Université catholique de Louvain
Faculté des sciences appliquées
Département d'ingénierie mathématique

Center for Systems Engineering and Applied Mechanics

Joint Spectral Radius : theory and approximations

Jacques Theys

Thesis submitted in partial fulfillment
of the requirements for the degree of
Docteur en sciences appliquées

Dissertation committee:
Valérie Berthé, Université de Montpellier II (France)
Vincent Blondel (Promotor), UCL
Jean-Didier Legat (Chair), UCL
Yurii Nesterov, UCL
Rodolphe Sepulchre, Université de Liège
Paul Van Dooren, UCL
Fabian Wirth, Hamilton Institute (Ireland)

May 30, 2005

Acknowledgements

I first would like to thank my promotor Vincent Blondel for accepting me as his first Ph.D student, and providing me with a challenging research subject. His constructive comments, his pragmatism and his initiative were essential in the realization of this thesis.

Several researchers contributed to this thesis. I would like to especially thank Alexander Vladimirov and Yurii Nesterov for their important collaborations. I am also very grateful to Fabian Wirth whose knowledge of the subject was of invaluable help. I learned a lot working with him. I am also grateful to Paul Van Dooren for clarifying matrix theoretical problems on several occasions. And I thank Vladimir Protasov for his numerous relevant comments on this work.

I am also indebted to the members of my jury for their careful analysis of my work, allowing for a rich feedback on this thesis, which translated into an exciting private defense. I wish every Ph.D. student to get such a feedback on his work.

This research was supported by the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. I also thank the Belgian National Research Fund (FNRS) for granting me a fellowship.

Two members of the CESAME successively had to stand my sense of humour all day long : Amaury Lendasse and Antoine Vandendorpe. My stay in the CESAME would not have been as pleasant without them. I also want to thank Vincent Canterini for helping me at the beginning of this thesis, and for his friendship. I also want to thank the other members of the CESAME for making it such a nice place (among others, Adrien Leygue, Agnès Provost, Luc Moens, Jonathan de Halleux and Jean-Charles Delvenne). I also thank the secretaries team consisting in Isabelle, Michou, Dominique and Lydia, for making everything so easy to organize.

Last but not least, I thank Marie for standing by me throughout these years.

Contents

Notation glossary

Basics

$\mathbb{N}$	set of nonnegative integer numbers
$\mathbb{Z}$	set of integer numbers
$\mathbb{R}$	field of real numbers
$\mathbb{C}$	field of complex numbers
j	complex unit, $j = \sqrt{-1}$
$\mathcal{R}e(z)$	real part of z
$\mathcal{I}m(z)$	complex part of z
$\bar{z}$	complex conjugate of z
$\|z\|$	modulus of z
$\lceil x \rceil$	ceiling of x : the smallest integer $\geq x$
$\iff$	if and only if
$:=$	equal by definition to
δ_{ij}	Kronecker delta: equals 1 if $i = j$ and zero otherwise
$\\| \cdot \\|_p$	p-norm $(1 \leq p \leq +\infty)$
$\\| \cdot \\|_2$	Euclidean norm (vectors) / spectral norm
$\\| \cdot \\|_1$	1-norm (vectors and induced matrix norm)
$\\| \cdot \\|_1$	∞-norm (vectors and induced matrix norm)
$\\| \cdot \\|_P$	Ellipsoid norm (vectors and induced matrix norm)

Matrix theory

I_k	$k \times k$ identity matrix
X^T	transpose of matrix X
X^*	conjugate transpose of matrix X
$x_{i,j}$	element located at the i^{th} row and the j^{th} column of X
$\det X$	determinant of square matrix X
$trace(X)$	trace of square matrix X
$\lambda_k(X)$	k-th eigenvalue of matrix X
$\sigma_{\max}(X)$	maximal singular value of matrix X
$\sigma_{\min}(X)$	minimal singular value of matrix X
$\otimes$	Kronecker product
$X^{\otimes k}$	k^{th} Kronecker power of matrix X

Introduction

Historical background

The formal definition of the joint spectral radius was first introduced by Rota and Strang in the 60's. In the 90's, Daubechies & Lagarias defined the generalized spectral radius, and Berger & Wang proved later these two values to be equal for bounded sets of matrices.

The notion of joint spectral radius of a set of matrices appears in a wide range of contexts and has led to a large number of contributions. A survey presenting a list of over hundred related contributions was compiled by Gilbert Strang in [78].

We will give fully detailed definitions in this work, but we can give a foretaste of the things to come. The spectral radius of a matrix characterizes the growth rate of successive powers of the matrix. When considering a set of matrices, we can characterize the maximal growth rate of the various products that can be composed using matrices from the set. This is the joint spectral radius of this set.

This extension of the notion of spectral radius of a matrix to a set of matrices is fairly natural. However, it gives rise to some serious issues. This value can indeed be shown to be hard to compute, and also leads to problems that have been found to be undecidable. We will detail these points in this work.

Thesis Outline

Chapter 1 first introduces basic definitions about matrices. Then we move on to the definition of the joint and generalized spectral radii, as well as their elementary properties. We also define the lower spectral radius of a set of matrices, which is the infimum counterpart to the joint spectral radius, which itself deals with the supremum taken on all possible products. We then state founding results for the remainder of the thesis.

Chapter 2 presents three different areas where the joint spectral radius has been found to play a role. These are:

- the coordination of autonomous agents,
- the continuity of wavelet functions,
- the capacity of codes.

We explain what these problems consist in, and detail how the joint spectral radius arises in these contexts. The joint spectral radius can be used to determine if, under certain assumptions, a group of autonomous agents will eventually converge to the same direction. The answer to this question is of the type yes/no, but the joint spectral radius can also be used as a quantitative performance measure. For example, it can be used to quantify the continuity of wavelet basis functions (via the Hölder exponent). It can also express the capacity of a binary code, that is the amount of information it is able to transmit over a given number of bits.

Chapter 3 is devoted to the dynamical systems aspect of our study. We introduce switched linear systems and define their stability. We then make the link between stability and the joint spectral radius. This link may be considered as induced by order results, but we provide a clear statement that the joint spectral radius rules the stability of a discrete time switched linear system. We then discuss briefly discuss new notions of zero-reachability, that can be linked to the lower spectral radius, defined in Chapter 1.

This being done, we look at the role played by common quadratic Lyapunov functions in the study of the stability of such systems. We conclude this chapter by discussing the link between switched linear systems in discrete time and in continuous time. Our publications related to this chapter are [14, 15]:

- *Discrete and continuous stability of switched linear systems.*
  Blondel Vincent, Theys Jacques.
  22nd Benelux Meeting on Systems and Control, Lommel, Belgium, March 19-21, 2003.
- *On the relations between discrete and continuous time stability for switched linear systems.*
  Blondel Vincent, Theys Jacques.
  Proceedings of the Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Katholieke Universiteit Leuven, Belgium, July 5-9, 2004.

Chapter 4 details several results that illustrate the complexity of the computation of the joint spectral radius. We begin by stating the *finiteness conjecture*, and we provide a self-contained counterproof, by showing that there exist counterexamples to it. This result is of interest, since the only counterproof available appears to be the consequence of a more complex (and rather unrelated at first sight) result. In the following section, we present a non-algebraicity result by Kozyakin, that further highlights the difficulty of the problem. For this result, we point out a flaw in the original proof by Kozyakin, and provide a correction for it. We present the notion of extremal norm and we then discuss further the existence of extremal norms, which can make the computation of the joint spectral radius somewhat easier. And we conclude by stating results by Blondel & Tsitsiklis about the NP-hardness of the considered computation. Our publications related to this chapter are [11, 10, 7]:

- *An elementary counterexample to the finiteness conjecture.*
  Blondel Vincent, Theys Jacques, Vladimirov Alexander.
  SIAM Journal on Matrix Analysis and Applications, volume 24, number 4, pages 963-970 , 2003.
- *Switched systems that are periodically stable may be unstable.*
  Blondel Vincent, Theys Jacques, Vladimirov Alexander.
  Proceedings of the Fifteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2002).
- *When is a pair of matrices stable ?*
  Blondel Vincent, Theys Jacques, Tsitsiklis John.
  Contribution to the book : "Unsolved problems in Mathematical Systems and Control Theory", by V.Blondel and A.Megretski, Princeton University Press.

In Chapter 5, we present an extensive study of a specific pair of matrices that is central in this thesis:

$$\left\{ A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\ A_1 = \alpha \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\} ,$$

with α a parameter in $\mathbb{R}$. We show that, for this specific pair of matrices, the joint spectral radius is achieved by matrix products of a very specific nature. In particular, we prove that these are balanced words, which enables for an important improvement in the computation of the joint spectral radius for this family of matrices. Indeed, the computation time is not exponential any more, but polynomial. For this family of matrices, we also detail a method for constructing the level curve of an extremal norm. A publication gathering the results from this chapter is expected soon.

Chapter 6 deals with the problem of effectively computing the joint spectral radius. We first look at what methods the theory enables us to apply directly and review some of the known computational refinements. We then study specific methods in the case of matrices with non-negative entries. The following method uses a modification of the Lyapunov equations to define an approximation of the joint spectral radius. We give a proved bound on how precise this approximation is in general, before detailing specific cases where it can be shown to be equal to the actual value of the joint spectral radius. Our publications related to this chapter are [16, 17, 18]:

- *Approximations of the rate of growth of switched linear systems.*
  Blondel Vincent, Nesterov Yurii, Theys Jacques.
  Lecture Notes in Computer Science (Publisher: Springer-Verlag GmbH), ISSN: 0302-9743, volume 2993 / 2004. Hybrid Systems: Computation and Control: 7th International Workshop, HSCC 2004, Philadelphia, PA, USA, March 25-27, 2004. Chapter: pp. 173 - 186.
- *Computing the joint spectral radius of a set of matrices.*
  Blondel Vincent, Nesterov Yurii, Theys Jacques.
  23rd Benelux Meeting on Systems and Control, Helvoirt, The Netherlands, March 17-19, 2004.
- *On the accuracy of the ellipsoid norm approximation of the joint spectral radius.*
  Blondel Vincent, Nesterov Yurii, Theys Jacques.
  Linear Algebra and its Applications, 394, pages 91-107, 2005.

We will then conclude with the possible follow-ups to this work, discussing the work that could be done to gain further knowledge about the joint spectral radius.

Related publications

In a chronological order, this thesis gave rise to the following publications and communications:

- *An elementary counterexample to the finiteness conjecture.*
  Blondel Vincent, Theys Jacques, Vladimirov Alexander.
  SIAM Journal on Matrix Analysis and Applications, volume 24, number 4, pages 963-970 , 2003.
- *Switched systems that are periodically stable may be unstable.*
  Blondel Vincent, Theys Jacques, Vladimirov Alexander.
  Proceedings of the Fifteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2002).
- *When is a pair of matrices stable ?*
  Blondel Vincent, Theys Jacques, Tsitsiklis John.
  Contribution to the book: "Unsolved problems in Mathematical Systems and Control Theory", by V.Blondel and A.Megretski, Princeton University Press.
- *Discrete and continuous stability of switched linear systems.*
  Blondel Vincent, Theys Jacques.
  22nd Benelux Meeting on Systems and Control, Lommel, Belgium, March 19-21, 2003.
- *Approximations of the rate of growth of switched linear systems.*
  Blondel Vincent, Nesterov Yurii, Theys Jacques.
  Lecture Notes in Computer Science (Publisher: Springer-Verlag GmbH), ISSN: 0302-9743, volume 2993 / 2004. Hybrid Systems: Computation and Control: 7th International Workshop, HSCC 2004, Philadelphia, PA, USA, March 25-27, 2004. Chapter: pp. 173 - 186.
- *Computing the joint spectral radius of a set of matrices.*
  Blondel Vincent, Nesterov Yurii, Theys Jacques.
  23rd Benelux Meeting on Systems and Control, Helvoirt, The Netherlands, March 17-19, 2004.
- *On the relations between discrete and continuous time stability for switched linear systems.*
  Blondel Vincent, Theys Jacques.
  Proceedinfs of the Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Katholieke Universiteit Leuven, Belgium, July 5-9, 2004.
- *On the accuracy of the ellipsoid norm approximation of the joint spectral radius.*
  Blondel Vincent, Nesterov Yurii, Theys Jacques.
  Linear Algebra and its Applications, 394, pages 91-107, 2005.

1

The joint spectral radius : definition

This chapter is devoted to the formal definition of the joint spectral radius. This must be done with care, since this is the notion that will be discussed extensively throughout this thesis. First, we will define the widely used notion of spectral radius of a matrix, along with some of its properties. Then, we will show how this notion can be naturally extended to the *joint spectral radius* of a set of matrices. We will also make the link between properties of the spectral radius and analog properties of the joint spectral radius. We will present other useful properties of the joint spectral radius in later chapters of this thesis.

1.1 Norm and spectral radius of a matrix

In this section, we present the definitions of basic notions, that are related to single matrices. The first one is the notion of matrix norm, which is critical in proving several properties that will be presented later on in this work. After this, we will introduce the spectral radius, and various (in)equalities involving norms and the spectral radius. Both the spectral radius and the norm of a matrix will be shown to play a similar role in the definition of the joint spectral radius, in Section 1.2.

Although some of the notions we will be dealing with do not require the matrices to be square, we will always be considering square matrices in this work. Indeed, since we will be studying products of matrices chosen in a set in arbitrary order, these have to be square for this to make sense. Also, we will be considering real matrices, while many of the reasonings presented in this thesis would work as well with complex numbers.

1.1.1 Definition of a matrix norm

Let us consider a square matrix $A \in \mathbb{R}^{n\times n}$. A matrix norm $\|A\|$ is a function associating to this matrix a nonnegative real number, satisfying the following properties:

- nonnegativity

$$\forall A \in \mathbb{R}^{n\times n}, \|A\| > 0 \text{ if } A \neq 0 \text{ and } \|A\| = 0 \iff A = 0 \ ,$$

- linearity

$$\|kA\| = |k| \ \|A\|, \forall k \in \mathbb{R}, \forall A \in \mathbb{R}^{n\times n} \ ,$$

- the triangular inequality

$$\|A + B\| \leq \|A\| + \|B\|, \quad \forall A, B \in \mathbb{R}^{n\times n} \ ,$$

Moreover, we want the norms we use to have the following property.

- submultiplicativity

$$\|AB\| \leq \|A\| \ \|B\|, \ \forall A, B \in \mathbb{R}^{n\times n} \ .$$

This property will be required in the rest of this work. To this aim, we consider *induced matrix norms*. Such a matrix norm $\|.\|_{ind}$ is derived from a vector norm $\|.\|_{vec}$, as follows:

$$\|A\|_{ind} = \max_{\|x\|\neq 0} \frac{\|Ax\|_{vec}}{\|x\|_{vec}} \ ,$$

which, by linearity, can be reduced without loss of generality to

$$\|A\|_{ind} = \max_{\|x\|_{vec}=1} \|Ax\|_{vec} \ .$$

Any induced matrix norm is submultiplicative. Indeed,

$$\begin{aligned} \|AB\| = \max_{\|x\|\neq 0} \frac{\|ABx\|}{\|x\|} &= \max_{\|x\|\neq 0} \frac{\|ABx\|}{\|Bx\|} \frac{\|Bx\|}{\|x\|} \\ &\leq \max_{\|y\|\neq 0} \frac{\|Ay\|}{\|y\|} \max_{\|x\|\neq 0} \frac{\|Bx\|}{\|x\|} = \|A\|\|B\| \ . \end{aligned}$$

Similarly, all induced matrix norms are *consistent*. This means that any vector norm $\|.\|_{vec}$ and its induced operator norm $\|.\|_{ind}$ satisfy, for any matrix A and any vector x,

$$\|Ax\|_{vec} \leq \|A\|_{ind}\|x\|_{vec} \ .$$

This is straightforward from the definition above.

It is known that any two matrix norms $\|.\|_a$ and $\|.\|_b$ can be said to be *equivalent*. This means that, for any norms $\|.\|_a$ and $\|.\|_b$, there exist $0 < k_2 < k_1$ such that

$$k_2\|A\|_b \leq \|A\|_a \leq k_1\|A\|_b, \ \forall A \in \mathbb{R}^{n\times n} \ . \tag{1.1}$$

For induced norms, this equivalence comes from the equivalence between vectors norms, using the definition of an induced norm. The equivalence between vector norms can be justified as follows: if $\|.\|_f$ and $\|.\|_g$ are norms on $\mathbb{R}^2$, then the unit ball $\|.\|_f = 1$ is compact and hence $\|.\|_g$ is bounded on it (norms are continuous). So $\exists M : \|x\|_g \leq M$ for all x with $\|x\|_f \leq 1$. This implies that $\|x\|_g \leq M\|x\|_f, \forall x$. And vice versa for $\|x\|_f \leq M'\|x\|_g$.

For simplicity, throughout this thesis, we will use the same notation $\|A\|$ for matrix norms and vector norms, as their use should be understandable from the context. We will also sometimes be more specific and use the 1-norm $\|.\|_1$, the 2-norm $\|.\|_2$ or the ∞-norm $\|.\|_\infty$. We may remind that, for a vector $x \in \mathbb{R}^n$ and $p \in \mathbb{R}$, $\|x\|_p$ is defined as:

$$\|x\|_p = \sqrt[p]{\sum_i |x_i|^p} \ .$$

And so, the following norms are specific cases of this definition:

- $\|x\|_1$ is sum of the modulus of the components of x:

$$\|x\|_1 = \sum_i |x_i| \ ,$$

- $\|x\|_\infty$ is the maximum modulus of the components of x:

$$\|x\|_\infty = \max_i |x_i| \ .$$

- $\|x\|_2$ is the standard Euclidian norm of x:

$$\|x\|_2 = \sqrt{\sum_i |x_i|^2} \ .$$

The corresponding induced matrix norms are the following:

$$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \max_j \sum_i |a_{i,j}|$$
$$\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_i \sum_j |a_{i,j}| \ ,$$

while $\|A\|_2$ in known as the *spectral norm* of A. This denomination will be explained after we define the spectral radius and singular values of a matrix, in the next subsection.

Another norm we will be using is the *ellipsoid norm*, the level curves of which are ellipsoids. Given a symmetric positive definite (with positive eigenvalues, which we define in the next section) matrix P, denoted by $P \succ 0$, we define:

$$\|x\|_P = \sqrt{x^T P x} \; .$$

1.1.2 Definition of the spectral radius

First, let us recall that an *eigenvalue* of a matrix A is a scalar $\lambda \in \mathbb{C}$ such that:

$$\exists v \neq 0 : Av = \lambda v \; ,$$

where v is an *eigenvector* of A. The eigenvalues of A are the zeros of its *characteristic polynomial*:

$$p(A, \lambda) = \det(A - \lambda I) \; ,$$

where I is the identity matrix.

The spectral radius of a matrix is defined as the largest modulus of its eigenvalues:

$$\rho(A) = max\{|\lambda| : Av = \lambda v\}. \tag{1.2}$$

A square matrix can always be put into its *Jordan form* J, as follows:

$$\forall A, \exists T \in \mathbb{C}^{n \times n} \text{ invertible } : A = TJT^{-1} \; .$$

The matrix J is block-diagonal. These blocks are called Jordan blocks, and are associated to the eigenvalues λ_i of A. They are of the form

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 \dots & 0 \\ 0 & \lambda_i & 1 \dots & 0 \\ \vdots & & \ddots & 1 \\ 0 & \dots & 0 & \lambda_i \end{pmatrix} \; .$$

There can be several such blocks of various sizes associated to a given eigenvalue, depending on its algebraic and geometric multiplicities. We do not need to develop this further.

We also mention the singular values $\sigma_i(A)$ of a matrix A, defined as the square roots of the eigenvalues of A^*A, which are real:

$$\sigma_i(A) = \lambda_i(A^*A)^{1/2} \; .$$

Any matrix A can be transformed in the following way (see [45]):

$$A = U\Sigma V^* \; ,$$

where Σ is a diagonal matrix which contains the singular values of A, while U and V are *unitary* matrices: $U^* = U^{-1}$ and $V^* = V^{-1}$ (U^* denotes the conjugate transpose). Singular values are always real non-negative, so $\Sigma^* = \Sigma$.

At this point, we know enough to explain why $\|A\|_2$ is called the spectral norm. We can indeed write:

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \max_{\|x\|_2=1} \sqrt{x^* A^* A x} \ .$$

And, using the decomposition $A = U\Sigma V^*$, we get

$$\|A\|_2 = \max_{\|x\|_2=1} \sqrt{x^* V \Sigma^* U^* U \Sigma V^* x} = \max_{\|x\|_2=1} \sqrt{x^* V \Sigma^2 V^* x} \ .$$

Then, denoting $y = V^* x$, we have

$$\|A\|_2 = \max_{\|y\|_2=1} \sqrt{y^* \Sigma^2 y} = \max_{\|y\|_2=1} \sqrt{\sum_i y_i^2 \sigma_i^2} \ .$$

This expression is clearly maximal when y is made of only zeros and a single one, corresponding to the maximal singular value of A, yielding the value

$$\|A\|_2 = \sqrt{\max_i \sigma_i^2} \ .$$

This can also be expressed as

$$\|A\|_2 = \rho(A^* A)^{1/2} \ ,$$

that is, the square root of the spectral radius of $A^* A$.

1.1.3 Properties of the spectral radius

Besides its formal definition, the spectral radius can be shown to have interesting properties, which we explain in the following.

Spectral radius of a matrix power

This first result is quite simple and will be used in the upcoming discussions. The eigenvalues of the k^{th} power A^k of a matrix A are the k^{th} power λ^k of those λ of A. Indeed, one can see that:

$$Av = \lambda v \Rightarrow A^k v = \lambda^k v \ .$$

Consequently, from the definition (1.2) above,

$$\rho(A^k) = \rho(A)^k \ . \tag{1.3}$$

We refer the reader to [36] for a formal proof of this result.

Convergence condition

For any matrix $A \in \mathbb{R}^{n\times n}$, the spectral radius rules the convergence to zero of the successive powers A^k of A. More precisely:

Lemma 1.1.

$$\lim_{k\to\infty} A^k = 0 \iff \rho(A) < 1 \ . \tag{1.4}$$

Proof. Let us suppose that $A^k \to 0$, and that v is an eigenvector, associated to the eigenvalue λ. Thus $Av = \lambda v$, and $A^k v = \lambda^k v$, which tends to 0 for $k \to \infty$ only if $|\lambda| < 1$. As this must be true for any λ of A, we conclude that $\rho(A) < 1$ is necessary.

On the other hand, we know there exists an invertible T such that $A = TJT^{-1}$, with J the Jordan form of A. Using the relation

$$A^k = (TJT^{-1})^k = TJ^kT^{-1} \ ,$$

we conclude that the convergence to zero of successive powers of A is equivalent to the convergence to zero of the successive powers of all Jordan blocks of A. And it is a standard result (see [36]) that the d^{th} power of a $(k \times k)$ Jordan block $J(\lambda_i)$ is, for $d \geq k-1$:

$$[J(\lambda_i)]^d = \begin{bmatrix} \lambda_i^d & \binom{d}{1}\lambda_i^{d-1} & \cdots & \binom{d}{k-1}\lambda_0^{d-k+1} \\ & \lambda_i^d & \ddots & \vdots \\ & & \ddots & \binom{d}{1}\lambda_i^{d-1} \\ 0 & & & \lambda_i^d \end{bmatrix} \tag{1.5}$$

where

$$\binom{d}{j} = \frac{d(d-1)\dots(d-j+1)}{1(2)\dots(j)} \ .$$

This can easily be shown by induction.

So, if we suppose that $\rho(A) < 1$ (and so, $|\lambda_i| < 1, \forall i$) it is clear from the above expression that the powers of all the Jordan blocks of A converge to zero. On the other hand, for any $|\lambda_i| \geq 1$, we do not have convergence to zero. $\square$

From this convergence property, we can view the spectral radius of a matrix as a quantity describing the asymptotic behavior of successive powers of this matrix.

We take here the opportunity to mention that an alternative proof is presented in [45] (5.6.12, p. 298). It uses the fact that, if $\rho(A) < 1$, there exists a matrix norm $\|.\|_*$ such that $\|A\|_* < 1$ (by Lemma 5.6.10 from [45]). Then, by submultiplicativity, $\|A^k\|_* \leq \|A\|_*^k \to 0$, as $k \to \infty$.

This result stating the existence of a norm $\|.\|_*$ such that $\|A\|_* < 1$ when $\rho(A) < 1$ is worth picking out. We will see later that there exists a similar result for the sets of matrices. Section 4.3.1 presents the definition of what is called an extremal norm of a set of matrices, and Section 4.3 discusses the existence of such norms. Namely, Theorem 4.16 is analogous to this norm existence for a single matrix.

Inequality with the norm

The following inequality will play a key role in the study of the joint spectral radius (to be defined later in this chapter).

Let $A \in \mathbb{R}^{n\times n}$, and $\rho(A) = \lambda^*$, with the corresponding eigenvector $v^* \in \mathbb{C}^n$, such that $Av^* = \lambda^* v^*$. Let us define a matrix $V \in \mathbb{C}^{n\times n}$, of which the columns are the column vector v^* duplicated n times. Using this construction, we have

$$AV = \lambda^* V \Rightarrow |\lambda^*|\ \|V\| = \|\lambda^* V\| = \|AV\| \leq \|A\|\ \|V\|\ ,$$

using the submultiplicativity property of the matrix norm, which is mentioned earlier in Section 1.1. So, if $\|.\|$ is any submultiplicative matrix norm, and if $A \in \mathbb{R}^{n\times n}$, then

$$\rho(A) \leq \|A\|\ . \tag{1.6}$$

The spectral radius as a limit of norms

From the inequality (1.6), we can derive, using property (1.3), $\forall k \in \mathbb{N}$:

$$\begin{aligned} \rho(A^k) &\leq \|A^k\| \\ \rho(A)^k &\leq \|A^k\| \\ \rho(A) &\leq \|A^k\|^{1/k}\ . \end{aligned} \tag{1.7}$$

Now, for any matrix A, and for any constant $\varepsilon > 0$, we can define the matrix $\tilde{A} = [\rho(A)+\varepsilon]^{-1}A$. By construction, the spectral radius of $\tilde{A}$, $\rho(\tilde{A})$, is less than 1 and the successive powers of $\tilde{A}$ converge to the zero matrix, that is:

$$\rho(\tilde{A}) < 1 \Rightarrow \lim_{k\to\infty} \|\tilde{A}^k\| = 0\ .$$

From this, we have

$$\begin{aligned} \forall A \in \mathbb{R}^{n\times n}, \forall\varepsilon > 0, \exists N(A,\varepsilon) : \forall k \geq N, \|\tilde{A}^k\| &< 1 \\ \|A^k\| &< (\rho(A)+\varepsilon)^k \\ \|A^k\|^{1/k} &< \rho(A)+\varepsilon\ . \end{aligned}$$

Considering equation (1.7), we can then write $\forall A \in \mathbb{R}^{n\times n}, \forall\varepsilon > 0, \exists N(A,\varepsilon)$:

$$\forall k \geq N, \quad \rho(A) \leq \|A^k\|^{1/k} < \rho(A)+\varepsilon\ .$$

This means that

$$\rho(A) = \lim_{k \to \infty} \|A^k\|^{1/k} \ . \tag{1.8}$$

Here, we have shown how the spectral radius of a matrix can be viewed as a limit of norms. Moreover, this relation confirms what we said earlier, after expressing the convergence condition for the successive powers of matrix. Indeed, the spectral radius of A gives the asymptotic growth rate of the norm of $\|A^k\|$:

$$\text{as } k \to \infty, \ \|A^k\| \sim \rho(A)^k \ .$$

Note that from this property (1.8), we can re-derive the relation (1.3). Indeed, we can write

$$\begin{aligned} \rho(A^n) &= \lim_{k \to \infty} \|A^{nk}\|^{1/k} \\ &= \lim_{k' \to \infty} \|A^{k'}\|^{n/k'} \ , \text{posing } k' = nk \\ &= \Big(\lim_{k' \to \infty} \|A^{k'}\|^{1/k'} \Big)^n \ , \end{aligned}$$

which shows again that

$$\rho(A^n) = \rho(A)^n \ . \tag{1.9}$$

This confirms that Equation (1.8) can serve as an equivalent definition of the spectral radius of a matrix. One can indeed show that both definitions would yield the same properties.

1.2 Joint spectral radius of a set of matrices

The spectral radius of a matrix has been introduced and some of its properties have been reviewed. One might wonder whether such a notion holds for a set of matrices. This section presents a generalization of the notion of spectral radius to a set of matrices. Although this extension seems fairly natural, it gives rise to some issues. Namely, one has to understand the meaning (if any) of the spectral radius of a set of matrices. Also, its possible link with a notion of convergence of products of matrices, like the property in Lemma 1.1, has to be checked. These points will be discussed later in this work.

From a historical perspective, the formal definition of the joint spectral radius was first introduced by Rota & Strang in the 60's (see [72]). In the 90's, Daubechies & Lagarias defined the generalized spectral radius (see [26]), and Berger & Wang (see [4]) proved later these two values to be equal for *bounded* sets of matrices.

At this point, we can insist on the difference between a finite set and a bounded set. A set $\mathcal{M}$ is said to be *finite* if it consists in a finite number of matrices, while it is *bounded* if there exists an upper bound K on the norms of the matrices of $\mathcal{M}$. As a simple illustration in $\mathbb{R}$, a segment $[a, b]$ (with $a \neq b$) is bounded but contains an infinite number of points.

The notion of joint spectral radius of a set of matrices appears in a wide range of contexts and has led to a number of recent contributions (see, e.g., [12, 13, 28, 42, 51, 79, 81, 83]). A survey presenting a list of over hundred related contributions was carried out by Gilbert Strang in [78].

We here present the definitions of those values, and complete these with more recent results that help to put things into perspective.

1.2.1 Definitions

There are two natural generalizations of the notion of spectral radius to a set of matrices. The first is based on the spectral radius and the second on matrix norms.

We consider a (non necessarily bounded) set $\mathcal{M}$ of $n \times n$ matrices A_i :

$$\mathcal{M} = \{A_i : i \in I\} ,$$

where I is the set of indices. These matrices are linear operators from $\mathbb{R}^n$ to $\mathbb{R}^n$. As introduced before, we will denote by $\|.\|$ any submultiplicative matrix norm.

Generalized spectral radius - GSR

Using similar notations to those found in [26], we first define

$$\rho_k(\mathcal{M}) := \sup \left\{ \rho \left(\prod_{i=1}^{k} M_i \right) : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k \right\} . \tag{1.10}$$

For a given k, $\rho_k(\mathcal{M})$ represents the largest possible spectral radius of all products of k matrices chosen freely in the set $\mathcal{M}$. This serves as a basis for the definition of the *generalized spectral radius*, which is

$$\rho(\mathcal{M}) := \limsup_{k \to \infty} \left(\rho_k(\mathcal{M})\right)^{1/k} . \tag{1.11}$$

The generalized spectral radius is thus the maximal asymptotic spectral radius of the products of matrices that can be constructed using the set $\mathcal{M}$. The $1/k$ exponent serves as a normalization and allows for the GSR to be interpreted as a growth rate (of the spectral radius). Note that this definition only relies on the notion of spectral radius, and takes the supremum over all possible products in the considered set.

Joint spectral radius - JSR

We will define the joint spectral radius in a similar way as we did with the generalized spectral radius, but this time, using the property (1.8), which was

$$\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k} \ .$$

Considering the spectral radius as a limit of norms, this leads to

$$\hat{\rho}_k(\mathcal{M}, \|.\|) := \sup\left\{ \left\| \prod_{i=1}^{k} M_i \right\| : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k \right\} \ , \tag{1.12}$$

which is very similar to (1.10): here, $\hat{\rho}_k(\mathcal{M})$ represents the largest possible norm of all products of k matrices chosen in the set $\mathcal{M}$. This is all we need in order to define the *joint spectral radius* :

$$\hat{\rho}(\mathcal{M}) := \limsup_{k\to\infty} \left(\hat{\rho}_k(\mathcal{M}, \|.\|)\right)^{1/k} \ . \tag{1.13}$$

The generalized spectral radius is thus the maximal asymptotic norm of the products of matrices that can constructed using the set $\mathcal{M}$. As before, the $1/k$ exponent is for normalization and allows for the JSR to be interpreted as a growth rate (of the norm). We see that, thanks to the equivalence between matrix norms expressed in equation (1.1), the definition (1.12) does not depend on the norm we use. Indeed, rewriting this definition using equivalent norms yields

$$k_2 \hat{\rho}_k(\mathcal{M}, \|.\|_b) \leq \hat{\rho}_k(\mathcal{M}, \|.\|_a) \leq k_1 \hat{\rho}_k(\mathcal{M}, \|.\|_b) \ ,$$

and the constants k_1 and k_2 will have no influence on the value $\hat{\rho}(M)$, because of the $1/k$ exponent when $k \to \infty$. And so the joint spectral radius is independent on the choice of the norm (induced or not). Further, this will lead us to drop the reference to the norm in the above notations for $\hat{\rho}_k(\mathcal{M})$, when not necessary for the comprehension of the reasoning.

Let us remark that both definitions of the joint and generalized spectral radius are equivalent to the original spectral radius when the set $\mathcal{M}$ trivially consists of only one matrix. And, as was shown for the spectral radius of a matrix, we also have a convergence criterion thanks to the joint spectral radius. More precisely, if $\rho(\mathcal{M}) < 1$, then all products converge to the zero matrix. This will be formally explained in Section 3.2.1.

1.2.2 Properties

Multiplication by a scalar

First, let us note that multiplying all the matrices of $\mathcal{M}$ with a scalar μ (considered to be real, in this text) also multiplies the JSR and the GSR in the same way:

$$\forall \mu \in \mathbb{R},\ \hat{\rho}(\mu\mathcal{M}) = \mu\hat{\rho}(\mathcal{M}) ,$$
$$\rho(\mu\mathcal{M}) = \mu\rho(\mathcal{M}) .$$

This is straightforward from the definitions.

Equality between JSR and GSR and continuity

The definitions of the JSR and the GSR are very similar. But despite all links between matrix norms and the spectral radius, it is not trivial whether these two values are actually equal or not. At first, they have been conjectured to be equal for bounded sets of matrices ([26]). And, as shown by Berger & Wang in [4], this equality holds for finite sets of matrices and, even stronger, for bounded sets of matrices :

Theorem 1.2. *For any bounded set $\mathcal{M}$ of matrices,*

$$\rho(\mathcal{M}) = \hat{\rho}(\mathcal{M}) .$$

Elsner has provided an alternative proof of this result in [34], of which we present here an abstract. The outline presented here is not meant to be a fluid reading of the full proof, as the full details of it can be found in the above reference. We are more interested in the structure of the proof, as well as detailing the intermediate results, namely Lemma 1.6, which will be used later.

Proof (abstract). This proof successively shows the following lemmas.

Lemma 1.3.
$$\rho(\mathcal{M}) = \inf_{\|\cdot\|} \sup_{A_i \in \mathcal{M}} \|A_i\| .$$

This lemma is important, as it shows that the joint spectral radius can be seen as the infimum over all possible matrix norms of the largest norm of the matrices in the set. This property has been noted in several papers, such as [34, 57, 72]. We will make use of this property later, namely in Section 6.5, to approximate the JSR. We will also study, in Chapter 4, a theoretical construction (by Kozyakin) of a particular extremal norm (see 4.14 for the definition of an extremal norm). Moreover, this property teaches us that the joint spectral radius of a set $\mathcal{M}$ is unchanged when taking the convex hull defined by the elements of $\mathcal{M}$ or its closure. Indeed, it is clear that

$$\sup_{A_i \in \mathcal{M}} \|A_i\| = \sup_{A_i \in conv\mathcal{M}} \|A_i\| = \sup_{A_i \in cl\mathcal{M}} \|A_i\| .$$

We now continue with the proof of Theorem 1.2, and introduce the following lemma.

Lemma 1.4. *Let $\|.\|$ denote a vector norm on $\mathbb{C}^k$ and its operator norm in the space of $(k \times k)$ matrices. There exists a constant C depending on $\|.\|$ such that for any $z \in \mathbb{C}^k$, $\|z\| = 1$, and any $(k \times k)$ matrix A with $\|A\| \leq 1$ and eigenvalues $\lambda_1, \dots, \lambda_k$, the following inequality holds:*

$$\min_i |1 - \lambda_i| \leq C\|Az - z\|^{1/k} .$$

We left the reference to $\mathbb{C}^k$ in this last lemma, as it is the way it was presented by Elsner, although we consider matrices in $\mathbb{R}^k$. These two lemmas are used to prove the following, which is a first restricted version of the targeted theorem:

Lemma 1.5. *Let $\mathcal{M} = \{A_1, \dots, A_m\}$ be finite, product bounded, and such that $\hat{\rho}(\mathcal{M}) = 1$. Then $\rho(\mathcal{M}) = \hat{\rho}(\mathcal{M})$.*

A set of matrices is *product bounded* if there exists a finite upper bound on the norms of all the possible matrix products that can be generated using this set.

Lemma 1.6. *Let $\mathcal{M}$ be a bounded set of complex $(k \times k)$ matrices, $\rho(\mathcal{M}) = 1$. If $\mathcal{M}$ is not product bounded, then there is an invertible matrix S and $1 \leq n_1 < k$ such that, $\forall A \in \mathcal{M}$,*

$$S^{-1}AS = \begin{pmatrix} A_{(2)} & * \\ 0 & A_{(1)} \end{pmatrix} ,$$

where $A_{(1)}$ is $(n_1 \times n_1)$.

This lemma is a noteworthy result by itself. The existence of a matrix S putting simultaneously all matrices of $\mathcal{M}$ into an upper block triangular form means that all matrices share a common invariant subspace of dimension $k - n_1$. The set $\mathcal{M}$ is then called a *reducible* set (this notion will be formally defined in Section 4.3.1).

At this point, these lemmas allow to prove the theorem for a finite set of matrices. Then, a last lemma is needed.

Lemma 1.7. *If $\mathcal{M}$ is a bounded set of real $(k \times k)$ matrices, then*

$$\hat{\rho}(\mathcal{M}) = \sup\{\hat{\rho}(\tilde{\mathcal{M}}) : \tilde{\mathcal{M}} \subset \mathcal{M},\ \tilde{\mathcal{M}} \text{ finite } \} .$$

This allows to deduce the theorem for bounded sets. □

In [44], a consequence of this theorem is pointed out. We just recall the following before giving the idea:

Definition 1.8. *Let X be a topological space, and f a function from X into the extended real numbers $\overline{\mathbb{R}}$; $f : X \to \overline{\mathbb{R}}$. Then:*

- *If $\{x \in X \mid f(x) > \alpha\}$ is an open set in X for all $\alpha \in \mathbb{R}$, then f is said to be lower semicontinuous.*

- *If $\{x \in X \mid f(x) < \alpha\}$ is an open set in X for all $\alpha \in \mathbb{R}$, then f is said to be upper semicontinuous.*

We can say that $\rho(\mathcal{M})$ is a lower-semicontinuous function in the entries of $\mathcal{M}$, as it is a supremum of continuous functions (eigenvalues of products). Similarly, $\hat{\rho}(\mathcal{M})$ is upper-semicontinuous in the entries of $\mathcal{M}$, as it is an infimum of continuous functions (norms of products). The equality yields that $\rho(\mathcal{M})$ is both upper and lower-semicontinuous, so it is continuous in the entries of $\mathcal{M}$. This continuity is also to be found in the work of Barabanov ([3]).

This continuity result is also obtained by Wirth in [85], as well as the local Lipschitz continuity of the GSR on the set of compact *irreducible sets* of matrices (see Section 4.3.1 for a definition of irreducible sets).

Knowing the above Theorem 1.2, when considering bounded sets of matrices in this thesis, we may use the terminology joint spectral radius, while using the definition of the generalized spectral radius. On the other hand, the equality for unbounded sets of matrices does not hold in general. We present a counterexample below.

Counterexample for unbounded sets

Here is an example showing that the equality $\rho(\mathcal{M}) = \hat{\rho}(\mathcal{M})$ does not hold in general for an unbounded set of matrices. First consider the finite (and bounded) set:

$$\mathcal{M}_1 = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \right\} .$$

It yields $\rho(\mathcal{M}_1) = \hat{\rho}(\mathcal{M}_1) = 1$. And the similar unbounded set

$$\mathcal{M}_2 = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}, \ldots \right\}$$

yields $\rho(\mathcal{M}_2) = 1$, while $\hat{\rho}(\mathcal{M}_2) = +\infty$. Although these values can be considered as straightforward to compute, we provide the explicit computations. Indeed, we feel it appropriate to provide some details about the computation of the JSR, at this point in the thesis.

We first study the set:

$$\mathcal{M}_1 = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \right\}$$

with N an arbitrarily chosen natural number. We will use the following notation:

$$M_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \text{ for } k = 0 \ldots N .$$

Let us note that the product of two such matrices like

$$\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} \tag{1.14}$$

is still a matrix of the very same form, with $\gamma = \alpha + \beta$.

In order to determine the joint spectral radius of the set $\mathcal{M}$, using the definitions (1.12) and (1.13), we have to choose a norm. As any matrix norm is suitable for our needs, we will use either the maximum absolute row sum norm $\|.\|_\infty$ or the maximum absolute column sum norm $\|.\|_1$, because they are easy to evaluate in our case. Indeed, we have

$$\left\| \begin{matrix} 1 & \alpha \\ 0 & 1 \end{matrix} \right\|_1 = \left\| \begin{matrix} 1 & \alpha \\ 0 & 1 \end{matrix} \right\|_\infty = 1 + \alpha \ .$$

Knowing this, we just have to form a product of matrices of the same type as the one presented in (1.14), achieving the largest norm. So, we aim for the largest possible value of γ, which is obtained by taking the largest possible values for α and β (because $\gamma = \alpha + \beta$).

As seen in the definition (1.12), we have to take the supremum of such a norm on all possible matrix products of a given length k. Recalling what was said a few lines before, we know that this supremum is here achieved by always using the matrix M_N in all the considered products. And so, we have

$$\begin{aligned} \hat{\rho}_k(\mathcal{M}, \|.\|_1) = \|(M_N)^k\|_1 &= \left\| \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}^k \right\|_1 \\ &= \left\| \begin{pmatrix} 1 & kN \\ 0 & 1 \end{pmatrix} \right\|_1 \\ &= 1 + kN \ . \end{aligned} \tag{1.15}$$

Using the definition (1.13), this result (1.15) leads to

$$\hat{\rho}(\mathcal{M}) = \limsup_{k\to\infty} (1 + kN)^{1/k} = 1 \ .$$

The computation of $\rho(\mathcal{M})$, as defined in (1.10) and (1.11), is much easier. Indeed, as can be seen from (1.14), all possible matrix products of length k admit a spectral radius equal to 1 and so, $\rho(\mathcal{M}) = 1$. In this first example, $\rho(\mathcal{M}) = \hat{\rho}(\mathcal{M})$ effectively holds. This was expected, since $\mathcal{M}$ is here bounded.

Let us now derive the effective counterexample where $\rho(\mathcal{M}) \neq \hat{\rho}(\mathcal{M})$. It comes from a rather natural generalization of the above example. We just need to extend the set to an unbounded one using the same definition and letting $N \to \infty$. Now the set $\mathcal{M}$ is :

$$\mathcal{M} = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}, \ldots \right\} \ .$$

The value which was previously referred to as N in the case of the finite set and which we will denote by N', can now be taken as large as we want. In particular, in the expression (1.15), where we had to choose the largest N available, N' is not bounded any more. So,

$$\forall k, \forall N' \in \mathbb{N},\ \hat{\rho}_k(\mathcal{M}, \|.\|_1) \geq 1 + kN' \ .$$

Note that this expression $\to \infty$ as $N' \to \infty$, for a fixed k. Consequently, $\hat{\rho}_k(\mathcal{M}, \|.\|_1)$ is actually unbounded. The question is to know how this affects the value of $\hat{\rho}(\mathcal{M})$. To find out, in this case, having no upper bound on N', we can do the computation while arbitrarily choosing for any k that $N' := N'(k) \geq 2^k$. This is still an underestimation of $\hat{\rho}_k(\mathcal{M}, \|.\|_1)$, but it will prove to be sufficient for our proof :

$$\begin{aligned}\hat{\rho}(\mathcal{M}) &\geq \limsup_{k\to\infty} \left(1 + kN'\right)^{1/k} \\ &\geq \limsup_{k\to\infty} \left(1 + k\ 2^k\right)^{1/k} \geq \limsup_{k\to\infty} \left(2^k\right)^{1/k} = 2.\end{aligned}$$

Instead of choosing $N' \geq 2^k$, we could have chosen any $N' \geq a^k$ (for any $a \in \mathbb{R}$), which would yield $\hat{\rho}(\mathcal{M}) \geq a, \forall a \in \mathbb{R}$, proving $\hat{\rho}(\mathcal{M})$ to be actually unbounded. On the other hand, the spectral radius of any matrix product is still equal to 1, as was the case when considering the finite set of matrices. Consequently,

$$\rho(\mathcal{M}) = 1 \ .$$

This concludes the example where, for an unbounded set $\mathcal{M}$ of matrices, the joint and generalized spectral radius do not coincide :

$$\rho(\mathcal{M}) = 1 \neq \hat{\rho}(\mathcal{M}) = +\infty \ . \tag{1.16}$$

Four members inequality

We now present an important inequality, as it will be used extensively in order to determine lower and upper bounds of the joint spectral radius. This inequality was first presented in [26] (see Lemma 3.1 therein), but its proof was completed later on by the same authors in [28]. We present here the whole idea of their proof.

Lemma 1.9. *For any set of matrices $\mathcal{M}$ and any $k \geq 1$,*

$$\rho_k(\mathcal{M})^{1/k} \leq \rho(\mathcal{M}) \leq \hat{\rho}(\mathcal{M}) \leq \hat{\rho}_k(\mathcal{M})^{1/k} \ , \tag{1.17}$$

independently of the induced (submultiplicative) norm used to define $\hat{\rho}_k(\mathcal{M})$.

Proof. Let us start with the first inequality: $\rho_k(\mathcal{M})^{1/k}$ has to be $\leq \rho(\mathcal{M})$. Let us recall the definition (1.10):

$$\rho_k(\mathcal{M}) := \sup\left\{\rho\left(\prod_{i=1}^{k} M_i\right) : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k\right\} .$$

From this definition, we deduce that $\left(\rho_k(\mathcal{M})\right)^j \leq \rho_{kj}(\mathcal{M})$. Indeed, all matrix products that can be expressed as the j^{th} power of a product of k matrices, are included in the set of products of kj matrices. And the property (1.9), $\rho(A^n) = \rho(A)^n$ does the rest. We can then rewrite this inequality as:

$$\left(\rho_k(\mathcal{M})\right)^{1/k} \leq \left(\rho_{kj}(\mathcal{M})\right)^{1/kj} ,$$

which yields, letting $j \to \infty$,

$$\left(\rho_k(\mathcal{M})\right)^{1/k} \leq \limsup_{j\to\infty} \left(\rho_{kj}(\mathcal{M})\right)^{1/kj} .$$

At this point, we can make use of the following :

$$\limsup_{j\to\infty} \left(\rho_{kj}(\mathcal{M})\right)^{1/kj} \leq \limsup_{k\to\infty} \left(\rho_k(\mathcal{M})\right)^{1/k} ,$$

as the left member terms are a subsequence of the right member terms. The second member matches the definition (1.11) of $\rho(\mathcal{M})$, and we deduce that

$$\left(\rho_k(\mathcal{M})\right)^{1/k} \leq \rho(\mathcal{M}) .$$

The next inequality, $\rho(\mathcal{M}) \leq \hat{\rho}(\mathcal{M})$, is a direct consequence of the property $\rho(A) \leq \|A\|$ and the definitions of $\rho_k(\mathcal{M})$, $\hat{\rho}_k(\mathcal{M})$ and the definitions of $\rho(\mathcal{M})$ and $\hat{\rho}(\mathcal{M})$. In short, we have successively:

$$\forall k,\ \sup\left\{\rho\left(\prod_{i=1}^{k} M_i\right) : M_i \in \mathcal{M}\right\} \leq \sup\left\{\left\|\prod_{i=1}^{k} M_i\right\| : M_i \in \mathcal{M}\right\} ,$$
$$\text{and this leads to } \ \rho_k(\mathcal{M})^{1/k} \leq \hat{\rho}_k(\mathcal{M})^{1/k},\ \forall k ,$$

leading to the expected $\rho(\mathcal{M}) \leq \hat{\rho}(\mathcal{M})$.

We finally deal with the last inequality, $\hat{\rho}(\mathcal{M}) \leq \hat{\rho}_k(\mathcal{M}, \|.\|)^{1/k}$, the proof of which is in 2 steps:

1. First, let us suppose there exists a finite bound on the norm of the matrices M_i belonging to the set $\mathcal{M}$: $\|M_i\| \leq K,\ \forall M_i \in \mathcal{M}$. We pick two natural numbers l and k such that $l \geq k$, rewrite l as $mk + j$, with $0 \leq j \leq k-1$, and, using the submultiplicativity property mentioned in 1.1, we can write

$$\begin{aligned}\|M_1 \dots M_l\| &\leq \prod_{i=0}^{m-1} \left(\|M_{ik+1} M_{ik+2} \dots M_{ik+k}\|\right) \prod_{i=1}^{j} \left(\|M_{mk+i}\|\right) \\ &\leq (\hat{\rho}_k(\mathcal{M}, \|.\|))^m K^j \text{ , by the definitions of } \rho_k \text{ and } K \text{ .}\end{aligned}$$

In order for the left member to pass from $\|M_1 \dots M_l\|$ to $\hat{\rho}(\mathcal{M})$, we will successively take the supremum over all products of length l, raise to the power $1/l$ and then take the lim sup of the expression for $l \to \infty$.

$$\begin{aligned}\hat{\rho}_l(\mathcal{M}) &\leq (\hat{\rho}_k(\mathcal{M}, \|.\|))^m K^j = (\hat{\rho}_k(\mathcal{M}, \|.\|))^{\frac{l-j}{k}} K^j \\ \hat{\rho}_l(\mathcal{M})^{1/l} &\leq (\hat{\rho}_k(\mathcal{M}, \|.\|))^{\frac{1}{k} - \frac{j}{kl}} K^{\frac{j}{l}} \\ \hat{\rho}(\mathcal{M}) &\leq (\hat{\rho}_k(\mathcal{M}, \|.\|))^{\frac{1}{k}} \text{ .}\end{aligned}$$

2. We now have to deal with the case where a bound such as K defined above does not exist. In such a case, it is clear that $\hat{\rho}_1(\mathcal{M}, \|.\|) = \infty$. Note that this is only possible for an unbounded set. If $\hat{\rho}_k(\mathcal{M}, \|.\|) = \infty,\ \forall k \geq 1$, then the desired inequality holds. We may then assume that $\exists k : \hat{\rho}_k(\mathcal{M}, \|.\|) < \infty$. At this point in the proof, we skip some technical details, which would make the thread of the ideas hard to follow. To be more precise, the following implication

$$\exists k : \hat{\rho}_k(\mathcal{M}, \|.\|) < \infty \Rightarrow \exists N : \forall k \geq N,\ \hat{\rho}_k(\mathcal{M}, \|.\|) < \infty$$

is the main object of the publication [28]. Its only proof extends onto several pages and we will take it for granted.
This property allows us to deduce that if $\exists\ k : \hat{\rho}_k(\mathcal{M}, \|.\|) < \infty$, then $\exists\ l > N$ with l coprime to k such that $\hat{\rho}_l(\mathcal{M}, \|.\|) < \infty$. Also, for any $n > kl$, there exist integers t and s such that $n = tk + sl$ with $t > 0$ and $0 \leq s < k$. We can then write

$$\begin{aligned}\hat{\rho}_n(\mathcal{M}, \|.\|)^{1/n} &\leq \left(\hat{\rho}_k(\mathcal{M}, \|.\|)^t \hat{\rho}_l(\mathcal{M}, \|.\|)^s\right)^{1/n} \\ &\leq \left(\hat{\rho}_k(\mathcal{M}, \|.\|)^{(n-sl)/k} \hat{\rho}_l(\mathcal{M}, \|.\|)^s\right)^{1/n} \text{ .}\end{aligned}$$

Letting $n \to \infty$, we get

$$\hat{\rho} \leq \hat{\rho}_k(\mathcal{M}, \|.\|)^{1/k} \text{ .}$$

□

We may remark that this last inequality can be used to simplify the definition (1.13), which was

$$\hat{\rho}(\mathcal{M}) := \limsup_{k \to \infty} \left(\hat{\rho}_k(\mathcal{M}, \|.\|)\right)^{1/k} \text{ ,}$$

in order to get rid of the lim sup and replace it with a more convenient lim :

$$\hat{\rho}(\mathcal{M}) := \lim_{k\to\infty} \left(\hat{\rho}_k(\mathcal{M}, \|.\|)\right)^{1/k} .$$

So, from now on, we may omit lim sup in this definition (and use lim).

Reducible sets of matrices

We will be using the following notion:

Definition 1.10. *A set $\mathcal{M}$ of matrices is* irreducible *if only the trivial subspaces $\{0\}$ and $\mathbb{R}^n$ are invariant under all matrices $A_i \in \mathcal{M}$. Otherwise, it is* reducible.

We refer to lemma 1.6, where we discussed the existence of an invertible S such that, $\forall A_i \in \mathcal{M}$,

$$S^{-1}A_iS = \begin{pmatrix} A_{(2)} & * \\ 0 & A_{(1)} \end{pmatrix} .$$

Clearly, in this case, there is a nontrivial subspace that is left invariant by all the matrices of the set.

Reciprocally, the reducible sets of matrices have the following important property:

Proposition 1.11. *If a set $\mathcal{M}$ of matrices is reducible, then there exists an invertible matrix T such that all the matrices of $\mathcal{M}$ can be put simultaneously in a block upper-triangular form:*

$$\forall A_i \in \mathcal{M}, T^{-1}A_iT = \begin{pmatrix} B_i & * \\ 0 & C_i \end{pmatrix} ,$$

with B_i and C_i square, and $*$ is any matrix with suitable dimensions. The joint spectral radius $\rho(\{A_i\})$ is equal to

$$\max\left(\rho(\{B_i\}), \rho(\{C_i\})\right) .$$

What is new here is the last part of the proposition. Indeed, the existence of a specific base under which the matrices can be expressed in upper block triangular form is straightforward from the definition of reducibility.

Proof. To see how to prove the claim about the JSR of $\{A_i\}$, we use the fact that the spectrum of a matrix of the type $\begin{pmatrix} B & * \\ 0 & C \end{pmatrix}$ is the union of the spectra of B and C, since

$$\det \begin{pmatrix} B-\lambda I & * \\ 0 & C-\lambda I \end{pmatrix} = \det(B-\lambda I)\det(C-\lambda I) ,$$

where the I matrices are the identities with the suitable dimensions. The second property to use is that

$$\forall i, j, \begin{pmatrix} B_i & * \\ 0 & C_i \end{pmatrix} \begin{pmatrix} B_j & * \\ 0 & C_j \end{pmatrix} = \begin{pmatrix} B_i B_j & * \\ 0 & C_i C_j \end{pmatrix} .$$

As a consequence, the products between the B_i's and the C_i's do not interfere, and the largest spectral radius among the products in $\{A_i\}$ will be the maximum of the spectral radii among the products in $\{B_i\}$ or $\{C_i\}$. □

1.3 *Lower* spectral radius of a set of matrices

1.3.1 Definition

As presented by Gurvits in an appendix of [42] or by Blondel & Tsitsiklis in [79], the developments presented above can be extended to a minimum version. These references use different terminologies, respectively *joint/generalized spectral subradius* and *lower spectral radius*. This time, instead of taking the sup in the definitions (1.10) and (1.12), we take the inf. And in (1.11) and (1.13), we take lim inf instead of lim sup. This leads to

$$\rho_{*k}(\mathcal{M}) := \inf \left\{ \rho \left(\prod_{i=1}^{k} M_i \right) : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k \right\} \tag{1.18}$$

$$\rho_{*}(\mathcal{M}) := \liminf_{k\to\infty} \left(\rho_{*k}(\mathcal{M})\right)^{1/k} \tag{1.19}$$

for the generalized spectral *sub*radius and

$$\hat{\rho}_{*k}(\mathcal{M}) := \inf \left\{ \left\| \prod_{i=1}^{k} M_i \right\| : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k \right\} \tag{1.20}$$

$$\hat{\rho}_{*}(\mathcal{M}) := \liminf_{k\to\infty} \left(\hat{\rho}_{*k}(\mathcal{M}, \|.\|)\right)^{1/k} \tag{1.21}$$

for the joint spectral *sub*radius. We will see that there is no need any more to distinguish the joint from the generalized spectral subradius, so that both can be called lower spectral radius.

At this point, we may stress on the difference between the JSR and the lower spectral radius. The JSR rules the convergence to zero of all the possible matrix products and can be seen as a *stability* measure. On the other hand, the lower spectral radius expresses the possibility of finding a matrix sequence that tends to zero. It can consequently be considered as a *stabilizability* measure. In Chapter 2, we will be presenting applications of the JSR. We could think of similar applications for the lower spectral radius. However, as this is not the main subject of this thesis, we will remain focused on the JSR.

1.3.2 Properties

In a similar way as the equality $\rho(\mathcal{M}) = \hat{\rho}(\mathcal{M})$ presented in section 1.2.2 for bounded sets, we now prove that these two values are equal. The proof is similar to the idea presented in [42], yet giving more details.

Lemma 1.12. *For any (not necessarily bounded) set of matrices $\mathcal{M}$,*

$$\rho_*(\mathcal{M}) = \hat{\rho}_*(\mathcal{M}) \ . \tag{1.22}$$

This confirms that $\hat{\rho}_*(\mathcal{M})$ does not depend on the norm chosen in its definition, just like in the case of the joint spectral radius.

Proof. Knowing that $\rho(A) \leq \|A\|$, we successively get that

$$\begin{aligned}
\rho_{*k}(\mathcal{M}) &\leq \hat{\rho}_{*k}(\mathcal{M}), \forall k \in \mathbb{N}_0 \text{ , for any set } \mathcal{M} \\
\rho_{*k}^{1/k}(\mathcal{M}) &\leq \hat{\rho}_{*k}^{1/k}(\mathcal{M}) \\
\inf_k \rho_{*k}^{1/k}(\mathcal{M}) &\leq \inf_k \hat{\rho}_{*k}^{1/k}(\mathcal{M}) \\
\rho_*(\mathcal{M}) &\leq \hat{\rho}_*(\mathcal{M}) \ .
\end{aligned} \tag{1.23}$$

Let us now arbitrarily choose a number $n_0 \in \mathbb{N}$. For any $\varepsilon > 0$, we can find a product $A_{i_1} \dots A_{i_{n_0}}$, with $A_i \in \mathcal{M}$, such that its spectral radius is close enough to $\rho_{*n_0}(\mathcal{M})$, that is to say

$$\rho_{*n_0}(\mathcal{M}) + \varepsilon \geq \rho(A_{i_1} \dots A_{i_{n_0}}) \ .$$

We have

$$\hat{\rho}_{*n_0 k}^{(1/n_0 k)}(\mathcal{M}) \leq \|(A_{i_1} A_{i_2} \dots A_{i_{n_0}})^k\|^{(1/n_0 k)}, \ \forall k \ .$$

We now let $k \to \infty$, and this leads to

$$\begin{aligned}
\liminf_{k\to\infty} \left(\hat{\rho}_{*n_0 k}^{(1/n_0 k)}(\mathcal{M})\right) \leq \lim_{k\to\infty} \left(\|(A_{i_1} \dots A_{i_{n_0}})^k\|^{(1/n_0 k)}\right) &= \rho(A_{i_1} \dots A_{i_{n_0}})^{\frac{1}{n_0}} \\
&= \rho_{*n_0}(\mathcal{M})^{1/n_0} \ .
\end{aligned}$$

Furthermore, about the first member of this inequality, we can write, using the definition (1.21),

$$\hat{\rho}_*(\mathcal{M}) = \liminf_{k\to\infty} \left(\hat{\rho}_{*k}^{(1/k)}(\mathcal{M})\right) \leq \liminf_{k\to\infty} \left(\hat{\rho}_{*n_0 k}^{(1/n_0 k)}(\mathcal{M})\right) \ .$$

And so,

$$\hat{\rho}_*(\mathcal{M}) \leq \rho_{*n_0}(\mathcal{M})^{1/n_0}, \ \forall n_0 \in \mathbb{N} \ .$$

This implies

$$\hat{\rho}_*(\mathcal{M}) \leq \liminf_{n\to\infty} \rho_{*n}(\mathcal{M})^{1/n} = \rho_*(\mathcal{M}) \ . \tag{1.24}$$

The inequality (1.24), associated with the inequality (1.23), yields

$$\rho_* = \hat{\rho}_* \ .$$

□

1.4 Conclusion

This concludes the chapter devoted to the definition of the joint spectral radius. At this point, we have presented all the notions that are needed to understand the next two chapters. Chapter 2 is devoted to presenting the motivation behind the study of this notion, while Chapter 3 shows how the joint spectral radius relates to the stability of discrete time switched linear systems.

2

Motivations for the joint spectral radius

Now that the joint spectral radius has been introduced and formally defined, we can describe some practical uses that can be made of it. This chapter is devoted to the presentation of several research areas that involve the joint spectral radius. We will consider applications in the following areas:

- the coordination of autonomous agents,
- the continuity of wavelet basis functions,
- the capacity of codes.

These are detailed in the next sections.

We may also give references to the reader who might want to have a look at other applications. Information about *De Rham* curves, as well as subdivision and corner cutting algorithms can be found in [31, 2, 58, 70]. More generally, subdivision algorithms also make use of the joint spectral radius at some point, as can be seen in [59, 33, 21, 69].

2.1 Coordination of autonomous agents

Instead of detailing a long bibliography about multi-agent systems, we will limit ourselves to papers by Moreau ([63]) and Jadababaie, Lin & Morse ([48]). Indeed, these papers present sufficient details for anyone willing to make the link between this topic and the joint spectral radius.

We begin with a few elementary notions of graph theory. These are not absolutely necessary to understand the remainder of this chapter, but they provide a sound theoretical basis.

Definition 2.1. *A* directed graph *is a pair* $(\mathcal{V}, \mathcal{E})$*, where*

- $\mathcal{V}$ *is a nonempty, finite set. Its elements are called vertices.*

- *$\mathcal{E}$ is a subset of $\mathcal{V} \times \mathcal{V}$ such that $\forall j \in \mathcal{V}, (j, j) \notin \mathcal{E}$. An element (i, j) of $\mathcal{E}$ is an edge from node i to node j.*

Given a directed graph $(\mathcal{V}, \mathcal{E})$ and a nonempty subset of vertices $\mathcal{L} \subseteq \mathcal{V}$, the neighbor set *$\mathcal{N}(\mathcal{L}, \mathcal{E})$ is the set of vertices $k \in \mathcal{V} \backslash \mathcal{L}$ for which there is a $l \in \mathcal{L}$ such that $(k, l) \in \mathcal{E}$.*
A weighted directed graph *is a triple $(\mathcal{V}, \mathcal{E}, w)$, where $(\mathcal{V}, \mathcal{E})$ is a directed graph, and $w : \mathcal{E} \to \mathcal{R}^+$ associates a nonnegative weight w_{kl} to every edge (k, l).*

First, note that we excluded self-loops (edges that connect a vertex to itself). The motivation behind these definitions is that the connectivity between the agents of a multi-agent system can be described as a directed graph. More precisely, the agents can be seen as the vertices, while the edges are the communication channels between the agents. The neighbors of an agent will be the agents that are able to send information to it.

Let us consider a finite number, say n, of autonomous agents in the plane. These agents are assumed to be just points and are labeled from 1 to n. They are all moving with the same fixed speed, but with different and variable headings, denoted by θ_i. Note that the values of $\theta_i(t)$ are naturally supposed to lie in $[0, 2\pi[$. This is represented in Figure 2.1.

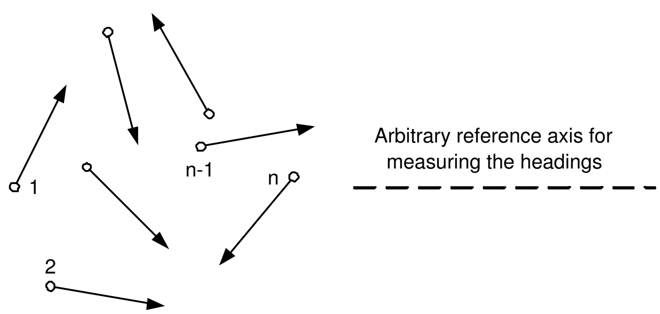

Fig. 2.1. Representation of n autonomous agents evolving in the plane.

The evolution of the whole system is considered in discrete time. At each time step, every element moves according to its velocity, and the direction of this velocity can be modified. The updated direction of an agent is given by taking a weighted average of its own direction and the directions of its neighbors. We will use the notation $\mathcal{N}_i(t)$ to denote the neighbors of agent i at time t, instead of the heavier $\mathcal{N}(\{i\}, \mathcal{E}(t))$.

From what was said above, the evolution law of the direction of an agent can be expressed as follows :

$$\theta_i(t+1) = \frac{\theta_i(t) + \sum_{j \in \mathcal{N}_i(t)} w_{ji}(t)\theta_j(t)}{1 + \sum_{j \in \mathcal{N}_i(t)} w_{ji}(t)} \ , \tag{2.1}$$

where $w_{ji}(t)$ denotes the weight of the link from vertex j to vertex i at instant t.

Of course, this evolution law depends on the chosen modeling. The above model refers to the Vicsek model, defined in [80] and studied in [48]. It is a linear approximation of the Kuramoto equation (see [86, 74]):

$$\dot{\theta}_i = \omega_i + \sum_{j=1}^{N} K_{ij} \sin(\theta_j - \theta_i), \ \ i = 1, \ldots, N \ .$$

The Kuramoto equation models a population of oscillators with sinusoidal coupling terms. We will not study this equation, but we simply mention that it is widely used in the field of the synchronization of coupled oscillators. The study of coupled oscillators has applications in several domains, such as engineering and physics (see [75]). It is also useful for modeling the swarming of some living beings, such as fish, birds or bacteria.

In this work, we will stick to the simpler model (2.1). Note that the denominator is intended for normalization. In matrix form, this gives:

$$\theta(t+1) = A_t \theta(t) \ ,$$

where A_t is given by :

$$(A_t)_{i,k} = \begin{cases} \frac{1}{1+\sum_{j \in \mathcal{N}_i(t)} w_{ji}(t)} & \text{if } k = i, \\ \frac{w_{ki}}{1+\sum_{j \in \mathcal{N}_i(t)} w_{ji}(t)} & \text{if } (k,i) \in \mathcal{E}(t), \\ 0 & \text{otherwise } . \end{cases}$$

Note that, the graph being directed, the *evolution matrix* A_t is not necessarily symmetric.

The following example illustrates the various notions introduced so far.

Example 2.2. Suppose that we have three agents and that the connectivity between them at time t is given by the weighted graph shown in Figure 2.2. The evolution law at this time t is then given by

$$\begin{aligned} \theta_1(t+1) &= \frac{\theta_1(t) + 4\ \theta_2(t)}{1+4} \\ \theta_2(t+1) &= \frac{\theta_1(t) + \theta_2(t) + 0.5\ \theta_3(t)}{1+1+0.5} \\ \theta_3(t+1) &= \theta_3(t) \ , \end{aligned}$$

which can be rewritten in matrix form :

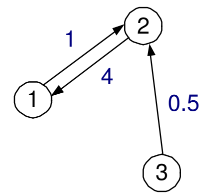

Fig. 2.2. Weighted graph

$$\theta(t+1) = \begin{pmatrix} 0.2 & 0.8 & 0 \\ 0.4 & 0.4 & 0.2 \\ 0 & 0 & 1 \end{pmatrix} \theta(t) \ .$$

As an agent moves, its relation with the other agents is bound to change with time. So, the connectivity of the graph evolves, and the evolution matrix A_t is likely to change at every time step. We are thus dealing with a linear time-varying system of the type

$$x(t+1) = A_t x(t) \ , \text{ where } \forall t, A_t \in \mathcal{M} \ .$$

In the problem at hand, the matrix A_t is *stochastic* : it is square, its rows all sum to 1, its entries are non-negative, and positive on the diagonal. This property plays an essential role in the study of this system. Indeed, it is easy to see that the products of stochastic matrices are stochastic matrices. Consequently, the spectral radii of all matrix products are equal to 1, and the joint spectral radius is then also equal to 1. Thus, under the current formulation, we are not facing a stability problem: the issue is not to determine if the headings of the system converge to zero. Note that the link between the joint spectral radius and the stability of a switched linear systems is detailed in Chapter 3.

The question here is then to know whether all variables (the angles/headings) will stabilize and/or reach the same value. This means that the issue is to determine whether all agents will eventually tend to move in the same direction.

In order to answer this question, we may mention results that are presented in [63]. In this paper, Moreau establishes that all the agents of such a system will eventually evolve in the same direction, under the following condition:

- there exists a T such that the union of any T consecutive graphs (in $[t_0, t_0 + T]$) is weakly connected.

A directed graph is *weakly connected* if it has a vertex from which there exists a directed path to every other vertex. The above criterion ensures that the connectivity of the successive graphs is sufficient for the information to spread and to let all agents converge to the same direction.

Up to this point, one could think that the joint spectral radius is not of much help here, since we know it is equal to 1, by construction. But, in [48], a result is presented that allows more to be said. Indeed, any stochastic matrix A_t leaves the vector $\mathbf{1}$ (all ones) invariant. That is, $A_t\mathbf{1} = \mathbf{1}$, and span$\{\mathbf{1}\}$ is an A_t-invariant subspace. As a reminder, the *linear span* of a set of vectors $(v_1, \ldots, v_k)$ in a vector space $\mathbb{V}$ is the subspace generated by

$$\left\{ \sum_i r_i v_i : r_i \in \mathbb{R} \right\} .$$

It can then be shown that, for any $(n-1)\times n$ matrix P having a kernel spanned by $\mathbf{1}$, the equation

$$PA_t = \tilde{A}_t P$$

has a unique solution $\tilde{A}_t$. And $\tilde{A}_t$ is $(n-1) \times (n-1)$ with

$$\{\lambda_i(A_t)\} = \{1\} \cup \{\lambda_i(\tilde{A}_t)\} .$$

This implies that, for any succession of A_t, we have

$$PA_iA_{i-1}\ldots A_2A_1 = \tilde{A}_i PA_{i-1}\ldots A_2A_1 = \cdots = \tilde{A}_i\tilde{A}_{i-1}\ldots\tilde{A}_2\tilde{A}_1 P .$$

We can now use the fact that P has a full row-rank and $P\mathbf{1} = 0$. From this, it can be deduced that $A_iA_{i-1}\ldots A_2A_1$ converging to a matrix of the type $\mathbf{1}c$ (c being a row vector) is equivalent to the convergence to zero of $\tilde{A}_i\tilde{A}_{i-1}\ldots\tilde{A}_2\tilde{A}_1$. And such a convergence to zero can be studied using the joint spectral radius of the set $\tilde{\mathcal{M}}$ that can be constructed directly from $\mathcal{M}$, using the same transformation as above, $\tilde{\mathcal{M}} = \left\{ \tilde{A}_t : PA_t = \tilde{A}_t P, A_t \in \mathcal{M} \right\}$. Note that these matrices $\tilde{A}_i$ are not stochastic and are not necessarily composed of non-negative elements.

2.2 Continuity of wavelet functions

The results presented in this section mainly come from [44], which presents the link between the continuity of wavelet functions and the joint spectral radius.

To keep things short, we can say that a wavelet is a function $\psi : \mathbb{R} \to \mathbb{R}$ such that its dilations and translations provide an orthogonal basis for $L^2(\mathbb{R})$. That is, starting with $\psi(x)$, the functions $\psi(2^j x - k)$ $(j, k \in \mathbb{N})$ can serve as a basis of decomposition for any Ł^2 function $\mathbb{R} \to \mathbb{R}$. The indices j rule the dilations and the indices k rule the translations of the initial wavelet function. The analogy with $\sin(nx)$ and $\cos(nx)$ in the Fourier basis is clear.

The wavelet basis has now become widespread, because it appears to be superior to the usual Fourier decomposition in various domains. Moreover, the computation of the discrete wavelet transform has a $O(n)$ complexity in the size n of the treated vector, while its Fourier counterpart, the discrete Fourier transform is $O(n \log n)$ (see [77]).

As an example among many, there are now image compression standards making use of the wavelet transform. The FBI uses wavelet based compression to store its digitized fingerprints. This database contains tens of millions of images and wavelets have allowed for a significant decrease in the disk space these pictures use, with an acceptable loss of quality. Compared to this, the JPEG standard, which uses the Discrete Cosine Transform, introduced unacceptable blocky artefacts, as illustrated in Figure 2.3.

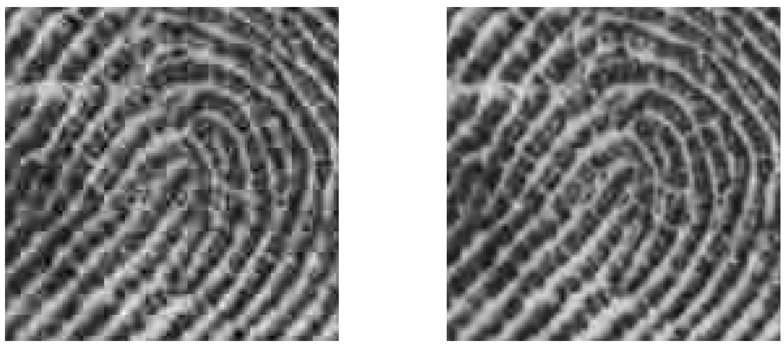

Fig. 2.3. Two different compression methods for the same image, at the same compression ratio. The left one uses the JPG standard, involving cosine transforms, and introduces blocky artifacts, prejudicial to the legal value of this print. The right one uses wavelet transforms and solves that issue.

Many other examples of wavelets applications could be presented, but the point here is not to convince the reader of the usefulness of wavelets, but to show how the joint spectral radius can be used to study them. The reader interested in a more complete historical background on wavelets is referred to [38]. And more technical details about continuity (not presented here) are to be found in [23], as well as in [71].

2.2.1 Basics of wavelets

We now define, from a mathematical point of view, how wavelets can be constructed. To do so, we use well-known results, without referring to a specific paper for every property. We refer the reader to standard references for wavelets, such as [24, 25, 27]. This being done, we will detail how the joint spectral radius comes into play in their study.

First, coefficients $(c_0, \ldots, c_N)$ have to be chosen in $\mathbb{C}$, as well as the number of these coefficients. These coefficients are used to express a *dilation equation*:

$$\varphi(x) = \sum_{k=0}^{N} c_k \varphi(2x-k) \ , \qquad \text{where the support of } \varphi(x) \text{ is } [0, N] \ . \tag{2.2}$$

Solving this dilation equation yields the *scaling function* $\varphi(x)$. Further computation is needed to express the wavelet function $\psi(x)$ as a function of $\varphi(x)$. We do not present this here, and we just use the final expression. The wavelet function $\psi(x)$ is obtained directly from $\varphi(x)$ by reversing the order of the coefficients and alternating their signs :

$$\psi(x) = \sum_{k=0}^{N} (-1)^k c_{N-k} \varphi(2x-k) \ . \tag{2.3}$$

The choice of the coefficients $(c_0, \ldots, c_N)$ strongly influences the functions $\varphi(x)$ and $\psi(x)$. Note that these are not chosen at random, as could be guessed. The usual conditions on the values of the c_k are the following :

$$\sum_k c_{2k} = \sum_k c_{2k+1} = 1 \ , \tag{2.4}$$

$$\sum_{k=0} c_k \bar{c}_{k+2j} = 2\delta_{0j} \ . \tag{2.5}$$

The second condition ensures that the family $\psi(2^j x - k)$ generated using these coefficients is orthogonal. Considered conjointly, these conditions guarantee a unique, compactly supported, integrable scaling function $\varphi(x)$. The only degree of freedom that remains about $\varphi(x)$ is that it can be multiplied by a scalar and still satisfy the dilation equation. But this does not change the shape of the function, nor its continuity.

Let us now take a look at an example, in order to shed some light on the described process.

Example 2.3. Let us pick the coefficients as $c_0 = c_1 = 1$. Note that this is about the simplest possible choice for these coefficients, given the constraints (2.4) and (2.5). We then have to solve

$$\varphi(x) = \varphi(2x) + \varphi(2x-1) \ , \tag{2.6}$$

which gives $\varphi(x) = \chi_{[0,1[}$, known as the *box function*, represented in Figure 2.4. Once $\varphi(x)$ is known, it remains to compute $\psi(x)$:

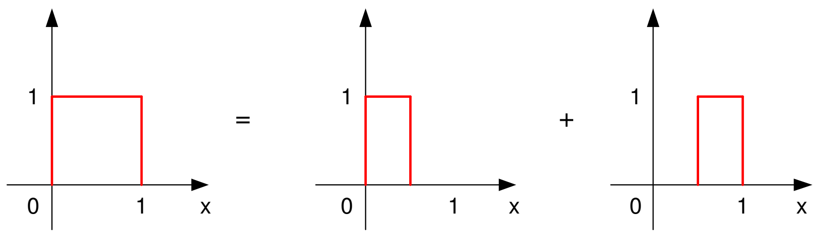

Fig. 2.4. The box function $\chi(x)_{[0,1[}$ is on the left. The whole figure shows how this function satisfies Equation (2.6).

$$\psi(x) = \varphi(2x) - \varphi(2x-1) \ ,$$

which is known as the *Haar wavelet*, depicted in Figure 2.5. This wavelet generates an orthonormal basis for $L^2(\mathbb{R})$, when dilated and translated as explained before.

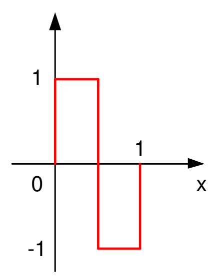

Fig. 2.5. Haar wavelet.

2.2.2 Matrix form of the formalism

We will now rewrite the above developments in a matrix fashion. As the support of $\varphi(x)$ is $[0, N]$, its information can be contained in the vector function

$$v(x) = \begin{pmatrix} \varphi(x) \\ \varphi(x+1) \\ \vdots \\ \varphi(x+N-1) \end{pmatrix}, \ x \in [0,1] \ .$$

Now, taking x to be in $[0, 1/2]$, we show how $v(x)$ can be expressed as a linear functions of $v(2x)$. Taking into account the fact that $\varphi(x)$ is zero outside its domain $[0, N]$, the dilation equation yields :

$$\begin{aligned}
\varphi(x) &= c_0\ \varphi(2x) \\
\varphi(x+1) &= c_0\ \varphi(2x+2) + c_1\ \varphi(2x+1) + c_2\ \varphi(2x) \\
&\vdots \\
\varphi(x+N-1) &= c_{N-1}\ \varphi(2x+N-1) + c_N\ \varphi(2x+N-2)\ .
\end{aligned}$$

And similarly, for $x \in [1/2, 1]$, we have

$$\begin{aligned}
\varphi(x) &= c_0\ \varphi(2x) + c_1\ \varphi(2x-1) \\
\varphi(x+1) &= c_0\ \varphi(2x+2) + c_1\ \varphi(2x+1) + c_2\ \varphi(2x) + c_3\ \varphi(2x-1) \\
&\vdots \\
\varphi(x+N-1) &= c_N\ \varphi(2x+N-2)\ .
\end{aligned}$$

In the above expressions, all the arguments of the function $\varphi(.)$ are supposed to be in $[0, N]$, otherwise the corresponding term is zero. This allows us to deduce that there are two $(N \times N)$ matrices T_0 and T_1 such that

$$v(x) = \begin{cases} T_0 v(2x) & \text{for } x \in [0, 1/2]\ , \\ T_1 v(2x-1) & \text{for } x \in [1/2, 1]\ . \end{cases} \tag{2.7}$$

One can check that $v(1/2) = T_0 v(1) = T_1 v(0)$. And to be precise, the general (i, j) element of T_0 and T_1 is respectively c_{2i-j-1} and c_{2i-j}, still with the constraints that $2i - j - 1$ and $2i - j$ have to be between 0 and N.

In the following, we will be dealing with binary expansions of real numbers that lie in $[0, 1]$. This leads to introduce τx as the fractional part of $2x$: $(2x)$ mod 1, or explicitly

$$\tau x = \begin{cases} 2x & \text{for } x \in\ [0, 1/2[\ , \\ 2x - 1 & \text{for } x \in\ ]1/2, 1]\ . \end{cases}$$

Note that $1/2$ has been intentionally left out because it admits two different binary expansions, namely $0.10000\ldots$ and $0.01111\ldots$, and so $\tau(1/2)$ is not defined. More generally, a number that admits two different binary expansions is called *dyadic*. This is the case for any number admitting a finite binary development (i.e. ending in $1000\ldots$). Indeed, $10000\ldots$ can be replaced by $01111\ldots$.

Further, we will denote the binary expansion of a real number $x \in [0, 1]$ as follows : $x = .d_1 d_2 d_3 \ldots$. Using this notation, equation (2.7) can be rewritten as

$$v(x) = T_{d_1} v(\tau x)\ .$$

Now, if we take two different dyadic values $x < y$, with y sufficiently close to x so as to have their m first digits in common, we can write

$$\begin{aligned} v(y) - v(x) &= T_{d_1}(v(\tau y) - v(\tau x)) \\ &= T_{d_1} T_{d_2}(v(\tau^2 y) - v(\tau^2 x)) \\ &\vdots \\ &= T_{d_1} \dots T_{d_m}(v(\tau^m y) - v(\tau^m x)) \ . \end{aligned}$$

This formulation will allow us to to link the continuity of the wavelet function to a joint spectral radius being less than 1.

We begin with the necessity condition. If we assume the continuity of the φ wavelet function, then v is also continuous, and taking $y \to x$, the number m of common digits between x and y tends to infinity. In the same time, if v is continuous, the difference $v(y) - v(x)$ has to tend to zero. As a consequence, a necessary condition for continuity is that the matrix products of the type $T_{d_1} \dots T_{d_m}$ yield zero when applied to a difference of v functions of dyadic values. That is to say that we can restrict our study of these operators to the subspace

$$W = \operatorname{span}\{v(x) - v(y) : x, y \in [0,1] \text{ are dyadic}\} \ .$$

Thus, all the products of the matrices $T_0|_W$ and $T_1|_W$ (T_0 and T_1 restricted to W) must tend to zero as their length increases. This means that the joint spectral radius of this pair of matrices has to be less than 1, as explained in section 1.2, when we defined the joint and generalized spectral radii:

$$\rho\left(T_0|_W, T_1|_W\right) < 1 \ .$$

We just established a necessary condition for the continuity of $\varphi(x)$.

Conversely, it can be shown that this criterion $\rho(T_0|_W, T_1|_W) < 1$ is also a sufficient condition for the continuity of $\varphi(x)$. We refer the reader to [22, 43] for details about this.

Note that under a well-chosen basis transformation, both matrices $T_0|_W$ and $T_1|_W$ can be reduced to square matrices M_0 and M_1 acting on $\mathbb{R}^{n'}$, where $n' = \dim(W)$. This gives the following necessary and sufficient condition for the continuity of $\varphi(x)$ ([23]):

$$\rho\left(T_0|_W, T_1|_W\right) = \rho(M_0, M_1) < 1 \ .$$

From a practical point of view, the subspace W can be determined by using the fact that it is the smallest subspace containing $v(1) - v(0)$ and invariant under T_0 and T_1. Moreover, if the condition (2.4) is satisfied, it implies that all columns of T_0 and T_1 sum to 1. And so, the subspace $V = \{y \in \mathbb{C}^N : y_1 + \cdots + y_N = 1\}$ is invariant under both T_0 and T_1. From this, it can be deduced that $W \subseteq V$ and $\rho(T_0|_W, T_1|_W) \leq \rho(T_0|_V, T_1|_V)$.

Example 2.4. This example presents what is known as the *Daubechies wavelet*. Let us pick as coefficients the values

$$(c_0, c_1, c_2, c_3) = (3/5, 6/5, 2/5, -1/5) \ .$$

These do satisfy both conditions expressed in (2.4) and (2.5). The solution of the corresponding dilation equation is represented in Figure 2.6, as well as the associated wavelet function. The associated matrices are

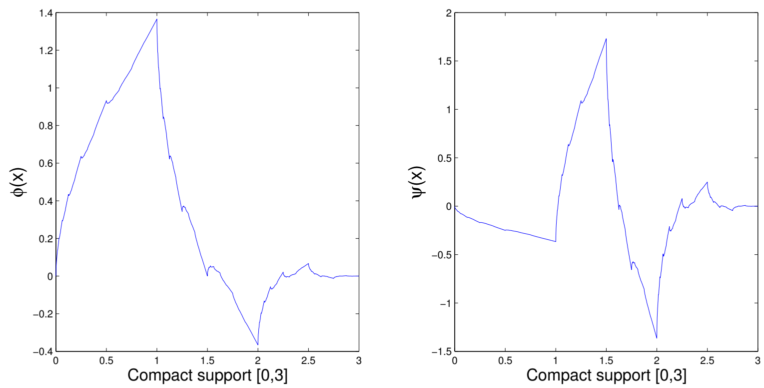

Fig. 2.6. Daubechies wavelet with coefficients $(3/5, 6/5, 2/5, -1/5)$. The left-hand plot represents $\varphi(x)$, solution of the dilation equation using these coefficients, while the right-hand shows the corresponding wavelet function $\psi(x)$.

$$T_0 = \begin{pmatrix} c_0 & 0 & 0 \\ c_2 & c_1 & c_0 \\ 0 & c_3 & c_2 \end{pmatrix} = \begin{pmatrix} 3/5 & 0 & 0 \\ 2/5 & 6/5 & 3/5 \\ 0 & -1/5 & 2/5 \end{pmatrix} \ ,$$
$$T_1 = \begin{pmatrix} c_1 & c_0 & 0 \\ c_3 & c_2 & c_1 \\ 0 & 0 & c_3 \end{pmatrix} = \begin{pmatrix} 6/5 & 3/5 & 0 \\ -1/5 & 2/5 & 6/5 \\ 0 & 0 & -1/5 \end{pmatrix} \ .$$

The restriction of these matrices to the subspace V is straightforward. This subspace is the plane with equation $x + y + z = 1$. In this plane, we use as (arbitrary) basis vectors $v_1 = (1, 0, -1)$ and $v_2 = (0, 1, -1)$. We then have

$$\begin{aligned}
T_0 v_1 &= (3/5, -1/5, -2/5) = 3/5 v_1 - 1/5 v_2 \\
T_0 v_2 &= (0, 3/5, -3/5) = 3/5 v_2 \\
T_1 v_1 &= (6/5, -7/5, 1/5) = 6/5 v_1 - 7/5 v_2 \\
T_1 v_2 &= (3/5, -4/5, 1/5) = 3/5 v_1 - 4/5 v_2
\end{aligned}$$

This yields the following pair of matrices, operating inside the plane:

$$M_0 = \begin{pmatrix} 3/5 & 0 \\ -1/5 & 3/5 \end{pmatrix}, \; M_1 = \begin{pmatrix} 6/5 & 3/5 \\ -7/5 & -4/5 \end{pmatrix} .$$

It remains to compute the joint spectral radius of this pair of matrices. This will be detailed in chapter 6. For now, we will merely say that $\rho(M_0, M_1)$ is found to be ≈ 0.66, for any of the two above matrix pairs. The continuity result does not depend on the chosen basis of the subspace W. In this example, this allows us to deduce that φ and ψ are continuous for the considered parameters.

To conclude this section, it can be noted that the joint spectral radius is useful to determine the value of the *Hölder exponent* of the wavelet function. As a reminder, the Hölder exponent somehow expresses the degree of continuity of a function. More precisely, a complex-valued map φ on a set $U \subset \mathbb{C}$ is called uniformly Hölder continuous with Hölder exponent $0 < \alpha \leq 1$, if there exists a constant $C_\alpha \in \mathbb{R}$ such that

$$|\varphi(x) - \varphi(y)| \leq C_\alpha |x - y|^\alpha, \forall x, y \in U .$$

A value $\alpha > 1$ would make the function φ constant, because we could then write $\frac{|\varphi(x)-\varphi(y)|}{|x-y|} \leq C_\alpha |x-y|^{\alpha-1}, \forall x, y \in U$. And so, letting $y \to x$, we see that the derivative $\varphi'(x) = 0, \forall x$.

The interesting result here is that the joint spectral radius $\rho(M_0, M_1)$ yields a threshold value for the Hölder exponent of $\varphi(x)$. Indeed, $\varphi(x)$ is Hölder continuous for any $0 \leq \alpha < -\log_2 \rho(M_0, M_1)$, while not Hölder continuous for $\alpha > -\log_2 \rho(M_0, M_1)$. In the example at hand, we then have Hölder continuity for approximately

$$\alpha < -\log_2 0.66 \cdots \approx 0.6 .$$

Further details about this relation and its proof, including details about the case $\alpha = -\log_2 \rho(M_0, M_1)$, can be found in [22, 43]. The Hölder continuity is also treated in [57].

2.3 Differences in codes

This section presents notions and notations taken from [62], where more details are given, as well as numerous useful examples. Some of these examples are reproduced here, in order to provide a better comprehension. Although we have skipped technical details from the original paper by Moison et al., we felt it was appropriate to explain their approach with some details. Thanks to this, this section is rather self-contained.

2.3.1 Notion of code and capacity

The joint spectral radius also arises in a totally different area: the storage of digital information on magnetic-recording devices.

One of the main concerns about storing data is to keep the probability of read/write errors low. When storing bits, these are organized in *sequences* of a given length n. The set of all possible sequences used to do so is called a n-bit *code*, which we denote by $\mathcal{C}$. To give an example, bits can be organised as sequences of 8 bits, called bytes. The code here would be the set of all 2^8 possible bytes.

The error rate of a given code can be shown to be related to the smallest distance between its allowed sequences. We will not enter into details here, such as what measure to use to define the distance between bits sequences. Although, it is clear to realize that the sequence $w_1 = 00001011$ is easier to distinguish from $w_2 = 11101100$ than from $w_3 = 00001010$.

To do so, the *difference* between such sequences can be defined. For two sequences $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ where $u_i, v_i \in \{0, 1\}$, the difference is $u - v = (u_1 - v_1, \ldots, u_n - v_n)$ where $(u-v)_i = u_i - v_i \in \{-, 0, +\}$. Using this definition, we see that

$$
\begin{aligned}
w_1 - w_2 &= (-, -, -, 0, 0, -, +, +) \ , \\
w_1 - w_3 &= (0, 0, 0, 0, 0, 0, 0, +).
\end{aligned}
$$

Using the euclidian vector norm, we have

$$\|w_1 - w_3\| < \|w_1 - w_2\| \ .$$

This confirms our first impression that w_1 was somehow closer to w_3 than to w_2. This phenomenon motivates the desire to avoid certain pairs of sequences in the same code, although the criterion used to determine these pairs is not as simple as the example above. So, codes have to be defined that avoid specified differences between their sequences.

But what has to be kept in mind is the role of a code : to transmit information. In this perspective, a key notion is the *capacity* of a code. Let us first give a formal definition of it.

Let D be the set of the ternary sequences (made of 0,-,+) representing the differences that we forbid to exist between the patterns in the code. In short, D is called the set of *forbidden difference patterns*. We say that the n-bit code $\mathcal{C}$ *avoids* D if

$$\forall u, v \in \mathcal{C}, \forall i \leq j \in [1, n], \ u_{[i,j]} - v_{[i,j]} \notin D \ ,$$

where $[i, j]$ represents $\{i, \ldots, j\}$ and $u_{[i,j]} = u_i, \ldots, u_j$.

From this, we can define the largest number of n-bit sequences avoiding the forbidden differences defined by D as

$$\delta_n(D) := \max\{|\mathcal{C}| : \mathcal{C} \text{ avoids } D\} \ .$$

The *capacity* of D is then defined as

$$\operatorname{cap}(D) := \log_2 \left[\lim_{n\to\infty} (\delta_n(D))^{1/n} \right]$$

So, the capacity of the channel described by D is defined as the exponential growth rate of the maximal number of sequences whose differences avoid a given D, as the length of the considered sequences increases. It can be seen as the richness of the code, that is, its ability to transmit more or less information for a given number of transmitted symbols. Of course, the aim of studying such a notion is to determine how to achieve the best possible transmission rate, taking into consideration the disallowed difference patterns imposed by D.

2.3.2 Examples

Let us now illustrate this with examples.

Trivial case $D = \emptyset$

In the simple case where no difference has to be avoided, that is $D = \emptyset$, the number $\delta_n(\emptyset)$ of possible sequences of bits of length n is naturally 2^n. And so, the capacity of such an unrestricted binary code is $\log_2(2) = 1$.

Case $D = \{++, +-\}$

Here, the differences $++$ and $+-$ are forbidden. It can be checked that the following codes avoid D :

$$\mathcal{C}_2 = \begin{bmatrix} 00 \\ 10 \end{bmatrix}, \; \mathcal{C}_3 = \begin{bmatrix} 000 \\ 001 \\ 100 \\ 101 \end{bmatrix}, \; \mathcal{C}_4 = \begin{bmatrix} 0000 \\ 0010 \\ 1000 \\ 1010 \end{bmatrix} .$$

Moreover, there exist no larger codes for these lengths that avoid D. One can also see that if a code $\mathcal{C}$ avoids D, then at least one of every two adjacent columns in $\mathcal{C}$ must be constant. If we add that codes of the type $\mathcal{C}_n = \{u_1 0 u_3 0 \ldots 0 u_n\}$ for n odd, or $\mathcal{C}_n = \{u_1 0 u_3 0 \ldots 0 u_n\}$ for n even, effectively avoid D, we can conclude that

$$\delta_n(D) = 2^{\lceil n/2 \rceil} ,$$

and so $\operatorname{cap}(D) = 0.5$.

2.3.3 Link with the joint spectral radius

Without entering into the details of the proof provided in [62], we mention here the main results making the link between the capacity of a code and the joint spectral radius of a set of matrices.

A *joint pattern*, which denotes a set of two patterns $\{p, p'\}$, both of the same length which we denote by m, is disallowed for a difference set D if

$$p - p' \in D \text{ or } p' - p \in D \ .$$

Now, given D, the notation $\mathcal{J}(D)$ will stand for the set of all disallowed joint patterns.
As an example : if $D = \{++, +-\}$, then $\mathcal{J}(D) = \{\{11, 00\}, \{10, 01\}\}$.

We will call a set $M \subsetneqq \{0,1\}^m$ a *representing set for* $\mathcal{J}(D)$ if it intersects every set in $\mathcal{J}(D)$. Such a representing set will be called *minimal* if none of its strict subsets is itself a representing set for $\mathcal{J}(D)$. The collection of all minimal representing sets of $\mathcal{J}(D)$ is denoted by $\mathcal{M}(D)$.
Let us illustrate this definition using the same example as above, with $D = \{++, +-\}$ and $\mathcal{J}(D) = \{\{11, 00\}, \{10, 01\}\}$. Then, from the definition, all the sets that contain at least one element from $\{11, 00\}$ and one element from $\{10, 01\}$ are representing:

$$\{11, 10\}, \{11, 00, 10\}, \{00, 10, 01\}, \{11, 00, 10, 01\}, \dots \ .$$

These are minimal if none of their strict subsets is itself representing. For example, $\{11, 00, 10\}$ is not minimal, since it contains $\{11, 10\}$, which is also representing. Clearly, this leads us to the following minimal representing sets, gathered in $\mathcal{M}(\{++, +-\})$:

$$\mathcal{M}(\{++, +-\}) = \{\{11, 10\}, \{11, 01\}, \{00, 10\}, \{00, 01\}\} \ .$$

The following part makes use of bipartite graphs. A *bipartite graph* (L, R, E) is composed of :

- L : a set of left vertices,
- R : a set of right vertices,
- E : a set of edges (l, r), each connecting a left vertex l to a right vertex r .

We will denote by G_m bipartite graphs of the following type :

- $L = R = \{0,1\}^{m-1}$, so that the vertices are labeled with $(m-1)$-bit sequences,
- E such that $(l_1, \dots, l_{m-1})$ is connected to $(r_1, \dots, r_{m-1})$ if the bit sequence $l_2, \dots, l_{m-1}$ is equal to $r_1, \dots, r_{m-2}$. Such edges are identified with the m-bit sequence that is the merging of both vertices :

$$(l_1, \dots, l_{m-1}, r_m) = (l_1, r_1, \dots, r_{m-1}) \ .$$

For example, the graph G_3 is presented in Figure 2.7.

Now, we are able to define the graph G_M corresponding to a set $M \subseteq \{0,1\}^m$. This is just done by removing from G_m the edges corresponding to the sequences of M. Figure 2.8 represents the graph $G_{\{100,111\}}$ as an example.

Bipartite graphs can be put in cascade. To do so, the right vertices of one graph correspond to the left vertices of the next one, and so on. For clarity, Figure 2.9 shows such a cascaded graph. By merging the labels of the edges used to go from one the leftmost vertices to one of the rightmost vertices, we

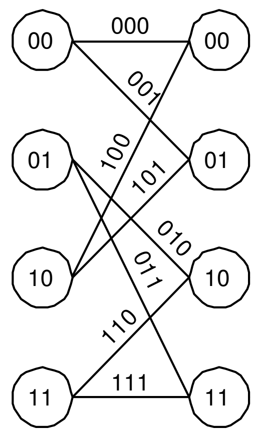

Fig. 2.7. The bipartite graph G_3. Vertices are all 2-bit sequences. One can see that the label on an edge linking two vertices does begin and end with the labels of these vertices.

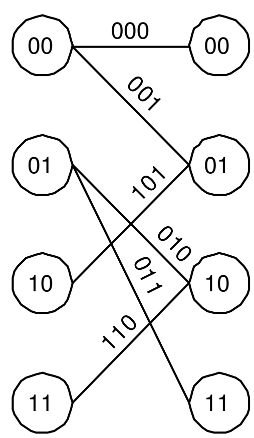

Fig. 2.8. The bipartite graph $G_{\{100,111\}}$.

read a bit sequence. The property of interest here is the following : the defined bit sequence is composed of the labels of the edges. This is illustrated in figure 2.10.

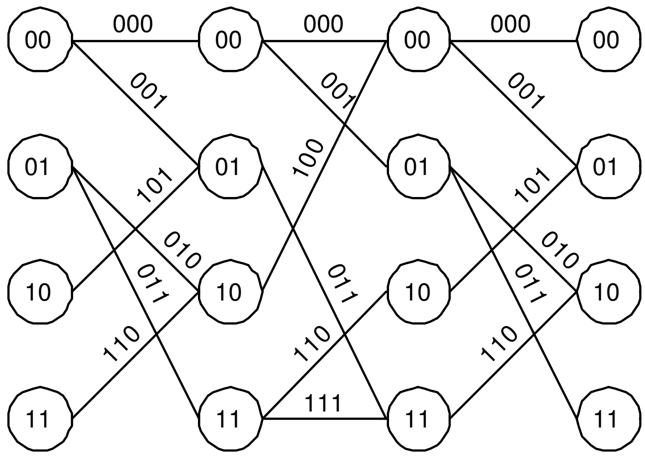

Fig. 2.9. Cascaded bipartite graphs. This presents the cascade consisting of $\{G_{\{100,111\}}, G_{\{101,010\}}, G_{\{100,111\}}\}$.

As a consequence, by using graphs of the type G_M for the appropriate $M \in \mathcal{M}(D)$, it is possible to avoid the use of certain bit patterns. Namely, in this way we generate only sequences avoiding D. So, counting a number of bit sequences avoiding D can be brought down to counting the number of paths from the leftmost part to the rightmost part of a cascaded bipartite graph.

Counting such a number of paths in a graph is done using *adjacency matrices*: a graph can be associated to an adjacency matrix A, in which $A_{ij} = 1$ if the vertex i is connected to the vertex j, and $A_{ij} = 0$ otherwise. We denote by A_G the adjacency matrix of a bipartite graph G. When dealing with cascaded graphs, we just need to take the product of the adjacency matrices corresponding to the consecutive bipartite graphs. The number of paths from the leftmost vertex i to the rightmost vertex j is equal to the (i, j) element of this product. And the total number of paths from the left end to the right end is given by the sum of all elements of this product. For a matrix A, such a procedure yields a norm, which we will here call

$$\|A\|_f = \sum_{i,j=1}^{n} |A_{ij}| \; .$$

Note that this is not an induced norm, and it consequently illustrates the need for the joint spectral radius to be defined for all matrix norms, and not only for induced norms. We are, however, on the safe side, since all matrix norms (and not only induced norms) are equivalent, as said in Section 1.2.

So, for a given set D of forbidden differences, it is of interest to define the set $\Sigma(D)$ in the following way :

$$\Sigma(D) := \{A_{G_M} : M \in \mathcal{M}(D)\} \ .$$

Then finding the arrangement of bipartite graphs yielding the maximum number of paths is equivalent to finding the best possible arrangement of matrices A_{G_M} so that their product yields the most paths, thus achieving the largest $\|.\|_f$ norm. As this is the definition of $\hat{\rho}_k$, this computation is equivalent to the determination of the joint spectral radius of this set of matrices, $\rho(\Sigma(D))$. This implies that the capacity of a code avoiding D is given by

$$\operatorname{cap}(D) = \log \rho(\Sigma(D)) \ .$$

2.3.4 Illustration

$D = \{+-\}$

Let us take the following set of forbidden differences $D = \{+-\}$. Then the set of forbidden pairs is

$$\mathcal{J}(D) = \{\{10, 01\}\} \ .$$

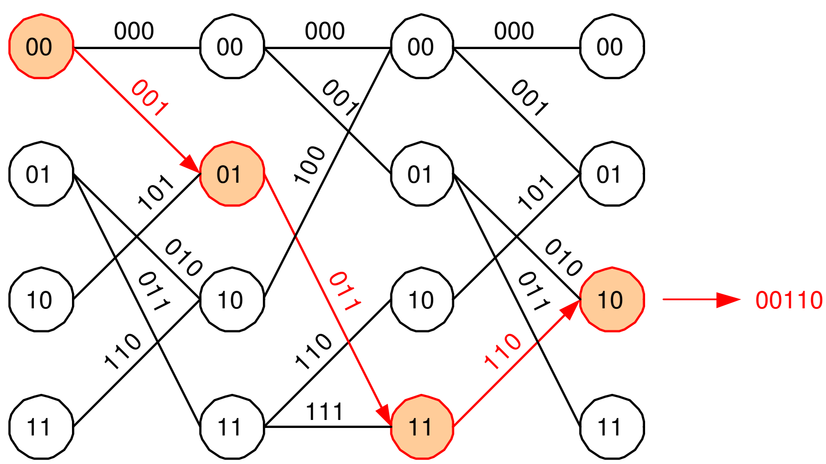

Fig. 2.10. Cascaded bipartite graphs. This presents the cascade consisting of $\{G_{\{011,111\}}, G_{\{101,010\}}, G_{\{011,111\}}\}$. The highlighted edges and vertices show a path from the left end to the right end of the cascaded graph. We start with the two bits of the first vertex, and then add a bit for each edge we use. This defines the bit sequence 00110. Clearly, every edge label 001, 011, 110 is a subword of the defined sequence.

The collection of all minimal representing sets of $\mathcal{J}(D)$ is then

$$\mathcal{M}(D) = \{\{10\}, \{01\}\} \ .$$

We then have

$$\Sigma(D) = \{G_{\{10\}}, G_{\{01\}}\} \ .$$

The graph $G_{\{10\}}$ is represented in Figure 2.11. Its adjacency matrix turns out

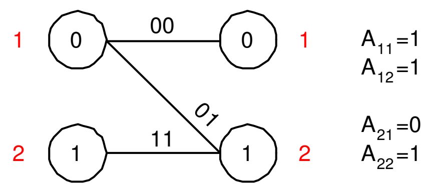

Fig. 2.11. Bipartite graph $G_{\{10\}}$. The vertices have been numbered in order to write the adjacency matrix. $A_{ij} = 1$ if there is a vertex from i on the left to j on the right.

to be $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Proceeding in this way with the other graph, this yields a set $\Sigma(D)$ of two (2×2) matrices,

$$\Sigma(D) = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\},$$

the joint spectral radius of which can be proved to be $\frac{1+\sqrt{5}}{2}$ (this will be done in this work, see namely Proposition 6.20). And so, the capacity of a code avoiding the difference $\{+-\}$ is $\text{cap}(D) = \log_2(\frac{1+\sqrt{5}}{2}) \approx 0.6942$.

$D = \{+-0\}$

Let us take another set of forbidden differences: $D = \{+-0\}$. Then the set of forbidden pairs is

$$\mathcal{J}(D) = \{\{100, 010\}, \{101, 011\}\} \ .$$

The collection of all minimal representing sets of $\mathcal{J}(D)$ is then

$$\mathcal{M}(D) = \{\{100, 101\}, \{100, 011\}, \{010, 101\}, \{010, 011\}\} \ ,$$

and we have

$$\Sigma(D) = \{G_{\{100,101\}}, G_{\{100,011\}}, G_{\{010,101\}}, G_{\{010,011\}}\} \ .$$

The graph $G_{\{100,101\}}$ is represented in Figure 2.12. Its adjacency matrix turns out to be $\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$.

Proceding in this way with the other graphs, this yields a set $\Sigma(D)$ of four (4×4) matrices,

$$\Sigma(D) = \left\{ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \right\},$$

the joint spectral radius of which can also be approximated. It is found to be 1.618, which also is the spectral radius of the third matrix.

2.4 Conclusion

In this chapter, we have considered applications that serve as motivations for the study of the joint spectral radius. One could argue that the only concern about the JSR is actually to know whether it is less than 1 or not. Indeed, if the only question is to determine the continuity of a wavelet function or the convergence to zero of a switched linear system, this is true.

However, the precise value of the JSR can be a valuable measure of performance. For example, for wavelet functions, the value of the JSR provides us with the Hölder exponent of the function, which is more informative than just

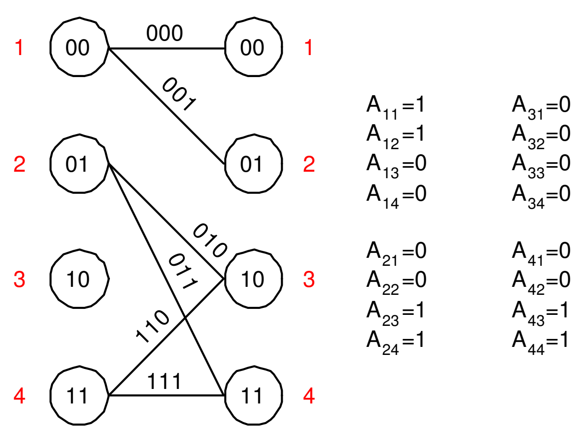

Fig. 2.12. Bipartite graph $G_{\{100,101\}}$. The vertices have been numbered before writing the adjacency matrix. $A_{ij} = 1$ if there is a vertex from i on the left to j on the right.

its (dis)continuity. And for switched linear systems, the JSR tells us how fast a system will converge or diverge. Similarly, the application of the JSR to the capacity of codes clearly shows that we need to compute the JSR as precisely as possible in order to compare different codes.

At this point, it is worth noting that two different problems, which are a stability issue and a continuity issue, can both be expressed as the computation of a joint spectral radius. This shows that the computation of a JSR is behind several different problems. Of course, we do not claim to have fully covered the scope of the applications of the joint spectral radius. We refer the reader to Gilbert Strang's survey paper [78] for a more extensive list of applications.

The examples presented in this chapter show that there is an interest for computation and approximation algorithms for the JSR. This problem is a hard one, as the results presented in Chapter 4 show. We will be focusing on the effective computation of the JSR in Chapter 6.

3

Dynamical systems approach of the joint spectral radius

The previous chapters were devoted to introducing the definition of the joint spectral radius, as well as presenting some of its potential applications. We now turn to the main aspect of our study : the dynamical systems.

In this chapter, we will show how the joint spectral radius rules the stability of discrete time switched linear systems, also known as discrete linear inclusions (DLI). In this context, we will also study the relations between the existence of a common quadratic Lyapunov function for a DLI and its stability. And after this, we will discuss the possible correspondence between the stability of discrete time and continuous time switched linear systems.

3.1 Switched linear systems and discrete linear inclusions

We first introduce switched linear systems in continuous time, before moving to discrete time, which is the main subject of this thesis. Let us consider a set $\mathcal{M}$ of matrices A_i and pick an initial point at $t = 0$, $x(0)$. A switched linear system is a dynamical system of the type

$$\dot{x}(t) \in \left\{ A_{i(t)} x(t) : A_{i(t)} \in \mathcal{M} \right\} \ .$$

This notation means that, at every instant, the matrix A_i defining the evolution of the system can be replaced by another one from the set $\mathcal{M}$. The function $i(t)$ is the *switching signal*, which is a function of t that can also depend explicitly on $x(t)$. Note that it is not important, at this stage, to know which element of $\mathcal{M}$ is chosen at each step. A function $x(t)$ satisfying the above equation for some $i(t)$ is a *trajectory* of the system in the $\mathbb{R}^n$ space. At this stage, we define the notion of *Zeno behavior* : a system has a Zeno behavior if it admits an infinite number of switchings in a finite interval of time. We will not consider such behaviors in this thesis, as we will assume that a number of switchings tending to $+\infty$ corresponds to $t \to \infty$. We will not discuss conditions on the function $i(t)$ that would be necessary for the corresponding trajectory $x(t)$ to be differentiable. This is motivated by the fact that this thesis focuses on discrete time systems, for which such issues are irrelevant.

The discrete time counterpart is very similar. Let us again consider a set $\mathcal{M}$ of matrices A_i, which can also be seen as a set of $\mathbb{R}^n \to \mathbb{R}^n$ operators. Note that these matrices are not, for the time being, linked in any way to those that were used in the continuous time case. We will discuss later extensively how to relate the discrete time matrices to their continuous time counterparts (see Section 3.5).

We take an arbitrary point $x_0 \in \mathbb{R}$ as initial point. We then define the following recurrence:

$$x_{t+1} \in \{A_{i_t} x_t : A_{i_t} \in \mathcal{M}\} \ . \tag{3.1}$$

Again, this notation means that, at each time step t, any matrix A_i can be used in order to determine the following value x_{t+1}. The sequence (i_t) is the *switching signal*. Its values depend on t but could also depend on x_t. As before, it is not important at this stage to know which element of $\mathcal{M}$ is chosen at each step. A series $(x_0, x_1, \ldots, x_k, \ldots)$ satisfying the inclusion (3.1),

$$x_{t+1} = A_{i_t} x_t, \quad \text{for some } (i_t) \text{ sequence},$$

is a *trajectory* in the $\mathbb{R}^n$ space. Note that the set of all possible switching signals defines a whole set of possible trajectories.

We will call *discrete linear inclusion* the set $(x_k) : k \geq 0$ of all possible sequences of vectors in $\mathbb{R}^n$ generated using the relation (3.1). This set will be denoted by DLI($\mathcal{M}$).

The interpretation of a DLI is quite simple. We are facing a dynamical system in discrete time where, at each time step, we have to choose a linear evolution law from a given set of operators. Some of the first questions that arise in such a case is: "Do these different trajectories converge ? if so, to what point ? and under what circumstances ?". In order to treat these questions, we first need to formally introduce the notion of stability.

In some cases, the study of the stability of such a system is defined as the convergence to the origin for a given switching sequence i_t. The aim could also be to find a class of functions i_t that stabilizes the systems associated to a given set $\mathcal{M}$ of matrices. In our case, stability will essentially mean that all the trajectories converge to the origin, for any switching signal. The following section details this aspect.

3.1.1 Different notions of stability

This section first presents four different definitions of stability, as presented by Gurvits in [42]. Although only two of these will actually be used afterwards, we feel it appropriate to present all of them. We will then present the stability defined by Kozyakin in [51].

AAS : Absolute Asymptotic Stability

Absolute asymptotic stability is the first notion of stability and can be considered as the simplest one, because it is the most intuitive. As could have been guessed by its designation, this is the strongest form of stability. It states that any trajectory (x_k) taken from DLI($\mathcal{M}$) converges to the origin:

$$\lim_{k\to\infty} x_k = 0,\ \forall (i_t)\ .$$

As this is supposed to hold for any initial vector x_0, it is equivalent to saying that all matrix products converge to the zero matrix :

$$\lim_{k\to\infty} A_{i_k} A_{i_{k-1}} \dots A_{i_1} = 0,\ \forall (i_t)\ .$$

PAS : Periodic Asymptotic Stability

This is a restriction of the previous notion to periodic sequences. It states that all periodic matrix products converge to the zero matrix. Formally, a discrete linear inclusion is PAS if any finite matrix product of the type

$$A_\omega = A_{i_m} A_{i_{m-1}} \dots A_{i_1}$$

satisfies

$$\lim_{k\to\infty} (A_\omega)^k = 0$$

for any choice of a pattern ω of any finite period length m. Equivalently, this also means that $\rho(A_\omega) < 1$.

We will see later on that it is nontrivial to know whether PAS implies AAS or not, while AAS clearly implies PAS. This implication PAS $\Rightarrow$ AAS was conjectured in what has been known as the *finiteness conjecture*. It is now proved that this conjecture does not hold in general, as will be presented in full details in Section 4.1.

MAS : Markov Asymptotic Stability

To understand this stability notion, we first need to recall the notion of a Markov chain. Consider a sequence of random variables $\{X_k\}$, where k is the index, taking values in $\mathbb{N}$. It is a Markov chain if the probability law conditioning the value X_k (X at time k) only depends on X_{k-1}. In other words, the transition probability from a state to the next one is memoryless. In Figure 3.1, we present an example of a system generating such a sequence.

We now take any Markov chain on the finite state space $\{1, 2, \ldots, N\}$, where all transition probabilities are positive. Let the sequence of successive states of such a chain be denoted by $\{s(k) : k \geq 1\}$. Now consider a DLI generated by a set $\mathcal{M} = \{A_1, \ldots, A_N\}$. It is MAS if, for any initial $x_0 \in \mathbb{R}^n$,

$$\lim_{k\to\infty} A_{s(k)} A_{s(k-1)} \ldots A_{s(1)}\, x_0 = 0$$

holds with probability 1.

To be complete, we may mention at this point the *top Lyapunov exponent* $\bar{\rho}(\mathcal{M})$. It is defined in [79] as follows:

$$\bar{\rho}(\mathcal{M}) := \lim_{k\to\infty} \frac{1}{k} E\left[\log(\|A_1 \ldots A_k\|)\right] \ .$$

If all matrices occur with the same probability $1/N$, then this definition reduces to the logarithm of the geometric mean of all possible products of k matrices.

$$\bar{\rho}(\mathcal{M}) := \lim_{k\to\infty} \log \left(\prod_{d_1,\ldots,d_k=1}^{N} \|A_{d_1} \ldots A_{d_k}\|^{\frac{1}{N^k k}} \right) \ ,$$

where $\prod_{d_1 \ldots d_k}$ stands for all the possible combinations of indices. As this expression contains the geometric mean of the norms of all products of a given length k, it is clearly inferior to the log of the joint spectral radius: $\bar{\rho}(\mathcal{M}) \leq \log \rho(\mathcal{M})$. Similar considerations show us that the Lyapunov exponent is superior to the lower spectral radius : $\bar{\rho}(\mathcal{M}) \geq \rho_*(\mathcal{M})$.

Although Markov asymptotic stability would be worth studying, it will not be detailed in this thesis. Our work is essentially focused on guaranteing the

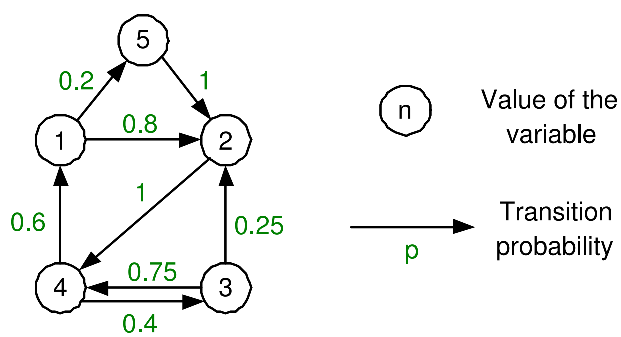

Fig. 3.1. Example of a system generating a sequence of values that satisfy the definition of a Markov chain. The random variables can take the values 1 to 5. When it is 1 at step k, it may take values 2 or 5 at step $k+1$, with respective probabilities 0.8 and 0.2, and so on for the other values. The transition probabilities only depend on the present value of the variable, and not on its past values.

convergence of the trajectories to the origin, without any probabilistic considerations.

SAS : Shuffle Asymptotic Stability

A sequence of matrices is called *generic* if each matrix A_i of the set $\mathcal{M}$ occurs an infinite number of times in the sequence. We extend this definition to a trajectory : a sequence $(x_k : k \geq 0)$ is called generic if the corresponding sequence $(A_{i_k} : k \geq 0)$ is generic. The SAS demands that every generic trajectory converges to 0, that is:

$$\lim_{k \to \infty} x_k = 0, \text{ for any initial point } x_0 \text{ and a generic } (x_k) \ .$$

Although we will not be dealing with generic sequences or trajectories in the remainder of this thesis, this definition calls for an immediate comment. One should remark that any sequence composed by picking matrices in a finite set $\mathcal{M}$ can always be considered as *asymptotically generic* for a subset $\mathcal{M}' \subseteq \mathcal{M}$.

Let us clarify this. Take any infinite sequence $A_{\bar{\omega}}$ built with $\mathcal{M} = \{A_1, \ldots, A_N\}$, where $\bar{\omega}$ denotes an infinite sequence of indices. Then, if it is not generic relatively to $\mathcal{M}$, we denote as $\mathcal{K}$ the set of indices of the matrices A_i that are repeated only a finite number of times in $A_{\bar{\omega}}$. And then, the sequence can be split into two parts:

$$A_{\bar{\omega}} = A_{\bar{\omega}'} A_{\tilde{\omega}} \ ,$$

where $A_{\bar{\omega}'}$ is still an infinite sequence of indices and $A_{\tilde{\omega}}$ is a finite suffix containing all occurrences of the matrices $A_i, i \in \mathcal{K}$. From this construction, it is clear that $A_{\bar{\omega}'}$ is generic relatively to $\mathcal{M}' = \mathcal{M} \backslash \mathcal{K}$. We can add to this observation that a finite suffix in an infinite matrix product does not alter the asymptotic behavior of the trajectories (unless of course a zero product is formed, which we choose not to consider here). This leads us to conclude that

$$\text{DLI}(\mathcal{M}) \text{ is AAS} \iff \forall \mathcal{M}' \subseteq \mathcal{M}, \text{DLI}(\mathcal{M}') \text{ is SAS} \ .$$

Exponential Stability ([51])

The following definitions use notations that are consistent with the rest of this thesis. Minor adaptations have been made from the original ones in [51].

Definition 3.1. *A system*

$$x_{t+1} \in \{A_{i_t} x_t : i_t \in \{1, \ldots, m\}\}$$

is said to be absolutely stable *with respect to a matrix class* $\mathcal{U} = \{A_1, \ldots, A_m\}$ *if, for a* $c = c(\mathcal{U})$*, the following inequality holds for any sequence* (i_t)*:*

$$\forall n,\ \|A_{i_n}A_{i_{n-1}}\dots A_{i_1}\ x\| \leq c\|x\|\ .$$

The system is said to be absolutely exponentially stable *with respect to a matrix class $\mathcal{U}$ if for a $c = c(\mathcal{U})$ and a $q = q(\mathcal{U}) < 1$, the following inequality holds for any sequence (i_t):*

$$\forall n,\ \|A_{i_n}A_{i_{n-1}}\dots A_{i_1}\ x\| \leq cq^n\|x\|\ .$$

We can comment a bit on this definition. The first stability describes trajectories that only have to remain bounded in time, while the second demands a convergence to zero of all trajectories (because $q < 1$). Consequently, we can consider this absolute exponential stability as equivalent to the absolute asymptotic stability defined earlier. Indeed, in case of AAS, the joint spectral radius can be identified to play the same role as this value q. More details about this last claim are to be found in Section 4.3.3.

3.1.2 Link between PAS and AAS

From the four definitions AAS-PAS-MAS-SAS, it is clear that AAS implies all three other kinds of stability, including PAS. Reciprocally, it is not trivial to determine whether PAS implies AAS or not. It has been conjectured in 1995, in what we will keep calling the *Finiteness Conjecture* by Lagarias & Wang in [52] in 1995 (see also [42]). This conjecture will be detailed and studied in Section 4.1. In short, it states that the joint spectral radius of a set of matrices can always be achieved by a periodic product of matrices. This conjecture has been proved to be false by Bousch & Mairesse in [19] in 2002. We have published a more elementary and self-contained proof a bit later in [11]. It is presented in full details in section 4.1. To conclude this discussion for now, we can state that we now know that PAS does not imply AAS in general. Some comments about this property will be made in Section 3.2.2, after Lemma 3.6, which gives the matrix expressions corresponding to these two different stabilities. Regarding the other types of stability (MAS and SAS), they were cited as a reference, but we will not study them. We will now concentrate on AAS and PAS and their link to the joint spectral radius. Later on, we will use the exponential stability as well.

3.2 Spectral radius and stability

There is a strong presumption that there should be a link between the joint spectral radius of a set of matrices and the stability of the associated dynamical system, which appears to be a discrete linear inclusion, as presented in section 3.1. This link is more or less apparent in some of the various publications cited so far. Without being exhaustive, we can mention works by Daubechies, Lagarias, Wang, Gurvits in [26, 42, 52]. However, this relation still had to be put into its simplest form. We here provide a proof of this link as a stand-alone result.

In the following subsection, we still distinguish the JSR from the GSR, as they do not play the same role in determining the stability of a DLI.

3.2.1 Joint spectral radius and AAS

The first property concerns the joint spectral radius and the AAS. We remind that the joint spectral radius has been defined using a norm, but its value does not depend on the norm we use, as presented in definition (1.13) :

$$\hat{\rho}(\mathcal{M}) := \lim_{k\to\infty} \left(\hat{\rho}_k(\mathcal{M}, \|.\|)\right)^{1/k},$$

$$\text{where} \quad \hat{\rho}_k(\mathcal{M}, \|.\|) = \sup_{A_1,\dots,A_k\in\mathcal{M}} \|A_k \cdots A_1\|.$$

Note that, like we said after Lemma 1.9, we use lim instead of lim sup in the definition of the JSR. And, for the ease of notation, we will take the liberty of dropping the reference to the norm in $\hat{\rho}_k(\mathcal{M})$. The link between the JSR and absolute asymptotic stability is the following :

Proposition 3.2. *For any bounded set $\mathcal{M}$ of matrices,*

$$\hat{\rho}(\mathcal{M}) < 1 \iff DLI(\mathcal{M}) \text{ is } AAS\ .$$

Proof. This proposition is proved by the following lemmas. □

Lemma 3.3. *For any set $\mathcal{M}$ of matrices,*

$$\hat{\rho}(\mathcal{M}) < 1 \Rightarrow DLI(\mathcal{M}) \text{ is } AAS\ . \tag{3.2}$$

Proof. Let us start from $\hat{\rho}(\mathcal{M}) < 1$. This means, from the definition, that

$$\lim_{k\to\infty} \left(\hat{\rho}_k(\mathcal{M})\right)^{1/k} < 1\ .$$

We can therefore pose

$$\lim_{k\to\infty} \left(\hat{\rho}_k(\mathcal{M})\right)^{1/k} = \gamma, \text{ with } \gamma < 1\ .$$

So, by the definition of lim, and noting that $\gamma < \frac{\gamma+1}{2} < 1$, we can state that

$$\exists N \in \mathbb{N} : \forall k \geq N,\ \hat{\rho}_k(\mathcal{M})^{1/k} < \frac{\gamma+1}{2} < 1\ ,$$

which can be rewritten as

$$\exists N \in \mathbb{N} : \forall k \geq N,\ \hat{\rho}_k(\mathcal{M}) < \left(\frac{\gamma+1}{2}\right)^k < 1\ .$$

As $\lim_{k\to\infty} \left(\frac{\gamma+1}{2}\right)^k = 0$, and using the fact that $\hat{\rho}_k(\mathcal{M})$ cannot be negative, we deduce that

$$\lim_{k\to\infty} \hat{\rho}_k(\mathcal{M}) = 0\ .$$

Then, taking back the definition (1.12), which was

$$\hat{\rho}_k(\mathcal{M}) := \sup\left\{ \left\| \prod_{i=1}^{k} M_i \right\| : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k \right\}\ ,$$

we have, for any product of matrices $M_i \in \mathcal{M}$, $0 \leq \left\| \prod_{i=1}^{k} M_i \right\| \leq \hat{\rho}_k(\mathcal{M})$, with $\hat{\rho}_k(\mathcal{M}) \to 0$ for $k \to \infty$. So, for any matrix product,

$$\lim_{k\to\infty} \left\| \prod_{i=1}^{k} M_i \right\| = 0\ .$$

And, by definition of a matrix norm, recalling that $\|A\| = 0 \iff A = 0$,

$$\lim_{k\to\infty} \prod_{i=1}^{k} M_i = 0\ ,$$

which is the definition of AAS. So far, we have proved that

$$\hat{\rho}(\mathcal{M}) < 1 \Rightarrow \text{DLI}(\mathcal{M}) \text{ is AAS}\ .$$

□

We now prove the converse implication, limiting ourselves to irreducible sets. We know that reducible sets of matrices can be decomposed into several irreducible sets, and those can be studied independently, as presented in Proposition 1.11.

Lemma 3.4. *For an irreducible and bounded set $\mathcal{M}$ of matrices,*

$$\hat{\rho}(\mathcal{M}) \geq 1 \Rightarrow DLI(\mathcal{M}) \textit{ is not AAS}\ . \tag{3.3}$$

Proof. If the set $\mathcal{M}$ is compact, we can directly use Barabanov's result, which is detailed in Theorem 4.15. Is shows that there exists a norm $\|.\|_*$ such that

$$\forall x, \exists A \in \mathcal{M} : \|Ax\|_* \geq \|x\|_*\ .$$

So, for $\|x_0\|_* = 1$, there is $A(0) \in \mathcal{M}$ such that, for $x_1 = A(0)x_0$, we have $\|x_1\|_* \geq 1$. Using the same theorem again, we know there is $A(1) \in \mathcal{M}$ such that, for $x_2 = A(1)x_1$, we have $\|x_2\|_* \geq 1$. Repeating this procedure, we can build a sequence of points that clearly does not converge to the origin.

Let us now consider the case where $\mathcal{M}$ is bounded but not compact. In this case, let us first define a sequence of δ_k such that $0 < \delta_k < 1$ and

$$\lim_{N\to\infty} \prod_{k=1}^{N} \delta_k = 1/2 \ .$$

Now pick x_0 such that $\|x_0\|_* = 1$. As above, we make use of Theorem 4.15, this time applied to $cl\mathcal{M}$, to prove the existence of a matrix $A(0)$ such that $\|x_1\|_* \geq 1$ for $x_1 = A(0)x_0$. But $A(0)$ is now only guaranteed to exist in the closure of $\mathcal{M}$, $cl\mathcal{M}$. Yet, by continuity, for any $0 < \delta_1 < 1$, we can pick a matrix $\tilde{A}(0) \in \mathcal{M}$ such that

$$\left\| \left(\tilde{A}(0) - A(0) \right) x_0 \right\|_* \leq 1 - \delta_1 \ .$$

From this, we define $x_1 = \tilde{A}(0)x_0$, which yields

$$\begin{aligned} \|x_1\|_* &= \left\| A(0)x_0 - \left(A(0) - \tilde{A}(0) \right) x_0 \right\|_* \\ &\geq \|A(0)x_0\|_* - \left\| \left(A(0) - \tilde{A}(0) \right) x_0 \right\|_* = 1 - (1 - \delta_1) = \delta_1 \ . \end{aligned}$$

The next step is to repeat the same procedure. We use the existence of $A(1) \in cl\mathcal{M}$ such that, for $x_2 = A(1)x_1$, $\|x_2\|_* \geq \|x_1\|_*$ and then, pick $\tilde{A}(1) \in \mathcal{M}$ such that $\left\| \left(\tilde{A}(1) - A(1) \right) x_1 \right\| \leq 1 - \delta_2$. Using this construction iteratively, we can build a sequence (x_i) such that

$$\|x_N\|_* \geq \prod_{k=1}^{N} \delta_k \ .$$

By the choice of the δ_k, we know that $\|x_N\|_*$ does not tend to zero, as $N \to \infty$:

$$\lim_{N\to\infty} \|x_N\|_* \geq \lim_{N\to\infty} \prod_{k=1}^{N} \delta_k = \frac{1}{2} \ .$$

Consequently, DLI($\mathcal{M}$) is not AAS, which concludes the proof. □

We may note that the equivalence (3.3) does not hold in general for unbounded sets of matrices, as can be seen in the following example.

Example 3.5. Consider the unbounded set

$$\mathcal{M} = \left\{ \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix}, \alpha \in \mathbb{R} \right\} \ .$$

First, we can notice that any product containing two matrices of the type $\begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix}$ is zero, as

$$\begin{pmatrix} 0 & \alpha_1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \alpha_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \ \forall \alpha_1, \alpha_2 \in \mathbb{R} \ .$$

So, the non-zero products of length k are of the type

$$\begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix} \left(\frac{1}{2}\right)^{k-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{k-1} \text{ or } \left(\frac{1}{2}\right)^{k} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{k} \ .$$

Those are all converging to zero, as $k \to \infty$. The system generated by these matrices is consequently AAS. On the other hand, the JSR is unbounded. Indeed,

$$\hat{\rho}_k(\mathcal{M})^{1/k} = \sup_{\alpha \in \mathbb{R}} \left\| \frac{1}{2^{k-1}} \begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix} \right\|^{1/k}$$

is unbounded, for any k. And so, the JSR cannot be bounded. This concludes the example.

3.2.2 Joint spectral radius and PAS

Now that the link between the joint spectral radius and AAS is established, we look for conditions establishing the PAS. We know that $AAS \Rightarrow PAS$, but it appears that the joint spectral radius does not yield any necessary condition for PAS. However, the periodic asymptotic stability of a discrete linear inclusion DLI($\mathcal{M}$) can be linked to the spectral radius of the matrices A_ω associated to finite words ω.

More formally, consider a set $\mathcal{M}$ of matrices M_i, where $i \in I$ and I is the index set. Associating words to matrix products as we did earlier, we can state the following equivalence, as presented in [42].

Lemma 3.6.

$$\rho(A_\omega) < 1, \forall \textit{ finite } \omega \iff DLI(\mathcal{M}) \textit{ is } PAS \ .$$

Proof. The proof is straightforward, considering the following relation, presented earlier as Lemma 1.1 : if A is a $n \times n$ matrix,

$$\lim_{k \to \infty} A^k = 0 \iff \rho(A) < 1 \ . \tag{3.4}$$

So, saying that $\forall$ finite $\omega, \lim_{k\to\infty} A_\omega^k = 0$ (this is the definition of PAS) is equivalent to saying that $\forall$ finite $\omega, \rho(A_\omega) < 1$. □

To summarize, for bounded sets, we have the following implications and correspondences :

$$\begin{array}{ccc} AAS & \Rightarrow & PAS \\ \Updownarrow & & \Updownarrow \\ \hat{\rho}(\mathcal{M}) < 1 & \Rightarrow & \rho(A_\omega) < 1, \forall \omega \ . \end{array}$$

From the above correspondences, we can understand what the property expressed in the *Finiteness Conjecture* would have implied (at this time, it is known to be false). From a dynamical systems point of view, the consequent implication would have been PAS $\Rightarrow$ AAS. We refer the reader to Section 4.1 for the exact statement of this conjecture, where we discuss it extensively, as well as its counterproof.

3.2.3 Generalized spectral radius and AAS

As shown in subsection 1.2.2, when the set $\mathcal{M}$ is finite, the joint and generalized spectral radii do match, i.e. $\bar{\rho}(\mathcal{M}) = \hat{\rho}(\mathcal{M})$. In this case, we have the analog to the relation (3.2) :

Lemma 3.7. *For a bounded set $\mathcal{M}$ of matrices,*

$$\rho(\mathcal{M}) < 1 \iff DLI(\mathcal{M}) \text{ is } AAS \ .$$

Let us get back to the more general case where $\bar{\rho}(\mathcal{M}) \neq \hat{\rho}(\mathcal{M})$ is possible. As we know from property (1.17), $\bar{\rho}(\mathcal{M}) \leq \hat{\rho}(\mathcal{M})$ for any set $\mathcal{M}$ of linear operators. It is then clear that $\hat{\rho}(\mathcal{M}) < 1 \Rightarrow \bar{\rho}(\mathcal{M}) < 1$ and we have

$$\text{DLI}(\mathcal{M}) \text{ is AAS} \Rightarrow \bar{\rho}(\mathcal{M}) < 1 \ .$$

But the implication $\bar{\rho}(\mathcal{M}) < 1 \Rightarrow \text{DLI}(\mathcal{M})$ is AAS, is not true in general for unbounded sets $\mathcal{M}$, as can be seen in this counterexample from [26] involving the set

$$\mathcal{M} = \left\{ \begin{pmatrix} 1/2 & 1 \\ 0 & 1/2 \end{pmatrix}, \begin{pmatrix} 1/2 & 2 \\ 0 & 1/2 \end{pmatrix}, \ldots, \begin{pmatrix} 1/2 & 2^n \\ 0 & 1/2 \end{pmatrix}, \ldots \right\} ,$$

and leading to $\rho(\mathcal{M}) = \frac{1}{2} \neq \hat{\rho}(\mathcal{M}) = +\infty$. Indeed, this is a case where $\rho(\mathcal{M}) = \frac{1}{2} < 1$ but the matrix product of fixed length k which yields the supremum found in the definition of $\hat{\rho}(\mathcal{M})$ tends, at $k \to \infty$, to the limit matrix $\begin{pmatrix} 0 & +\infty \\ 0 & 0 \end{pmatrix}$, which is far from converging to zero. As it does not send every $x \in \mathbb{R}^2$ to $(0,0)$, the set $\mathcal{M}$ is not AAS. This leads to the conclusion that the GSR cannot be used to determine the stability of discrete linear inclusions, in the case of an unbounded set of matrices.

3.3 Lower spectral radius and zero-reachability

3.3.1 Definitions of zero-reachability

In a similar way as the joint spectral radius with absolute asymptotic stability, the lower spectral radius (defined in section 1.3) can be linked to a notion of zero-reachability. We will present two definitions of this zero-reachability, in a similar fashion as what we did for the earlier AAS and PAS. We will denote them by Absolute Zero Reachability and Periodic Zero Reachability (AZR and PZR).

The main difference between the previous definitions and the ones that we are about to give is that the first ones implied a convergence to 0 *for all products*, while these new ones will only need *the existence of at least one product* converging to 0.

AZR : Absolute Zero Reachability

We define AZR as follows. There *exists at least one* switching sequence (i_k), such that, for any initial point x_0, the corresponding trajectory (x_k) given by DLI($\mathcal{M}$) converges to the origin :

$$\exists (i_k),\ \forall x_0 : \lim_{k\to\infty} x_k = 0 \ .$$

This comes to saying that at least one matrix product converges to the zero matrix. So, this means that there exists at least one switching signal (i_k) such that

$$\lim_{k\to\infty} A_{i_k} A_{i_{k-1}} \dots A_{i_1} = 0 \ .$$

PZR : Periodic Zero Reachability

As before when PAS was a restriction of AAS, we define PZR as a restriction of AZR to periodic sequences. For a system to be PZR, there must exist a periodic matrix product that converges to the zero matrix. So, let us consider a finite sequence $(i_1, i_2, \dots, i_m)$, which can be used as a period. The corresponding matrix product is

$$A_\omega = A_{i_m} A_{i_{m-1}} \dots A_1 \ .$$

In this context, a DLI is PZR if

$$\exists \text{ finite } \omega : \lim_{k\to\infty} (A_\omega)^k = 0 \ .$$

3.3.2 New implications

We now present relations that can be found to exist between the lower spectral radius and these zero-reachabilities. Using the same notations as before, let $\mathcal{M}$ be a set of $n \times n$ square matrices: $\mathcal{M} = \{A_1, A_2, \ldots\}$. And let $A_\omega = A_{\omega_k} A_{\omega_{k-1}} \ldots A_{\omega_1}$ be the matrix associated to a word $\omega = \omega_1 \omega_2 \ldots \omega_k$ of length $|\omega| = k$.

PZR and $\rho(A_\omega) < 1$

We first show a simple equivalence that will help us clarify the situation. Let us suppose that DLI($\mathcal{M}$) is PZR, i.e. there exists a finite word ω such that $\lim_{n\to\infty}(A_\omega)^n = 0$. We can now use the very same property as the one we used before, presented in relation (3.4) :

$$\lim_{k\to\infty} A^k = 0 \iff \rho(A) < 1 \ .$$

As a consequence, we have the similar result that PZR is equivalent to the existence of a finite word ω such that $\rho(A_\omega) < 1$.

Link between PZR and AZR

Previously, we have seen that AAS is the strongest form of stability and implies PAS, while PAS does not automatically imply AAS. We will see now that the situation is a bit different about AZR and PZR. Indeed, AZR and PZR are equivalent, as showed in the following lemma.

Lemma 3.8. *The above definitions directly imply that*

$$PZR \iff AZR \ .$$

Proof. First, let us assume that the system is PZR. Then there exists a periodic trajectory converging to 0, and this trajectory can be regarded as the one that converges to 0 in the definition of AZR. So, we immediately have

$$PZR \Rightarrow AZR \ .$$

On the other hand, AZR implies PZR, and this can be shown as follows. The main idea is to take the series achieving the stability in the definition of AZR and truncate it wisely to generate a periodic sequence also converging to zero. To this aim, let us consider a non-periodic infinite sequence of indices $(i_1, i_2, \ldots, i_k, \ldots)$. We will use the following notation : $A_{\omega_k} = A_{i_k} A_{\omega_{k-1}}$ with $A_{\omega_1} = A_{i_1}$. Let us suppose that we have AZR, that is

$$\lim_{k\to\infty} \|A_{\omega_k}\| = 0 \ .$$

As this is a convergent product, we know that

$$\forall \varepsilon > 0 \ , \ \exists N \in \mathbb{N} : \forall n \geq N \ , \ \|A_{\omega_n}\| < \varepsilon \ .$$

So, let us pick an arbitrary $\varepsilon^* < 1$ and the corresponding N^*. Using the sub-multiplicativity of the matrix norm, we can write

$$\|(A_{\omega_{N^*}})^l\| \leq \|A_{\omega_{N^*}}\|^l < (\varepsilon^*)^l \ , \text{which } \rightarrow 0 \ \text{ as } \ l \rightarrow \infty \ .$$

So,

$$\|(A_{\omega_{N^*}})^l\| \rightarrow 0 \ \text{ as } \ l \rightarrow \infty \ ,$$

which is the condition for PZR, and thus

$$AZR \Rightarrow PZR \ .$$

This concludes the proof. □

3.4 Stability and Common Quadratic Lyapunov Functions

In the context of dynamical systems, and more precisely the study of their stability, a widely used notion is the Lyapunov function. A Lyapunov function is a scalar function, which we will denote by V. In our case, this function is defined on the space of the variables x. When studying the stability of the origin, the interesting properties of V are the following:

- $V(x) > 0, \forall x \neq 0$,
- the derivatives of V along the trajectories of the system must be ≤ 0 (at least in a domain D around the origin). This means that V must be non-increasing as time evolves.

Such a function is used to prove the stability of the trajectories around the origin. Indeed, it guarantees that once the trajectories reaches a level set of V that is included in the domain D, they do not leave it. If we have a strictly decreasing V, then the origin is attracting the trajectories. This is the case we will be considering from now on.

The existence of a Lyapunov function is clearly a sufficient condition for the stability of a system. On the other hand, the existence of an extremal norm $\|.\|_{Bar}$ is guaranteed by Theorem 4.15 for switched linear systems generated by bounded irreducible sets of matrices (see Definition 4.14 for the definition of an extremal norm). And so, in case of stability, this extremal norm is actually a Lyapunov function, as it decreases along the trajectories of the system and is positive except at the origin (thanks to the norm properties). As a consequence, in the case at hand, the existence of a Lyapunov function is also necessary.

The main issue about Lyapunov functions is to find one, as nothing tells us which form this function has. Once a candidate Lyapunov function is known, we just need to check if it satisfies the required conditions. In this context, it can be useful to restrict ourselves to a specific type of Lyapunov functions, that could be easier to compute. This computational aspect has led us to consider the case of the *quadratic* Lyapunov functions, presented in the following section.

3.4.1 Quadratic Lyapunov function of a (switched) linear system

Before discussing further the existence of a quadratic Lyapunov function for discrete linear inclusions, let us illustrate such functions for linear systems. We will present continuous-time systems before moving to discrete-time systems.

Let us consider a quadratic scalar function of a vector $x \in \mathbb{R}^n$, which can be expressed as

$$V(x) = x^T P x \ ,$$

where P is a $n \times n$ real matrix. It is a quadratic Lyapunov function for the system

$$\dot{x}(t) = Ax(t)$$

if it satisfies the following criteria:

- $V(x) > 0$ for $x \neq 0$ and $V(x) = 0$ if $x = 0$. This means that P has to be a symmetric positive definite matrix, which we denote by $P \succ 0$.
  A symmetric matrix indeed has real eigenvalues, and it is positive (semi)definite if its eigenvalues are positive (non negative).
- $\frac{\partial V}{\partial t} < 0$ along the trajectories, which implies that

$$\begin{aligned} \dot{V}(x) = \quad & \dot{x}^T P x + x^T P \dot{x} < 0, \ \forall x \neq 0 \\ & x^T A^T P x + x^T P A x < 0, \ \forall x \neq 0 \ . \end{aligned}$$

So, the matrix $A^T P + PA$ must be negative definite ($\prec 0$).

These conditions rule the quadratic Lyapunov function of a linear system in continuous-time. In this simple case, the existence of a quadratic Lyapunov function is equivalent to the stability criterion in continuous time $\Re(\lambda(A)) < 0$ (see [82]) . Thus, to any stable matrix we can associate a quadratic Lyapunov function.

Now, let us extend this to a *switched* linear system, as described in Section 3.1. As a reminder, the evolution is then ruled by an equation of the type

$$\dot{x}(t) \in \left\{ A_{i(t)} x(t) : i(t) \in \{1, \ldots, m\} \right\} \ .$$

In this new case, the above conditions for $V(x) = x^T P x$ to be a Lyapunov function must hold simultaneously for all matrices A_i in the set $\mathcal{M}$:

- $P \succ 0$, this condition does not change.
- $\frac{\partial V}{\partial t} < 0$ along any possible trajectory, which implies that

$$A_i^T P + P A_i \prec 0, \forall i \ .$$

This last condition ensures that, under any switching strategy, the value of V decreases. The function $V(x)$ is then called a *common quadratic Lyapunov function* (CQLF), as this must be a valid quadratic Lyapunov function for each of the linear systems considered separately.

Now, for discrete-time linear systems, we have very similar conditions. Indeed, again considering a function of the type $V(x) = x^T P x$, we still have $P \succ 0$, but the evolution law $x_{t+1} = A x_t$ yields slightly different conditions for the decrease of V. We have :

$$\begin{aligned} V(x_{t+1}) &< V(x_t), \ \forall x_t \neq 0 \\ x_{t+1}^T P x_{t+1} &< x_t^T P x_t, \ \forall x_t \neq 0 \\ x_t^T A^T P A x_t &< x_t^T P x_t, \ \forall x_t \neq 0 \\ x_t^T (A^T P A - P) x_t &< 0, \ \forall x_t \neq 0 \\ A^T P A - P &\prec 0 \ . \end{aligned}$$

These quadratic Lyapunov conditions are again equivalent to the stability criterion, which is $|\lambda(A)| < 1$ in discrete time (mentioned in Section 1.1.3, Lemma 1.1). This equivalence is a standard result and can be found in [82] (Theorem 4.5 from Chapter VII therein). And consequently, here in discrete time, as previously for the continuous time case, to any stable matrix we can associate a quadratic Lyapunov function.

And, generalizing this to switched linear systems in discrete time (that is, DLI), this last relation would have to hold simultaneously for all matrices A_i of the set :

$$A_i^T P A_i - P \prec 0, \forall i \ .$$

Together with $P \succ 0$, these are the conditions for $x^T P x$ to be a *common quadratic Lyapunov function (CQLF)* of a DLI generated by a set of matrices $\{A_1, \ldots, A_i, \ldots\}$. The following section details what such Lyapunov functions can teach us about the stability of switched linear systems..

3.4.2 Non necessity in continuous time : an example

Although the existence of a CQLF clearly guarantees the stability of a switched linear system, it is not a necessary condition. For continuous-time systems, this can be seen in the following example, which was presented by Dayawansa & Martin in [30]. To be complete, we mention that Brockett did publish a similar counter-example much earlier, in [20].

Consider the continuous-time switched system

$$\dot{x} = A_i^{(c)} x, \quad A_i^{(c)} \in \left\{ A_1^{(c)}, A_2^{(c)} \right\} ,$$

generated by the following two matrices, the second of which depends on the real parameter a:

$$A_1^{(c)} = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} \text{ and } A_2^{(c)} = \begin{pmatrix} -1 & -a \\ 1/a & -1 \end{pmatrix} . \tag{3.5}$$

Note that the notation $^{(c)}$ is there to remind the reader that we are dealing with a continuous-time system. Moreover, both matrices are individually stable, as their eigenvalues have negative real parts. Without this property, the instability of the switched system would be trivial. As an illustration, trajectories generated by these matrices from an arbitrary starting point are shown in Figure 3.2.

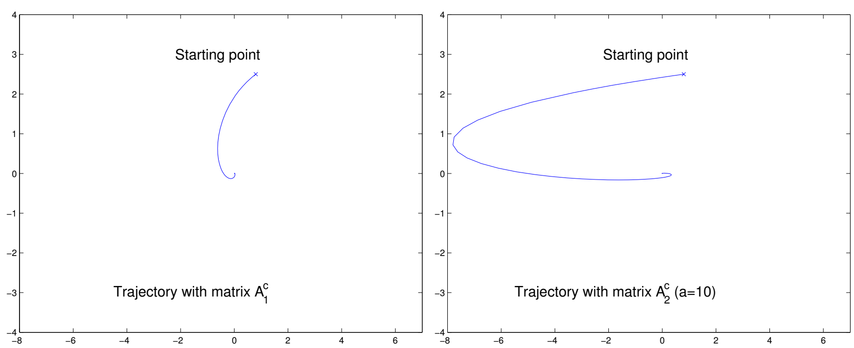

Fig. 3.2. Sample trajectories generated by $A_1^{(c)}$ (on the left) and $A_2^{(c)}$ with $a = 10$ (on the right) from the arbitrary initial point $x_0 = (0.8, 2.5)$.

For specific values of a, the trajectories of this system are shown to converge to the origin for any switching between the two evolution laws. To do so, the authors define a "worst case" switching strategy, constructed by choosing at every instant the matrix giving the direction pointing the farthest away from the origin, as illustrated in Figure 3.3. Using this rule, the (x_1, x_2) plane is divided into 4 zones by 2 lines containing the origin. These 2 lines are the places where $A_1^{(c)}x$ is parallel to $A_2^{(c)}x$. The authors then show, by integration, that this "worst case" strategy is stable (i.e. goes to 0) for a range of the parameter a including the interval $[1, 10]$. So, for this range of a, any switching strategy other than the one described generates a trajectory that is closer to the origin, and hence also going to 0. In particular, this range of stability of a

includes the value $a = 10$, which we will use further. A simulation depicting a trajectory of the system using this law is presented in Figure 3.4 for $a = 10$.

Now, to determine the existence of a CQLF for this system, consider the following parametrization:

$$V(x_1, x_2) = x_1^2/2 + px_1x_2 + qx_2^2/2 ,$$

so that $V = \frac{1}{2} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & p \\ p & q \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Then the decrease of V along the trajectories implies

$$\nabla V^T A_i^{(c)} x < 0, \forall x, \text{ for } i = 1, 2 .$$

In this expression, ∇V stands for the *gradient* of V, and the scalar product of ∇V and $A_i^{(c)} x$ gives the variation of V in the direction $A_i^{(c)} x$. These conditions give

$$\begin{pmatrix} x_1 + px_2 \\ px_1 + qx_2 \end{pmatrix}^T A_i^{(c)} x < 0, \forall x, \text{ for } i = 1, 2 .$$

Explicitly, with $A_2^{(c)}$, this gives

$$\begin{pmatrix} x_1 + px_2 \\ px_1 + qx_2 \end{pmatrix}^T \begin{pmatrix} -1 & -a \\ 1/a & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} < 0, \forall x ,$$

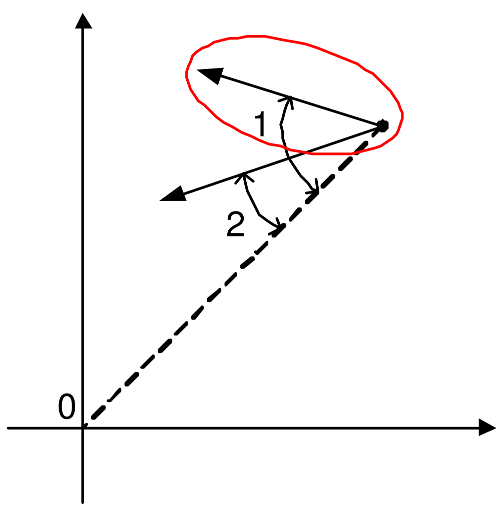

Fig. 3.3. The considered angles are denoted by 1 and 2. They are, respectively, the angles formed by $A_1^{(c)}(x, y)$ and $A_2^{(c)}(x, y)$ with the vector from (x, y) to the origin $(0, 0)$. The idea is to constantly be using the matrix giving the largest angle, as is the case for angle 1 in this figure.

which can be rewritten as

$$x_1^2\left(\frac{p}{a}-1\right)+x_1x_2\left(\frac{q}{a}-a-2p\right)+x_2^2(-pa-q)<0\ ,\forall x. \tag{3.6}$$

Note that we just need to replace a by 1 to obtain the same relation with $A_1^{(c)}$:

$$x_1^2\,(p-1)+x_1x_2\,(q-1-2p)+x_2^2(-p-q)<0\ ,\forall x. \tag{3.7}$$

These two quadratic forms must be simultaneously negative, for any value of (x_1, x_2). Standard computations show that (3.6) and (3.7) respectively lead to:

$$p^2+\frac{1}{8}\left(\frac{q}{a}-3a\right)^2<a^2$$
$$p^2+\frac{1}{8}\,(q-3)^2<1\ ,$$

which define the interiors of two ellipsoids, when considered in the (p, q) plane. This is illustrated in Figure 3.5. If these ellipsoids intersect, they provide adequate values for the parameters p and q, and thus a CQLF exists. The threshold

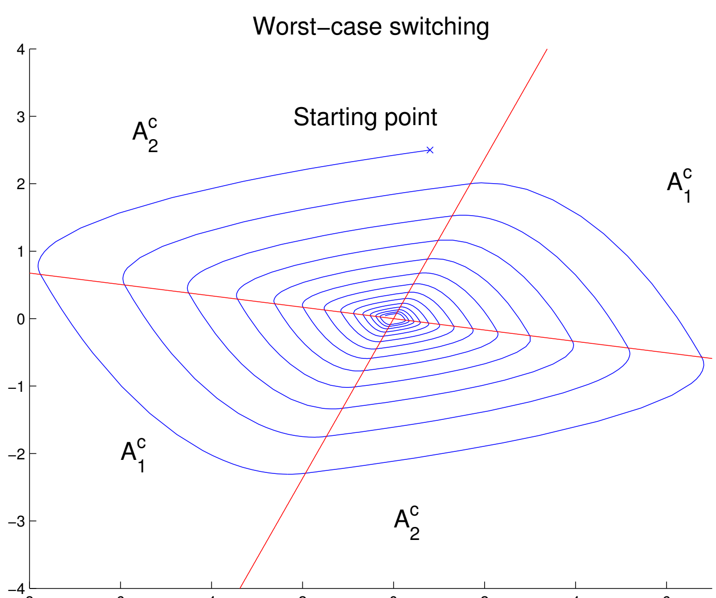

Fig. 3.4. The (x_1, x_2) plane is divided radially into four zones. The most diverging strategy is to apply $A_1^{(c)}$ in two of these, and $A_2^{(c)}$ in the other two. Any other switching will yield a trajectory approaching the origin faster than the one depicted here. Here, the simulation was run for $a = 10$.

value for this to happen can be shown to be $a = 3 + \sqrt{8}$. For lower values of a, there exists a CQLF, while there is none for $a > 3 + \sqrt{8}$. This value is still less than 10, which was shown earlier to generate a stable system. Consequently, this shows that the values of a in $[3 + \sqrt{8}, 10]$ generate stable systems, yet not admitting any CQLF.

3.4.3 Non necessity in discrete time : an example

The previous example shows that a continuous-time switched linear system can be stable without admitting a CQLF. We now turn to discrete-time systems to see if the same property holds.

The first idea that comes to mind is to directly derive a discrete-time system from the continuous-time one by Dayawansa & Martin presented above. To this aim, we will here use a transformation called *linear fractional transformation* (or bilinear transformation, or Möbius transformation). Note that the choice of this transformation is not trivial. A full discussion of the transformations from discrete to continuous time (and vice versa) will be carried out in Section 3.5. So, we will here limit ourselves to giving the expression of this linear fractional transformation.

This transformation can be found in [50] and is widespread. Indeed, it appears in several references from various domains (see, for example, [37], which deals with systems approximation). It is considered to link discrete-time and continuous-time systems together. Having a discrete-time system expressed as

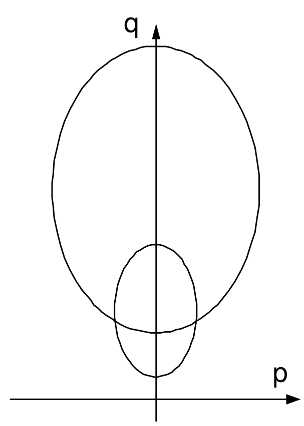

Fig. 3.5. Representation of the ellipsoids defined by inequalities (3.4.2). In the plane (p, q), they are centered respectively around (0,3) and $(0, 3a^2)$. A CQLF exists if and only if the interiors of these two ellipsoids do intersect, then providing adequate values for the parameters p and q.

$$x_{k+1} = A_i^{(d)} x_k, \quad A_i^{(d)} \in \{A_1^{(d)}, \ldots, A_n^{(d)}\} ,$$

the continuous-time analog to this system is defined as

$$\dot{x} = A_i^{(c)} x, \quad A_i^{(c)} \in \{A_1^{(c)}, \ldots, A_n^{(c)}\}$$

with

$$A_i^{(c)} = (A_i^{(d)} - I)(A_i^{(d)} + I)^{-1} .$$

Let us apply this transformation to the example (3.5) with $a = 6$ and see what happens to the system in discrete-time evolution. The matrices become

$$A_i^{(d)} = (I - A_i^{(c)})^{-1}(I + A_i^{(c)}) .$$

So,

$$\begin{aligned} A_1^{(d)} &= \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}\right] \\ &= \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0.4 & -0.2 \\ 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -0.2 & -0.4 \\ 0.4 & -0.2 \end{pmatrix} , \end{aligned}$$

$$\begin{aligned} A_2^{(d)} &= \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 & -6 \\ 1/6 & -1 \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & -6 \\ 1/6 & -1 \end{pmatrix}\right] \\ &= \begin{pmatrix} 2 & 6 \\ -1/6 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -6 \\ 1/6 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0.4 & -1.2 \\ 1/30 & 0.4 \end{pmatrix} \begin{pmatrix} 0 & -6 \\ 1/6 & 0 \end{pmatrix} = \begin{pmatrix} -0.2 & -2.4 \\ 1/15 & -0.2 \end{pmatrix} . \end{aligned}$$

We now have the discrete-time system

$$x_{k+1} = A_i^{(d)} x_k, \quad A_i^{(d)} \in \{A_1^{(d)}, A_2^{(d)}\} ,$$

generated by the two matrices

$$A_1^{(d)} = \begin{pmatrix} -0.2 & -0.4 \\ 0.4 & -0.2 \end{pmatrix} \text{ and } A_2^{(d)} = \begin{pmatrix} -0.2 & -2.4 \\ 1/15 & -0.2 \end{pmatrix} .$$

This switched system is stable. A way to see this is to compute its joint spectral radius and check that it is less than 1. At this point, we have not yet discussed how to compute it, as it is the subject of chapter 6. However, for now, we can say that it is numerically found to be in the interval

$$0.9275 \leq \hat{\rho} \leq 0.9510 \ .$$

Moreover, this system does not admit any CQLF. Indeed, we have seen that the initial continuous time system has no CQLF and we use the fact that the linear fractional transformation preserves the existence of a CQLF (this will be shown in Section 3.5).

Note that we could have used the value $a = 10$ for the previous example. Then, instead of providing a discrete-time example of a stable switched linear system having no CQLF, we would have ended up discovering that this linear fractional transformation does not preserve the stability of the switched linear system. This issue will be studied in Section 3.5.

3.5 Link between discrete time and continuous time systems

Many of the references on the stability of switched systems discuss the stability of switched linear systems from the continuous time perspective, like the survey papers [29], [32] or [54] do. There, the stability of discrete-time systems is often considered as a natural extension of the results for continuous time systems. This extension is often implicitly based on the standard *linear fractional transformation* of the type

$$z \rightarrow \frac{1+z}{1-z} \ .$$

Back in Section 3.4.3, when we tried to apply this type of transformation to examples used in [30], we noticed differences between discrete-time and continuous-time systems. In this section, we present the observations we have gathered about this issue. Some of these were presented in [14, 15]. We begin with well-known transformations between discrete and continuous time and then we will discuss the properties we would like theses transformations to have.

3.5.1 The linear fractional transformation

This transformation has been used earlier in this chapter, and we here give details about it, as well as its properties.

Definition

Having a continuous-time switched system expressed as

$$\dot{x} = A_i^{(c)} x, \quad A_i^{(c)} \in \{A_1^{(c)}, \ldots, A_n^{(c)}\} \ ,$$

the discrete-time analog to this system is defined as

$$x_{k+1} = A_i^{(d)} x_k, \quad A_i^{(d)} \in \{A_1^{(d)}, \dots, A_n^{(d)}\} ,$$

with

$$A_i^{(d)} = (I - A_i^{(c)})^{-1}(I + A_i^{(c)}) .$$

We now review the properties that can be derived from this transformation.

Lyapunov-preserving

This transformation has the nice property of preserving the existence of a common quadratic Lyapunov function for the switched system. Indeed, let us consider the Lyapunov equation in the continuous time case,

$$A^{(c)^T} P^{(c)} + P^{(c)} A^{(c)} = -Q^{(c)} ,$$

with $Q^{(c)} \succ 0$, which is another way of saying that $A^{(c)^T} P^{(c)} + P^{(c)} A^{(c)} \prec 0$. Assuming the invertibility of $(A_i^{(d)} + I)$, the following change of variables (see [50]), where I is the identity matrix,

$$\begin{aligned} A_i^{(c)} &= (A_i^{(d)} - I)(A_i^{(d)} + I)^{-1} \\ Q^{(c)} &= 2(A^{(d)^T} + I)^{-1} Q^{(d)} (A^{(d)} + I)^{-1} \\ P^{(c)} &= P^{(d)} , \end{aligned}$$

yields, after going through some algebra,

$$A^{(d)^T} P^{(d)} A^{(d)} - P^{(d)} = -Q^{(d)} .$$

This transformation is such that $P^{(d)}$ and $Q^{(d)}$ are positive definite whenever $P^{(c)}$ and $Q^{(c)}$ are. So, the above expression corresponds to the definition of the common quadratic Lyapunov function of a discrete-time switched linear system, as it is equivalent to $A^{(d)^T} P^{(d)} A^{(d)} - P^{(d)} \prec 0$.

It can be noted that the eigenvalues $\lambda_i^{(d)}$ of $A_i^{(d)}$ can be derived immediately from those $\lambda_i^{(c)}$ of $A_i^{(c)}$, because $A_i^{(d)}$ is a function of $A_i^{(c)}$, and vice-versa. Indeed, we have:

$$\begin{aligned} A_i^{(c)} &= f(A^{(d)}) = (A_i^{(d)} - I)(A_i^{(d)} + I)^{-1} \\ A_i^{(d)} &= f(A^{(c)}) = (I - A_i^{(c)})^{-1}(I + A_i^{(c)}) . \end{aligned}$$

And so, using the fact $I = (A_i^{(d)})^0 = (A_i^{(c)})^0$, we can relate the eigenvalues in the following way, using the spectral mapping theorem (see [36], Vol.I, chap.V):

$$\lambda_i^{(c)} = f(\lambda_i^{(d)}) = (\lambda_i^{(d)} - 1)(\lambda_i^{(d)} + 1)^{-1}$$
$$\lambda_i^{(d)} = f(\lambda_i^{(c)}) = (1 - \lambda_i^{(c)})^{-1}(1 + \lambda_i^{(c)}) \ .$$

This shows that A_c matrices with negative real part eigenvalues correspond to A_d matrices with eigenvalues of modulus inferior to 1. So, stable matrices in continuous time correspond to stable matrices in discrete-time. As a particular case, the stability of a unique matrix is preserved when changing from discrete to continuous and vice-versa. But we will now show that this is not the case for the stability of a switched system in its whole.

Not Stability-preserving

This transformation does not preserve the stability of a switched system when going from continuous to discrete time. To show this, let us consider again the example (3.5) presented earlier in Section 3.4.3. This time, we pick $a = 10$ and see what happens to the system in discrete-time evolution. The matrices become, as said before,

$$A_i^{(d)} = (I - A_i^{(c)})^{-1}(I + A_i^{(c)}) \ .$$

So,

$$\begin{aligned} A_1^{(d)} &= \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}\right] \\ &= \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0.4 & -0.2 \\ 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -0.2 & -0.4 \\ 0.4 & -0.2 \end{pmatrix} \ , \end{aligned}$$

$$\begin{aligned} A_2^{(d)} &= \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 & -10 \\ 0.1 & -1 \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & -10 \\ 0.1 & -1 \end{pmatrix}\right] \\ &= \begin{pmatrix} 2 & 10 \\ -0.1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -10 \\ 0.1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0.4 & -2 \\ 0.02 & 0.4 \end{pmatrix} \begin{pmatrix} 0 & -10 \\ 0.1 & 0 \end{pmatrix} = \begin{pmatrix} -0.2 & -4 \\ 0.04 & -0.2 \end{pmatrix} \ . \end{aligned}$$

We now have the discrete-time system

$$x_{k+1} = A_i^{(d)} x_k, \quad A_i^{(d)} \in \{A_1^{(d)}, A_2^{(d)}\} \ ,$$

generated by the two matrices

$$A_1^{(d)} = \begin{pmatrix} -0.2 & -0.4 \\ 0.4 & -0.2 \end{pmatrix} \text{ and } A_2^{(d)} = \begin{pmatrix} -0.2 & -4 \\ 0.04 & -0.2 \end{pmatrix} .$$

This system is unstable, as there exist switching strategies leading to trajectories moving away from the origin. A first way to see this is to compute its joint spectral radius to see if it is greater than 1. At this point, we have not yet discussed how to compute it, as it is the subject of chapter 6. However, for now, we can say that it is indeed numerically found to be in the interval

$$1.228 \leq \hat{\rho} \leq 1.245 .$$

To reach the same conclusion without using future results, one can see that the simple matrix product

$$A_1^{(d)} A_2^{(d)} = \begin{pmatrix} 0.024 & 0.88 \\ -0.088 & -1.56 \end{pmatrix}$$

has -0.0265 and -1.5095 as eigenvalues. Consequently, the value $\sqrt{1.5095} \approx 1.228$ provides a lower bound on the joint spectral radius. This confirms that $\hat{\rho} > 1$. To sum up, simply repeating this pattern $A_1^{(d)} A_2^{(d)}$ yields a diverging trajectory. Of course, there is no need to bother about the existence of a CQLF, since the system is unstable. This concludes the example, as we have a system that is stable in continuous time, yet unstable in discrete time.

3.5.2 The matrix exponential

Definition

The matrix exponential is also a natural candidate for applying continuous-time systems onto discrete-time systems. We know that the integration of the equation $\dot{x} = Ax$ leads to solutions of the type $x = e^{At} x_0$. In a similar way, a natural way of discretizing the trajectories of $\dot{x} = Ax$ is to define the series (x_t) as $x_{t+1} = e^{A} x_t$. This leads to the following transformation from continuous to discrete time :

$$A_i^{(d)} = e^{A_i^{(c)}}, \quad i = 1, \ldots, n .$$

Although this transformation can look quite natural, is has drawbacks, which we present just below. First, it is not bijective because of periodicity of the complex exponential function. And there are also examples where a d-stable (short for *stable in discrete time*) system can be associated to a c-unstable system (short for *unstable in continuous time*), as we will show a bit further.

Non-bijective

The non-bijectivity property is due to the fact that $e^{i\theta} = \cos(\theta) + i\sin(\theta)$. This implies that several distinct matrices $A_i^{(d)}$ can have the same corresponding $A_i^{(c)}$ matrix. Indeed, let us take two pairs of complex conjugate eigenvalues with the same real part and a complex part differing by 2π , let us say

$$\begin{aligned} \lambda_{A\pm} &= -\alpha \pm \beta i \\ \lambda_{B\pm} &= -\alpha \pm (\beta + 2\pi)i \ . \end{aligned}$$

Let us now consider the two diagonal matrices

$$I_A = \begin{pmatrix} \lambda_{A_+} & 0 \\ 0 & \lambda_{A_-} \end{pmatrix} \text{ and } I_B = \begin{pmatrix} \lambda_{B_+} & 0 \\ 0 & \lambda_{B_-} \end{pmatrix} \ .$$

Because $\lambda_{A_\pm}$ and $\lambda_{B_\pm}$ only differ by $\pm 2i\pi$, we have

$$e^{I_A} = \begin{pmatrix} e^{\lambda_{A_+}} & 0 \\ 0 & e^{\lambda_{A_-}} \end{pmatrix} = \begin{pmatrix} e^{\lambda_{B_+}} & 0 \\ 0 & e^{\lambda_{B_-}} \end{pmatrix} = e^{I_B} \ .$$

As this work is written with real matrices in mind, we may use an invertible matrix T representing a base change to create real matrices A and B as follows:

$$\begin{aligned} A &= T I_A T^{-1} \\ B &= T I_B T^{-1} \ . \end{aligned}$$

Then, we can again check that these two matrices do have the same image using the exponential transformation:

$$e^A = e^{T I_A T^{-1}} = T e^{I_A} T^{-1} = T e^{I_B} T^{-1} = e^{T I_B T^{-1}} = e^B \ .$$

In this way, we can construct the following pair of matrices:

$$C_A = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix}, C_B = \begin{pmatrix} -1 & \sqrt{(n^*)} \\ -\sqrt{(n^*)} & -1 \end{pmatrix}$$

where we have chosen $n^* = 4\pi^2 + 4\pi + 1$. These matrices satisfy $e^{C_A} = e^{C_B}$. Indeed, the eigenvalues are found to be:

$$\begin{aligned} \lambda_{A\pm} &= -1 \pm i \\ \lambda_{B\pm} &= -1 \pm (1 + 2\pi)i \ . \end{aligned}$$

Note that here, the base change matrix would be

$$T = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \ .$$

This concludes our example where $C_A \neq C_B$, but $e^{C_A} = e^{C_B}$.

Not Lyapunov-preserving

While the linear fractional transformation described earlier was especially tailored to preserve the existence of a common quadratic Lyapunov function, the same cannot be said about the matrix exponential. We provide an example of a d-stable pair of matrices $\{e^{A_1^{(c)}}, e^{A_2^{(c)}}\}$ which admits a CQLF, while the c-stable pair $\{A_1^{(c)}, A_2^{(c)}\}$ does not. This non-existence has already been established in Section 3.4.2. We recall that we had considered the following pair of matrices:

$$A_1^{(c)} = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} , \; A_2^{(c)} = \begin{pmatrix} -1 & -a \\ 1/a & -1 \end{pmatrix} \text{with the real parameter } a \; .$$

We know that for $a \geq 3 + \sqrt{8}$, the switched system does not admit a CQLF, while it is stable for values of a including $[1, 10]$. Let us pick any value of a in the interval where we have stability without CQLF, say $a = 6$, for example. Then the pair $\{e^{A_1^{(c)}}, e^{A_2^{(c)}}\}$ is equal to

$$e^{A_1^{(c)}} = \begin{pmatrix} 0.1988 & -0.3096 \\ 0.3096 & 0.1988 \end{pmatrix} , e^{A_2^{(c)}} = \begin{pmatrix} 0.1988 & -1.8574 \\ 0.0516 & 0.1988 \end{pmatrix} \; ,$$

and we have the CQLF $V(x) = x^T P x$, where

$$P = \begin{pmatrix} 2.7532 & -1.7378 \\ -1.7378 & 15.6606 \end{pmatrix} \; .$$

Note that P is symmetric positive definite. And it satisfies

$$e^{A_i^{(c)}T} P e^{A_i^{(c)}} - P = -Q_i \quad \text{for } i = 1, 2 \; ,$$

where Q_i is also symmetric positive definite.

The opposite case is simpler, as the existence of a CQLF for a continuous time system immediately yields a CQLF for the discrete time system obtained by the exponential transformation. Indeed, the trajectories of this discrete time system are a subset of those of the continuous time system, and so the Lyapunov function of the continuous time system is still a Lyapunov function for the discrete time system.

Not Stability-preserving

First, it can be noted that every c-stable system is mapped to a d-stable system by the matrix exponential. The argument behind this claim is quite elementary. Indeed, any discrete time trajectory generated by $x_{t+1} \in \left\{ e^{A_i^{(c)}} \right\} x_t$ is itself a discretization of the trajectories of the system $\dot{x}(t) \in \left\{ A_i^{(c)} \right\} x(t)$. This is illustrated in Figure 3.6. Consequently, if all continuous trajectories converge to the origin, then so do all discrete trajectories as well.

The converse is false. Too see this, we provide an example of a d-stable pair of matrices $\{e^{A_1^{(c)}}, e^{A_2^{(c)}}\}$, while $\{A_1^{(c)}, A_2^{(c)}\}$ is c-unstable. Indeed, consider the following pair of matrices :

$$A_1^{(c)} = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}, A_2^{(c)} = \begin{pmatrix} -0.5 & -15 \\ 1 & -1/15 \end{pmatrix} .$$

The d-stability is obtained via the calculation of the joint spectral radius of the corresponding discrete time pair :

$$\left\{ e^{A_1^{(c)}}, e^{A_2^{(c)}} \right\} .$$

This value happens to be strictly less than 1. More precisely, it can be shown to lie in the interval [0.85 0.92]. The c-instability is obtained via numerical simulation. For instance, one can check that applying alternatively $A_2^{(c)}$ for a duration of 0.5 and then $A_1^{(c)}$ for a duration of 0.7 yields the following product:

$$\exp\left(0.7A_1^{(c)}\right)\ \exp\left(0.5A_2^{(c)}\right) \approx \begin{pmatrix} -0.20 & -1.11 \\ -0.03 & -1.10 \end{pmatrix} .$$

The eigenvalues of this matrix are approximately -1.15 and -0.16. This proves that repeating this sequence leads to divergence. And this is not even the worst

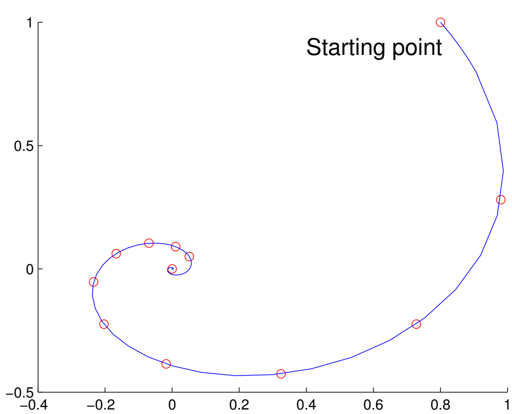

Fig. 3.6. This depicts both discrete and continuous time trajectories starting from the initial point (0.8,1). The curve is the continuous time trajectory generated by some matrix, while the circles are the trajectory generated by the matrix exponential of this matrix. The second trajectory is clearly a discretized version of the first one.

possible (i.e. most diverging) switching strategy. We can therefore conclude that the matrix exponential does not allow to deduce stability results from the discrete time domain to the continuous one. Moreover, we can note that the switched system generated by $\left\{e^{A_1^{(c)}}, e^{A_2^{(c)}}\right\}$ admits a CQLF in this specific case.

Comment about commutativity

At this point, we can highlight the role played by the non-commutativity of the matrices. As a reminder, we say that two matrices A and B *commute* when $AB = BA$. Let us consider a switched linear system in discrete time generated by a set of matrices that commute pairwise. The stability if this system is easy to determine, since the order of the matrix products has no impact on the result. Consequently, any finite product mixing A's and B's is equal to a product of the form $A^{k_1}B^{k_2}$, with $k_1, k_2 \in \mathbb{N}$. So, for this system to be AAS, we need $A^{k_1}B^{k_2}$ to converge to the zero matrix, as $k_1 + k_2 \to \infty$. As this must hold whatever the repartition between k_1 and k_2, it can quickly be seen that a necessary and sufficient condition for AAS is that both matrices A and B have a spectral radius less than 1.

On the other hand, for a continuous time switched linear system, we can make use of a similar property. If $AB = BA$, it is known that $\exp(A)\exp(B) = \exp(A + B) = \exp(B + A) = \exp(B)\exp(A)$. This relation simplifies the integration of the trajectories, as, for any switching, $x(t)$ can be shown to be equal to a product of the type

$$x(t) = \exp(At_A)\exp(Bt_B)x(0) \ , \ \text{with } t = t_A + t_B \ .$$

For this system to be stable for any switching strategy, $x(t)$ has to tend to zero, as $t \to \infty$ (whatever the repartition between t_A and t_B). As for discrete time systems, it can be seen that a necessary and sufficient condition for the stability of the system is that both matrices A and B are individually stable (i.e. have negative real part eigenvalues).

From this discussion, it appears that the non-commutativity is a key phenomenon that makes the stability of a switched linear system harder to determine than the stability of individual matrices.

3.5.3 Stability induced by discretization

We can add the following observation to the previous considerations. For any pair of stable matrices (not necessarily a c-stable pair) in continuous time, there exists a discretization step τ^* such that the pair $\{e^{\tau^* A_1}, e^{\tau^* A_2}\}$ is d-stable. Formally stated:

$$\forall A_1^{(c)}, A_2^{(c)} : \Re(\lambda(A_1^{(c)})), \Re(\lambda(A_2^{(c)})) < 0, \ \exists \tau^* : \rho\{e^{\tau^* A_1^{(c)}}, e^{\tau^* A_2^{(c)}}\} < 1 \ .$$

Indeed, as $A_1^{(c)}$ and $A_2^{(c)}$ are stable matrices, the trajectories they each describe in continuous time both converge to the origin. That is :

$$\forall i, \forall \epsilon, \exists \tau_i : \forall \tau > \tau_i, \|e^{\tau A_i}\| < \epsilon \ .$$

So, choosing an $\epsilon < 1$, we get values for τ_1, τ_2. We now choose

$$\tau^* > \max(\tau_1, \tau_2) \ ,$$

so that $\|e^{\tau A_1}\| < \epsilon$ and $\|e^{\tau A_2}\| < \epsilon$, which suffices to prove the d-stability, using the submultiplicativity of the norm. This establishes the following proposition. To formulate it, recall that a matrix is called *Hurwitz* if all its eigenvalues have a negative real part.

Proposition 3.9. *Consider a switched linear system generated by Hurwitz matrices* $\{A_1, \dots, A_m\}$. *There exists a discretization step* τ *that makes the corresponding switched discrete-time system, generated by* $\{e^{\tau A_1}, \dots, e^{\tau A_m}\}$, *stable.*

Now, to go one step further, we wonder if there are time systems that lead to d-stable systems, whatever τ we τ we choose. This is what we present next.

$\{A_1, A_2\}$ c-unstable, $\{e^{\tau A_1}, e^{\tau A_2}\}$ d-stable $\forall \tau$

Consider the pair of matrices we used earlier (from [30]), using a real parameter a, in the range $[1, \infty[$:

$$A_1 = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} -1 & -a \\ 1/a & -1 \end{pmatrix}$$

We already presented the study of the c-stability of the continuous-time switched linear system generated by this pair of matrices, according to the value of a. Defining a worst-case switching strategy, we obtained stability for a range of a including $[1, 10]$. More precisely, it comes out from this analysis that there exists a threshold value (let us call it a^*) such that, for values of $a < a^*$, this worst-case trajectory is a kind of deformed spiral converging to the origin (see Figure 3.7). For values of $a > a^*$, the trajectory is a diverging spiral, moving away from the origin.

We may now consider the specific value a^* of the parameter a such that the trajectory is a closed curve, at the limit case between the converging and diverging spirals (the actual value of a^* is not of much interest here, but from our calculations, a^* should be lying around 11.2). If we now consider the switched dynamical system generated using this value, we get an *unstable* system, in the sense that is does not converge to the origin.

In order to achieve this, the switching strategy (worst-case) is to apply A_1 during a time interval of length T_1 and then A_2 during T_2. These values

T_1 and T_2 are determined by the positions of the "switching lines", shown as dashed lines in figure 3.7. The slopes of these lines are given by $\nu_\pm = \frac{a+1\pm\sqrt{a^2+6a+1}}{2a}$. Note that the values T_1 and T_2 repeat themselves periodically, thanks to the periodicity of the angular velocity (in polar coordinates) of the trajectory (period π).

Let us consider a discretization of the system using the matrix exponential as follows:

$$\begin{aligned} A_1 &\to e^{\tau A_1} \\ A_2 &\to e^{\tau A_2} \ , \end{aligned}$$

where τ is the discretization step. As we have seen before, this parameter plays an essential role in the stability of the so constructed discrete-time system and this is true in this case too. Indeed, in order for its switching to take place precisely on the switching lines (as it does for the continuous-time system), we need to have a value of τ that divides exactly T_1 and T_2 (note : τ does not need to be integer). So, there must exist integers m, n such that

$$\begin{aligned} T_1 &= m\tau \\ T_2 &= n\tau \ . \end{aligned}$$

This is the case if and only if $\frac{T_1}{T_2}$ is rational. Therefore, if this ratio is not rational, then any switching strategy for the discrete-time system will lead to switchings that take place besides the switching lines (or, to say the least, not every time on the switching lines) . This means that the trajectory of this discrete-time system will not remain on the closed curve defined above. And, this curve being the worst case trajectory, the system will inevitably converge to the origin.

But for the considered system, using the value a_*, we do not know if $\frac{T_1}{T_2}$ is rational. This is not a major issue, since we may therefore consider two cases:

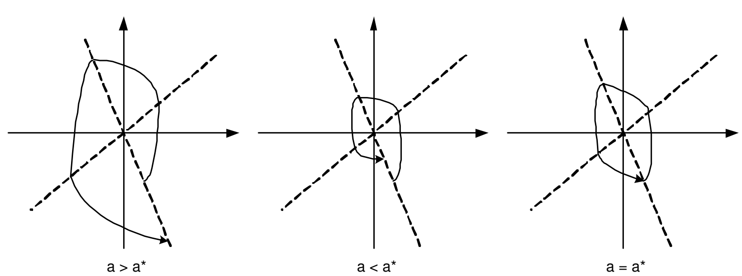

Fig. 3.7. Various behaviors according to the value of the real parameter a. The equations of the dashed lines are $x_2 = \nu_\pm x_1$.

- If $\frac{T_1}{T_2}$ is not rational, we are done, since the above consideration holds.
- If $\frac{T_1}{T_2}$ is rational, then we can consider the system generated by A_1 and $\sqrt{2}A_2$, which admits exactly the same trajectories as (A_1, A_2) in continuous-time, but for which the zones where A_2 is used are gone through at a velocity multiplied by $\sqrt{2}$. The corresponding times T_1 and T_2 are respectively left unchanged and divided by $\sqrt{2}$. This new system has the desired property, since $\frac{T_1}{\sqrt{2}T_2}$ is not rational, because $\frac{T_1}{T_2}$ is supposed to be.

So, in both cases, there exists a pair of matrices that generate a continuous-time switched linear system that is not converging to the origin, while the discretized system converges to the origin, for any value of the discretization step τ.

3.5.4 Summary

At this point, we judge useful to present a summary of what is known so far about the two transformations we have been discussing. We present in Figures 3.8 the possible cases we can encounter when transforming discrete time systems into continuous time ones and vice versa, using the linear fractional transformation. Figure 3.9 shows similar correspondences, but for the matrix exponential.

3.5.5 Other transformations

So, this far, we have not found a function sending every c-stable switched linear system to a d-stable switched linear system and vice versa. The natural question that arises then is to know whether such a function even exists. In order to find other transformations than the two presented before, we have to summarize the properties we want them to have. At first, we may want the function to be analytic. This restriction to analytic functions is motivated by the fact that we want to be able to express this function effectively, without restricting ourselves to trivial families of functions. Asking for an analytic function seems like a reasonable tradeoff for us.

Moreover, this function being analytic, it has a convergent power series development. So, there would be a natural way of determining how the eigenvalues evolve under this transformation as done for equation (3.8), using the spectral mapping theorem:

$$A_i^{(d)} = f(A^{(c)}) = (I - A_i^{(c)})^{-1}(I + A_i^{(c)})$$

$$\lambda_i^{(d)} = f(\lambda_i^{(c)}) = (1 - \lambda_i^{(c)})^{-1}(1 + \lambda_i^{(c)}) \ .$$

This means that the function transforming the matrices does transform their eigenvalues in the same way. As the desired transformation has to be efficient even for a trivial pair of equal matrices $\{A^{(c)}, A^{(c)}\}$, it must send a c-stable matrix on a d-stable matrix. So, in the complex plane, it must apply eigenvalues with a negative real part onto the unit disk, as represented in Figure 3.10. Note that we said *onto* and not *into*, because we want all the points from the left half plane to have an image in the unit disk, and vice versa.

3.5.6 Possible lead for proving the non-existence of the ideal transformation

At the time of writing this thesis, the quest for a *perfect* transformation is still ongoing. By perfect, we mean a transformation that would allow us to switch at will between discrete and continuous systems, while preserving the stability of the corresponding switched systems. Moreover, one could ask for the existence of a CQLF to be preserved as well. Stated formally, we are looking for a transformation $T : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ with T invertible, such that :

$$\begin{aligned} \forall \mathcal{M} \subset \mathbb{R}^{n\times n},\ \mathcal{M} \text{ is c-stable } &\Rightarrow T(\mathcal{M}) \text{ is d-stable },\\ T^{-1}(\mathcal{M}) \text{ is c-stable } &\Leftarrow \mathcal{M} \text{ is d-stable }. \end{aligned}$$

To be precise, when we say $\mathcal{M}$ is c-stable, we mean *the switched linear system in continuous-time generated by the matrices in $\mathcal{M}$ is stable.*

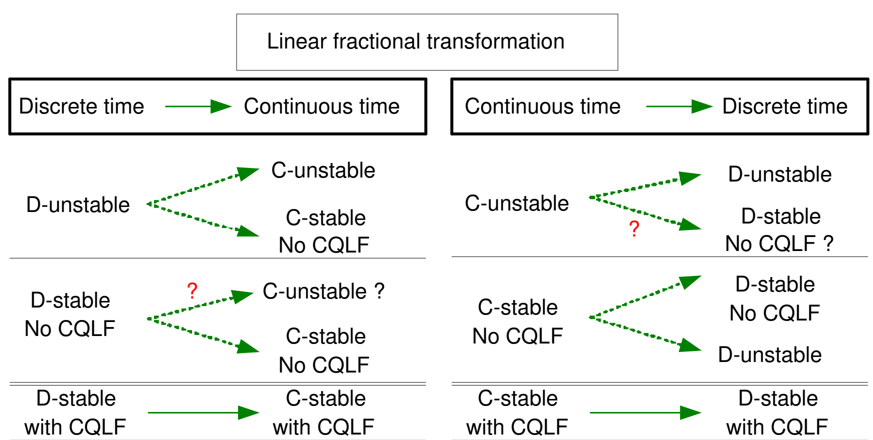

Fig. 3.8. Tables gathering the possible correspondences between continuous and discrete time systems using the linear fractional transformation. Solid arrows show implications that are automatic. On the other hand, dashed arrows express that there are several possible cases (each of which is proved to exist, as there is an example for it in this work). The question marks show the only situation for which we do not have any example, but we have no proof of impossibility either. We can see here that no link exists between the cases with and without CQLF (we know that CQLF are preserved by the linear fractional transformation).

We have not managed to conclude to the existence (by providing the explicit transformation) or the non-existence (by proving it impossible) of such a transformation. However, we here present results that we believe could be of interest in this research. As said in the previous section, we know that the eigenvalues lying inside the unit circle (for discrete time systems) must me mapped onto negative real part eigenvalues (for continuous time systems). This allows us

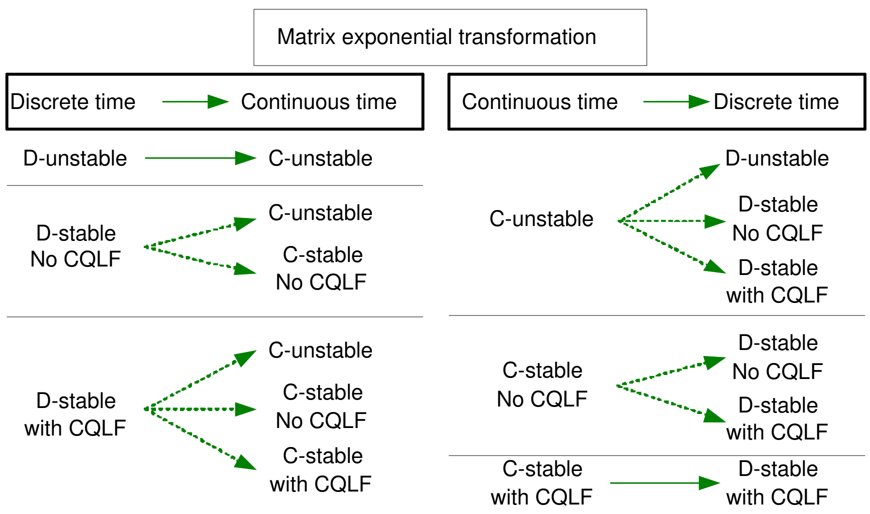

Fig. 3.9. Tables gathering the possible correspondences between continuous and discrete time systems using the matrix exponential transformation. Solid arrows show implications that are automatic. On the other hand, dashed arrows express that there are several possible cases (each of which is proved to exist, as there is an example for it in this work).

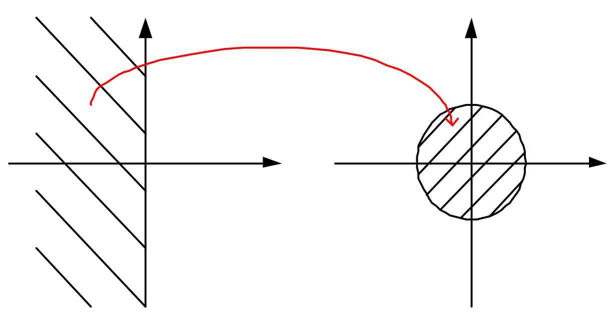

Fig. 3.10. Desired behavior of the wanted transformation. The left half-plane must be mapped inside the unit disk, and the right half-plane outside the unit disk.

to characterize more precisely the general form of the transformation we are looking for. We first present a known result of complex analysis.

Theorem 3.10 ([64]). *Every bijective analytic function preserving the upper half plane has the form*

$$f(z) = \frac{az+b}{cz+d} \ , \ a,b,c,d \in \mathbb{R} \ ,$$

with $ad - bc \geq 0$.

This result enables us to state the following result.

Corollary 3.11. *Every bijective analytic function preserving the left half plane has the form*

$$f(z) = \frac{az+bi}{d-ciz} \ , \ a,b,c,d \in \mathbb{R} \ ,$$

with $ad - bc \geq 0$.

Proof. A transformation preserving the left half plane can be seen as the composition a $-\pi/2$ rotation around the origin, a transformation preserving the upper half plane and a $\pi/2$ rotation around the origin. This is illustrated in Figure 3.11. And, as the $\pi/2$ rotation is simply a multiplication by i, this

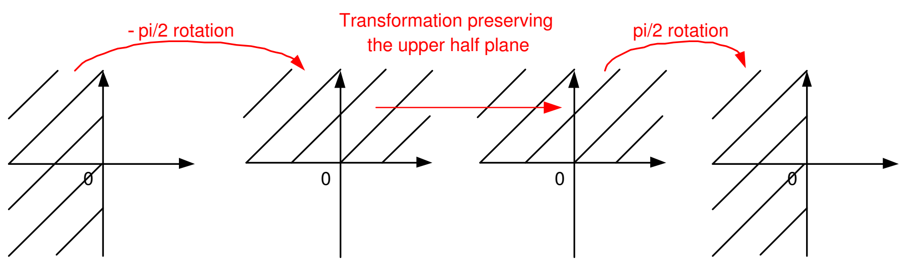

Fig. 3.11. Composition of transformations that gives a transformation applying the right half plane onto itself. We successively carry out a $-\pi/2$ rotation around the origin, a transformation preserving the upper half plane of the type $\frac{az+bi}{d-ciz}$ and a $\pi/2$ rotation around the origin.

composition of transformations is given by

$$f(z) = i \ \frac{a(-iz)+b}{c(-iz)+d} \ ,$$

which immediately yields the result. □

This leads us to the following result.

Proposition 3.12. *Every bijective analytic function mapping the unit circle onto the left half plane has the form*

$$f(z) = \frac{z(a+bi)+(a-bi)}{z(d-ci)-(d+ci)} \ , \ a,b,c,d \in \mathbb{R} \ ,$$

with $ad - bc \geq 0$.

Proof. All the analytic functions mapping the unit circle can be seen as the composition of the function

$$z \to \frac{z+1}{z-1}$$

and a transformation preserving the left half plane, of the type described in Corollary 3.11. The resulting composition has then the following form:

$$f(z) = \frac{a\frac{z+1}{z-1}+bi}{d-ci\frac{z+1}{z-1}} \ ,$$

which we can rewrite as

$$\begin{aligned} f(z) &= \frac{a(z+1)+bi(z-1)}{d(z-1)-ci(z+1)} \\ &= \frac{z(a+bi)+(a-bi)}{z(d-ci)-(d+ci)} \end{aligned}$$

□

To conclude, it should be possible to make use of this specific form to deduce (for example) that there does not exist an analytic transformation with the desired properties.

3.6 Conclusion

This concludes the chapter devoted to the link between the JSR and the stability of switched linear systems. We can now relate the issue of knowing if $\rho(\mathcal{M}) < 1$ to the absolute asymptotic stability of the switched system generated by $\mathcal{M}$. The notion of common quadratic Lyapunov function has also been presented. We saw that it is clearly a sufficient condition for the stability of a DLI, but not a necessary one. This will be used in Chapter 6, to define an approximation of the JSR.

We also studied the correspondence between discrete time and continuous time systems, and we listed the possible stability changes due to the conversion between discrete and continuous. We detailed what kind of transformation between these two time domains we were looking for, without finding a suitable one that would preserve the stability.

4

Computational issues - Complexity

This chapter presents the issues that make the computation of the joint spectral radius a difficult task. First, we present the finiteness conjecture (in short, FC), with a self-consistent counterproof. To be more precise, we should use a different name, such as *finiteness property*, since this is not a conjecture any more. But, as it has often been referred to as *finiteness conjecture*, we will stick to this terminology.

Then, we detail two results originally proved by Kozyakin, the first of which has been re-worked. These state a non-algebraicity property and the existence of an extremal norm. Then, we will end this chapter by unfolding NP-hardness of the computation of the JSR and undecidability results by Blondel & Tsitsiklis.

4.1 The Finiteness Conjecture

4.1.1 Introduction

The Lagarias-Wang finiteness conjecture was introduced in 1995 (in [52]) in connection with problems related to the computation of the joint/generalized spectral radius of finite sets of matrices. Before stating this conjecture, we remind the reader how we defined the GSR $\rho(\mathcal{M})$ of a set $\mathcal{M}$ of matrices in (1.11) as

$$\rho(\mathcal{M}) := \limsup_{k\to\infty} \left(\rho_k(\mathcal{M})\right)^{1/k} ,$$

where

$$\rho_k(\mathcal{M}) := \sup\left\{\rho\left(\prod_{i=1}^{k} M_i\right) : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k\right\} .$$

Conjecture 4.1 (Finiteness conjecture in [52]). Let $\mathcal{M}$ be a finite set of matrices. Then there exists some finite $k \geq 1$ and a matrix product $A = A_1 \dots A_k$ with $A_i \in \mathcal{M}$ such that the generalized spectral radius is achieved by this product:

$$\rho(A)^{1/k} = \rho(\mathcal{M}) \ .$$

This means that, in the four members inequality stated in Lemma 1.9, there is a finite k for which $\rho_k(\mathcal{M})^{1/k} = \rho(\mathcal{M})$.

As this was stated for finite sets of matrices, the joint and generalized spectral radii are identical (see JSR = GSR in Section 1.2.2). So, the notation $\rho(\mathcal{M})$ might as well designate the joint spectral radius in this chapter, without loss of generality. We will keep using the terminology joint spectral radius below.

This conjecture has strong implications on the computation of the joint spectral radius. Indeed, our concern is to know, given a set $\mathcal{M}$, if there exists an effective procedure for deciding whether $\rho(\mathcal{M}) < 1$. If the conjecture is true, then we may consider the following algorithm:

For $k = 1, 2, \ldots$, compute the values $\hat{\rho}_k(\mathcal{M})$ and $\rho_k(\mathcal{M})$:

- if $\rho(\mathcal{M}) < 1$, we know that we will have $\hat{\rho}_k(\mathcal{M}) < 1$ for some k (this will be justified later, in Section 6.1.1),
- if $\rho(\mathcal{M}) \geq 1$, then the FC teaches us that $\exists k^* : \rho_{k^*}(\mathcal{M})^{1/k^*} = \rho(\mathcal{M})$, so that $\rho_{k^*}(\mathcal{M}) \geq 1$.

In both cases, there is a finite value k such that the algorithm stops and provides the correct answer.

We can view the FC in terms of the dynamical system interpretation given in Chapter 3. It means that the convergence to zero of all trajectories associated to periodic products of matrices implies the convergence to zero for all possible products, including the non-periodic ones. This is precisely the implication PAS $\Rightarrow$ AAS, discussed in Sections 3.1.2 and 3.2.2 .

Although we will see that the FC does not hold in general, there are cases where it is valid. Before proceeding, we recall that a *polytope norm* is a norm the unit ball B of which is a polytope. In [42], Gurvits shows conditions under which PAS implies AAS, which are consequently conditions for the finiteness conjecture to hold :

Lemma 4.2 (Gurvits). *For any polytope norm $\|.\|$, if a set $\mathcal{M} = \{A_i : 1 \leq i \leq m\}$ satisfies*

$$\|A_i\| \leq 1, \forall i$$

and DLI($\mathcal{M}$) is PAS, then it is also AAS.

This means that finite sets of matrices admitting a polytope extremal norm do not violate the FC (extremal norms are defined in Section 4.3.1). Lagarias & Wang further generalized this result to *piecewise analytic norms* ([52]): these norms have a unit ball B the boundary of which is contained in the zero set of a holomorphic function f defined on an open set $\Omega \subseteq \mathbb{C}^n$ containing B, such that $f(0) \neq 0$.

However, in order to make use of these results, we need to know whether an arbitrary set $\mathcal{M}$ admits such a specific extremal norm or not. It can be seen in Section 4.3 that bounded irreducible sets of matrices always admit extremal norms, yet nothing proves that polytope or piecewise analytic norms are always possible.

Despite what has been believed for some time after the publication of the above results, the FC is not valid in general. It has been proved to be false in [19] by Bousch & Mairesse. The existence of a counterexample is proved therein by using iterated function systems, topical maps and Sturmian sequences. The proof relies in part on a particular fixed point theorem known as Mañé's lemma, which is not relevant to detail here. In this thesis, we provide an alternative proof, in which we show that there are uncountably many values of the real parameter α for which the pair of matrices

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \alpha \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\} \tag{4.1}$$

does not satisfy the finiteness conjecture. This was published in [11] and presented in [10]. Our proof is not constructive as we do not exhibit any particular value of α for which the corresponding pair of matrices violates the finiteness conjecture. The problem of finding an explicit counterexample and the problem of determining if there even exist matrices with rational entries that violate the conjecture remain open questions. As compared to the proof in [19], our proof has the advantage of being self-contained and fairly elementary, as it uses only elementary facts from linear algebra.

4.1.2 Proof outline

We first briefly outline our proof. We define

$$A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} ,$$

and

$$A_0^\alpha = \frac{1}{\rho_\alpha} A_0, \quad A_1^\alpha = \frac{\alpha}{\rho_\alpha} A_1 ,$$

with $\rho_\alpha = \rho(\{A_0, \alpha A_1\})$. Since $\rho(\lambda \mathcal{M}) = |\lambda|\, \rho(\mathcal{M})$, the joint spectral radius of the set $\mathcal{M}_\alpha = \{A_0^\alpha, A_1^\alpha\}$ is equal to one by construction. Let $I = \{0, 1\}$ be a two letter alphabet and let $I^+ = \{0, 1, 00, 01, 10, 11, 000, \ldots\}$ be the set of finite nonempty words. We will also denote the empty word by $\emptyset$ and use the notation $I^* = I^+ \cup \{\emptyset\}$. To the word $w = w_1 \ldots w_t \in I^+$ we associate the products $A_w = A_{w_1} \ldots A_{w_t}$ and $A_w^\alpha = A_{w_1}^\alpha \ldots A_{w_t}^\alpha$. From the above definitions, we have:

$$\forall \alpha, \forall w \in I^+, \rho(A_w^\alpha) \le 1 .$$

A word $w \in I^+$ will be said *optimal* for some α if $\rho(A_w^\alpha) = 1$. We use J_w to denote the set of α's for which $w \in I^+$ is optimal. If the finiteness conjecture is true, then for any α, there exists a finite word that is optimal, and so the union of the sets J_w for $w \in I^+$ covers the real line. We will show that this union does not cover the interval $[0, 1]$, proving that the finiteness conjecture does not hold. This summarizes the proof.

The next sections are organized as follows. In Section 4.1.4, we show that if two words $u, v \in I^+$ are essentially equal, then $J_u = J_v$. Two words $u, v \in I^+$ are *essentially equal* if the periodic infinite words $U = uu\ldots$ and $V = vv\ldots$ can be decomposed as $U = xww\ldots$ and $V = yww\ldots$ for some $x, y, w \in I^+$. Words that are not essentially equal are *essentially different*. Obviously, if u and v are essentially different, then so are cyclic permutations of u and v. We show in the same section that the sets J_u and J_v are disjoint if u and v are essentially different. This part of the proof requires some properties of infinite words presented in Section 4.1.3. The proof is then almost complete. To conclude, we observe in Section 4.1.5 that the sets $J_w \cap [0, 1]$ are closed sub-intervals of $[0, 1]$. There are countably many words in I^+ and so

$$\cup_{w \in I^+} (J_w \cap [0, 1])$$

is a countable union of disjoint closed sub-intervals of $[0, 1]$. Except for a trivial case that we can exclude here, there are always uncountably many points in $[0, 1]$ that do not belong to such a countable union. Each of these points provides a particular counterexample to the finiteness conjecture.

4.1.3 Palindromes in infinite words

The *length* of a word $w = w_1 \ldots w_t \in I^*$ is equal to $t \geq 0$ and is denoted by $|w|$. The *mirror image* of w is the word $\tilde{w} = w_t \ldots w_1 \in I^*$. A *palindrome* is a word in I^* that is identical to its mirror image. In particular, the empty word is a palindrome. For $u, v \in I^*$, we write $u > v$ if u is lexicographically larger than v, that is, $u_i = 1$, $v_i = 0$ for some $i \geq 1$ and $u_j = v_j$ for all $j < i$. This is only a partial order since words of different lengths are not comparable, as 101000 and 1010, for example.

Given a finite word $w \in I^+$, a factor f of w is a word $f \in I^+$ such that there exist $x, y \in I^*$ (x and/or y can be $\emptyset$) such that $w = xfy$. This definition extends naturally to infinite words : if w is infinite ($\in I^\infty$), then $f \in I^+$ a finite factor of w if there exist $x \in I^*$ and $y \in I^\infty$ such that $w = xfy$. For an infinite word U, we denote by $F(U)$ the set of all finite factors of U.

Lemma 4.3. *Let $u, v \in I^+$ be two words that are essentially different. We denote $U = uuu\ldots$ and $V = vvv\ldots$ their periodic repetitions. Then there exists a pair of words $0p0$ and $1p1$ in the set $F(U) \cup F(V)$ such that $p \in I^*$ is a palindrome.*

Proof. Let m and n be the minimal periods of U and V respectively. The values of m and n are invariant under cyclic permutations of u and v. Let us use induction on $m+n$. The result is obvious for $m+n=2$ since in this case U and V must be equal to $111\ldots$ and $000\ldots$ and we may then take $p=\emptyset$. Consider now $u,v \in I^+$ with $m+n>2$. If the words 00 and 11 both belong to the set $F(U) \cup F(V)$ then we can set $p=\emptyset$. So, assume without loss of generality that 11 does not belong to $F(U) \cup F(V)$. We may also assume that both u and v begin with 0; otherwise, we can take appropriate cyclic permutations of u and v. At this point, u and v can be factorized in a unique way by factors $0' = 0$ and $1' = 01$.

In the new alphabet $\{0', 1'\}$, the resulting words u' and v' are still essentially different and the minimal periods m' and n' of $U' = u'u' \ldots$ and $V' = v'v' \ldots$ satisfy $m'+n' < m+n$. By the induction hypothesis, the property is satisfied for $m'+n'$ and there exists a pair of words $0'q'0'$ and $1'q'1'$ in $F(U') \cup F(V')$ such that $q' = \tilde{q'}$. Let q be the word obtained from q' by replacing $0'$ by 0 and $1'$ by 01. From $1'q'1' \in F(U') \cup F(V')$ we get

$$01q01 \in F(U) \cup F(V) \ .$$

And since $0'q'0' \in F(U') \cup F(V')$ we have either $0'q'0'0' \in F(U') \cup F(V')$ or $0'q'0'1' \in F(U') \cup F(V')$, and both cases imply that

$$0q00 \in F(U) \cup F(V) \ .$$

We can now define $p = q0$ and observe that $0p0, 1p1 \in F(U) \cup F(V)$.

To conclude, we have to show that if q' is a palindrome in $\{0', 1'\}$, then $q0$ is a palindrome in $\{0, 1\}$. Again, we use induction, this time on $|q'|$. For $|q'| = 0, 1$ the statement is obviously true. Suppose that $|q'| \geq 2$. Then $q' = 0's'0'$ or $q' = 1's'1'$ for $s' \in \{0', 1'\}^*$, and s' is a palindrome in $\{0', 1'\}$. We then have that either $q0 = 0s00$ or $q0 = 01s010$. By the induction hypothesis, as $|s'| < |q'|$, $s0$ is a palindrome in $\{0, 1\}$. It follows that $q0$ is also a palindrome. $\square$

Corollary 4.4. *Let $u, v \in I^+$ be two essentially different words and let $U = uuu\ldots$ and $V = vvv\ldots$. Then there exist words $a, b, x, y \in I^+$ satisfying $|x| = |y|$, $x > y$, $\tilde{x} > \tilde{y}$, $x > \tilde{y}$, $\tilde{x} > y$, and a palindrome $p \in I^*$ such that*

$$U = apxpxp\ldots \quad \textit{and} \quad V = bpypyp\ldots \ ,$$

or one of the words U and V, say U, can be decomposed as

$$U = apxpxp\cdots = bpypyp\ldots \ .$$

Proof. By Lemma 4.3, there exists a pair of words $0p0$ and $1p1$ in the set $F(U) \cup F(V)$ such that p is a palindrome. Without loss of generality, assume that $1p1$ occurs in U. Then it occurs in U infinitely many times because U is periodic. Let us write

$$U = a'1p1d1p1d\ldots$$

and, analogously,

$$W = b'0p0f0p0f\ldots,$$

where W is either U or V. Without loss of generality we may assume $|d| = |f|$; otherwise, we can always take $d' = d1p1d\ldots1p1d$ instead of d and $f' = f0p0f\ldots0p0f$ instead of f in such a way that $|f'| = |d'|$. It remains to set $a := a'1$, $b := b'0$, $x := 1d1$, and $y := 0f0$. These clearly satisfy the required properties, as $1d1 > 0f0$. □

4.1.4 Optimal words are essentially equal

For a given word $w \in I^+$ we define $J_w = \{\alpha : \rho(A_w^\alpha) = 1\}$. This is, for a given word w, the set of values of α such that w if optimal for the pair $\{A_0^\alpha, A_1^\alpha\}$. Our goal in this section is to prove that J_u and J_v are equal when u and v are essentially equal, and have otherwise empty intersection.

Lemma 4.5. *Let $u, v \in I^+$ be two essentially equal words. Then $J_u = J_v$.*

Proof. Assume $u, v \in I^+$ are essentially equal. Then $U = uu\ldots$ and $V = vv\ldots$ can be written as $U = ss\ldots$ and $V = tt\ldots$ with $|s| = |t|$ and t a cyclic permutation of s. The spectral radius satisfies $\rho(AB) = \rho(BA)$ and so the spectral radius of a product of matrices is invariant under cyclic permutations of the product factors. From this it follows that $\rho(A_s^\alpha) = \rho(A_t^\alpha)$ and hence u is optimal whenever v is. □

We now set ourselves to prove that essentially different words induce disjoint intervals J. We need two preliminary lemmas before proving this result.

Lemma 4.6. *For any word $w \in I^+$ we have*

$$A_{\tilde{w}} - A_w = k(w)T,$$

where $k(w)$ is an integer and

$$T = A_0A_1 - A_1A_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Moreover, $k(w)$ is positive if and only if $w > \tilde{w}$.

Proof. Let us prove by induction that

$$A_w = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A_{\tilde{w}} = \begin{pmatrix} d & b \\ c & a \end{pmatrix}.$$

First, this is trivial for $w = 0$ and $w = 1$. Notice also that

$$A_0 \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ c & d \end{pmatrix}$$

and

$$\begin{pmatrix} d & b \\ c & a \end{pmatrix} A_0 = \begin{pmatrix} d & b+d \\ c & a+c \end{pmatrix},$$

so that

$$A_0 \begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} d & b \\ c & a \end{pmatrix} A_0 = \begin{pmatrix} (a+c)-d & 0 \\ 0 & d-(a+c) \end{pmatrix} .$$

Identical computations for A_1 yield a similar result. From this it follows that $A_{\tilde{w}} - A_w = k(w)T$. Now, to determine the sign of $k(w)$, we see that

$$A_0 \begin{pmatrix} a & b \\ c & d \end{pmatrix} A_1 - A_1 \begin{pmatrix} d & b \\ c & a \end{pmatrix} A_0 = \begin{pmatrix} a+b+c & 0 \\ 0 & -(a+b+c) \end{pmatrix} .$$

So, $\forall w, A_{0w1} - A_{1\tilde{w}0} = (a+b+c)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Moreover, the simultaneous left and right multiplication of this relation by either A_0 or A_1 does not change the right member. Consequently, any $w > \tilde{w}$ will give a positive $k(w)$. $\square$

We say that a matrix A *dominates* B if $A \geq B$ componentwise and $\mathrm{tr}A > \mathrm{tr}B$ (tr denotes the trace). And it is quick to check that the eigenvalues of a 2×2 matrix A are given by

$$\frac{\mathrm{tr}A \pm \sqrt{(\mathrm{tr}A)^2 - 4\det A}}{2} .$$

For all words w, the matrix A_w satisfies $\det(A_w) = 1$ and $\mathrm{tr}(A_w) \geq 2$, ensuring that the argument of the square root is never negative. The spectral radius is obviously obtained by choosing the plus sign in the above expression. We therefore have $\rho(A_u) > \rho(A_v)$ whenever A_u dominates A_v.

Lemma 4.7. *For any word of the form $w = psq$, where $s > \tilde{s}$ and $q < \tilde{p}$, the matrix $A_{w'}$ with $w' = p\tilde{s}q$ dominates A_w.*

Proof. We have

$$A_{w'} - A_w = A_p(A_{\tilde{s}} - A_s)A_q = k(s)A_pTA_q ,$$

with $k(s) > 0$ because $s > \tilde{s}$. We now consider the product A_pTA_q. We see that, as $\tilde{p} > q$, these two words may share their common first digits before $\tilde{p}$ contains a 1 where q contains a 0. So, we can have products of the types

$$\ldots A_0TA_0 \ldots \text{ or } \ldots A_1TA_1 \ldots .$$

These do not interfere, since $A_i T A_i = T$, for $i = 0, 1$. We then necessarily reach a product of the type

$$\dots A_1 T A_0 \dots = \dots \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \dots .$$

The dots here stand for the remaining A_0's and A_1's. These will not affect the resulting product, as they only contain 0's and 1's, and they will preserve the strict positivity of the first diagonal element. This finishes the proof, as $A_{w'} \geq A_w$ componentwise, the first diagonal element being strictly larger. $\square$

Let $w = psq$. If $s > \tilde{s}$ and $q < \tilde{p}$, we say that $s \to \tilde{s}$ is a *dominating flip*. We are now ready to prove the main result of this section.

Lemma 4.8. *Let $u, v \in I^+$ be two essentially different words. Then $J_u \cap J_v = \emptyset$.*

Proof. Let $u, v \in I^+$ be two words that are essentially different. We assume without loss of generality that neither $U = uu\dots$ nor $V = vv\dots$ is equal to $00\dots$ or $11\dots$ because $11\dots$ is not optimal for any $\alpha \in [0, 1]$ and $00\dots$ is only optimal for $\alpha = 0$, but no other word is optimal for $\alpha = 0$. In order to prove the result, we will show that, if such words u and v satisfy $\rho(A_u^\alpha) = \rho(A_v^\alpha)$ for some value of α, then there exists a third word w satisfying $\rho(A_w^\alpha) > \rho(A_u^\alpha)$. And so, neither u nor v is actually optimal.

By Corollary 4.4, there exist words $a, b, x, y \in I^+$ satisfying $|x| = |y|$, $x > y$, $\tilde{x} > \tilde{y}, x > \tilde{y}, \tilde{x} > y$, and a palindrome $p \in I^*$ such that

$$U = apxpxp\dots \quad \text{and} \quad V = bpypyp\dots$$

or

$$U = apxpxp\dots = bpypyp\dots .$$

Since neither U nor V is equal to $00\dots$ or $11\dots$, the matrices A_{xp} and A_{yp} are strictly positive, as they both contain A_0 and A_1.

Let us consider the word $xpxpxpypypyp$. Setting $s = xpy$, we make a dominating flip in this word and get the word $xpxp\tilde{y}p\tilde{x}pypyp$. Then we set $s = xp\tilde{y}p\tilde{x}py$ and make another dominating flip. As a result, we get the word $xp\tilde{y}pxpyp\tilde{x}pyp$. The matrix $A_{xpxpxpypypyp}$ is dominated by the matrix $A_{xp\tilde{y}pxpyp\tilde{x}pyp}$. Analogously, any matrix $A_s A_v A_r$, $v \in I^*$, is dominated by the matrix $A_{s'} A_v A_{r'}$ where $s = xpxpxp$, $r = ypypyp$, $s' = xp\tilde{y}pxp$, and $r' = yp\tilde{x}pyp$.

Let us denote the linear operators $A \to A_s A A_r$ and $A \to A_{s'} A A_{r'}$ acting in $\mathbb{R}^4$, as well as their (4×4) matrices, by L and L', respectively. It is known that $L = A_r^T \otimes A_s$ and $L' = (A'_r)^T \otimes A'_s$, where $\otimes$ is used to denote the Kronecker (tensor) product (see e.g. [46] Lemma 4.3.1). As a reminder, for $A, B \in \mathbb{R}^{2\times 2}$:

$$A \otimes B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix} \in \mathbb{R}^{4\times 4} .$$

As $A_r, A_s, A_{r'}, A_{s'}$ are strictly positive, both L and L' are also strictly positive. The minimal closed convex cone in $\mathbb{R}^4$ containing all matrices A_v, $v \in I^*$, is the cone of all nonnegative 2×2-matrices. Indeed, any nonnegative matrix X with $\det(X) = 0$ can be approximated by matrices of the form βA_w, $\beta > 0$, $w \in I^*$. In particular, this is true for the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Hence $L' \geq L$ elementwise and $L \neq L'$. These considerations allow us to use the Perron-Frobenius theory (see, for instance, Problem 8.15 in [45]) to get $\rho(L') > \rho(L)$. Moreover, the spectral radius of a Kronecker product is known to be the product of the spectral radii ([46], Theorem 4.2.12), and so

$$\rho(L') = \rho(A_{s'})\rho(A_{r'}) > \rho(A_s)\rho(A_r) = \rho(L) \ .$$

Since the performed flips do not change the average proportion of matrices A_0 and A_1 in the product, we can also write, noting the dependance on α (it is not an exponent):

$$\rho(L^\alpha) = \rho(A_s^\alpha)\rho(A_r^\alpha) \quad \text{and} \quad \rho(L'^\alpha) = \rho(A_{s'}^\alpha)\rho(A_{r'}^\alpha)$$

for each $\alpha > 0$, where L^α and L'^α are defined analogously to L and L'. Suppose now that both u and v are optimal: $\rho(A_u^\alpha) = \rho(A_v^\alpha) = 1$. Then $\rho(A_s^\alpha) = \rho(A_r^\alpha) = 1$, because they are periodic patterns of respectively U and V (or both patterns of U). Hence, $\rho(L^\alpha) = 1$ and consequently $\rho(L'^\alpha) > 1$. Then, either $\rho(A_{s'}^\alpha) > 1$ or $\rho(A_{r'}^\alpha) > 1$, which is a contradiction. Two essentially different words cannot be optimal for the same α. □

4.1.5 Finiteness conjecture

We are now ready to prove the main result.

Theorem 4.9. *There are uncountably many values of the real parameter α for which the pair of matrices*

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ \alpha \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\}$$

does not satisfy the finiteness conjecture.

Proof. It is clearly equivalent to prove the statement for the matrices A_0^α and A_1^α. For $\alpha = 0$, all optimal words w are essentially equal to 0. For any other word w, the set $J_w \cap [0,1]$ can be written as

$$J_w \cap [0,1] = \{\alpha \in (0,1] : \rho(A_w^\alpha) = 1\}$$

or, equivalently,

$$J_w \cap [0,1] = \left\{ \alpha \in (0,1] : \left(\rho(A_w)\alpha^{|w|_1}\right)^{\frac{1}{|w|}} = \sup_{v \in I^+} \left(\rho(A_v)\alpha^{|v|_1}\right)^{\frac{1}{|v|}} \right\} . \quad (4.2)$$

In this expression, $|w|_1$ denotes the number of occurrence of 1's in the word w. Associated to $w \in I^+$ we define the affine function

$$h_w(\beta) = \frac{1}{|w|}(\ln \rho(A_w) + |w|_1 \beta) ,$$

and let

$$h(\beta) = \sup_{w \in I^+} h_w(\beta).$$

Passing to the logarithmic scale in the expression (4.2), we identify the h functions to get:

$$J_w \cap [0,1] = \{e^\beta : \beta \in \mathbb{R}, h_w(\beta) = h(\beta)\} \cap [0,1]. \quad (4.3)$$

The functions h_w are affine, h is convex and continuous, and $h(\beta) \geq h_w(\beta)$ for all $w \in I^+$ and $\beta \in \mathbb{R}$. From this, it follows that the set $\{\beta \in \mathbb{R} : h_w(\beta) = h(\beta)\}$ is an interval of the real line. This interval is the zero set of a continuous function and it is therefore closed. From (4.3) we conclude that $J_w \cap [0,1]$ is a closed subinterval of $[0,1]$.

We conclude this proof using a simple argument : there is a countable infinite number of finite words, while the number of disjoint closed intervals needed to pave the real interval $[0,1]$ is uncountable (unless this is a single interval which is, obviously, not the case here).

□

Our proof is not constructive in the sense that we do not exhibit any explicit value of α. The problem of determining if there is a pair of *rational* matrices that do not satisfy the finiteness conjecture therefore remains open.

We believe that we could be facing a situation where nearly all values of the parameter α lead to a pair satisfying the FC. More formally, when picking a value α at random in $[0,1]$, the probability of the generated pair to satisfy the FC seems to be equal to 1. From a different perspective, the set of values α_{sat} satisfying the FC in $[0,1]$ would be of measure 1. And the set of values α_{unsat} violating the FC in $[0,1]$ would be of measure zero. This raises the question to know whether, in general, the sets of m matrices satisfying the FC are dense in $\mathbb{R}^{mn^2}$. This aspect is being studied at the time of writing this thesis and will hopefully be the subject of a future publication.

4.2 Non-algebraicity

In [51], Kozyakin shows that there does not exist any algebraic criterion of absolute stability for switched linear systems. We have to mention that his initial proof contains a flaw, and so cannot be admitted as is. After e-mail contacts, we learned that the author was aware of the mistake and had published a correction somewhere else, but only in Russian. Consequently, we feel the need to present our own corrected version here. The same author also proves the existence of an extremal norm for switched linear systems, by giving a construction for it. We detail here these results. For the definitions of the notions of stability used here, we refer the reader to Section 3.1.1.

Non-algebraicity property

Any set composed of m real $(n \times n)$ matrices can be seen as a point in the space $\mathbb{R}^{mn^2}$. Therefore, a subset of $\mathbb{R}^{mn^2}$ is a set of m-uples of matrices. We define two particular sets of this kind:

- $S(m, n)$ is the set of all m-uples of matrices generating an absolutely stable DLI.
- $E(m, n)$ is the set of all m-uples of matrices generating an absolutely exponentially stable DLI.

Analyzing the stability of a DLI can then be regarded as finding satisfactory characterizations of the sets $S(m, n)$ and/or $E(m, n)$, depending on the type of stability. Their structure should yield criteria of absolute (exponential) stability. To this aim, a property that we would like this structure to have is *semi-algebraicity.*

Definition 4.10. *We denote by $x = (x_1, \ldots, x_L)$ an element of $\mathbb{R}^L$. We call a polynomial of the variable x a finite sum of the form $p(x) = \sum p_{i_1 \ldots i_L} x_1^{i_1} \ldots x_L^{i_L}$, where the $p_{i_1 \ldots i_L}$ are numerical coefficients. A set U has the* SA-property *if there exist finitely many polynomials $p_1(x), \ldots, p_k(x), p_{k+1}(x), \ldots, p_{k+l}(x)$ such that U is equal to the set of elements x defined by the conditions*

$$p_1(x) > 0, \ \ldots, \ p_k(x) > 0$$
$$p_{k+1}(x) = 0, \ \ldots, \ p_{k+l}(x) = 0 \ .$$

In short, U has to be definable by a finite number of polynomial constraints. Finally, a set is semi-algebraic *if it is a union of a finite number of set having the SA-property.*

The semi-algebraicity of a set implies the existence of a criterion allowing to decide whether an element belongs to this set, using finitely many arithmetic operations. This notion is an extension of the more restrictive *algebraicity.* From the following definition, an algebraic set is also semi-algebraic.

Definition 4.11. *A set U is* algebraic *if there exist finitely many polynomials $p_1(x), \ldots, p_k(x)$ such that U is equal to the set of elements x defined by the conditions*

$$p_1(x) = 0, \ \ldots, \ p_k(x) = 0 \ .$$

In short, U has to be the set of common zeros of a finite number of polynomials.

4.2.1 The result

In the specific case $m = 1$, the sets E and S contain single matrices as elements. And the stability of a single matrix is straightforward, in the sense that the eigenvalues of a matrix can be determined efficiently. So, it is possible to decide whether a given matrix singleton $\{A\}$ belongs to $S(1, n)$ and/or to $E(1, n)$. The following result deals with the general case $m \geq 2$.

Theorem 4.12 (Kozyakin). *Consider $m, n \geq 2$. If a subset U of the space $\mathbb{R}^{mn^2}$ satisfies the condition*

$$E(m, n) \subseteq U \subseteq S(m, n),$$

then it is not semi-algebraic.

More specifically, the subsets $E(m, n)$ and $S(m, n)$ of the space $\mathbb{R}^{mn^2}$ are not semi-algebraic.

We are mainly interested in the last claim, about $E(m, n)$ and $S(m, n)$. It teaches us that, in the general case, given a discrete linear inclusion, there is no procedure involving a finite number of operations that allows to decide whether it is stable or not. This shows that the stability of a DLI is hard to determine. We now detail the proof(s) of this theorem.

Proof (main idea presented in [51]). The proof is carried out for the case $m = n = 2$, because a proof for this case implies the result for all cases with larger m or n. Two families of matrices $G(t)$ and $H(t)$ depending on a real parameter $t \in [-1, 1]$ are defined. Kozyakin proposes to use

$$G(t) = (1 - t^4) \begin{pmatrix} 1 & -\frac{t}{\sqrt{1-t^2}} \\ 0 & 0 \end{pmatrix}, \ H(t) = (1 - t^4) \begin{pmatrix} \sqrt{1 - t^2} & -t \\ t & \sqrt{1 - t^2} \end{pmatrix} \ .$$

The set W of all pairs of the type $\{G(t), H(t)\}$ is an algebraic set of $\mathbb{R}(2, 2)$. Now suppose we have a semi-algebraic set U satisfying $E(m, n) \subseteq U \subseteq S(m, n)$. Then, $U \cap W$ is also semi-algebraic. It is then made use of a result by Whitney (referring to [61]) implying that a neighborhood of the pair $\{G(0), H(0)\}$ in $U \cap W$ must be either empty, or consist of finitely many connectivity components (this notion will be made clearer further). But the chosen $G(t)$ and $H(t)$ provide infinitely many connectivity components, thus proving false the initial assumption (the semi-algebraicity of U). $\square$

Pointing out the issue

We now give the details from the initial proof and show the flaw that needs a correction. The two proposed families of matrices are:

$$G(t) = (1-t^4)\begin{pmatrix} 1 & -\frac{t}{\sqrt{1-t^2}} \\ 0 & 0 \end{pmatrix},\ H(t) = (1-t^4)\begin{pmatrix} \sqrt{1-t^2} & -t \\ t & \sqrt{1-t^2} \end{pmatrix} .$$

One can see that the matrices $G(t)$ are projection matrices, while $H(t)$ are rotation matrices. The point of the proof is that:

- when $t_n = \sin(\frac{\pi}{n})$, the system generated by the pair $\{G(t_n), H(t_n)\}$ is unstable for n sufficiently large.
- when $t'_n = \sin(\frac{\pi}{2n+1})$, the system generated by the pair $\{G(t'_n), H(t'_n)\}$ is stable.

The two points imply that in any neighborhood of $t = 0$, there are infinitely many stable and unstable pairs of matrices $\{G(t), H(t)\}$. And as the t_n and t'_n are alternating, $U \cap W$ contains infinitely many distinct connectivity components, which contradicts the semi-algebraicity of U.

We reprove here the first point, but we disprove the second one. So, both cases generate unstable systems and the whole proof found in the appendix of [51] has to be rethought. Taking back the original construction, let us consider the following product for the first point, with $t_n = \sin(\frac{\pi}{n})$:

$$\begin{aligned} G(t_n)H(t_n)^{n-1} &= (1-t_n^4)\begin{pmatrix} 1 & -\frac{t_n}{\sqrt{1-t_n^2}} \\ 0 & 0 \end{pmatrix}(1-t_n^4)^{n-1}\begin{pmatrix} \sqrt{1-t_n^2} & -t_n \\ t_n & \sqrt{1-t_n^2} \end{pmatrix}^{n-1} \\ &= \frac{(1-t_n^4)^n}{\sqrt{1-t_n^2}}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \sqrt{1-t_n^2} & -t_n \\ t_n & \sqrt{1-t_n^2} \end{pmatrix}\begin{pmatrix} \sqrt{1-t_n^2} & -t_n \\ t_n & \sqrt{1-t_n^2} \end{pmatrix}^{n-1} \\ &= \frac{(1-t_n^4)^n}{\sqrt{1-t_n^2}}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} . \end{aligned}$$

It remains to study the behavior of the factor $\frac{(1-t_n^4)^n}{\sqrt{1-t_n^2}}$ for $n \to \infty$. It can effectively be shown to be > 1 for a sufficiently large n. Indeed, as $n \to \infty$, $t_n \to 0$, and

$$\lim_{t_n \to 0} \frac{(1-t_n^4)^n}{\sqrt{1-t_n^2}} = \lim_{t_n \to 0} \frac{1-nt_n^4}{1-\frac{1}{2}t_n^2} = 1 . \tag{4.4}$$

We now know this factor tends to 1 as $n \to \infty$. Now we want to prove that it tends to 1 from above, that is

$$\exists N : \forall n \geq N, \frac{(1-t_n^4)^n}{\sqrt{1-t_n^2}} > 1 .$$

Using (4.4) and

$$t_n = \sin(\frac{\pi}{n}) \approx \frac{\pi}{n}, \text{ as } n \to \infty \ ,$$

we can verify that $1 - nt_n^4 > 1 - \frac{1}{2}t_n^2$, for sufficiently large n. Indeed, this last inequality becomes, for $n \to \infty$ ($t_n \to 0$),

$$nt_n^4 < \frac{1}{2}t_n^2 \ .$$

This is straightforward from $nt_n^4 \approx \frac{\pi^4}{n^3}$ and $\frac{1}{2}t_n^2 \approx \frac{\pi^2}{2n^2}$, as $\frac{1}{n^3} < \frac{1}{n^2}$, for large n. This shows that, for the pairs of the type $\{G(t_n), H(t_n)\}$ with n large enough, there exists a diverging product.

Regarding the second point, when $t'_n = \sin(\frac{2\pi}{2n+1})$, nearly the same can be said. In this case, we use the following product:

$$\begin{aligned} G(t'_n)H(t'_n)^{2n} &= (1 - t'^4_n)\begin{pmatrix} 1 & -\frac{t'_n}{\sqrt{1-t'^2_n}} \\ 0 & 0 \end{pmatrix}(1 - t'^4_n)^{2n}\begin{pmatrix} \sqrt{1-t'^2_n} & -t'_n \\ t'_n & \sqrt{1-t'^2_n} \end{pmatrix}^{2n} \\ &= \frac{(1 - t'^4_n)^{2n+1}}{\sqrt{1-t'^2_n}}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \sqrt{1-t'^2_n} & -t'_n \\ t'_n & \sqrt{1-t'^2_n} \end{pmatrix}\begin{pmatrix} \sqrt{1-t'^2_n} & -t'_n \\ t'_n & \sqrt{1-t'^2_n} \end{pmatrix}^{2n} \\ &= \frac{(1 - t'^4_n)^{2n+1}}{\sqrt{1-t'^2_n}}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ . \end{aligned}$$

Here too, it remains to study the behavior of a scalar factor: $\frac{(1-t'^4_n)^{2n+1}}{\sqrt{1-t'^2_n}}$, which can actually also be shown to be > 1 for a sufficiently large n. This is in contradiction with the stability that was claimed. To prove the divergence, similarly as before,

$$\lim_{n\to\infty} \frac{(1 - t'^4_n)^{2n+1}}{\sqrt{1-t'^2_n}} \sim \lim_{n\to\infty} \frac{1 - (2n+1)t'^4_n}{1 - \frac{1}{2}t'^2_n} = 1 \ . \tag{4.5}$$

But, for similar reasons as in the first case, $(2n+1)t'^4_n < \frac{1}{2}t'^2_n$ for $t'_n \to 0$, so $1 - (2n+1)t'^4_n > 1 - \frac{1}{2}t'^2_n$ and the factor is > 1, as in the previous case. So, this value also tends to 1, and also from above. Figure 4.1 provides an illustration of the behavior of both factors.

Origin of the mistake

The above observations contradict Lemma 4 of the Appendix of [51], which is itself induced by Lemma 3 and Lemma 2 therein. A thorough analysis effectively shows us a mistake in Lemma 2. Let us detail this. The author introduces families of matrices of the same nature as those presented above, using a different parametrization, with $t = \sin(\phi)$:

$$P(\phi) = \begin{pmatrix} 1 & -\tan(\phi) \\ 0 & 0 \end{pmatrix} , \; R(\phi) = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} .$$

As was G before, P is a projection matrix, while R is a rotation matrix as was H (hence the choice of the letters). The following equation is then used. Note that this equation is not proved in the paper, but it can quickly be done just by developing it explicitly:

$$(PR^m P)(\phi) = \frac{cos[(m+1)\phi]}{cos(\phi)} P(\phi) . \tag{4.6}$$

This Lemma 2 says:

Lemma 4.13 (Lemma 2 from [51] - false). *Let* $\phi = \frac{2\pi}{2n+1}$. *Then*

$$(PR^m P)(\phi) = \lambda_{m,n} P(\phi) ,$$

where $\lambda_{m+n,n} = \lambda_{m,n}, |\lambda_{m,n}| \leq 1$.

Using equation (4.6), we can indeed write

$$(PR^m P)\left(\frac{2\pi}{2n+1}\right) = \frac{cos\left[(m+1)\frac{2\pi}{2n+1}\right]}{cos\left(\frac{2\pi}{2n+1}\right)} P\left(\frac{2\pi}{2n+1}\right) .$$

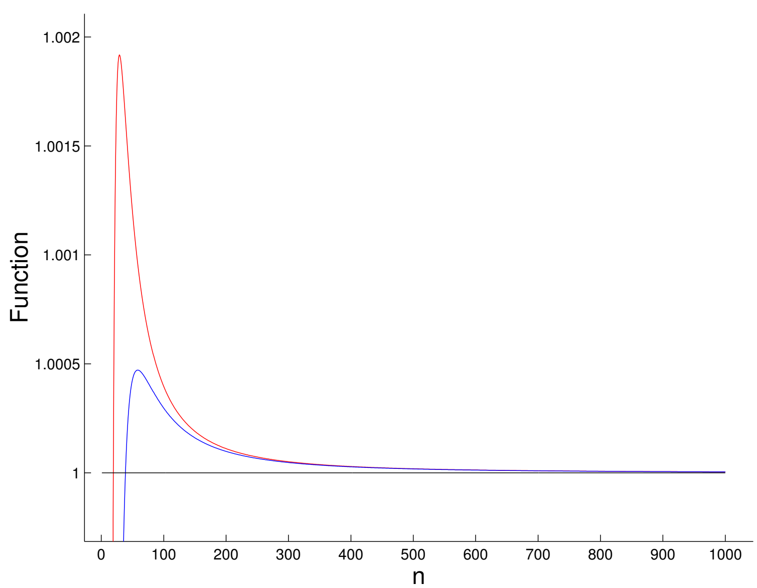

Fig. 4.1. Both factors $\frac{(1-t_n^4)^n}{\sqrt{1-t_n^2}}$ and $\frac{(1-t_n'^4)^{2n+1}}{\sqrt{1-t_n'^2}}$ as a function of n. The upper is the one using t_n, the lower one for t'_n. They are larger than 1 for high values of n.

So, $\lambda_{m,n}$ can be identified to $\frac{cos\left[(m+1)\frac{2\pi}{2n+1}\right]}{cos\left(\frac{2\pi}{2n+1}\right)}$. The problem is that this value cannot be guaranteed to be ≤ 1. One can indeed consider a value $m = 2n$, which gives

$$\frac{cos\left[(m+1)\frac{2\pi}{2n+1}\right]}{cos\left(\frac{2\pi}{2n+1}\right)} = \frac{1}{cos\left(\frac{2\pi}{2n+1}\right)} > 1, \text{ for } n \geq 1 \ .$$

This observation voids the lemma and another example is needed.

Alternative proof

The simplest way of patching the proof is to provide a correct example of a matrix pair $\{G(t_n), H(t_n)\}$ actually admitting an infinite number of "jumps" between stability and instability as $t_n \to 0$.

We will take back the same families of matrices. Matrices $P(\phi)$ and $Q(\phi)$ can still be interpreted in terms of projections and rotations. We identify

$$t = \sin(\phi) \ ,$$

but both notations will coexist. Let us denote by ϕ_k the angle $\frac{2\pi}{k}$, so that this *unified* notation is valid for even and odd values of k (we may denote it by ϕ below for clarity).

We now need to construct G and H by multiplying P and Q by suitable scalar factors, let us say f_G and f_H. These factors must be chosen so that the pair $\{G(\phi), H(\phi)\}$ can be alternatively stable and unstable as $\phi \to 0$. To see this, let us consider the following generic product, for any $l \in \mathbb{N}_0$:

$$\begin{aligned}
&G(\phi)H^l(\phi)G(\phi) = \dots \\
&= f_G \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ 0 & 0 \end{pmatrix} f_H^l \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}^l f_G \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ 0 & 0 \end{pmatrix} \\
&= f_G f_H^l \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}^{l+1} f_G \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ 0 & 0 \end{pmatrix} \\
&= f_G f_H^l \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos[(l+1)\phi] & -\sin[(l+1)\phi] \\ \sin[(l+1)\phi] & \cos[(l+1)\phi] \end{pmatrix} f_G \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ 0 & 0 \end{pmatrix} \\
&= f_G f_H^l \begin{pmatrix} \cos[(l+1)\phi] & -\sin[(l+1)\phi] \\ 0 & 0 \end{pmatrix} f_G \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ 0 & 0 \end{pmatrix} \\
&= f_G f_H^l \cos[(l+1)\phi] \ G(\phi) \\
&= \gamma(l,\phi) \ G(\phi) \ .
\end{aligned}$$

This shows that any product of the form $G(\phi)H^l(\phi)G(\phi)$ is equivalent to $G(\phi)$ times a scalar, which we denote by $\gamma(l,\phi)$. As any product can be written in the form:

$$G(\phi)H^{l_1}(\phi)G(\phi)H^{l_2}(\phi)G(\phi)\dots H^{l_m}(\phi)G(\phi)\ ,$$

and then reduced using the above relation, we only have to study the stability of $G(\phi)H^l(\phi)G(\phi)$, for any l. More precisely, we need to know when $\gamma(l,\phi)$ is >1 or <1.

So, it remains to choose f_G and f_H that suit our needs. Looking back at the limits considered in equations (4.4) and (4.5), we want to try $f_H = 1-t^4$ (as before) and $f_G = \frac{1-t^4}{1-\frac{3\pi}{2}t^3}$, where t stands for $\sin(\phi)$. We now have to study the behavior of the factor

$$\gamma(l,\phi) = \frac{(1-t^4)^{l+1}}{1-\frac{3\pi}{2}t^3}\cos[(l+1)\phi]$$

as $t \to 0$ for any l in the cases of even and odd k.

• In the case k is even, let us say $k = 2k'$, we have $\phi = \frac{2\pi}{k} = \frac{\pi}{k'}$ and the ratio is

$$\gamma(l,\phi) = \frac{(1-t^4)^{l+1}}{1-\frac{3\pi}{2}t^3}\cos[(l+1)\frac{\pi}{k'}]\ .$$

As we may choose any strategy to prove instability, we may pick an arbitrary l. We then take $l = k'-1$, so

$$\gamma(k'-1,\phi) = \frac{(1-t^4)^{k'}}{1-\frac{3\pi}{2}t^3}\cos(\pi) = -\frac{(1-t^4)^{k'}}{1-\frac{3\pi}{2}t^3}\ .$$

Using this, $\gamma(k'-1,\phi) \to 1$ as $t \to 0$, and, as before, its absolute value tends to 1 *from above*. Indeed,

$$\gamma(k'-1,\phi) \sim \frac{1-k't^4}{1-\frac{3\pi}{2}t^3}\ ,$$

where $k' = \pi/\phi \sim \pi/t$, because $\sin t \sim t$ as $t \to 0$. So,

$$\gamma(k'-1,\phi) \sim \frac{1-\pi t^3}{1-\frac{3\pi}{2}t^3} > 1\ .$$

• In the case k is odd, let us say $k = 2k''+1$, we have $\phi = \frac{2\pi}{2k''+1}$ and the ratio is

$$\gamma(l,\phi) = \frac{(1-t^4)^{l+1}}{1-\frac{3\pi}{2}t^3}\cos\left[(l+1)\frac{2\pi}{2k''+1}\right]\ .$$

A first observation teaches us that this also tends to 1 as $k'' \to \infty$. This time, as we want to prove stability, we must prove $\gamma(l,\phi) < 1$ for any value of l and for any $\phi = \frac{2\pi}{2k''+1}$ for k'' large enough. So, we may not pick an arbitrary l as in the previous case. We have to prove that $\forall l$, and for t close enough to 0,

$$\frac{(1-t^4)^{l+1}}{1-\frac{3\pi}{2}t^3}\cos\left[(l+1)\frac{2\pi}{2k''+1}\right] < 1\ .$$

Without developing cumbersome details, we can see that, for any fixed l, the factor $\frac{(1-t^4)^{l+1}}{1-\frac{3\pi}{2}t^3}$ tends to 1, while the absolute value cos factor is ≤ 1. If it is < 1, then we are done. It remains to study the case when $\cos\left[(l+1)\frac{2\pi}{2k''+1}\right] = \pm 1$, that is when $\frac{2(l+1)}{2k''+1} \in \mathbb{N}$. Let us pose

$$\frac{2(l+1)}{2k''+1} = r \in \mathbb{N}\ .$$

So, $l = rk'' + \frac{r}{2} - 1$. The absolute value of the whole factor to study can be rewritten

$$\frac{(1-t^4)^{rk''+\frac{r}{2}}}{1-\frac{3\pi}{2}t^3}\ ,$$

where the cos has disappeared because it is ± 1. This tends to 1 from below, as can be seen by using this simple development:

$$\frac{1-(rk''+\frac{r}{2})t^4}{1-\frac{3\pi}{2}t^3}\ ,$$

and using $t \sim \frac{2\pi}{2k''+1} \sim \frac{\pi}{k''}$ for large k:

$$\frac{1-r\frac{\pi^4}{k''^3}-\frac{r\pi^4}{2k''^4}}{1-\frac{3\pi^4}{2k''^3}}\ .$$

Any choice of $r(k'')$ leads to this ratio being < 1 for $k'' \to \infty$. Figure 4.2 backs up our claim.

Remark

If we took $l = 2k''$ (trying to find the worst l in a construction analog to the even case, with the cos function being equal to 1), this would yield the following:

$$\gamma(2k'',\phi) = \frac{(1-t^4)^{2k''+1}}{1-\frac{3\pi}{2}t^3}\cos(2\pi) = \frac{(1-t^4)^{2k''+1}}{1-\frac{3\pi}{2}t^3}\ .$$

For any fixed k'', $\gamma(2k'',\phi) \to 1$ as $t \to 0$, and it tends to 1 *from under*. Indeed,

$$\gamma(2k'',\phi) \sim \frac{1-(2k''+1)t^4}{1-\frac{3\pi}{2}t^3}\ ,$$

where $2k''+1 = 2\pi/\phi \sim 2\pi/t$, because $\sin t \sim t$ as $t \to 0$. So,

$$\gamma(2k'', \phi) \sim \frac{1 - 2\pi t^3}{1 - \frac{3\pi}{2} t^3} < 1 ,$$

still giving stable products.

This concludes the proof of the non semi-algebraic property of any set U such that

$$E(m, n) \subseteq U \subseteq S(m, n) .$$

4.3 Existence of an extremal norm in case of stability

4.3.1 Definition of extremal norms

We now present a definition of the extremality property of matrix norms, which will be used in this chapter and later ones. For more theoretical considerations about this aspect, an extensive discussion of extremal norms by Wirth can be found in [85], where the statement of Barabanov's theorem also appears.

As mentioned in and after the proof of Lemma 1.1 (Section 1.1.3), there exist norms that play a specific role relatively to sets of matrices. We now

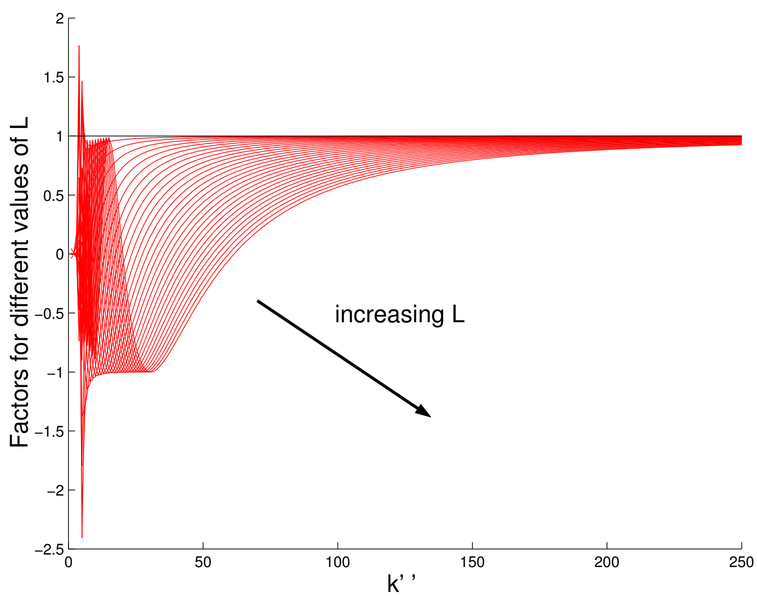

Fig. 4.2. Plots of the factor $\frac{(1-t^4)^{l+1}}{1-\frac{3\pi}{2}t^3} \cos\left[(l+1)\frac{2\pi}{2k''+1}\right]$ for values of l from 1 to 30. The larger l values yield the lower curves. They all tend to 1 from below as $k'' \to \infty$. The jump above 1 for small values of k'' does not matter, since we study the behavior for $k'' \to \infty$.

present an extension of this property to matrix sets. We know, from the four members inequality, that, for a set of matrices $\mathcal{M} = \{A_i, i \in I\}$,

$$\sup_{i \in I} \|A_i\| \geq \hat{\rho}(\mathcal{M}) \ .$$

Definition 4.14. *A norm $\|.\|$ is called* extremal *(for a given set $\mathcal{M}$) if it satisfies*

$$\|A\| \leq \hat{\rho}(\mathcal{M}), \forall A \in \mathcal{M} \ .$$

From the above property, it is clear that, for a given norm, this inequality cannot be strict simultaneously for all matrices $A \in \mathcal{M}$.

In [3], Barabanov shows the following :

Theorem 4.15 (Barabanov). *If a set $\mathcal{M}$ of matrices is compact and irreducible, then there exists a vector norm $\|.\|_{Bar}$ such that*

$$\forall x, \forall A_i \in \mathcal{M}, \ \|A_i x\|_{Bar} \leq \rho(\mathcal{M})\|x\|_{Bar} \ ,$$
$$\forall x, \exists A_i \in \mathcal{M} : \ \|A_i x\|_{Bar} = \rho(\mathcal{M})\|x\|_{Bar} \ .$$

This result is coherent with lemma 1.6, as non product-bounded sets are shown to be reducible. Hence, an irreducible set is guaranteed to be product bounded. We call a norm such as in theorem 4.15 a *Barabanov norm*. It is an extremal norm, as can quickly be seen from the first property in the above theorem. Moreover, the second property adds that, from any point x, a Barabanov norm gives a criterion to choose the matrix A_i achieving the JSR.

For completeness, we may add that Elsner & Friedland have come up with norm conditions for the convergence of infinite products, generated by a pair of matrices (in [35]).

4.3.2 Theorem on the existence of extremal norms

The following theorem proves the existence of an extremal norm (as defined in the previous section). Yet, nothing can be said about the shape of this norm. Nonetheless, the existence itself of an extremal norm is the source of a nontrivial result about the convergence of the approximations of the JSR that will be presented in Section 6.1.1.

Theorem 4.16 (Kozyakin). *(a) The system $x_{t+1} \in \{A_{i_t} x_t : i_t \in \{1, \ldots, m\}\}$ with $\mathcal{M} = \{A_1, A_2, \ldots, A_m\}$ is absolutely stable iff there exists an induced norm $\|.\|_*$ in $\mathbb{R}^n$ such that*

$$\|A_1\|_*, \|A_2\|_*, \ldots, \|A_m\|_* \leq 1 \ .$$

(b) The system $x_{t+1} \in \{A_{i_t} x_t : i_t \in \{1, \ldots, m\}\}$ with $\mathcal{M} = \{A_1, A_2, \ldots, A_m\}$ is absolutely exponentially stable iff there exists an induced norm $\|.\|_*$ in $\mathbb{R}_n$ such that $\exists\ q < 1$:

$$\|A_1\|_*, \|A_2\|_*, \ldots, \|A_m\|_* \leq q \ .$$

Proof. The *if* part of both parts of the theorem is clear, as we illustrate for the (b) part, using the submultiplicativity of the norm to state that

$$\|A_1\|_*, \|A_2\|_*, \ldots, \|A_m\|_* \leq q \Rightarrow \|A_{i_n} A_{i_{n-1}} \ldots A_{i_1} \ x\| \leq q^n \|x\| \ .$$

The same argument can be used for the (a) part, with $q = 1$.

Now let us consider the *only if* part of the theorem. Let us suppose that there exists a $q < 1$ such that, for any sequence of matrices A_i, we have

$$\|A_{i_n} A_{i_{n-1}} \ldots A_{i_1} \ x\| \leq c q^n \|x\| \ .$$

The existence of the wanted norm will be a consequence of the following construction. We define the κ function as follows, using the above norm $\|.\|$:

- $\kappa_0(x) = \|x\|$,
- $\kappa_n(x) = q^{-n} \max_{B_i \in \mathcal{M}} \|B_1 B_2 \ldots B_n x\| \quad (n \geq 1)$.

Then we define the norm $\|.\|_*$ by

$$\|x\|_* = \sup_{n \geq 0} \kappa_n(x) \ .$$

We may remark that the definition of κ_n is somewhat similar to the joint spectral radius. Indeed, we are here again choosing matrix products that maximize a norm, but this time also involving the point x. The condition on q guarantees the boundedness of κ_n, for any n. We can check that $\|.\|_*$ itself satisfies all the properties of a norm. This is done right after this proof, to keep the present reasoning as short and clear as possible.

We can now say that $\forall A_k \in \mathcal{M}, \forall n \in \mathbb{N}$, we have

$$\begin{aligned}
\max_{B_i \in \mathcal{M}} \|B_1 B_2 \ldots B_n A_k x\| &\leq \max_{B_i \in \mathcal{M}} \|B_1 B_2 \ldots B_n B_{n+1} x\| \\
q^n \kappa_n(A_k x) &\leq q^{n+1} \kappa_{n+1}(x) \\
\kappa_n(A_k x) &\leq q \kappa_{n+1}(x) \ .
\end{aligned}$$

And taking the $\sup_{n \geq 0}$, we have

$$\|A_k x\|_* \leq q \sup_{n \geq 1} \kappa_n(x) \leq q \|x\|_* \ .$$

and so, $\frac{\|A_k x\|_*}{\|x\|_*} \leq q$. Consequently, the matrix norm induced by the vector norm $\|.\|_*$ defined above has the desired property. □

The desired matrix norm is induced by the above defined vector norm:

$$||A_k||_* = \max_x \frac{\|A_k x\|_*}{\|x\|_*} = \max_x \frac{\sup_{n \geq 0}(q^{-n} \max_{B_i} \|B_1 \ldots B_n A_k x\|)}{\sup_{n \geq 0}(q^{-n} \max_{B_i} \|B_1 \ldots B_n x\|)} \ .$$

Checking the norm properties of $\|.\|_*$ (vector norm)

- Subadditivity:
$$\|x_1 + x_2\|_* \leq \|x_1\|_* + \|x_2\|_* \ .$$
We have $\|x_1 + x_2\|_* = \sup_{n\geq 0} \kappa_n(x_1 + x_2)$. If the sup is achieved for $n = 0$, then the inequality $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$ suffices to prove the property. Let us assume the sup is achieved for $n = \tilde{n} \neq 0$. We can write
$$\begin{aligned}
\|x_1 + x_2\|_* &= \sup_{n\geq 0} \kappa_n(x_1 + x_2) = \kappa_{\tilde{n}}(x_1 + x_2) \\
&= q^{-\tilde{n}} \max_{B_i} \|B_1 \dots B_{\tilde{n}}(x_1 + x_2)\| \\
&= q^{-\tilde{n}} \|B_1^* \dots B_{\tilde{n}}^*(x_1 + x_2)\| \quad \text{where } B_1^* \dots B_{\tilde{n}}^* \text{ achieves the max} \\
&\leq q^{-\tilde{n}} \|B_1^* \dots B_{\tilde{n}}^* x_1\| + q^{-\tilde{n}} \|B_1^* \dots B_{\tilde{n}}^* x_2\| \\
&\leq \kappa_{\tilde{n}}(x_1) + \kappa_{\tilde{n}}(x_2) \\
&\leq \|x_1\|_* + \|x_2\|_* \ .
\end{aligned}$$
- Nonzero except in $x = 0$:
$$\|x\|_* = 0 \iff x = 0 \ .$$
This property is obvious from the definition of $\|.\|_*$, since
$$x = 0 \Rightarrow \|x\|_* = 0$$
and
$$\|x\|_* = 0 \Rightarrow \kappa_0(x) = 0 \Rightarrow x = 0 \ .$$
Yet, we can even show the equivalence between $\|.\|_*$ and the norm $\|.\|$ we used in the first place. From the definition of $\|x\|_*$, we know that $\|x\| \leq \|x\|_*$. And, considering
$$\begin{aligned}
\|A_{i_n} A_{i_{n-1}} \dots A_{i_1} \ x\| &\leq c q^n \|x\| \\
q^{-n} \|A_{i_n} A_{i_{n-1}} \dots A_1 \ x\| &\leq c\|x\| \\
\text{taking the } \sup_{A_i}, \quad \kappa_n(x) &\leq c\|x\| \\
\text{taking the } \sup_{n}, \quad \|x\|_* &\leq c\|x\| \ .
\end{aligned}$$
This leads to $\|x\|_* \leq \max(c, 1)\|x\|$ (this $\max(c, 1)$ comes from the case $n = 0$). Gathering the results, we have
$$\|x\| \leq \|x\|_* \leq \max(c, 1)\|x\| \ ,$$
which also proves that $\|x\|_* = 0 \iff x = 0$.
- Linearity:
$$\|kx\|_* = |k|\|x\|_* \ .$$
This property is straightforward to check.

4.3.3 Notes on Kozyakin's norm

We have seen that q must satisfy the property described in Theorem 4.16,

$$\forall A_i \in \mathcal{M},\ \forall n,\ \|A_{i_n} A_{i_{n-1}} \dots A_{i_1}\ x\| \leq q^n \|x\| \ ,$$

in order to keep

$$\sup_{n \geq 0} \kappa_n(x)$$

bounded. Otherwise $\|x\|_*$ is unbounded and is therefore not a norm. This leads us to the following short remark.

Characterization of the exponential growth rate using the JSR

From the definitions, we clearly cannot have $q < \rho(\mathcal{M})$. On the other hand, $q > \rho(\mathcal{M})$ is always acceptable, since the κ_n functions would tend to zero for larger n, trivially ensuring that κ_n is bounded. In this perspective, Theorem 4.16 completes Theorem 4.15. Both can be combined in the following statement.

Proposition 4.17. *For any finite set of matrices $\mathcal{M} = \{A_1, \dots, A_m\}$, for any $\varepsilon > 0$, there exists a matrix norm $\|.\|_*$ such that*

$$\|A_1\|_*, \|A_2\|_*, \dots, \|A_m\|_* \leq \rho(\mathcal{M}) + \varepsilon \ .$$

If the set is irreducible, then there exists a matrix norm $\|.\|_*$ such that

$$\|A_1\|_*, \|A_2\|_*, \dots, \|A_m\|_* \leq \rho(\mathcal{M}) \ .$$

Proof. Consider a finite set $\mathcal{M}$ of matrices of joint spectral radius ρ^*. If we divide it by a quantity r larger than ρ^*, then the JSR of $\mathcal{M}/r$ is less than 1:

$$\forall r > \rho^*,\ \rho(\mathcal{M}/r) = \frac{\rho^*}{r} < 1 \ .$$

One can then apply Theorem 4.16 : there exists a $q < 1$ and a norm $\|.\|_*$ such that $\left\|\frac{A_i}{r}\right\|_* \leq q, \forall i$ (i.e. $\max_i \left\|\frac{A_i}{r}\right\|_* \leq 1, \forall x$). By linearity of the norm and the joint spectral radius, this allows us to deduce that

$$\forall r > \rho(\mathcal{M}), \exists \|.\|_* : \max_i \|A_i\|_* \leq r \ .$$

One can replace r by $\rho(\mathcal{M}) + \varepsilon$ to obtain the exact formulation of the proposition.
The second statement is simply Barabanov's theorem. $\square$

This confirms that the joint spectral radius can be seen as the infimum over all possible matrix norms of the largest norm of the matrices in the set, as already announced back in Section 4.3.1. As a reminder, we called *extremal* a norm achieving this infimum for the set (not every set of matrices possesses an extremal norm, see [83] for a discussion of this issue). We have seen that Kozyakin described the theoretical construction of such an extremal norm, but this construction is not explicit and partly relies on the a-priori knowledge of the numerical value of the joint spectral radius.

One can of course not hope enumerating all possible matrix norms for computing the joint spectral radius, but we can enumerate particular sets of norms. This is what will be done in Section 6.5.

Remark 4.18. In the proof of Theorem 4.16, it is established that absolute exponential stability implies the existence of a *vector* norm $\|.\|_*$ such that

$$\|A_1 x\|_*, \|A_2 x\|_*, \ldots, \|A_m x\|_* \leq q\|x\|_*, \ \forall x \ .$$

We could have expressed the inequalities in Theorem 4.16 and Proposition 4.17 using vector norms, as here above. Further on, we will make use of such a vector norm version of Proposition 4.17.

The question that remains is: when exactly can we choose $q = \rho(\mathcal{M})$? This is equivalent to deciding if a set $\mathcal{M}$ with $\rho(\mathcal{M}) = 1$ is product bounded. As we will see in Section 4.4, results from [79] show that this is undecidable.

Let us begin with a simpler example. For a single matrix, for instance, in order to chose q equal to the spectral radius of the matrix, we need to exclude the case where the largest eigenvalue has a geometric multiplicity different from its algebraic multiplicity. Differently said, the matrix must admit a diagonal form (at least for the part concerning the spectral radius). Otherwise, its Jordan form (defined in Section 1.1.2) has at least one block of size ≥ 2 for $\rho(A)$. We illustrate such a block of size 2:

$$J_A = \begin{pmatrix} \rho(A) & 1 \\ 0 & \rho(A) \end{pmatrix} \ .$$

And one may note that

$$\begin{pmatrix} \rho(A) & 1 \\ 0 & \rho(A) \end{pmatrix}^n = \begin{pmatrix} \rho(A)^n & n\rho(A)^{n-1} \\ 0 & \rho(A)^n \end{pmatrix} = \rho(A)^n \begin{pmatrix} 1 & \frac{n}{\rho(A)} \\ 0 & 1 \end{pmatrix} \ .$$

And, more generally, for Jordan blocks of larger size, $\left\| \frac{J_A^n}{\rho(A)^n} \right\|$ is not bounded as $n \to \infty$, as can be seen from the general expression of the power of a Jordan block in the proof of Lemma 1.1. So this prevents us from using $q = \rho(A)$ in such as case. Although, we may choose q as close to $\rho(A)$ (from above) as we want, say $q = 1 + \epsilon$. Indeed, the growth mentioned here is linear in n ($\frac{n}{\rho(A)}$), while q determines an exponential growth rate (which is always stronger than a linear one).

For sets of matrices, we use a different approach, by considering a *normalized* set of matrices: the initial set $\mathcal{M}$ is divided by its joint spectral radius, $\mathcal{M}' = M/\rho(\mathcal{M})$, so that $\rho(\mathcal{M}') = 1$ (we assume that $\rho(\mathcal{M}) \neq 0$). This normalization may look like a restriction, but we do not lose any generality by doing this, since we are only multiplying the set by a scalar, as explained earlier in Section 1.2.2. We know that

$$\text{a value } q \text{ is acceptable } \iff \kappa_n(x) \text{ bounded } \forall n \text{ (for bounded } |x|) \ .$$

Now, choosing precisely $q = 1$, it implies that, for a given x,

$$\sup_n \max_{B_i \in A} \|B_1 B_2 \dots B_n x\|$$

must be bounded.

Families of matrices having a JSR equal to 1 and generating products that are bounded in norm are called *non-defective* (see Guglielmi and Zennaro [40, 41]). So, non-defective sets of matrices allow $q = \rho(\mathcal{M})$ and do admit an extremal norm.

Extremality **property of Kozyakin's matrix norm**

In the case q can be chosen to be equal to $\rho(\mathcal{M})$, we can deduce the following property.

Lemma 4.19. *If the value q, appearing in theorem 4.16, can be chosen to be equal to the joint spectral radius $\rho(\mathcal{M})$ of a compact set $\mathcal{M}$, then*

$$\exists k : \|A_k\|_* = \rho(\mathcal{M}) \ .$$

Proof. Let us start from the assumption that there exists a c such that:

$$\|A_{i_n} A_{i_{n-1}} \dots A_{i_1} \ x\| \leq c\rho(\mathcal{M})^n \|x\| \ .$$

Then, by theorem 4.16, there exists a norm $\|.\|_*$ in $\mathbb{R}_n$ such that

$$\|A_1\|_*, \|A_2\|_*, \dots, \|A_m\|_* \leq \rho(\mathcal{M}) \ .$$

Now, from the the four members inequality (1.17), we know that:

$$\hat{\rho}(\mathcal{M}) \leq \hat{\rho}_k(\mathcal{M}, \|.\|)^{1/k}, \ \forall k \ .$$

For $k = 1$, this means that

$$\hat{\rho}(\mathcal{M}) \leq \hat{\rho}_1(\mathcal{M}, \|.\|) = \sup_i \|A_i\| \ .$$

This inequality holds for any matrix norm, including the norm $\|.\|_*$ considered here:

$$\sup_i \|A_i\|_* \geq \rho(\mathcal{M}) \ .$$

We conclude that

$$\rho(\mathcal{M}) = \sup_i \|A_i\|_* \ .$$

The set being compact, there exists a matrix achieving this supremum. $\square$

This shows that, under the above assumption, Kozyakin's norm is an extremal norm. However, it is not necessarily a Barabanov norm, as defined in Section 4.3.1. Although its construction methodology is detailed, it is still impossible to compute in general, since it requires the knowledge of the exact value of the joint spectral radius. This actually brings us back to our initial problem.

4.4 NP-hardness of the problem

We conclude this chapter by presenting results that show NP-hardness and undecidability results. Although we will not be directly using them afterwards, we feel they are needed to understand the difficulty of the problem at hand. The decidability is a common notion. In the following, a problem/question is said to be *decidable* if it can be answered using a finite number of computational steps.

In [79], Tsitsiklis & Blondel show the following:

Definition 4.20. *The function $\rho(\mathcal{M})$ is said to be* polynomial-time approximable *if there exists an algorithm $\rho^*(\mathcal{M}, \varepsilon)$ returning an approximation $\rho^*(\mathcal{M})$ of $\rho(\mathcal{M})$ with a relative error of at most ε in time polynomial in the sizes of $\mathcal{M}$ and ε:*

$$|\rho^*(\mathcal{M}) - \rho(\mathcal{M})| \leq \varepsilon |\rho(\mathcal{M})| \ .$$

The *size* of $\mathcal{M}$ and ε denotes the bit size of their description.

Theorem 4.21 (Tsitsiklis-Blondel). *Unless $P = NP$, the joint/generalized spectral radius $\rho(\mathcal{M})$ of a pair $\mathcal{M}$ of matrices is not polynomial-time approximable. This is true even for a pair of matrices with $\{0, 1\}$ entries.*

This theorem leads to two corollaries that explain its implications.

Corollary 4.22. *Unless $P = NP$, the joint/generalized spectral radius of a pair of $(n \times n)$ matrices with $\{0, 1\}$ entries is not approximable with relative error 2^{-k} in a number of operations that is polynomial in n and k ($k \in \mathbb{N}_0$).*

Corollary 4.23. *It is NP-hard to decide whether all products of two given real matrices A_0 and A_1 are bounded.*

These results show in abundance how high the computational cost of the computation of the JSR can be. We now complete these with decidability results. In [13], the following results are detailed:

Theorem 4.24 (Blondel-Tsitsiklis). *The following problems are undecidable:*

- *Given a finite set $\mathcal{M}$ of $(n \times n)$ matrices with rational entries, determine if $\rho(\mathcal{M}) \leq 1$.*
- *Given a finite set $\mathcal{M}$ of $(n \times n)$ matrices with rational entries, determine if the set of all matrix products is bounded.*

These undecidability and NP-hardness results call for a comment. They imply that, in the general case, there is no easily computable extremal norm. Otherwise, this norm would provide an efficient way of determining the boundedness of the matrix products generated by a given set.

The same authors present a more general survey of complexity results in systems theory and control in [12]. Note that the first problem in the theorem above looks like it could have been used to solve the Finiteness Conjecture, but it cannot. Indeed, as we said in Section 4.1.1, the FC being true would imply the decidability of $\rho(\mathcal{M}) < 1$, while here we are talking about $\rho(\mathcal{M}) \leq 1$. However, this first problem of 4.24 implies the following:

Corollary 4.25. *For any finite set $\mathcal{M}$ of square matrices $\in Q^{n\times n}$, there does not exist in general an effectively computable $k^*(\mathcal{M})$, giving the length of the shortest product achieving the JSR:*

$$\rho_{k^*}(\mathcal{M}) = \rho(\mathcal{M}) \ .$$

4.5 Conclusion

From this chapter, it could seem that computing the JSR is impossible. The situation is not that bad, as there are several cases where it can actually be computed exactly, or with a non-exponential computational cost. These cases include namely matrices with non-negative elements, upper-triangular matrices and a pair of the type $\{A, A^T\}$. We do not list them exhaustively here, since they are detailed in Chapter 6. In this chapter, we also present tentative algorithms to approach the JSR efficiently in the general case. In chapter 5, we present a nontrivial example for which we can compute the JSR effectively, thanks to an understanding of the specific structure of the products achieving it.

5

Study of a specific pair of matrices

In this chapter, we study in detail the example that was originally used in our counterproof of the Finiteness Conjecture (presented in Section 4.1). To facilitate the exposition, we take the liberty of reminding some notions that were already used in this work. By doing so, we hope to make this section easier to read and more self-contained.

5.1 Binary words

As the example we are presenting in this section involves a pair of matrices, all the products will be represented as binary sequences, or binary *words*. We denote by I the alphabet $\{0, 1\}$, by I^+ the set of finite nonempty words, and by I^∞ the set of nonempty finite and infinite words. As an example, if we denote the matrices by A_0 and A_1, then the product $A_0A_1A_1A_0A_1$ will be denoted by A_ω, where $\omega = 01101$. The i^{th} element of ω is denoted by $\omega(i)$. Here, we would have $\omega(2) = 1$. We now present various properties of binary words, many of which can be found in [55].

The *length* of a word ω is its number of elements, and will be denoted by $|\omega|$. Similarly, the notations $|\omega|_0$ and $|\omega|_1$ respectively denote the number of 0's and the number of 1's in ω. We can also define the *1-ratio* of a finite word ω as follows:

$$\forall \omega \in I^+, r_1(\omega) := \frac{|\omega|_1}{|\omega|} \ .$$

Example 5.1.

$$\begin{aligned} \omega = 001011001 &\rightarrow |\omega| = 9,\ |\omega|_0 = 5,\ |\omega|_1 = 4 \\ &\rightarrow r_1(\omega) = 4/9. \end{aligned}$$

For infinite words, some caution is required, as the 1-ratio might not converge to any value.

$$\forall \omega \in I^\infty, r_1(\omega) := \lim_{k\to\infty} \frac{|\omega_k|_1}{|\omega_k|}, \text{ if this limit exists },$$

where ω_k represents the sequence of the k first elements of ω.

Note that these two definitions (for finite and infinite words) are coherent, as they will give equal values for an infinite periodic word, whether one considers its finite period or the infinite word in its whole.

Example 5.2.

$$\begin{aligned}\omega = 001001001001\ldots &\to |\omega| = |\omega|_0 = |\omega|_1 = +\infty \\ &\to r_1(\omega) = 1/3.\end{aligned}$$

The main issue about computing the 1-ratio arises when one has to deal with a non-periodic infinite word. Indeed, in such a case, it could happen that the above limit does not exist. Just consider a word composed of one '0', followed by 2 '1', then 4 '0', and so on, each time alternating the digit and doubling the number of occurrences. Such a word clearly does not admit any r_1. Consequently, there is a need for a way of determining whether the words we will be considering in the following discussions do actually yield a value for r_1.

Up to now, we have been dealing with words, and we now need to define what we mean by *subwords*. We say that u is a subword of w, or $u \subset w$, if

$$\exists k : \ \forall\ 1 \le j \le |u|, \quad w(k+j) = u(j)\ .$$

We now recall the notion of *balanced* word. There exist several equivalent definitions, and we present here the one that appears useful for our purpose.

Definition 5.3. *A binary word w is balanced if and only if*

$$\forall n \in \mathbb{N}, \quad \forall u, v \subset w : |u| = |v| = n, \quad -1 \le |u|_1 - |v|_1 \le 1\ .$$

This means that all subwords of the same length must contain nearly the same number of 1's, with an allowed variation of one. This rules out, for example, the presence of both 00 and 11 in a balanced word. Note that all subwords of a balanced word are themselves balanced. Infinite words that are balanced can be either periodic or not. In the latter case, they are called *Sturmian* words.

An interpretation of the balanced property is the following. Let us consider the real plane $\mathbb{R}^2$, gridded with lines corresponding to integer values. Let us choose a slope s and plot the line $y = sx$. Every time this line cuts a vertical line, it yields 0, and when it cuts horizontal lines, it yields 1. This is illustrated in Figure 5.1. One can show that a rational value for s yields a periodic balanced word, while an irrational s leads to a Sturmian word.

5.2 The pair of matrices

We now present the specific example we will be studying. Let us consider the following pair of matrices :

$$A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \; A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \; .$$

This matrix pair was already used in Section 4.1 for proving the existence of a counterexample to the finiteness conjecture. It appears to have specific properties, which we develop here.

For $\alpha \in \mathbb{R}$, we denote by ρ_α the joint spectral radius of the pair $\{A_0, \alpha A_1\}$. We also define normalized matrices

$$A_0^\alpha = \frac{A_0}{\rho_\alpha} \; , A_1^\alpha = \frac{\alpha A_1}{\rho_\alpha} \; .$$

so that $\rho(A_0^\alpha, A_1^\alpha) = 1$.

For a word $\omega \in I^+$, A_ω^α is the corresponding product of matrices A_0^α and A_1^α. Let us denote by OS_α the set of binary sequences that achieve the joint spectral radius for a given α:

$$OS_\alpha = \left\{ \omega \in I^\infty : \lim_{k \to \infty} \rho(A_{\omega_k}^\alpha)^{1/|\omega_k|} = 1 \right\} \; .$$

These sequences will be referred to as *optimal sequences*. These optimal words have a 1-ratio that we denote by $\mu_1(\alpha)$:

$$\mu_1(\alpha) = \{ r_1(\omega) \in \mathbb{R} : \omega \in OS_\alpha \} \; .$$

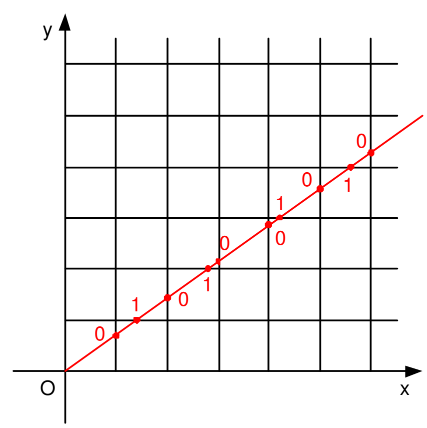

Fig. 5.1. The line $y = sx$ through $\mathbb{R}^2$, yielding 0's and 1's when cutting integer coordinates.

Later, we will state and prove several properties of this function $\mu_1(\alpha)$. For example, we will show that, for any α, the value $\mu_1(\alpha)$ is unique. But we first need to recall some notations and results from Section 4.1:

- For a given word $\omega \in I^+$, we define $J_\omega = \{\alpha : \rho(A^\alpha_\omega) = 1\}$, the set of values of α such that the given word ω is optimal for the pair of matrices $\{A^\alpha_0, A^\alpha_1\}$.
- Two words $u, v \in I^+$ are said to be *essentially equal* if the periodic infinite words $U = uu\ldots$ and $V = vv\ldots$ can be decomposed as $U = xzz\ldots$ and $Y = yzz\ldots$ for some $x, y, z \in I^+$.
  Words that are not essentially equal are *essentially different*.

And from lemmas 4.5 and 4.8, we know:

- If $u, v \in I^+$ are essentially equal, then $J_u = J_v$.
  If $u, v \in I^+$ are essentially different, then $J_u \cap J_v = \emptyset$.

5.3 The 1-ratio of the optimal words

Before studying the properties of the function $\mu_1(\alpha)$, we need to know whether the optimal words we will be considering in this section always admit a well-defined 1-ratio. So, it proves useful to focus on the particular structure of the optimal word(s) for a given value of α.

We prove that, for the specific pair we are studying, there always exists an optimal word that is balanced. More precisely, for a fixed α, among all possible sequences achieving the joint spectral radius, we can always find one that is balanced (or even Sturmian, if it is not periodic).

We now present notions needed to understand the proof of Proposition 5.10.

5.3.1 Lexicographic order

An *order* can be defined between binary words, which is only partial. Indeed, in the following definition, words of different lengths cannot be compared. Given two binary words u and v with $|u| = |v|$, we say that $u > v$ if and only if

$$\exists\ k < |u| \in \mathbb{N} : \begin{array}{l} \bullet\ \forall i \leq k, u(i) = v(i), \\ \bullet\ u(k+1) > v(k+1),\ \text{that is } u(k+1) = 1, v(k+1) = 0\ . \end{array}$$

As an example, $1011001 > 1010100$.

The word obtained by reversing the order of the elements of ω will be denoted $\tilde{\omega}$. For example, $\omega = 001011001$ gives $\tilde{\omega} = 100110100$. A word for which $\omega = \tilde{\omega}$ is called a *palindrome*, such as 1001001.

5.3.2 Reminder of previous results

In Section 4.1, Lemma 4.7, we defined a *dominating flip* as follows:

Definition 5.4. *Consider a word $\omega = psq$, with $s > \tilde{s}$, $q < \tilde{p}$ (and so $|p| = |q|$). Making the change from $\omega = psq$ to $\omega' = p\tilde{s}q$ is called a* dominating flip.

Here, the word $\omega' = p\tilde{s}q$ *dominates* ω, because $A_{\omega'} \geq A_\omega$ element-wise, with $\text{trace}(A_{\omega'}) > \text{trace}(A_\omega)$. This implies $\rho(A_{\omega'}) > \rho(A_\omega)$.

As 0 and 1 play symmetric roles, we naturally have the same result, reversing the order of the lexicographic inequalities. This result can consequently be extended in the following way.

Lemma 5.5. *For any word of the form $\omega = psq$, where $s < \tilde{s}$ and $q > \tilde{p}$, the matrix $A_{\omega'}$ with $\omega' = p\tilde{s}q$ dominates A_ω.*

The proof is similar in every aspect to the proof of Lemma 4.7, but we give it here to explain how it comes to the same conclusion. As a reminder, the matrix T appearing below is defined as

$$T := A_0A_1 - A_1A_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Proof. We have $A_{\omega'} - A_\omega = k(s)A_pTA_q$, with $k(s) < 0$ because $s < \tilde{s}$. As long as the last bits of p will coincide with the first ones of q, the computation of the product A_pTA_q around T will yield products of the type A_iTA_i, which are equal to T. But, as $q > \tilde{p}$, we will get a product

$$A_0TA_1 = \begin{pmatrix} -1 & -1 \\ -1 & 0 \end{pmatrix} .$$

As $k(s) < 0$, we can conclude that $A_{\omega'}$ dominates A_ω. □

Note that in the original proof, we had $k(s) > 0$ because $s > \tilde{s}$, and $A_1TA_0 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, also leading to all non-negative elements.

5.3.3 Proof of existence

The proof is divided into the following lemmas, Lemma 5.6 dealing with necessity and Lemma 5.8 with sufficiency.

Lemma 5.6. *If a word admits a dominating flip, then it is unbalanced.*

Proof. First, let us consider a word and suppose that such a dominating flip exists. Then this word contains $\omega = psq$ such that $s > \tilde{s}$, $q < \tilde{p}$. We can rewrite these "inequalities" :

As $q < \tilde{p}$, there exist words $a, b, c \in I^*$, with $|b| = |c|$, such that

$\tilde{p} = a1b$
$q = a0c.$

On the other hand, as $s > \tilde{s}$, there exist words $d, e, f \in I^*$, with $|e| = |f|$, such that

$s = d1e$
$\tilde{s} = d0f$ (Note that $d1e = \tilde{f}0\tilde{d}$).

We can now write $\omega = psq$ as $\omega = \tilde{b}1\tilde{a}sa0c$. First replacing s by $d1e$, we get $\omega = \tilde{b}1\tilde{a}d1ea0c$, which contains $1\tilde{a}d1$. And if we replace s by $s = \tilde{f}0\tilde{d}$, we get $\omega = \tilde{b}1\tilde{a}\tilde{f}0\tilde{d}a0c$, which in turn contains $0\tilde{d}a0$. So, ω contains both $1\tilde{a}d1$ and $0\tilde{d}a0$. This is enough to deduce that ω in unbalanced. □

In order to prove the converse implication, we need the following property. It can be found in the literature ([55]), but we feel it is appropriate to prove it here.

Lemma 5.7. *If a sequence is unbalanced then it contains both* $0p0$ *and* $1p1$ *for some palindrome* p. *Moreover,* $0p0$ *and* $1p1$ *are disjoint.*

Proof. Let u be unbalanced and let U and V be the two shortest words of equal length in u whose weights differ by 2. Then they have the form $U = 0a0$ and $V = 1b1$. Let us prove that $a = b = \tilde{a} = \tilde{b}$.

We begin with $a = b$. We denote by a_i and b_i the elements of a and b. If $a_1 = 0$ and $b_1 = 1$ then 00 and 11 occur in u and they are the shortest pair. If $a_1 = 1$ and $b_1 = 0$, then we can cut out two first characters of U and V and get a shorter pair of words with the same difference of weights. And so on. This settles that $a = b$, unless 00 and 11 are present in U, in which case $a = b = \emptyset$.

Suppose now that a is of length n. Then U can be written as $0a_1 \dots a_n 0$ and $V = 1a_1 \dots a_n 1$, as we know that $a = b$. Now, if $a_1 \neq a_n$, then U begins with 00 and V ends with 11. Assuming $a_1 = a_n$, we can proceed further with a_2 and a_{n-1}, and so on. Step by step, this settles that a is a palindrome, otherwise both 00 and 11 are present in u, which is the case $a = \emptyset$.

We now prove that U and V are disjoint. In order to do so, suppose they are not. Then, u contains a sequence of the type $1d0e1f0$, where $d0e = p$ and $e1f = p$. And p being a palindrome, we have that $d0e = e1f = \tilde{e}0\tilde{d} = \tilde{f}1\tilde{e}$. The second equality is impossible, because e and $\tilde{e}$ have the same length ($|e| = |\tilde{e}|$), implying that $1 = 0$ in order to be satisfied. □

We can now state the following:

Lemma 5.8. *If a word is unbalanced, then it admits a dominating flip.*

Proof. Consider an unbalanced sequence. From Lemma 5.7, it must correspond to one of these two cases :

- it contains a word of the form $1p1\omega0p0$. Then, we can read it as asb, with $a = 1p$, $s = 1\omega0$, $b = p0$. Clearly, $s > \tilde{s}$ and $b < \tilde{a}$ and so lemma 4.7 can be applied.
- it contains a word of the form $0p0\omega1p1$. In this case, we read it as asb, with $a = 0p$, $s = 0\omega1$, $b = p1$. Then, $s < \tilde{s}$ and $b > \tilde{a}$ and Lemma 5.5 can be applied.

Put together, Lemmas 5.6 and 5.8 prove the following.

Theorem 5.9. *A binary word admits a dominating flip if and only if it is unbalanced.*

This is enough for our needs, since it is clear that balanced words admit a 1-ratio. Indeed, they can be associated to a line going through the $\mathbb{R}^2$ plane, as shown in Figure 5.1. And such a line does have a slope, which is a function of μ_1 (μ_1 is the tangent of the angle this line makes with the real axis). Consequently, we can deduce that the μ_1 function we will be discussing always exists. Note that we do not formally need to associate balanced sequences to a line in the $\mathbb{R}^2$ plane to prove the existence of the 1-ratio. This could indeed be proved directly from the definition of a balanced sequence. But we think that the representation in $\mathbb{R}^2$ gives a better intuition. This existence allows us to formulate the following:

Proposition 5.10. $\mu_1(\alpha)$ *is nonempty for the pair of matrices*

$$\left\{ A_0^\alpha = \frac{A_0}{\rho_\alpha},\ A_1^\alpha = \frac{\alpha A_1}{\rho_\alpha} \right\} .$$

5.3.4 Non-decreasing property

Theorem 5.11. $\mu_1(\alpha)$ *is non-decreasing :*

$$\forall \alpha_1 < \alpha_2,\ \sup\{\mu_1(\alpha_1)\} \leq \inf\{\mu_1(\alpha_2)\} .$$

Proof. Let us suppose the assertion is false. There would then exist two values α_1 and α_2 such that $\alpha_1 < \alpha_2$ and $\mu_1(\alpha_1) > \mu_1(\alpha_2)$. Let us denote by ω_1 and ω_2 words (which we first suppose finite) of respective 1-ratios $h_1 := \mu_1(\alpha_1)$ and $h_2 := \mu_1(\alpha_2)$ achieving the joint spectral radius for these two values of α. According to our definition of optimality, this means that $\rho(A_{\omega_1}^{\alpha_1}) = \rho(A_{\omega_2}^{\alpha_2}) = 1$. Let us denote by m and n the respective lengths of ω_1 and ω_2. The number of A_1 in A_{ω_1} is mh_1, which must be an integer, by definition. The same for A_{ω_2} where there are nh_2 occurrences of A_1. We can now develop :

$$1 = \rho(A_{\omega_1}^{\alpha_1}) = \frac{\alpha_1^{mh_1}}{\rho_{\alpha_1}^m} \rho(A_{\omega_1}) \tag{5.1}$$

$$1 = \rho(A_{\omega_2}^{\alpha_2}) = \frac{\alpha_2^{nh_2}}{\rho_{\alpha_2}^n} \rho(A_{\omega_2}) . \tag{5.2}$$

Now, we may use the word ω_2 with the set associated to the value α_1 and vice-versa ω_1 with α_2. Taking into account the respective numbers of occurrences of A_1 in ω_1 and ω_2, this leads to:

$$\begin{aligned}\rho(A_{\omega_2}^{\alpha_1}) &= \frac{\alpha_1^{nh_2}}{\rho_{\alpha_1}^n}\rho(A_{\omega_2})\\ \rho(A_{\omega_1}^{\alpha_2}) &= \frac{\alpha_2^{mh_1}}{\rho_{\alpha_2}^m}\rho(A_{\omega_1})\ .\end{aligned}$$

Using the values of $\rho(A_{\omega_1})$ and $\rho(A_{\omega_2})$ from (5.1) and (5.2), and taking into account the normalization, we can rewrite :

$$\begin{aligned}\rho(A_{\omega_2}^{\alpha_1}) &= \frac{\alpha_1^{nh_2}}{\rho_{\alpha_1}^n}\rho(A_{\omega_2}^{\alpha_2})\frac{\rho_{\alpha_2}^n}{\alpha_2^{nh_2}} = \left(\frac{\alpha_1}{\alpha_2}\right)^{nh_2}\left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right)^n\\ \rho(A_{\omega_1}^{\alpha_2}) &= \frac{\alpha_2^{mh_1}}{\rho_{\alpha_2}^m}\rho(A_{\omega_1}^{\alpha_1})\frac{\rho_{\alpha_1}^m}{\alpha_1^{mh_1}} = \left(\frac{\alpha_2}{\alpha_1}\right)^{mh_1}\left(\frac{\rho_{\alpha_1}}{\rho_{\alpha_2}}\right)^m\ .\end{aligned}$$

Now, from the optimality assumption, that is ω_1 is optimal for α_1 and ω_2 for α_2, we know that $\rho(A_{\omega_2}^{\alpha_1}) \leq 1$ and $\rho(A_{\omega_1}^{\alpha_2}) \leq 1$, which leads to :

$$\begin{aligned}\left(\frac{\alpha_1}{\alpha_2}\right)^{nh_2}\left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right)^n &\leq 1\\ \left(\frac{\alpha_2}{\alpha_1}\right)^{mh_1}\left(\frac{\rho_{\alpha_1}}{\rho_{\alpha_2}}\right)^m &\leq 1\ ,\end{aligned}$$

or again, after simplification,

$$\begin{aligned}\left(\frac{\alpha_1}{\alpha_2}\right)^{h_2}\left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right) &\leq 1\\ \left(\frac{\alpha_2}{\alpha_1}\right)^{h_1}\left(\frac{\rho_{\alpha_1}}{\rho_{\alpha_2}}\right) &\leq 1\ .\end{aligned}$$

We can now conclude by noticing that we have

$$\left(\frac{\alpha_2}{\alpha_1}\right)^{h_1} \leq \left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right) \leq \left(\frac{\alpha_2}{\alpha_1}\right)^{h_2}\ ,$$

which is impossible for $\alpha_2 > \alpha_1$ and $h_1 > h_2$, because the left member is larger than the right member.

The case where ω_1 and/or ω_2 is infinite is the same, but for some details of notation. We rewrite the above proof, but this time assuming that w_1 is an infinite word. Let us denote by n the length of ω_2. In A_{ω_2}, there are nh_2

occurrences of A_1. According to our definition of optimality, this means that $\lim_{k\to\infty} \rho(A_{\omega_{1k}}^{\alpha_1})^{1/k} = 1 = \rho(A_{\omega_2}^{\alpha_2})$. We can now develop:

$$1 = \lim_{k\to\infty} \rho(A_{\omega_{1k}}^{\alpha_1})^{1/k} = \frac{\alpha_1^{h_1}}{\rho_{\alpha_1}} \lim_{k\to\infty} \rho(A_{\omega_{1k}})^{1/k} \tag{5.3}$$

$$1 = \rho(A_{\omega_2}^{\alpha_2}) = \frac{\alpha_2^{nh_2}}{\rho_{\alpha_2}^n} \rho(A_{\omega_2}) \tag{5.4}$$

Now, we may use the word ω_2 with the set associated to the value α_1 and vice-versa ω_1 with α_2. Taking into account the respective numbers of occurrences of A_1 in ω_1 and ω_2, this leads to:

$$\rho(A_{\omega_2}^{\alpha_1}) = \frac{\alpha_1^{nh_2}}{\rho_{\alpha_1}^n} \rho(A_{\omega_2})$$

$$\lim_{k\to\infty} \rho(A_{\omega_{1k}}^{\alpha_2})^{1/k} = \frac{\alpha_2^{h_1}}{\rho_{\alpha_2}} \lim_{k\to\infty} \rho(A_{\omega_{1k}})^{1/k}$$

Using the values of $\lim_{k\to\infty} \rho(A_{\omega_{1k}})^{1/k}$ and $\rho(A_{\omega_2})$ from (5.3) and (5.4), and taking into account the normalization, we can rewrite :

$$\rho(A_{\omega_2}^{\alpha_1}) = \frac{\alpha_1^{nh_2}}{\rho_{\alpha_1}^n} \frac{\rho_{\alpha_2}^n}{\alpha_2^{nh_2}} \rho(A_{\omega_2}^{\alpha_2}) = \left(\frac{\alpha_1}{\alpha_2}\right)^{nh_2} \left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right)^n$$

$$\lim_{k\to\infty} \rho(A_{\omega_{1k}}^{\alpha_2})^{1/k} = \frac{\alpha_2^{h_1}}{\rho_{\alpha_2}} \frac{\rho_{\alpha_1}}{\alpha_1^{h_1}} \lim_{k\to\infty} \rho(A_{\omega_{1k}}^{\alpha_1})^{1/k} = \left(\frac{\alpha_2}{\alpha_1}\right)^{h_1} \left(\frac{\rho_{\alpha_1}}{\rho_{\alpha_2}}\right)$$

Now, from the optimality assumption, that is ω_1 is optimal for α_1 and ω_2 for α_2, we know that $\rho(A_{\omega_2}^{\alpha_1}) \leq 1$ and $\lim_{k\to\infty} \rho(A_{\omega_{1k}}^{\alpha_2})^{1/k} \leq 1$, which leads to :

$$\left(\frac{\alpha_1}{\alpha_2}\right)^{nh_2} \left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right)^n \leq 1$$

$$\left(\frac{\alpha_2}{\alpha_1}\right)^{h_1} \left(\frac{\rho_{\alpha_1}}{\rho_{\alpha_2}}\right) \leq 1$$

and, just as for the finite case, after simplification, we can write

$$\left(\frac{\alpha_2}{\alpha_1}\right)^{h_1} \leq \left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right) \leq \left(\frac{\alpha_2}{\alpha_1}\right)^{h_2}$$

which is impossible for $\alpha_2 > \alpha_1$ and $h_1 > h_2$, because the left member is larger than the right member.

The case where both ω_1 and ω_2 are infinite can be dealt with in exactly the same way. $\square$

5.3.5 The value $\mu_1(\alpha)$ is unique

Theorem 5.12. *Words that achieve the joint spectral radius for the same α have the same 1-ratio. This means that $\mu_1(\alpha)$ consists of a single value.*

Proof. From Lemma 4.8, we know that if two words are optimal for a given value of the parameter α, then they are essentially equal. This implies that they (or their periodic infinite extensions) have the same 1-ratio. So, for a given value α^*, there is only one value for $\mu_1(\alpha^*)$. $\square$

Lemma 5.13. *If, for $\alpha_1 < \alpha_2$, the corresponding optimal words ω_1 and ω_2 have the same 1-ratio $\mu_1(\omega_1) = \mu_1(\omega_2)$, then they are essentially equal.*

Proof. Suppose that, for two values $\alpha_1 < \alpha_2$, we had optimal words ω_1 and ω_2 of the same 1-ratio $\mu_1(\omega_1) = \mu_1(\omega_2)$ but essentially different. This would mean that $\rho(A_{\omega_1}^{\alpha_1}) = \rho(A_{\omega_2}^{\alpha_2}) = 1$ and also that $\rho(A_{\omega_1}^{\alpha_2}) < 1$ and $\rho(A_{\omega_2}^{\alpha_1}) < 1$, since the words being essentially different, they cannot be optimal for the same value of α. Taking back the same computations as in the proof of Theorem 5.11, we get, with $r := \mu_1(\omega_1) = \mu_1(\omega_2)$:

$$\left(\frac{\alpha_2}{\alpha_1}\right)^r < \left(\frac{\rho_{\alpha_2}}{\rho_{\alpha_1}}\right) < \left(\frac{\alpha_2}{\alpha_1}\right)^r ,$$

which is again impossible. $\square$

5.3.6 Main result

Putting the previous results together, we state the following general result:

Theorem 5.14. *Consider the pair of matrices*

$$\left\{ A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\ A_1 = \alpha \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\} ,$$

where $\alpha \in \mathbb{R}$. Then, for any value α, there is one and only one optimal 1-ratio. Moreover, any balanced sequence with that ratio is optimal.

This balanced property allows for very fast and very precise computation of the JSR of the pair $\{A_0, A_1\}$ for any value of α, since the number of balanced words of length k grows polynomially with k (see [60, 5]). This polynomial order can also be guessed from the following observations about the grid approach to generate balanced words.

For a fixed length k, there are only $(k+1)$ possible 1-ratios (from 0 to 1, in $1/k$ steps). However , there could exist several essentially different balanced words for the same 1-ratio. But we can generate all of them (up to a cyclic permutation) by generating all possible lines through a grid in $\mathbb{N}^2$ of appropriate size $(k \times k)$. This method was already described in Section 5.1. When using lines $y = sx$, the word generated by two nearly equal slopes s are identical if these lines pass between the same integer points and cut the grid lines in the same order. In Figure 5.2, we show two pairs of lines illustrating this.

When continuously increasing the slope s, the generated word is modified only if the line $y = sx$ goes on the other side of a grid vertex. Consequently, the number of slopes to consider is not greater than the number of vertices in the grid. Moreover, the staircase-shaped green line on the same figure shows where we can stop checking the intersections, as we have reached k bits. The number of vertices involved drops down to the order $k^2/4$.

Consequently, knowing that the optimal product is balanced, we are able to compute the JSR using much longer products, yielding a higher precision for a lower computational cost.

5.3.7 Probable devil staircase property

We now use the following definition

Definition 5.15. *A devil staircase is a function defined on an interval* $[a, b]$ *having the following properties:*

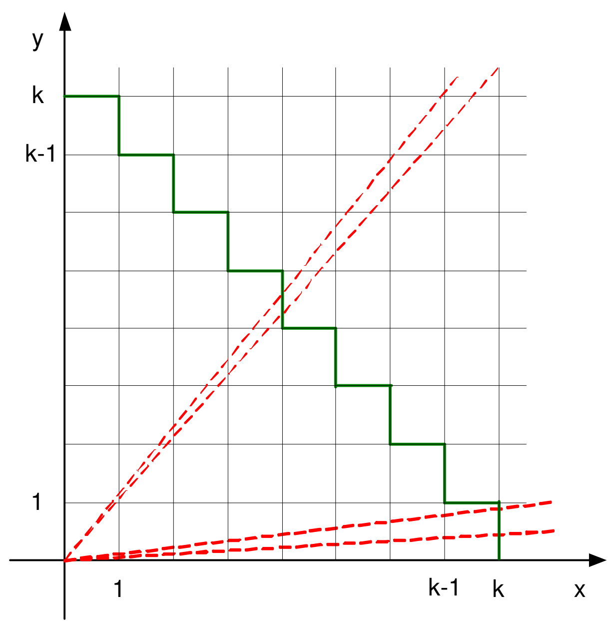

Fig. 5.2. The staircase (from top left to bottom right) shows where the lines of the form $y = sx$ cuts the coordinate grid for the k^{th} time. Each pair of dashed lines shows two different lines that generate the same binary sequence.

- *$f : [a, b] \to \mathbb{R}$ is continuous,*
- *f is non-decreasing on $[a, b]$,*
- *$f(a) < f(b)$,*
- *the derivative f' is 0 everywhere on $[a, b]$ except on a set of zero measure.*

This definition leads us to the following conjecture:

Conjecture 5.16. The curve $\mu_1(\alpha)$ is a devil staircase.

Although key points of the above definition have not been formally established, the plot on Figure 5.3 allows to believe that this conjecture holds. It remains

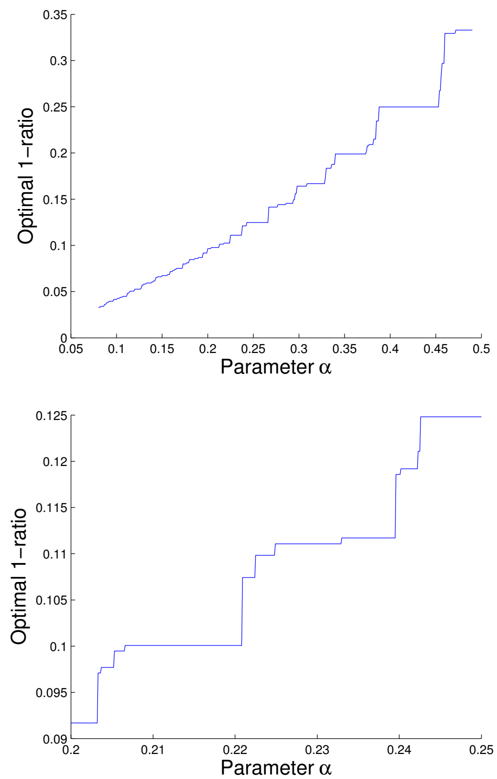

Fig. 5.3. Plot of the computer optimal 1-ratio as a function of the parameter α. The second graph presents a zoom into the first one. Plotting such a graph quickly leads to numerical limitations.

to be proved that:

- $\mu_1(\alpha)$ is continuous in α,

- the derivative in α of $\mu_1(\alpha)$ is zero everywhere except on a set of zero-measure.

We therefore conjecture that these two properties are true, hoping to provide a proof in a future publication.

5.4 Note about optimal words

We have already seen that, for the specific case $\alpha = 1$, the optimal matrix product is A_0A_1 and $\mu_1(1) = 1/2$. It is quite clear that, as α decreases, this optimal 1-ratio will decrease, until it reaches 0 for $\alpha = 0$. From a discussion with Nicola Guglielmi, it appears that he found the product A_0A_1 to be optimal for $\alpha \in [4/5, 1]$, with a rather numerical approach. Actually, by symmetry, it must also be optimal for $\alpha \in [1, 5/4]$, but we are only discussing the values $\alpha \leq 1$, precisely thanks to this symmetry. We will here formally prove the following.

Proposition 5.17. *Considering the pair of matrices expressed above,*

- *For* $\alpha \in [0.8, 1]$*, the product* A_0A_1 *is optimal.*
- *For* $\alpha < 0.8$*, the product* A_0A_1 *is not optimal.*

Before proving this statement, we provide an illustration of this behavior in Figure 5.4.

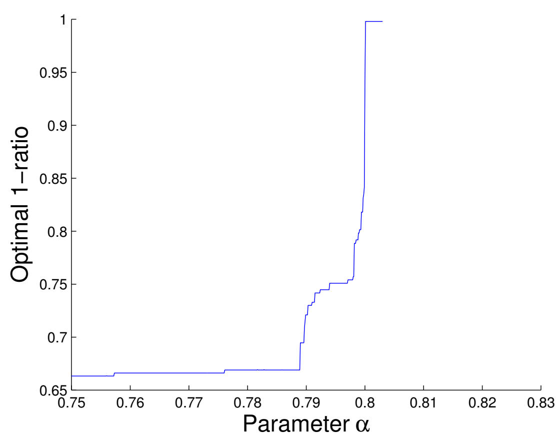

Fig. 5.4. Plot of the computer optimal 1-ratio as a function of the parameter α around the threshold value 0.8. This figure was obtained assuming the optimal words are balanced and considering words of length up to 1000. At $\alpha = 0.8$, a jump to 1 is observed.

Proof. To prove the optimality of the product $A_0 A_1$ when $\alpha \in [4/5, 1]$, we use a method that was shown to us by Vladimir Protasov. This method was elaborated in [65] and applied to the exact computation of the JSR and lower spectral radius for special $n \times n$ matrices arising in various problems in [68] and [70]. The implementation of the method for arbitrary 2×2 matrices is done in [76].

We have to consider normalized matrices, so that their JSR is equal to 1 if $A_0 A_1$ is optimal. We use the notation

$$A_0' = \frac{A_0}{\sqrt{\rho(A_0 A_1)}}, \ A_1' = \frac{A_1}{\sqrt{\rho(A_0 A_1)}} \ .$$

Keep in mind that these matrices still depend on α. More precisely,

$$\sqrt{\rho(A_0 A_1)} = \sqrt{\alpha}\frac{1+\sqrt{5}}{2} \ .$$

Further, we will be using the notation $\phi = \frac{1+\sqrt{5}}{2}$. This pair of matrices is irreducible, and consequently, under the optimality assumption, there exists a bounded convex set M (the interior of which contains the origin) such that

$$conv(A_0' M, A_1' M) \subseteq \underbrace{\rho(A_0', A_1')}_{=1} M \ .$$

This result appears in [65]. To study the optimality of the product $A_0' A_1'$, we will assume that we have M, and we will try to determine whether specific points lie on its border ∂M. We will se that this leads us to the level curve of an extremal norm.

Let us start with the leading (or *Perron-Frobenius*) eigenvector v of $A_1' A_0'$, which satisfies, by assumption

$$A_1' A_0' v = v \ .$$

We may assume v to lie on ∂M, because its length can be chosen in this manner. We then deduce that $A_0' v$ also lies on ∂M. Indeed, $A_1' A_0' v = v$ lies obviously on ∂M, and from the definition of an extremal norm, if $(A_0' v)$ did not lie on ∂M, so would not $A_1'(A_0' v)$. We have the situation depicted in Figure 5.5.

We can determine the formal expression of $A_1' A_0'$, as well as v. Indeed, we have

$$A_1' A_0' = \frac{\alpha}{\phi\sqrt{\alpha}} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \frac{1}{\phi\sqrt{\alpha}} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \frac{1}{\phi^2} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \ .$$

The spectral radius of this matrix is 1, by construction, and its leading eigenvector is $v = (1, \phi)^T$. We will denote this point by B:

$$OB = (1, \phi) \ .$$

Note that the second eigenvector of this matrix is orthogonal to OB. One can indeed check that $A'_1A'_0$ admits $(-\phi, 1)$ as an eigenvector, associated to the eigenvalue $\frac{2-\phi}{\phi^2}$.

The image of the point B by A'_0 will be denoted by C and is

$$OC = \frac{1}{\phi\sqrt{\alpha}}(1+\phi, \phi) = \frac{1}{\sqrt{\alpha}}(\phi, 1) \ .$$

We will use these points B and C as starting points for our construction.

Let us now assume that the segment BC is totally included in ∂M. We compute its image by the matrices A'_0 and A'_1. We have, introducing the new points D and E:

- $A'_0 : B \to C$, by construction,
- $A'_0 : C \to D$, with $OD = \frac{1}{\alpha}(\phi, 1/\phi)$,
- $A'_1 : B \to E$, with $OE = \sqrt{\alpha}(1/\phi, \phi)$,
- $A'_1 : C \to B$, by construction.

We then choose to extend the segments BE and CD so that they reach the axes respectively in N and K. The coordinates of these new points can be computed to be:

$$ON = (0, \phi\sqrt{\alpha}\frac{\phi-1}{\phi-\sqrt{\alpha}}) \ ,$$
$$OK = (\frac{\phi}{\sqrt{\alpha}}\frac{\phi-1}{\phi\sqrt{\alpha}-1}, 0) \ .$$

This defines a polygon in the positive quadrant, $ONBCK$, which is represented in Figure 5.6.

The question is now to determine for which values of α this polygon is mapped into itself by both matrices A'_0 and A'_1. To determine its image by A'_0, we observe that:

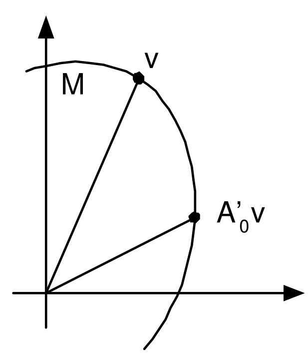

Fig. 5.5. Eigenvector v of $A'_1A'_0$ and its image by A'_0.

- $A_0' : N \to N'$, with $ON' = A_0'\, ON = (\frac{\phi-1}{\phi-\sqrt{\alpha}}, \frac{\phi-1}{\phi-\sqrt{\alpha}})$,
- $A_0' : B \to C$, by construction,
- $A_0' : C \to D$, by construction,
- $A_0' : K \to K'$, with $OK' = (\frac{1}{\alpha}\frac{\phi-1}{\phi\sqrt{\alpha}-1}, 0)$.

The polygon $ON'CDK'$ is the image of $ONBCK$. It is represented in Figure 5.7. The inclusion $ON'CDK' \subseteq ONBCK$, together with the inclusion we will see later, is a sufficient condition for $ONBCK$ to be the level curve of an extremal norm. In this aim, we have to check that N' and K' actually lie inside $ONBCK$. This leads us to the following conditions:

- for K', we need $|OK'| \leq |OK|$, because both lie on the x-axis. This holds if
$$\frac{1}{\alpha}\frac{\phi-1}{\phi\sqrt{\alpha}-1} \leq \frac{\phi}{\sqrt{\alpha}}\frac{\phi-1}{\phi\sqrt{\alpha}-1}\ ,$$
which simplifies to $\frac{1}{\phi} \leq \sqrt{\alpha}$, or again $\alpha \geq \frac{1}{\phi^2} \approx 0.382$.
- for N', we know that ON' has a $\pi/4$ angle with the x-axis. Moreover, we know that OB and OC are respectively above and below $\pi/4$. So, we just

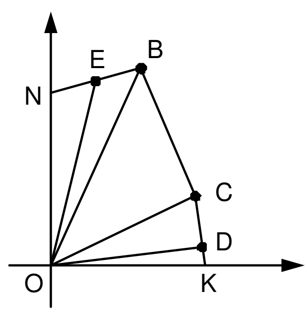

Fig. 5.6. Positions of the points B, C, D and E, and the extension to the points N and K to define a polygon.

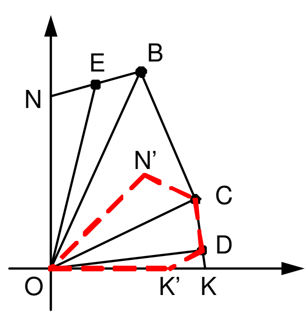

Fig. 5.7. Polygon $ON'CDK'$.

have to check that N' lies closer to the origin than the segment BC. Some calculus leads to the condition $\sqrt{\alpha} \leq \frac{\phi^2+1}{2\phi}$, or again $\alpha \leq 5/4$.

Now, we also have to determine the image of the polygon $ONBCK$ by A'_1. To do so, we observe that:

- $A'_1 : N \to N''$, with $ON'' = (0, \alpha\frac{\phi-1}{\phi-\sqrt{\alpha}})$,
- $A'_1 : B \to E$, by construction,
- $A'_1 : C \to B$, by construction,
- $A'_1 : K \to K''$, with $OK'' = (\frac{\phi-1}{\phi\sqrt{\alpha}-1}, \frac{\phi-1}{\phi\sqrt{\alpha}-1})$.

The polygon $ON''EBK''$ is the image of $ONBCK$. It is represented in Figure 5.8. The inclusion $ON''EBK'' \subseteq ONBCK$, together with the earlier

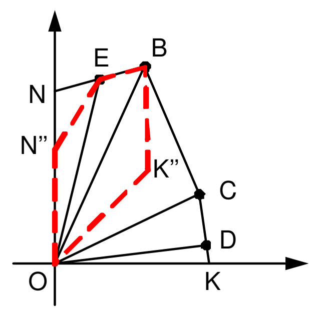

Fig. 5.8. Polygon $ON''EBK''$.

$ON'CDK' \subseteq ONBCK$, is a sufficient condition for us to have found the level curve of an extremal norm. In this aim, we have to check that N'' and K'' actually lie inside $ONBCK$. This leads us to the following conditions:

- for N'', we need $|ON''| \leq |ON|$, because both lie on the y-axis. This holds if
$$\alpha\frac{\phi-1}{\phi-\sqrt{\alpha}} \leq \phi\sqrt{\alpha}\frac{\phi-1}{\phi-\sqrt{\alpha}} \ ,$$
which simplifies to $\sqrt{\alpha} \leq \phi$, or again $\alpha \leq \phi^2 \approx 2.618$.
- for K'', we know that OK'' has a $\pi/4$ angle with the x-axis. Moreover, we know that OB and OC are respectively above and below $\pi/4$. So, we just have to check that K' lies closer to the origin than the segment BC. Some calculus leads to the condition $\sqrt{\alpha} \geq \frac{2\phi}{\phi^2+1}$, or again $\alpha \geq 4/5$.

Summing up all the above conditions, we know that the polygon $ONBCK$ is the level curve of an extremal norm when

$$4/5 \leq \alpha \leq 5/4 \ .$$

This symmetry is consistent with what we predicted earlier.

We now have to prove that, for any $\alpha < \frac{4}{5}$, one has $\rho(A_0'A_1') < \rho(A_0', A_1')$, so that the periodic sequence $(A_0'A_1')$ is no longer optimal. To do this, it suffices to show that no other extremal norm can be constructed for $\alpha < 4/5$, when assuming that $A_1'A_0'$ achieves the JSR. Let us again consider the pair of normalized matrices

$$\left\{A_0' = \frac{1}{\phi\sqrt{\alpha}}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, A_1' = \frac{\alpha}{\phi\sqrt{\alpha}}\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}\right\} .$$

Yet this time, if $\alpha < 4/5$, we only know that $\rho(A_0', A_1') \geq 1$, because $A_1'A_0'$ is not known to be optimal.

Let us assume $\rho(A_0', A_1') = 1$ even if $\alpha < 4/5$. We deduce that there again exists a convex body M such that v and $A_0 v \in \partial M$, with v the leading eigenvector of $A_1'A_0'$. As before, v is denoted by OB and $A_0'v$ is OC. We are now interested in the angle formed by OC and BC (denoted by 1 in the figure). This is depicted in Figure 5.9.

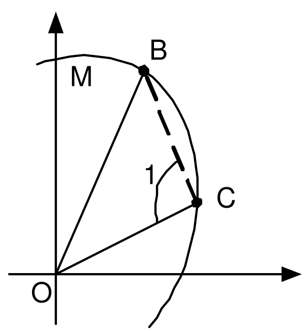

Fig. 5.9. The eigenvector v of $A_1'A_0'$ is at point B and its image by A_0' is point C. We are interested in the value of the angle between OC and BC.

The vector OC is $\frac{1}{\sqrt{\alpha}}(\phi, 1)$, while

$$CB = (1, \phi) - \frac{1}{\sqrt{\alpha}}(\phi, 1) = (1 - \frac{\phi}{\sqrt{\alpha}}, \phi - \frac{1}{\sqrt{\alpha}}) .$$

These vectors are found to be perpendicular when

$$
\begin{aligned}
\phi(1-\frac{\phi}{\sqrt{\alpha}})+1(\phi-\frac{1}{\sqrt{\alpha}}) &= 0 \\
\phi-\frac{\phi^2}{\sqrt{\alpha}}+\phi-\frac{1}{\sqrt{\alpha}} &= 0 \\
2\sqrt{\alpha}\phi-\phi^2-1 &= 0 \\
\alpha=\left(\frac{\phi^2+1}{2\phi}\right)^2 &= \frac{5}{4}\ .
\end{aligned}
$$

We look at orthogonality because we saw earlier that the second eigenvector of $A'_1A'_0$ is orthogonal to OB.

Similarly, BC is perpendicular to $OB=(1,\phi)$ if

$$
\begin{aligned}
1(1-\frac{\phi}{\sqrt{\alpha}})+\phi(\phi-\frac{1}{\sqrt{\alpha}}) &= 0 \\
1-\frac{\phi}{\sqrt{\alpha}}+\phi^2-\frac{\phi}{\sqrt{\alpha}} &= 0 \\
\sqrt{\alpha}-2\phi+\sqrt{\alpha}\ \phi^2 &= 0 \\
\alpha=\left(\frac{2\phi}{\phi^2+1}\right)^2 &= \frac{4}{5}\ .
\end{aligned}
$$

For values of $\alpha<4/5$, the angle between BC and OB is larger than $\pi/2$, as illustrated in Figure 5.10. In this case, we can decompose the vector OC into

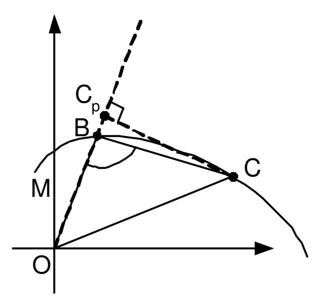

Fig. 5.10. Projection C_p of the point C onto the line OB.

OC_p+C_pC, where OC_p is longer that OB, because of the considered angle being larger than $\pi/2$. We know that $OC=\frac{1}{\sqrt{\alpha}}(\phi,1)$ is an eigenvector of $A'_0A'_1$ associated to the eigenvalue 1. This means that $(A'_0A'_1)^kOC=OC,\forall k$, even for $k\to\infty$. On the other hand, if we apply $(A'_1A'_0)$ several times on OB, we get a similar result, as $(A'_1A'_0)^kOB=OB,\forall k$, even for $k\to\infty$. Now, if we apply $(A'_1A'_0)^k$ to OC, we have

$$\begin{aligned}(A_1'A_0')^k OC &= (A_1'A_0')^k (OC_p + C_pC)\\ &= OC_p + (A_1'A_0')^k C_pC.\end{aligned}$$

The first term clearly comes from the fact that OB is an eigenvector of $(A_1'A_0')$ associated to the eigenvalue 1. And, as C_pC is orthogonal to OB, it is also an eigenvector of $(A_1'A_0')$, but this time associated to the eigenvalue $\frac{2-\phi}{\phi^2} \approx 0.146$. Consequently, $(A_1'A_0')^k C_pC \to 0$ as $k \to \infty$, and so OC is mapped outside the initial polygon. We can conclude that OB and OC cannot lie simultaneously on the level curve of an extremal norm, and so $A_1'A_0'$ is not optimal. This holds for any value of α such that the considered angle between OB and BC is larger than $\pi/2$, that is for $\alpha < 0.8$. The same argument can be used as well for $\alpha > 1.25$. This concludes the proof. □

5.5 Conclusion

This concludes this chapter devoted to the specific family that we fully studied:

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ \alpha \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\} .$$

For this pair, we proved that the binary words achieving the JSR are balanced, which a very strong property, allowing for an efficient computation of the JSR. An important question that remains open is to determine other classes of sets of matrices for which this property holds.

We then showed a construction of an extremal norm for this pair of matrices, using a geometrical method due to Protasov, and proved the optimality of the product A_0A_1 for (and only for) $\alpha \in [0.8, 1]$. From private communications, it appears that this last result was simultaneously proved by Guglielmi and Zennaro (no reference is available for their proof, at the time of writing).

6

Approximation of the JSR

In this chapter, we tackle the problem of computing the joint spectral radius of a set of matrices. The computability issue of the spectral radius of sets of matrices has been raised in [8] and [52]. And, more generally, the problem of computing approximations of the joint spectral radius is dealt with in a number of recent contributions (among others, [39, 56, 84, 65]). This problem is a hard one, as explained in Chapter 4. As an illustration, it is proved in [79] that, unless $P = NP$, there is no polynomial-time approximation algorithm for the spectral radius of two matrices. We have also shown that $\rho(\mathcal{M})$ is not an algebraic function on $\mathbb{R}^{mn^2}$, the space of m-tuples of $(n \times n)$ matrices. Consequently, we will only be trying to approximate the JSR, and not to compute it exactly.

For convenience, we will be considering bounded sets of matrices, so that both definitions of the JSR and GSR can be used. We will then be free to use either norms or spectral radii in our computations.

We first detail relatively simple bounds on the JSR, that can be derived from its definition. These yield what can be considered as a brute force method for the computation of the JSR. Although not very subtle, it gives good enough approximations. Then, we will refine this method by presenting observations made by Gripenberg and Maesumi, that allow for significant reductions of computation time. However, this does not change the global complexity of the computations.

We then prove a result of independent interest: the largest spectral radius of the matrices in the convex hull of $\mathcal{M} = \{A_1, \ldots, A_m\}$ is a lower bound on the spectral radius of $\mathcal{M}$:

$$\max_{0 \leq \lambda_i \leq 1,\, \sum \lambda_i = 1} \rho\Big(\sum_i \lambda_i A_i\Big) \leq \rho(\mathcal{M}).$$

After this, we study matrices consisting of only non-negative elements. In this specific case, we prove a simple bound for the spectral radius of sets of matrices that have *non-negative entries.* The spectral radius of the matrix S

whose entries are the componentwise maximum of the entries of the matrices in $\mathcal{M}$ satisfies

$$\frac{\rho(S)}{m} \leq \rho(\mathcal{M}) \leq \rho(S)$$

where m is the number of matrices in the set. While the upper bound on $\rho(\mathcal{M})$ is valid for bounded sets, the lower bound requires the set to be finite. This result does not depend on the size of the matrices. These are very simple to compute upper and lower bounds on the JSR, but not very sharp approximations.

The following approximation is based on the possible existence of a common quadratic Lyapunov function for the switched linear system generated by $\mathcal{M}$. This will be seen to be equivalent to the computation of an ellipsoid norm. This approximation has the advantage that it can be expressed as a convex optimization problem for which efficient algorithms exist. This approximation, denoted by ρ_L, satisfies

$$\frac{1}{\sqrt{n}}\,\rho_L \leq \rho \leq \rho_L$$

where n is the dimension of the matrices. This approximation has a polynomial-time cost, is easy to compute and is guaranteed to be within a factor $\sqrt{n}$ of the exact value. After establishing this result in 2003 and having it published ([18], also presented in [16, 17]), it has been brought to our attention that Ando & Shih had published similar results, yet with a different proof, in [1]. We present here our approach.

We then describe particular classes of matrices for which this approximation is equal to the joint spectral radius, such as symmetric matrices and triangular matrices. The problem of characterizing exactly the sets of matrices for which equality holds is a question that remains open. We then present a case where this approximation is proved to be different from the exact value of the JSR.

The approximation ρ_L is then extended by also considering products of matrices for the existence of a CQLF, instead of simply using the matrices from the set. After this, we generalize our CQLF method to express the computation of the JSR as the determination of an extremal norm.

The question remains open to find better approximations at a reasonable computational cost. In particular, the approximations presented here have relative errors that increase with the size or number of the matrices. In [9], Blondel & Nesterov have pursued this analysis and derived approximations of arbitrary accuracy, thus improving the guarantee on the approximation of the joint spectral radius. For completeness, we briefly present these results.

We conclude this chapter with the detailed study of the pair of matrices we introduced in Section 4.1. We try to understand what characterizes the matrix products that achieve the JSR, using concepts borrowed from discrete mathematics and the theory of words.

6.1 Theoretical bounds

6.1.1 Brute force

Referring back to Section 1.2.2, we recall the four-members inequality (1.17):

$$\rho_k(\mathcal{M})^{1/k} \leq \rho(\mathcal{M}) \leq \hat{\rho}(\mathcal{M}) \leq \hat{\rho}_k(\mathcal{M})^{1/k} \ ,$$

simplified thanks to $\hat{\rho}(\mathcal{M}) = \rho(\mathcal{M})$:

$$\rho_k(\mathcal{M})^{1/k} \leq \rho(\mathcal{M}) \leq \hat{\rho}_k(\mathcal{M})^{1/k} \ .$$

Then,

$$\rho_k(\mathcal{M})^{1/k} = \sup \left\{ \rho\left(\prod_{i=1}^{k} M_i \right) : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k \right\}^{1/k}$$

and

$$\hat{\rho}_k(\mathcal{M})^{1/k} = \sup \left\{ \left\| \prod_{i=1}^{k} M_i \right\| : M_i \in \mathcal{M} \text{ for } 1 \leq i \leq k \right\}^{1/k}$$

are respectively a lower bound and an upper bound for $\rho(\mathcal{M})$, for any k. The spectral radius can then easily be approximated to any desired accuracy. Indeed, these bounds can be evaluated for increasing values of k and lead to arbitrary close approximations of ρ. This can be shown thanks to the existence of an extremal norm, the existence of which we studied in Section 4.3. The rate of convergence of the upper bound can be given by the following result (which we use here for real matrices), taken from [84]:

Lemma 6.1 (Wirth). *Consider a bounded set $\mathcal{M} \subset \mathbb{R}^{n \times n}$.*

- *If there exists an extremal norm $\|.\|_*$ for $\mathcal{M}$, then there exists a constant $K > 0$, depending on $\mathcal{M}$, such that, $\forall k \geq 1$,*

$$\left| \frac{1}{k} \log \hat{\rho}_k(\mathcal{M}) - \log \rho(\mathcal{M}) \right| < K k^{-1} \ .$$

- *Otherwise, there exists a constant $K > 0$ such that, $\forall k \geq 1$,*

$$\left| \frac{1}{k} \log \hat{\rho}_k(\mathcal{M}) - \log \rho(\mathcal{M}) \right| < K \frac{1 + \log k}{k} \ .$$

Note that, in [84], there is no $1/k$ factor preceding $\log \hat{\rho}_k(\mathcal{M})$ in these inequalities. This is simply because we do not include the $1/k$ power into the definition of $\hat{\rho}_k(\mathcal{M})$, while it is done in [84]. We now justify the first point of this lemma, to underline the role played by extremal norms.

Proof (First point). Starting from $1 \leq \frac{\hat{\rho}_k(\mathcal{M})^{1/k}}{\rho(\mathcal{M})}$, we have, taking the logarithm and using the equivalence between norms (which guarantees the existence of the c constant used hereunder):

$$
\begin{aligned}
0 \leq \frac{1}{k}\log\hat{\rho}_k(\mathcal{M}) - \log\rho(\mathcal{M}) &= \frac{1}{k}\log \sup_{M_i\in\mathcal{M}} \left\| \prod_{i=1}^{k} M_i \right\| - \log\rho(\mathcal{M}) \\
&\leq \frac{1}{k}\log \sup_{M_i\in\mathcal{M}} c \left\| \prod_{i=1}^{k} M_i \right\|_* - \log\rho(\mathcal{M}) \\
&= \frac{1}{k}\log c + \frac{1}{k}\log \sup_{M_i\in\mathcal{M}} \left\| \prod_{i=1}^{k} M_i \right\|_* - \log\rho(\mathcal{M}) \ .
\end{aligned}
$$

The last two terms are equal because of the extremality property of the norm $\|.\|_*$. Using the notation $K = \log c$, we end up with

$$\frac{1}{k}\log\hat{\rho}_k(\mathcal{M}) - \log\rho(\mathcal{M}) \leq \frac{K}{k} \ .$$

□

However, this convergence is not monotonous in general, as there is no guarantee that, for any k,

$$\hat{\rho}_{k+1}(\mathcal{M}) \leq \hat{\rho}_k(\mathcal{M}) \ .$$

This is confirmed by our experience. We have seen that $\hat{\rho}_k(\mathcal{M})$, as well as $\rho_k(\mathcal{M})$, may have an oscillating behavior. This is illustrated in Figure 6.1 for a pair of matrices that exhibits both behaviors simultaneously:

$$\left\{ \begin{pmatrix} -1 & -\sqrt{3} \\ 0.9\sqrt{3} & -0.9 \end{pmatrix}, \begin{pmatrix} -0.9 & -0.9\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} \right\} \ .$$

We can, however, see that there is a weak monotonicity in k of these approximations. By this, we mean that a word of a length that is a multiple of an other word will yield an approximation that is at least as close to the JSR.

Lemma 6.2. *For any integers $k, l \geq 1$, we have*

$$\rho_k(\mathcal{M})^{1/k} \leq \rho_{kl}(\mathcal{M})^{1/kl} \leq \rho(\mathcal{M}) \leq \hat{\rho}_{kl}(\mathcal{M})^{1/kl} \leq \hat{\rho}_k(\mathcal{M})^{1/k} \ .$$

Proof (elementary). As was said earlier (in the proof of the four members inequality), all matrix products that can be expressed as the j^{th} power of a product of k matrices, are included in the set of products of kj matrices. Using the property (1.9), $\rho(A^n) = \rho(A)^n$ this leads to

$$\rho_k(\mathcal{M})^j \leq \rho_{kj}(\mathcal{M}) \ .$$

On the contrary, when a norm is involved, its submultiplicativity provides the reverse inequality:

$$\begin{aligned}\hat{\rho}_{kl}(\mathcal{M}) = \sup_{M_j} \|M_1 \dots M_{kl}\| &\leq \sup_{M_j} \prod_{i=0}^{l-1} \|M_{ik+1} M_{ik+2} \dots M_{ik+k}\| \\ &\leq \prod_{i=0}^{l-1} \sup_{M_j} \|M_1 M_2 \dots M_k\| = \hat{\rho}_k(\mathcal{M})^l \ .\end{aligned}$$

□

Although we have a guaranteed convergence of these basic approximations, they lead to expensive calculations. Indeed, the computational cost is exponential relatively to the desired precision.

6.1.2 Computational refinements

We now present various results that enable speeding up the computation of the JSR, based on the above brute force approach.

Maesumi

In [56], Maesumi reduces the number of products that appear in the naive computation of the lower bound ρ'_k described above. This is done using this property:

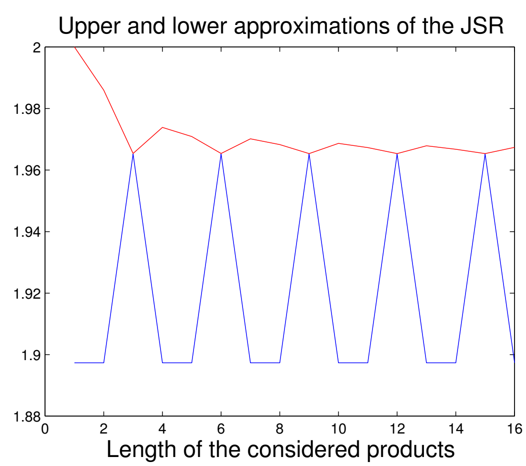

Fig. 6.1. Example of approximations $\hat{\rho}_k(\mathcal{M})$ and $\rho_k(\mathcal{M})$ of the joint spectral radius of a pair of matrices, for increasing k. From this picture, we can tell that there is a product of length 3 that is already close to optimal, as both approximations ρ_3 and $\hat{\rho}_3$ are very close.

Lemma 6.3. *Given a product of square matrices $M_1 \dots M_m$, its eigenvalues are invariant under cyclic permutations of the matrices. Hence its spectral radius is also unchanged.*

Proof. As we have seen in Section 1.1.2, the eigenvalues are the zeros of the characteristic polynomial. So, to prove the claim, we just have to prove that the characteristic polynomial of a matrix product is left unchanged by cyclic permutations. Moreover, all cyclic permutations can be considered as a commutation of two matrices. Indeed, the permutation

$$(M_1 \dots M_k)(M_{k+1} \dots M_m) \to (M_{k+1} \dots M_m)(M_1 \dots M_k)$$

is equivalent to $AB \to BA$, where we have posed

$$A = (M_1 \dots M_k),\ B = (M_{k+1} \dots M_m)\ .$$

So, it remains for us to prove that the characteristic polynomials of AB and BA are equal. Although this result is well-known, we present a proof of it, since we want this work to be as self-contained as possible. Let us consider the following relations, inspired by the *Schur complement*, where I and 0 respectively denote the $n \times n$ unit matrix and the $n \times n$ zero matrix :

$$\begin{pmatrix} \lambda I & A \\ B & I \end{pmatrix} \begin{pmatrix} I & 0 \\ -B & I \end{pmatrix} = \begin{pmatrix} \lambda I - AB & A \\ 0 & I \end{pmatrix}\ ,$$
$$\begin{pmatrix} I & A \\ B & \lambda I \end{pmatrix} \begin{pmatrix} I & -A \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & 0 \\ B & \lambda I - BA \end{pmatrix}\ .$$

Taking the determinants of those expressions, and keeping in mind that we are dealing with square matrices, we have:

$$\det \begin{pmatrix} \lambda I & A \\ B & I \end{pmatrix} \times 1 = \det \big(\lambda I - AB\big) \times 1\ ,$$
$$\det \begin{pmatrix} I & A \\ B & \lambda I \end{pmatrix} \times 1 = \det \big(\lambda I - BA\big) \times 1\ .$$

Both left members are equal, since

$$\begin{pmatrix} I & 0 \\ 0 & \lambda I \end{pmatrix} \begin{pmatrix} \lambda I & A \\ B & I \end{pmatrix} \begin{pmatrix} I/\lambda & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & A \\ B & \lambda I \end{pmatrix}\ ,$$
$$\lambda^n \det \begin{pmatrix} \lambda I & A \\ B & I \end{pmatrix} \lambda^{-n} = \det \begin{pmatrix} I & A \\ B & \lambda I \end{pmatrix}\ ,$$

and we obtain

$$\det(\lambda I - AB) = \det(\lambda I - BA)\ ,$$

which are respectively the characteristic polynomials of AB and BA. □

This property can be taken into account during the computation of the JSR, in order to avoid duplicate computation of cyclic permutations of the same product. As a product of length m admits m different cyclic permutations, this method provides a reduction of the number of products to consider by a factor m. This is even improved by eliminating the products that can be seen to be a power of a shorter one. As an example, $AABAAB \dots AAB$ of length $3k$ is a power of the product AAB, and they satisfy $\rho(AAB)^k = \rho(AABAAB \dots AAB)$ (as shown in Section 1.1.3).

The main contribution from [56] is to explicitly develop the needed combinatorics and number theory to precisely quantify the number of matrix products that can be neglected thanks to the above properties. In our perspective, what we have to remember from these results is that the total number of product to consider remains exponential, yet allowing for quicker computations.

Gripenberg

In [39], two aspects of the computation are treated. The first one is the characterization of the cases where the JSR is simply the largest spectral radius of the matrices in the set. This will be the object of Section 6.5.3. The second aspect consists in the presentation of an algorithm that yields successive intervals where the value of the JSR is guaranteed to lie. These intervals can me made arbitrarily small, thus providing arbitrary precision for the computation of the JSR. Using the following notations:

$$\Pi(X) = \prod_{i=1}^{m} A_i, \quad p(X) = \min_{1 \le j \le m} \left\| \prod_{i=1}^{j} A_i \right\|^{1/j} ,$$

with $X = (A_1, \dots, A_m) \in \mathcal{M}^m$, this result is proved:

Theorem 6.4 (Gripenberg). *Let $\mathcal{M}$ be a finite set of $(n \times n)$ matrices, with $n \ge 2$. Pick an arbitrary $\delta \ge 0$ and define:*

$$\begin{aligned} T_1 &= \mathcal{M} \ , \\ \alpha_1 &= \max_{A_i \in \mathcal{M}} \rho(A_i) \ , \\ \beta_1 &= \max_{A_i \in \mathcal{M}} \|A_i\| \ . \end{aligned}$$

And, for $m \ge 2$, recursively,

$$\begin{aligned} T_m &= \left\{ (X, A_i) \in \mathcal{M}^m | \ X \in T^{m-1}, A_i \in \mathcal{M}, p((X, A_i)) > \alpha_{m-1} + \delta \right\} \ , \\ \alpha_m &= \max \left\{ \alpha_{m-1}, \sup_{Y \in T_m} \rho\left(\Pi(Y)\right)^{1/m} \right\} \ , \\ \beta_m &= \min \left\{ \beta_{m-1}, \max \left\{ \alpha_m + \delta, \sup_{Y \in T_m} p(Y) \right\} \right\} \ . \end{aligned}$$

Then $\alpha_m \leq \rho(\mathcal{M}) \leq \beta_m, \forall m \geq 1,$ *and* $\lim_{m\to\infty}(\beta_m - \alpha_m) \leq \delta.$

The trick here is to save on the number of products to consider by eliminating those that are not likely to achieve the best bounds. This is done in the construction process of T_m.

Although this result has the interesting property of allowing the user to determine the eventual precision, no rate of convergence is proved. That is to say that there is no guarantee on the computational cost of the method. Moreover, this convergence rate depends on the choice of the norm in $p(X)$. This choice should itself be dependent on the set $\mathcal{M}$. This issue can somehow be compared with the determination of an extremal norm, which is also a hard task. Even the choice of the δ is not trivial, since it has implications on the behavior of the algorithm.

Despite these shortcomings, numerical experiments carried out by Gripenberg tend to show that this method can deliver good approximations while evaluating a reasonable number of matrix products.

6.2 The value $\tilde{\rho}$

In the context of studying different ways of approaching the JSR, we define $\tilde{\rho}(\mathcal{M})$ as the largest spectral radius of the convex combinations of the matrices in the set $\mathcal{M}$.

Definition 6.5. *Let* $\mathcal{M} = \{A_1, \dots, A_m\}$.

$$\tilde{\rho}(\mathcal{M}) := \max_{\sum_{i=0}^{m} \alpha_i = 1,\ \alpha_i \geq 0} \rho\left(\sum \alpha_i A_i\right).$$

6.2.1 Bounds for $\tilde{\rho}$

To understand what can be expected from this value $\tilde{\rho}$, let us give it some bounds.

Lower bound

Directly from the definition, $\tilde{\rho}(\mathcal{M})$ is at least as large as the largest spectral radius of the matrices in the set:

$$\tilde{\rho}(\mathcal{M}) \geq \max_i(\rho(A_i)), A_i \in \mathcal{M} .$$

To achieve this, just choose $\alpha_i = 1$ for the corresponding i.

Upper bound

A bit more interestingly, we now prove that $\tilde{\rho}(\mathcal{M})$ is a lower bound for the JSR.

Proposition 6.6. *Let* $\mathcal{M} = \{A_1, \dots, A_m\}$. *Then*

$$\tilde{\rho}(\mathcal{M}) := \max_{\sum_{i=0}^m \alpha_i = 1,\ \alpha_i \geq 0} \rho\left(\sum \alpha_i A_i\right) \leq \rho(\mathcal{M}).$$

Proof. We have, using the inequality between the spectral radius and a norm $\rho(.) \leq \|.\|$, along with the triangular inequality,

$$\begin{aligned} \forall \alpha_i \geq 0,\ \rho\left(\sum_i \alpha_i A_i\right) &\leq \left\|\sum_i \alpha_i A_i\right\| \\ &\leq \sum_i \|\alpha_i A_i\| = \sum_i \alpha_i \|A_i\| \\ &\leq \max_i (\|A_i\|) \quad , \text{as } \textstyle\sum_i \alpha_i = 1 \ . \end{aligned}$$

Now, we know from Proposition 4.17 that, $\forall \varepsilon$, there exists a norm $\|.\|_*$ such that $\forall i, \|A_i\|_* < \rho(\mathcal{M}) + \varepsilon$. So, if $\rho(\mathcal{M}) < 1$, then there exists a norm $\|.\|_*$ such that $\max_i \|A_i\|_* < 1$. We can then deduce from the previous inequality that $\tilde{\rho}(\mathcal{M}) < 1$. In short,

$$\rho(\mathcal{M}) < 1 \Rightarrow \tilde{\rho}(\mathcal{M}) < 1 \ , \tag{6.1}$$

and we deduce, using linearity and a *normalization method* (see note below), that $\tilde{\rho}(\mathcal{M}) \leq \rho(\mathcal{M})$.

Normalization method using linearity:

Let us now consider a set of matrices $\mathcal{M}$, which we normalize using a value $y > \rho(\mathcal{M})$ so that $\rho(\mathcal{M}/y) < 1$. Relation (6.1) implies $\tilde{\rho}(\mathcal{M}/y) < 1$ and so

$$\tilde{\rho}(\mathcal{M}) < y \ .$$

To sum up, we have that, for any set $\mathcal{M}$ of matrices, for any y such that $\rho(\mathcal{M}) < y$,$\tilde{\rho}(\mathcal{M}) < y$. This proves that

$$\rho(\mathcal{M}) \geq \tilde{\rho}(\mathcal{M}) \ .$$

This reasoning will be referred later as *linearization method* or simply *linearity.*

□

As a particular case, we also have the following relation:

Lemma 6.7. *For any finite set* $\mathcal{M} = \{A_1, \dots, A_m\}$,

$$\rho(\mathcal{M}) \geq \frac{1}{m}\rho(A_1 + A_2 + \cdots + A_m) \ .$$

We prove below, by providing an example, that there are pairs of 2×2 matrices for which the inequality in Proposition 6.6 is strict.

Example 6.8. Consider the pair

$$A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \ .$$

We can see that

- $\rho(A_0) = \rho(A_1) = 1$,
- $\rho(A_0, A_1) = 1.6180$ (this value will be justified further),
- $\tilde{\rho} = 1.5$ (for $\alpha = 0.5$).

Here, we clearly have $\tilde{\rho}(A_0, A_1) < \rho(A_0, A_1)$.

6.2.2 $\tilde{\rho}$ and stability

Sufficient condition

The condition $\tilde{\rho}(\mathcal{M}) < 1$ is not sufficient for the system to be stable. Otherwise, we would have $\tilde{\rho}(\mathcal{M}) < 1 \ \Rightarrow \rho(\mathcal{M}) < 1$, meaning that $\tilde{\rho}(\mathcal{M}) \geq \rho(\mathcal{M})$ (by linearity). Together with the converse inequality proved above, this would mean $\tilde{\rho}(\mathcal{M}) = \rho(\mathcal{M})$, which is false.

Necessary condition

If a system is stable, then $\rho(\mathcal{M}) < 1$ (see Section 3.2.1). And so, from Proposition 6.6, $\tilde{\rho}(\mathcal{M}) < 1$.

As we will see later, the value $\tilde{\rho}(\mathcal{M})$ is not necessarily a close approximation of $\rho(\mathcal{M})$. But its evaluation prior to computing the JSR can be sufficient to establish instability.

6.3 Matrices with non-negative entries

In this section, we introduce a way of approximating the joint spectral radius of a set of matrices with non-negative entries. To be complete, we must mention that the results here can be extended to sets of matrices that leave a cone invariant (see [9]). Clearly, positive matrices are a particular case of this, since they leave $\mathbb{R}^n_+$ invariant. The result we present here will be obtained by using Proposition 6.6.

Theorem 6.9. *Let $\mathcal{M} = \{A_1, \dots, A_m\}$ be a set of matrices with non-negative entries and define $S_{ij} = \max_{1 \leq k \leq m} (A_k)_{ij}$. We have*

$$\frac{\rho(S)}{m} \leq \rho(\mathcal{M}) \leq \rho(S). \tag{6.2}$$

Proof. For matrices M_1, M_2 with non-negative entries such that $M_2 \geq M_1$ componentwise, it is known that $\rho(M_2) \geq \rho(M_1)$ ([36], Vol.2, chap. XIII). Knowing this, and using Proposition 6.6, we can write

$$S \leq \sum_{k=1}^{m} A_k \Rightarrow \rho(S) \leq \rho\left(\sum_{k=1}^{m} A_k\right) = m\rho\left(\frac{\sum_{k=1}^{m} A_k}{m}\right) \leq m\rho(\mathcal{M}) \ .$$

This proves the first inequality of (6.2).

To prove the second inequality, we see that, as the elements are non-negative, for any sequence $\omega = (\omega_1, \dots, \omega_k)$ of k indices, the product $A_\omega = A_{\omega_1} \dots A_{\omega_k}$ satisfies $(S^k)_{ij} \geq (A_\omega)_{ij}$. As above, this allows us to deduce, $\forall \omega : |\omega| = k$,

$$S^k \geq A_\omega \Rightarrow \limsup_{k \to \infty} \|S^k\|^{1/k} \geq \limsup_{k \to \infty} \left(\max_{\omega : |\omega| = k} \|A_\omega\|^{1/k}\right) \ .$$

This proves that $\rho(S) \geq \rho(\mathcal{M})$. ☐

Corollary 6.10. *For a set $\mathcal{M} = \{A_1, \dots, A_m\}$ of matrices with non-negative entries,*

$$\rho(\mathcal{M}) \leq \rho\left(\sum_{k=1}^{m} A_k\right) \ .$$

Note that there is no better constant value than the factor m for the bound $\rho(S) \leq m\rho(\mathcal{M})$, as the equality $\rho(S)/m = \rho(\mathcal{M})$ can be realized for particular matrices. For $m = 2$, consider the following example.

Example 6.11.

$$\mathcal{M} = \left\{ A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\} \ ,$$

for which

$$\rho(S) = \rho\left[\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right] = 2 \ .$$

We just have to see that $A^2 = A$, $B^2 = B$, $AB = B$ and $BA = A$. This proves that any product generated by $\{A, B\}$ is either A or B. And $\rho(A) = \rho(B) = 1$ implies $\rho(A, B) = 1$. This concludes the example as we have $\rho(S) = 2\rho(\mathcal{M})$.

There exists a similar construction for the cases $m \geq 3$.

6.4 The value ρ_L, alias CQLF or ellipsoid norm approximation

For the clarity of the presentation, the value $\rho_L(\mathcal{M})$ will here be defined and discussed for pairs of matrices. Our results remain valid for finite sets of matrices in general, by just extending the notations to the appropriate (finite) number of matrices.

Let us first recall the condition for the existence of a common quadratic Lyapunov function, which was (referring to Section 3.4.1)

$$\exists P \succ 0: \; A_i^T P A_i - P \prec 0, \forall A_i \in \mathcal{M} \; ,$$

where the notations $\succ, \succeq, \prec, \preceq$ stand for positive and negative (semi) definite. We may also use the notation A^* instead of A^T, to keep our developments valid for a possible use of the results for complex matrices.

We now define

$$\rho_L(\mathcal{M}) := \inf\{\lambda \in \mathbb{R}^+ \text{ such that } \exists P \succ 0: \begin{array}{l} A_0^* P A_0 \preceq \lambda^2 P \\ A_1^* P A_1 \preceq \lambda^2 P \; .\} \end{array} \tag{6.3}$$

In the above conditions, we use $\preceq$, but we know that using $\prec$ would give the same value for ρ_L, because we take the inf on all λ satisfying the constraints. We use $\mathbb{R}^+$ in the definition, because the case $\rho_L = 0$ is trivial : $\mathcal{M}$ would then consist in zero matrices.

As these are a generalization of the Lyapunov conditions, we can already tell that for $\rho_L(M) \leq 1$, the system admits a CQLF and so is stable.

6.4.1 Bounds for ρ_L

We now detail what can be said about this value. From the definition, we express bounds for it in terms of the spectral radius and singular values of the matrices A_0 and A_1.

Lower bound

Considering again the inequality $A_0^* P A_0 \preceq \lambda^2 P$, we take an eigenvector V_{A_0} of A_0, associated to the eigenvalue λ_{A_0}, and we multiply the condition on the left by $V_{A_0}^*$ and on the right by V_{A_0}, which yields

$$\begin{aligned} V_{A_0}^* A_0^* P A_0 V_{A_0} &\preceq \lambda^2 V_{A_0}^* P V_{A_0} \\ \lambda_{A_0}^* \lambda_{A_0} (V_{A_0}^* P V_{A_0}) &\preceq \lambda^2 (V_{A_0}^* P V_{A_0}) \end{aligned}$$

As P is positive definite, we obtain $\lambda_{A_0}^* \lambda_{A_0} \leq \lambda^2$ for any eigenvalue λ_{A_0} of A_0. This inequality holding for any $P \succ 0$, ρ_L must be greater than the largest modulus of the eigenvalues of A_0.

$$\rho_L{}^2 \geq \rho(A_0)^2$$

The very same holds for A_1 and we also have $\rho_L{}^2 \geq \rho(A_1)^2$. So,

$$\rho_L \geq \max\left(\rho(A_0), \rho(A_1)\right) \ .$$

And, more generally

$$\rho_L \geq \max_i \left(\rho(A_i)\right) \ . \tag{6.4}$$

Upper bound

We can get an immediate upper bound by choosing to use the identity matrix as an acceptable P in the definition of ρ_L . It is indeed positive definite. The inequality becomes

$$A_0 A_0^* \preceq \lambda^2 I \ ,$$

meaning that $\sigma(A_0)^2$, the square of the largest singular value of A_0, is an acceptable value for $\rho_L{}^2$. The same argument with $P = I$ for A_1 leads to $\rho_L{}^2 \geq \sigma(A_1)^2$. We then deduce that $\lambda = \max(\sigma(A_0), \sigma(A_1))$ is always an acceptable value for ρ_L and so, the actual value of ρ_L satisfies:

$$\rho_L \leq \max\left(\sigma(A_0), \sigma(A_1)\right) \ .$$

We generalize to

$$\rho_L \leq \max_i \left(\sigma(A_i)\right) \ .$$

6.4.2 Illustration

On figure 6.2, the values of the different approximations of the joint spectral radius are presented for the following pair of matrices, as a function of the parameter a:

$$A_1 = \begin{pmatrix} -0.2 & -0.4 \\ 0.4 & -0.2 \end{pmatrix}, A_2 = \begin{pmatrix} -0.2 & -0.4a \\ 0.4/a & -0.2 \end{pmatrix} \ .$$

Without resorting to numerical computations, the same example as before can be solved formally.

Example 6.12. Consider the pair

$$A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \ .$$

We have

- $\rho(A_0) = \rho(A_1) = 1$,

- $\tilde{\rho}(A_0, A_1) = 1.5$ (for $\alpha = 0.5$),
- $\sigma(A_0) = \sigma(A_1) = \frac{1+\sqrt{5}}{2} \Rightarrow \rho_L(A_0, A_1) \leq \frac{1+\sqrt{5}}{2} \approx 1.6180$,
- $\rho(A_0, A_1) = \frac{1+\sqrt{5}}{2}$

In this example, $A_1 = A_0^T$, so $\rho(A_0A_1) = \rho(A_0A_0^T)$, and it can be shown (see Proposition 6.20) that $\rho(A_0, A_1) = \rho(A_0A_1)^{1/2}$ in this particular case. This implies that

$$\rho(A_0, A_1) = \sigma(A_0) \ .$$

6.4.3 ρ_L and stability

We now study the link between the stability of a switched linear system and the value of the associated ρ_L. As can be expected from the above comments, $\rho_L = 1$ will be a threshold value.

Sufficient condition

Let us suppose that $\rho_L < 1$. As equation (6.4) shows, if a matrix A_i satisfies $A_i^* P A_i \leq \lambda^2 P$ for a definite positive P, then $\rho(A_i)^2 \leq \lambda^2$. This leads to the implication

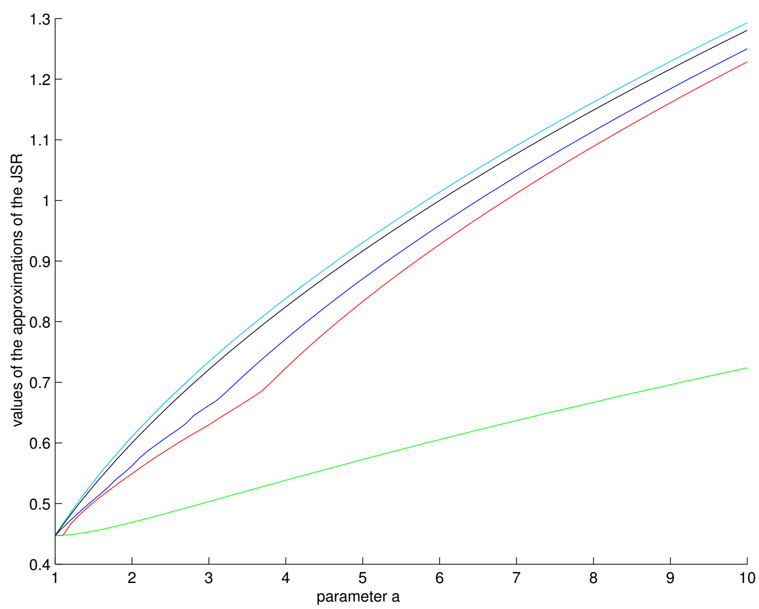

Fig. 6.2. Different approximations of the joint spectral radius. Blue and red curves show the upper and lower bounds on the actual value of the JSR, far underneath is $\tilde{\rho}$, and above are bounds for ρ_L.

$$\rho_L < 1 \Rightarrow \begin{cases} \rho(A_0)^2 < 1 \ , \\ \rho(A_1)^2 < 1 \ . \end{cases}$$

Moreover, for any product X of the matrices A_0 and A_1, denoted by a function $h : \mathbb{N} \to \{0,1\}$, we have

$$\begin{aligned} X^* P X &= \Big(\prod_{i=1}^{n} A_{h(i)}\Big)^* P \Big(\prod_{i=1}^{n} A_{h(i)}\Big) \\ &= \Big(\prod_{i=2}^{n} A_{h(i)}\Big)^* \underbrace{A^*_{h(1)} P A_{h(1)}}_{\leq {\rho_L}^2 P} \Big(\prod_{i=2}^{n} A_{h(i)}\Big) \\ &\leq \dots \\ &\leq {\rho_L}^{2n} P \end{aligned}$$

And so, $\rho(X)^2 \leq {\rho_L}^{2n} < 1$. This is sufficient to deduce that $\rho(A_0, A_1) < 1$. We can express this implication as

$$\rho_L(A_0, A_1) < 1 \Rightarrow \rho(A_0, A_1) < 1 \ . \tag{6.5}$$

Using the linearity property, as in the normalization method detailed earlier, we get that, for any pair $\{A_0, A_1\}$ of matrices,

$$\rho_L(A_0, A_1) \geq \rho(A_0, A_1) \ .$$

Or again,

$$\rho_L(\mathcal{M}) \geq \rho(\mathcal{M}) \ . \tag{6.6}$$

Necessary condition

Common quadratic Lyapunov functions (CQLF) do not fully characterize the stability of switched systems. There are indeed stable systems for which there does not exist a CQLF. Such an example has already been detailed in Section 3.4.3. Consequently, the condition $\rho_L < 1$ is not necessary for stability.

So, we want to define another approximation of the JSR that is "better" than (6.3). By *better*, we mean another upper bound on the JSR that is less than ρ_L, so that this approximation can be < 1 while $\rho_L \geq 1$ in a stable system. An idea can be to consider products of matrices A_i in the definition (6.3) of our approximation, rather than restricting ourselves to purely A_0 and A_1. This will be done in Section 6.5.5.

6.5 Interpretation of ρ_L in terms of matrix norms

We have seen earlier that extremal norms are closely related to the notion of joint spectral radius.

We saw in Theorem 4.15 that irreducible sets $\mathcal{M}$ of matrices admit a norm $\|.\|_*$ having the following properties:

1. $\|A_i x\|_* \leq \rho(\mathcal{M})\ \|x\|_*$, $\forall x$ and $\forall A_i \in \mathcal{M}$;
2. $\forall x, \exists A_i \in \mathcal{M} : \|A_i x\|_* = \rho(\mathcal{M})\ \|x\|_*$.

Considering the matrix norm induced by this vector norm, and also denoting it by $\|.\|_*$, we have

$$\|A_i\|_* \leq \rho(\mathcal{M}), \forall i\ . \tag{6.7}$$

As previously discussed in Section 4.3.3 (Proposition 4.17), the joint spectral radius can be seen as the infimum over all possible matrix norms of the largest norm of the matrices in the set. We already saw that Kozyakin had described the theoretical construction of such an extremal norm. But it is practically useless, because it is not explicit and partly relies on the a-priori knowledge of the numerical value of the joint spectral radius. And one can of course not hope enumerating all possible matrix norms in order to compute the joint spectral radius. We can however enumerate particular sets of norms. We now show that the value ρ_L defined previously does just that.

6.5.1 The ellipsoid norm approximation : definition

Let us define an approximation of the joint spectral radius that would be obtained by finding, among all ellipsoid norms $\|\cdot\|_P$, one that minimizes $\max_i \|A_i\|_P$.

Before proceeding, we briefly recall the definition of the ellipsoid (or quadratic) norms. Let P be a symmetric positive definite matrix, $P \succ 0$. The vector P-norm is defined as $\|x\|_P = \sqrt{x^T P x}$. Associated to this vector norm, there is an induced matrix norm:

$$||A_i||_P = \max_{|x|\neq 0} \frac{\|A_i x\|_P}{\|x\|_P} = \max_{|x|\neq 0} \frac{\sqrt{x^T A_i^T P A_i x}}{\sqrt{x^T P x}}\ . \tag{6.8}$$

Note that we use the notation $\|.\|_P$ for both the vector and matrix norms, as there is no ambiguity regarding which one is applied. Let us now define the *ellipsoid norm approximation* of the joint spectral radius by:

$$\rho_P(\mathcal{M}) = \inf_{P\succ 0} \max_{A_i\in\mathcal{M}} \|A_i\|_P.$$

The infimum on ellipsoid norms cannot be lower than the infimum on all possible norms and so it immediately follows that

$$\rho(\mathcal{M}) \leq \rho_P(\mathcal{M})\ .$$

The ellipsoid norm approximation can be computed as follows. First, its definition (6.8) implies that

$$\begin{aligned}\forall x,\ \sqrt{x^T A_i^T P A_i x} &\leq \|A_i\|_P \sqrt{x^T P x}\\ \forall x,\ x^T A^T P A x &\leq \|A\|_P^2\ x^T P x\\ \forall x,\ x^T (A_i^T P A_i - \|A_i\|_P^2 P) x &\leq 0\\ A_i^T P A_i - \|A_i\|_P^2 P &\preceq 0\ .\end{aligned}$$

One can therefore think of $\|A_i\|_P$ as the smallest scalar value γ for which $A_i^T P A_i \preceq \gamma^2 P$ for some $P \succ 0$. The ellipsoid norm approximation of a set $\mathcal{M} = \{A_1, \ldots, A_m\}$ is thus equal to the smallest scalar γ for which there is a solution $P \succ 0$ to $A_i^T P A_i \preceq \gamma^2 P$, $\forall i$. This is exactly the same definition as (6.3) and so

$$\rho_P(\mathcal{M}) = \rho_L(\mathcal{M})\ .$$

From now on, the denomination *ellipsoid norm approximation* will also designate the value ρ_L.

6.5.2 Quality of the approximation

A natural question to ask is how good this approximation is in the general case. In the next section, we describe situations for which the approximation is equal to the joint spectral radius, and we provide an example for which the approximation is larger than the joint spectral radius.

6.5.3 The ellipsoid norm approximation : special cases

We prove in this section that the joint spectral radius and the ellipsoid norm approximation are equal (and are equal to the maximal spectral radius of the matrices in the set) in the following situations: all matrices are symmetric, all matrices are triangular or, more generally, the Lie algebra associated to the matrices is solvable. We close the section with an example for which the spectral radius and its approximation are different. We start with the case of symmetric matrices:

Proposition 6.13. *For a set of symmetric matrices, the joint spectral radius and its ellipsoid norm approximation are equal. Their value is the largest spectral radius of the matrices in the set.*

Proof. Using the identity I as matrix P, we get

$$\|A_i\|_I = \inf\{\gamma : A_i^2 \preceq \gamma^2 I\}\ ,$$

which implies $\|A_i\|_I = \rho(A_i)$. And knowing that $\rho(A_i) \leq \inf_{P \succ 0} \|A_i\|_P$, we actually have the equality $\rho(A_i) = \inf_{P \succ 0} \|A_i\|_P$, achieved for $P = I$. Finally, we have

$$\rho_L(\mathcal{M}) = \inf_{P \succ 0} \max_i \|A_i\|_P = \max_i \rho(A_i)\ .$$

□

In order to derive our result for triangular matrices, we first establish a discrete-time analog to a continuous-time result by Liberzon, Hespanha & Morse ([53]) on the existence of a common quadratic Lyapunov function for switched linear systems.

Lemma 6.14. *Let $\mathcal{M}$ be the set $\{A_1, \ldots, A_m\}$ and consider the discrete-time switched linear system*

$$x_{k+1} = A_{i_k}\, x_k \quad A_{i_k} \in \mathcal{M}.$$

If the system is stable and the matrices are upper-triangular, then there exists a common quadratic Lyapunov function in the form of a diagonal matrix.

Proof. Let $\{A_i, \ldots, A_m\}$ be a set of upper-triangular (possibly complex) matrices and P the candidate Lyapunov function (diagonal, real):

$$A_i = \begin{pmatrix} a^i_{11} & a^i_{12} & \ldots & a^i_{1n} \\ 0 & a^i_{22} & \ldots & a^i_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a^i_{nn} \end{pmatrix}, \quad P = \begin{pmatrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & p_n \end{pmatrix}, p_k > 0, \forall k \ .$$

For P to be a Lyapunov function of $x_{k+1} = A_i x_k$ (fixed A_i), the following relation has to hold:

$$P - A_i^* P A_i \succ 0 \ .$$

Developing $P - A_i^* P A_i$, we get:

$$\begin{pmatrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & p_n \end{pmatrix} - \begin{pmatrix} {a^i_{11}}^* & 0 & \ldots & 0 \\ {a^i_{12}}^* & {a^i_{22}}^* & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ {a^i_{1n}}^* & {a^i_{2n}}^* & \ldots & {a^i_{nn}}^* \end{pmatrix} \begin{pmatrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & p_n \end{pmatrix} \begin{pmatrix} a^i_{11} & a^i_{12} & \ldots & a^i_{1n} \\ 0 & a^i_{22} & \ldots & a^i_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a^i_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & p_n \end{pmatrix} - \begin{pmatrix} {a^i_{11}}^* p_1 & 0 & \ldots & 0 \\ {a^i_{12}}^* p_1 & {a^i_{22}}^* p_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ {a^i_{1n}}^* p_1 & {a^i_{2n}}^* p_2 & \ldots & {a^i_{nn}}^* p_n \end{pmatrix} \begin{pmatrix} a^i_{11} & a^i_{12} & \ldots & a^i_{1n} \\ 0 & a^i_{22} & \ldots & a^i_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a^i_{nn} \end{pmatrix}$$

which yields

$$\begin{pmatrix} (1 - |{a^i}_{11}|^2) p_1 & -{a^i_{11}}^* a^i_{12} p_1 & \ldots \\ -a^i_{11} {a^i_{12}}^* p_1 & -|{a^i}_{12}|^2 p_1 + (1 - |{a^i}_{22}|^2) p_2 & \ldots \\ \vdots & \vdots & \ddots \\ -a^i_{11} {a^i_{1n}}^* p_1 & -a^i_{12} {a^i_{1n}}^* p_1 - a^i_{22} {a^i_{2n}}^* p_2 & \ldots \end{pmatrix} . \tag{6.9}$$

This matrix is Hermitian, and so its leading principal minors are real (see [45]).

As A_i is assumed to be stable, $a^i_{jj} < 1, \forall j$. The first diagonal element in (6.9) is therefore positive, for any value of p_1. Let it be chosen as 1. Moreover, the value of p_2 can be chosen in such a way that the (2×2) leading principal minor is positive. Indeed, p_2 only appears in its last diagonal element, and its coefficient $(1 - |a^i{}_{22}|^2)$ is positive, as $a^i{}_{22} < 1$. So, taking p_2 such that

$$\begin{vmatrix} (1 - |a^i{}_{11}|^2) & -a^i_{11}{}^* a^i_{12} \\ -a^i_{11} a^i_{12}{}^* & -|a^i{}_{12}|^2 + (1 - |a^i{}_{22}|^2) p_2 \end{vmatrix} > 0$$

is possible, and simple developments give the following condition:

$$p_2 > \frac{1}{(1 - |a^i{}_{22}|^2)} \left[\frac{\left(|a^i_{11}||a^i_{12}|\right)^2}{(1 - |a^i{}_{11}|^2)} + |a^i{}_{12}|^2 \right] \ .$$

We can define in this way a p_2 that satisfies this for all matrices A_i of the set by choosing

$$p_2 > \max_i \frac{1}{(1 - |a^i{}_{22}|^2)} \left[\frac{(|a^i_{11}||a^i_{12}|)^2}{(1 - |a^i{}_{11}|^2)} + |a^i{}_{12}|^2 \right] \ .$$

The same argument shows that we can successively choose the values of $p_3, \ldots, p_n$ in a way such that the leading principal minors of (6.9) are all positive, and this for any matrix A_i of the set. Indeed, let the leading principal minor of order k be > 0. Then, the leading principal minor of order $k + 1$ can be made > 0 too, because p_{k+1} only appears in its last diagonal term, with a strictly positive coefficient. So, taking p_{k+1} large enough is sufficient. The finiteness of the elements of A_i guarantees us that such a value p_{k+1} exists and is finite.

A Hermitian matrix H is positive definite if and only if all its leading principal minors are positive ([45]), and so we can deduce that the Hermitian matrix appearing in (6.9) is indeed positive definite, for any i. So, the P matrix built in this way is a common quadratic Lyapunov function for the set $\mathcal{M}$. □

Corollary 6.15. *For a set $\mathcal{M}$ of triangular matrices A_i, the joint spectral radius of $\mathcal{M}$ is equal to* $\max_i \rho(A_i)$ *and also to its ellipsoid norm approximation.*

Proof. From lemma 6.14, it follows that, for a set of stable upper-triangular matrices A_i, there exists a positive definite P_* such that $\|A_i\|_{P_*} < 1, \forall i$. This is equivalent to expressing

$$\max_i \rho(A_i) < 1 \Rightarrow \exists P_* \succ 0 : \max_i \|A_i\|_{P_*} < 1 \ .$$

By linearity (using the *normalization method*), this implies that

$$\max_i \rho(A_i) \geq \max_i \|A_i\|_{P_*} \ .$$

On the other hand, we know that the joint spectral radius is greater or equal to the largest spectral radius of the matrices in the set, that is $\rho(\mathcal{M}) \geq \max_i \rho(A_i)$. So, summing up, we have

$$\rho(\mathcal{M}) \geq \max_i \rho(A_i) \geq \max_i \|A_i\|_{P_*} \geq \rho_L(\mathcal{M}) \ .$$

As ρ_L is an over-approximation of $\rho(\mathcal{M})$, the four members of this inequality are equal. This yields $\rho(\mathcal{M}) = \max_i \rho(A_i)$ and $\rho_L(\mathcal{M}) = \rho(\mathcal{M})$. □

Lie algebraic conditions

We now generalize the previous result to a more general class of sets of matrices. This development is very similar to the one presented in [53]. Let us recall the following notations and definitions.
The *commutator* of the matrices A, B is the matrix given by

$$[A, B] = AB - BA \ .$$

Although it as already been used earlier in this work, we recall the following definition: the *linear span* of a set of vectors $(v_1, \ldots, v_k)$ in a vector space $\mathbb{V}$ is the subspace generated by

$$\left\{ \sum_i r_i v_i : r_i \in \mathbb{R} \right\} \ .$$

Definition 6.16. *The* Lie algebra $\{A_0, A_1\}_{LA}$ *is the linear span of the set of all possible combinations of commutators of* A_0 *and* A_1*:*

$$\{A_0, A_1, [A_0, A_1], [A_0, [A_0, A_1]], [A_1, [A_0, A_1]], \ldots\} \ .$$

The commutator series *of the Lie algebra* $\{A_0, A_1\}_{LA}$ *is the sequence of subalgebras recursively defined by* $\mathfrak{g}^0 = \{A_0, A_1\}_{LA}$*, and* $\mathfrak{g}^{k+1} = [\mathfrak{g}^k, \mathfrak{g}^k]$*. We have*

$$\mathfrak{g}^0 \supseteq \mathfrak{g}^1 \supseteq \mathfrak{g}^2 \supseteq \ldots \ ,$$

and if $\mathfrak{g}^k = \mathfrak{g}^{k+1}$ *then all subsequent subalgebras are also equal to* $\mathfrak{g}^k$*.*

A Lie Algebra is solvable *if its commutator series* $\mathfrak{g}^k$ *vanishes for some* k *(i.e.* $\mathfrak{g}^k$ *reaches the zero matrix for some* k*).*

An often used example of solvable Lie algebra is the vector space of upper-triangular matrices. It is easy to check that the sequence of subalgebras $\mathfrak{g}^k$ is the set of upper-triangular matrices whose elements on the diagonal at distance less than k from the main diagonal are all zero :

$$\mathfrak{g}^0 = \left\{ \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix},\ a_{ij} \in \mathbb{C}^{n \times n} \right\},$$

$$\mathfrak{g}^1 = \left\{ \begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ 0 & 0 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix},\ a_{ij} \in \mathbb{C}^{n \times n} \right\}, \dots$$

We make use of the following result (cited in [53], referring to [73]), which we state for $\mathbb{R}$, but would also hold for $\mathbb{C}$.

Lemma 6.17. *Let $\mathfrak{g}$ be a solvable Lie algebra over $\mathbb{R}$, and let ρ be a representation of $\mathfrak{g}$ on the vector space $\mathbb{R}^n$ (n finite). Then there exists a basis $\{v_1, \dots, v_n\}$ of $\mathbb{R}^n$ such that for each $X \in \mathfrak{g}$ the matrix of $\rho(X)$ in that basis takes the upper-triangular form*

$$\begin{pmatrix} \lambda_1(X) & \dots & * \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n(X) \end{pmatrix},$$

where the $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of the matrix $\rho(X)$.

This means that, under the solvability assumption, all the matrices of a set can be put simultaneously into upper-triangular form. We make use of this property to establish the following result.

Theorem 6.18. *Let $\mathcal{M} = \{A_1, \dots, A_m\}$ and consider the switched linear system*

$$x_{k+1} = A_{i_k} x_k,\ A_{i_k} \in \mathcal{M}.$$

If all matrices in $\mathcal{M}$ have a spectral radius less than 1 and the Lie algebra associated to $\mathcal{M}$ is solvable, then the system has a common quadratic Lyapunov function.

Proof. From lemma 6.17, we know that, if $\{A_i : A_i \in \mathcal{M}\}_{LA}$ is solvable, then there exists a (possibly complex) invertible matrix T such that

$$A_i = T^{-1} \tilde{A}_i T, \text{ with } \tilde{A}_i \text{ upper-triangular}, \forall i \ .$$

This introduction of complex values does not change the main argument.

The transformation from A_i to $\tilde{A}_i$ does not alter the eigenvalues and so $\rho(\tilde{A}_i) < 1$. By corollary 6.15, we know that this system is stable. Using this stability, Lemma 6.14 shows that there exists a real common quadratic Lyapunov function $\tilde{P}$ in diagonal form for such a set of matrices $\tilde{\mathcal{M}} = \{\tilde{A}_1, \dots, \tilde{A}_n\}$.

From this $\tilde{P}$, we can deduce the form of the corresponding P for the non-upper-triangular set $\mathcal{M} = \{A_1, \ldots, A_n\}$. To this aim, we rewrite the condition for the existence of a common quadratic Lyapunov function.

$$\begin{aligned}
\tilde{A_i}^* \tilde{P} \tilde{A}_i - \tilde{P} &\prec 0 \\
(TA_iT^{-1})^* \tilde{P} TA_iT^{-1} - \tilde{P} &\prec 0 \\
T^{*-1} A_i^* T^* \tilde{P} TA_iT^{-1} - \tilde{P} &\prec 0 \\
A_i^*(T^* \tilde{P} T) A_i - (T^* \tilde{P} T) &\prec 0 \ .
\end{aligned}$$

And we get $P = T^* \tilde{P} T$. As $\tilde{P}$ is positive definite, so is P. Moreover, $\tilde{P}$ being diagonal, $T^* \tilde{P} T$ is actually Hermitian, but is not guaranteed to be real, as T can be complex. Let us then denote

$$-R := A_i^*(T^* \tilde{P} T) A_i - (T^* \tilde{P} T)$$

where R is, by construction, Hermitian positive definite. We can write, by separating the real and imaginary parts,

$$P = \mathbb{R}(P) + i\mathbb{I}(P) \quad \text{and} \quad R = \mathbb{R}(R) + i\mathbb{I}(R) \ .$$

As P and R are Hermitian, $\mathbb{R}(P)$, $\mathbb{R}(R)$ are symmetric positive definite and $\mathbb{I}(P)$, $\mathbb{I}(R)$ are skew-symmetric. We can rewrite

$$A_i^*(\mathbb{R}(P) + i\mathbb{I}(P))A_i - (\mathbb{R}(P) + i\mathbb{I}(P)) = -(\mathbb{R}(R) + i\mathbb{I}(R)) \ ,$$

and taking the real part,

$$A_i^* \mathbb{R}(P) A_i - \mathbb{R}(P) = -\mathbb{R}(R) \ .$$

As a consequence, $\mathbb{R}(P)$ is a real common quadratic Lyapunov function for the solvable Lie algebra $\{A_i : A_i \in \mathcal{M}\}_{LA}$. □

Corollary 6.19. *If the Lie algebra associated to the set* $\mathcal{M} = \{A_1, \ldots, A_m\}$ *is solvable, then the joint spectral radius of* $\mathcal{M}$ *is equal to* $\max_i \rho(A_i)$ *and also to its ellipsoid norm approximation.*

Proof. Indeed, Theorem 6.18 allows us to deduce that $\rho(\mathcal{M}) < 1 \Rightarrow \rho_L(\mathcal{M}) < 1$, which yields, by linearity,

$$\rho(\mathcal{M}) \geq \rho_L(\mathcal{M}) \ ,$$

allowing to deduce the strict equality, thanks to the already known $\rho(\mathcal{M}) \leq \rho_L(\mathcal{M})$. Here again, we already know that $\rho(\mathcal{M}) \geq \max_i \rho(A_i)$, and Theorem 6.18 teaches us that $\max_i \rho(A_i) < 1 \Rightarrow \rho(\mathcal{M}) < 1$. By linearity, this leads to

$$\max_i \rho(A_i) \geq \rho(\mathcal{M}) \ .$$

And we deduce $\rho(\mathcal{M}) = \max_i \rho(A_i)$. □

We have equality between the joint spectral radius and its ellipsoid norm approximation when the Lie algebra is solvable. One could wonder whether the solvability of the Lie algebra is necessary for this equality to hold. This is not the case. In order to exhibit a counter-example, we first prove a property of independent interest.

Proposition 6.20. *The joint spectral radius of the pair $\{A, A^T\}$ is equal to its ellipsoid norm approximation and to the largest singular value of A, $\sigma(A)$.*

Proof. We have the already proved inequalities

$$\rho(A, A^T) \leq \rho_L(A, A^T) \leq \sigma(A) \ .$$

On the other hand, we can write

$$\rho(A, A^T) \geq \rho(AA^T)^{1/2} = \sigma(A) \ ,$$

because AA^T is among the products of length 2, taken from the set $\{A, A^T\}$. This eventually yields $\rho(A, A^T) = \rho_L(A, A^T) = \sigma(A)$. □

Example 6.21. As an illustration, consider the matrices

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ A^T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

It is easy to check that the Lie algebra associated to these matrices is not solvable. On the other hand it follows from the above proposition that for this pair of matrices the spectral radius and its ellipsoid norm approximation are equal (and are equal to $\sigma(A) = \frac{1+\sqrt{5}}{2} \simeq 1.618$). This provides the counter-example mentioned above.

We close this section with a numerical example of two matrices for which we do not have equality between the spectral radius and its ellipsoid norm approximation. Let us consider the following matrices:

$$A_1 = \begin{pmatrix} 1 & 2\ a_1 \\ -2/a_1 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 2\ a_2 \\ -2/a_2 & 1 \end{pmatrix} \ .$$

Assuming that $a_2 \geq a_1 \geq 1$, extensive calculations (presented in Appendix A) show that the approximation ρ_L is such that

$$\rho_L(A_1, A_2) \geq \sqrt{1 + 4\ a_2/a_1} \ .$$

For $a_1 = 1, a_2 = 2$, the joint spectral radius can be shown to be strictly less than 2.8584 by using an exhaustive calculation of all the products of 5 matrices. This is strictly less than $\sqrt{1 + 4 \times 2/1} = 3$, so here $\rho < \rho_L$. The gap between

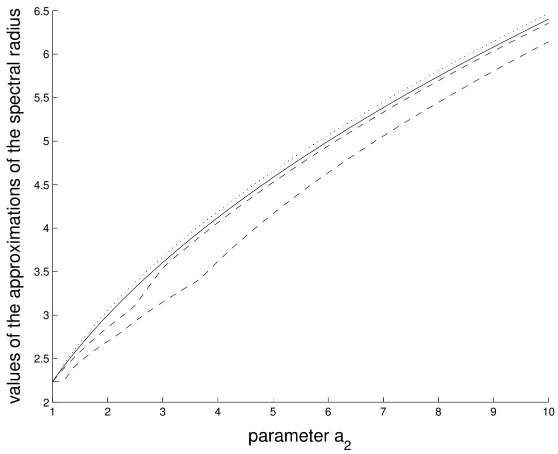

Fig. 6.3. The spectral radius and its ellipsoid norm approximation as functions of the real parameter a_2 (a_1 is fixed to 1). The two lowest curves (dashed) represent upper and lower bounds on the exact value of the spectral radius (computed with words of length 6). The middle curve (solid) represents $\sqrt{1+4a_2}$. The highest curve (dotted) represents the ellipsoid norm approximation.

the spectral radius and its approximation can be seen on figure 6.3 for $a_1 = 1$ and varying a_2.

From Corollary 6.19, we know that

$$\begin{aligned} \text{Solvability of the Lie algebra} &\Rightarrow \rho(\mathcal{M}) = \max_i \rho(A_i) \ , \\ &\Rightarrow \rho(\mathcal{M}) = \rho_L(\mathcal{M}) \ . \end{aligned}$$

We may wonder whether these two equalities are linked together.

$$\rho(\mathcal{M}) = \rho_L(\mathcal{M}) \ \stackrel{?}{\Leftarrow\Rightarrow} \ \rho(\mathcal{M}) = \max_i \rho(A_i) \ .$$

The implication from left to right is false, as can be seen from Example 6.21 above. The implication from right to left looks more sensible at first sight, but we provide hereunder a pair of matrices that seems to be usable as a counterexample.

Example 6.22.

$$A_1 = \begin{pmatrix} 1 & 1.5 \\ 0 & -0.5 \end{pmatrix}, A_2 = \begin{pmatrix} -0.5 & 0 \\ 1 & 0 \end{pmatrix} \ .$$

For this pair, ρ_L is found to be 1.1289, while $\rho(A_1) = 1$ and $\rho(A_1, A_2)$ is equal to 1. This can be shown by constructing an extremal norm, using the method presented in Section 5.4. This method is based on the properties of extremal norms (defined in Section 4.3.1) and will be easier to understand after a first reading of the corresponding section. To summarize the method, we here know that $\rho(A_1, A_2) \geq 1$ because $\rho(A_1) = 1$. To prove that $\rho(A_1, A_2) = 1$, we construct a set that is left invariant by $\{A_1, A_2\}$. If this succeeds, then this proves that the pair of matrices is product bounded, and so $\rho(A_1, A_2) \leq 1$. This is enough to deduce that $\rho(A_1, A_2) = 1$.

To keep the notations simple, we write the vectors on a line, knowing that they are to be treated as columns when multiplied by a matrix. First, we take the leading eigenvector of A_1, which is $v_1 = (1, 0)$ and see how it is affected by A_1 and A_2. For example, $A_1(1, 0) = (1, 0)$ and $A_2(1, 0) = (-0.5, 1)$. Then, we apply A_1 and A_2 to these new points. This is summarized in figure 6.4.

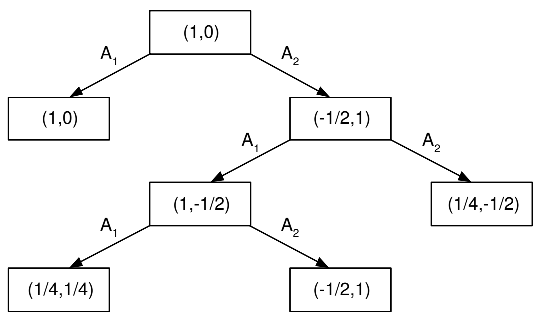

Fig. 6.4. This tree presents the effect of successive products of the matrices A_1 and A_2 on the vector $(1, 0)$. We stop developing it when we find a point that is already present in the tree. Alternatively, if the found point is inside the convex hull defined by the previous points, we also stop. In this manner, we find that the points $(1, 0)$, $(-1/2, 1)$, $(1, -1/2)$ and their symmetric counterparts about the origin form a invariant set under $\{A_1, A_2\}$.

From this, we deduce that the polygon defined by the points $(1, 0)$, $(-1/2, 1)$, $(1, -1/2)$ and their symmetric counterparts about the origin is invariant under $\{A_1, A_2\}$. After reading Section 4.3.1, it will become clear that this is the level set of an extremal norm. We represent this polygon in figure 6.5. This concludes the example.

6.5.4 Guaranteed precision of the ellipsoid norm approximation

The ellipsoid norm approximation of the joint spectral radius can be shown to have a guaranteed precision. The argument is simple: Let ρ be the spectral radius of the set $\{A_1, \ldots, A_m\}$. We know by Proposition 4.17 that, $\forall r > \rho(\mathcal{M})$, there exists a vector norm $\|.\|_*$ for which $\|A_i x\|_* \leq r\|x\|_*$ for all x and i. This norm can be seen as a Lyapunov function of the dynamical system. Thanks to the usual norm properties (namely the triangular inequality), we know that any *level curve* $\{x : \|x\|_* = c\}$ of this norm defines a closed convex set $\{x : \|x\|_* \leq c\}$. And a closed convex set can be approximated by ellipsoids. The quality of these approximations can be measured and provides a guaranteed precision for the approximation.

We start by describing the quality of the best possible ellipsoids (the result below is known as John's theorem; it is stated in [47], referring to [49]).

Theorem 6.23 (John). *Let $K \subset \mathbb{R}^n$ be a compact convex set with nonempty interior. Then there is an ellipsoid E with center c such that the following inclusions hold:*

$$E \subseteq K \subseteq c + n(E - c) \ .$$

If K is symmetric about the origin ($K = -K$), the constant n can be changed into $\sqrt{n}$:

$$E \subseteq K \subseteq \sqrt{n}E \ .$$

Here, $c + n(E - c)$ is the set of points $\{c + n(x - c) : x \in E\}$, which is the dilation of E by a factor n with the center c. This ellipsoid is called John's ellipsoid, and it turns out to be the ellipsoid of maximal volume contained in K. We illustrate this in Figure 6.6.

Knowing this, we can now prove:

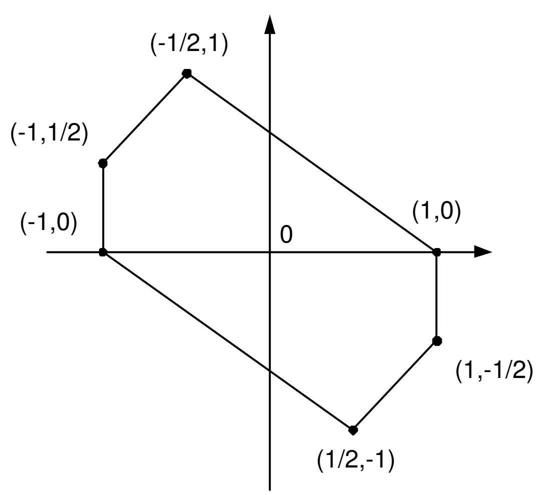

Fig. 6.5. Representation of the polygon that is left invariant by matrices A_1 and A_2. The points $(1, 0)$, $(-1/2, 1)$, $(1, -1/2)$ have been found through the algebra described earlier, and we added their symmetric counterparts.

Theorem 6.24. *Let ρ be the joint spectral radius of a finite set of matrices of dimension n. The ellipsoid norm approximation ρ_L of the joint spectral radius satisfies*

$$\rho_L(\mathcal{M})/\sqrt{n} \leq \rho(\mathcal{M}) \leq \rho_L(\mathcal{M}) \ .$$

Proof. The norm mentioned above is symmetric about the center, as

$$\|x\|_* = \|-x\|_*, \forall x \ .$$

We can apply Theorem 6.23. It guarantees us that, whatever the norm $\|.\|_*$ is, there exists a quadratic norm $\|x\|_P = \sqrt{x^T P x}$ (of which level curves are ellipsoids) such that

$$\|x\|_P \leq \|x\|_* \leq \sqrt{n}\|x\|_P \ .$$

And, from Proposition 4.17 and Remark 4.18, we know that $\forall q > \rho(\mathcal{M})$, there is a norm $\|.\|_*$ that satisfies

$$\|A_i x\|_* \leq q\|x\|_*, \forall x, \forall i \ .$$

We can now write

$$\begin{gathered}
\forall q > \rho(\mathcal{M}), \forall x, \forall i, \quad \|A_i x\|_P \leq \|A_i x\|_* \leq q\|x\|_* \leq q\|x\|_P\sqrt{n} \\
\forall q > \rho(\mathcal{M}), \forall x, \forall i, \quad \|A_i x\|_P \leq q\|x\|_P\sqrt{n} \\
\forall q > \rho(\mathcal{M}), \forall x, \forall i, \quad x^T A_i^T P A_i x \leq q^2 n \ x^T P x \\
\forall q > \rho(\mathcal{M}), \forall i, \quad A_i^T P A_i - q^2 n P \preceq 0 \ .
\end{gathered}$$

Thus, the approximation ρ_L, which was shown to be equivalent to

$$\rho_L(\mathcal{M}) = \inf_{P \succ 0} \max_{A_i \in \mathcal{M}} \|A_i\|_P \ ,$$

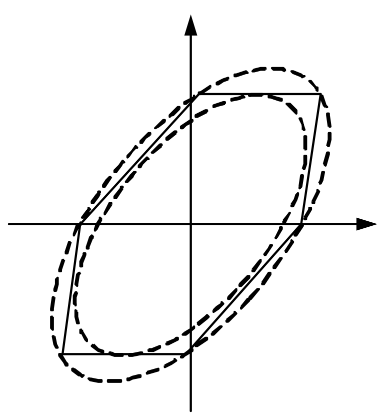

Fig. 6.6. A convex body bounded frow inside and outside by two ellipsoids.

satisfies

$$\forall q > \rho(\mathcal{M}), \quad \rho_L \leq q\sqrt{n} \ .$$

This approximation will, at worst, result in the value $\rho(\mathcal{M})\sqrt{n}$. On the other hand, taking back equation (6.6), we also know that $\rho_L(\mathcal{M})$ is an over-approximation of $\rho(\mathcal{M})$. This settles the second inequality. Summing up, this gives:

$$\rho(\mathcal{M}) \leq \rho_L(\mathcal{M}) \leq \rho(\mathcal{M})\sqrt{n} \ .$$

□

6.5.5 Extension of ρ_L to longer products

The value $\rho_L(\mathcal{M})$ has a fixed value, and it can be computed up to any precision. Once we have found it, there is no way of computing further in order to get closer to the JSR. We can try to achieve this by directly considering the set $\mathcal{M}^k$ of 2^k products of a fixed length k built from the initial matrices.

As before in this thesis, we associate the product $A_w = A_{w_1} \cdots A_{w_k}$ to the word $w = w_1 w_2 \cdots w_k$, $w_i \in \{0, 1\}$, and $|w|$ denotes the length of w. For a given k, we define the following value:

$$\rho_{L_k}(A_0, A_1) = \inf_{\lambda, X > 0} \{ \sqrt[2k]{\lambda} : A_w^* X A_w \leq \lambda X, \forall w, |w| = k \} \ .$$

The case $k = 1$ corresponds to the approximation ρ_L above.

The new extended set $\mathcal{M}^k$ we consider is not restrictive, in the sense that we will still be able to generate any arbitrary product of (A_0, A_1) from this new set. So, the approximations of the JSR we will get from it are at least as good as the ones from $\mathcal{M}$. We now prove this more formally (again starting from a pair of matrices, but larger sets work as well).

Proposition 6.25. *For all pairs of matrices* (A, B)*, we have*

$$\forall k, l \geq 1, \ \rho(A, B) \overbrace{\leq}^{(a)} \rho_{L_{kl}}(A, B) \overbrace{\leq}^{(b)} \rho_{L_k}(A, B) \overbrace{\leq}^{(c)} \rho_L(A, B) \ .$$

Proof. Let us now prove inequality (b). We will consider $A_w^* X A_w$, taken from the definition of $\rho_{L_k}(A, B)$, with $|w| = kl$. We use the supposed-known value of $\rho_{L_l}(A, B)$, which we denote by ρ_L^*. We can write these successive inequalities:

$$\begin{aligned} A_w^* X A_w &\leq {\rho_L^*}^{2l} A_{w-l}^* X A_{w-l} \leq {\rho_L^*}^{4l} A_{w-2l}^* X A_{w-2l} \leq \dots \\ &\leq {\rho_L^*}^{2kl} X \ , \end{aligned}$$

denoting by A_{w-n} the matrix associated to the word w of which we cut the last n elements. So, for the very same X that achieves the value ρ_L^* in the definition of $\rho_{L_l}(A,B)$, we can already get a λ equal to ρ_L^*. And other words of length kl might achieve lower values of λ. This proves that $\rho_{L_{kl}}(A,B) \leq \rho_{L_l}(A,B)$.

Inequality (c) is proved in just the same way. It is indeed a particular case of (b), with $l = 1$.

Inequality (a) remains to be proved. We do not need to prove it specifically with a right member of the form $\rho_{L_{kl}}(A,B)$, just for $\rho_{L_l}(A,B)$, $\forall l$. Let us suppose that we have a finite value l such that $\rho_{L_l}(A,B) < 1$. Then clearly, the system made of all matrix products of length l is stable, because it admits a CQLF. And as we said earlier, the set of matrices $\mathcal{M}^l$ can simulate all products built from $\mathcal{M}$. This proves the convergence to the origin of all the possible trajectories of the switched system generated by $\mathcal{M} = (A,B)$, and so $\rho(A,B) < 1$. We now have shown that, $\forall l$,

$$\rho_{L_l}(A,B) < 1 \Rightarrow \rho(A,B) < 1$$

Using again our *normalization method*, we get $\rho_l(A,B) \geq \rho(A,B)$. $\square$

Remark 6.26. In [6], Bliman & Ferrari-Trecate show that, for any stable system, there exists a finite k^* such that all words of length k^* admit a CQLF, even if the initial matrices do not. Put in our context, it is equivalent to saying that

$$\text{system is stable} \Rightarrow \exists k^* : \rho_{L_{k^*}}(A_0, A_1) < 1 \ ,$$

or again

$$\rho(A_0, A_1) < 1 \Rightarrow \exists k^* : \rho_{L_{k^*}}(A_0, A_1) < 1 \ .$$

Using linearity, we deduce that, in this case,

$$\forall y > \rho(A_0, A_1), \exists k^* : \rho_{L_{k^*}} < y \ .$$

So, for the proper value of k, $\rho_{L_k}(A_0, A_1)$ can be arbitrarily close to $\rho(A_0, A_1)$.

But these ρ_k^* are not guaranteed to reach exactly $\rho(\mathcal{M})$ for a finite k^*. If we were to prove that $\exists k^* : \rho_{L_{k^*}}(A_0, A_1) \leq \rho(A_0, A_1)$, this would provide an algorithm for effectively computing the JSR, which is impossible (see Section 4.4). To clarify the situation, we now prove that there are pairs for which $\rho \neq \rho_{L_k}, \forall k$, although ρ_{L_k} can approximate ρ arbitrarily closely. This is done by providing an example.

Example 6.27. Consider a set made of twice the same matrix:

$$\mathcal{M} = \{A_0, A_0\}, \quad \text{with} \quad A_0 = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix} \ .$$

We immediately obtain $\rho(A_0, A_0) = \rho(A_0) = 1$ and $\sigma(A_0) = 1.2808$. The newly defined approximation is

$$\rho_{L_k}(A_0, A_0) = \inf_{\lambda, X>0} \{ \sqrt[2k]{\lambda} : (A_0^k)^T X A_0^k \leq \lambda X \} .$$

To compute this, we make use of the relation $A_0^k = \begin{pmatrix} 1 & \frac{k}{2} \\ 0 & 1 \end{pmatrix}$, so this system does not admit any Quadratic Lyapunov function, and $\rho_{L_k} > 1, \forall k$. The reason here is that, while the spectral radius is equal to one, there is still a divergence of the system. This divergence is polynomial, while the spectral radius gives the exponential behavior of the system. This polynomial divergence is due to the fact that the geometric multiplicity of the spectral radius is different from its algebraic multiplicity (that is to say, the matrix is proportional to a Jordan block).

Quality of the approximation ρ_{L_k}

We proved earlier that the approximation ρ_L is at most within a factor $\sqrt{n}$ of $\rho(\mathcal{M})$. Here, the result will be different, as we consider products of k matrices. We have $\forall x, \forall |w| = k, \|A_w x\|_* \leq q^k \|x\|_*$. By the same deduction as what was made for ρ_L in Theorem 6.24, we know that $\rho_{L_k} \leq q \sqrt[2k]{n}$. We may express the error factor between ρ and ρ_{L_k} as $(1 + \varepsilon)$ with

$$\begin{aligned} 1 + \varepsilon &= \sqrt[2k]{n} \\ (1 + \varepsilon)^{2k} &= n \\ 2k &= \ln n / \ln (1 + \varepsilon) \end{aligned}$$

Figure 6.7 plots this evolution of k as a function of the desired ε for $n = 2$.

6.6 Approximation of extremal norms

6.6.1 Method 1 : converging to an invariant set

Initial context

Guglielmi & Zennaro have tackled the problem of approximating the JSR in a different way. One of the aspects of their work in [40] is to determine an extremal norm for a set of matrices. They consider a set of real matrices with non-negative elements of the form

$$\hat{\mathcal{G}} = \left\{ \{\hat{A}^{(i)}\}_{i \in \mathcal{I}}, \{\hat{B}^{(j)}\}_{j \in \mathcal{J}} \right\} ,$$

with

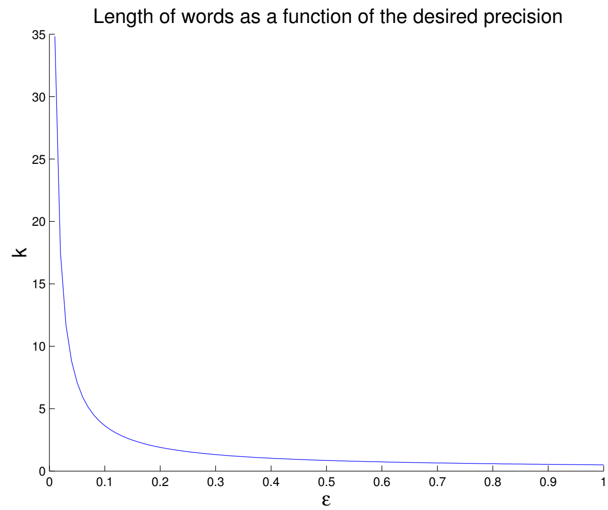

Fig. 6.7. Evolution of the value of k needed to guarantee a given precision ε on ρ_{L_k}. Here, $n = 2$.

$$\hat{A}^{(i)} = \begin{pmatrix} 1 & \alpha_i \\ 0 & a_i \end{pmatrix}, \; \hat{B}^{(j)} = \begin{pmatrix} b_j & \beta_j \\ 0 & 1 \end{pmatrix} .$$

This set has a joint spectral radius equal to 1, and they want to prove its nondefectiveness (i.e. that it is product bounded) under the condition $a_i, b_j < 1$. To do so, it is sufficient to prove the existence of an extremal norm. With a JSR equal to 1, this equates to finding a norm such that

$$\begin{aligned} \forall A^{(i)}, B^{(j)} \in \hat{\mathcal{G}}, \quad & \|A^{(i)}x\| \leq \|x\| \\ & \|B^{(j)}x\| \leq \|x\| . \end{aligned}$$

This is indeed sufficient to prove product boundedness.

Guglielmi & Zennaro then come up with a polytope $\mathcal{P}$ which they prove to be the unit ball of an extremal norm of the set of matrices. To do so, they show that all the vertices of $\mathcal{P}$ are mapped by all the matrices of $\hat{\mathcal{G}}$ into $\mathcal{P}$ itself. This polytope is illustrated in Figure 6.8.

It is symmetric about the origin O and its vertices are $x_1 = (\gamma, 1), x_2 = (\delta, 1), x_3 = (\delta + \theta, 0)$, with

$$\begin{aligned} \theta &= \sup_{i \in \mathcal{I}} \frac{\alpha_i}{1 - a_i} , \\ \gamma &= \inf_{j \in \mathcal{J}} \frac{\beta_j}{1 - b_j} , \\ \delta &= \sup_{j \in \mathcal{J}} \frac{\beta_j}{1 - b_j} . \end{aligned}$$

The values $a_i = 1$ and $b_j = 1$ are threshold values, as expected. However, it is not trivial to see how the authors built this polytope in practice. We investigate here a way of finding such a level curve.

Principle

We first need to normalize the matrices, so that the JSR is equal to 1. Moreover, we need the set to be non-defective (or irreducible). Under these two assumptions, all possible matrix products will remain bounded, even the product(s) achieving the JSR, as proved formally in [40]. And, as concerns matrix products that are not optimal, they just tend to zero. Consequently, we can try to apply all 2^k products of length k (if we study a pair of matrices) to an arbitrary initial set of points, as the unit circle or the unit square, for example. The resulting set of points should remain bounded, thanks to the non-defectiveness property. This should converge to a fixed set, the convex envelope of which should provide the level curve of a norm (at least close to) extremal.

Application

We take a pair from Guglielmi & Zennaro's example, and assign the arbitrary values $a_i = 0.6$, $\alpha_i = 0.2$, $b_j = 0.6$, $\beta_j = 1.5$:

$$\hat{A} = \begin{pmatrix} 1 & 0.2 \\ 0 & 0.6 \end{pmatrix}, \ \hat{B} = \begin{pmatrix} 0.6 & 1.5 \\ 0 & 1 \end{pmatrix} .$$

This pair is already normalized, as these matrices are upper-triangular, and their diagonal elements are ≤ 1. We can use the result above to compute

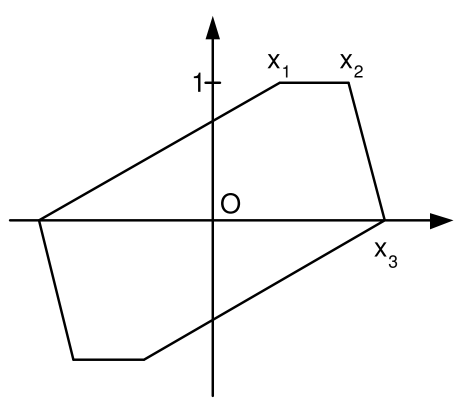

Fig. 6.8. Shape of the polytope proved to be the unit ball of an extremal norm. The coordinates of x_1, x_2, x_3 are: $x_1 = (\gamma, 1), x_2 = (\delta, 1), x_3 = (\delta + \theta, 0)$.

$$\theta = \frac{0.2}{1 - 0.6} = 0.5 \ ,$$
$$\gamma = \frac{1.5}{1 - 0.6} = 3.75 \ ,$$
$$\delta = \frac{1.5}{1 - 0.6} = 3.75 \ .$$

Due to the specific nature of this example, γ and δ are identical. A unit ball of extremal norm is thus given by the polyhedron defined by $x_1 = x_2 = (3.75, 1), x_3 = (4.25, 0)$ and their opposites. It is represented in Figure 6.9. To carry out the computation we proposed above, we need a starting set of

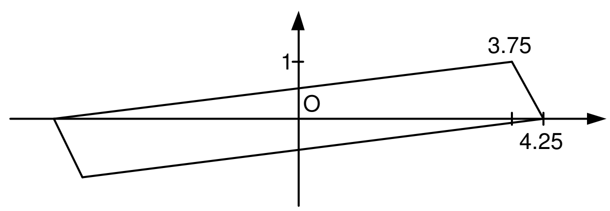

Fig. 6.9. Shape of the polytope proved to be the unit ball of an extremal norm for the specific pair with $a_i = 0.6$, $\alpha_i = 0.2$, $b_j = 0.6$, $\beta_j = 1.5$. The coordinates of the vertices are $x_1 = x_2 = (3.75, 1), x_3 = (4.25, 0)$.

points. Of course, we may not start from the above polytope, as this is precisely what we are trying to determine. Moreover, by construction, we already know that all the image points will lie inside it, and thus will not teach us anything. We chose to start from the unit circle. We find the image of the circle shown in Figure 6.10. To clear up things and speed up the computations, we can, at each step, take only the vertices of the convex hull defined by all the obtained points. Then, we get the picture presented in Figure 6.11. This simplification allows the computation of much longer products. In figure 6.12, we present the result for products of length 100. The convex hull we find in this way is undoubtedly similar to the polytope in Figure 6.9.

6.6.2 Method 2 : computing the level curve of the norm

The second method is in the same spirit as the one above. We will also carry out many matrix products and deduce the information we need. We try to approximate the *Kozyakin norm*, the construction of which is detailed in the proof of Theorem 4.16. We will again consider an irreducible set of matrices, so that it is non-defective and admits an extremal norm. Otherwise there would be no point in searching for one.

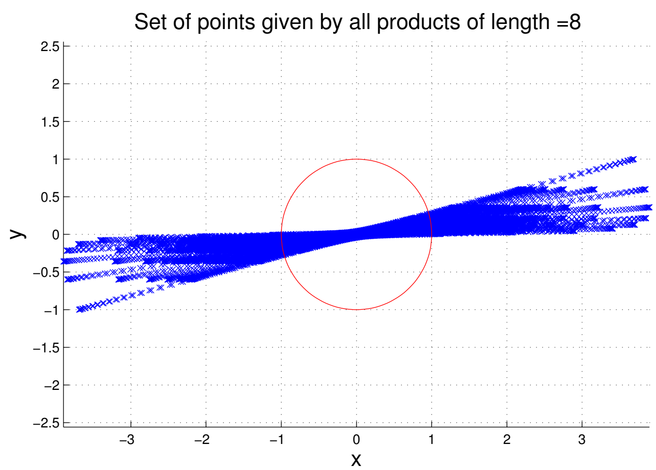

Fig. 6.10. The starting set is the unit circle. The cloud of points represents the image of this discretized circle by all products of length 8.

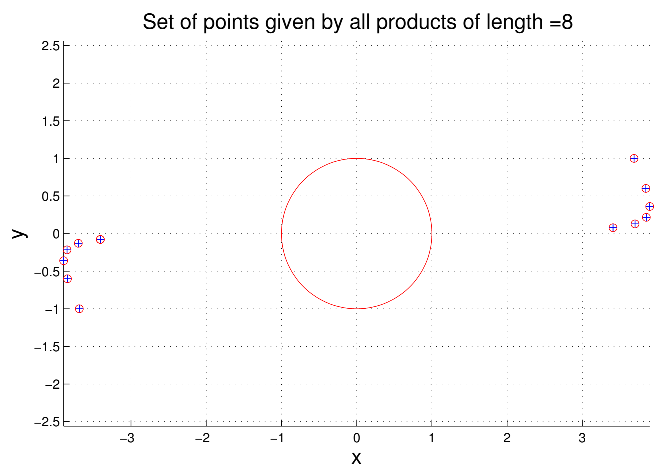

Fig. 6.11. The starting set is the unit circle. The points represent the vertices of the convex hull defined by the image of the discretized unit circle by all products of length 8.

Principle

We start from an initial set of points. For each point x, we compute the largest norm v of the image of this point by matrix products. Here, we arbitrarily choose to use the usual Euclidian norm (or 2-norm). Because of finite computing capacities, we are limited to a finite length, let us say k. This means that we will compute

$$v(x) = \max \left\{ \left\| \prod_{i=1}^{k} A_i x \right\| : A_i \in \mathcal{M} \right\} .$$

Then, by plotting the normalized points $x/v(x)$, we get a set of points x^* for which $v(x^*) \approx 1$. In figure 6.13, we represent the result of this computation for the example above:

$$\hat{A} = \begin{pmatrix} 1 & 0.2 \\ 0 & 0.6 \end{pmatrix}, \ \hat{B} = \begin{pmatrix} 0.6 & 1.5 \\ 0 & 1 \end{pmatrix} .$$

The left figure shows that the image can lie within the unit circle, but the corresponding points are dismissed and are assigned a Kozyakin norm equal to 1. This is the mere application of the construction described in the proof of Theorem 4.16. Now, it remains to take the inverse of these values to obtain the approximated level curve of the Kozyakin norm. This is presented in Figure 6.14. We can see a similarity between this approximation and the polytope norm that was previously shown to be extremal for the studied pair of matrices.

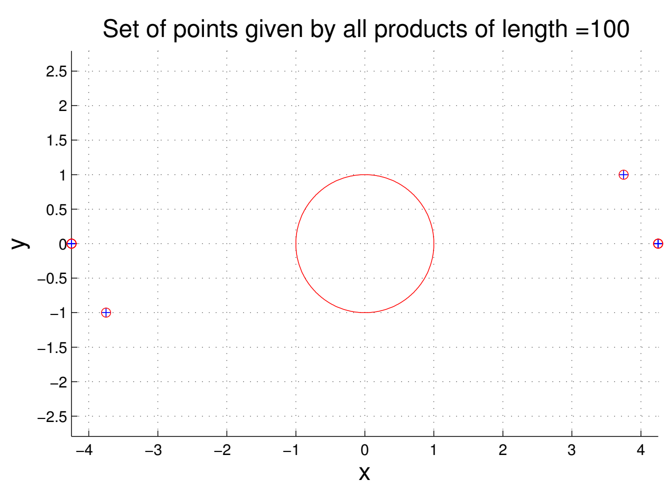

Fig. 6.12. The starting set is the unit circle. The points represent the vertices of the convex hull defined by the image of the discretized unit circle by all products of length 8.

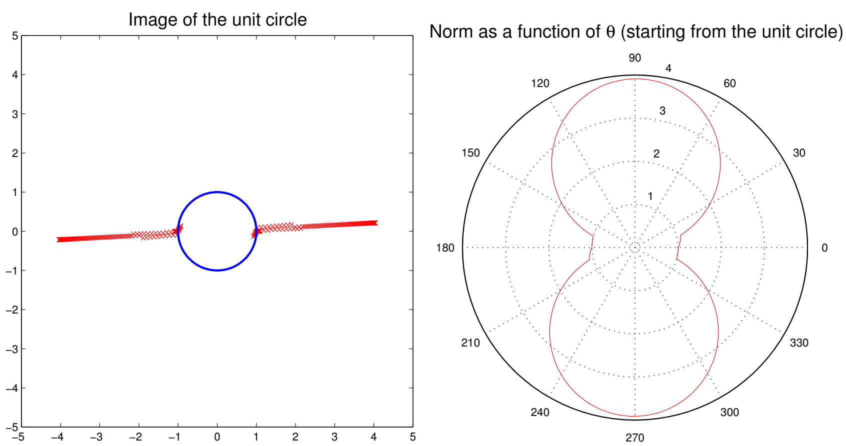

Fig. 6.13. On the left : image of the unit circle. On the right : deduced Kozyakin norm of the image as a function of the angle. We used products of length 10.

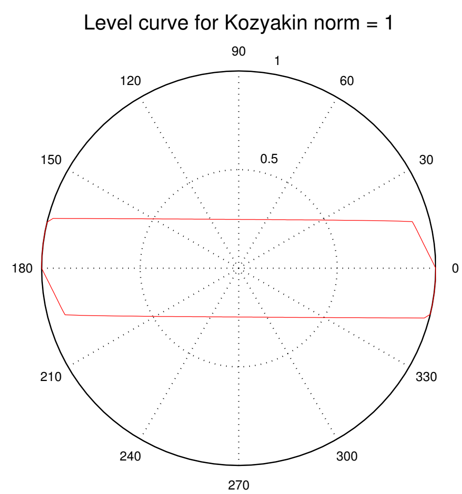

Fig. 6.14. The normalized points $x^* = x/v(x)$ for the pair $\hat{A}, \hat{B}$. This is an approximation of the level curve of the Kozyakin norm.

6.7 Kronecker products

We now present a totally different approach to the approximation of the JSR. This one involves Kronecker (tensor) product, denoted by $\otimes$. As a reminder, for $A, B \in \mathbb{R}^{2\times 2}$:

$$A \otimes B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix} \in \mathbb{R}^{4\times 4} \ .$$

More generally, the Kronecker product of a $(m_1 \times n_1)$ matrix times a $(m_2 \times n_2)$ matrix is a $(m_1 m_2 \times n_1 n_2)$ matrix. The Kronecker product has interesting properties, such as:

- $(A \otimes B)^T = A^T \otimes B^T$
- $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

These properties are simple to check. And if we denote the Kronecker power of a matrix by

$$A^{\otimes k} = A \otimes A \otimes \cdots \otimes A \quad (k \text{ times}) \ ,$$

we also have

- $(A^{\otimes k})^T = (A^T)^{\otimes k}$
- $A^{\otimes k} B^{\otimes k} = (AB)^{\otimes k}$.

Moreover, it is easy to see that the spectral radius of the k^{th} Kronecker power of a matrix is the k^{th} power of the matrix:

$$\rho(A^{\otimes k}) = \rho(A)^k \ .$$

Combining these properties, we make the following observation. The spectral radius of a product of k^{th} powers of matrices can be shown to be

$$\begin{aligned} \rho\left(A_{i_1}^{\otimes k} A_{i_2}^{\otimes k} \dots A_{i_j}^{\otimes k}\right) &= \rho\left((A_{i_1} A_{i_2} \dots A_{i_j})^{\otimes k}\right) \\ &= \rho\left(A_{i_1} A_{i_2} \dots A_{i_j}\right)^k \ . \end{aligned} \tag{6.10}$$

As a consequence,

Lemma 6.28. *When raising a set $\mathcal{M} = (A_1, A_2, \dots)$ to the k^{th} Kronecker power to get*

$$\mathcal{M}^{\otimes k} = \left(A_1^{\otimes k}, A_2^{\otimes k}, \dots\right) \ ,$$

its JSR is raised to the power k:

$$\rho\left(\mathcal{M}^{\otimes k}\right) = \rho\left(\mathcal{M}\right)^k \ .$$

Proof. As we consider bounded sets, the JSR and GSR are identical. Then, applying the definition of the GSR (using spectral radii), the property is immediate from relation (6.10). $\square$

We now recall the result presented in Lemma 6.7:

$$\rho(\mathcal{M}) \geq \frac{1}{m}\, \rho\left(\sum_{i=1}^{m} A_i\right) ,$$

and the result from Corollary 6.10:

$$\rho(\mathcal{M}) \leq \rho\left(\sum_{i=1}^{m} A_i\right) ,$$

for a set $\mathcal{M} = \{A_1, \dots, A_m\}$ of matrices with non-negative entries. Combining these with Lemma 6.28, we get the following relation:

$$\frac{1}{m}\, \rho\left(\sum_{i=1}^{m} A_i^{\otimes k}\right) \leq \rho(\mathcal{M}^{\otimes k}) \leq \rho\left(\sum_{i=1}^{m} A_i^{\otimes k}\right) .$$

The center member is equal to $\rho(\mathcal{M})^k$. Then, taking the k^{th} root of the equation gives

$$\left(\frac{1}{m}\right)^{1/k} \rho\left(\sum_{i=1}^{m} A_i^{\otimes k}\right)^{1/k} \leq \rho(\mathcal{M}) \leq \rho\left(\sum_{i=1}^{m} A_i^{\otimes k}\right)^{1/k} .$$

This proves that, for increasing k, the value $\rho\left(\sum_{i=1}^{m} A_i^{\otimes k}\right)^{1/k}$ converges to the JSR, because $\left(\frac{1}{m}\right)^{1/k} \to 1$ as $k \to \infty$.

From the computational cost point of view, there is no miracle here, as the size of the considered matrices grows exponentially with k. The advantage of using this relation to determine the JSR is to have an a priori guarantee on the precision of the obtained approximation. Fore more details about this kind of approach to the approximation of the JSR, we refer the reader to [9]. Blondel & Nesterov present therein results for matrices leaving a cone invariant, as well as a *lifting* method that turns any set of matrices into a cone preserving one. To be complete, more work done around ellipsoid norms and Kronecker products can be found in [66, 67, 87].

6.8 Conclusion

This concludes the chapter devoted to the computation of the JSR. We have seen that we could deduce some information about the value of the JSR directly from its definition. We presented the approximation $\tilde{\rho}$, which is quick to compute, yet not necessarily very close to the actual value of the JSR. The approximation ρ_L then proved to be useful, namely to identify cases where the JSR $\rho(\mathcal{M})$ is equal to the largest spectral radius of the matrices in $\mathcal{M}$.

However, the cases where $\rho(\mathcal{M}) = \rho_L(\mathcal{M})$ and $\rho(\mathcal{M}) = \sup_{A\in\mathcal{M}} \rho(A)$ do not coincide. We were able to prove that ρ_L is at worst $\sqrt{n}\rho(\mathcal{M})$, thanks to the John ellipsoid theorem. The interpretation of ρ_L in terms of an ellipsoid norm allowed us to link it to the more general search for an extremal norm.

Conclusion

In this thesis, we reviewed the definition and the bibliographical background of the joint spectral radius. We first discussed some of its elementary properties that could be derived directly from the definitions.

We then presented three different contexts where the joint spectral radius comes into play. It is shown to play a key role in the study of multi-agent dynamical systems, where it can be used to determine whether all the agents will eventually move in the same direction. It is also shown to describe the continuity of wavelet functions, which have become a widely used basis for function decomposition. And it does not only give a yes/no answer to the continuity of the wavelet, but it appears to provide the Hölder exponent of the wavelet function. The JSR also appears in information theory, and more specifically for expressing the capacity of codes. The joint spectral radius is used there as a performance measure.

After this, we made the connection between the joint spectral radius and the stability of switched linear systems. We formally established the equivalence between $\rho(\mathcal{M}) < 1$ and the absolute asymptotic stability of the discrete linear inclusion associated to the set $\mathcal{M}$. Although this result is not new, as it is a consequence of known results, we had never seen it stated so explicitly.

We also considered how discrete time switched linear systems relate to their continuous time counterparts. We have explored the connections between the stabilities of discrete time and continuous time systems. We raised more questions than we provided answers about this topic, and it is still an open problem to determine whether a transformation exists that maps stable discrete time systems to stable continuous time systems. Such a transformation would provide a strong justification to narrowing the study of switched systems to either continuous time or discrete time. On the other hand, if one could prove that this transformation does not exist, then this would prove that both types of systems deserve different approaches, and they might not obey the same rules for stability.

We then compiled a small bibliography of results proving that the computation of the joint spectral radius is a difficult problem. We presented a non-algebraicity result by Kozyakin, to which we provide an original correction, as its initial proof contains a flaw. We also discussed criteria for the existence of an extremal norm. Such an extremal norm can be useful to compute the joint spectral radius, but it appears to be as hard to compute exactly as the JSR itself. This is coherent, since the computation of the JSR is supposed to be NP-hard, so we are not expecting it to reduce to a simple (easily computable/solvable) problem. We then concluded this aspect by reviewing NP-hardness and undecidability results.

We then tackled the problem of effectively computing the joint spectral radius, in a more general case. We first looked at what the theory taught us, and what kind of method it allowed. From this, it appears that the exponential computational cost cannot be avoided, as the numerical tricks that can be used do not allow for a decrease in the global complexity. We provided an easy way of approximating the joint spectral radius of matrices with non-negative entries, and showed that this approximation is within a factor at most m of the exact value, where m is the number of matrices in the set. After this, we introduced a polynomial-time approximation ρ_L of the joint spectral radius, based on a modification of the Lyapunov equation. Our approximation is easy to compute and is guaranteed to be within a factor $\sqrt{n}$ of the exact value, where n is the dimension of the matrices. We described particular classes of matrices for which ρ_L is equal to the joint spectral radius. These classes includes upper-triangular matrices and, more generally, matrices that can be put simultaneously under a triangular form. It appears that, for those cases, the joint spectral radius is actually the largest spectral radius of the matrices in the set. The problem of characterizing exactly the sets of matrices for which this equality holds is a question that remains open. Answering this question requires a better understanding of the underlying mechanisms that rule the interactions between the different matrices. To this aim, a point of view worth considering would be to characterize the various eigenvectors of the matrices in the set, and to see if their differences/similarities have an influence on the JSR.

One of the contributions of this work is the extensive study of the specific pair of matrices

$$\left\{ A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\ A_1 = \alpha \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\} .$$

For this type of pair, we may say that we solved the problem of computing the JSR. Indeed, we managed to prove that, for this pair of matrices, the matrix products achieving the JSR do possess a specific property : the binary words associated to these products are balanced. This gives a good understanding of the structure of these *optimal* words. This opens a new range of questions to answer, as :

- are there other families of matrices that admit optimal balanced words ?
- what makes the balanced words so special ?

We also depicted a geometrical method for building the level curve of an extremal norm for a pair of matrices of the same type as the one above. We feel it would be interesting to apply this method to other sets of matrices to determine if it is effective. Moreover, a study of the shape of these level curves as a function of the matrices composing the set might provide some insight into the interactions between the matrices that lead to the JSR. At this point, we have not studied this construction method in detail, but it would be interesting to determine the typical number of geometrical operations it demands for effectively terminating, in a general case (or at least using several different examples).

A

The existence of a CQLF is not necessary for stability

Let us consider the following pair of matrices :

$$A_1 = \begin{pmatrix} -0.2 & -0.4\ a_1 \\ 0.4/a_1 & -0.2 \end{pmatrix}, A_2 = \begin{pmatrix} -0.2 & -0.4\ a_2 \\ 0.4/a_2 & -0.2 \end{pmatrix} .$$

Assuming that $a_2 > a_1 > 1$, extensive calculations will show that the approximation $\hat{\rho}$ is such that

$$\hat{\rho}(A_1, A_2) \geq \sqrt{0.04 + 0.16\ a_2/a_1} .$$

For $a_1 = 1, a_2 = 2$, the joint spectral radius can be shown to be strictly less than 0.5717, using an exhaustive calculation involving all matrix products of length 5. This is strictly less than $\sqrt{0.04 + 0.16\ 2/1} = \sqrt{0.36} = 0.6$, so here $\rho < \hat{\rho}$.

The detailed calculations are as follows:

Let us consider the matrix $A = \begin{pmatrix} -0.2 & -0.4\ a \\ 0.4/a & -0.2 \end{pmatrix}$ and a positive definite matrix X. We now study the validity of the inequality $A^*XA - \lambda^2 X \leq 0$.

$$\begin{pmatrix} -0.2 & 0.4/a \\ -0.4\ a & -0.2 \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \begin{pmatrix} -0.2 & -0.4\ a \\ 0.4/a & -0.2 \end{pmatrix} - \lambda^2 \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \leq 0$$

$$\begin{pmatrix} -0.2\ x_{11} + 0.4/a\ x_{12} & -0.2\ x_{12} + 0.4/a\ x_{22} \\ -0.4\ ax_{11} - 0.2\ x_{12} & -0.4\ ax_{12} - 0.2\ x_{22} \end{pmatrix} \begin{pmatrix} -0.2 & -0.4\ a \\ 0.4/a & -0.2 \end{pmatrix} - \lambda^2 \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \leq 0$$

$$\begin{pmatrix} (0.04 - \lambda^2)x_{11} - 0.16/a\ x_{12} + 0.16/a^2 x_{22} & 0.08\ ax_{11} - (0.12 + \lambda^2)x_{12} - 0.08/a\ x_{22} \\ 0.08\ ax_{11} - (0.12 + \lambda^2)x_{12} - 0.08/a\ x_{22} & 0.16a^2 x_{11} + 0.16\ ax_{12} + (0.04 - \lambda^2)\ x_{22} \end{pmatrix} \leq 0$$

First of all, for this matrix to be negative semi-definite, its diagonal elements have to be ≤ 0 (note that this is necessary and not sufficient). Furthermore, we can impose $x_{11} = 1$ without loss of generality. So, we get the following inequalities, both of which have to be satisfied:

$$\lambda^2 \geq 0.04 - 0.16\frac{x_{12}}{a} + 0.16\frac{x_{22}}{a^2} \quad =: b_1(a) \ ,$$
$$\lambda^2 \geq 0.04 + 0.16\frac{a\ x_{12}}{x_{22}} + 0.16\frac{a^2}{x_{22}} \quad =: b_2(a) \ .$$

For the ease of notation, we will denote the two right members of these inequalities by $b_1(a)$ and $b_2(a)$. We also have to remember that we are working with a pair of matrices $\{A_1, A_2\}$, and so these inequalities must hold for $a = a_1$ and $a = a_2$. This allows us to write that

$$\hat{\rho}^2(A_1, A_2) \geq \max(b_1(a_1), b_2(a_1), b_1(a_2), b_2(a_2)) \ . \tag{A.1}$$

At first, we will determine when $b_1(a_1) > b_1(a_2)$ and when $b_2(a_1) > b_2(a_2)$ (still assuming that $a_2 > a_1$).

1.

$$\begin{aligned} b_1(a_1) > b_1(a_2) \Rightarrow 0.04 - 0.16\frac{x_{12}}{a_1} + 0.16\frac{x_{22}}{a_1^2} &> 0.04 - 0.16\frac{x_{12}}{a_2} + 0.16\frac{x_{22}}{a_2^2} \\ x_{22}(\frac{1}{a_1^2} - \frac{1}{a_2^2}) &> x_{12}(\frac{1}{a_1} - \frac{1}{a_2}) \\ \frac{x_{12}}{x_{22}} &< \frac{1}{a_1} + \frac{1}{a_2} \quad (\text{as } a_2 > a_1) \ . \end{aligned}$$

2.

$$\begin{aligned} b_2(a_1) > b_2(a_2) \Rightarrow 0.04 + 0.16\frac{a_1\ x_{12}}{x_{22}} + 0.16\frac{a_1^2}{x_{22}} &> 0.04 + 0.16\frac{a_2\ x_{12}}{x_{22}} + 0.16\frac{a_2^2}{x_{22}} \\ \frac{a_1\ x_{12}}{x_{22}} + \frac{a_1^2}{x_{22}} &> \frac{a_2\ x_{12}}{x_{22}} + \frac{a_2^2}{x_{22}} \\ x_{12}(a_1 - a_2) &> a_2^2 - a_1^2 \quad (\text{as } x_{22} > 0) \\ x_{12} &< -a_2 - a_1 \end{aligned}$$

Taking into account that $x_{22} > 0$, this divides the (x_{12}, x_{22}) plane into 3 zones, which are depicted in figure A.1.

In zone 1, $b_1(a_1) > b_1(a_2)$ and $b_2(a_1) > b_2(a_2)$, so, to determine the value of the right member of (A.1), we only have to compare $b_1(a_1)$ with $b_2(a_1)$. Let us check that in this case, $b_1(a_1) > b_2(a_1)$:

$$0.04 - 0.16\frac{x_{12}}{a_1} + 0.16\frac{x_{22}}{a_1^2} >^? 0.04 + 0.16\frac{a_1\ x_{12}}{x_{22}} + 0.16\frac{a_1^2}{x_{22}} \ ,$$
$$-\frac{x_{12}}{a_1} + \frac{x_{22}}{a_1^2} >^? \frac{a_1\ x_{12}}{x_{22}} + \frac{a_1^2}{x_{22}} \ .$$

We can replace x_{12} by $(-2a_1 - \epsilon)$, with $\epsilon > 0$, because we are in the case $x_{12} < -a_2 - a_1$, and $a_2 > a_1$. This yields :

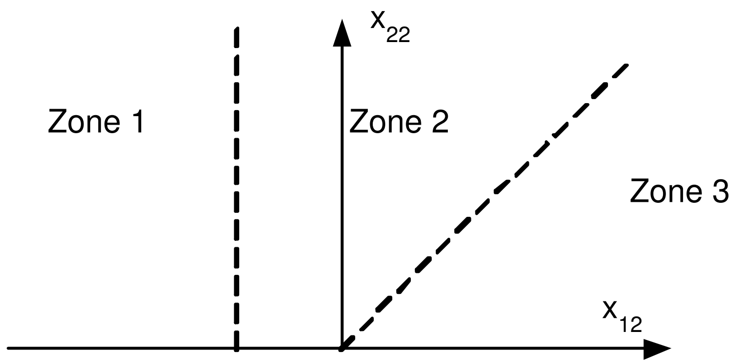

Fig. A.1. The 3 zones of the plane (x_{12}, x_{22}).

$$\frac{2a_1+\epsilon}{a_1}+\frac{x_{22}}{a_1^2} >^? \frac{a_1(-2a_1-\epsilon)}{x_{22}}+\frac{a_1^2}{x_{22}} \ ,$$
$$\frac{2a_1+\epsilon}{a_1}+\frac{x_{22}}{a_1^2} >^! \frac{-a_1\epsilon}{x_{22}}-\frac{a_1^2}{x_{22}} \ ,$$

which is true, as the left member is positive and the right member is negative. This proves that, in zone 1, $b_1(a_1)$ is the largest value. It now remains to minimize it among all possible values of (x_{12}, x_{22}) of zone 1. This gives (without the 0.04 term and the 0.16 factor) :

$$\inf_{(x_{12},x_{22})\in\ Zone1} -\frac{x_{12}}{a_1}+\frac{x_{22}}{a_1^2} \ ,$$

added to the constraint that $x_{22} > x_{12}^2$ because $X > 0$. Now, for a fixed x_{12}, the best x_{22} is the smallest possible, that is $x_{22} = x_{12}^2$ in the limit case. Another element is that $x_{12} < 0$, so we may want to take the smallest one in absolute value, as is appears with the minus sign in the expression. So, we get $x_{12} = -a_1 - a_2$ and

$$\hat{\rho}^2 \geq 0.04 + 0.16\ [\frac{a_1+a_2}{a_1}+\frac{(-a_1-a_2)^2}{a_1^2}] \ ,$$
$$\hat{\rho}^2 \geq 0.04 + 0.16\ \frac{2a_1^2+3a_1a_2+a_2^2}{a_1^2} > 0.04 + 0.16\ \frac{a_2}{a_1} \ .$$

In zone 2, $b_1(a_1) > b_1(a_2)$ and $b_2(a_1) < b_2(a_2)$. For the same reasons as detailed before, it remains to compare $b_1(a_1)$ with $b_2(a_2)$. In details, it gives :

$$\begin{aligned}
b_1(a_1) > b_2(a_2) \iff 0.04 - 0.16\frac{x_{12}}{a_1} + 0.16\frac{x_{22}}{a_1^2} &> 0.04 + 0.16\frac{a_2\ x_{12}}{x_{22}} + 0.16\frac{a_2^2}{x_{22}} \\
-\frac{x_{12}}{a_1} + \frac{x_{22}}{a_1^2} &> \frac{a_2\ x_{12}}{x_{22}} + \frac{a_2^2}{x_{22}} \\
-x_{12}x_{22}a_1 + x_{22}^2 &> a_1^2 a_2\ x_{12} + a_1^2 a_2^2 \\
x_{22}^2 - a_1^2 a_2^2 &> a_1 x_{12}(x_{22} + a_1 a_2) \\
x_{22} - a_1 a_2 &> a_1 x_{12}\ .
\end{aligned}$$

In this case, there are two sub-zones, as shown in figure A.2 :

- zone 2a, where $b_1(a_1) > b_2(a_2)$ (when $x_{22} > a_1 x_{12} + a_1 a_2$)
- zone 2b, where $b_1(a_1) < b_2(a_2)$ (when $x_{22} < a_1 x_{12} + a_1 a_2$)

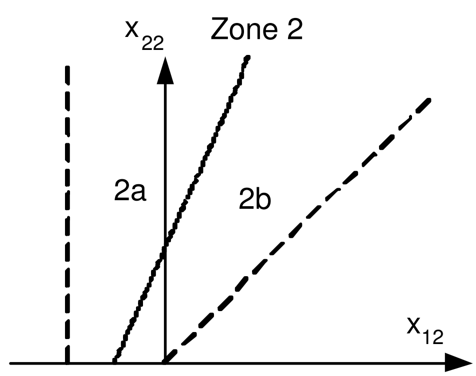

Fig. A.2. The 2 sub-zones of zone 2 in the plane (x_{12}, x_{22}).

In case 2a, the value to study is $b_1(a_1)$. The goal is then to minimize $-\frac{x_{12}}{a_1} + \frac{x_{22}}{a_1^2}$ over values of (x_{12}, x_{22}) in Zone 2a. The constraints are : $x_{12} < \frac{x_{22}}{a_1} - a_2$ and $x_{12}^2 < x_{22}$. We can notice that the function to minimize is similar to the first contraint and we immediately get that $-\frac{x_{12}}{a_1} + \frac{x_{22}}{a_1^2} > \frac{a_2}{a_1}$. So,

$$\hat{\rho}^2 \geq b_1(a_1) > 0.04 + 0.16\frac{a_2}{a_1}\ .$$

In case 2b, the value to study is $b_2(a_2)$. The goal is then to minimize $\frac{a_2\ x_{12}}{x_{22}} + \frac{a_2^2}{x_{22}}$ over values of (x_{12}, x_{22}) in Zone 2b. The constraints are : $x_{12} > \frac{x_{22}}{a_1} - a_2$, $x_{12}^2 < x_{22}$ and $\frac{x_{12}}{x_{22}} < \frac{1}{a_1} + \frac{1}{a_2}$. For a fixed $x_{22} > 0$, $b_2(a_2)$ is an increasing function of x_{12}. As we have $x_{12} > \frac{x_{22}}{a_1} - a_2$, we deduce that

$$\frac{a_2\ x_{12}}{x_{22}} + \frac{a_2^2}{x_{22}} > \frac{a_2\ (\frac{x_{22}}{a_1} - a_2)}{x_{22}} + \frac{a_2^2}{x_{22}} = \frac{a_2}{a_1}\ ,$$

which yields

$$\hat{\rho}^2 \geq b_2(a_2) > 0.04 + 0.16\frac{a_2}{a_1}\ .$$

In zone 3, $b_1(a_1) < b_1(a_2)$ and $b_2(a_1) < b_2(a_2)$. For the same reasons as detailed before, it remains to compare $b_1(a_2)$ with $b_2(a_2)$. Let us check that in this case, $b_2(a_2) > b_1(a_2)$:

$$0.04 + 0.16\frac{a_2\ x_{12}}{x_{22}} + 0.16\frac{a_2^2}{x_{22}} >^? 0.04 - 0.16\frac{x_{12}}{a_2} + 0.16\frac{x_{22}}{a_2^2}\ ,$$
$$\frac{a_2\ x_{12}}{x_{22}} + \frac{a_2^2}{x_{22}} >^? -\frac{x_{12}}{a_2} + \frac{x_{22}}{a_2^2}\ .$$

We can notice that the left member is increasing in x_{12} while the right member is decreasing in x_{12}. So we replace x_{12} by its minimal value, that is $\frac{a_1+a_2}{a_1 a_2}x_{22}$. This yields :

$$\frac{a_1 + a_2}{a_1} + \frac{a_2^2}{x_{22}} >^? -\frac{a_1 + a_2}{a_1 a_2^2}x_{22} + \frac{x_{22}}{a_2^2}\ ,$$
$$1 + \frac{a_2}{a_1} + \frac{a_2^2}{x_{22}} >^! -\frac{a_2}{a_1 a_2^2}x_{22}\ ,$$

which is true, as the left member is positive and the right member is negative. This proves that, in zone 3, $b_2(a_2)$ is the largest value. It now remains to minimize it among all possible values of (x_{12}, x_{22}) of zone 3. This gives (without the 0.04 term and the 0.16 factor) :

$$\inf_{(x_{12},x_{22})\in\ Zone3} \frac{a_2\ x_{12}}{x_{22}} + \frac{a_2^2}{x_{22}}\ ,$$

added to the constraint that $x_{22} > x_{12}^2$ because $X > 0$. Now, for a fixed x_{12} (> 0 because we are in zone 3), the best x_{22} is the largest possible, that is $x_{22} = x_{12}\frac{a_1 a_2}{a_1+a_2}$ in the limit case. So, we get

$$\hat{\rho}^2 \geq 0.04 + 0.16\ [\frac{a_1 + a_2}{a_1} + \frac{(a_1 + a_2)a_2^2}{x_{12}a_1 a_2}]\ ,$$
$$\hat{\rho}^2 \geq 0.04 + 0.16\ [1 + \frac{a_2}{a_1} + \frac{(a_1 + a_2)a_2}{x_{12}a_1}] > 0.04 + 0.16\ \frac{a_2}{a_1}\ .$$

This exhaustive study shows that a necessary condition for the inequality $A^*XA - \lambda^2 X \leq 0$ to be satisfied with $X > 0$ is that $\lambda^2 > 0.04 + 0.16\ \frac{a_2}{a_1}$, and so

$$\hat{\rho}^2(A_1, A_2) \geq 0.04 + 0.16\ \frac{a_2}{a_1}\ .$$

This concludes the exhaustive calculations detailing the non-necessity of the existence of a CQLF for the stability of a discrete time switched linear system.

References

1. Tsuyoshi Ando and Mau-hsiang Shih. Simultaneous contractibility. *SIAM J. Matrix Anal. Appl.*, 19(2):487–498, 1998.
2. G. Aumann. Subdivision of linear corner cutting curves. *J. Geom. Graph.*, 1(2):91–103, 1997.
3. N.E. Barabanov. Lyapunov indicator of discrete inclusions. I. II. III. *Autom. Remote Control*, 49(2,3,5):152–157, 283–287, 558–565, 1988.
4. Marc A. Berger and Yang Wang. Bounded semigroups of matrices. *Linear Algebra and its Applications*, (166):21–27, 1992.
5. J. Berstel and M. Pocchiola. A geometric proof of the enumeration formula for sturmian words. *Inter. J. Algebra Comput.*, 3:349–355, 1993.
6. Pierre-Alexandre Bliman and Giancarlo Ferrari-Trecate. Stability analysis of discrete-time switched systems through Lyapunov functions with nonminimal states. In *IFAC Conference on the Analysis and Design of Hybrid Systems (ADHS03)*, 2003.
7. Vincent D. Blondel and Alexander Megretski. *Unsolved Problems in Mathematical Systems & Control Theory*, chapter Problem 10.2 : When is a pair of matrices stable ?, by Blondel Vincent, Theys Jacques, Tsitsiklis John., pages 304–308. Princeton University Press, 2004.
8. Vincent D. Blondel and Alexander Megretski. *Unsolved Problems in Mathematical Systems & Control Theory.* Princeton University Press, 2004.
9. Vincent D. Blondel and Yurii Nesterov. Computationally efficient approximations of the joint spectral radius. *SIAM Journal on Matrix Analysis and Applications*, 2005.
10. Vincent D. Blondel, Jacques Theys, and Alexander A. Vladimirov. Switched systems that are periodically stable may be unstable. In *Fifteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2002)*, 2002.
11. Vincent D. Blondel, Jacques Theys, and Alexander A. Vladimirov. An elementary counterexample to the finiteness conjecture. *SIAM J. Matrix Anal. Appl.*, 24(4):963–970, 2003.
12. Vincent D. Blondel and John N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000.

13. Vincent D. Blondel and John N. Tsitsiklis. The boundedness of all products of a pair of matrices is undecidable. *Systems and Control Letters*, 41(2):135–140, 2000.
14. Theys Jacques. Blondel Vincent. Discrete and continuous stability of switched linear systems. In *22nd Benelux Meeting on Systems and Control, Lommel, Belgium, March 19-21, 2003.*, 2003.
15. Theys Jacques. Blondel Vincent. On the relations between discrete and continuous time stability for switched linear systems. In *Proceedinfs of the Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Katholieke Universiteit Leuven, Belgium, July 5-9, 2004.*, 2004.
16. Theys Jacques. Blondel Vincent, Nesterov Yurii. Approximations of the rate of growth of switched linear systems. In *Lecture Notes in Computer Science, ISSN: 0302-9743, volume 2993 / 2004. Hybrid Systems: Computation and Control: 7th International Workshop, HSCC 2004, Philadelphia, PA, USA, March 25-27, 2004.*, pages 173–186. Springer-Verlag GmbH, ISBN: 3-540-21259-0, 2004.
17. Theys Jacques. Blondel Vincent, Nesterov Yurii. Computing the joint spectral radius of a set of matrices. In *23rd Benelux Meeting on Systems and Control, Helvoirt, The Netherlands, March 17-19, 2004.*, 2004.
18. Theys Jacques. Blondel Vincent, Nesterov Yurii. On the accuracy of the ellipsoid norm approximation of the joint spectral radius. *Linear Algebra and its Applications*, 394:91–107, 2005.
19. T. Bousch and J. Mairesse. Asymptotic height optimization for topical ifs, tetris heaps, and the finiteness conjecture. *Journal of the American Mathematical Society*, (15):77–111, 2002.
20. Roger W. Brockett. Optimization theory and the converse of the circle criterion. In *National Electronics Conference*, volume 21, 1965.
21. A.S. Cavaretta, W. Dahmen, and C.A. Micchelli. Stationary subdivision. *Mem. Amer. Math. Soc.*, 93(453):186, 1991.
22. David Collela and Christopher Heil. The characterization of continuous, four-coefficient scaling functions and wavelets. *IEEE Trans. Inf. Th., Special Issue on Wavelet Transforms and Multiresolution Signal Analysis*, 38:876–881, 1992.
23. David Collela and Christopher Heil. Characterization of scaling functions. I. Continuous solutions.. *SIAM J. Matrix Anal. Appl.*, 15:496–518, 1994.
24. Ingrid Daubechies. Orthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:909–996, 1988.
25. Ingrid Daubechies and Jeffrey C. Lagarias. Two-scale difference equations: I. Existence and global regularity of solutions. *SIAM J. Math. Anal.*, 22:1388–1410, 1991.
26. Ingrid Daubechies and Jeffrey C. Lagarias. Sets of matrices all infinite products of which converge. *Linear Algebra Appl.*, 161:227–263, 1992.
27. Ingrid Daubechies and Jeffrey C. Lagarias. Two-scale difference equations: II. Local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.*, 23:1031–1079, 1992.
28. Ingrid Daubechies and Jeffrey C. Lagarias. Corrigendum/addendum to: Sets of matrices all infinite products of which converge. *Linear Algebra Appl.*, 327(1-3):69–83, 2001.

29. G. Davrazos and N.T. Koussoulas. A Review of Stability Results for Switched and Hybrid Systems. In *9th Mediterranean Conference on Control and Automation (MED2001)*, 2001.
30. Wijesuriya P. Dayawansa and C.F. Martin. A converse Lyapunov theorem for a class of dynamical systems which undergo switching. *IEEE Transactions on Automatic Control*, 44(4):751–760, 1999.
31. G. De Rham. Un peu de mathematique a propos d'une courbe plane. *Revue de Mathematiques Elementaires. II.*, 4(5):678–689, 1947.
32. Raymond A. Decarlo, Michael S. Branicky, Pettersson Stefan, and Bengt Lennartson. Perspectives and Results on the Stability and Stabilizability of Hybrid Systems. *Proceedings of the IEEE*, 88(7):1069–1082, 2000.
33. N. Dyn, J.A. Gregory, and D. Levin. Analysis of linear binary subdivision schemes for curve design. *Constr.Approx.*, 7:127–147, 1991.
34. Ludwig Elsner. The Generalized Spectral-Radius Theorem : An Analytic-Geometric Proof. *Linear Algebra Appl.*, 220:151–159, 1995.
35. Ludwig Elsner and S. Friedland. Norm Conditions for Convergence of Infinite Products. *Linear Algebra Appl.*, 250:133–142, 1997.
36. J.R. Gantmacher. *Matrix Theory, Vol. I and II.* Chelsea Publishing Co., 1960.
37. Keith Glover. All optimal Hankel-norm approximations if linear multivariable systems and their L^{∞}-error bounds. *International Journal of Control*, 39(6):1115–1193, 1984.
38. Amara L. Graps. An Introduction to Wavelets. *IEEE Computational Sciences and Engineering*, 2(2):50–61, 1995.
39. Gustaf Gripenberg. Computing the joint spectral radius. *Linear Algebra Appl.*, 234:43–60, 1996.
40. N. Guglielmi and M. Zennaro. On the asymptotic properties of a family of matrices. *Linear Algebra Appl.*, 322(1-3):169–192, 2001.
41. N. Guglielmi and M. Zennaro. On the limit products of a family of matrices. *Linear Algebra Appl.*, 363:11–27, 2003.
42. Leonid Gurvits. Stability of discrete linear inclusion. *Linear Algebra and its Applications*, 231:47–85, 1995.
43. Christopher Heil and David Collela. *Dilation equations and the smoothness of compactly supported wavelets, Wavelets : Mathematics and Applications*, pages 161–200. J. Benedetto and M. Frazier, eds, CRC Press, 1993.
44. Christopher Heil and Gilbert Strang. Continuity of the joint spectral radius: Application to wavelets. In *Bojanczyk, Adam (ed.) et al., Linear algebra for signal processing. Papers based on lectures presented at the IMA workshop, held at IMA, University of Minnesota, Minneapolis, MN, USA, April 6-10, 1992. New York, NY: Springer-Verlag. IMA Vol. Math. Appl. 69, 51-61* . 1995.
45. Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.
46. Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.
47. Ralph Howard. The John Ellipsoid Theorem. Lecture at the University of South Carolina, November 1997.
48. Ali Jadbabaie, Jie Lin, and A. Stephen Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

49. F. John. *Extremum problems with inequalities as subsidiary conditions, Studies and Essays Presented to R.Courant on his 60th birthday*, pages 187–204. Interscience Publishers, Inc., New-York, N.Y., January 8, 1948.
50. Thomas Kailath. *Linear Systems.* Prentice Hall, Inc, Eaglewood Cliffs, NJ, 1980.
51. V.S. Kozyakin. Algebraic unsolvability of problem of absolute stability of desynchronized systems. *Autom. Remote Control*, 51(6):754–759, 1990.
52. Jeffrey. C. Lagarias and Yang Wang. The finiteness conjecture for the generalized spectral radius of a set of matrices. *Linear Algebra and its Applications*, 214:17–42, 1995.
53. Daniel Liberzon, João P. Hespanha, and A. Stephen Morse. Stability of switched systems: a lie-algebraic condition. *Systems ans Control Letters*, 37(3):117–122, 1999.
54. Daniel Liberzon and A.Stephen Morse. Basic problems in stability and design of switched systems. *IEEE Control Systems Magazine*, October 1999.
55. M. Lothaire. *Algebraic combinatorics on words.* Encyclopedia of Mathematics and Its Applications. 90. Cambridge: Cambridge University Press. xiii, 504 p., 2002.
56. Mohsen Maesumi. An efficient lower bound for the generalized spectral radius of a set of matrices. *Linear Algebra Appl.*, 240:1–7, 1996.
57. Mohsen Maesumi. Joint Spectral Radius and Hölder Regularity of Wavelets. *Computers and Mathematics with Applications.*, 40:145–155, 2000.
58. E. Mainar and J.M. Pena. Error analysis of corner cutting algorithm. *Numer. Algorithms.*, 22(1):41–52, 1999.
59. C.A. Micchelli and H. Prautzsch. Uniform refinement of curves. *Linear Alg.Appl.*, 114/115:841–870, 1989.
60. Filippo Mignosi. On the number of factors of sturmian words. *Theor. Comput. Sci.*, 82(1):71–84, 1991.
61. J. Milnor. *Singular Points of Complex Hypersurfaces.* Princeton University Press, New Jersey, 1969.
62. B.E. Moison, A. Orlitsky, and P.H. Siegel. On codes that avoid specified differences. *IEEE Trans. Inform. Theory*, 47:433–442, 2001.
63. Luc Moreau. Stability of multi-agent systems with time-dependent communication links. *IEEE Trans. Automat. Control*, 50(2):169– 182, 2005.
64. Zeev Nehari. *Conformal Mapping.* Dover Publications, Inc. New-York, 1952.
65. V.Yu. Protasov. The joint spectral radius and invariant sets of the several linear operators. *Fundamentalnaya i Prikladnaya Matematika.*, 2(1):205–231, 1996.
66. V.Yu. Protasov. The generalized joint spectral radius. The geometric approach. *Izvestiya Akademii nauk. Seriya Matematicheskaya.*, 61(5):99–136, 1997.
67. V.Yu. Protasov. A generalization of the joint spectral radius: the geometrical approach. *Facta Univ. Ser. Math. Inform.*, 13:19–23, 1998.
68. V.Yu. Protasov. Asymptotics of the partition function. *Mat. Sb.*, 191(3):65–98, 2000. translation in Sb. Math. 191 (2000), no. 3-4, 381–414.
69. V.Yu. Protasov. The stability of subdivision operator at its fixed point. *SIAM Journal of Math. Analysis.*, 33(2):448–460, 2001.
70. V.Yu. Protasov. On the regularity of de Rham curves. *Izvestiya Mathematics.*, 68(3):27–68, 2004.

71. V.Yu. Protasov. Refinement equations and corresponding linear operators. *to appear in the International Journal of Wavelets, Multiresolution and Information Processing.*, 2005.
72. Gian-Carlo Rota and W.Gilbert Strang. A note on the joint spectral radius. *Nederl. Akad. Wet., Proc., Ser. A*, 63:379–381, 1960.
73. H. Samelson. *Notes on Lie Algebras.* Van Nostrand Reinhold Co., New York., 1969.
74. Strogatz S.H. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D*, 143:1–20, 2000.
75. Strogatz S.H. *Sync: The Emerging Science of Spontaneous Order.* Hyperion, 2003.
76. I. Sheipak. Nontrivial fractals in the plane and linear operators with joint spectral radius equal to 1. *Math. Notes.*, 63(5):701–705, 1998.
77. Gilbert Strang. Wavelet transforms versus fourier transforms. *Bull. Amer. Math. Soc.*, 28:288–305, 1993.
78. Gilbert Strang. The joint spectral radius. commentary by gilbert strang on paper number 5. *Collected works of Gian-Carlo Rota*, 2001.
79. John N. Tsitsiklis and Vincent D. Blondel. The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible - to compute and to approximate. *Math. Control Signals Syst.*, 10(1):31–40, 1997.
80. T. Vicsek, A. Czirok, E. Ben Jacob, I. Cohen, and O. Schochet. Novel type of phase transitions in a system of self-driven particles. *Physical Review Letters*, 75:1226–1229, 1995.
81. Alexander Vladimirov, Ludwig Elsner, and Wolf-Jürgen Beyn. Stability and paracontractivity of discrete linear inclusions. *Linear Algebra Appl.*, 312(1-3):125–134, 2000.
82. J.L. Willems. *Stability Theory of Dynamical Systems.* Thomas Nelson & Sons Ltd., 1970.
83. Fabian Wirth. Dynamics of time-varying discrete-time linear systems: Spectral theory and the projected system. *SIAM J. Control Optimization*, 36(2):447–487, 1998.
84. Fabian Wirth. On the calculation of real time-varying stability radii. *Int. J. Robust & Nonlinear Control*, 8:1043–1058, 1998.
85. Fabian Wirth. The generalized spectral radius and extremal norms. *Linear Algebra Appl.*, 342(1-3):17–40, 2002.
86. Kuramoto Y. Cooperative dynamics of oscillator community. *Progress of Theoretical Physics Supplement*, 79:223–240, 1984.
87. D.X. Zhow. The p-norm joint spectral radius for even integers. *Methods Appl.Anal.*, 5:39–54, 1998.