the subspace $\mathcal{H}$ with $\dim(\mathcal{H}) = 2$, their intersection can be checked by considering only three sets at a time [256].*

Remark A.3.3. *Similarly to how the trace of P can be fixed to a positive value without loss of generality in $\mathbb{S}_2^{++}(\mathbb{R})$, its determinant can also be fixed to a positive value without loss of generality in $\mathbb{S}_2^{++}(\mathbb{R})$. In fact, the proposed graphical criterion can also be interpreted inside of $\mathcal{D}$ as the Klein disk model of the sheet of the hyperboloid associated with $\det(\cdot) = 1$ contained in $\mathbb{S}_2^{++}(\mathbb{R})$ [57].*

Theorem A.3.2 (Geometry of the solutions). *$A \in \mathbb{R}^{2\times 2}$ is Hurwitz if and only if $\mathcal{Q}_A$ is the interior of an ellipse. If so, $\mathcal{Q}_A \subseteq \mathcal{D}$. Similarly, A is Schur if and only if $\mathcal{R}_A$ is the interior of an ellipse. If so, $\mathcal{R}_A \subseteq \mathcal{D}$.*

Proof. $\boxed{\Rightarrow}$ Let $A \in \mathbb{R}^{2\times 2}$ be a Hurwitz (resp. Schur) matrix, i.e. $\mathcal{Q}_A \cap \mathcal{D} \neq \emptyset$ (resp. $\mathcal{R}_A \cap \mathcal{D} \neq \emptyset$) according to Theorem A.3.1. It is first shown by contradiction that $\mathcal{Q}_A \subseteq \mathcal{D}$ (resp. $\mathcal{R}_A \subseteq \mathcal{D}$).

Assuming that there exists $z' \in \mathcal{D}$ and $z'' \notin \mathcal{D}$ such that $z', z'' \in \mathcal{Q}_A$ (resp. $\mathcal{R}_A$); by convexity, for all $t \in [0,1]$, $tz' + (1-t)z'' \in \mathcal{Q}_A$ (resp. $\mathcal{R}_A$). The norm of $tz' + (1-t)z''$ being continuous with respect to t, the intermediate value theorem provides $z^* \in \mathcal{Q}_A$ (resp. $z^* \in \mathcal{R}_A$) such that $\|z^*\|_2 = 1$. Its associated P_{z^*} belongs to $\mathbb{S}_2^{+}(\mathbb{R}) \setminus \mathbb{S}_2^{++}(\mathbb{R})$, so there exists $v \in \mathbb{R}^2 \setminus \{0\}$ a vector in the kernel of P_{z^*}. However, this yields $v^\top \mathcal{L}_A^c(P_{z^*})v = 0$ and $v^\top \mathcal{L}_A^d(P_{z^*})v = v^\top A^\top P_{z^*} A v$, where $A^\top P_{z^*} A \in \mathbb{S}_2^{+}(\mathbb{R})$ by congruence [48]. In both cases this is in contradiction with $\mathcal{L}_A^{c,d}(P_{z^*}) \in \mathbb{S}_2^{--}(\mathbb{R})$, i.e. with $z^* \in \mathcal{Q}_A$ (resp. $z^* \in \mathcal{R}_A$).

Since $\mathcal{Q}_A$ (resp. $\mathcal{R}_A$) is a subset of $\mathcal{D}$, the boundary of $\mathcal{Q}_A$ (resp. $\mathcal{R}_A$) is necessarily bounded as well. Moreover, (A.11) (resp. (A.12)) guarantees that this boundary is a quadratic curve. The only bounded quadratic curve is the ellipse [136], meaning $\mathcal{Q}_A$ (resp. $\mathcal{R}_A$) is the interior of an ellipse (contained in $\mathcal{D}$) if A is Hurwitz (resp. Schur).

$\boxed{\Leftarrow}$ It is shown that if A is not Hurwitz (resp. not Schur), then $\mathcal{Q}_A$ (resp. $\mathcal{R}_A$) is either an empty set, or is unbounded, guaranteeing that $\mathcal{Q}_A$ (resp. $\mathcal{R}_A$) cannot be the interior of an ellipse. The eigenvalues of A are denoted λ_1, λ_2. A proof by cases is performed.

If $\Re(\lambda_1) = \Re(\lambda_2) = 0$ (resp. $|\lambda_1| = |\lambda_2| = 1$).
In the continuous-time case, either $\lambda_1 = \lambda_2 = 0$, and v in the kernel of A guarantees $v^\top \mathcal{L}_A^c(P)v = 0$ for all P, or there exists trajectories of $\dot{x} = Ax$ following a limit cycle, and

no QLF can be strictly decreasing along these trajectories [149, 289]. Either way, there are no symmetric $P \in \mathbb{R}^{2\times 2}$ such that $\mathcal{L}_A^c(P) \in \mathbb{S}_2^{--}(\mathbb{R})$, so $\mathcal{Q}_A = \emptyset$. In the discrete-time case, since A is a real matrix, there exists $\theta \in [0, 2\pi)$ such that $\lambda_1 = e^{\mathbf{i}\theta}$, $\lambda_2 = e^{-\mathbf{i}\theta}$, hence $\lambda_1\lambda_2 = 1$, and Lemma 3.4 of [44] guarantees that there are no symmetric $P \in \mathbb{R}^{2\times 2}$ such that $\mathcal{L}_A^d(P) \in \mathbb{S}_2^{--}(\mathbb{R})$, so $\mathcal{R}_A = \emptyset$.

If $\Re(\lambda_1) < \Re(\lambda_2) = 0$ (resp. $|\lambda_1| < |\lambda_2| = 1$).

Since A is a real matrix, in that case the two distinct eigenvalues are necessarily real. Taking v_2 the eigenvector of A associated with λ_2, it is easily noticed that for all symmetric $P \in \mathbb{R}^{2\times 2}$, $v_2^\top \mathcal{L}_A^{c,d}(P) v_2 = 0$, so there exits no P such that $\mathcal{L}_A^{c,d}(P) \in \mathbb{S}_2^-(\mathbb{R})$, hence $\mathcal{Q}_A = \emptyset$ (resp. $\mathcal{R}_A = \emptyset$).

If $\Re(\lambda_1), \Re(\lambda_2) \geq 0$ (resp. $|\lambda_1|, |\lambda_2| \geq 1$).

Then $-A$ (resp. A^{-1}) is in the closure of the set of Hurwitz (resp. Schur) matrices. By symmetry of everything proven so far with respect to the hyperplane $\mathrm{Tr}(\cdot) = 0$, all the P (if they exist) such that $\mathcal{L}_{-A}^c(P) \in \mathbb{S}_2^{++}(\mathbb{R})$, i.e. $\mathcal{L}_A^c(P) \in \mathbb{S}_2^{--}(\mathbb{R})$ (resp. $\mathcal{L}_{A^{-1}}^d(P) \in \mathbb{S}_2^{++}(\mathbb{R})$, i.e. $\mathcal{L}_A^d(P) \in \mathbb{S}_2^{--}(\mathbb{R})$ by congruence [48]) are contained in $\mathbb{S}_2^{--}(\mathbb{R})$, which does not intersect $\mathcal{H}$, and finally $\mathcal{Q}_A = \emptyset$ (resp. $\mathcal{R}_A = \emptyset$).

If $\Re(\lambda_1) < 0 < \Re(\lambda_2)$ (resp. $|\lambda_1| < 1 < |\lambda_2|$).

Since A is a real matrix, in that case the two distinct eigenvalues are necessarily real. The eigenvalues of $A^\top$ are the same as those of A, and their associated eigenvectors are taken real, not collinear, and denoted v_1 and v_2. Clearly, $P_1 = v_1 v_1^\top \in \mathbb{S}_2^+(\mathbb{R}) \setminus (\mathbb{S}_2^{++}(\mathbb{R}) \cup \{0\})$ is such that $\mathcal{L}_A^{c,d}(P_1) \in \mathbb{S}_2^-(\mathbb{R})$ and $P_2 = -v_2 v_2^\top \in \mathbb{S}_2^-(\mathbb{R}) \setminus (\mathbb{S}_2^{--}(\mathbb{R}) \cup \{0\})$ is such that $\mathcal{L}_A^{c,d}(P_2) \in \mathbb{S}_2^-(\mathbb{R})$. By convexity, for all $t \in [0,1]$, $P(t) = tP_1 + (1-t)P_2$ is such that $\mathcal{L}_A^{c,d}(P(t)) \in \mathbb{S}_2^-(\mathbb{R})$. By continuity of $\mathrm{Tr}(P(t))$ with respect to $t \in [0,1]$, the intermediate value theorem provides $t^* \in (0,1)$ such that $\mathrm{Tr}(P(t^*)) = 0$. Since v_1 and v_2 are not collinear, $P(t^*) \neq 0$. Now if $\mathcal{Q}_A = \emptyset$ (resp. $\mathcal{R}_A = \emptyset$) the proof is finished. Otherwise, there exists $z \in \mathcal{Q}_A$ (resp. $z \in \mathcal{R}_A$), and it is easily checked that for all $q > 0$, $\mathcal{L}_A^{c,d}(qP(t^*) + P_z) \in \mathbb{S}_2^{--}(\mathbb{R})$. Yet, it is also easily verified that for all $q > 0$, $qP(t^*) + P_z \in \mathcal{H}$, hence $\mathcal{Q}_A$ (resp. $\mathcal{R}_A$) is unbounded. □

A last straightforward proposition is given below, stating that adding a Hurwitz (resp. Schur) matrix to the set $\{A_i\}_{1\leq i\leq m}$ never increases the CQLF problem difficulty if this matrix is proportional to I_2.

Proposition A.3.1. *For all $\alpha < 0$ and $\beta \in (-1, 1)$, $\mathcal{D} = \mathcal{Q}_{\alpha I_2} = \mathcal{R}_{\beta I_2}$.*

Proof. Given a symmetric positive-definite matrix $P_z \in \mathcal{H}$, $\alpha I_2 P_z + \alpha P_z I_2 = 2\alpha P_z$ is negative-definite if and only if $\alpha < 0$, which translates to $\mathcal{D} = \mathcal{Q}_{\alpha I_2}$; and $\beta^2 I_2 P_z I_2 - P_z = (\beta^2 - 1)P_z$ is negative-definite if and only if $\beta \in (-1, 1)$, which translates to $\mathcal{D} = \mathcal{R}_{\beta I_2}$. □

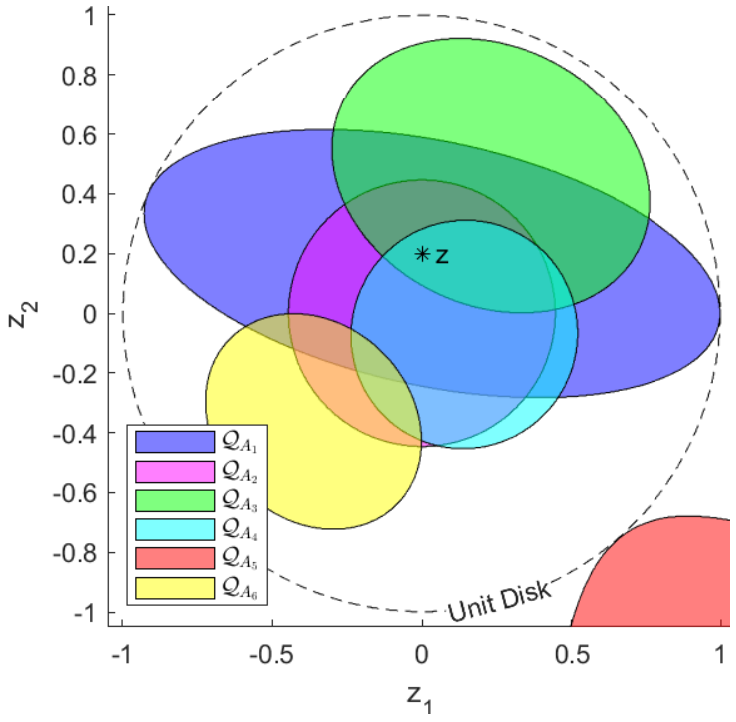


Figure A.2: Plot of the graphical criterion applied to the set of matrices (A.17) in the continuous-time setting.

A.4 Illustrative examples

Example A.4.1. *The graphical criterion is first applied in the continuous-time setting to the following matrices:*

$$A_1 = \begin{pmatrix} -10 & 0 \\ 2 & -0.5 \end{pmatrix} \quad A_2 = \begin{pmatrix} -1 & -2 \\ 2 & -1 \end{pmatrix}$$
$$A_3 = \begin{pmatrix} 0 & 0.5 \\ -1 & -1 \end{pmatrix} \quad A_4 = \begin{pmatrix} -3 & 5 \\ -7 & -2 \end{pmatrix} \tag{A.17}$$
$$A_5 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad A_6 = \begin{pmatrix} 0 & -0.5 \\ 0.2 & -0.3 \end{pmatrix}$$

The resulting plot is presented in Figure A.2. It can be noticed that the set $\mathcal{Q}_{A_5}$ is outside $\mathcal{D}$, hence no $P \in \mathbb{S}_2^{++}(\mathbb{R})$ such that $\mathcal{L}_{A_5}^c(P) \in \mathbb{S}_2^{--}(\mathbb{R})$ can be found, and A_5 is not Hurwitz. From there, it is clear that no CQLF can be found to the set of matrices (A.17). However, since $\mathcal{Q}_{A_1} \cap \mathcal{Q}_{A_2} \cap \mathcal{Q}_{A_3} \cap \mathcal{Q}_{A_4} \cap \mathcal{D}$ and $\mathcal{Q}_{A_1} \cap \mathcal{Q}_{A_2} \cap \mathcal{Q}_{A_4} \cap \mathcal{Q}_{A_6} \cap \mathcal{D}$ are not empty, Theorem A.3.1 guarantees that a CQLF can be found to $\{A_1, A_2, A_3, A_4\}$ and $\{A_1, A_2, A_4, A_6\}$. Moreover, as $\mathcal{Q}_{A_3}$ and $\mathcal{Q}_{A_6}$ are non-intersecting ellipses, there are no CQLF to the sets of matrices containing $\{A_3, A_6\}$.

Graphically, taking $z = (0; 0.2)$ provides $z \in \mathcal{Q}_{A_1} \cap \mathcal{Q}_{A_2} \cap \mathcal{Q}_{A_3} \cap \mathcal{Q}_{A_4} \cap \mathcal{D}$, hence the positive-definite matrix P_z given by

$$P_z = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \tag{A.18}$$

is such that $\mathcal{L}^c_{A_i}(P_z) \in \mathbb{S}_2^{--}(\mathbb{R})$ *for* $i = 1, \ldots, 4$. *This is verified below:*

$$\lambda_{\max}(A_1^\top P_z + P_z A_1) \approx -0.9995 < 0 \tag{A.19}$$
$$\lambda_{\max}(A_2^\top P_z + P_z A_2) \approx -1.1056 < 0 \tag{A.20}$$
$$\lambda_{\max}(A_3^\top P_z + P_z A_3) \approx -0.1101 < 0 \tag{A.21}$$
$$\lambda_{\max}(A_4^\top P_z + P_z A_4) \approx -0.8657 < 0 \tag{A.22}$$
$$\lambda_{\max}(A_5^\top P_z + P_z A_5) \approx +12.621 \geq 0 \tag{A.23}$$
$$\lambda_{\max}(A_6^\top P_z + P_z A_6) \approx +0.2085 \geq 0 \tag{A.24}$$

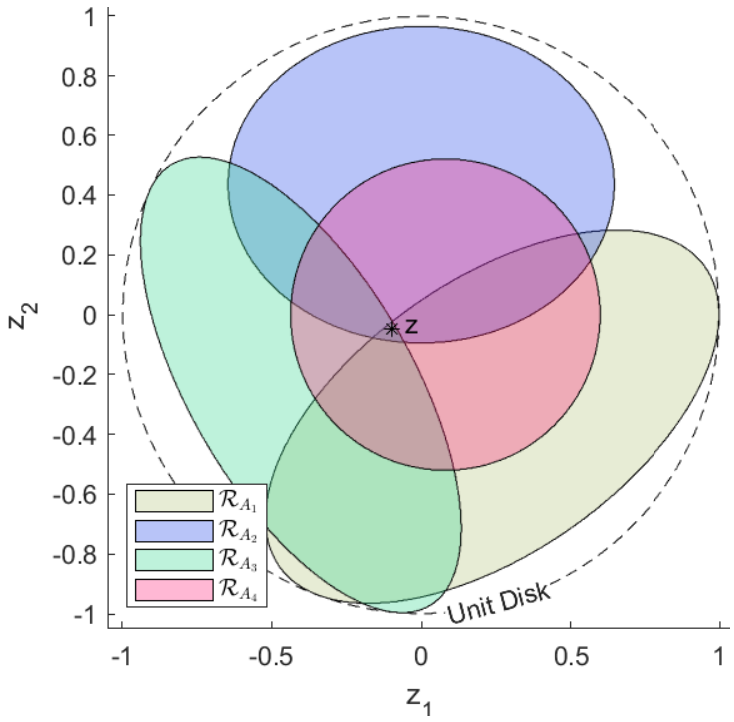


Figure A.3: Plot of the graphical criterion applied to the set of matrices (A.25) in the discrete-time setting.

Example A.4.2. *The graphical criterion is now applied in the discrete-time setting to the following matrices:*

$$A_1 = \begin{pmatrix} -0.5 & 0 \\ -0.7 & 0.5 \end{pmatrix} \quad A_2 = \begin{pmatrix} 0.8 & 0.4 \\ -0.4 & 0.2 \end{pmatrix}$$
$$A_3 = \begin{pmatrix} -0.4 & 1 \\ 0.2 & 0.3 \end{pmatrix} \quad A_4 = \begin{pmatrix} 0.5 & -0.4 \\ 0.5 & 0.5 \end{pmatrix} \tag{A.25}$$

The resulting plot is presented in Figure A.3. This time, there exists $z = (-0.1; -0.05) \in \mathcal{R}_{A_1} \cap \mathcal{R}_{A_2} \cap \mathcal{R}_{A_3} \cap \mathcal{R}_{A_4} \cap \mathcal{D}$, *hence Theorem A.3.1 guarantees that the positive-definite matrix* P_z *given by*

$$P_z = \begin{pmatrix} 0.9 & -0.05 \\ -0.05 & 1.1 \end{pmatrix} \tag{A.26}$$

is such that $\mathcal{L}^d_{A_i}(P_z) \in \mathbb{S}_2^{--}(\mathbb{R})$ *for* $i = 1, \ldots, 4$. *This is also verified below:*

$$\lambda_{\max}(A_1^\top P_z A_1 - P_z) \approx -0.0387 < 0 \tag{A.27}$$
$$\lambda_{\max}(A_2^\top P_z A_2 - P_z) \approx -0.0446 < 0 \tag{A.28}$$
$$\lambda_{\max}(A_3^\top P_z A_3 - P_z) \approx -0.0386 < 0 \tag{A.29}$$
$$\lambda_{\max}(A_4^\top P_z A_4 - P_z) \approx -0.3580 < 0 \tag{A.30}$$

A.5 MATLAB code

Figure A.2 is obtained with:

```
A = {[-10 0;2 -0.5],[-1 -2;2 -1],[0 0.5;-1 -1],[-3 5;-7 -2],[1 2;3 4],[0 -0.5;0.2
    -0.3]};
CQLF_criterion(A,'c','colorList',{'b','m','g','c','r','y'},'lineStyle','-');
```

Figure A.3 is obtained with:

```
A = {[-0.5 0;-0.7 0.5],[0.8 0.4;-0.4 0.2],[-0.4 1;0.2 0.3],[0.5 -0.4;0.5 0.5]};
CQLF_criterion(A,'d','opacity',0.3,'lineStyle','-');
```

The function *CQLF_criterion* is given thereafter:

```
function CQLF_criterion(A,time,varg)
arguments
A % A cell array of 2x2 real matrices
time='c'; % 'c' for continuous, 'd' for discrete
varg.resolution = 1500; % plot resolution
varg.colorList = arrayfun(@(x)rand(1,3),1:numel(A),'UniformOutput',false); % cell
    array specifying a color for each system
varg.opacity = 0.5; % patches opacity
varg.lineStyle = 'none'; % outline of the patches
end
z1 = linspace(-2,2,varg.resolution);
z2 = linspace(-2,2,varg.resolution);
[Z1,Z2] = meshgrid(z1,z2);
hold on;
for i=1:numel(A)
[a11,a12,a21,a22] = deal(A{i}(1,1),A{i}(1,2),A{i}(2,1),A{i}(2,2));
f = @(x,y)(x*(a11-a22)+y*(a12+a21)+a11+a22)+sqrt((x*(a12-a21)+y*(a11+a22)+a12+a21)
    .^2+(x*(a11+a22)+y*(a21-a12)+a11-a22).^2)<0;
g = @(x,y)(x*(a11^2+a12^2-a21^2-a22^2)/2+y*(a11*a21+a12*a22)+(a11^2+a12^2+a21^2+a22
    ^2)/2-1+sqrt((x*(a21^2+a12^2-a11^2-a22^2+2)/2+y*(a12*a22-a11*a21)+(a22^2+a12^2-
    a21^2-a11^2)/2).^2+(x*(a11*a12-a21*a22)+y*(a11*a22+a12*a21-1)+a11*a12+a21*a22)
    .^2))<0;
val = ((time=='c')*f(Z1,Z2)+(time=='d')*g(Z1,Z2));
C = contourc(z1,z2,val,[1 1]);
l = sprintf('$$\\mathcal{%s}_{A_%d}$$',(time=='c')*'Q'+(time=='d')*'R',i);
patch(C(1,2:end),C(2,2:end),varg.colorList{i},'FaceAlpha',varg.opacity,'LineStyle',
    varg.lineStyle,'DisplayName',l);
end
contour(Z1,Z2,(Z1.^2+Z2.^2<1),[1 1],'--','LineColor','k','HandleVisibility','off');
legend('Location','southwest','Interpreter','latex');
axis([[-1.05,1.05,-1.05,1.05]]);
pbaspect([1 1 1]);
end
```

Appendix B

Orthogonal projection of convex sets with a $\mathcal{C}^1$ boundary

B.1 Introduction

In analytical geometry, given a family of curves $(\mathcal{C}_t)_{t\in\mathbb{R}}$ defined on the plane $\mathbb{R}^2$ by

$$\mathcal{C}_t : F(x, y, t) = 0 \tag{B.1}$$

with F a differentiable function, the envelope of $(\mathcal{C}_t)_{t\in\mathbb{R}}$ is defined as the set of points $(x, y) \in \mathbb{R}^2$ such that [83, 219]

$$\exists t \in \mathbb{R}, \quad F(x, y, t) = 0 \qquad \frac{\partial F}{\partial t}(x, y, t) = 0 \tag{B.2}$$

The well-known envelope theorem, mainly used in economics and optimization [3, 59, 191, 175], provides conditions for the envelope of a family of curves $(\mathcal{C}_t)_{t\in\mathbb{R}}$ to coincide with a single curve tangent to all of the $\mathcal{C}_t$. Under some circumstances, this curve is also the boundary of the region filled by $(\mathcal{C}_t)_{t\in\mathbb{R}}$, and despite this characterization being visually clear (Figure B.1), the authors have not been able to find a satisfying topological discussion on this matter in the literature [192, 142].

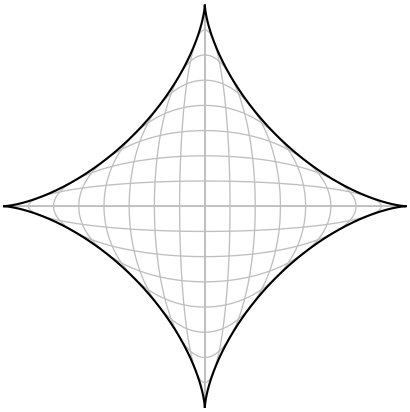


Figure B.1: The black astroid in the picture above can be seen as the envelope of the family of gray curves $(\mathcal{C}_t)_{t\in(0,1)}$ defined by $\mathcal{C}_t : (x/(1-t))^2 + (y/t)^2 = 1$

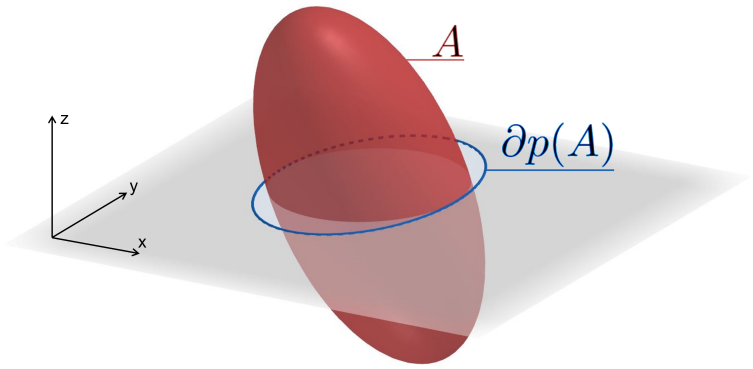


Figure B.2: A, the 3-dimensional ellipsoid in red, is a convex and compact set of $\mathbb{R}^3$. $p(A)$, the 2-dimensional ellipsoid with a blue outline, is the projection of A along the z-axis onto the xy-plane represented in gray, and $\partial p(A)$ is the boundary of this projection.

Now, given A a convex set of $\mathbb{R}^3$ with a boundary characterized by $F(x, y, z) = 0$ where F is differentiable, one can intuitively see by the envelope theorem how characterizing the boundary of A projected along the z-axis onto the xy-plane relates to the partial derivative of F with respect to z vanishing (Figure B.2). Moreover, the function F can be obtained from μ_A, the Minkowski functional associated with A, usually with the relation $F = \mu_A - 1$ [179]. In a more general setting, with E a Euclidean space and A a compact and convex set of E with a differentiable boundary and a non-empty interior, the aim of this appendix is to elucidate the link between the partial derivatives of μ_A and the boundary of the orthogonal projection of A onto the linear subspaces of E. Leveraging results from convex analysis [234], a system of equations for the orthogonal projection of A onto any linear subspace of E is obtained. This appendix can also be found in the arXiv deposit [22].

In this appendix, E denotes a Hilbert space of finite dimension over $\mathbb{R}$. Let $z \in E$, the hyperplane H defined by $H = \mathrm{Ker}(\langle z|\cdot\rangle)$ is called a supporting hyperplane of A at $x \in \partial_E(A)$ if for all $y \in E$, $\mu_A(y) \geq \mu_A(x) + \langle z|y - x\rangle$. For all $x \in E$, there exists a unique $x_{\mathcal{V}} \in \mathcal{V}$ and a unique $x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp$ such that $x = x_{\mathcal{V}} + x_{\mathcal{V}^\perp}$. From now on, $p_{\mathcal{V}}$ always denotes the map $x \mapsto x_{\mathcal{V}}$, that is to say the orthogonal projection along $\mathcal{V}^\perp$ onto $\mathcal{V}$, with $\mathcal{V} \neq \{0\}$.

> A always denotes a convex, bounded set of E with $0 \in \mathrm{intr}_E(A)$. A is said to have a differentiable boundary if $\mu_A \in \mathcal{C}^1(E \setminus \{0\}, \mathbb{R})$.

B.2 Preliminary results

As stated in the introduction, the partial derivatives of the equation of the boundary of A (a notion of convex analysis) are related to the boundary of the orthogonal projection of A onto the linear subspaces of E (a topological consideration). The main purpose of these preliminary results is to draw a link from convex analysis to topology via the Minkowski functionals associated with A. In particular, these preliminary results mainly focus on the link between the

gradient of μ_A and a topological characterization of the supporting hyperplanes of A (Corollary B.2.1, Corollary B.2.2 and Figure B.3). The topological characterization of the supporting hyperplanes of A then provides a characterization of the boundary of the projection of A onto $\mathcal{V}$ (Lemma B.2.6 and Figure B.5), which can finally be linked back to the gradient of μ_A.

Considering that A has a differentiable boundary (i.e. $\mu_A \in \mathcal{C}^1(E \setminus \{0\}, \mathbb{R})$), unicity of the supporting hyperplanes of A is demonstrated using the following result from convex analysis.

Property B.2.1. *Let $f \in \mathcal{C}^1(U, \mathbb{R})$ be a convex function. For all $x \in E$, we have:*

$$\{z \in E \,:\, \forall y \in U,\, f(y) \geq f(x) + \langle z | y - x \rangle\} = \left\{ \frac{\partial}{\partial x} f(x) \right\} \tag{B.3}$$

Proof. See Theorem 25.1 at page 242 of [234]. □

Remark B.2.1. *The set on the left-hand side of (B.3) contains the subgradients of f at x and is not necessarily a singleton when f is not differentiable at x.*

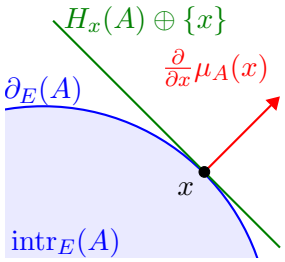
Indeed, if A has a differentiable boundary, then μ_A is a $\mathcal{C}^1$ convex function, and if for all $y \in E$, $\mu_A(y) \geq \mu_A(x) + \langle z | y - x \rangle$, then $\mathrm{Ker}(\langle z | \cdot \rangle)$ is by definition a supporting hyperplane of A at $x \in \partial_E(A)$. Unicity of the supporting hyperplanes of A is obtained from the unicity of such z. This links the supporting hyperplanes of A with the gradient of μ_A (Figure B.3a).

Corollary B.2.1 (The gradient characterization)**.** *If A has a differentiable boundary, then the hyperplane orthogonal to $\frac{\partial}{\partial x}\mu_A(x)$ is only supporting hyperplane of A at $x \in \partial_E(A)$. From now on, this supporting hyperplane is denoted $H_x(A)$. Formally, for all $x \in \partial_E(A)$, the following holds:*

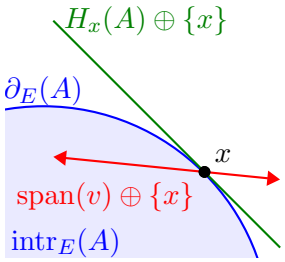
$$H_x(A) = \mathrm{Ker}\left(\langle \frac{\partial}{\partial x} \mu_A(x) | \cdot \rangle \right) \tag{B.4}$$

Now that the supporting hyperplane of A at x is linked with the gradient of μ_A at x, the previous results are now leveraged to obtain a topological characterization of the supporting hyperplanes of A. For a convex shape with a differentiable boundary, the supporting hyperplane at a boundary point of this shape is the only hyperplane that, once translated to this point, does not intersect the interior of the shape (Figure B.3b). Lemma B.2.1 provides the fact that a supporting hyperplane of A never intersects the interior of A, and Lemma B.2.3 provides the fact that, if A has a differentiable boundary, then any affine vector line going through $x \in \partial_E(A)$ that is not included in the supporting hyperplane of A at x will cross the interior of A.

Lemma B.2.1. *If H is a supporting hyperplane of A at $x \in \partial_E(A)$, then $(H \oplus \{x\}) \cap \mathrm{intr}_E(A) = \emptyset$.*



(a) Illustration of Corollary B.2.1



(b) Illustration of Corollary B.2.2

Figure B.3: Illustration of the gradient characterization (Corollary B.2.1) and of the topological characterization (Corollary B.2.2) of the supporting hyperplane of A at $x \in \partial_E(A)$ when A has a differentiable boundary.

Proof. By definition of the supporting hyperplane, for all $h \in H$ the following inequality holds $\mu_A(x+h) \geq \mu_A(x)$. Moreover since $x \in \partial_E(A)$, then $\mu_A(x) = 1$, which provides $\mu_A(x+h) \geq 1$, hence $(H \oplus \{x\}) \subseteq \mu_A^{-1}([1, +\infty))$, yet $\mathrm{intr}_E(A) = \mu_A^{-1}([0, 1))$. □

Lemma B.2.2. *By parallelism, if $(H \oplus \{x\}) \cap \mathrm{intr}_E(A) = \emptyset$, then $(H \oplus \{x\}) \cap (H \oplus \mathrm{intr}_E(A)) = \emptyset$ as well.*

Proof. This statement is proved by contraposition.
If there exists $y \in (H \oplus \{x\}) \cap (H \oplus \mathrm{intr}_E(A))$, then there exists $z \in \mathrm{intr}_E(A)$ and $h_1, h_2 \in H$ such that $y = x + h_1 = z + h_2$, providing $z = x + (h_1 - h_2)$ where $(h_1 - h_2) \in H$, hence $z \in (H \oplus \{x\}) \cap \mathrm{intr}_E(A)$. □

Lemma B.2.3. *Suppose A has a differentiable boundary. If $v \notin H_x(A)$, then*

$$(\mathrm{span}(v) \oplus \{x\}) \cap \mathrm{intr}_E(A) \neq \emptyset \tag{B.5}$$

Proof. This statement is proved by contraposition.
Suppose $(\mathrm{span}(v) \oplus \{x\}) \cap \mathrm{intr}_E(A) = \emptyset$ and consider the function $\phi(t) = \mu_A(x + tv)$. Since $\mu_A \in \mathcal{C}^1(E \setminus \{0\}, \mathbb{R}_+)$ is a convex function, then $\phi \in \mathcal{C}^1(\mathbb{R}, \mathbb{R}_+)$ is convex as well. Moreover, since $(\mathrm{span}(v) \oplus \{x\}) \cap \mathrm{intr}_E(A) = \emptyset$, then for all $t \in \mathbb{R}$, $\phi(t) \geq 1$. Yet $\phi(0) = 1$, $t = 0$ is therefore a minimum for ϕ, which implies $\phi'(0) = 0$. However, $\phi'(0) = \langle \frac{\partial}{\partial x}\mu_A(x) | v \rangle$, hence $v \in \mathrm{Ker}(\langle \frac{\partial}{\partial x}\mu_A(x) | \cdot \rangle)$. □

From the Lemmas B.2.1 and B.2.3, the following necessary and sufficient condition can be stated, providing a topological characterization of supporting hyperplanes (Figure B.3b) on top of their analytical one (obtained in Corollary B.2.1):

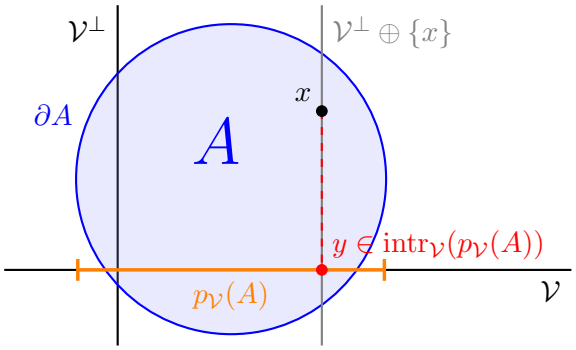


Figure B.4: If x is in the interior of A, then $y = p_\mathcal{V}(x)$ is in the interior of $p_\mathcal{V}(A)$, that is to say $y \in \mathrm{intr}_\mathcal{V}(p_\mathcal{V}(A))$, and $\mathcal{V}^\perp \oplus \{x\}$ crosses the boundary of A multiple times.

Corollary B.2.2 (The topological characterization)**.** *If A has a differentiable boundary, then $H_x(A)$ contains exactly the directions coming from x that never intersect the interior of A. Formally, for all $x \in \partial_E(A)$, the following holds:*

$$H_x(A) = \{v \in E \,:\, (\mathrm{span}(v) \oplus \{x\}) \cap \mathrm{intr}_E(A) = \emptyset\} \tag{B.6}$$

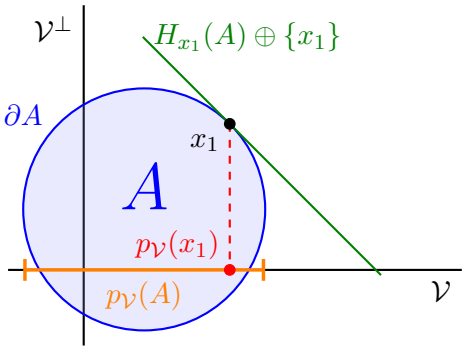
Before linking the topological characterization of the supporting hyperplanes of A with the boundary of the projection $p_\mathcal{V}(A)$ of A onto $\mathcal{V}$, two topological results on the orthogonal projection of A are stated. The first one simply states that the interior of the projection of A is the projection of the interior of A (Lemma B.2.4). The second one states that the projection of the closure of A is also the projection of the boundary of A (Lemma B.2.5). Both are easy to understand visually with the help of Figure B.4.

Lemma B.2.4. *If A has a differentiable boundary, then $p_\mathcal{V}(\mathrm{intr}_E(A)) = \mathrm{intr}_\mathcal{V}(p_\mathcal{V}(A))$.*

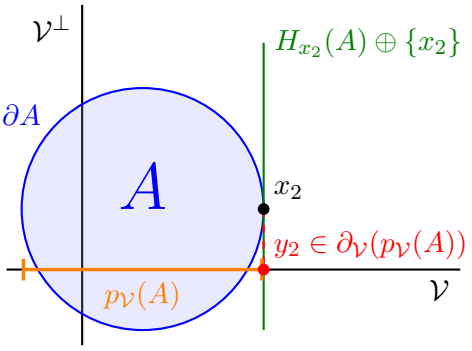
Proof. This statement is proved by double inclusion.

$\boxed{\subseteq}$ This inclusion is a direct consequence of $p_\mathcal{V}$ being an open map from E to $\mathcal{V}$.

$\boxed{\supseteq}$ Let $y \in \mathrm{intr}_\mathcal{V}(p_\mathcal{V}(A))$, and $x \in A$ such that $p_\mathcal{V}(x) = y$. If $x \in \mathrm{intr}_E(A)$ there is nothing to prove. If $x \in \partial_E(A)$, the following will show by contradiction that $(\mathcal{V}^\perp \oplus \{x\}) \cap \mathrm{intr}_E(A) \neq \emptyset$, which, thanks to Lemma B.2.3, is equivalent to the existence of $v \in \mathcal{V}^\perp$ such that $v \notin H_x(A)$. By contradiction, it is assumed that $\mathcal{V}^\perp \subseteq H_x(A)$. By the hyperplane separation theorem, A is contained on one side of $H_x(A) \oplus \{x\}$, hence there is $v \in H_x(A)^\perp \setminus \{0\}$ such that for all $t \in \mathbb{R}_+^*$, $x + tv \notin A \oplus H_x(A)$, therefore $x + tv \notin A \oplus \mathcal{V}^\perp$, and finally $p_\mathcal{V}(x + tv) \notin p_\mathcal{V}(A)$. However, since $\mathcal{V}^\perp \subseteq H_x(A)$, then $v \in \mathcal{V}$, hence $p_\mathcal{V}(x + tv) = y + tv$. Since $y \in \mathrm{intr}_\mathcal{V}(p_\mathcal{V}(A))$, by definition of the interior there exists $\delta \in \mathbb{R}_+^*$ such that $\mathcal{B}_\mathcal{V}(y, \delta) \subseteq p_\mathcal{V}(A)$, so in particular there exists $\epsilon \in (0, \delta)$ such that $p_\mathcal{V}(x + \epsilon v) = y + \epsilon v \in p_\mathcal{V}(A)$, which contradicts that for all $t \in \mathbb{R}_+^*$, $p_\mathcal{V}(x + tv) \notin p_\mathcal{V}(A)$. Finally $(\mathcal{V}^\perp \oplus \{x\}) \cap \mathrm{intr}_E(A) \neq \emptyset$. □



(a) $H_{x_1}(A)$ is the supporting hyperplane of A at $x_1 \in \partial A$. The projection of x_1 along $\mathcal{V}^\perp$ onto $\mathcal{V}$, denoted $p_\mathcal{V}(x_1)$, is generally unrelated to $H_{x_1}(A)$.



(b) If $\mathcal{V}^\perp \subseteq H_{x_2}(A)$ with $x_2 \in \partial A$, then $y_2 = p_\mathcal{V}(x_2)$ is also on the boundary of $p_\mathcal{V}(A)$, that is to say $y_2 \in \partial_\mathcal{V}(p_\mathcal{V}(A))$

Figure B.5: Illustration of the supporting hyperplanes relation to the orthogonal projection of a convex shape. This relation is formalized in Lemma B.2.6.

Lemma B.2.5. *The following equality holds:* $p_\mathcal{V}(\mathrm{cls}_E(A)) = p_\mathcal{V}(\partial_E(A))$.

Proof. This statement is proved by double inclusion.

$\boxed{\subseteq}$ Let $y \in p_\mathcal{V}(\mathrm{cls}_E(A))$, and $x \in \mathrm{cls}_E(A)$ such that $y = p_\mathcal{V}(x)$. If $x \in \partial_E(A)$ there is nothing to prove. If $x \in \mathrm{intr}_E(A)$, by definition of the interior there exists $\epsilon \in \mathbb{R}_+^*$ such that $\mathcal{B}_E(x, \epsilon) \subseteq A$. Let $v \in \left(\mathcal{B}_E(0, \epsilon) \cap \mathcal{V}^\perp\right) \setminus \{0\}$, which guarantees $x + v \in \mathrm{intr}_E(A)$. Since A is bounded, there exists $t \in (1, +\infty)$ such that $x + tv \notin \mathrm{cls}_E(A)$. Considering the Minkowski functional $\mu_{A \oplus \{-x\}}$, $x+v \in \mathrm{intr}_E(A)$ translates to $\mu_{A\oplus\{-x\}}(v) < 1$, and $x+tv \notin \mathrm{cls}_E(A)$ translates to $\mu_{A\oplus\{-x\}}(tv) > 1$. By continuity of $\mu_{A\oplus\{-x\}}$ the intermediate value theorem provides the existence of $t^* \in (1, t)$ such that $\mu_{A\oplus\{-x\}}(t^*v) = 1$, hence $x + t^*v \in \partial_E(A)$. Moreover $x + t^*v \in \mathcal{V}^\perp \oplus \{y\}$, meaning $p_\mathcal{V}(x + t^*v) = y$ (Figure B.4).

$\boxed{\supseteq}$ This inclusion is a direct consequence of the inclusion $\partial_E(A) \subseteq \mathrm{cls}_E(A)$. □

With the help of the previous results, the supporting hyperplanes relation to the boundary of the orthogonal projection of A onto $\mathcal{V}$ can be formally stated. Intuitively, when $y \in \mathcal{V}$ is at the boundary of $p_\mathcal{V}(A)$, the supporting hyperplane at the pre-image of y by $p_\mathcal{V}$ includes $\mathcal{V}^\perp$, the direction of the projection. Reciprocally, when there is such an alignment, that is to say when $\mathcal{V}^\perp$ is contained in the supporting hyperplane of the pre-image of y by $p_\mathcal{V}$, then $y \in \mathcal{V}$ is at the boundary of $p_\mathcal{V}(A)$ (see Figure B.5). More exactly, the following Lemma holds.

Lemma B.2.6. *Let A be closed and have a differentiable boundary. If $y \in p_\mathcal{V}(A)$, then the following statements are equivalent:*

1. $y \in \partial_\mathcal{V}(p_\mathcal{V}(A))$

2. $\{x \in \partial_E(A) : p_{\mathcal{V}}(x) = y\}$ *is convex*

3. $\exists x \in \partial_E(A) \,|\, \begin{cases} p_{\mathcal{V}}(x) = y \\ \mathcal{V}^\perp \subseteq H_x(A) \end{cases}$

Proof. This statement is proved by a circular chain of implications.
The notation $B = \{x \in \partial_E(A) : p_{\mathcal{V}}(x) = y\}$ is used in this proof as a shorthand.

$\boxed{1. \Rightarrow 2.}$ This implication is proved by contraposition.
Suppose B is not empty and not convex, hence there exists $z \in \text{hull}(B) \setminus B$. Clearly $\partial_E(A) \subseteq \text{cls}_E(A)$, and the following inclusion is easily checked:

$$\text{hull}(B) \subseteq \text{hull}\,\{x \in \text{cls}_E(A) : p_{\mathcal{V}}(x) = y\} \tag{B.7}$$

moreover, taking $x_1, x_2 \in \{x \in \text{cls}_E(A) : p_{\mathcal{V}}(x) = y\}$, by linearity of $p_{\mathcal{V}}$, we have for all $t \in [0,1]$, $(tx_1 + (1-t)x_2) \in \{x \in \text{cls}_E(A) : p_{\mathcal{V}}(x) = y\}$, which finally provides the convexity of $\{x \in \text{cls}_E(A) : p_{\mathcal{V}}(x) = y\}$, hence:

$$\text{hull}(B) \subseteq \{x \in \text{cls}_E(A) : p_{\mathcal{V}}(x) = y\} \tag{B.8}$$

This provides the following:

$$\begin{aligned} \text{hull}(B) \setminus B &\subseteq \{x \in \text{cls}_E(A) : p_{\mathcal{V}}(x) = y\} \setminus \{x \in \partial_E(A) : p_{\mathcal{V}}(x) = y\} \\ &\subseteq \{x \in \text{intr}_E(A) : p_{\mathcal{V}}(x) = y\} \end{aligned} \tag{B.9}$$

This provides $z \in \text{intr}_E(A)$ with $p_{\mathcal{V}}(z) = y$. By definition of the interior, there exists $\epsilon \in \mathbb{R}_+^*$ such that $\mathcal{B}_E(z, \epsilon) \subseteq A$. For all $h \in \mathcal{B}_E(0, \epsilon)$, $p_{\mathcal{V}}(z + h) = y + p_{\mathcal{V}}(h)$, and since $\|p_{\mathcal{V}}\|_2 = 1$, then $p_{\mathcal{V}}(h) \in \mathcal{B}_{\mathcal{V}}(0, \epsilon)$, hence $p_{\mathcal{V}}(\mathcal{B}_E(z, \epsilon)) \subseteq \mathcal{B}_{\mathcal{V}}(y, \epsilon) \subseteq p_{\mathcal{V}}(A)$. This finally provides $y \in \text{intr}_{\mathcal{V}}(p_{\mathcal{V}}(A))$.

$\boxed{2. \Rightarrow 3.}$ Since A is closed, then, by Lemma B.2.5, $y \in p_{\mathcal{V}}(\partial_E(A))$, hence $B \neq \emptyset$. Let $x \in B$ and $v \in \mathcal{V}^\perp$. The following will show by contradiction that $t \in \mathbb{R}$, $x + tv \notin \text{intr}_E(A)$.
Suppose, without loss of generality, that there exists $t \in \mathbb{R}_+^*$ such that $x + tv \in \text{intr}_E(A)$. Since A is bounded, with the help of the intermediate value theorem (similarly to Lemma B.2.5), there exists $t^* \in (1, +\infty)$ such that $x + t^*tv \in \partial_E(A)$. This provides $x \in B$, $x + t^*tv \in B$, and $x + tv \notin B$, yet B should be convex, so there is a contradiction (Figure B.4). This provides $\text{span}(v) \cap \text{intr}_E(A) = \emptyset$, hence by Corollary B.2.2, $v \in H_x(A)$.

$\boxed{3. \Rightarrow 1.}$ Let $x \in \partial_E(A)$ be such that $p_{\mathcal{V}}(x) = y$ and $\mathcal{V}^\perp \subseteq H_x(A)$. Lemma B.2.2 provides $(H_x(A) \oplus \{x\}) \cap (H_x(A) \oplus \text{intr}_E(A)) = \emptyset$, hence $(\mathcal{V}^\perp \oplus \{x\}) \cap (\mathcal{V}^\perp \oplus \text{intr}_E(A)) = \emptyset$. Moreover the following equalities hold:

$$\begin{aligned} (\mathcal{V}^\perp \oplus \{x\}) \cap (\mathcal{V}^\perp \oplus \text{intr}_E(A)) &= p_{\mathcal{V}}^{-1}(\{y\}) \cap p_{\mathcal{V}}^{-1}(p_{\mathcal{V}}(\text{intr}_E(A))) \\ &= p_{\mathcal{V}}^{-1}(\{y\}) \cap p_{\mathcal{V}}^{-1}(\text{intr}_{\mathcal{V}}(p_{\mathcal{V}}(A))) \quad \text{[Lemma B.2.4]} \\ &= p_{\mathcal{V}}^{-1}(\{y\} \cap \text{intr}_{\mathcal{V}}(p_{\mathcal{V}}(A))) \end{aligned} \tag{B.10}$$

Hence $\{y\} \cap \text{intr}_{\mathcal{V}}(p_{\mathcal{V}}(A)) = \emptyset$, that is to say $y \notin \text{intr}_{\mathcal{V}}(p_{\mathcal{V}}(A))$, providing $y \in \partial_{\mathcal{V}}(p_{\mathcal{V}}(A))$. □

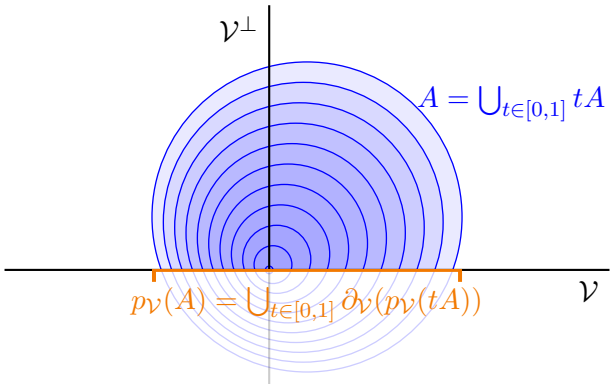


Figure B.6: Illustration of Lemma B.2.7, where A is assumed to be closed

Lastly, the projection of A onto $\mathcal{V}$ can be seen as the union of the boundaries of the projections of tA onto $\mathcal{V}$ with $t \in [0,1]$ (see Figure B.6). In the next section, the following Lemma will provide a way to go from a statement on the boundary of the projection to a statement on the whole projection $p_\mathcal{V}(A)$.

Lemma B.2.7. *The following equality holds:* $p_\mathcal{V}(\mathrm{cls}_E(A)) = \bigcup_{t\in[0,1]} \partial_\mathcal{V}(p_\mathcal{V}(\mathrm{cls}_E(tA)))$

Proof. $\mu_{p_\mathcal{V}(\mathrm{cls}_E(A))}$ denotes the Minkowski functional of $p_\mathcal{V}(\mathrm{cls}_E(A))$ defined over $\mathcal{V}$. The following equalities hold:

$$\begin{aligned}
p_\mathcal{V}(\mathrm{cls}_E(A)) &= \mathrm{cls}_\mathcal{V}(p_\mathcal{V}(\mathrm{cls}_E(A))) && [p_\mathcal{V} \text{ continuous}] \\
&= \mu^{-1}_{p_\mathcal{V}(\mathrm{cls}_E(A))}([0,1]) \\
&= \bigcup_{t\in[0,1]} \mu^{-1}_{p_\mathcal{V}(\mathrm{cls}_E(A))}(\{t\}) \\
&= \bigcup_{t\in[0,1]} \partial_\mathcal{V}(t p_\mathcal{V}(\mathrm{cls}_E(A))) \\
&= \bigcup_{t\in[0,1]} \partial_\mathcal{V}(p_\mathcal{V}(t\,\mathrm{cls}_E(A))) && [\text{linearity of } p_\mathcal{V}] \\
p_\mathcal{V}(\mathrm{cls}_E(A)) &= \bigcup_{t\in[0,1]} \partial_\mathcal{V}(p_\mathcal{V}(\mathrm{cls}_E(tA)))
\end{aligned} \tag{B.11}$$

□

B.3 Characterization of the orthogonal projection

The main result of this appendix consists in obtaining a system of equations that characterizes the orthogonal projection of the closure of A on a linear subspace $\mathcal{V} \neq \{0\}$ when A has a differentiable boundary. To obtain this system of equations, the following Minkowski functional of

two variables is introduced:

$$\begin{aligned} \psi_A : \mathcal{V} \times \mathcal{V}^\perp &\to \mathbb{R} \\ \left(x_\mathcal{V}, x_{\mathcal{V}^\perp}\right) &\mapsto \mu_A\left(x_\mathcal{V} + x_{\mathcal{V}^\perp}\right) \end{aligned} \tag{B.12}$$

From now on, $\frac{\partial \psi_A}{\partial x_\mathcal{V}}$ denotes the partial derivative of ψ_A with respect to $x_\mathcal{V}$ and $\frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}$ denotes the partial derivative of ψ_A with respect to $x_{\mathcal{V}^\perp}$.

The link between the partial derivatives of ψ_A and the boundary of the orthogonal projection of A onto the linear subspaces of E is explicitly written and leveraged in the proof of this characterization.

Theorem B.3.1. *If A is a compact and convex set of E with a differentiable boundary and $0 \in \mathrm{intr}_E(A)$, then, for all projection $p_\mathcal{V}$ such that $\mathcal{V} \neq \{0\}$, the following equality holds:*

$$p_\mathcal{V}(A) = \left\{ x_\mathcal{V} \in \mathcal{V} \,:\, \exists x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp \mid \begin{cases} \psi_A\left(x_\mathcal{V}, x_{\mathcal{V}^\perp}\right) \leq 1 \\ x_\mathcal{V} + x_{\mathcal{V}^\perp} \neq 0 \Rightarrow \frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}\left(x_\mathcal{V}, x_{\mathcal{V}^\perp}\right) = 0 \end{cases} \right\} \tag{B.13}$$

Proof. If $t = 0$ then $tA = \{0\} = p_\mathcal{V}(tA)$, hence for all $x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp$, the equality $\psi_A(0, x_{\mathcal{V}^\perp}) = \mu_{p_\mathcal{V}(A)}(0)$ holds, and there is nothing to prove.

If $t \in \mathbb{R}_+^*$, thanks to Lemma B.2.6, the following equivalence holds:

$$y \in \partial_\mathcal{V}(p_\mathcal{V}(tA)) \Leftrightarrow \exists x \in \partial_E(tA) \mid \begin{cases} p_\mathcal{V}(x) = y \\ \mathcal{V}^\perp \subseteq H_x(tA) \end{cases} \tag{B.14}$$

For all $x \in \partial_E(tA)$, $H_x(tA) = \mathrm{Ker}(\langle \frac{\partial}{\partial x}\mu_{tA}(x)|\cdot\rangle)$, and since $t \neq 0$, then $\frac{\partial}{\partial x}\mu_{tA}(x) = \frac{\partial}{\partial x}\mu_A(x)$. Moreover for all $h \in E$, $x_\mathcal{V}, h_\mathcal{V} \in \mathcal{V}$ and $x_{\mathcal{V}^\perp}, h_{\mathcal{V}^\perp} \in \mathcal{V}^\perp$ such that $x = x_\mathcal{V} + x_{\mathcal{V}^\perp}$ and $h = h_\mathcal{V} + h_{\mathcal{V}^\perp}$, the following equality holds:

$$\langle \frac{\partial}{\partial x}\mu_A(x)|h\rangle = \frac{\partial \psi_A}{\partial x_\mathcal{V}}\left(x_\mathcal{V}, x_{\mathcal{V}^\perp}\right) h_\mathcal{V} + \frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}\left(x_\mathcal{V}, x_{\mathcal{V}^\perp}\right) h_{\mathcal{V}^\perp} \tag{B.15}$$

hence the following equivalences hold:

$$\begin{aligned} y \in \partial_\mathcal{V}(p_\mathcal{V}(tA)) \Leftrightarrow \exists x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp \mid \begin{cases} y + x_{\mathcal{V}^\perp} \in \partial_E(tA) \\ \frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}(y, x_{\mathcal{V}^\perp}) = 0 \end{cases} \\ \text{i.e. } y \in \partial_\mathcal{V}(p_\mathcal{V}(tA)) \Leftrightarrow \exists x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp \mid \begin{cases} \psi_A(y, x_{\mathcal{V}^\perp}) = t \\ \frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}(y, x_{\mathcal{V}^\perp}) = 0 \end{cases} \end{aligned} \tag{B.16}$$

For $t = 1$, this last equivalence provides the link between the partial derivatives of ψ_A and the boundary of the orthogonal projection of A onto the linear subspaces of E.

Finally, Lemma B.2.7 provides:

$$\begin{aligned} p_{\mathcal{V}}(A) &= p_{\mathcal{V}}(\mathrm{cls}_E(A)) \\ &= \bigcup_{t\in[0,1]} \partial_{\mathcal{V}}(p_{\mathcal{V}}(tA)) \\ &= \{0\} \cup \bigcup_{t\in(0,1]} \left\{ x_{\mathcal{V}} \in \mathcal{V} \,:\, \exists x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp \,|\, \begin{cases} \psi_A\left(x_{\mathcal{V}}, x_{\mathcal{V}^\perp}\right) = t \\ \frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}\left(x_{\mathcal{V}}, x_{\mathcal{V}^\perp}\right) = 0 \end{cases} \right\} \end{aligned} \tag{B.17}$$

$$p_{\mathcal{V}}(A) = \left\{ x_{\mathcal{V}} \in \mathcal{V} \,:\, \exists x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp \,|\, \begin{cases} \psi_A\left(x_{\mathcal{V}}, x_{\mathcal{V}^\perp}\right) \leq 1 \\ x_{\mathcal{V}} + x_{\mathcal{V}^\perp} \neq 0 \Rightarrow \frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}\left(x_{\mathcal{V}}, x_{\mathcal{V}^\perp}\right) = 0 \end{cases} \right\}$$

which concludes the proof. □

Given a compact and convex set of E with a differentiable boundary and a non-empty interior, there exists a translation so that the origin of E is in the interior of the translated set, hence this new set is absorbing. Given a good translation of A, the main result of this appendix can therefore be extended without difficulty to a more general setting where A simply denotes a compact and convex set of E with a differentiable boundary and a non-empty interior.

Corollary B.3.1. *Keeping the assumptions of Theorem B.3.1, the following equality holds:*

$$\mu_{p_{\mathcal{V}}(A)}(x) = \begin{cases} \inf \left\{ t \in \mathbb{R}_+^* \,:\, \exists x_{\mathcal{V}^\perp} \in \mathcal{V}^\perp \,|\, \begin{cases} \psi_A\left(x, x_{\mathcal{V}^\perp}\right) \leq t \\ x + x_{\mathcal{V}^\perp} \neq 0 \Rightarrow \frac{\partial \psi_A}{\partial x_{\mathcal{V}^\perp}}\left(x, x_{\mathcal{V}^\perp}\right) = 0 \end{cases} \right\} & \text{if } x \in \mathcal{V} \\ +\infty & \text{if } x \notin \mathcal{V} \end{cases} \tag{B.18}$$

Moreover, if V and $V^\perp$ denote the matrices whose columns are resp. formed by $(v_1, \ldots, v_m)$ a basis to $\mathcal{V}$ and $(v_{m+1}, \ldots, v_n)$ a basis to $\mathcal{V}^\perp$, then the following equality holds:

$$\mu_{PA}(y) = \mu_{p_{\mathcal{V}}(A)}(Vy) \tag{B.19}$$

with $P = \begin{bmatrix} I_m & 0 \end{bmatrix} \begin{bmatrix} V & V^\perp \end{bmatrix}^{-1}$ and where $y \in \mathbb{R}^m$ is expressed in the $(v_1, \ldots, v_m)$ basis.

Proof. Equation (B.18) is easily derived by replacing the interval $[0, 1]$ by the interval $[0, t]$ in the proof of Theorem B.3.1. Equation (B.19) is a trivial consequence of $p_{\mathcal{V}}(A) = VPA$. □

B.4 Illustrative example

An implicit parametric equation to the projection of the unit ball of norm 4 of $\mathbb{R}^3$ (denoted A) onto the plane $H : x + y + z = 0$ is derived as an illustrative example of Theorem B.3.1.

The Minkowski functional of A is given by:

$$\mu_A(x, y, z) = \sqrt[4]{x^4 + y^4 + z^4} \tag{B.20}$$

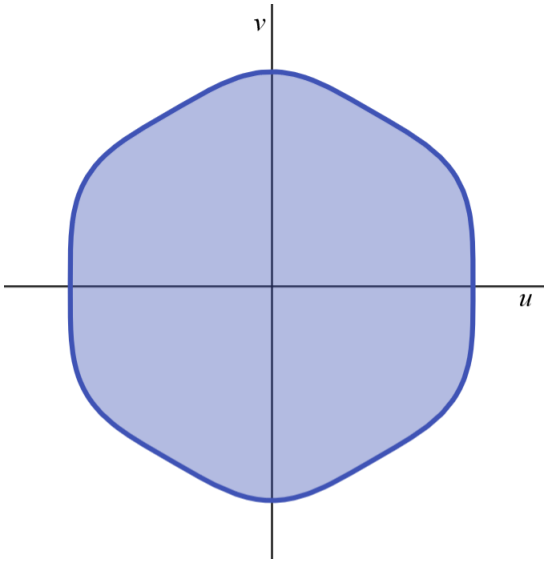


Figure B.7: Shape of the projection of the unit ball of norm 4 of $\mathbb{R}^3$ onto the plane $H : x+y+z = 0$

After the following orthonormal change of basis:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & \sqrt{2/3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} \tag{B.21}$$

where w is chosen such that $H : w = 0$, the following function ψ_A is introduced:

$$\psi_A(u,v,w) = \mu_A \left(\sqrt{\frac{2}{3}}v + \frac{1}{\sqrt{3}}w, \frac{1}{\sqrt{2}}u - \frac{1}{\sqrt{6}}v + \frac{1}{\sqrt{3}}w, -\frac{1}{\sqrt{2}}u - \frac{1}{\sqrt{6}}v + \frac{1}{\sqrt{3}}w \right) \tag{B.22}$$

For all $(u,v,w) \neq (0,0,0)$, its partial derivative with respect to w is given by:

$$\begin{aligned} &\frac{\partial \psi_A}{\partial w}(u,v,w) \\ &= \frac{\partial}{\partial w} \sqrt[4]{\left(\sqrt{\frac{2}{3}}v + \frac{1}{\sqrt{3}}w\right)^4 + \left(\frac{1}{\sqrt{2}}u - \frac{1}{\sqrt{6}}v + \frac{1}{\sqrt{3}}w\right)^4 + \left(-\frac{1}{\sqrt{2}}u - \frac{1}{\sqrt{6}}v + \frac{1}{\sqrt{3}}w\right)^4} \\ &= \left(\frac{1}{3}w^3 + (u^2+v^2)w - \frac{\sqrt{2}}{2}u^2 v + \frac{\sqrt{2}}{6}v^3\right) \psi_A^{-3}(u,v,w) \end{aligned} \tag{B.23}$$

Since $\psi_A^{-3}(u,v,w) > 0$, studying w such that $\frac{\partial \psi_A}{\partial w}(u,v,w) = 0$ is equivalent to the study of the solutions to the depressed cubic equation :

$$X^3 + 3(u^2+v^2)X + \frac{\sqrt{2}}{2}v\left(v^2 - 3u^2\right) = 0 \tag{B.24}$$

which discriminant is given by:

$$\Delta = -\left(108(u^2+v^2)^3 + \frac{27}{2}v^2\left(v^2-3u^2\right)^2\right) \tag{B.25}$$

It is easily verified that $\Delta \leq 0$, hence there is only one real root w^* satisfying (B.24), and it is given by Cardano's formula [288]:

$$w^* = \sqrt[3]{-\frac{\sqrt{2}}{4}v\left(v^2-3u^2\right)-\sqrt{\delta(u,v)}}+\sqrt[3]{-\frac{\sqrt{2}}{4}v\left(v^2-3u^2\right)+\sqrt{\delta(u,v)}} \tag{B.26}$$

where:

$$\delta(u,v) = \frac{1}{8}v^2\left(v^2-3u^2\right)^2+\left(u^2+v^2\right)^3 \tag{B.27}$$

Finally, Theorem B.3.1 provides that the projection of A onto H is given by the $(u,v) \in \mathbb{R}^2$ satisfying:

$$\psi_A\left(u, v, \sqrt[3]{-\frac{\sqrt{2}}{4}v\left(v^2-3u^2\right)-\sqrt{\delta(u,v)}}+\sqrt[3]{-\frac{\sqrt{2}}{4}v\left(v^2-3u^2\right)+\sqrt{\delta(u,v)}}\right) \leq 1 \tag{B.28}$$

which is plotted in Figure B.7.

Appendix C

Résumé détaillé en français

La théorie du contrôle est un domaine à l'intersection des mathématiques et de l'ingénierie, axé sur la compréhension et la manipulation du comportement des systèmes dynamiques capables de recevoir des entrées et pour lesquels les signaux de sortie peuvent être mesurés. Cette discipline a reçu une grande attention au cours du siècle dernier en raison de ses nombreuses applications, car elle joue un rôle crucial en génie électrique, mécanique et chimique [45, 184, 12], en robotique [182], en aéronautique [131]... Elle a également été appliquée à l'économie et à la finance [172, 232], à la biologie [134] ou même au domaine récemment très actif des grands modèles de langage (LLM, pour Large Language Models) [37]. La théorie du contrôle s'intéresse principalement à la conception de procédures génériques permettant d'obtenir des lois de contrôle, c'est-à-dire des signaux d'entrée soigneusement construits pour régir le comportement d'un système dynamique. Les systèmes étudiés par la théorie du contrôle sont généralement modélisés à l'aide d'équations différentielles ou d'équations aux différences, comme par exemple dans la représentation d'état suivante :

$$\delta x(t) = f(x(t), u(t), w(t), t) \tag{C.1a}$$
$$y(t) = h_1(x(t), u(t), w(t), t) \tag{C.1b}$$
$$z(t) = h_2(x(t), u(t), w(t), t) \tag{C.1c}$$

- $x(t) \in \mathbb{R}^{n_x}$ est l'état (interne) du système ;
- $u(t) \in \mathbb{R}^{n_u}$ est l'entrée du système pour laquelle une loi de contrôle est généralement conçue ;
- $w(t) \in \mathbb{R}^{n_w}$ est un vecteur de signaux d'entrées exogènes, comprenant généralement des perturbations et des signaux de référence ;
- $y(t) \in \mathbb{R}^{n_y}$ est le vecteur des sorties mesurées du système ;
- $z(t) \in \mathbb{R}^{n_e}$ est l'erreur de suivi des quantités contrôlées (par rapport à leur référence), également appelée vecteur des sorties régulées ;

L'équation (C.1a) est la partie dynamique du système, avec δ l'opérateur de décalage, défini par :

- $\delta x(t) = \dot{x}(t)$ dans le cas du temps continu (et $t \in \mathbb{R}$), ce qui conduit à une équation différentielle ;
- $\delta x(t) = x(t+1)$ dans le cas du temps discret (et $t \in \mathbb{Z}$), ce qui conduit à une équation aux différences.

Ce manuscrit étudie principalement les classes spéciales de systèmes non linéaires connues sous le nom de modèles T-S (Takagi-Sugeno) et LPV (Linéaire à Paramètres Variants). Dans ces deux cadres, les défis habituels du contrôle non linéaire sont abordés en utilisant une réécriture convexe du système non linéaire (C.1). Cette réécriture, qui peut être locale ou globale, conduit à une sorte de modèle d'espace d'état LTV (Linéaire Temps-Variant) dans lequel la dépendance temporelle n'est pas explicitement obtenue, mais est implicitement représentée à l'aide de signaux extrinsèques et intrinsèques au système, rassemblés dans un *vecteur d'ordonnancement* $\theta \in \Theta$, généralement considéré comme connu ou estimé en temps-réel. En termes pratiques, ces systèmes peuvent être représentés à l'aide de la représentation d'état suivante :

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} = \begin{pmatrix} A(\theta(t)) & B_1(\theta(t)) & B_2(\theta(t)) \\ C_1(\theta(t)) & D_{11}(\theta(t)) & D_{12}(\theta(t)) \\ C_2(\theta(t)) & D_{21}(\theta(t)) & D_{22}(\theta(t)) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} \tag{C.2}$$

En tirant parti des propriétés de convexité du modèle ci-dessus, parfois avec des hypothèses sur le taux de variation de θ, les cadres T-S et LPV parviennent à obtenir des résultats qui ne sont pas très éloignés du cadre linéaire habituel. Dans un certain sens, les cadres T-S et LPV affinent les résultats classiquement obtenues pour les inclusions différentielles linéaire (LDI), puisque le fait d'ignorer les informations fournies par le vecteur d'ordonnancement θ dans (C.2) conduit à l'inclusion différentielle linéaire suivante :

$$\begin{pmatrix} \delta x(t) \\ y(t) \\ z(t) \end{pmatrix} \in \left\{ \begin{pmatrix} A(\theta) & B_1(\theta) & B_2(\theta) \\ C_1(\theta) & D_{11}(\theta) & D_{12}(\theta) \\ C_2(\theta) & D_{21}(\theta) & D_{22}(\theta) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ w(t) \end{pmatrix} : \theta \in \Theta \right\} \tag{C.3}$$

Ce manuscrit développe une variété de résultats pour ces systèmes sur un large éventail de sujets, par exemple la modélisation des systèmes, le contrôle par ordonnancement des gains, ou le diagnostic des défauts. Les principales contributions de cette thèse portent sur l'introduction d'outils géométriques qui n'avaient pas été utilisés auparavant dans le contexte des systèmes T-S et LPV. En particulier :

- Il est montré que les coordonnées barycentriques jouent un rôle clé dans la modélisation des systèmes T-S par l'approche par secteur non linéaire (*Chapitre 3*).
- Il est montré que des complexes polyédraux peuvent également être utilisés pour obtenir des modèles T-S non convexes (*Chapitre 4*).
- Il est établi que les interpolations de Bézier permettent d'obtenir une compréhension géométrique des multisommes du cadre T-S (*Chapitre 5*).
- Il est aussi démontré qu'une hypothèse de Lipschitz sur le vecteur d'ordonnancement d'un système LPV borne toutes les matrices de transition d'état que l'on peut obtenir dans le futur, ce qui conduit à des résultats utiles pour caractériser le futur proche de ces systèmes (*Chapitre 6*).
- Une approche ensembliste est explorée pour la détection de défauts, en utilisant la fonctionnelle de Minkowski d'un ensemble (*Chapitre 7*).
- Enfin, la modélisation des saturations ainsi que d'autres phénomènes (zones mortes, hystérésis) affectant localement les actionneurs d'un système est abordée à l'aide d'outils géométriques, comme la fonctionnelle de Minkowski (*Chapitre 8*).

Le contenu de chaque chapitre de la thèse est résumé plus en détail ci-après :

Chapitre 1 : Introduction générale
Le domaine du contrôle est présenté de manière générique. Ce chapitre insiste en particulier sur la représentation d'état des systèmes non-linéaires et linéaires, en évoquant l'existence des systèmes par inclusion différentielle. Les définitions classiques de la stabilité et le formalisme Lyapunov sont rappelés. Le design d'un contrôleur linéaire basé observateur est également abordé en fin de chapitre, avant l'annonce du plan général du manuscript.

Chapitre 2 : Introduction au formalisme Takagi-Sugeno
Le cadre T-S est présenté plus en détail. En particulier, ce chapitre énumère les résultats classiques concernant les inégalités matricielles linéaires (LMI), et fournit notamment les conditions habituelles de stabilité et de stabilisation des modèles T-S, à la fois dans le cas du temps continu et du temps discret.

Chapitre 3 : Modélisation convexe des systèmes de Takagi-Sugeno
Un modèle T-S est généralement dérivé d'un système non linéaire par le biais d'une méthodologie connue sous le nom d'approche par secteur non linéaire. Cette méthodologie fournit des modèles T-S qui représentent exactement les systèmes non linéaires avec des non-linéarités bornées. Cependant, cette méthodologie suppose que ces non-linéarités sont spécifiquement bornées par une boîte. Dans ce chapitre, il est révélé que l'approche par secteur non linéaire s'appuie sur des coordonnées barycentriques. Les coordonnées barycentriques étant déjà largement étudiées, il est possible d'introduire de la flexibilité dans la forme de l'ensemble délimité, en lui permettant d'être pris dans une large classe de formes convexes. Cette flexibilité a de fortes implications à la fois en termes de complexité du modèle T-S résultant ainsi que de son conservatisme intrinsèque. Ce chapitre a été publié sous forme d'article dans [20].

Chapitre 4: Modélisation non-convexes des systèmes de Takagi-Sugeno
Ce chapitre revient sur l'approche par secteur non linéaire introduite au chapitre 3 et l'applique par morceaux, ce qui conduit à l'introduction de modèles T-S non convexes. Ces modèles doivent faire l'objet d'une attention particulière si l'on veut en tirer des conditions LMI non conservatives, car il est facile de rendre leur non-convexité inutile. Des conditions LMI de stabilité bénéficiant pleinement de la non-convexité de ces modèles sont fournies dans ce chapitre.

Chapitre 5: Les interpolations de Bézier dans le formalisme Takagi-Sugeno
Les sommes multiples sont omniprésentes dans le cadre de T-S. Cependant, il est bien connu que la plupart des termes de ces sommes multiples sont redondants. Ce chapitre propose de réécrire explicitement les sommes multiples de manière non redondante en utilisant des polynômes de Bernstein. Cette idée simple met en lumière certaines propriétés géométriques sous-jacentes au cadre T-S, qui repose en fait fortement sur les schémas d'interpolation de Bézier. Certains résultats du cadre T-S sont revisités sous cette nouvelle perspective, y compris des résultats déjà publiés par l'auteur sous la forme d'un papier de conférence [18].

Chapitre 6: Anticiper le futur proche d'un système LPV
Les modèles LPV diffèrent des modèles LTV par leur dépendance temporelle qui n'est pas *explicite*, mais *implicite*, car obtenue par l'intermédiaire d'un vecteur d'ordonnancement θ. Cette différence essentielle est malheureusement suffisamment importante pour rendre la plupart des résultats du cadre LTV inapplicables dans un cadre LPV. Cependant, en supposant un taux de variation limité de θ, il est possible de quantifier l'écart maximal entre un avenir anticipé pour θ et son évolution future réelle, ce qui conduit à des résultats similaires à LTV

pour l'avenir proche des modèles LPV. Après avoir présenté deux outils puissants, à savoir l'intégrale multiplicative de Volterra et la norme logarithmique pondérée d'une matrice, ce chapitre propose une méthodologie pour discrétiser *exactement* les modèles LPV, ainsi que pour anticiper l'évolution de certaines propriétés structurelles, telles que la contrôlabilité et l'observabilité. Certains résultats de ce chapitre sont adaptés de papiers de conférence publiés par l'auteur [17, 24].

Chapitre 7: Localisation des défauts par approche ensembliste
La détection et la localisation des défauts consistent à détecter si un système est sujet à un défaut (la détection), ainsi qu'à identifier quel défaut est actif (la localisation). On y parvient généralement en synthétisant des signaux, appelé résidus, via la mesure de l'entrée et de la sortie du système. En l'absence de défaut, les valeurs des résidus devraient être centrées sur 0. Cependant, tout système est sujet à de légères perturbations, et un bon système de détection et de localisation des défauts doit être robuste aux bruits bénins affectant le système et se répercutant sur les valeurs mesurées des résidus. La détection et la localisation des défauts par approche ensembliste consiste à traiter les perturbations du système comme des entrées bornées, ce qui permet de construire des seuils de détection sur les signaux résiduels. L'identification d'un défaut à l'aide de cette méthodologie est généralement réalisée en utilisant des classes spécifiques d'ensembles convexes (par exemple, les zonotopes ou les ellipsoïdes). Plutôt que de se concentrer sur de telles classes, ce chapitre décrit en termes abstraits un schéma de détection et de localisation des défauts ensembliste pour les systèmes linéaires incertains. Ce chapitre introduit la fonction de Minkowski d'un ensemble pour obtenir des résultats théoriques qui peuvent être appliqués à une large classe de méthodologies ensemblistes de la littérature.

Chapitre 8: Modélisation des phénomènes de saturations et de zone-mortes
Les saturations, les zones mortes, les bandes mortes ou les effets d'hystérésis sont des défauts usuels affectant les actionneurs et les capteurs d'un système. Ce chapitre présente des méthodologies permettant de les modéliser et d'obtenir des garanties théoriques sur le comportement d'un système sous leur influence.

- Une large classe de saturations d'actionneurs est modélisée dans le cadre T-S, avec un nombre réduit de modèles locaux par rapport à la littérature existante. Cet objectif est atteint en tirant parti de la fonction de Minkowski introduite au chapitre 7. Cette partie du chapitre a été publiée sous la forme d'un papier de conférence dans [21].
- Une modélisation unifiée des zones mortes, des bandes mortes et des effets d'hystérésis est proposée et étudiée pour les systèmes linéaire temps invariant (LTI). Des résultats asymptotiques de majoration des normes des trajectoires d'un système en présence de ces défauts sont obtenus. Cette partie du chapitre a été publiée sous la forme d'un papier de conférence dans [23].

Chapitre 9: Conclusions et perspectives
Ce chapitre conclut la thèse et offre quelques perspectives pour les travaux futurs.

Appendice A: Critère graphique pour les fonctions de Lyapunov quadratiques communes aux systèmes du second ordre
Un critère graphique très simple est proposé dans cette annexe pour obtenir une fonction de Lyapunov quadratique commune à un ensemble de systèmes LTI du second ordre à coefficients réels. En termes simples, le critère associe chaque matrice Hurwitz (resp. Schur) de $\mathbb{R}^{2\times 2}$

à l'intérieur d'une ellipse sur un plan à deux dimensions. Si l'intersection de toutes les ellipses associées à l'ensemble des matrices Hurwitz (resp. Schur) est non vide, on peut affirmer sans perte de généralité qu'il existe une fonction de Lyapunov quadratique commune à cet ensemble. Toutes les fonctions de Lyapunov quadratiques communes existantes peuvent en fait être extraites de cette intersection.

Appendice B: Projections d'ensembles convexes de frontière $\mathcal{C}^1$
Étant donné un espace Euclidien, cet appendice élucide le lien topologique entre les dérivées partielles de la fonctionnelle de Minkowski associée à un ensemble convexe ayant une frontière $\mathcal{C}^1$ et un intérieur non vide, et la frontière de sa projection orthogonale sur les sous-espaces linéaires de l'espace euclidien. Un système d'équations pour ces projections orthogonales est obtenu à partir de ce lien topologique. Cet appendice correspond au dépôt arXiv [22].

Liste des publications

Revues internationales avec comité de lecture

- G. Bainier, B. Marx, J.-C. Ponsart (2024). Common Quadratic Lyapunov Functions for Sets of Second-Order Linear Systems: A Simple Graphical Criterion. *IEEE Control Systems Letters*. https://doi.org/10.1109/LCSYS.2024.3418672
- G. Bainier, B. Marx, J.-C. Ponsart (2024). Generalized nonlinear sector approaches for Takagi-Sugeno models. *Fuzzy Sets and Systems, 476*. https://doi.org/10.1016/j.fss.2023.108791

Travail en cours d'évaluation

- G. Bainier, B. Marx, J.-C. Ponsart. A Unified Set-Membership Approach to Fault Isolation using the Minkowski functional. *International Journal of Robust and Nonlinear Control.*

Conférences nationales et internationales avec comité de lecture

- G. Bainier, B. Marx, J.-C. Ponsart (2024). Bézier Controllers and Observers for Takagi-Sugeno Models. *2024 American Control Conference, Toronto, Canada.* https://doi.org/10.23919/ACC60939.2024.10644414
- G. Bainier, B. Marx, J.-C. Ponsart (2024). Modeling a broad class of actuator saturations using Takagi-Sugeno models with a reduced number of local models. *2024 American Control Conference, Toronto, Canada.* https://doi.org/10.23919/ACC60939.2024.10644172
- G. Bainier, B. Marx, J.-C. Ponsart (2024). A unified modeling of dead-zone, dead-band, hysteresis, and other faulty local behaviors of actuators and sensors. *Safeprocess 2024, Ferrara, Italy.* https://doi.org/10.1016/j.ifacol.2024.07.298
- G. Bainier, J.-C. Ponsart, B. Marx (2022). Anticipating the loss of unknown input observability for sampled LPV systems. *16th European Workshop on Advanced Control and Diagnosis, ACD 2022, Nancy, France.* https://doi.org/10.1007/978-3-031-27540-1_2
- G. Bainier, B. Marx, J.-C. Ponsart (2022). Bounding the Trajectories of Continuous-Time LPV Systems with Parameters known in Real Time. *5th IFAC Workshop on Linear Parameter Varying Systems, LPVS 2022, Montréal, Canada.* https://doi.org/10.1016/j.ifacol.2022.11.292

Dépôt arXiv

- G. Bainier, B. Marx, J.-C. Ponsart (2023). Orthogonal Projection of Convex Sets with a Differentiable Boundary. https://doi.org/10.48550/arXiv.2302.08937

Bibliography

[1] J. Abonyi, R. Babuska, and F. Szeifert. Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 32(5):612–621, Oct. 2002.

[2] O. Adrot, D. Maquin, and J. Ragot. Bounding approach to fault detection of uncertain dynamic systems. *IFAC Proceedings Volumes*, 33(11):1131–1136, June 2000.

[3] S. N. Afriat. Theory of maxima and the method of Lagrange. *SIAM Journal on Applied Mathematics*, 20(3):343–357, May 1971.

[4] M. Akar, A. Paul, M. Safonov, and U. Mitra. Conditions on the stability of a class of second-order switched systems. *IEEE Transactions on Automatic Control*, 51(2):338–340, Feb. 2006.

[5] C. D. Aliprantis and K. C. Border. *Topological vector spaces. In: Infinite Dimensional Analysis*, page 161–236. Springer Berlin Heidelberg, 1999.

[6] M. Althoff, O. Stursberg, and M. Buss. Computing reachable sets of hybrid systems using a combination of zonotopes and polytopes. *Nonlinear Analysis: Hybrid Systems*, 4(2):233–249, May 2010.

[7] F. Amato. *Robust control of linear systems subject to uncertain time-varying parameters*. Lecture Notes in Control and Information Sciences. Springer, Berlin, Germany, 2006 edition, Mar. 2005.

[8] G. Angelis. *System analysis, modelling and control with polytopic linear models*. PhD thesis, Technische Universiteit Eindhoven, 2001.

[9] F. Anstett, G. Millérioux, and G. Bloch. Polytopic observer design for LPV systems based on minimal convex polytope finding. *Journal of Algorithms & Computational Technology*, 3(1):23–43, Mar. 2009.

[10] P. Apkarian, P. Gahinet, and G. Becker. Self-scheduled H_∞ control of linear parameter-varying systems: a design example. *Automatica*, 31(9):1251–1261, Sept. 1995.

[11] C. Arino and A. Sala. Design of multiple-parameterisation PDC controllers via relaxed conditions for multi-dimensional fuzzy summations. In *2007 IEEE International Fuzzy Systems Conference*. IEEE, June 2007.

[12] K. J. Astrom and R. M. Murray. *Feedback systems: An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ, Apr. 2008.

[13] H. Atoui, O. Sename, V. Milanes, and J. J. Martinez Molina. Advanced LPV-YK control design with experimental validation on autonomous vehicles. working paper or preprint, Apr. 2022.

[14] H. Atoui, O. Sename, V. Milanes, and J.-J. Martinez-Molina. Toward switching/interpolating LPV control: A review. *Annual Reviews in Control*, 54:49–67, 2022.

[15] T. Azuma, R. Watanabe, K. Uchida, and M. Fujita. A new LMI approach to analysis of linear systems depending on scheduling parameter in polynomial forms. *auto*, 48(4):199, Apr. 2000.

[16] M. Baake and U. Schlägel. The Peano-Baker series. *Proceedings of the Steklov Institute of Mathematics*, 275(1):155–159, Dec. 2011.

[17] G. Bainier, B. Marx, and J.-C. Ponsart. Bounding the trajectories of continuous-time LPV systems with parameters known in real time. *IFAC-PapersOnLine*, 55(35):67–72, 2022.

[18] G. Bainier, B. Marx, and J.-C. Ponsart. Bézier controllers and observers for Takagi-Sugeno models. In *2024 American Control Conference (ACC)*, page 4753–4758. IEEE, July 2024.

[19] G. Bainier, B. Marx, and J.-C. Ponsart. Common quadratic Lyapunov functions for sets of second-order linear systems: A simple graphical criterion. *IEEE Control Systems Letters*, 8:1889–1894, 2024.

[20] G. Bainier, B. Marx, and J.-C. Ponsart. Generalized nonlinear sector approaches for Takagi-Sugeno models. *Fuzzy Sets and Systems*, 476:108791, Jan. 2024.

[21] G. Bainier, B. Marx, and J.-C. Ponsart. Modelling a broad class of actuator saturations using takagi-sugeno models with a reduced number of local models. In *2024 American Control Conference (ACC)*, page 1701–1706. IEEE, July 2024.

[22] G. Bainier, B. Marx, and J.-C. Ponsart. Orthogonal projection of convex sets with a differentiable boundary. *arXiv*, 2024.

[23] G. Bainier, B. Marx, and J.-C. Ponsart. A unified modelling of dead-zone, dead-band, hysteresis, and other faulty local behaviors of actuators and sensors. *IFAC-PapersOnLine*, 58(4):682–687, 2024.

[24] G. Bainier, J.-C. Ponsart, and B. Marx. Anticipating the loss of unknown input observability for sampled LPV systems. In *European Workshop on Advanced Control and Diagnosis (ACD)*, page 11–21. Springer Nature Switzerland, 2022.

[25] P. Baranyi. The generalized TP model transformation for T-S fuzzy model manipulation and generalized stability verification. *IEEE Transactions on Fuzzy Systems*, 22(4):934–948, Aug. 2014.

[26] P. Bauer, K. Premaratne, and J. Duran. A necessary and sufficient condition for robust asymptotic stability of time-variant discrete systems. *IEEE Transactions on Automatic Control*, 38(9):1427–1430, 1993.

[27] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, Nov. 1998.

[28] A. Benavoli and D. Piga. A probabilistic interpretation of set-membership filtering: Application to polynomial systems through polytopic bounding. *Automatica*, 70:158–172, Aug. 2016.

[29] M. Bernal and T. M. Guerra. Generalized nonquadratic stability of continuous-time Takagi-Sugeno models. *IEEE Transactions on Fuzzy Systems*, 18(4):815–822, Aug. 2010.

[30] M. Bernal, T. M. Guerra, and A. Kruszewski. A membership-function-dependent approach for stability analysis and controller synthesis of Takagi-Sugeno models. *Fuzzy Sets and Systems*, 160(19):2776–2795, Oct. 2009.

[31] M. Bernal, R. Marquez, V. Estrada-Manzo, and B. Castillo-Toledo. Nonlinear output regulation via Takagi-Sugeno fuzzy mappings: A full-information LMI approach. In *2012 IEEE International Conference on Fuzzy Systems*. IEEE, June 2012.

[32] M. Bernal, A. Sala, Z. Lendek, and T. M. Guerra. *Analysis and Synthesis of Nonlinear Control Systems: A Convex Optimisation Approach*. Springer International Publishing, 2022.

[33] D. Bertsekas and I. Rhodes. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, 16(2):117–128, Apr. 1971.

[34] T. Besselmann, J. Löfberg, and M. Morari. Explicit LPV-MPC with bounded rate of parameter variation. *IFAC Proceedings Volumes*, 42(6):7–12, 2009.

[35] S. Bezzaoucha, B. Marx, D. Maquin, and J. Ragot. State and parameter estimation for time-varying systems: a Takagi-Sugeno approach. *IFAC Proceedings Volumes*, 46(2):761–766, 2013.

[36] S. Bezzaoucha, B. Marx, D. Maquin, and J. Ragot. State and output feedback control for Takagi-Sugeno systems with saturated actuators. *International Journal of Adaptive Control and Signal Processing*, 30(6):888–905, nov 2015.

[37] A. Bhargava, C. Witkowski, M. Shah, and M. Thomson. What's the magic word? a control theory of LLM prompting. *arXiv*, 2024.

[38] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics N°169. Springer, 1 edition, 1997.

[39] T. J. Bird, H. C. Pangborn, N. Jain, and J. P. Koeln. Hybrid zonotopes: A new set representation for reachability analysis of mixed logical dynamical systems. *Automatica*, 154:111107, Aug. 2023.

[40] F. Blanchini and S. Miani. Convex sets and their representation. In *Set-Theoretic Methods in Control*, pages 93–119. Springer International Publishing, 2015.

[41] J. Blesa, V. Puig, and J. Saludes. Robust fault detection using polytope-based set-membership consistency test. In *2010 Conference on Control and Fault-Tolerant Systems (SysTol)*, page 726–731. IEEE, Oct. 2010.

[42] J. Blesa, V. Puig, and J. Saludes. Identification for passive robust fault detection using zonotope-based set-membership approaches. *International Journal of Adaptive Control and Signal Processing*, 25(9):788–812, Apr. 2011.

[43] J. Blesa, V. Puig, J. Saludes, and R. M. Fernández-Cantí. Set-membership parity space approach for fault detection in linear uncertain dynamic systems. *International Journal of Adaptive Control and Signal Processing*, 30(2):186–205, Mar. 2014.

[44] N. Bof, R. Carli, and L. Schenato. Lyapunov theory for discrete time systems. *arXiv*, 2018.

[45] W. Bolton. *Mechatronics: Electronic control systems in mechanical and electrical engineering*. Pearson Education, London, England, 6 edition, Feb. 2015.

[46] I. Bongartz. *Over-approximative Reduction of Polytopes in the Context of Hybrid Systems Reachability Analysis*. Bachelor's thesis, RWTH Aachen University, 2016.

[47] U. Boscain and M. Sigalotti. Introduction to controllability of non-linear systems. In S. Dipierro, editor, *Contemporary Research in Elliptic PDEs and Related Topics*. Springer, 2019.

[48] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.

[49] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, England, Mar. 2004.

[50] P. Braun, L. Grune, and C. M. Kellett. *(in-)stability of differential inclusions*. SpringerBriefs in mathematics. Springer Nature, Cham, Switzerland, 1 edition, July 2021.

[51] A. Bregon. *Integration of FDI and DX Techniques within Consistency-based Diagnosis with Possible Conflicts*. PhD thesis, Universidad de Valladolid, 05 2010.

[52] C. Briat. *Linear Parameter-Varying and Time-Delay Systems: Analysis, Observation, Filtering & Control*. Springer Berlin Heidelberg, 2015.

[53] E. M. Bronstein. Approximation of convex sets by polytopes. *Journal of Mathematical Sciences*, 153(6):727–762, Sept. 2008.

[54] S. Broschart and D. J. Scheeres. On the implementation of spacecraft hovering under reduced-order dead-band control. In *Journal of Guidance, Control, and Dynamics*, 08 2007.

[55] H. Bruus and K. Flensburg. *Many-body quantum theory in condensed matter physics*. Oxford Graduate Texts. Oxford University Press, London, England, Sept. 2004.

[56] V. Cachia. *La formule de Trotter-Kato : approximation des semi-groupes en normes d'opérateur et de trace*. Theses, Université de la Méditerranée - Aix-Marseille II, Apr. 2001.

[57] J. W. Cannon, W. J. Floyd, R. Kenyon, W. R. Parry, et al. Hyperbolic geometry. *Flavors of geometry*, 31(59-115):2, 1997.

[58] Y.-Y. Cao and Z. Lin. Robust stability analysis and fuzzy-scheduling control for nonlinear systems subject to actuator saturation. *IEEE Transactions on Fuzzy Systems*, 11(1):57–67, feb 2003.

[59] M. Carter. *Foundations of mathematical economics*. MIT press, 2001.

[60] M. Chadli and T. M. Guerra. LMI solution for robust static output feedback control of discrete Takagi–Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, 20(6):1160–1165, Dec. 2012.

[61] M. Chadli, D. Maquin, and J. Ragot. Nonquadratic stability analysis of Takagi-Sugeno models. In *Proceedings of the 41st IEEE Conference on Decision and Control*. IEEE, 2002.

[62] C.-T. Chen. *Linear System Theory and Design*. The Oxford Series in Electrical and Computer Engineering. Oxford University Press, New York, NY, 3 edition, Aug. 1998.

[63] J. Chen and R. J. Patton. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer US, 1999.

[64] J. Chen, S. Xu, B. Zhang, Y. Chu, and Y. Zou. New relaxed stability and stabilization conditions for continuous-time T-S fuzzy models. *Information Sciences*, 329:447–460, Feb. 2016.

[65] A. Cherifi, K. Guelton, and L. Arcese. Quadratic design of D-stabilizing non-PDC controllers for quasi-LPV/T-S models. *IFAC-PapersOnLine*, 48(26):164–169, 2015.

[66] G. Chesi, A. Garulli, A. Tesi, and A. Vicino. Polynomially parameter-dependent Lyapunov functions for robust stability of polytopic systems: an LMI approach. *IEEE Transactions on Automatic Control*, 50(3):365–370, Mar. 2005.

[67] M. Chilali, P. Gahinet, and P. Apkarian. Robust pole placement in lmi regions. *IEEE Transactions on Automatic Control*, 44(12):2257–2270, 1999.

[68] E. Chow and A. Willsky. Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control*, 29(7):603–614, July 1984.

[69] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth analysis and control theory*, volume 178. Springer Science & Business Media, 2008.

[70] C. Combastel. An extended zonotopic and Gaussian Kalman filter (EZGKF) merging set-membership and stochastic paradigms: Toward non-linear filtering and fault detection. *Annual Reviews in Control*, 42:232–243, 2016.

[71] C. Combastel, V. Puig, T. Raïssi, and T. Alamo. Set-membership methods applied to FDI and FTC. *International Journal of Adaptive Control and Signal Processing*, 30(2):150–153, Jan. 2016.

[72] G. Dahlquist. *Stability and error bounds in the numerical integration of ordinary differential equations*. Kungl. tekniska hogskolans Handlingar. Almqvist & Wiksells, Uppsala, 1959.

[73] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Apr. 2002.

[74] J. de Kleer, A. K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56(2–3):197–222, Aug. 1992.

[75] J. Deckert, M. Desai, J. Deyst, and A. Willsky. F-8 DFBW sensor failure identification using analytic redundancy. *IEEE Transactions on Automatic Control*, 22(5):795–803, Oct. 1977.

[76] C. Desoer. Slowly varying system $\dot{x} = A(t)x$. *IEEE Transactions on Automatic Control*, 14(6):780–781, Dec. 1969.

[77] C. Desoer and H. Haneda. The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. *IEEE Transactions on Circuit Theory*, 19(5):480–486, 1972.

[78] D. Ding, R. Cooper, and D. Spaeth. Optimized joystick controller. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2004.

[79] S. X. Ding. *Model-Based Fault Diagnosis Techniques*. Springer London, 2013.

[80] J. D. Dollard and C. N. Friedman. *Product Integration with Application to Differential Equations*. Cambridge University Press, Dec. 1984.

[81] F. Duhem. Die dauernden Aenderungen und die Thermodynamik. *Zeitschrift für Physikalische Chemie*, 24U(1):666–666, Sept. 1897.

[82] D. Efimov and T. Raïssi. Design of interval observers for uncertain dynamical systems. *Automation and Remote Control*, 77(2):191–225, Feb. 2016.

[83] L. P. Eisenhart. *A treatise on the differential geometry of curves and surfaces*. Ginn, 1909.

[84] A. El-Amrani, B. Boukili, A. E. Hajjaji, and A. Hmamed. H∞ model reduction for T-S fuzzy systems over finite frequency ranges. *Optimal Control Applications and Methods*, 39(4):1479–1496, Apr. 2018.

[85] L. J. Elias, F. A. Faria, R. Araujo, and V. A. Oliveira. Stability analysis of Takagi-Sugeno systems using a switched fuzzy Lyapunov function. *Information Sciences*, 543:43–57, Jan. 2021.

[86] V. Estrada Manzo. *Estimation and control of descriptor systems*. PhD thesis, Université de Valenciennes et du Hainaut-Cambresis, Oct. 2015.

[87] C.-H. Fang, Y.-S. Liu, S.-W. Kau, L. Hong, and C.-H. Lee. A new LMI-based approach to relaxed quadratic stabilization of T-S fuzzy control systems. *IEEE Transactions on Fuzzy Systems*, 14(3):386–397, June 2006.

[88] F. A. Faria, G. N. Silva, and V. A. Oliveira. Reducing the conservatism of LMI-based stabilisation conditions for TS fuzzy systems using fuzzy Lyapunov functions. *International Journal of Systems Science*, 44(10):1956–1969, Oct. 2013.

[89] G. Farin. Triangular Bernstein-Bézier patches. *Computer Aided Geometric Design*, 3(2):83–127, Aug. 1986.

[90] G. Farin. Introductory material. In *Curves and Surfaces for CAGD*. Elsevier, 2002.

[91] A. Fellhauer. Approximation of smooth functions using Bernstein polynomials in multiple variables. *arXiv*, 2016.

[92] G. Feng. Stability analysis of discrete-time fuzzy dynamic systems based on piecewise Lyapunov functions. *IEEE Transactions on Fuzzy Systems*, 12(1):22–28, Feb. 2004.

[93] R. P. Feynman and A. R. Hibbs. *Quantam mechanics and path integrals*. Dover Books on Physics. Dover Publications, Mineola, NY, July 2010.

[94] M. Fiacchini, C. Prieur, and S. Tarbouriech. On the computation of set-induced control Lyapunov functions for continuous-time systems. *SIAM Journal on Control and Optimization*, 53(3):1305–1327, Jan. 2015.

[95] A. F. Filippov. Differential equations whose right-hand side is discontinuous on intersecting surfaces. *Differentsial'nye Uravneniya*, 15(10):1814–1823, 1979.

[96] M. Foupouagnigni and M. Mouafo Wouodjié. On multivariate Bernstein polynomials. *Mathematics*, 8(9):1397, Aug. 2020.

[97] P. M. Frank. Fault diagnosis in dynamic systems via state estimation - a survey. In S. G. Tzafestas S, Singh M, editor, *System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches*, page 35–98. Springer Netherlands, 1987.

[98] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. *Automatica*, 26(3):459–474, May 1990.

[99] J. Gallier. *Schur Complements and Applications. In: Geometric Methods and Applications*, page 431–437. Springer New York, 2011.

[100] S. Gentil, J. Montmain, and C. Combastel. Combining FDI and AI approaches within causal-model-based diagnosis. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 34(5):2207–2221, Oct. 2004.

[101] J. Gertler. Analytical redundancy methods in fault detection and isolation - survey and synthesis. *IFAC Proceedings Volumes*, 24(6):9–21, Sept. 1991.

[102] J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. CRC Press, Nov. 2017.

[103] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley. KKT conditions, first-order and second-order optimization, and distributed optimization: Tutorial and survey. *arXiv*, 2021.

[104] M. Gianino. Adaptive control of systems with dead zones. *Master Thesis, Department of Automatic Control - Lund Institute of Technology*, 1994.

[105] P.-L. Giscard, K. Lui, S. J. Thwaite, and D. Jaksch. An exact formulation of the time-ordered exponential using path-sums. *Journal of Mathematical Physics*, 56(5), May 2015.

[106] M. Goel, A. Maciejewski, V. Balakrishnan, and R. Proctor. Failure tolerant teleoperation of a kinematically redundant manipulator: an experimental study. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 33(6):758–765, Nov. 2003.

[107] R. Goldman. Curvature formulas for implicit curves and surfaces. *Computer Aided Geometric Design*, 22(7):632–658, Oct. 2005.

[108] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 3 edition, Oct. 1996.

[109] T. H. Gronwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *The Annals of Mathematics*, 20(4):292, July 1919.

[110] B. Grünbaum. *Convex Polytopes*. Springer New York, 2003.

[111] K. Guelton, T. Bouarar, and N. Manamanni. Robust dynamic output feedback fuzzy Lyapunov stabilization of Takagi–Sugeno systems—a descriptor redundancy approach. *Fuzzy Sets and Systems*, 160(19):2796–2811, Oct. 2009.

[112] K. Guelton, N. Manamanni, C.-C. Duong, and D. Koumba-Emianiwe. Sum-of-squares stability analysis of Takagi-Sugeno systems based on multiple polynomial Lyapunov functions. *International Journal of Fuzzy Systems*, 15:1–8, 03 2013.

[113] T.-M. Guerra and M. Bernal. A way to escape from the quadratic framework. In *2009 IEEE International Conference on Fuzzy Systems*. IEEE, Aug. 2009.

[114] T. M. Guerra, M. Bernal, K. Guelton, and S. Labiod. Non-quadratic local stabilization for continuous-time Takagi–Sugeno models. *Fuzzy Sets and Systems*, 201:40–54, Aug. 2012.

[115] T. M. Guerra, A. Jaadari, J. Pan, and A. Sala. Some refinements for non quadratic stabilization of continuous TS models. In *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. IEEE, June 2011.

[116] T. M. Guerra, A. Kruszewski, and J. Lauber. Discrete Tagaki–Sugeno models for control: Where are we? *Annual Reviews in Control*, 33(1):37–47, Apr. 2009.

[117] T. M. Guerra, R. Marquez, A. Kruszewski, and M. Bernal. H_∞ LMI-based observer design for nonlinear systems via Takagi–Sugeno models with unmeasured premise variables. *IEEE Transactions on Fuzzy Systems*, 26(3):1498–1509, June 2018.

[118] S. Hafstein and P. Giesl. Review on computational methods for Lyapunov functions. *Discrete and Continuous Dynamical Systems - Series B*, 20(8):2291–2331, Aug. 2015.

[119] A. Halder. Smallest ellipsoid containing p-sum of ellipsoids with application to reachability analysis. *IEEE Transactions on Automatic Control*, 66(6):2512–2525, June 2021.

[120] V. Hassani, T. Tjahjowidodo, and T. N. Do. A survey on hysteresis modeling, identification and control. *Mechanical Systems and Signal Processing*, 49(1-2):209–233, Dec. 2014.

[121] J. Hauser, S. Sastry, and G. Meyer. Nonlinear control design for slightly non-minimum phase systems: Application to v/STOL aircraft. *Automatica*, 28(4):665–679, July 1992.

[122] D. Henrion, S. Tarbouriech, and D. Arzelier. LMI approximations for the radius of the intersection of ellipsoids. In *Proceedings of the 37th IEEE Conference on Decision and Control*. IEEE, 1998.

[123] R. Hermann and A. Krener. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, 22(5):728–740, Oct. 1977.

[124] T. Hernández-Cortés, J. A. Meda Campaña, L. A. Páramo Carranza, and J. C. Gómez Mancilla. A simplified output regulator for a class of Takagi-Sugeno fuzzy models. *Mathematical Problems in Engineering*, 2015:1–18, 2015.

[125] C. Hoffmann. *Linear parameter-varying control of systems of high complexity*. Regelungstechnik. Verlag Dr. Hut GmbH, Munich, Germany, June 2016.

[126] C. Hoffmann and H. Werner. Complexity of implementation and synthesis in linear parameter-varying control. *IFAC Proceedings Volumes*, 47(3):11749–11760, 2014.

[127] H. Hogbe-Nlend. *Bornologies and functional analysis*. North-Holland Mathematics Studies. North-Holland, Jan. 1977.

[128] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Apr. 1991.

[129] G.-D. Hu and G.-D. Hu. A relation between the weighted logarithmic norm of a matrix and the Lyapunov equation. *Bit Numerical Mathematics*, 40(3):606–610, 2000.

[130] G.-D. Hu and M. Liu. The weighted logarithmic matrix norm and bounds of the matrix exponential. *Linear Algebra and its Applications*, 390:145–154, Oct. 2004.

[131] D. G. Hull et al. *Fundamentals of airplane flight mechanics*, volume 19. Springer, 2007.

[132] D. Ichalal, B. Marx, S. Mammar, D. Maquin, and J. Ragot. How to cope with unmeasurable premise variables in Takagi–Sugeno observer design: Dynamic extension approach. *Engineering Applications of Artificial Intelligence*, 67:430–435, Jan. 2018.

[133] D. Ichalal, B. Marx, J. Ragot, and D. Maquin. State estimation of Takagi–Sugeno systems with unmeasurable premise variables. *IET Control Theory & Applications*, 4(5):897–908, May 2010.

[134] P. A. Iglesias and B. P. Ingalls, editors. *Control theory and systems biology*. MIT Press, London, England, Sept. 2009.

[135] J. Y. Ishihara, H. T. M. Kussaba, and R. A. Borges. Existence of continuous or constant Finsler's variables for parameter-dependent systems. *IEEE Transactions on Automatic Control*, 62(8):4187–4193, Aug. 2017.

[136] J. S. Jackson. *Geometrical conic sections; An elementary treatise in which the conic sections are defined as the plane sections of a cone, and treated by the method of projections*. MacMillan and co, London, England, 1872.

[137] A. Jadbabaie. A reduction in conservatism in stability and L_2 gain analysis of Takagi-Sugeno fuzzy systems via linear matrix inequalities. *IFAC Proceedings Volumes*, 32(2):5451–5455, July 1999.

[138] T. Johansen and R. Babuska. Multiobjective identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, 11(6):847–860, Dec. 2003.

[139] T. Johansen, R. Shorten, and R. Murray-Smith. On the interpretation and identification of dynamic Takagi-Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, 8(3):297–313, June 2000.

[140] M. Johansson, A. Rantzer, and K.-E. Arzen. Piecewise quadratic stability of fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 7(6):713–722, 1999.

[141] M. Jorgensen. Volumes of n-dimensional spheres & ellipsoids. 2014.

[142] L. Jottrand. *Shadow Boundaries of Convex Bodies*. Theses, University College London, 2013.

[143] M. Jungers, G. S. Deaecto, and J. C. Geromel. Bounds for the remainders of uncertain matrix exponential and sampled-data control of polytopic linear systems. *Automatica*, 82:202–208, 2017.

[144] M. Jungers, R. C. Oliveira, and P. L. Peres. MPC for LPV systems with bounded parameter variations. *International Journal of Control*, 84(1):24–36, Jan. 2011.

[145] G. Kalai. Linear programming, the simplex algorithm and simple polytopes. *Mathematical Programming*, 79(1-3):217–233, Oct. 1997.

[146] D. Kapsalis, O. Sename, V. Milanes, and J. J. Molina. A reduced LPV polytopic look-ahead steering controller for autonomous vehicles. *Control Engineering Practice*, 129:105360, Dec. 2022.

[147] I. Karafyllis and A. Chaillet. Lyapunov conditions for uniform asymptotic output stability and a relaxation of Barbălat's lemma. *Automatica*, 132:109792, Oct. 2021.

[148] S.-W. Kau, H.-J. Lee, C.-M. Yang, C.-H. Lee, L. Hong, and C.-H. Fang. Robust fuzzy static output feedback control of T-S fuzzy systems with parametric uncertainties. *Fuzzy Sets and Systems*, 158(2):135–146, Jan. 2007.

[149] H. K. Khalil. *Nonlinear Systems*. Pearson, Upper Saddle River, NJ, 3 edition, dec 2001.

[150] H. K. Khalil. *Nonlinear Control*. Pearson, Upper Saddle River, NJ, Jan. 2014.

[151] S. H. Kim and P. Park. Observer-based relaxed H_∞ control for fuzzy systems using a multiple Lyapunov function. *IEEE Transactions on Fuzzy Systems*, 17(2):477–484, Apr. 2009.

[152] C. King and M. Nathanson. On the existence of a common quadratic Lyapunov function for a rank one difference. *Linear Algebra and its Applications*, 419(2–3):400–416, Dec. 2006.

[153] K. Kiriakidis. Nonlinear modeling by interpolation between linear dynamics and its application in control. *Journal of Dynamic Systems, Measurement, and Control*, 129(6):813–824, Jan. 2007.

[154] M. Klug, E. B. Castelan, V. J. Leite, and L. F. Silva. Fuzzy dynamic output feedback control through nonlinear Takagi-Sugeno models. *Fuzzy Sets and Systems*, 263:92–111, Mar. 2015.

[155] J. W. Ko, W. I. Lee, and P. Park. Stabilization for Takagi–Sugeno fuzzy systems based on partitioning the range of fuzzy weights. *Automatica*, 48(5):970–973, May 2012.

[156] S. R. Kou, D. L. Elliott, and T. J. Tarn. Observability of nonlinear systems. *Information and Control*, 22(1):89–99, Feb. 1973.

[157] A. Kruszewski, A. Sala, T. Guerra, and C. Arino. A triangulation approach to asymptotically exact conditions for fuzzy summations. *IEEE Transactions on Fuzzy Systems*, 17(5):985–994, Oct. 2009.

[158] A. Kruszewski, R. Wang, and T. M. Guerra. Nonquadratic stabilization conditions for a class of uncertain nonlinear discrete time TS fuzzy models: A new approach. *IEEE Transactions on Automatic Control*, 53(2):606–611, Mar. 2008.

[159] A. B. Kurzhanskii. *Ellipsoidal calculus for estimation and control*. IIASA Birkhauser Boston, Laxenburg, Austria Boston, 1997.

[160] A. Kwiatkowski. *LPV modeling and application of LPV controllers to SI engines*. PhD thesis, Technischen Universitat Hamburg-Harburg, Nov. 2008.

[161] H.-K. Lam. *Polynomial Fuzzy Model-Based Control Systems*. Springer International Publishing, 2016.

[162] A. Lasota and A. Strauss. Asymptotic behavior for differential equations which cannot be locally linearized. *Journal of Differential Equations*, 10(1):152–172, July 1971.

[163] C. Le Guernic and A. Girard. Reachability analysis of linear systems using support functions. *Nonlinear Analysis: Hybrid Systems*, 4(2):250–262, May 2010.

[164] D. H. Lee and Y. H. Joo. On the generalized local stability and local stabilization conditions for discrete-time Takagi-Sugeno fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 22(6):1654–1668, Dec. 2014.

[165] D. H. Lee and D. W. Kim. Relaxed LMI conditions for local stability and local stabilization of continuous-time Takagi-Sugeno fuzzy systems. *IEEE Transactions on Cybernetics*, 44(3):394–405, Mar. 2014.

[166] Z. Lendek and J. Lauber. Local stability of discrete-time TS fuzzy systems. *IFAC-PapersOnLine*, 49(5):7–12, 2016.

[167] C. Li, J. Zhou, B. Fu, P. Kou, and J. Xiao. T-S fuzzy model identification with a gravitational search-based hyperplane clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 20(2):305–317, Apr. 2012.

[168] D. Li and Y. Xi. The feedback robust MPC for LPV systems with bounded rates of parameter changes. *IEEE Transactions on Automatic Control*, 55(2):503–507, Feb. 2010.

[169] P. Li, A.-T. Nguyen, H. Du, Y. Wang, and H. Zhang. Polytopic LPV approaches for intelligent automotive systems: State of the art and future challenges. *Mechanical Systems and Signal Processing*, 161:107931, Dec. 2021.

[170] D. Lin and M. J. Er. A new approach for stabilizing nonlinear systems with time delay. In *2001 European Control Conference (ECC)*. IEEE, Sept. 2001.

[171] H. Lin and P. J. Antsaklis. Stability and stabilizability of switched linear systems: A survey of recent results. *IEEE Transactions on Automatic Control*, 54(2):308–322, Feb. 2009.

[172] M. Lines, editor. *Nonlinear Dynamical Systems in Economics*. Springer Vienna, 2005.

[173] X. Liu and H. Zhang. Stability analysis of uncertain fuzzy large-scale system. *Chaos, Solitons & Fractals*, 25(5):1107–1122, Sept. 2005.

[174] J.-C. Lo and M.-L. Lin. Observer-based robust H_∞ control for fuzzy systems using two-step procedure. *IEEE Transactions on Fuzzy Systems*, 12(3):350–359, June 2004.

[175] K.-G. Löfgren. On envelope theorems in economics: Inspired by a revival of a forgotten lecture. *Research Papers in Economics*, 2011.

[176] A. N. D. Lopes, V. J. S. Leite, L. F. P. Silva, and K. Guelton. Anti-windup TS fuzzy PI-like control for discrete-time nonlinear systems with saturated actuators. *International Journal of Fuzzy Systems*, 22(1):46–61, jan 2020.

[177] S. Lozinskii. *Error Estimates for the Numerical Integration of Ordinary Differential Equations, I.* STL trans. series. Space Technology Laboratories, 1962.

[178] A. D. Luca, G. Oriolo, and M. Vendittelli. Control of wheeled mobile robots: An experimental overview. In *Ramsete*, pages 181–226. Springer Berlin Heidelberg, 2001.

[179] D. Luenberger. *Optimization by vector space methods*. Wiley, New York, 1968.

[180] M. Luo, F. Sun, and H. Liu. Hierarchical structured sparse representation for T-S fuzzy systems identification. *IEEE Transactions on Fuzzy Systems*, 21(6):1032–1043, Dec. 2013.

[181] A. Lyapunov. *General Problem of the Stability of Motion*. Control Theory and Applications Series. Taylor & Francis, 1992.

[182] K. M. Lynch and F. C. Park. *Modern robotics*. Cambridge University Press, Cambridge, England, May 2017.

[183] F.-R. López-Estrada, D. Rotondo, and G. Valencia-Palomo. A review of convex approaches for control, observation and safety of linear parameter varying and Takagi-Sugeno systems. *Processes*, 7(11):814, Nov. 2019.

[184] R. H. Macmillan. *An introduction to the theory of control in mechanical engineering*. Cambridge University Press, Cambridge, England, May 2016.

[185] R. S. Mah, G. M. Stanley, and D. M. Downing. Reconcillation and rectification of process flow and inventory data. *Industrial & Engineering Chemistry Process Design and Development*, 15(1):175–183, Jan. 1976.

[186] B. Marx. A descriptor Takagi-Sugeno approach to nonlinear model reduction. *Linear Algebra and its Applications*, 479:52–72, Aug. 2015.

[187] O. Mason. On common quadratic Lyapunov functions for stable discrete-time LTI systems. *IMA Journal of Applied Mathematics*, 69(3):271–283, June 2004.

[188] F. Mazenc, T. N. Dinh, and S. I. Niculescu. Interval observers for discrete-time systems: interval observers for discrete-time systems. *International Journal of Robust and Nonlinear Control*, 24(17):2867–2890, June 2013.

[189] J. A. Meda-Campana, J. C. Gomez-Mancilla, and B. Castillo-Toledo. Exact output regulation for nonlinear systems described by Takagi–Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, 20(2):235–247, Apr. 2012.

[190] M. Milanese, J. Norton, H. Piet-Lahanier, and É. Walter, editors. *Bounding Approaches to System Identification*. Springer US, 1996.

[191] P. Milgrom and I. Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–610, 2002.

[192] J. W. Milnor. *Topology from the differentiable viewpoint*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, Princeton, NJ, Nov. 1997.

[193] A. Molchanov and Y. Pyatnitskiy. Criteria of asymptotic stability of differential and difference inclusions encountered in control theory. *Systems & Control Letters*, 13(1):59–64, July 1989.

[194] P. Montagnier, R. J. Spiteri, and J. Angeles. The control of linear time-periodic systems using Floquet–Lyapunov theory. *International Journal of Control*, 77(5):472–490, Mar. 2004.

[195] H. Moodi and M. Farrokhi. On observer-based controller design for Sugeno systems with unmeasurable premise variables. *ISA Transactions*, 53(2):305–316, Mar. 2014.

[196] L. Mozelli, R. Palhares, F. Souza, and E. Mendes. Reducing conservativeness in recent stability conditions of TS fuzzy systems. *Automatica*, 45(6):1580–1583, June 2009.

[197] L. A. Mozelli, R. M. Palhares, and G. S. Avellar. A systematic approach to improve multiple Lyapunov function stability and stabilization conditions for fuzzy systems. *Information Sciences*, 179(8):1149–1162, Mar. 2009.

[198] R. Murray-Smith and T. A. Johansen. *Multiple Model Approaches to Modelling and Control*. Taylor and Francis, London, 1997.

[199] A. M. Nagy, G. Mourot, B. Marx, J. Ragot, and G. Schutz. Systematic multimodeling methodology applied to an activated sludge reactor model. *Industrial & Engineering Chemistry Research*, 49(6):2790–2799, Feb. 2010.

[200] L. Narici and E. Beckenstein. *Topological Vector Spaces*. Chapman and Hall/CRC, July 2010.

[201] L. S. Nelson. Fault detection and diagnosis in chemical and petrochemical processes, 1979.

[202] A. Nguyen, A. Dequidt, and M. Dambrine. Anti-windup based dynamic output feedback controller design with performance consideration for constrained Takagi-Sugeno systems. *Engineering Applications of Artificial Intelligence*, 40:76–83, apr 2015.

[203] T. V. Nguyen, T. Mori, and Y. Mori. Existence conditions of a common quadratic Lyapunov function for a set of second-order systems. *Transactions of the Society of Instrument and Control Engineers*, 42(3):241–246, 2006.

[204] B. Noack, F. Pfaff, and U. D. Hanebeck. Optimal Kalman gains for combined stochastic and set-membership state estimation. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, page 4035–4040. IEEE, Dec. 2012.

[205] H. Ohtake, K. Tanaka, and H. Wang. Polar coordinate fuzzy model and stability analysis. In *Proceedings of the 2003 American Control Conference, 2003.* IEEE.

[206] H. Ohtake, K. Tanaka, and H. Wang. Fuzzy modeling via sector nonlinearity concept. In *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*. IEEE, 2001.

[207] V. Ojleska, T. Kolemishevska-Gugulovska, and G. Dymirkovsky. Recent advances in analysis and control design for switched fuzzy systems: A review. In *2012 6th IEEE International Conference on Intelligent Systems*. IEEE, Sept. 2012.

[208] V. Ojleska and G. Stojanovski. Switched fuzzy systems: overview and perspectives. In *9th International PhD Workshop on Systems and Control: Young Generation Viewpoint, Slovenia*, volume 1. Citeseer, 2008.

[209] A. Packard and G. Becker. Quadratic stabilization of parametrically-dependent linear systems using parametrically-dependent linear, dynamic feedback. *Advances in robust and nonlinear control systems*, pages 29–36, 1992.

[210] E. Palacios and A. Titli. Pole placement in LMI region with Takagi-Sugeno fuzzy systems. *IFAC Proceedings Volumes*, 36(12):243–248, July 2003.

[211] J.-T. Pan, T. M. Guerra, S.-M. Fei, and A. Jaadari. Nonquadratic stabilization of continuous T–S fuzzy models: LMI solution for a local approach. *IEEE Transactions on Fuzzy Systems*, 20(3):594–602, June 2012.

[212] J. Park, J. Kim, and D. Park. LMI-based design of stabilizing fuzzy controllers for nonlinear systems described by Takagi–Sugeno fuzzy model. *Fuzzy Sets and Systems*, 122(1):73–82, Aug. 2001.

[213] R. Patton and J. Chen. A review of parity space approaches to fault diagnosis. *IFAC Proceedings Volumes*, 24(6):65–81, Sept. 1991.

[214] D. Peaucelle, D. Arzelier, O. Bachelier, and J. Bernussou. A new robust-stability condition for real convex polytopic uncertainty. *Systems & Control Letters*, 40(1):21–30, May 2000.

[215] S. Pettersson and B. Lennartson. An LMI approach for stability analysis of nonlinear systems. In *1997 European Control Conference (ECC)*. IEEE, July 1997.

[216] S. Pettersson and B. Lennartson. Exponential stability of hybrid systems using piecewise quadratic Lyapunov functions resulting in LMIS. *IFAC Proceedings Volumes*, 32(2):4810–4815, July 1999.

[217] S. Ploix and O. Adrot. Parity relations for linear uncertain dynamic systems. *Automatica*, 42(9):1553–1562, Sept. 2006.

[218] D. Poole. Representing diagnosis knowledge. *Annals of Mathematics and Artificial Intelligence*, 11(1–4):33–50, Mar. 1994.

[219] H. Pottmann and M. Peternell. Envelopes - computational theory and applications. *Proceedings of Spring Conference on Computer Graphics*, April 2009.

[220] F. Preisach. Über die magnetische Nachwirkung. *Zeitschrift für Physik*, 94(5-6):277–302, May 1935.

[221] F. P. Preparata and M. I. Shamos. *Computational Geometry*. Springer New York, 1985.

[222] ProofWiki. Supremum of sum equals sum of suprema. *proofwiki.org/wiki/Supremum_of_-Sum_equals_Sum_of_Suprema*.

[223] V. Puig. Fault diagnosis and fault tolerant control using set-membership approaches: Application to real case studies. *International Journal of Applied Mathematics and Computer Science*, 20(4):619–635, Dec. 2010.

[224] D. Qu, Z. Huang, Y. Zhao, G. Song, K. Yi, and X. Zhao. Nonlinear state estimation by extended parallelotope set-membership filter. *ISA Transactions*, 128:414–423, Sept. 2022.

[225] V. Raghuraman and J. P. Koeln. Set operations and order reductions for constrained zonotopes. *Automatica*, 139:110204, May 2022.

[226] S.-A. Raka and C. Combastel. Fault detection based on robust adaptive thresholds: A dynamic interval approach. *Annual Reviews in Control*, 37(1):119–128, Apr. 2013.

[227] S. V. Raković. Control Minkowski-Lyapunov functions. *Automatica*, 128:109598, June 2021.

[228] A. Ramezanifar, J. Mohammadpour, and K. Grigoriadis. Sampled-data control of LPV systems using input delay approach. In *IEEE Conference on Decision and Control*. IEEE, Dec. 2012.

[229] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987.

[230] V. Reppa and A. Tzes. Fault-detection relying on set-membership techniques for an atomic force microscope. *IFAC Proceedings Volumes*, 42(8):1186–1191, 2009.

[231] B.-J. Rhee and S. Won. A new fuzzy Lyapunov function approach for a Takagi-Sugeno fuzzy control system design. *Fuzzy Sets and Systems*, 157(9):1211–1228, May 2006.

[232] G. G. Rigatos. *State-Space Approaches for Modelling and Control in Financial Engineering*. Springer International Publishing, 2017.

[233] D. Robert, O. Sename, and D. Simon. An H_∞ LPV design for sampling varying controllers: Experimentation with a T-inverted pendulum. *IEEE Transactions on Control Systems Technology*, 18(3):741–749, May 2010.

[234] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Dec. 1970.

[235] D. Rotondo. *Advances in Gain-Scheduling and Fault Tolerant Control Techniques*. Springer International Publishing, 2018.

[236] W. J. Rugh and J. S. Shamma. Research on gain scheduling. *Automatica*, 36(10):1401–1425, Oct. 2000.

[237] D. Saifia, M. Chadli, S. Labiod, and T. M. Guerra. Robust H_∞ static output feedback stabilization of T-S fuzzy systems subject to actuator saturation. *International Journal of Control, Automation and Systems*, 10(3):613–622, jun 2012.

[238] A. Sala and C. Arino. Asymptotically necessary and sufficient conditions for stability and performance in fuzzy control: Applications of Polya's theorem. *Fuzzy Sets and Systems*, 158(24):2671–2686, Dec. 2007.

[239] A. Sala and C. Arino. Relaxed stability and performance conditions for Takagi-Sugeno fuzzy systems with knowledge on membership function overlap. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 37(3):727–732, June 2007.

[240] A. Sala and C. Arino. Relaxed stability and performance LMI conditions for Takagi-Sugeno fuzzy systems with polynomial constraints on membership function shapes. *IEEE Transactions on Fuzzy Systems*, 16(5):1328–1336, Oct. 2008.

[241] S. Sastry. *Lyapunov Stability Theory. In: Nonlinear Systems*, page 182–234. Interdisciplinary Applied Mathematics, Springer New York, 1999.

[242] H. H. Schaefer and M. P. Wolff. Locally convex topological vector spaces. In *Topological Vector Spaces*, pages 36–72. Springer New York, 1999.

[243] E. Schechter. Convexity. In *Handbook of Analysis and Its Foundations*, pages 302–325. Elsevier, 1997.

[244] C. Scherer. LPV control and full block multipliers. *Automatica*, 37(3):361–375, Mar. 2001.

[245] C. Scherer and S. Weiland. Linear matrix inequalities in control. *Lecture Notes, Dutch Institute for Systems and Control, Delft, The Netherlands*, 3(2), 2000.

[246] C. W. Scherer. Mixed H_2/H_∞ control for time-varying and linear parametrically-varying systems. *International Journal of Robust and Nonlinear Control*, 6:929–952, 1996.

[247] R. Schneider. Basic convexity. In *Convex Bodies The Brunn-Minkowski Theory*, pages 1–73. Cambridge University Press, Oct. 2013.

[248] F. Schweppe. Recursive state estimation: Unknown but bounded errors and system inputs. *IEEE Transactions on Automatic Control*, 13(1):22–28, Feb. 1968.

[249] F. Schweppe. *Uncertain dynamic systems: Modelling, Estimation, Hypothesis Testing, Identification and Control*. Prentice-Hall, Englewood Cliffs, N.J, 1973.

[250] J. K. Scott, D. M. Raimondo, G. R. Marseglia, and R. D. Braatz. Constrained zonotopes: A new tool for set-based estimation and fault detection. *Automatica*, 69:126–136, July 2016.

[251] J. Seo, D. Chung, C. G. Park, and J. G. Lee. The robustness of controllability and observability for discrete linear time-varying systems with norm-bounded uncertainty. *IEEE Transactions on Automatic Control*, 50(7):1039–1043, July 2005.

[252] J. S. Shamma and M. Athans. Analysis of gain scheduled control for nonlinear plants. *IEEE Transactions on Automatic Control*, 35(8):898–907, 1990.

[253] J. S. Shamma and M. Athans. Guaranteed properties of gain scheduled control for linear parameter-varying plants. *Automatica*, 27(3):559–564, May 1991.

[254] R. Shorten and K. Narendra. Necessary and sufficient conditions for the existence of a common quadratic Lyapunov function for two stable second order linear time-invariant systems. In *Proceedings of the 1999 American Control Conference*, volume 2, pages 1410–1414, 1999.

[255] R. Shorten and K. Narendra. Necessary and sufficient conditions for the existence of a common quadratic Lyapunov function for m stable second order linear time-invariant systems. In *Proceedings of the 2000 American Control Conference*. IEEE, 2000.

[256] R. Shorten and K. Narendra. Necessary and sufficient conditions for the existence of a common quadratic Lyapunov function for a finite number of stable second order linear time-invariant systems. *International Journal of Adaptive Control and Signal Processing*, 16(10):709–728, Nov. 2002.

[257] L. M. Silverman and H. E. Meadows. Controllability and observability in time-variable linear systems. *SIAM Journal on Control*, 5(1):64–73, Feb. 1967.

[258] D. Silvestre. Constrained convex generators: A tool suitable for set-based estimation with range and bearing measurements. *IEEE Control Systems Letters*, 6:1610–1615, 2022.

[259] D. Silvestre. Exact set-valued estimation using constrained convex generators for uncertain linear systems. *IFAC-PapersOnLine*, 56(2):9461–9466, 2023.

[260] A. Slavík. *Product integration. Its history and applications*. Matfyzpress, 2007.

[261] S. Sojoudi, J. Lavaei, and A. G. Aghdam. Robust controllability and observability degrees of polynomially uncertain systems. *Automatica*, 45(11):2640–2645, Nov. 2009.

[262] E. Sontag and Y. Wang. Lyapunov characterizations of input to output stability. *SIAM Journal on Control and Optimization*, 39(1):226–249, Jan. 2000.

[263] R. Stanley. *Enumerative combinatorics: volume 1*. Cambridge university press, 2023.

[264] R. P. Stanley. *Enumerative Combinatorics*. Springer US, 1986.

[265] P. Sun and R. M. Freund. Computation of minimum-volume covering ellipsoids. *Operations Research*, 52(5):690–706, Oct. 2004.

[266] W. A. Sutherland. *Introduction to metric and topological space*. Oxford University Press, London, England, Aug. 1975.

[267] H. Suzukia and T. Sugie. MPC for LPV systems with bounded parameter variation using ellipsoidal set prediction. In *2006 American Control Conference*. IEEE, 2006.

[268] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-15(1):116–132, Jan. 1985.

[269] K. Tan, K. M. Grigoriadis, and F. Wu. Output-feedback control of LPV sampled-data systems. *International Journal of Control*, 75(4):252–264, Jan. 2002.

[270] K. Tanaka, T. Hori, and H. O. Wang. New parallel distributed compensation using time derivative of membership functions: a fuzzy Lyapunov approach. In *Proceedings of the 40th IEEE Conference on Decision and Control*, CDC-01. IEEE, 2001.

[271] K. Tanaka, T. Hori, and H. O. Wang. A multiple Lyapunov function approach to stabilization of fuzzy control systems. *IEEE Transactions on Fuzzy Systems*, 11(4):582–589, Aug. 2003.

[272] K. Tanaka and H. O. Wang. *Fuzzy Control Systems Design and Analysis*. John Wiley & Sons, Inc., Sept. 2001.

[273] K. Tanaka, H. Yoshida, H. Ohtake, and H. O. Wang. A sum of squares approach to stability analysis of polynomial fuzzy systems. In *2007 American Control Conference*. IEEE, July 2007.

[274] K. Tanaka, H. Yoshida, H. Ohtake, and H. O. Wang. A sum-of-squares approach to modeling and control of nonlinear dynamical systems with polynomial fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 17(4):911–922, Aug. 2009.

[275] G. Tao and P. V. Kokotovic. Adaptive control of plants with unknown dead-zones. *IEEE Transactions on Automatic Control*, 39(1):59–68, 1994.

[276] G. Tao and P. V. Kokotovic. *Adaptive control of systems with actuator and sensor nonlinearities*. John Wiley & Sons, Nashville, TN, May 1996.

[277] G. Tao and F. L. Lewis, editors. *Adaptive control of nonsmooth dynamic systems*. Springer, London, England, Sept. 2001.

[278] S. Tarbouriech, G. Garcia, J. M. G. da Silva, and I. Queinnec. *Stability and Stabilization of Linear Systems with Saturating Actuators*. Springer London, 2011.

[279] E. S. Tognetti, R. C. L. F. Oliveira, and P. L. D. Peres. Selective H_2 and H_∞ stabilization of Takagi–Sugeno fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 19(5):890–900, Oct. 2011.

[280] R. Tóth, P. V. den Hof, and P. Heuberger. Discretisation of linear parameter-varying state-space representations. *IET Control Theory & Applications*, 4(10):2082–2096, Oct. 2010.

[281] R. Tóth, F. Felici, P. Heuberger, and P. V. den Hof. Crucial aspects of zero-order hold LPV state-space system discretization. *IFAC Proceedings Volumes*, 41(2):4952–4957, 2008.

[282] L. Travé-Massuyès and T. Escobet. *Bridge: Matching Model-Based Diagnosis from FDI and DX Perspectives*, page 153–175. Springer International Publishing, 2019.

[283] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus. *Control Theory for Linear Systems*. Springer London, 2001.

[284] S.-H. Tsai and Y.-W. Chen. A novel identification method for Takagi-Sugeno fuzzy model. *Fuzzy Sets and Systems*, 338:117–135, May 2018.

[285] H. Tuan, P. Apkarian, T. Narikiyo, and Y. Yamamoto. Parameterized linear matrix inequality techniques in fuzzy control system design. *IEEE Transactions on Fuzzy Systems*, 9(2):324–332, Apr. 2001.

[286] R. Tóth. *Modeling and Identification of Linear Parameter-Varying Systems*. Springer Berlin Heidelberg, 2010.

[287] P. Van Der Kloet and F. Neerhoff. Behaviour of dynamic eigenpairs in slowly-varying systems. In *Proc. NDES*, volume 99, pages 15–17, 1999.

[288] B. L. van der Waerden. *Algebra*. Springer, New York, NY, 1 edition, Oct. 2003.

[289] M. Vidyasagar. *Nonlinear Systems Analysis*. Society for Industrial and Applied Mathematics, Jan. 2002.

[290] V. Volterra and B. Hostinský. *Opérations infinitésimales linéaires, applications aux équations différentielles et fonctionnelles*. Gauthier-Villars, 1938.

[291] H. Wang, K. Tanaka, and M. Griffin. An approach to fuzzy control of nonlinear systems: stability and design issues. *IEEE Transactions on Fuzzy Systems*, 4(1):14–23, 1996.

[292] J. Warren. Barycentric coordinates for convex polytopes. *Advances in Computational Mathematics*, 6(1):97–108, Dec. 1996.

[293] J. Warren. On the uniqueness of barycentric coordinates. In *Contemporary Mathematics, Proceedings of AGGM02*, 2003.

[294] J. Warren, S. Schaefer, A. N. Hirani, and M. Desbrun. Barycentric coordinates for convex sets. *Advances in Computational Mathematics*, 27(3):319–338, Dec. 2006.

[295] A. P. White, G. Zhu, and J. Choi. *Linear Parameter-Varying Control for Engineering Applications*. Springer London, 2013.

[296] D. Wilson. Convolution and Hankel operator norms for linear systems. *IEEE Transactions on Automatic Control*, 34(1):94–97, 1989.

[297] W. Wonham. On pole assignment in multi-input controllable linear systems. *IEEE transactions on automatic control*, 12(6):660–665, 1967.

[298] C. Wu, I. Kanevskiy, and M. Margaliot. k-contraction: Theory and applications. *Automatica*, 136:110048, Feb. 2022.

[299] H.-N. Wu and K.-Y. Cai. H_2 guaranteed cost fuzzy control for uncertain nonlinear systems via linear matrix inequalities. *Fuzzy Sets and Systems*, 148(3):411–429, 2004.

[300] L. Wu, X. Su, P. Shi, and J. Qiu. Model approximation for discrete-time state-delay systems in the ts fuzzy framework. *IEEE Transactions on Fuzzy Systems*, 19(2):366–378, Apr. 2011.

[301] L. Xiaodong and Z. Qingling. New approaches to H_∞ controller designs based on fuzzy observers for T-S fuzzy systems via LMI. *Automatica*, 39(9):1571–1582, Sept. 2003.

[302] X. Yang and J. K. Scott. A comparison of zonotope order reduction techniques. *Automatica*, 95:378–384, Sept. 2018.

[303] R. Yedavalli. Conditions for the existence of a common quadratic Lyapunov function via stability analysis of matrix families. In *Proceedings of the 2002 American Control Conference*. IEEE, 2002.

[304] J. Yoneyama, M. Nishikawa, H. Katayama, and A. Ichikawa. Output stabilization of Takagi–Sugeno fuzzy systems. *Fuzzy Sets and Systems*, 111(2):253–266, Apr. 2000.

[305] S.-H. Yoo and B.-J. Choi. A balanced model reduction for T-S fuzzy systems with uncertain time varying parameters. In *Computational and Information Science*, pages 148–153. Springer Berlin Heidelberg, 2004.

[306] L. Zaccarian and A. R. Teel. *Modern Anti-windup Synthesis: Control Augmentation for Actuator Saturation*. Princeton University Press, 2011.

[307] X. Zhan. *Matrix Inequalities*. Springer Berlin Heidelberg, 2002.

[308] J. Zhang, W.-B. Xie, M.-Q. Shen, and L. Huang. State augmented feedback controller design approach for T-S fuzzy system with complex actuator saturations. *International Journal of Control, Automation and Systems*, 15(5):2395–2405, Sept. 2017.

[309] S. Zhang, V. Puig, and S. Ifqir. Robust fault detection using set-based approaches for LPV systems: Application to autonomous vehicles. *IFAC-PapersOnLine*, 55(6):31–36, 2022.

[310] Z. Zhang, S. Xu, and B. Zhang. Asymptotic tracking control of uncertain nonlinear systems with unknown actuator nonlinearity. *IEEE Transactions on Automatic Control*, 59(5):1336–1341, May 2014.

[311] L. Zhao and L. Li. Robust stabilization of T-S fuzzy discrete systems with actuator saturation via PDC and non-PDC law. *Neurocomputing*, 168:418–426, nov 2015.

[312] H. Zheng, W.-B. Xie, H.-K. Lam, and L. Wang. Membership-function-dependent stability analysis and local controller design for T–S fuzzy systems: A space-enveloping approach. *Information Sciences*, 548:233–253, Feb. 2021.

[313] Y. Zi-zong and G. Jin-hai. Some equivalent results with Yakubovich's S-lemma. *SIAM Journal on Control and Optimization*, 48(7):4474–4480, Jan. 2010.

[314] G. M. Ziegler. *Lectures on Polytopes*. Springer New York, 1995.

Résumé

Cette thèse explore la théorie du contrôle non linéaire par le prisme des modèles LPV (Linéaires à Paramètres Variants) et T-S (Takagi-Sugeno), en se concentrant sur l'amélioration de leur flexibilité et de leur efficacité. Les principales contributions de cette thèse portent sur l'introduction d'outils géométriques qui n'avaient pas été utilisés auparavant dans ces contextes. Il est notamment montré que les coordonnées barycentriques jouent un rôle clé dans la modélisation des systèmes T-S par l'approche par secteur non linéaire (*Chapitre 3*), et que des complexes polyédraux peuvent également être utilisés pour obtenir des modèles T-S non convexes (*Chapitre 4*). En outre, il est établi que les interpolations de Bézier permettent d'obtenir une compréhension géométrique des multisommes du cadre T-S (*Chapitre 5*). Il est aussi démontré qu'une hypothèse de Lipschitz sur le vecteur d'ordonnancement d'un système LPV borne toutes les matrices de transition d'état que l'on peut obtenir dans le futur, ce qui conduit à des résultats utiles pour caractériser le futur proche de ces systèmes (*Chapitre 6*). Ces derniers résultats sont obtenus en utilisant l'intégrale-produit de Volterra et la norme logarithmique pondérée d'une matrice. Une approche ensembliste est explorée pour la détection de défauts, en utilisant la fonctionnelle de Minkowski d'un ensemble (*Chapitre 7*). Enfin, la modélisation des saturations ainsi que d'autres phénomènes (zones mortes, hystérésis) affectant localement les actionneurs d'un système est abordée à l'aide d'outils géométriques, comme la fonctionnelle de Minkowski (*Chapitre 8*).

Mots-clés: Systèmes non linéaires, modèles de Takagi-Sugeno, modèles linéaires à paramètres variants, méthode de Lyapunov, inégalités matricielles linéaires

Abstract

This thesis explores nonlinear control theory using LPV (Linear Parameter-Varying) and T-S (Takagi-Sugeno) models, focusing on improving the flexibility and efficiency of these frameworks. The main contributions of this thesis concern the introduction of geometric tools that had not previously been used in these contexts. In particular, it is shown that barycentric coordinates play a key role in the modeling of T-S systems using the nonlinear sector approach (*Chapter 3*), and that polyhedral complexes can also be used to obtain non-convex T-S models (*Chapter 4*). It is also established that Bézier interpolations provide a geometric understanding of the multi-sums involved in the T-S framework (*Chapter 5*). It is demonstrated that a Lipschitz assumption on the scheduling vector of a LPV system allows to bound all its potential state transition matrices in the future, leading to useful results to characterize the near-future of these systems (*Chapter 6*). These later results are obtained using Volterra's product integration and the weighted logarithmic norm of a matrix. Additionally, a set-membership approach is explored for fault detection purposes, using the Minkowski functional of a set (*Chapter 7*). Finally, the modeling of saturations and other phenomena (dead zones, hysteresis) locally affecting the actuators of a system is approached using geometric tools, such as the Minkowski functional (*Chapter 8*).

Keywords: Nonlinear systems, Takagi-Sugeno models, linear parameter varying models, Lyapunov method, linear matrix inequalities