Raphaël M. Jungers

The Joint Spectral Radius

Theory and Applications

March 5, 2009

Springer

To my parents

Non scholae sed vitae discimus

Preface

This monograph is based on the Ph.D. Thesis of the author [58]. Its goal is twofold: First, it presents most research work that has been done during his Ph.D., or at least the part of the work that is related with the joint spectral radius. This work was concerned with theoretical developments (part I) as well as the study of some applications (part II).
As a second goal, it was the author's feeling that a survey on the state of the art on the joint spectral radius was really missing in the literature, so that the first two chapters of part I present such a survey. The other chapters mainly report personal research, except Chapter 5 which presents an important application of the joint spectral radius: the continuity of wavelet functions.

The first part of this monograph is dedicated to theoretical results. The first two chapters present the above mentioned survey on the joint spectral radius. Its minimum-growth counterpart, the *joint spectral subradius,* is also considered. The next two chapters point out two specific theoretical topics, that are important in practical applications: the particular case of nonnegative matrices, and the Finiteness Property.

The second part considers applications involving the joint spectral radius. We first present the continuity of wavelet. We then study the problem of the capacity of codes submitted to forbidden difference constraints. Then we go to the notion of overlap-free words, a problem that arises in combinatorics on words. We then end with the problem of trackability of sensor networks, and show how the theoretical results developed in the first part allow to solve this problem efficiently.

Brussels, March 2009 *R. Jungers*

Acknowledgements

The research reported here has benefitted from scientific interactions with many researchers among which Pierre-Antoine Absil, Noga Alon, Paul Bell, Francine Blanchet-Sadri, Stephen Boyd, Jean-Charles Delvenne, François Glineur, Jean-Loup Guillaume, Julien Hendrickx, Tzvetan Ivanov, Alexandre Megretski, Yurii Nesterov, Justin Palumbo, Pablo Parrilo, Jean-Jacques Quisquater, Alexander Razborov, Eugeny Shatokhin, Paul Van Dooren, Jan Willems,.... May them be thanked for that. In particular, my thanks go to my advisor Vincent Blondel, and Vladimir Protasov, with whom I have been collaborating since my master's thesis.

I also would like to thank sincerely the members of my Ph.D. jury, whose comments have greatly improved the present text: these are Vincent Blondel, Yurii Nesterov, Pablo Parrilo, Jean-Eric Pin, Vladimir Protasov, Michel Rigo, Jean-Pierre Tignol, Paul Van Dooren. I also heartily thank Julien Hendrickx for a careful reading of a preliminary version.

The monograph version of the thesis has been completed while I was in the Universit Libre de Bruxelles (Computer Science Department). I would like to thank the whole team sincerely for its great hospitality.

I would also like to thank sincerely my family and friends for their constant support. In particular, my thanks go to Carole, Chloé and Léonard.

This work has been supported by a F.R.S.-FNRS fellowship. In many occasions It has also benefited from the "Pôle d'Attraction Interuniversitaire DYSCO" initiated by the Belgian federal state and the "Action de Recherche Concertée on Large Graphs and Networks" initiated by the Belgian French Community.

Contents

Introduction

The *joint spectral radius* characterizes the maximal asymptotic growth rate of a point submitted to a switching linear system in discrete time. In the last decades it has been the subject of intense research due to its role in the study of wavelets, switching systems, approximation algorithms, curve design, and many other topics. In parallel with these practical engineering applications, beautiful theoretical challenges have arisen in the effort to understand the joint spectral radius. These two facts make the study of the joint spectral radius a dream of a subject for a Ph.D. thesis, but perhaps a not so easy task.

Indeed by its natural essence, this notion appears in a number of very different fields of mathematics. For instance, since its definition uses norms and eigenvalues, the joint spectral radius is undoubtedly a linear algebra concept, but not only. It has been defined for purposes of analysis of dynamical systems, and the boost of research on this topic came in the middle 90's from its use in numerical analysis: the joint spectral radius appeared to be the concept needed to determine the continuity of wavelets, a tool of high practical importance nowadays. But the range of applications in which the joint spectral radius has proved useful is much wider; it goes from number theory to network security management, from combinatorics on words to signal processing, etc...

Also, the spectrum of theoretical problems one has to cope with when analyzing the joint spectral radius is wide. In order to solve these problems, results from very different disciplines have been put together: Dynamical systems theory, numerical analysis, theoretical computer science and computability theory, abstract algebra and group theory, graph theory, convex optimization and semidefinite programming (SDP), combinatorics, are a few examples of fields of mathematics that have proved helpful for improving our understanding of problems related to the joint spectral radius. A beautiful example is the contribution of SDP-programming whose usefulness to approximate a joint spectral radius has been progressively understood in the last ten years. This particular contribution is still a subject of research on itself, and seems by now not only

to be a state-of-the-art way of approximating the joint spectral radius, but also to bring interesting insight on the very nature of the joint spectral radius.

Undoubtedly, this profusion of different fields of mathematics that have been involved in "the joint spectral radius conquest" does not make its understanding easy. Many researchers with they own (very) personal background, conventions, motivations, notations and definitions have made progress that one who wants to properly understand the joint spectral radius cannot honestly ignore. However, the ideas behind the mathematical constructions are sometimes simpler than they look at first sight. In view of this, we provide in the first part of this monograph a survey on the subject.
In the theoretical survey, which constitutes the first two chapters, we tried to be exhaustive, self-contained, and easily readable at the same time. In order to do that, some proofs differ from the ones given in the literature. Also, the order of presentation of the results does not follow their chronological apparition in the literature, because it allowed sometimes to simplify the proofs. Finally, we decided to split the survey in two chapters: the first one is intended to help the reader to understand the notion of joint spectral radius, by describing its behavior without confusing him with long proofs and theoretical developments, while the second chapter brings the mathematical study of advanced results, and the rigorous demonstrations.

Outline. This monograph is separated in two parts, the first one is dedicated to theoretical and general problems on the joint spectral radius, while the second part is applications-oriented.

The first two chapters form the above mentioned survey: Chapter 1 presents elementary and fundamental results, while Chapter 2 is more involved, and brings the theory necessary to prove the fundamental theorems. In Chapter 1, we compare the results available for the joint spectral radius to its minimum-growth counterpart: the *joint spectral subradius*. Though very interesting and useful in practice, this latter quantity has received far less attention in the literature, perhaps because it has been introduced later. We had the feeling that a rigorous analysis of the basic behavior of this notion was missing.
The remainder of the monograph presents our personal research. We start with two particular theoretical questions: In chapter 3 we analyze the case of nonnegative integer matrices. We show that for these particular sets, it is possible to decide in polynomial time whether the joint spectral radius is exactly equal to zero, exactly equal to one, or larger than one. Moreover it is possible to precisely characterize the growth of the products in the case where the joint spectral radius is exactly equal to one.
In Chapter 4, we analyze the finiteness property. We show that this property holds for nonnegative rational matrices if and only if it holds for pairs of binary matrices. We give a similar result for matrices with negative entries, and we show that the property holds for pairs of 2×2 binary matrices.

The second part of this monograph presents applications of the joint spectral radius. We first present in Chapter 5 the continuity of wavelet functions. Then, in Chapter 6 we go to the capacity of codes submitted to forbidden differences constraints, that can be expressed in terms of a joint spectral radius. We propose two approximation algo-

rithms for the capacity, we show how to efficiently decide whether the capacity is zero, and exhibit a closely related problem that we prove to be NP-hard.
We then turn to a problem in combinatorics on words: estimating the asymptotic growth of the overlap-free language (Chapter 7). We show how this problem is related with the joint spectral radius and related quantities. Thanks to this, we provide accurate estimates for the rate of growth of the number of overlap-free words, a classical problem in combinatorics on words. We also provide algorithms to estimate the joint spectral subradius and the Lyapunov exponent that appear to perform extremely well in practice.
We finally analyze a problem related to graph theory and network security: we present the trackability of sensor networks (Chapter 8) and show how this problem is efficiently tractable.

Part I
Theory

Chapter 1
Basics

Abstract This chapter is the first part of the theoretical survey on the joint spectral radius. We first present precise definitions of the main concepts. We then show that these definitions are well posed, and we present some basic properties on the joint spectral radius. In the last section, we show that these notions are "useful", in the sense that they actually characterize the maximal and minimal growth rates of a switched dynamical system.
This chapter is meant to be a quick survey on the basic behavior of the joint spectral radius. Some of the results presented in this chapter require rather involved proofs. For this reason this chapter is not self-contained, and some proofs are postponed to the next one.
In this introductory chapter, we compare all results for the joint spectral radius to its minimum-growth counterpart: the joint spectral subradius.

A switched linear system in discrete time is characterized by the equation

$$\begin{aligned} x_{t+1} &= A_t x_t : \quad A_t \in \Sigma, \\ x_0 &\in \mathbb{R}^n, \end{aligned} \tag{1.1}$$

where Σ is a set of real $n \times n$ matrices. We would like to estimate the evolution of the vector x, and more particularly (if it exists) the asymptotic growth rate of its norm:

$$\lambda = \lim_{t \to \infty} ||x_t||^{1/t}.$$

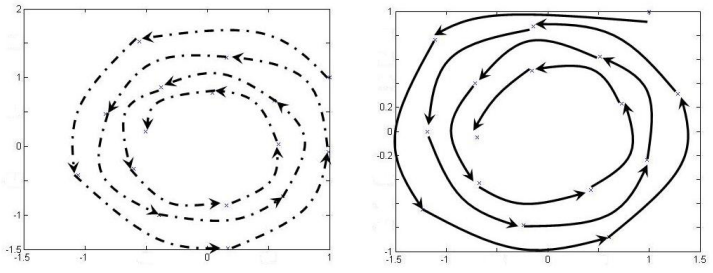
Clearly, one cannot expect that this limit would exist in general. Indeed, even in dimension one, it is easy to design a dynamical system and a trajectory such that the limit above does not exist. Thus a typical relevant question for such a system is the extremal rate of growth: given a set of matrices Σ, what is the maximal value for λ,

over all initial vectors x_0 and all sequences of matrices A_t? In the case of dynamical systems for instance, such an analysis makes a lot of sense. Indeed, by computing the maximal growth rate one can ensure the stability of the system, provided that this growth rate is less than one. We will see that the quantity characterizing this maximal rate of growth of a switched linear discrete time system is the *joint spectral radius,* introduced in 1960 by Rota and Strang [104]. Thanks to its interpretation in terms of dynamical systems, and for many other reasons that we will present later on, it has been widely studied during the last decades.

When the set of matrices consists in a single matrix A, the problem is simple: the maximal growth rate is the largest magnitude of the eigenvalues of A. As a consequence, a matrix is stable if and only if the magnitudes of its eigenvalues are less than one. However, if the set of matrices consists in more than just one matrix, the problem is far more complex: the matrices could well all be stable, while the system itself could be unstable! This phenomenon, which motivates the study of the joint spectral radius, is illustrated by the next example. Consider the set of matrices

$$\Sigma = \left\{ A_0 = \frac{2}{3}\begin{pmatrix} \cos 1.5 & \sin 1.5 \\ -2\sin 1.5 & 2\cos 1.5 \end{pmatrix}, A_1 = \frac{2}{3}\begin{pmatrix} 2\cos 1.5 & 2\sin 1.5 \\ -\sin 1.5 & \cos 1.5 \end{pmatrix} \right\}.$$

The dynamics of these matrices are illustrated in Figure 1.1(a) and (b), with the initial point $x_0 = (1,1)$. Since both matrices are stable ($\rho(A_0) = \rho(A_1) = 0.9428$, where $\rho(A)$, the *spectral radius* of A, is the largest magnitude of its eigenvalues) the trajectories go to the origin. But if one combines the action of A_0 and A_1 alternatively, a diverging behavior occurs (Figure 1.2). The explanation is straightforward: the spectral radius of A_0A_1 is equal to $1.751 > 1$.



(a) (b)

Fig. 1.1 Trajectories of two stable matrices

In practical applications, some other quantities can be of importance, as for instance the *minimal* rate of growth. This concept corresponds to the notion of *joint spectral subradius*. In this introductory chapter, we give definitions for these concepts, as well as some basic results. For the sake of conciseness, and to save time for the reader,

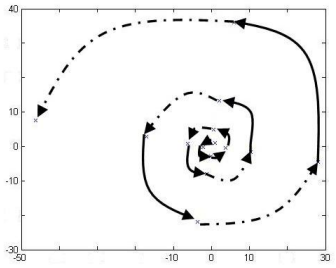


Fig. 1.2 Unstable behavior by combining two stable matrices

we decided not to recall too many basic facts or definitions from linear algebra. We instead refer the reader to classical reference books [45, 72].

In this chapter we first present precise definitions of the main concepts (Section 1.1). In Section 1.2 we show that these definitions are well posed, and we present some basic properties on the joint spectral radius and the joint spectral subradius. In the last section, we show that these notions are "useful", in the sense that they actually characterize the maximal and minimal growth rates of a switched dynamical system of the type (1.1). As the reader will discover, this is not so obvious.
Some of the results presented in this chapter require rather involved proofs. For this reason this chapter is not self-contained, and some proofs are postponed to Chapter 2. Nevertheless we had the feeling that a small chapter with all the basic results could be useful for the reader in order to summarize the basic properties of the joint spectral radius and the joint spectral subradius.

1.1 Definitions

The joint spectral radius characterizes the maximal asymptotic growth rate of the norms of long products of matrices taken in a set Σ. By a *norm*, we mean a function that to any matrix $A \in \mathbb{R}^{n\times n}$ associates a real number $||A||$ such that

- $||A|| \geq 0$, $||A|| = 0 \Leftrightarrow A = 0$,
- $\forall k \in \mathbb{R} : ||kA|| = |k|\,||A||$,
- $||A+B|| \leq ||A|| + ||B||$,
- $||AB|| \leq ||A||\,||B||$.

The latter condition, called *submultiplicativity* is not required in classical definitions of a norm, but in this monograph we will restrict our attention to them, so that all results involving norms have to be understood in terms of submultiplicative norms. Many norms are submultiplicative, and it is for instance the case of any norm induced by a vector norm. So, let $\|\cdot\|$ be a matrix norm, and $A \in \mathbb{R}^{n\times n}$ be a real matrix. It is well

known that the spectral radius of A, that is, the maximal modulus of its eigenvalues, represents the asymptotic growth rate of the norm of the successive powers of A:

$$\rho(A) = \lim_{t\to\infty} \|A^t\|^{1/t}. \tag{1.2}$$

This quantity does provably not depend on the norm used, and one can see that it characterizes the maximal rate of growth for the norm of a point x_t subject to a Linear Time Invariant dynamical system. In order to generalize this notion to a set of matrices Σ, let us introduce the following notation:

$$\Sigma^t \triangleq \{A_1 \dots A_t : A_i \in \Sigma\}.$$

Also, it is a common practice to denote by A^T the transpose of A. It will always be clear from the context whether A^T denotes the transpose of A of the classical exponentiation.
We define the two following quantities that are good candidates to quantify the "maximal size" of products of length t:

$$\hat{\rho}_t(\Sigma, \|\cdot\|) \triangleq \sup\{\|A\|^{1/t} : A \in \Sigma^t\},$$
$$\rho_t(\Sigma) \triangleq \sup\{\rho(A)^{1/t} : A \in \Sigma^t\}.$$

For a matrix $A \in \Sigma^t$, we call $||A||^{1/t}$ and $\rho(A)^{1/t}$ respectively the *averaged norm* and the *averaged spectral radius* of the matrix, in the sense that it is averaged with respect to the length of the product. We also abbreviate $\hat{\rho}_t(\Sigma, \|\cdot\|)$ into $\hat{\rho}_t(\Sigma)$ or even $\hat{\rho}_t$ if this is clear enough with the context. Rota and Strang introduced the *joint spectral radius* as the limit [104]:

$$\hat{\rho}(\Sigma) \triangleq \lim_{t\to\infty} \hat{\rho}_t(\Sigma, \|\cdot\|). \tag{1.3}$$

This definition is independent of the norm used by the equivalence of the norms in $\mathbb{R}^n$. Daubechies and Lagarias introduced the generalized spectral radius as [33]:

$$\rho(\Sigma) \triangleq \limsup_{t\to\infty} \rho_t(\Sigma).$$

We will see in the next chapter that for bounded sets of matrices these two quantities are equal. Based on this equivalence, we use the following definition:

Definition 1.1 *The* joint spectral radius *of a bounded set of matrices Σ is defined by:*

$$\rho(\Sigma) = \limsup_{t\to\infty} \rho_t(\Sigma) = \lim_{t\to\infty} \hat{\rho}_t(\Sigma).$$

Example 1.1. Let us consider the following set of matrices:

$$\Sigma = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \right\}.$$

The spectral radius of both matrices is one. However, by multiplying them, one can obtain the matrix

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

whose spectral radius is equal to two. Hence, $\rho(\Sigma) \geq \sqrt{2}$, since

$$\lim_{t\to\infty} \hat{\rho}_t(\Sigma) \geq \lim ||A^{t/2}||^{1/t} = \sqrt{2}.$$

Now, $\hat{\rho}_2 = \sqrt{2}$ (where we have chosen the maximum column-sum for the norm) and, as we will see below, $\hat{\rho}_t$ is an upper bound on ρ for any t. So we get $\rho(\Sigma) = \sqrt{2}$.

As the reader will see, the proof of the equivalence between the joint spectral radius and the generalized spectral radius necessitates some preliminary work so as to be presented in a natural way. Before to reach this proof, we continue to make the distinction between the joint spectral radius $\hat{\rho}(\Sigma)$ and the generalized spectral radius $\rho(\Sigma)$.

Let us now interest ourself to the minimal rate of growth. We can still define similar quantities, describing the minimal rate of growth of the spectral radius and of the norms of products in Σ^t. These notions were introduced later than the joint spectral radius ([52], see also [17]).

$$\begin{aligned} \check{\rho}_t(\Sigma, \|\cdot\|) &\triangleq \inf\{\|A\|^{1/t} : A \in \Sigma^t\}, \\ \underline{\rho}_t(\Sigma) &\triangleq \inf\{\rho(A)^{1/t} : A \in \Sigma^t\}. \end{aligned}$$

Then, the *joint spectral subradius* is defined as the limit:

$$\check{\rho}(\Sigma) \triangleq \lim_{t\to\infty} \check{\rho}_t(\Sigma, \|\cdot\|), \tag{1.4}$$

Which is still independent of the norm used by equivalence of the norms in $\mathbb{R}^n$. We define the *generalized spectral subradius* as

$$\underline{\rho}(\Sigma) \triangleq \lim_{t\to\infty} \underline{\rho}_t.$$

Again, we will see that for bounded sets of matrices these two quantities are equal, and we use the following definition:

Definition 1.2 *The* joint spectral subradius *of a set of matrices Σ is defined by:*

$$\check{\rho}(\Sigma) = \lim_{t\to\infty} \check{\rho}_t = \lim_{t\to\infty} \underline{\rho}_t.$$

Example 1.2. Let us consider the following set of matrices:

$$\Sigma = \left\{ \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 3 \end{pmatrix} \right\}.$$

The spectral radius of both matrices is greater than one. However, by multiplying them, one can obtain the zero matrix, and then the joint spectral subradius is zero.

The above examples are simple but, as the reader will see, the situation is sometimes much more complex.

1.2 Basic results

In this section, we review basic results on the joint spectral characteristics, that allow to understand what they are and what they are not. We first present the fundamental theorems, proving the equality for the joint and generalized spectral radii (resp. subradii). We then present basic properties of the joint spectral characteristics, some of which had to our knowledge not yet been formalized.

1.2.1 Fundamental theorems

The joint spectral radius

First, recall that we defined $\hat{\rho}$ as a limit, and not as a limsup. This is due to a classical result, known as *Fekete's Lemma*:

Lemma 1.1 *[43] Let $\{a_n\} : n \geq 1$ be a sequence of real numbers such that*

$$a_{m+n} \leq a_m + a_n.$$

Then the limit

$$\lim_{n\to\infty} \frac{a_n}{n}$$

exists and is equal to $\inf\{\frac{a_n}{n}\}$.

In the above lemma, the limit can be equal to $-\infty$, but this is not possible in our case since the sequence is nonnegative. We are now in position to prove the convergence:

Lemma 1.2 *For any bounded set $\Sigma \subset \mathbb{R}^{n\times n}$, the function $t \to \hat{\rho}_t(\Sigma)$ converges when $t \to \infty$. Moreover,*

$$\lim_{t\to\infty} \hat{\rho}_t(\Sigma) = \inf\{\hat{\rho}_t(\Sigma)\}.$$

Proof. Since the norms considered are submultiplicative, the sequence

$$\log\left(\sup\{\|A\| : A \in \Sigma^t\}\right) = \log \hat{\rho}_t^t$$

is subadditive. That is,

$$\log \hat{\rho}_{t+t'}^{t+t'} \leq \log \hat{\rho}_t^t + \log \hat{\rho}_{t'}^{t'}.$$

If for all t, $\hat{\rho}_t \neq 0$, then by Fekete's lemma,

$$\frac{1}{t} \log \hat{\rho}_t^t = \log \hat{\rho}_t$$

converges and is equal to $\inf \log \hat{\rho}_t$.
If there is an integer t such that $\hat{\rho}_t = 0$, then clearly, for all $t' \geq t, \hat{\rho}_{t'} = 0$, and the proof is done.

Unlike the maximal norm, the behavior of the maximal spectral radius, ρ_t is not as simple, and in general the limsup in the definition of $\rho(\Sigma)$ cannot be replaced by a simple limit. In the following simple example, $\limsup \rho_t = 1$, but $\lim \rho_t$ does not exist:

$$\Sigma = \left\{ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right\}.$$

Indeed, for any t, $\rho_{2t}(\Sigma) = 1$, but $\rho_{2t+1}(\Sigma) = 0$.

The Joint Spectral Radius Theorem.
It is well known that the spectral radius of a matrix satisfies $\rho(A^k) = \rho(A)^k$, $\rho(A) = \lim \|A^t\|^{1/t}$. One would like to generalize these relations to "inhomogeneous" products of matrices, that is, products where factors are not all equal to a same matrix A. This is possible, as has been proved in 1992 by Berger and Wang [5] in the so-called *Joint Spectral Radius Theorem*:

For bounded sets, the values $\hat{\rho}(\Sigma)$ and $\rho(\Sigma)$ are equal.

No elementary proof is known for this theorem. Elsner [41] provides a self-contained proof that is somewhat simpler than (though inspired by) the original one from [5]. Since both proofs use rather involved results on the joint spectral radius, we postpone an exposition of the proof to the next chapter. The reader can check that the elementary facts presented in the remainder of this chapter do not make use of this result.
Observe that the joint spectral radius theorem cannot be generalized to unbounded sets of matrices, as can be seen on the following example:

$$\Sigma = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \dots \right\}.$$

Indeed for this set we have $\rho(\Sigma) = 1$, while $\hat{\rho}(\Sigma) = \infty$.

The joint spectral subradius

Let us now consider the joint spectral subradius. It appears that now both $\underline{\rho}_t$ and $\check{\rho}_t$ converge:

Proposition 1.1 *For any set $\Sigma \subset \mathbb{R}^{n\times n}$, the function $t \to \check{\rho}_t(\Sigma)$ converges when $t \to \infty$, and*

$$\lim \check{\rho}_t(\Sigma) = \inf_{t>0} \check{\rho}_t(\Sigma).$$

Moreover, the function $t \to \underline{\rho}_t(\Sigma)$ converges when $t \to \infty$, and

$$\lim \underline{\rho}_t(\Sigma) = \inf_{t>0} \underline{\rho}_t(\Sigma).$$

Proof. Again the sequence $\log(\check{\rho}_n)$ is subadditive, which proves the first part. Let us now prove the second assertion. We define $\underline{\rho} = \liminf \underline{\rho}_t$ and we will show that the limit actually exists. Fix an $\varepsilon > 0$. For any sufficiently long t, we will construct a product of matrices $B \in \Sigma^t$ such that $||B||^{1/t} \leq \underline{\rho} + \varepsilon$, and thus $\rho(B)^{1/t} \leq \underline{\rho} + \varepsilon$. Indeed, for any norm $||\cdot||$ and any matrix A, the relation

$$\rho(A) \leq ||A||$$

always holds. In order to do that, we will pick a matrix A whose spectral radius is small enough, and we define $B = A^k C$, where C is a short product that does not perturb too much the norm of B.
By the definition of $\underline{\rho}$, there exist $T \in \mathbb{N}, A \in \Sigma^T$ such that $\rho(A)^{1/T} \leq \underline{\rho} + \varepsilon/4$. Since $\rho(A) = \lim ||A^k||^{1/k}$, one gets $\lim ||A^k||^{1/kT} \leq \underline{\rho} + \varepsilon/4$ and there exists an integer k_0 such that for all $k > k_0$, $||A^k||^{1/kT} \leq \underline{\rho} + \varepsilon/2$.
Let us first define a real number M such that for each length $t' \leq T$, there is a product C of length t' such that $||C|| \leq M$. Next, there is an integer T_0 large enough so that $M^{1/T_0} \leq (\underline{\rho} + \varepsilon)/(\underline{\rho} + \varepsilon/2)$.
Now, for any length $t > \max\{k_0 T, T_0\}$, we define $t' < T$ such that $t = kT + t'$, and we construct a product of length $t : B = A^k C$, such that $C \in \Sigma^{t'}$, and $||C|| \leq M$. Finally

$$||B||^{1/t} \leq (\underline{\rho} + \varepsilon/2) \frac{\underline{\rho} + \varepsilon}{\underline{\rho} + \varepsilon/2} \leq \underline{\rho} + \varepsilon.$$

We also have the equality between $\check{\rho}$ and $\underline{\rho}$; moreover in this case the set need not be bounded;

Theorem 1.1 *[111] For any set of matrices Σ,*

$$\lim_{t\to\infty} \inf\{\rho(A)^{1/t} : A \in \Sigma^t\} = \lim_{t\to\infty} \inf\{||A||^{1/t} : A \in \Sigma^t\} \triangleq \check{\rho}(\Sigma).$$

Proof. Clearly,

$$\liminf_{t\to\infty}\{\rho(A)^{1/t} : A \in \Sigma^t\} \leq \liminf_{t\to\infty}\{||A||^{1/t} : A \in \Sigma^t\}$$

because for any matrix A, $\rho(A) \leq ||A||$.
Now, for any matrix $A \in \Sigma^t$ with averaged spectral radius r close to $\underline{\rho}(\Sigma)$, the product $A^k \in \Sigma^{kt}$ is such that $||A^k||^{1/kt} \to r$ so that

$$\liminf_{k\to\infty}\{||A||^{1/kt} : A \in \Sigma^{kt}\} \leq r.$$

1.2.2 Basic properties

1.2.2.1 Scaling property

Proposition 1.2 *For any set $\Sigma \in \mathbb{R}^{n\times n}$ and for any real number α,*

$$\hat{\rho}(\alpha\Sigma) = |\alpha|\hat{\rho}(\Sigma),$$

$$\check{\rho}(\alpha\Sigma) = |\alpha|\check{\rho}(\Sigma).$$

Proof. This is a simple consequence of the relation $||\alpha A|| = |\alpha|\,||A||$.

1.2.2.2 Complex matrices vs. real matrices

From now on, all matrices are supposed to be real-valued. This is not a restriction as we can consider complex matrices acting on $\mathbb{C}^{n\times n}$ as real operators acting on $\mathbb{R}^{2n\times 2n}$.

1.2.2.3 Invariance under similarity

Proposition 1.3 *For any bounded set of matrices Σ, and any invertible matrix T,*

$$\rho(\Sigma) = \rho(T\Sigma T^{-1}).$$

$$\check{\rho}(\Sigma) = \check{\rho}(T\Sigma T^{-1}).$$

Proof. This is due to the fact that for any product $A_1 \dots A_t \in \Sigma^t$, the corresponding product in $T\Sigma T^{-1}$ is $TA_1 \dots A_t T^{-1}$, and has equal spectral radius.

1.2.2.4 The joint spectral radius as an infimum over all possible norms

The following result has been known for long, since it was already present in the seminal paper of Rota and Strang [104]. Nevertheless, it is very interesting, as it characterizes the joint spectral radius in terms of the matrices in Σ, without considering any product of these matrices. We give here a simple self-contained proof due to Berger and Wang [5].

Proposition 1.4 *For any bounded set Σ such that $\hat{\rho}(\Sigma) \neq 0$, the joint spectral radius can be defined as*

$$\hat{\rho}(\Sigma) = \inf_{||\cdot||} \sup_{A \in \Sigma} \{||A||\}.$$

From now on, we denote by Σ^* the *monoid generated by* Σ:

$$\Sigma^* \triangleq \cup_{t=0}^{\infty} \Sigma^t,$$

With $\Sigma^0 \triangleq I$. If we exclude Σ^0 from the above definition, we obtain Σ^+, the *semigroup generated by* Σ:

$$\Sigma^+ \triangleq \cup_{t=1}^{\infty} \Sigma^t.$$

Proof. Let us fix $\varepsilon > 0$, and consider the set $\tilde{\Sigma} = (1/(\hat{\rho} + \varepsilon))\Sigma$. Then, all products of matrices in $\tilde{\Sigma}^*$ are uniformly bounded, and one can define a norm $|\cdot|$ on $\mathbb{R}^n$ in the following way: $|x| = \max\{|Ax|_2 : A \in \tilde{\Sigma}^*\}$, where $|\cdot|_2$ is the Euclidean vector norm. Remark that in the above definition, the maximum can be used instead of the supremum, because $\rho(\tilde{\Sigma}) < 1$. The matrix norm induced by this latter vector norm, that is, the norm defined by

$$||A|| = \max_{|x|=1} \{|Ax|\},$$

clearly satisfies $\sup_{A \in \tilde{\Sigma}} \{||A||\} \leq 1$, and so $\sup_{A \in \Sigma} \{||A||\} \leq \hat{\rho} + \varepsilon$.

1.2.2.5 Common reducibility

We will say that a set of matrices is *commonly reducible,* or simply *reducible* if there is a nontrivial linear subspace (i.e. different from $\{0\}$ and $\mathbb{R}^n$) that is invariant under all matrices in Σ. This property is equivalent to the existence of an invertible matrix T that "block-triangularizes simultaneously" all matrices in Σ :

$$\Sigma \text{ reducible} \quad \Leftrightarrow \quad \exists T, n' : \forall A_i \in \Sigma, TA_iT^{-1} = \begin{pmatrix} B_i & C_i \\ 0 & D_i \end{pmatrix} : D_i \in \mathbb{R}^{n' \times n'}.$$

We will say that a set of matrices is *commonly irreducible,* or simply *irreducible* if it is not commonly reducible.

Proposition 1.5 *With the notations defined above, if Σ is bounded and reducible,*

$$\begin{aligned}\rho(\Sigma) &= \max\{\rho(\{B_i\}),\rho(\{D_i\})\},\\ \check{\rho}(\Sigma) &\geq \max\{\check{\rho}(\{B_i\}),\check{\rho}(\{D_i\})\},\\ \hat{\rho}(\Sigma) &= \max\{\hat{\rho}(\{B_i\}),\hat{\rho}(\{D_i\})\}.\end{aligned} \tag{1.5}$$

Proof. The first two relations follow from the invariance under similarity (Proposition 1.3), together with the following elementary facts:

$$\begin{pmatrix} B_1 & C_1 \\ 0 & D_1 \end{pmatrix} \cdot \begin{pmatrix} B_2 & C_2 \\ 0 & D_2 \end{pmatrix} = \begin{pmatrix} B_1B_2 & B_1C_2+C_1D_2 \\ 0 & D_1D_2 \end{pmatrix},$$

$$\rho\left(\begin{pmatrix} B & C \\ 0 & D \end{pmatrix}\right) = \max\{\rho(B),\rho(D)\}.$$

The third relation is more technical, and is proved by showing that extradiagonal blocks cannot increase the exponent of growth. We can suppose $A_i \in \Sigma$ block-triangular, still by invariance under similarity. Let us denote M the maximal joint spectral radius among the diagonal blocks:

$$M = \max\{\hat{\rho}(\{B_i\}),\hat{\rho}(\{D_i\})\}.$$

We define the norm $||\cdot||$ as the sum of the absolute values of the entries. Clearly $\hat{\rho}(\Sigma) \geq M$, and we now prove the reverse inequality.

Writing

$$A_i = \begin{pmatrix} 0 & C_i \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} B_i & 0 \\ 0 & D_i \end{pmatrix},$$

we have

$$||A_t \ldots A_1|| = ||B_t \ldots B_1|| + ||D_t \ldots D_1|| + ||\sum_{r=1}^{t} B_t \ldots B_r C_r D_{r-1} \ldots D_1||.$$

Now for any ε there is a natural T such that for all $t \geq T$,

$$\hat{\rho}_t(\{B_i\}),\hat{\rho}_t(\{D_i\}) < (M+\varepsilon)^t.$$

Thus, for t large enough we can bound each term in the summation above by $O((M+\varepsilon)^t)$:
if $T < r < t-T$, then

$$||B_t \ldots B_r C_r D_{r-1} D_1|| \leq ||C_r||(M+\varepsilon)^{t-1},$$

and in the other case (say, $r \leq T$, the other case is similar),

$$||B_t \ldots B_r C_r D_{r-1} D_1|| < ||C_r||(\hat{\rho}_1)^r (M+\varepsilon)^{t-r-1} = O((M+\varepsilon)^t).$$

Recall that $\hat{\rho}_1$ is the supremum of the norms of the matrices in Σ. Finally, $||A_t \dots A_1|| \leq 2(M+\varepsilon)^t + tO((M+\varepsilon)^t)$, and $\hat{\rho}(\Sigma) \leq M+\varepsilon$.

It is straightforward that the above proposition generalizes inductively to the case where there are more than two blocks on the diagonal.

In the above proposition, Equation (1.5) enlightens a fundamental difference between the joint spectral radius and the joint spectral subradius. For this latter quantity, the inequality cannot be replaced by an equality. This is due to the fact that the joint spectral subradius is the *minimum* growth of a quantity (the spectral radius) which is by essence a *maximum* (over all eigenvalues of a matrix). Consider the next example:

$$\Sigma = \left\{ \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \right\}.$$

The joint spectral subradius of the first diagonal entries is 2, and this is also the case for the set of the second diagonal entries. However, the joint spectral subradius of Σ is equal to $\sqrt{8} > 2$.

1.2.2.6 Three members inequalities

Proposition 1.6 *For any bounded set $\Sigma \in \mathbb{R}^{n\times n}$ and for any natural t,*

$$\rho_t(\Sigma) \leq \rho(\Sigma) \leq \hat{\rho}_t(\Sigma). \tag{1.6}$$

Proof. The left hand side inequality is due to the fact that $\rho(A^k) = \rho(A)^k$. The right hand side is from Fekete's lemma (Lemma 1.1).

Let us add that this has been generalized to unbounded sets to what is called the four members inequality [33, 35]:

$$\rho_t(\Sigma) \leq \rho(\Sigma) \leq \hat{\rho}(\Sigma) \leq \hat{\rho}_t(\Sigma).$$

For the joint spectral subradius, it appears that both quantities $\underline{\rho}_t$ and $\check{\rho}_t$ are in fact upper bounds:

Proposition 1.7 *For any bounded set $\Sigma \in \mathbb{R}^{n\times n}$ and for any natural t,*

$$\check{\rho}(\Sigma) \leq \underline{\rho}_t(\Sigma) \leq \check{\rho}_t(\Sigma).$$

Proof. The left hand side inequality is due to the fact that $\rho(A^k) = \rho(A)^k$, implying that $\underline{\rho}_{kt} \leq \underline{\rho}_t$. The right hand side is a straightforward consequence of the property $\rho(A) \leq ||A||$.

1.2.2.7 Closure and convex hull

Taking the closure or the convex hull of a set does not change its joint spectral radius. For the closure, we also prove the result for the generalized spectral radius, since it will be needed in further developments.

Proposition 1.8 *[111] For any bounded set* $\Sigma \in \mathbb{R}^{n\times n}$

$$\hat{\rho}(\Sigma) = \hat{\rho}(conv\,\Sigma) = \hat{\rho}(cl\,\Sigma),$$

$$\rho(\Sigma) = \rho(cl\,\Sigma).$$

Proof. For the convex hull, observe that for all $t > 0 : \hat{\rho}_t(\text{conv}\Sigma) = \hat{\rho}_t(\Sigma)$. Indeed, all products in $(\text{conv}\Sigma)^t$ are convex combinations of products in Σ^t, and are thus less or equally normed. The equalities for the closure hold because for all t, $\rho_t(\text{cl}\,\Sigma) = \rho_t(\Sigma)$, and $\hat{\rho}_t(\text{cl}\,\Sigma) = \hat{\rho}_t(\Sigma)$, by continuity of the norm and the eigenvalues.

We now show the counterpart for the joint spectral subradius. The property still holds for the closure, but not for the convex hull:

Proposition 1.9 *For any bounded set* $\Sigma \in \mathbb{R}^{n\times n}$

$$\check{\rho}(\Sigma) = \check{\rho}(cl\,\Sigma),$$

but the equality $\check{\rho}(\Sigma) = \check{\rho}(conv\,\Sigma)$ *does not hold in general.*

Proof. The equality $\check{\rho}_t(\text{cl}\Sigma) = \check{\rho}_t(\Sigma)$ still holds for all t by continuity of the norm and the matrix multiplication.
On the other hand, consider the simple example $\Sigma = \{1,-1\} \subset \mathbb{R}$. All products have norm one, and so $\check{\rho} = \rho = 1$, but $0 \in \text{conv}\,\Sigma$, and so $\check{\rho}(\text{conv}\,\Sigma) = 0$.

1.2.2.8 Continuity

We show here that the joint spectral radius of bounded sets of matrices is continuous in their entries. Recall that the Hausdorff distance measures the distance between sets of points in a metric space:

$$d(\Sigma,\Sigma') \triangleq \max\{\sup_{A\in\Sigma}\{\inf_{A'\in\Sigma'}||A-A'||\}, \sup_{A'\in\Sigma'}\{\inf_{A\in\Sigma}||A-A'||\}\}.$$

Proposition 1.10 *The joint spectral radius of bounded sets of matrices is continuous with respect to the Hausdorff distance in* $\mathbb{R}^{n\times n}$.
That is, for any bounded set of matrices $\Sigma \in \mathbb{R}^{n\times n}$, *and for any* $\varepsilon > 0$, *there is a* $\delta > 0$ *such that*

$$d(\Sigma,\Sigma') < \delta \Rightarrow |\hat{\rho}(\Sigma) - \hat{\rho}(\Sigma')| < \varepsilon.$$

Proof. Let us fix $\varepsilon > 0$. By Proposition 1.4, there exists a norm $||\cdot||$ such that

$$\hat{\rho}_1(\Sigma) = \sup\{||A|| : A \in \Sigma\} \leq \hat{\rho}(\Sigma) + \varepsilon/2.$$

Let us now pick a set Σ' close enough to Σ : $d(\Sigma, \Sigma') < \varepsilon/2$. By definition of the Hausdorff distance, we have

$$\forall A' \in \Sigma', \exists A \in \Sigma : ||A' - A|| < \varepsilon/2,$$

and we can bound the norm of any matrix in Σ' :

$$||A'|| = ||A + (A' - A)|| \leq \hat{\rho}(\Sigma) + \varepsilon/2 + \varepsilon/2 = \hat{\rho}(\Sigma) + \varepsilon.$$

By applying the same argument to Σ, we obtain $|\hat{\rho}(\Sigma) - \hat{\rho}(\Sigma')| \leq \varepsilon$.

Let us note that this proposition does not generalize to unbounded sets, as shown by the next example:

$$\Sigma = \left\{ \begin{pmatrix} 0 & 0 \\ \varepsilon & 0 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 0 & n \\ 0 & 0 \end{pmatrix} : n \in \mathbb{N} \right\}.$$

Indeed for $\varepsilon = 0$ we have $\hat{\rho}(\Sigma) = 0$, while for any $\varepsilon > 0$ we have $\hat{\rho}(\Sigma) = \infty$.
Let us add that Wirth has proved that the joint spectral radius is even locally Lipschitz continuous on the space of compact irreducible sets of matrices endowed with the Hausdorff topology [115, 117].

Surprisingly, a similar continuity result for the joint spectral subradius is not possible. It appears that this quantity is only lower semicontinuous:

Proposition 1.11 *The joint spectral subradius of bounded sets of matrices is lower semicontinuous with respect to the Hausdorff distance in $\mathbb{R}^{n\times n}$.*
That is, for any bounded set of matrices $\Sigma \in \mathbb{R}^{n\times n}$, and for any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$d(\Sigma, \Sigma') < \delta \Rightarrow \check{\rho}(\Sigma') < \check{\rho}(\Sigma) + \varepsilon.$$

Proof. Let us fix $\varepsilon > 0$. By Proposition 1.1, there exists a t and a product $A \in \Sigma^t$ such that

$$\rho(A)^{1/t} \leq \check{\rho}(\Sigma) + \varepsilon/2.$$

Let us now pick a set Σ' close enough to Σ. By continuity of the eigenvalues there exists a product $A' \in \Sigma'^t$ with averaged spectral radius $\rho(A')^{1/t} < \rho(A)^{1/t} + \varepsilon/2$, and $\check{\rho}(\Sigma') < \check{\rho}(\Sigma) + \varepsilon$.

To prove that the joint spectral subradius is not continuous, we introduce the following example.

Example 1.1 *Consider the set*

$$\Sigma = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ -\frac{1}{k} & 1 \end{pmatrix} \right\}.$$

Where $k \in \mathbb{N}$. When $k \to \infty$, the joint spectral subradius of these sets is equal to zero (the product $(A_1 A_0^k)^2$ is the zero matrix). However these sets tend to

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

whose joint spectral subradius is equal to 1. *Indeed any matrix in the semigroup is nonnegative, and has the lower right entry equal to one.*

1.2.2.9 Zero spectral radius

The case where the joint spectral radius (resp. joint spectral subradius) is equal to zero is of practical importance for obvious reasons. In order to state them, following to [44], we introduce the following definition, which holds for the rest of this monograph, unless specified otherwise. A *polynomial time algorithm* is an algorithm that takes an instance and delivers an answer "yes" or "no", after having performed a number of elementary operations that is bounded by a fixed polynomial in the size of the instance, where the size of the instance is its "bit size", that is, the number of bits necessary to encode it.
The following two results are not trivial. Their proofs are to be found in Chapter 2:

Proposition 1.12 *There is a polynomial time algorithm allowing to decide whether the joint spectral radius of a set of matrices is zero.*

Proposition 1.13 *There is no algorithm allowing to decide whether the joint spectral subradius of a set of matrices is zero, that is, this problem is undecidable.*

1.3 Stability of dynamical systems

As explained in the introduction, one possible use of the joint spectral radius is to characterize the maximal asymptotic behavior of a dynamical system. But is this exactly what we are doing, when we compute a joint spectral radius? The notion of stability of a dynamical system (like the system defined in Equation (1.1)) is somewhat fuzzy in the literature, and many different (and not equivalent) definitions appear. According to the natural intuition, and to the more commonly used definition, we introduce the next definition:

Definition 1.3 *A switched dynamical system*

$$\begin{aligned} x_{t+1} &= A_t x_t : \quad A_t \in \Sigma, \\ x_0 &\in \mathbb{R}^n, \end{aligned} \tag{1.7}$$

is stable *if for any initial condition* $x_0 \in \mathbb{R}^n$, *and any sequence of matrices* $\{A_t\}$, $\lim_{t\to\infty} x_t = 0$.

Clearly, if $\rho(\Sigma) < 1$, then the dynamical system is stable, because $x_t = Ax_0$, with $A \in \Sigma^t$, and so $|x_t| \leq ||A||\,|x_0| \to 0$. But the converse statement is less obvious: could the condition $\rho < 1$ be too strong for stability? Could it be that for any length, one is able to provide a product of this length that is not too small, but yet that any *actual trajectory,* defined by an infinite sequence of matrices, is bound to tend to zero? The next example shows that such a case appears with unbounded sets:

Example 1.2 *Let*

$$\Sigma = \left\{ A = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \cup \left\{ B_k = \begin{pmatrix} 0 & k \\ 0 & 0 \end{pmatrix}, k \in \mathbb{R} \right\}.$$

For any length t, $\hat{\rho}_t = \infty$, *but one can check easily that every infinite product tends to zero. To see this, observe that a left-infinite product must be of one of these three forms, each of which tends to zero*

$$\begin{aligned} ||\ldots AA|| &\approx (1/2)^t, \\ ||\ldots A\ldots AB_kA|| &\approx k(1/2)^{t-1}, \\ ||\ldots A\ldots AB_kA\ldots AB_{k'}A|| &= 0. \end{aligned}$$

The following theorem ensures that such a pathological situation does not appear with bounded sets:

Theorem 1.2 *[5] For any bounded set of matrices* Σ, *there exists a left-infinite product* $\ldots A_2A_1$ *that does not converge to zero if and only if* $\rho(\Sigma) \geq 1$.

The proof of this theorem is not trivial, and makes use of results developed in the next chapter. The reader will find a proof of this important result in Section 2.1.

This proves that the joint spectral radius rules the stability of dynamical systems:

Corollary 1.1 *For any bounded set of matrices* Σ, *the corresponding switched dynamical system is stable if and only if* $\rho(\Sigma) < 1$.

In the above theorem, the boundedness assumption cannot be removed, as shown by Example 1.2.

The equivalent problem for the joint spectral subradius is obvious: For any bounded set of matrices Σ, the corresponding switched dynamical system is stabilizable (*i.e.* there exists an infinite product of matrices whose norm tends to zero) if and only if $\check{\rho}(\Sigma) < 1$. Indeed, if $\check{\rho} < 1$, there exists a real γ, and a finite product $A \in \Sigma^t$ such that $||A|| \leq \gamma < 1$, and $\lim_{k\to\infty} A^k = 0$. On the other hand, if $\check{\rho} \geq 1$, then for all $A \in \Sigma^t : ||A|| \geq 1$, and so no long product of matrices tends to zero. There is however a

nontrivial counterpart to Corollary 1.1. To see this, let us rephrase Theorem 1.2 in the following corollary:

Corollary 1.2 *For any bounded set of matrices Σ, there is an infinite product of these matrices reaching the joint spectral radius. More precisely, there is a sequence of matrices $A_0, A_1, \ldots$ of Σ such that*

$$\lim_{t\to\infty} ||A_t \ldots A_1||^{1/t} = \rho(\Sigma).$$

Proof. The proof is a direct consequence of the proof of Theorem 1.2, see Section 2.1.

The idea of this corollary can be transposed to the following result on the joint spectral subradius:

Theorem 1.3 *For any (even unbounded) set of matrices Σ, there is an infinite product of these matrices reaching the joint spectral subradius:*

$$\exists A_{t_j} \in \Sigma : \lim_{i\to\infty} ||A_{t_i} \ldots A_{t_1}||^{1/t} = \check{\rho}(\Sigma).$$

Proof. Let Σ be a set of matrices. Thanks to the definition and Theorem 1.1, for every natural k there exists a product $B_k \in \Sigma^{n_k}$ of a certain length n_k such that

$$||B_k||^{1/n_k} < \check{\rho} + \frac{1}{2^k}.$$

Now the sequence $||B_t \ldots B_1||^{1/\sum_{1\leq k\leq t} n_k}$ tends to $\check{\rho}$. However, this only provides a product $\ldots A_2A_1$ such that $\liminf ||A_t \ldots A_1||^{1/t} = \check{\rho}$. In order to replace the liminf by a limit, for all k we define c_k to be the maximal norm of all the suffixes of B_k, and one can raise the matrix B_k to a sufficiently large power p_k such that for any suffix C of B_{k+1},

$$||CB_k^{p_k}||^{1/t} < c_{k+1}^{1/t} ||B_k^{p_k}||^{1/t} < \check{\rho} + \frac{1}{2^{k-1}},$$

and finally the sequence $||\Pi_t||$ converges, where Π_t is the suffix of length t of the left infinite product $\ldots B_2^{p_2} B_1^{p_1}$.

1.4 Conclusion

The goal of this chapter was to understand properly the notions of joint spectral radius and joint spectral subradius in a glance. As the reader has seen, even some basic facts, such as the equivalence between the joint and generalized spectral radii, require some advanced results. We have thus decided to postpone this proof to Chapter 2. There, the result will naturally follow from a careful study of a particular problem related to the joint spectral radius, namely the *defectiveness* of a set of matrices.

Further elementary properties of the joint spectral radius of sets of matrices can be found in [20, 94, 115, 116].

Chapter 2
Classical results and problems

Abstract In this chapter we review major results on the joint spectral radius. Our goal is to remain concise, but at the same time exhaustive and self-contained. We begin by analyzing in detail the growth of matrix products, and by presenting the concept of extremal norms. Existence of extremal norms is an encouraging result, since it is easier to evaluate the joint spectral radius when an extremal norm is available. We found it natural to follow with calculability/complexity theorems, which are on the other hand discouraging. In a subsequent section, we present methods of computation and approximation of the joint spectral radius. In view of the negative results of the second section, the reader shall not be surprised to find algorithms whose efficiency is often rather poor (at least theoretically). In the last section of this chapter we present a fascinating question: the finiteness property.

2.1 Defectivity and extremal norms

2.1.1 Defectivity

We start with a first result that sheds light on the growth of long matrix products. From the basic results in the previous chapter, we know that $\hat{\rho}_t$ goes to $\hat{\rho}$ as t goes to infinity, or more precisely:

$$\lim_{t\to\infty} \max\{\|A\|^{1/t} : A \in \Sigma^t\} = \hat{\rho}.$$

However, in some applications, one is interested in a more precise definition of the asymptotic growth: how does the quantity $\hat{\rho}_t^t/\hat{\rho}^t$ evolve with t? Another way to ask this question is: How does the maximal norm evolve when the joint spectral radius is equal to one?

Definition 2.1 *[49, 50] A set of matrices Σ is* nondefective *if there exists $K \in \mathbb{R}$ such that for all t,*

$$\sup\{||A|| : A \in \Sigma^t\} \leq K\hat{\rho}^t.$$

Defectivity appears to be a problem of crucial importance, as we will see all along this section. A first result states that Σ is nondefective if it is irreducible. This has been known for long, and several proofs are available in the literature [3, 41, 115]. We present here a new proof that is somewhat simpler and more intuitive. In the proof of the theorem, the notation $\Sigma^{\leq t}$ represents the products of length less than t in the semigroup Σ^*.

Theorem 2.1 *Let a bounded set Σ be irreducible, and $\hat{\rho}(\Sigma) \neq 0$. Then Σ is nondefective; that is, there is a constant K such that for all $A \in \Sigma^t$,*

$$||A|| \leq K\hat{\rho}^t.$$

Proof. We suppose without loss of generality that $\hat{\rho}(\Sigma) = 1$ (nondefectivity is clearly conserved under scalar multiplication). Let us define

$$V = \{x \in \mathbb{R}^n \mid \sup_{A \in \Sigma^*} |Ax| < \infty\}.$$

By construction, V is a linear subspace, invariant under the matrices in Σ. Since Σ is irreducible, we have $V = \mathbb{R}^n$, or $V = \{0\}$.

If $V = \{0\}$, then for each vector $x \in \mathbb{R}^n$, there exists a product $A \in \Sigma^t$ such that $|Ax| \geq 2|x|$.
We claim that this length t is bounded uniformly from above by a constant T over all x. Indeed, if it is not the case, we can define an increasing sequence $\{t_k\}$, and a sequence x_k of norm 1 such that for all $A \in \Sigma^{\leq t_k}$, $|Ax_k| < 2$. A subsequence of the x_k converges thus to a vector x of norm 1 such that for all $A \in \Sigma^*$, $|Ax| < 2$ and so $V \neq \{0\}$.
Finally if for all x there exists a matrix $A \in \Sigma^{\leq T}$ such that $|Ax| \geq 2|x|$, then $\hat{\rho} \geq 2^{1/T} > 1$, and we have a contradiction.

So, $V = \mathbb{R}^n$, but this implies that Σ^* is bounded.

Theorem 2.1 tells us that if a set of matrices is irreducible, then the quantity $\hat{\rho}_t^t / \hat{\rho}^t$ is bounded from above by a constant. Remark that the equivalent lower bound clearly always holds, by the three members inequality (1.6): For any set of matrices, and for all t,

$$1 \leq \hat{\rho}_t^t / \hat{\rho}^t.$$

2.1.2 Extremal norms

The nondefectivity of Σ allows for a powerful construction, known as *extremal norm*, that we now describe.

We know that the joint spectral radius can be defined as follows (Proposition 1.4):

$$\hat{\rho}(\Sigma) = \inf_{||\cdot|| \; A\in\Sigma} \sup \{||A||\}.$$

So the natural question arises to know whether there is a norm that actually realizes this infimum. This is exactly the concept of an extremal norm.

Definition 2.2 *A norm $||\cdot||$ on $\mathbb{R}^{n\times n}$ is* extremal *for a set of matrices Σ if for all $A \in \Sigma$,*

$$||A|| \leq \hat{\rho}(\Sigma).$$

Let us note that the above definition, together with the three members inequality (1.6) implies that for an extremal norm we have

$$\sup_{A\in\Sigma} ||A|| = \hat{\rho}.$$

Also, following Wirth [117], we introduce two similar notions for vector norms:

Definition 2.3 *A vector norm $|\cdot|$ is* extremal *for Σ if for all $x \in \mathbb{R}^n$, and for all $A \in \Sigma$,*

$$|Ax| \leq \hat{\rho}|x|.$$

A vector norm $|\cdot|$ is a Barabanov norm *for Σ if it is extremal, and if moreover for any vector $x \in \mathbb{R}^n$, there exists a matrix A in the closure of Σ such that*

$$|Ax| = \hat{\rho}|x|.$$

One can directly see that the matrix norm induced by an extremal vector norm would be an extremal matrix norm for Σ. So the first question is: "Does there always exist an extremal matrix norm?" Unfortunately, the answer to this question is negative in general, as can be shown with the following simple example:

Example 2.1 *Let us consider the following set:*

$$\Sigma = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\}.$$

The joint spectral radius of this set is the spectral radius of the matrix, that is, one. But there is no norm that takes the value one for such a matrix. Indeed, by submultiplicativity it would imply that

$$|| \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} || \leq 1,$$

for any k, which is impossible, by the equivalence between the norms in any finite dimensional vector space.

Note that the set Σ in the above example is defective. When it is not the case, the construction of an extremal norm appears to be possible. The following theorem is mainly due to Barabanov [3] for the existence of a Barabanov norm, and to Kozyakin [70] for the existence of an extremal norm.

Theorem 2.2 *[3, 70] A bounded set $\Sigma \in \mathbb{R}^{n\times n}$ admits an extremal norm if and only if it is nondefective.*
If Σ is moreover compact and irreducible, then it admits a Barabanov norm.

Proof. If $\hat{\rho}(\Sigma) = 0$ the theorem is clear, since the existence of an extremal norm, or the nondefectivity, are both equivalent to the fact that Σ only contains the zero matrix. We thus restrict our attention without loss of generality to sets of matrices Σ such that $\hat{\rho}(\Sigma) = 1$.

First if Σ admits an extremal norm then it is nondefective: indeed all the products are bounded by the submultiplicativity property.

We now suppose that Σ is nondefective and we prove the existence of an extremal norm in a constructive way. Let us take an arbitrary vector norm $|\cdot|$ in $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$, we define a new norm in the following way:

$$|x|_* = \sup_{A\in\Sigma^*} |Ax|.$$

Recall that by convention, the identity matrix I is in Σ^*. Taking this into account, one can verify that $|\cdot|_*$ is a norm, and, since by its definition $|Ax|_* \leq |x|_*$ holds for all $x \in \mathbb{R}^n$, this is an extremal vector norm. So it induces an extremal matrix norm $||\cdot||_*$ in $\mathbb{R}^{n\times n}$.

We now turn to the second part of the theorem, and suppose that Σ is compact and irreducible. We provide a Barabanov norm in a constructive way. Let us take an arbitrary vector norm $|\cdot|$ in $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$, we define a new norm as follows:

$$|x|_* = \limsup_{t\to\infty} \sup \{|Ax| : A \in \Sigma^t\}. \tag{2.1}$$

Again it is not difficult to prove that $|\cdot|_*$ is a norm. The critical point is that it is definite: $|x|_* = 0 \Rightarrow x = 0$. Indeed, if $|x|_* = 0$, $x \neq 0$, then the linear space generated by the set $\{Ax : A \in \Sigma^*\}$ is a nontrivial linear subspace, invariant under all the matrices in Σ. Moreover this subspace is not $\mathbb{R}^n$, for otherwise we would have $\hat{\rho}(\Sigma) < 1$. Indeed, by compacity of the unit ball, if for all $x : |x| = 1$,

$$\limsup_{t\to\infty} \sup \{|Ax| : A \in \Sigma^t\} = 0, \tag{2.2}$$

then there must exist a $T \in \mathbb{N}$ such that for all $A \in \Sigma^T$ $|Ax| < 1$, and the induced matrix norm provides

$$\hat{\rho} \leq \hat{\rho}_T < 1.$$

Finally we have a contradiction since Σ is assumed to be irreducible.
Now this is clear that for all $x \in \mathbb{R}^n$, and $A \in \Sigma$, $|Ax|_* \leq |x|_*$, and that, by compacity of Σ, for all x there exists an $A \in \Sigma$ such that $|Ax|_* = |x|_*$.

Clearly, the interest of the above theorem is rather theoretical. Indeed, the construction of the norm is not possible in general, as it requires for instance the knowledge of the joint spectral radius itself. However, the knowledge of the existence of an extremal norm can be useful in several cases: some algorithms have for instance been constructed that allow to approximate this extremal norm when it exists, in order to evaluate the joint spectral radius (see Section 2.3). Also, these very concepts allow to prove two fundamental results that we now present: the existence of an infinite product reaching the joint spectral radius, and the joint spectral radius theorem.

2.1.3 Proofs of the fundamental theorems

We start with the existence of an "optimal infinite product":

Theorem 1 *(Theorem 1.2) [5] For any bounded set of matrices Σ, all left-infinite products $\ldots A_2A_1$ converge to zero if and only if $\hat{\rho}(\Sigma) < 1$.*

Proof. $\Leftarrow$: If $\hat{\rho} < 1$, then $\sup\{||A||, A \in \Sigma^t\} \to 0$ when $t \to \infty$, and $A_t \ldots A_1 \to 0$.
$\Rightarrow$: Let us suppose that $\hat{\rho} \geq 1$, and show that the system is not stable. The statement is easy to prove if there exists a Barabanov norm. Indeed, for any $x_0 \in \mathbb{R}^n$, there exists a matrix $A \in \Sigma$ such that $|Ax_0| = |x_0|$. By iterating the construction one gets a trajectory satisfying $|x_t| = |x_0|$ and the system is not stable. So, in view of Theorem 2.2, the proof is done if Σ is compact and irreducible.
Let us now suppose that Σ is only bounded and irreducible. The closure $\mathrm{cl}\,\Sigma$ is compact, and has the same joint spectral radius as Σ (Proposition 1.8). So for any $x \in \mathbb{R}$, there exists an $A \in \mathrm{cl}\,\Sigma$ such that $|Ax| = |x|$. We will approximate each matrix A_i closer and closer with matrices $\tilde{A}_i \in \Sigma$ so that the norm of the whole approximated product $x_t = \tilde{A}_t \ldots \tilde{A}_1 x_0$ is larger than $1/2$. To do that, we define a sequence $0 < \delta_i < 1$ such that $\prod_0^\infty \delta_i = 1/2$. Let us pick an $x_0 \in \mathbb{R}_0^n$. For all $t \geq 1$, we define $A_t \in \mathrm{cl}\,\Sigma$ such that $|A_t x_{t-1}| = |x_{t-1}|$. We approximate each A_t with $\tilde{A}_t \in \Sigma$ such that $||\tilde{A}_t - A_t|| < 1 - \delta_t$. By induction, one can see that $|x_t| \geq \prod_0^t \delta_i$. Indeed, $|x_t| = |\tilde{A}_t x_{t-1}| = |A_t x_{t-1} - (A_t - \tilde{A}_t)x_{t-1}| \geq |x_{t-1}|(1-(1-\delta_t))$.
Finally, if Σ is commonly reducible, and $\hat{\rho}(\Sigma) \geq 1$, we know that there exists a transformation T such that the matrices $TAT^{-1} : A \in \Sigma$ are block-triangular, with each block irreducible, and the restriction of these matrices to the first block has a joint spectral radius greater than or equal to one. We can then apply the above result to this block.

The above proof provides as a direct corollary the existence of a left-infinite product whose average norm converges to the joint spectral radius (Corollary 1.2).

We are now almost in position to prove the joint spectral radius theorem. No elementary proof is known for this theorem. We will present a self-contained proof strongly inspired by the former proof of Berger and Wang [5], and by another published proof due to Elsner [41]. Before presenting it, we begin with a small technical lemma of linear algebra, that states that if a matrix A maps a unitary vector close to itself, then it has an eigenvalue close to 1.

Lemma 2.1 *[41] Let $||\cdot||$ be a matrix norm in $\mathbb{R}^{n\times n}$ induced by the vector norm $|\cdot|$. There is an absolute constant $C(n)>0$ such that for all $z\in\mathbb{R}^n$, $|z|=1$, and all $A\in\mathbb{R}^{n\times n}$, $||A||\leq 1$, there is an eigenvalue λ of A such that*

$$|1-\lambda|\leq C|Az-z|^{1/n}.$$

Theorem 2.3 ***(Joint Spectral Radius Theorem)***
For any bounded set of matrices Σ,

$$\limsup_{t\to\infty}\sup\{\rho(A)^{1/t}:A\in\Sigma^t\}=\lim_{t\to\infty}\sup\{||A||^{1/t}:A\in\Sigma^t\}\triangleq\rho(\Sigma).$$

Proof. We suppose without loss of generality that Σ is closed (taking the closure does not change $\rho(\Sigma)$ nor $\hat{\rho}(\Sigma)$ by Proposition 1.8) and that $\hat{\rho}=1$. Clearly, $\rho(\Sigma)\leq 1$. Let us first suppose that Σ is irreducible. Then, by Theorem 2.2, there exists a Barabanov norm.

Let us pick an $x_0\in\mathbb{R}^n, |x_0|=1$. By the definition of the Barabanov norm, there exists a sequence of matrices $A_1,A_2,\ldots$ such that for all t, $|x_t|=1$, where $x_t=A_t\ldots A_1x_0$. By using the compacity of the unit ball, we know that there exists a subsequence x_{t_i} converging to a vector y of norm one. So we have two lengths $t_i>t_j$ such that

$$|A_{t_i}\ldots A_{t_{j+1}}A_{t_j}\ldots A_1x_0-A_{t_j}\ldots A_1x_0|<\varepsilon,\quad |A_{t_j}\ldots A_1x_0|=1,$$

for any ε.
Setting $z=A_{t_j}\ldots A_1x_0$, we get a matrix $A\in\Sigma^*$ such that

$$|Az-z|<\varepsilon,$$

and we can conclude by Lemma 2.1 that A has an eigenvalue λ such that $|1-\lambda|=O(\varepsilon^{1/n})$, which implies that

$$\sup\{\rho(A):A\in\Sigma^*\}\geq 1.$$

Now, if Σ is not irreducible, since the joint spectral radius is the maximum over the joint spectral radii of each diagonal block (Proposition 1.5), one just has to apply the result to each irreducible block separately.

This ends the first section of this chapter. We have now a fair insight on the behavior of the joint spectral radius: Given a set of matrices, there is always a set of submatrices that is irreducible, and whose joint spectral radius has the same value. Remark that since the conditions for a set of matrices to be commonly reducible can be expressed as quantified polynomial equalities, the problem of finding irreducible components is computable with quantifier elimination methods (see [27]). So in some sense we could restrict our attention to irreducible matrices, for which there exists an extremal norm. However, even if one is able to find the irreducible component of a set of matrices leading to the joint spectral radius, this would not be sufficient to compute its value. Indeed, no constructive method is known for computing the extremal norm of an irreducible sets of matrices. We finally mention that Guglielmi et al. provides sufficient conditions for a set of matrices to admit an extremal norm which is a complex polytope [48]. These conditions are rather strong and are not checkable in practice.

2.2 Complexity

In view of the results in the previous section, the joint spectral radius could seem rather easy to compute: if the set of matrices is reducible it can be decomposed in smaller irreducible matrices without changing the joint spectral radius. And if the matrices are irreducible, there exists a matrix norm for which all the matrices in Σ have a norm smaller than ρ, trivially providing a tight upper bound on ρ, via the three members inequality (1.6). The reality is unfortunately not so easy. In this section we present not less than three results that show that the joint spectral radius is (at least theoretically) extremely hard to compute. These results explore three of the most discouraging arguments: NP-hardness, Turing-undecidability, and non algebraicity. As usually, the proof of these infeasibility results are somewhat artificial, and (to our opinion) the details of the proofs are of little interest for a common reader. Nevertheless, the ideas behind the proofs may give some insight on what is actually difficult in computing a joint spectral radius, and where is the limit between feasibility and infeasibility. For these reasons we limit ourselves to present the main ideas of the theorems, and we provide bibliographic references for the interested reader.
We add that for now on and unless explicitly stated we restrict our attention to finite sets of matrices.

2.2.1 NP-hardness

The first theorem we present is on the NP-hardness of the joint spectral radius approximation, and is valid even for binary matrices. In the following we call the *size* of a number ε its bit size, that is, for instance, if $\varepsilon = p/q$, its size is equal to $\log(pq)$.

Theorem 2.4 *[17] Unless $P = NP$, there is no algorithm that, given a set of matrices Σ and a relative accuracy ε, returns an estimate $\tilde{\rho}$ of $\rho(\Sigma)$ such that $|\tilde{\rho} - \rho| \leq \varepsilon\rho$ in a number of steps that is polynomial in the size of Σ and ε. This is true even if the matrices in Σ have binary entries.*

Proof. The proof proceeds by reduction of SAT whose NP-completeness is well-known [44].

2.2.2 Non algebraicity

The next theorem, due to Kozyakin [70][1], states that there is no algebraic criterion allowing to decide stability of a switched linear system. To state this theorem properly, we consider a finite set of m $n \times n$ matrices as a point $x \in \mathbb{R}^{mn^2}$. So we can talk about the *joint spectral radius of the point x* as the joint spectral radius of the associated set of matrices. We are interested in the set of all such points corresponding, for instance, to $\rho(x) < 1$. For these sets to be easily recognizable, one would like them to be expressed in terms of simple constraints, and for instance, polynomial constraints. That is the notion of *semi-algebraic sets*.

Definition 2.4 *A subset of $\mathbb{R}^n$ is* semi-algebraic *if it is a finite union of sets that can be expressed by a finite list of polynomial equalities and inequalities.*

Theorem 2.5 *[70,111] For all $m,n \geq 2$, the set of points $x \in \mathbb{R}^{mn^2}$ for which $\rho(x) < 1$ is not semi-algebraic.*
For all $m,n \geq 2$, the set of points $x \in \mathbb{R}^{mn^2}$ corresponding to a bounded semigroup (i.e. $\Sigma^*(x)$ bounded) is not semi-algebraic.s

Proof. Kozyakin exhibits a set of matrices depending on the parameter $t \in]0,1[$ and shows that the set of admissible values for t (that involve stability of the corresponding set of matrices) is not a finite number of intervals:

$$G(t) = (1-t^4)\begin{pmatrix} 1 & -\frac{t}{\sqrt{1-t^2}} \\ 0 & 0 \end{pmatrix}, \quad H(t) = (1-t^4)\begin{pmatrix} \sqrt{1-t^2} & -t \\ t & \sqrt{1-t^2} \end{pmatrix}. \tag{2.3}$$

[1] There is actually a flaw in the first version of the proof in [70]. A corrected proof can be found in [111].

The set $\Sigma(t) = \{G(t), H(t)\}$ is unstable for all $t = \sin(2\pi/(2k))$, while it is stable for all $t = \sin(2\pi/(2k+1))$ (see [111] for a proof). Since $W = \{\Sigma(t)\}$ is an algebraic set in $\mathbb{R}^8$, the intersection $W \cap E$ should be made of a finite number of connected components, if E was semi-algebraic. Taking E the set of points corresponding to stable sets, or to sets generating a bounded semigroup, we have a contradiction.

2.2.3 Undecidability

The results that we now present are in a sense even worse than the previous ones, since they teach us that there does not exist in general any algorithm allowing to compute a joint spectral radius in finite time:

Theorem 2.6 *[9, 19] The problem of determining, given a set of matrices Σ, if the semigroup generated by Σ is bounded is Turing-undecidable.*
The problem of determining, given a set of matrices Σ, if $\rho(\Sigma) \leq 1$ is Turing-undecidable.
These two results remain true even if Σ contains only nonnegative rational entries.

Proof. The proof proceeds by reduction from the PFA EMPTINESS problem (Probabilistic Finite state Automaton Emptiness problem), which is known to be undecidable [93]. In this problem, one is given a set of nonnegative rational matrices Σ and two nonnegative rational vectors v_1, v_2. The entries of these matrices and vectors, between zero and one, are interpreted as probabilities. A character is associated to each matrix in Σ; and to a word w (i.e. a sequence of characters) is associated the corresponding product $A_w \in \Sigma^*$. A word w is accepted if its corresponding probability $v_1^T A_w v_2$ is more than a certain given threshold λ. The problem to decide, given $\Sigma, v_1, v_2, \lambda$, whether there exists a word that is accepted is undecidable.

We end this section with an open problem of great practical interest:

Open question 1 *Is there an algorithm that, given a finite set of matrices Σ, decides whether $\rho < 1$?*

This question is important in practice, since it is equivalent to ask for the stability of the dynamical system ruled by the set Σ, in the sense of definition 1.3. We show in Section 2.4 a link between this problem and the famous *finiteness property.*

2.2.4 Similar results for the joint spectral subradius

For the sake of completeness, and because we have the feeling that it is worth to have in mind the different borders between feasibility and infeasibility, we briefly cite

the classical infeasibility results on the joint spectral subradius. They are based on a famous old result by Paterson on the *mortality problem.* In this problem, one is given a set of matrices Σ, and it is asked whether there exists a product of matrices in Σ^* that is equal to zero.

Theorem 2.7 *[92] The mortality problem is undecidable. This is true even for sets of* $2(n_p+1)$ 3×3 *matrices, where* n_p *is any number for which Post's correspondence problem is undecidable.*

Corollary 2.1 *The mortality problem is undecidable for sets of* 16 3×3 *matrices.*

Proof. Matiyasevitch and Sénizergues have shown that Post's correspondence problem is undecidable even for 7 pairs of words [80].

Corollary 2.2 *[18] The mortality problem is undecidable for* pairs *of* 48×48 *matrices.*

Proof. Given a set of m $n\times n$ matrices, Blondel and Tsitsiklis show how to construct a pair of $mn\times mn$ matrices that is mortal if and only if the former set is (see [18] for details).

This latter corollary allows us to prove the following theorem on the approximation of the joint spectral subradius. In order to derive a result as strong as possible, the authors of [17] define a wide class of approximation algorithms, and show that they do not exist for approximating the joint spectral subradius. An algorithm providing the value $\tilde{\rho}$ as an approximation of the joint spectral subradius $\check{\rho}$ of a given set is said to be a (K,L)-approximation algorithm if $|\tilde{\rho}-\check{\rho}|<K+L\check{\rho}$.

Theorem 2.8 *[17, Theorem 2] Let* n_p *be a number of pairs of words for which Post's correspondence problem is undecidable. Fix any* $K>0$ *and* $0<L<1$.

- *There exists no* (K,L)*-approximation algorithm for computing the joint spectral subradius of an arbitrary set* Σ*. This is true even for the special case where* Σ *consists of one* $(6n_p+7)\times(6n_p+7)$ *integer matrix and one* $(6n_p+7)\times(6n_p+7)$ *integer diagonal matrix.*
- *For the special cases where* Σ *consists of two integer matrices with binary entries, there exists no polynomial time* (K,L)*-approximation algorithm for computing the joint spectral subradius unless* $P=NP$.

2.3 Methods of computation

The results in the previous section are in no way good news. However, far from discouraging researchers of trying to approximate the joint spectral radius, it seems that

it has attracted a wealth of motivation in order to cope with these theoretical limitations. In the last decade, many different methods, of very different natures, have been proposed to evaluate the joint spectral radius. Some are heuristics, others rely on brute force methods, others are converging algorithms based on powerful theoretical results. The consequence of this abundance is twofold:
Firstly, the joint spectral radius appears to be easier to handle than one could think in view of the above results. Indeed, we do not have knowledge of a practical problem that is unsolved due to the hardness of a joint spectral radius computation. We will see in the next chapters that in some practical case where this computation is needed, the results of approximation algorithms are of remarkable accuracy.
Secondly, this abundance of available algorithms, each with their own (dis)-advantages, might trouble the practitioner, so that there is some need of classifying the different methods. This part of the research on the joint spectral radius is in our opinion not yet mature, and an exhaustive and precise classification would require a huge amount of work, of implementation and computation, as well as of theoretical investigations. Nevertheless, we give in this section a summary of some families of methods that have proved useful in practical applications. We first show how to decide if the joint spectral radius is exactly equal to zero, which is an important particular case; we then give direct arguments that allow in some situations to compute the joint spectral radius exactly. We next present general methods: branch-and-bound methods, the simple convex combinations method, a geometric method, and Lyapunov methods. Before this last important class of methods, we present a lifting procedure that, combined with other algorithms, allows to reach an arbitrary accuracy.

2.3.1 Zero spectral radius

A special case, important in practice, is when joint spectral characteristics are exactly equal to zero. There is a polynomial time algorithm to decide whether the joint spectral radius of a set of matrices is zero. This algorithm, mentioned in [53] without proof, is a corollary of the following proposition:

Proposition 2.1 *Let* $\Sigma = \{A_1, \ldots, A_m\} \subset \mathbb{R}^{n\times n}$, *Then* $\rho(\Sigma) = 0$ *if and only if*

$$\Sigma^n = \{0\}.$$

This proposition delivers a polynomial time algorithm to check whether a joint spectral radius is zero. Indeed, by defining iteratively:

$$X_0 = I \tag{2.4}$$

$$X_k = \sum_1^m A_i^T X_{k-1} A_i, \tag{2.5}$$

one has $X_n = \sum_{A \in \Sigma^n} A^T A$, and this matrix is computable in polynomial time. Moreover X_n is equal to zero if and only if $\Sigma^n = \{0\}$.

The proof of Proposition 2.1 is based on the following lemma:

Lemma 2.2 *If Σ is irreducible, then $\rho(\Sigma) > 0$.*

Proof. If Σ is irreducible, there exists a real number $\beta > 0$ such that for all x of norm 1, there exists a matrix $A \in \Sigma$ such that $|Ax| \geq \beta$. Indeed if it is not the case by compacity of the unit ball there must exist a vector $x \in \mathbb{R}^n, |x| = 1$ such that for all $A \in \Sigma$, $Ax = 0$, and Σ is not irreducible. Now chose $x_0 \in \mathbb{R}^n_0$, and pick $A_i \in \Sigma$ such that $|x_i| = |A_i x_{i-1}| \geq \beta |x_{i-1}|$. This implies that for any $t > 0$, $||A_t \dots A_1|| \geq \beta^t$, and $\rho \geq \beta > 0$.

We are now in position to prove Proposition 2.1

Proof. The if part is trivial.
The proof of the only if part is by induction on the dimension n: it is true for scalar matrices. Now suppose it is true for sets of matrices of dimension less than n. Let $\Sigma \in \mathbb{R}^{n \times n}, \rho(\Sigma) = 0$. By the previous lemma, we can suppose Σ reducible, and for all $A_i \in \Sigma$,

$$A_i = \begin{pmatrix} B_i & C_i \\ 0 & D_i \end{pmatrix} : D_i \in \mathbb{R}^{n' \times n'},$$

where $\rho(\{B_i\}) = \rho(\{D_i\}) = 0$. Now, consider a product of length n. By applying twice the induction hypothesis on n' and $n - n'$, we have:

$$A_n \dots A_{n'+1} A_{n'} \dots A_1 = \begin{pmatrix} 0 & C \\ 0 & D \end{pmatrix} \begin{pmatrix} B' & C' \\ 0 & 0 \end{pmatrix},$$

for some (potentially zero) matrices C, C', B', D and this latter product vanishes.

2.3.2 Direct arguments

In some cases, a direct argument allows one to compute the joint spectral radius exactly. We present some of these cases here. Other cases can be found in Section 4.3. Recall that a matrix A is said *normal* if $A^T A = A A^T$.

Proposition 2.2 *If Σ is a set of normal matrices, the joint spectral radius is equal to the largest spectral radius of the matrices in Σ.*

Proof. The matrix norm induced by the Euclidean vector norm is given by the largest singular value of the matrix. For normal matrices the largest singular value is also equal to the largest magnitude of the eigenvalues. Thus, $\max\{||A|| : A \in \Sigma\} = \max\{\rho(A) : A \in \Sigma\}$ and from the three members inequality it follows that $\rho(\Sigma) = \max\{\rho(A) : A \in \Sigma\}$.

Corollary 2.3 *[111, Proposition 6.13] If Σ is a set of symmetric matrices, the joint spectral radius is equal to the largest spectral radius of the matrices in Σ.*

Proof. Symmetric matrices are a particular case of normal matrices.

Proposition 2.3 *[111, Proposition 6.13] If Σ is a set of upper triangular matrices, the joint spectral radius is equal to the largest spectral radius of the matrices, that is, the largest absolute value of the diagonal entries.*

The proposition obviously also holds for lower triangular matrices.

Proof. We have seen (Proposition 1.5) that if every matrix A in Σ is block-diagonal with diagonal blocks $[A_{i,i}]$, the joint spectral radius is given by

$$\max_i \rho(\{[A_{i,i}] : A \in \Sigma\}).$$

Now, for triangular matrices, these blocks are just 1×1, and the joint spectral radius is the maximum of the diagonal entries of the matrices, that is, the largest spectral radius of the matrices.

Corollary 2.4 *If the matrices in Σ are* commonly upper triangularizable, *that is, if there exists an invertible matrix T such that for all $A \in \Sigma$, TAT^{-1} is upper triangular, then $\rho(\Sigma) = \rho_1(\Sigma)$.*

Recall that $\rho_1(\Sigma)$ denotes the maximal spectral radius of the matrices in Σ.

Proof. This is due to the fact that the joint spectral radius is invariant under similarity transformations.

We now present another corollary of Proposition 2.3 that is based on a famous result in algebra. We recall that the commutator $[A,B]$ of two matrices A and B is equal to $AB-BA$, and that the linear span of a set of vectors is $\text{span}\{v_1,\ldots,v_n\} = \{\sum \alpha_i v_i : \alpha_i \in \mathbb{R}\}$. We also need the following definitions:

Definition 2.5 *Let Σ be a set of matrices, the* Lie Algebra associated to Σ, *that we denote by $\{\Sigma\}_{LA}$ is the linear span of the set of all the combinations of commutators of matrices in Σ :*

$$\{\Sigma\}_{LA} = g = span\{[A,B],[A,[B,C]],\cdots : A,B,C \in \Sigma\}.$$

The descending sequence of ideals $g^{(k)}$ *of a Lie algebra g is defined inductively: $g^{(1)} = g$, $g^{(k+1)} = [g^{(k)}, g^{(k)}] \subset g^{(k)}$.*
If there exists a $k > 0$ such that $g^{(k)} = \{0\}$, then the Lie Algebra is said to be solvable.

We have the following theorem, known as Lie's Theorem (see [105]):

Theorem 2.9 *Let Σ be a finite set of matrices. If the Lie Algebra associated to Σ is solvable, then Σ is commonly upper triangularizable.*

Corollary 2.5 *Let Σ be a set of matrices. If Σ generates a solvable Lie algebra, then $\rho(\Sigma) = \rho_1(\Sigma)$.*

2.3.3 Branch and bound methods

The first method that comes to mind when one wants to compute the joint spectral radius is to apply the three-members inequality (1.6):

$$\rho_t(\Sigma) \leq \rho(\Sigma) \leq \hat{\rho}_t(\Sigma),$$

remembering that the successive bounds given by the left hand side as well as the right hand side tend to ρ when $t \to \infty$. So an immediate algorithm would be to fix a t, compute all products of length t, take the maximal spectral radius as a lower bound, and the maximal norm (for a fixed norm) as an upper bound. This algorithm can be iterated for increasing values of t, and will converge to the desired value.
The main problem in the previous algorithm is clearly the explosion of the number of products of length t that one needs to compute: there are m^t of them (m is the number of matrices in Σ).
Some ideas have been proposed to attenuate this exponential growth: Maesumi [77] observes that since the spectral radius of a product is invariant under cyclic permutations of the factors, one has to compute only $O(m^t/t)$ products. Gripenberg [46] proposes a branch and bound algorithm that allows to approximate asymptotically the joint spectral radius up to an a priori fixed absolute error. More precisely, given a set Σ and a desired precision δ, the algorithm computes iteratively successive bounds α_t and β_t such that $\alpha_t \leq \rho \leq \beta_t$ and $\lim \beta_t - \alpha_t < \delta$. The algorithm is a branch and bound algorithm in the sense that it builds longer and longer products, based on the ones previously constructed, but removing at each step unnecessary products, that is, products that are provably not necessary to reach the required accuracy.
Also, if the matrices in Σ have nonnegative entries, there is an obvious way of disregarding some products: if A, B are products of length t and $A \leq B$ (where the inequality has to be understood entrywise), then one does not have to keep A in order to have better and better approximations of the joint spectral radius. Indeed, in any product of length $T > t$, one can always replace the subproduct A with B, and by doing this the approximation of the joint spectral radius will be at least as good as with the other product.
As a matter of fact, it is clear that none of these algorithms provide approximations of the joint spectral radius in polynomial time, since this is *NP*-hard, even for nonnegative matrices. However, it is worth mentioning that in practice, these simple algorithms can sometimes provide good approximations of the joint spectral radius, especially if the number and the size of the matrices are not too large.

2.3.4 Convex combination method

The following result provides rapidly a lower bound on the joint spectral radius. Recall that a cone is said *proper* if it is closed, solid, convex and pointed.

Proposition 2.4 *[13] Let $\Sigma = \{A_1, \ldots, A_m\} \in \mathbb{R}^{n\times n}$ be an arbitrary set of matrices; the following simple lower bound on the joint spectral radius holds:*

$$\rho(A_1 + \cdots + A_m)/m \leq \rho(\Sigma).$$

If moreover the matrices in Σ leave a proper cone invariant, then

$$\rho(\Sigma) \leq \rho(A_1 + \cdots + A_m).$$

Proof. The first inequality comes from the fact that $(A_1 + \cdots + A_m)/m \in \operatorname{conv}\Sigma$. The second inequality comes from the fact that associated to an invariant cone K, there exists a norm $||\cdot||_K$, depending on K, such that for all $A, B \in \Sigma$, $||A||_K \leq ||A+B||_K$. This norm is given by

$$||A||_K = \max_{v\in K, w\in K^* ||v||, ||w||=1} w^T A v,$$

where K^* denotes the dual of the cone K (see [13] for details).

2.3.5 A geometric algorithm

If the set of matrices is nondefective it is possible to apply a specific algorithm due to Protasov [94]. The computation time of this algorithm is exponential in the dimension of the matrices, but it has some advantages: it provides a clear geometric interpretation in terms of the construction of an extremal norm, and in particular applications, it has been reported to converge remarkably fast [11]. Finally, in some cases it gives a criterion that allows stopping the algorithm and to compute exactly the joint spectral radius. The idea of the algorithm is to compute iteratively an approximation of the unit ball of the extremal norm, starting with an arbitrary polytope which is symmetric with respect to the origin.

We now briefly describe this algorithm. For all technical details we refer the reader to [94]. For the sake of simplicity we consider the case of two matrices, the case of an arbitrary number of matrices is treated in the same way.

Suppose $A_0, A_1 \in \mathbb{R}^n$ possess an extremal norm; one needs to find a number ρ^* such that $\left|\rho^* - \rho\right|/\rho < \varepsilon$, where $\varepsilon > 0$ is a given accuracy. Consider a sequence of convex polytopes $\{P_k\}$ produced as follows. $P_0 = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n, \quad \sum |x_i| \leq 1\right\}$. For any $k \geq 0$ the polytope P_{k+1} is an arbitrary polytope possessing the following properties: it

is symmetric with respect to the origin, has at most $q(\varepsilon) = C_n \varepsilon^{\frac{1-n}{2}}$ vertices, where C_n is an effective constant depending only on n, and $(1-\varepsilon)\bar{\Sigma}P_k \subset P_{k+1} \subset \bar{\Sigma}P_k$, where $\bar{\Sigma}X = \mathrm{Conv}\{\mathrm{A}_0\mathrm{X}, \mathrm{A}_1\mathrm{X}\}$.

After $T = \left\lceil \frac{3\sqrt{n}\ln\frac{c_2}{c_1}}{\varepsilon} \right\rceil$ steps the algorithm terminates. The value

$$\rho^* = \left(v_{T+1}\right)^{1/(T+1)}$$

gives the desirable approximation of the joint spectral radius. Here v_k is the largest distance from the origin to the vertices of the polytope P_k, c_1, c_2 are such that $c_1 \leq \rho^{-t}\hat{\rho}_t \leq c_2$. Each step requires to take the convex hull of two polytopes having at most $q(\varepsilon)$ vertices and requires the approximation of one polytope with $2q(\varepsilon)$ vertices by a polytope with $q(\varepsilon)$ vertices with accuracy ε. Both operations are known to be polynomial w.r.t. $\frac{1}{\varepsilon}$ [94] (the dimension n is fixed). The computational complexity of this algorithm is $C \cdot \varepsilon^{-\frac{n+1}{2}}$, where C is some constant.

In addition, suppose that by numerical observations we conjecture that ρ is attained by some product $A_w = A_{i_1} \ldots A_{i_T}$, *i.e.* $\rho = \rho(A_w)^{1/T}$. If during the calculations we find a polytope P such that $\bar{\Sigma}P \subset \rho(A_w)^{1/T}P$, then it occurs that $\rho = \rho(A_w)^{1/T}$. For the polytope P we take $P = P_k = \mathrm{Conv}\{\bar{\Sigma}^j v, -\bar{\Sigma}^j v,\ j = 0, \ldots, k\}$ for some integer k, where v is the eigenvector of A_w corresponding to the largest by modulo eigenvalue (assuming that this is real and unique).

2.3.6 Lifting methods to improve the accuracy

As we will see in subsection 2.3.7, it can be useful, given a set of matrices Σ to "lift" this set into another set Σ', that is to represent it with matrices acting in a higher dimensional space, such that the joint spectral radius is raised to a certain power d :

$$\rho(\Sigma') = \rho(\Sigma)^d.$$

A first method consists in using so-called Kronecker powers of matrices. This method has been recently improved with the so-called symmetric algebras. We will focus on this last (more efficient) method, but we give hereafter definitions of the Kronecker powers, for sake of completeness, and because they give an interesting insight to the symmetric algebra method.

Definition 2.6 ***Kronecker product.*** *Let $A, B \in \mathbb{R}^{n\times n}$. The* Kronecker product *of A and B is a matrix in $\mathbb{R}^{n^2\times n^2}$ defined as*

$$(A \otimes B) \triangleq \begin{pmatrix} A_{1,1}B & \dots & A_{1,n}B \\ \vdots & \vdots & \vdots \\ A_{n,1}B & \dots & A_{n,n}B \end{pmatrix}.$$

The k-th Kronecker power of A, *denoted* $A^{\otimes k}$, *is defined inductively as*

$$A^{\otimes k} = A \otimes A^{\otimes (k-1)} \quad A^{\otimes 1} = A.$$

We now introduce *symmetric algebras,* which requires some definitions. Corresponding to an $n$uple $\alpha \in \mathbb{N}^n$, we introduce the "α monomial" of a vector $x \in \mathbb{R}^n$ as the real number:

$$x^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}.$$

The *degree* of the monomial is $d = \sum \alpha_i$. We denote by $\alpha!$ the multinomial coefficient

$$\alpha! = \frac{d!}{\alpha_1! \dots \alpha_n!}.$$

We denote by N the number of different monomials of degree d :

$$N = \binom{n+d-1}{d}.$$

Definition 2.7 ***Symmetric algebra.*** *Let* $x \in \mathbb{R}^n$. *The* d-lift *of* x, *denoted* $x^{[d]}$, *is the vector in* $\mathbb{R}^N$, *indexed by all the possible exponents* α *of degree* d

$$x_\alpha^{[d]} = \sqrt{\alpha!} x^\alpha.$$

The d-lift *of the matrix* A *is the matrix* $A^{[d]} \in \mathbb{R}^N$ *associated to the linear map*

$$A^{[d]} : \quad x^{[d]} \to (Ax)^{[d]}.$$

The matrix $A^{[d]}$ *can be obtained via the following formula [90]:*

$$A_{\alpha\beta}^{[d]} = \frac{perA(\alpha,\beta)}{\sqrt{\mu(\alpha)\mu(\beta)}},$$

where perM denotes the permanent of the matrix M, *and* $\mu(\alpha)$ *is the product of the factorials of the entries of* α.

Denoting $\Sigma^{\otimes d} = \{A^{\otimes d} : A \in \Sigma\}$ and $\Sigma^{[d]} = \{A^{[d]} : A \in \Sigma\}$, we have the following properties for the Kronecker products and the d-lifts:

Proposition 2.5 *[13, 90] Let* $\Sigma \in \mathbb{R}^{n \times n}$ *and* $d \in \mathbb{N}_0$,

$$\rho(\Sigma)^d = \rho(\Sigma^d) = \rho(\Sigma^{\otimes d}) = \rho(\Sigma^{[d]}). \tag{2.6}$$

Proof. The first inequality is well known, while the two others come from the well known properties:

$$(AB)^{\otimes d} = A^{\otimes d} B^{\otimes d}$$
$$(AB)^{[d]} = A^{[d]} B^{[d]}.$$

Together with:

$$||A^{\otimes d}|| = ||A||^d,$$
$$||A^{[d]}|| = ||A||^d,$$

that holds when $||\cdot||$ is the spectral norm (i.e. the matrix norm induced by the standard Euclidean norm).

We will see in the next subsection how the above proposition is useful to obtain sharp approximations of the joint spectral radius.

2.3.7 Lyapunov methods

Let us recall a fundamental result presented in the previous chapter:

Proposition 2.6 *[104] For any bounded set Σ such that $\rho(\Sigma) \neq 0$, the joint spectral radius can be defined as*

$$\rho(\Sigma) = \inf_{||\cdot||} \sup_{A \in \Sigma} \{||A||\}.$$

This result is very strong, as it tells that in order to compute a joint spectral radius, one is not bounded to compute long products of matrices. It is sufficient to find a good norm to obtain an arbitrary close estimate of ρ via the formula

$$\rho \leq \max_{A \in \Sigma} ||A||.$$

So an alternative way to estimate the joint spectral radius is to look over all norms (or a sufficiently wide set of norms) the one that provides the tightest bound on ρ.
A family of norms that is well understood and classically used in engineering is the family of ellipsoidal norms:

Definition 2.8 *Let P be a symmetric positive definite matrix, the quantity*

$$|\cdot| : \mathbb{R}^n \to \mathbb{R} \quad x \to |x|_P = \sqrt{x^T P x}$$

is called the ellipsoidal (vector) norm *associated to P.*
The induced matrix norm

$$A \to \max_{|x|_P = 1} |Ax|_P$$

is called the ellipsoidal (matrix) norm *associated to P.*

One can check that these quantities are indeed norms, and since the matrix norm is induced by a vector norm, it is submultiplicative. The denomination "ellipsoidal" comes from the fact that the unit ball $E = \{x \in \mathbb{R}^n : |x|_P \leq 1\}$ is an ellipsoid. Ellipsoidal norms are well understood, and easy to use in practice, thanks to the following well known result:

Proposition 2.7 *Given a symmetric positive definite matrix P, the norm $||A||_P$ of a matrix A is the smallest $\gamma \in \mathbb{R}^+$ such that the following equation has a solution:*

$$A^T P A \preceq \gamma^2 P. \tag{2.7}$$

The computation of the minimal γ in (2.7) is easy. Indeed, it can be expressed as follows:

$$\gamma^2 = \max_{x^T P x = 1} x^T A^T P A x.$$

This problem can be solved by computing the Choleski factorization of the matrix $P = LL^T$, and then by posing $y = L^T x, x = L^{-1T} y$. One gets the following expression:

$$\gamma^2 = \max_{y^T y = 1} (L^{-1T} y)^T A^T P A (L^{-1T} y).$$

This problem just amounts to compute the spectral radius of a matrix:

$$\gamma^2 = \rho(L^{-1} A^T P A L^{-1T}),$$

which can be done easily with classical methods, like the power method.

In case such a norm exists for which $||A||_P < 1$, we naturally speak of a *quadratic Lypunov function,* where the term quadratic refers to the fact that the norm $|x|_P$ is a quadratic function in the entries of x. A *Lyapunov function* for a dynamical system is a function f such that $f(x_t)$ provably tends to zero when t tends to infinity, and such that $f(x_t) \to 0$ implies $x_t \to 0$. For a dynamical system ruled by the matrix A, this norm $||\cdot||_P$ is thus a Lyapunov function, ensuring that x_t tends to zero. Now the next proposition is straightforward, and allows one to derive an upper bound on the joint spectral radius of an arbitrary set of matrices.

Proposition 2.8 *For any set of matrices Σ, if there is a solution P to the following SDP program:*

$$\begin{aligned} A_i^T P A_i &\preceq \gamma^2 P \quad \forall A_i \in \Sigma \\ P &\succ 0, \end{aligned} \tag{2.8}$$

then $\rho(\Sigma) \leq \gamma$.

Proof. The above SDP program can be rewritten as $||A||_P \leq \gamma$ for all $A \in \Sigma$.

SDP programming has become a classical topic in mathematical engineering, and we do not provide here a survey on this technique. Let us just mention that this family

of problems are well understood and that it is possible to find a solution to these in polynomial time. For more information on SDP programming, we refer the reader to classical textbooks [4, 7, 8, 22, 86]. In case such a norm exists such that $||A||_P < 1$ for all $A \in \Sigma$, we speak of a *common quadratic Lyapunov function*. It is not difficult in practice to compute the minimal γ such that the SDP program (2.8) has a solution. Indeed, even though the first line of this program is not a linear matrix inequality because of the term $\gamma^2 P$, the minimum γ can be found by bisection.
How tight is such an approximation? That is, since we know (Proposition 2.6) that there exists a norm giving arbitrarily tight upper bounds on ρ, how accurately can we approximate this norm with an ellipsoidal norm? The answer is given by the following classical result:

Theorem 2.10 *("John's ellipsoid theorem" [56])*
Let $K \in \mathbb{R}^n$ be a compact convex set with nonempty interior. Then there is an ellipsoid E with center c such that the inclusions $E \subset K \subset n(E-c)+c$ hold. If K is symmetric about the origin ($K = -K$), the constant n can be changed into $\sqrt{n}$.

We are now able to present an important result, that provides two certifications of accuracy for a joint spectral radius estimation:

Theorem 2.11 *[2, 15] For an arbitrary set of m matrices $\Sigma \subset \mathbb{R}^{n\times n}$, the best ellipsoidal norm approximation ρ^* of its joint spectral radius ρ satisfies*

$$\frac{1}{\sqrt{n}}\rho^* \leq \rho \leq \rho^*,$$

$$\frac{1}{\sqrt{m}}\rho^* \leq \rho \leq \rho^*.$$

Proof. The first part is a simple application of John's ellipsoid theorem.

For the second part, consider the set $\tilde{\Sigma}$ of linear operators acting on symmetric matrices:

$$\tilde{\Sigma} = \quad \{\tilde{A} : \quad S \to A^T S A : A \in \Sigma\}.$$

Clearly, $\rho(\tilde{\Sigma}) = \rho(\Sigma)^2$. This set of linear operators leaves the cone of semidefinite positive symmetric matrices invariant. So defining

$$B = \sum_{\tilde{A}\in\tilde{\Sigma}} \tilde{A},$$

we can apply Proposition 2.4:

$$\frac{1}{m}\rho(B) \leq \rho(\tilde{\Sigma}) \leq \rho(B).$$

Observe that the spectral radius of the linear operator B can be represented as:

$$\inf_{\gamma>0,P}\{\gamma^2 : \sum_{A_i\in\Sigma} A_i^T P A_i \preceq \gamma^2 P, P \succ 0\}.$$

And a feasible pair (P,γ) for the above expression provides a solution to the SDP program (2.8). Finally,

$$\frac{1}{\sqrt{m}}\rho^* \leq \frac{1}{\sqrt{m}}\rho(B)^{1/2} \leq \rho(\Sigma).$$

Theorem 2.11 provides an efficient method for estimating the joint spectral radius within a fixed ratio (that cannot be chosen) of its actual value. This is a good step forward, but this method seems to fail if we want to compute more accurate estimates. Actually there is a way to do this, by using the lifting procedure defined in Subsection 2.3.6.

Putting Theorem 2.11 together with Proposition 2.5, we obtain:

Theorem 2.12 *[90] Let $\Sigma = \{A_1,\ldots,A_m\} \in \mathbb{R}^{n\times n}$. For any $d \in \mathbb{N}_0$, denoting $\rho^d_{ell,d}$ the best ellipsoidal approximation of the joint spectral radius of $\Sigma^{[d]}$, obtained by applying the SDP-program (2.8) to $\Sigma^{[d]}$, we have the following convergence properties:*

$$\binom{n+d-1}{d}^{-\frac{1}{2d}} \rho_{ell,d} \leq \rho(\Sigma) \leq \rho_{ell,d}, \tag{2.9}$$

$$m^{-\frac{1}{2d}} \rho_{ell,d} \leq \rho(\Sigma) \leq \rho_{ell,d}. \tag{2.10}$$

It appears that this method can still be improved by the following recent theorem due to Parrilo and Jadbabaie [89, 90].

Theorem 2.13 *[90] Let $p(x)$ be a strictly positive homogeneous multivariate polynomial in the n variables $x_1,\ldots,x_n$, of degree $2d$, and let $\Sigma \subset \mathbb{R}^{n\times n}$ be a set of matrices. If for all $A_i \in \Sigma$,*

$$p(A_i x) \leq \gamma^{2d} p(x),$$

then $\rho(\Sigma) \leq \gamma$.

Proof. Let us fix an arbitrary norm $|\cdot|$. By compactness of the unit ball, and because $p(x)$ is strictly positive on this ball, there exist two real numbers α,β such that for all $x \in \mathbb{R}^n$, $\alpha|x|^{2d} \leq p(x) \leq \beta|x|^{2d}$. So, for an arbitrary product of length $t : A_1\ldots A_t \in \Sigma^t$ one has:

$$\alpha|A_1\ldots A_t x|^{2d} \leq p(A_1\ldots A_t x) \leq \gamma^{2dt} p(x) \leq \beta\gamma^{2dt}|x|^{2d}.$$

Finally,

$$\rho^t \leq \sup_{x\in\mathbb{R}^n, A_i\in\Sigma} |A_1\ldots A_t x|/|x| \leq (\frac{\beta}{\alpha})^{1/2d}\gamma^t.$$

Unfortunately, no algorithm is known to optimize efficiently over sets of positive polynomials in general. However, a subclass of positive polynomials is easy to handle:

sums of squares. This is due to the following theorem, that has led to a by now classical way of relaxing optimization problems on positive polynomials: the Sum Of Squares relaxation (SOS relaxation) [28, 85, 91, 106]:

Theorem 2.14 *A homogeneous multivariate polynomial $p(x)$ of degree $2d$ is a sum of squares if and only if $p(x) = x^{[d]T} Q x^{[d]}$, where $x^{[d]}$ is a vector whose entries are (possibly scaled) monomials of degree d in the variables x_i, and Q is a symmetric positive semidefinite matrix.*

Putting Theorem 2.13 and Theorem 2.14 together, one obtains an SOS relaxation providing an upper bound on the joint spectral radius:

Theorem 2.15 *[90] Let $\Sigma \subset \mathbb{R}^{n\times n}$ be a finite set of matrices, and let $\gamma > 0$. If there exist P, Q_s such that the following polynomial equality holds:*

$$x^{[d]T}(\gamma^{2d} P - A_s^{[d]T} P A_s^{[d]}) x^{[d]} = x^{[d]T} Q_s x^{[d]} \quad \forall A_s \in \Sigma \qquad (2.11)$$
$$P, Q_s \succ 0,$$

then $\rho(\Sigma) < \gamma$.
Moreover the above condition can be stated as an SDP program.

Proof. Since $P \succ 0$, the polynomial $p(x) = x^{[d]T} P x^{[d]}$ is a strictly positive sum of squares. Hence, Equation (2.11) asks for the polynomial $\gamma^{2d} p(x) - p(A_s x)$ to be a sum of squares, and this can be expressed as an SDP program, as it only consists in linear relations between entries of the matrices P and Q. Finally, since a sum of squares is a positive polynomial, the hypotheses of Theorem 2.13 are satisfied, and the proof is done.

The above theorem provides an upper bound that is at least as good as the ellipsoidal approximation of Theorem 2.8, since a solution P for the SDP program (2.8) provides a solution to the SDP program (2.11) by defining $Q_s = \gamma^2 P - A_s^T P A_s$. We have thus the following result, where we put the SOS-approximations in comparison with the convex combination technique:

Theorem 2.16 *[90] Let $\Sigma = \{A_1, \ldots, A_m\} \in \mathbb{R}^{n\times n}$. For any $d \in \mathbb{N}_0$, let us denote $\rho^d_{SOS,d}$ the best SOS approximation of the joint spectral radius of $\Sigma^{[d]}$, obtained by applying the SDP-program (2.11) to $\Sigma^{[d]}$, and*

$$\rho^d_{conv,d} = \rho\Big(\sum_{A_i \in \Sigma^d} A_i\Big),$$

We have the following convergence properties:

$$m^{-\frac{1}{2d}} \rho_{SOS,d} \leq \rho(\Sigma) \leq \rho_{SOS,d}, \qquad (2.12)$$

$$\binom{n+d-1}{d}^{-\frac{1}{2d}} \rho_{SOS,d} \leq \rho(\Sigma) \leq \rho_{SOS,d}, \tag{2.13}$$

$$m^{-\frac{1}{2d}} \rho_{conv,2d} \leq \rho(\Sigma) \leq \rho_{conv,2d}. \tag{2.14}$$

Proof. Formulas (2.12) and (2.13) are straightforward consequences of the above theorems. The proof of Formula (2.14) follows the same idea as the proof of Theorem 2.11.

The computational cost of the different approximations obtained in the above theorem is $O(mn^{6d}\log 1/\varepsilon)$, where $\varepsilon = \frac{n}{2}\log d/d$ in the estimate (2.13), and $\varepsilon = 1 - m^{-\frac{1}{2d}}$ in the estimate (2.12).

Is the bound of the SOS relaxation (2.11) better than the bound of the common quadratic Lyapunov function (2.8)? That is, is it possible that

$$\gamma^2 P - A_s^T P A_s \not\succeq 0,$$

but yet

$$x^{[d]T}(\gamma^{2d} P - A_s^{[d]T} P A_s^{[d]}) x^{[d]} = x^{[d]T} Q_s x^{[d]} \quad \forall A_s \in \Sigma,$$
$$P, Q_s \succ 0,$$

for some $Q_s \succeq 0$? Recent numerical experiments ([90, table 2]) indicate that it is indeed the case for some sets of matrices. The question whether it is possible to have better bounds than (2.12) and (2.13) on the accuracy for the SOS approximation, is still open.

Open question 2 *Does the SOS approximation of the joint spectral radius guarantee more accurate bounds than presented in Theorem 2.16?*

2.3.8 Similar results for the joint spectral subradius and the Lyapunov exponent

Compared to the interest for the joint spectral radius estimation, very few exists in the literature on the estimation of the joint spectral subradius. In Chapter 7, we propose algorithms for approximating the joint spectral subradius and the Lyapunov exponent. These algorithms appear to perform very well in practice. See Chapter 7, Theorems 7.7, 7.8, and 7.9 for more information.

Recently, a more general class of methods, which have been called *conic programming methods* has been proposed. These methods encapsulate the ones described in Chapter 7. It has also been shown that similar methods can be applied to the joint spectral radius computation. See the recent preprints [14, 99] for more information.

2.3.8.1 Conclusion and discussion

This section on approximation algorithms is not intended to be exhaustive, but tries to present the main trends in the attempts to approximate the joint spectral radius, in such a way that the reader could easily implement the most efficient algorithms known by now.
An exhaustive analysis of the existing algorithms would be much longer. For instance, it is possible to interpret the symmetric algebra lifting in (at least) two other ways: First, it can be viewed as an application of another approximation algorithm developed by Protasov [95]. This algorithm had appeared previously in the literature, but we have preferred to introduce the point of view of symmetric algebras for several reasons: it is simple to apply and is based on well known algebraic constructions, thus allowing to focus easily on computational aspects, and it does not need additional assumptions (such as irreducibility).
Secondly, it can be shown (see [90]) that the symmetric algebra lifting is simply a symmetry-reduced version of the Kronecker liftings presented in [13], and that is why we decided not to expose this Kronecker method here.

2.4 The finiteness property

As we have seen, the three members inequality (1.6) provides a straightforward way to approximate the joint spectral radius to any desired accuracy: evaluate the upper and lower bounds on ρ for products of increasing length t, until ρ is squeezed in a sufficiently small interval and the desired accuracy is reached. Unfortunately, this method, and in fact any other general method for computing or approximating the joint spectral radius, is bound to be inefficient. Indeed, we know that, unless $P = NP$, there is no algorithm that even approximates with a priori guaranteed accuracy the joint spectral radius of a set of matrices in a time that is polynomial in the size of the matrices and the accuracy. And this is true even if the matrices have binary entries.
For some sets Σ, the right hand side inequality in the three members inequality is strict for all t. This is the case for example for the set consisting of just one matrix

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

Thus, there is no hope to reach the exact value of the joint spectral radius by simply evaluating the right hand side in the three members inequality. On the other hand, since $\rho(A^k) = \rho(A)^k$ the left hand side always provides the exact value when the set Σ consists of only one matrix and one can thus hope to reach the exact value of the joint spectral radius by evaluating the maximal spectral radii of products of increasing length. If for some t and $A \in \Sigma^t$ we have $\rho(A)^{1/t} = \rho(\Sigma)$, then the value of the joint

spectral radius is reached. Sets of matrices for which such a product exists are said to have the finiteness property.

Definition 2.9 *A set Σ of matrices is said to have the* finiteness property *if there exists some product $A = A_1 \dots A_t$ with $A_i \in \Sigma$ for which $\rho(\Sigma) = \rho(A)^{1/t}$.*

One of the interests of the finiteness property arises from its connection with the stability question for a set of matrices which is a problem of practical interest in a number of contexts. Recall from Definition 1.3 that a set of matrices Σ is *stable* if all long products of matrices taken from Σ converge to zero. As mentioned in Section 2.2, there are no known algorithms for deciding stability of a set of matrices and it is unknown if this problem is algorithmically decidable. We have also seen that stability of the set Σ is equivalent to the condition $\rho(\Sigma) < 1$ and we may therefore hope to decide stability as follows: for increasing values of t evaluate $\rho_t = \max\{\rho(A)^{1/t} : A \in \Sigma^t\}$ and $\hat{\rho}_t = \max\{||A||^{1/t} : A \in \Sigma^t\}$. Since we know that $\rho_t \leq \rho \leq \hat{\rho}_t$, as soon as a t is reached for which $\hat{\rho}_t < 1$ we stop and declare the set stable, and if a t is reached for which $\rho_t \geq 1$ we stop and declare the set unstable. This procedure will always stop unless $\rho = 1$ and $\rho_t < 1$ for all t. But this last situation never occurs for sets of matrices that satisfy the finiteness property and so we conclude:

Proposition 2.9 *Stability is algorithmically decidable for sets of matrices that have the finiteness property.*

It was first conjectured in 1995 by Lagarias and Wang that all sets of real matrices have the finiteness property [71]. This conjecture, known as the *finiteness conjecture*, has attracted intense attention and several counterproofs have been provided in recent years [16, 21, 69]. So far all proofs provided are nonconstructive, and all sets of matrices whose joint spectral radius is known exactly satisfy the finiteness property. In fact, all counterproofs describe sets of matrices in which there are counterexamples, but no such counterexamples have been exhibited yet.
The finiteness property is also known to hold in a number of particular cases including the case where the matrices are symmetric, or if the Lie algebra associated with the set of matrices is solvable, since in this case the joint spectral radius is simply equal to the maximum of the spectral radii of the matrices (Corollary 2.5; see Subsection 2.3.2 or [52, 74] for more information). The finiteness property also holds if the set of matrices admits a complex polytope extremal norm [48].

The definition of the finiteness property leads to a number of natural questions: When does the finiteness property holds? Is it decidable to determine if a given set of matrices satisfies the finiteness property? Do matrices with rational entries satisfy the finiteness property? Do matrices with binary entries satisfy the finiteness property? Some of these questions are studied in Chapter 4.

2.5 Conclusion

This closes the survey on the joint spectral radius. We have seen that, even though results in Section 2.1 were encouraging since they ensure the existence of an extremal norm (at least on a set of commonly irreducible submatrices), the joint spectral radius is hard to compute or approximate in theory. We have however presented algorithms that enable one to compute the joint spectral radius to arbitrary accuracy, but at the cost of an exponential time of computation. We have finally mentioned the finiteness property, that will be the central subject of a subsequent chapter.

Chapter 3
Nonnegative integer matrices

Abstract In this chapter, for a given finite set Σ of matrices with nonnegative integer entries we study the growth of

$$\hat{\rho}_t^t = \max\{\|A_1 \dots A_t\| : A_i \in \Sigma\}.$$

We show how to determine in polynomial time whether the growth with t is bounded, polynomial, or exponential, and we characterize precisely all possible behaviors.

3.1 Introduction

In this chapter[1], we focus on the case of *nonnegative integer* matrices and consider questions related to the growth of $\hat{\rho}_t^t$ with t. When the matrices have nonnegative integer entries, we will see that the following cases can possibly occur:

1. $\rho(\Sigma) = 0$. Then $\hat{\rho}_t^t(\Sigma)$ takes the value 0 for all values of t larger than some t_0 and so all products of length at least t_0 are equal to zero.
2. $\rho(\Sigma) = 1$ and the products of matrices in Σ are bounded, that is, there is a constant K such that $\|A_1 \dots A_t\| < K$ for all $A_i \in \Sigma$.
3. $\rho(\Sigma) = 1$ and the products of matrices in Σ are unbounded. We will show that in this case the growth of $\hat{\rho}_t^t(\Sigma)$ is polynomial.
4. $\rho(\Sigma) > 1$. In this case the growth of $\hat{\rho}_t^t(\Sigma)$ is exponential.

In the sequel we will mostly use the norm given by the sum of the magnitudes of all matrix entries. Of course, for nonnegative matrices this norm is simply given by the sum of all entries. Note that the situation $0 < \rho(\Sigma) < 1$ is not possible because the norm of a nonzero integer matrix is always larger than one. The four cases already

[1] The chapter presents research work that has been published in [62, 65].

occur when there is only one matrix in the set Σ. Particular examples for each of these four cases are given by the matrices:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The problem of distinguishing between the different cases has a long history. The polynomial-time decidability of the equality $\rho(\Sigma) = 0$ is shown in [53]. As mentioned by Blondel and Canterini [9], the decidability of the boundedness of products of nonnegative integer matrices follows from results proved in the 1970′s. Indeed, the finiteness of a semigroup generated by a finite set of matrices has been proved to be decidable independently by Jacob [55] and by Mandel and Simon [78]. It is clear that for integer matrices, finiteness of the semigroup is equivalent to its boundedness, and so boundedness is decidable for integer matrices. The decision algorithms proposed in [55] and [78] are based on the fact that if the semigroup is finite, then every matrix in the semigroup can be expressed as a product of length at most B of the generators, and the bound B only depends on the dimension of the matrices n and on the number of generators. The proposed algorithms consist in generating all products of length less than B; and checking whether new matrices are obtained by considering products of length $B+1$. The high value of the bound B does however lead to highly nonpolynomial algorithms and is therefore not practical. A sufficient condition for the unboundedness of $\hat{\rho}_t^t(\Sigma)$ has been derived recently for the case of binary matrices by Crespi et al. [31]. We will show in this chapter that the condition given there is also necessary. Moreover, we provide a polynomial algorithm that checks this condition, and thus we prove that boundedness of semigroups of integer matrices is decidable in polynomial time. Crespi et al. [31] also provide a criterion to verify the inequality $\rho(\Sigma) > 1$ for binary matrices and an algorithm based on that criterion. However, their algorithm is not polynomial. In this chapter, we present a polynomial algorithm for checking $\rho(\Sigma) > 1$ for sets of nonnegative integer matrices. Observe that it is not in contradiction with NP-hardness results of Chapter 2 since our algorithm allows only to check if ρ is larger than the particular value one. Let us recall that the same problem for other joint spectral characteristics (such as the joint spectral subradius) is proved to be NP-hard even for binary matrices. Therefore, the polynomial solvability of this question for the joint spectral radius is somewhat surprising.

The main results of this chapter can be summarized as follows. For any finite set of nonnegative integer $n \times n$ matrices Σ there is a polynomial algorithm that decides between the four cases $\rho = 0$, $\rho = 1$ and bounded growth, $\rho = 1$ and polynomial growth, $\rho > 1$ (see Theorem 3.1 and Theorem 3.2). Moreover, if $\rho(\Sigma) = 1$, then there exist constants C_1, C_2, k, such that $C_1 t^k \leq \hat{\rho}_t^t(\Sigma) \leq C_2 t^k$ for all t; the rate of growth k is an integer such that $0 \leq k \leq n-1$, and there is a polynomial time algorithm for computing k (see Theorem 3.3). This sharpens previously known results on the asymptotic behavior of the value $\hat{\rho}_t^t(\Sigma)$ for nonnegative integer matrices. We discuss this aspect in Section 3.6. Thus, for nonnegative integer matrices, the only case for which we cannot

decide the exact value of the joint spectral radius is $\rho > 1$. Once more, it is most likely that the joint spectral radius cannot be polynomially approximated in this case since it was proved that its computation is NP-hard, even for binary matrices.

The chapter is organized as follows. Section 3.2 contains some notation and auxiliary facts from graph theory. In Section 3.3 we establish a criterion for separating the three main cases $\rho(\Sigma) < 1, \rho(\Sigma) = 1$ and $\rho(\Sigma) > 1$. Applying this criterion we derive a polynomial algorithm that decides each of these cases. In Section 3.4 we present a criterion for deciding product boundedness and provide a polynomial time implementation of this criterion. In Section 3.5 we find the asymptotic behavior of the value $\hat{\rho}_t^t(\Sigma)$ as $t \to \infty$ for the case $\rho = 1$. We prove that this value is asymptotically equivalent to t^k for a certain integer k with $0 \leq k \leq n-1$ and show how to find the rate of growth k in polynomial time. Finally, in Section 3.6 we formulate several open problems on possible generalizations of those results to arbitrary matrices.

3.2 Auxiliary facts and notations

For two nonnegative functions $f(t), g(t)$ we use the standard notation $f(t) = O(g(t))$, which means that there is a positive constant C such that $f(t) \leq Cg(t)$ for all t. The functions f and g are said to be asymptotically equivalent, which we denote $f(t) \asymp g(t)$ if $f(t) = O(g(t))$ and $g(t) = O(f(t))$.

We shall consider each nonnegative $n \times n$ matrix A as the adjacency matrix of a directed weighted graph $G(A)$. This graph has n nodes enumerated from 1 to n. There is an edge from node i to node j if the (i, j) entry of the matrix is positive and the *weight* of this edge is then equal to the corresponding entry. This graph may have loops, i.e., edges from a node to itself, which correspond to diagonal entries. If we are given a family Σ of nonnegative integer matrices, then we have several weighted graphs on the same set of nodes $\{1, \ldots, n\}$. In addition we define the graph $G(\Sigma)$ associated to our family Σ as follows: There exists an edge in $G(\Sigma)$ from node i to node j if and only if there is a matrix $A \in \Sigma$ such that $A_{i,j} > 0$. The weight of this edge is equal to $\max_{A \in \Sigma} A_{i,j}$. We shall also use the graph G^2, whose n^2 nodes represent the ordered pairs of our initial n nodes, and whose edges are defined as follows: there is an edge from a node (i, i') to (j, j') if and only if there is a matrix $A \in \Sigma$ such that both $A_{i,j}$ and $A_{i',j'}$ are positive *for the same matrix*. The edges of G^2 are not weighted.

Products of matrices from Σ can be represented by *cascade graphs*. We now present this tool that will enable us to clarify many reasonings in subsequent proofs. In a cascade graph, a matrix $A \in \Sigma$ is represented by a bipartite graph with a left and a right set of nodes. The sets have identical size and there is an edge between the ith left node and the jth right node if $A_{i,j} > 0$. The weight of this edge is equal to the entry $A_{i,j}$. For instance, the non-weighted bipartite graph on Figure 3.1 represents the matrix

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

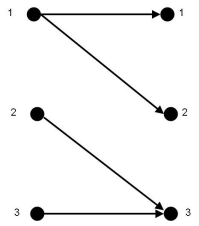


Fig. 3.1 A bipartite graph representing a binary matrix

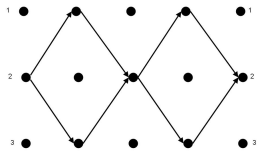


Fig. 3.2 A typical cascade graph

Now, for a given product of matrices $A_{d_1} \dots A_{d_t}$ we construct a cascade graph as follows: we concatenate the corresponding bipartite graphs in the order in which they appear in the product, with the right side of each bipartite graph directly connected to the left side of the following graph. For example, Figure 3.2 shows a cascade graph representing the product $A_0A_1A_0A_1$ of length four, with

$$A_0 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

We say that the bipartite graph at the extreme left side begins at level $t = 0$ and the one at the extreme right side ends at the last level. We note (i,t) to refer to the node i at level t. We say that there is a path from node i to node j if one is able to construct a cascade graph with a path from some node (i,t) to some node (j,t') for some $t < t'$ (that is, if there is a product with the $(i,j)-$entry larger than zero). A path is to be understood as a succession of edges from a level to the next level, i.e. always from left to right. One can check that the (i,j) entry of a matrix product of length t is equal to

the number of directed paths from the node $(i,0)$ to the node (j,t) in the corresponding cascade graph. We thus have a way of representing $\hat{\rho}_t^t(\Sigma)$ as the maximal total number of paths from extreme left nodes to extreme right nodes in cascade graphs of length t.

Two nodes of a graph are called connected if they are connected by a path (not necessarily by an edge). A directed graph is strongly connected if for any pair of nodes (i,j), i is connected to j. The following well known result states that we can partition the set of nodes of a directed graph in a unique way in strongly connected components, and that the links between those components form a tree [110].

Lemma 3.1 *For any directed weakly connected graph G there is a partition of its nodes in nonempty disjoint sets $V_1,\ldots,V_I$ that are strongly connected and such that no two nodes belonging to different partitions are connected by directed paths in both directions. Such a maximal decomposition is unique up to renumbering of the nodes. Moreover there exists a (non necessarily unique) ordering of the subsets V_s such that any node $i \in V_k$ cannot be connected to any node $j \in V_l$, whenever $k > l$. There is an algorithm to obtain this partition in $O(n)$ operations (with n the number of nodes).*

In this lemma, we suppose by convention that a node that is not strongly connected to any other node is itself a strongly connected subset, even if it does not have a self-loop. In such a case we will say that the corresponding set is a *trivial strongly connected subset*. Consider the graph $G(\Sigma)$ corresponding to a family of matrices Σ, as defined above. After possible renumbering, it can be assumed that the set of nodes is ordered, that is, for all nodes $i \in V_k$ and $j \in V_l$, if $k > l$ then $i > j$. In that case all the matrices of Σ have block upper-triangular form with I blocks corresponding to the sets $V_1,\ldots,V_I$ (I can be equal to one).

3.3 Deciding $\rho < 1$, $\rho = 1$, and $\rho > 1$

The goal of this section is to prove the following result.

Theorem 3.1 *For matrices with nonnegative integer entries there is a polynomial algorithm that decides the cases $\rho < 1, \rho = 1$ and $\rho > 1$.*

Proof. The proof will be split into several lemmas. The inequality $\rho < 1$ means that $\hat{\rho}_t^t$, the maximum number of paths in a cascade graph of length t tends to zero as $t \to \infty$. Now $\hat{\rho}_t^t$ is integer-valued, and this implies that for sufficiently large t there are no paths of this length in the graph $G(\Sigma)$ corresponding to the whole family Σ. This means that $G(\Sigma)$ has no cycle. So we get our first lemma :

Lemma 3.2 *For a finite set of nonnegative integer matrices Σ, we have $\rho(\Sigma) > 0$ if and only if the graph $G(\Sigma)$ has a cycle. In this case $\rho \geq 1$.*

This condition can be checked in $O(n)$ operations : one just has to find the strongly connected components of the graph $G(\Sigma)$ (a task that can be performed in $O(n)$ operations [110]); a cycle will be possible iff one of the subsets is nontrivial.

The problem of deciding between $\rho = 1$ and $\rho > 1$ is more difficult. Let us start with the following lemma.

Lemma 3.3 *Let Σ be an arbitrary finite set of real matrices. If $\rho(\Sigma) > 1$, then there is a product $A \in \Sigma^*$, for which $A_{i,i} > 1$ for some i. If the matrices are nonnegative, then the converse is also true.*

Proof. **Necessity**. Since $\rho(\Sigma) > 1$ it follows that there is a product $B \in \Sigma^*$ such that $\rho(B) > 1$. Let λ_1 be one eigenvalue of B of largest magnitude, so $|\lambda_1| = \rho(B) > 1$ and let $\lambda_2, \ldots, \lambda_n$ be the other eigenvalues. Since these eigenvalues rotate at different speeds when one takes the successive powers of B, there must be large values of t for which $\arg(\lambda) \approx 0$ for all eigenvalues λ of B^t, where $\arg(z)$ is the argument of the complex number z. More precisely, there exists a sufficiently large t such that $|\lambda_1|^t > 2n$ and $\arg(\lambda_k^t) \in [-\frac{\pi}{3}, \frac{\pi}{3}]$ for all $k = 1, \ldots, n$ (for a rigorous proof, see [114]). Therefore $\mathrm{Re}(\lambda_k^t) \geq \frac{1}{2}|\lambda_k^t|$ for all k. We have $\sum_{k=1}^{n} (B^t)_{k,k} = \mathrm{tr}\, B^t = \sum_{k=1}^{n} \lambda_k^t = \sum_{k=1}^{n} \mathrm{Re}\, \lambda_k^t \geq \frac{1}{2}|\lambda_1^t| > n$. Since the sum of the n numbers $(B^t)_{k,k}$ exceeds n, one of them must exceed 1.
Sufficiency. Since $A \in \Sigma^t$ has nonnegative elements, it follows that $\|A^k\| \geq (A^k)_{i,i}$, hence $\rho(A) \geq A_{i,i} > 1$. Now, by the three members inequality (1.6) $\rho(\Sigma) \geq [\rho(A)]^{1/t} > 1$.

Corollary 3.1 *For any finite set of nonnegative integer matrices Σ, we have $\rho(\Sigma) > 1$ if and only if there is a product $A \in \Sigma^*$ such that $A_{i,i} \geq 2$ for some i.*

Thus, the problem is reduced to testing if there is a product $A \in \Sigma^*$ that has a diagonal element larger or equal to 2. This is equivalent to the requirement that at least one of the following conditions is satisfied:

1. There is a cycle in the graph $G(\Sigma)$ containing at least one edge of weight greater than 2.
2. There is a cycle in the graph G^2 containing at least one node (i,i) (with equal entries) and at least one node (p,q) with $p \neq q$.

Indeed, if $A_{i,i} \geq 2$ for some $A \in \Sigma^*$, then either there is a path on the graph $G(\Sigma)$ from i to i that goes through an edge of weight ≥ 2 (first condition), or there are two different paths from i to i in the cascade graph corresponding to the product A, this is equivalent to the second condition. The converse is obvious. To verify Condition 1 one needs to look over all edges of $G(\Sigma)$ of weight ≥ 2 and to check the existence of a cycle containing this edge. This requires at most $O(n^3)$ operations. To verify Condition 2 one needs to look over all $\frac{1}{2}n^2(n-1)$ triples (i,p,q) with $p > q$ and for each of them check the existence in the graph G^2 of paths from (i,i) to (p,q) and from (p,q) to (i,i), which requires at most $O(n^2)$ operations. Thus, to test Condition 2 one needs to perform at most $O(n^5)$ operations. This completes the proof of Theorem 3.1.

Figure 3.2 shows a cascade graph with the second condition above satisfied: there are two paths from node 2 to node 2, and for every even t, the number of paths is multiplied by two.

The shortest cycle in the graph G^2 with the required properties has at most n^2 edges. It therefore follows that whenever $\rho > 1$, there is a product A of length less than n^2 such that $A_{i,i} \geq 2$ for some i. From this we deduce the following corollary.

Corollary 3.2 *Let Σ be a finite set of nonnegative integer matrices of dimension n. If $\rho(\Sigma) > 1$, then $\rho(\Sigma) \geq 2^{1/n^2}$.*

3.4 Deciding product boundedness

If $\rho = 1$, two different cases are possible: either the maximum norm of products of length t is bounded by a constant, or it grows less than exponentially with t. Deciding between these two cases is not trivial. Indeed, we have seen in Chapter 2 that this problem is undecidable for general matrices. In this section we present a simple criterion that allows us to decide whether the products are bounded, in the particular case of nonnegative integer matrices. Our reasoning will be split into several lemmas. We begin with a simple but crucial observation.

Lemma 3.4 *Let Σ be a finite set of nonnegative integer matrices with $\rho(\Sigma) = 1$. If there is a product $A \in \Sigma^*$ that has an entry larger than 1, then the graph $G(\Sigma)$ is not strongly connected.*

Proof. Let $A_{i,j} \geq 2$, that is, counting with weights, there are two paths from i to j in the same cascade graph. If there is another cascade graph with a path from j to i, then, concatenating the two cascade graphs, we can find two different paths from i to itself, and by corollary 3.1 $\rho(\Sigma) > 1$, which is a contradiction. Hence $G(\Sigma)$ is not strongly connected.

Consider the partition of the nodes of $G(\Sigma)$ into strongly connected sets $V_1, \ldots, V_I$ (see Lemma 3.1). Applying Lemma 3.4 we get the following corollaries.

Corollary 3.3 *Let Σ be a finite set of nonnegative integer matrices. If $\rho(\Sigma) = 1$, but the products of these matrices are not uniformly bounded, then there exists a permutation matrix P such that for all matrices A in Σ, $P^T AP$ is block upper triangular with at least two blocks.*

Proof. A graph is strongly connected if and only if no permutation puts the adjacency matrix in block triangular form.

Corollary 3.4 *Let Σ be a finite set of nonnegative integer matrices with joint spectral radius one. Then all products of these matrices restricted to any strongly connected set V_k are binary matrices.*

We are now able to prove the main result of this section. We first provide a result for the case of one matrix and then consider the case of several matrices.

Proposition 3.1 *Let A be a nonnegative integer matrix with $\rho(A) = 1$. The set $\{\|A^t\| : t \geq 1\}$ is unbounded if and only if there exists some $k \geq 1$, and a pair of indices (i, j) such that*

$$A^k_{i,i}, A^k_{i,j}, A^k_{j,j} \geq 1. \tag{3.1}$$

Proof. Sufficiency is easy: One can check that $(A^{kt})_{i,j} \geq t$ for any t, and hence $\hat{\rho}^t_i(\Sigma)$ is unbounded. Let us prove the necessity : Consider the partition in strongly connected subsets $V_1, \ldots, V_I$. By Corollary 3.3 we have $I \geq 2$.
We claim that there are two nontrivial sets V_a and V_b, $a < b$ that are connected by a path (there is a path from an element of V_a to an element of V_b). In order to prove this, we show that if any path in $G(\Sigma)$ intersects at most one nontrivial set, then their number must be bounded.
Let a path start from a set V_{a_1}, then go to V_{a_2} etc., until it terminates in V_{a_l}. We associate the sequence of indices $a_1 < \cdots < a_l$, $l \leq I$ to this path. As supposed, this sequence contains at most one nontrivial set, say V_{a_s}. There are at most K^l paths, counting with weights, corresponding to this sequence, where K is the largest number of edges between two given sets (still counting with weights). Indeed, each path of length $t > l$ begins with the only edge connecting V_{a_1} to V_{a_2} (since V_{a_1} is trivial), *etc.* until it arrives in V_{a_s} after $s-1$ steps (for each of the previous steps we had at most K variants), and the reasoning is the same if one begins by the end of the path, while, given a starting node in V_{a_s}, and a last node in the same set, there is at most one path between these two nodes, by Corollary 3.4. Since there are finitely many sequences $\{a_j\}_{j=1}^l$, $l \leq I$, we see that the total number of paths of length t is bounded by a constant independent of t, which contradicts the assumption.

Hence there are two nontrivial sets V_a and V_b, $a < b$ connected by a path. Let this path go from a node $i_1 \in V_a$ to $j_1 \in V_b$ and have length l. Since both graphs V_a and V_b are strongly connected, it follows that there is a cycle $i_1 \to \ldots \to i_p \to i_1$ in V_a and a path $j_1 \to \ldots \to j_q \to j_1$ in V_b, $p, q \geq 1$. Take now a number $s \in \{1, \ldots, p\}$ such that $l + s$ is divisible by p : $l + s = vp$, $v \in \mathbb{N}$. Take a nonnegative integer x such that $v + x$ is divisible by q : $v + x = uq$, $u \in \mathbb{N}$. Let us show that the matrix A^{upq} and the indices $i = i_{p-s+1}, j = j_1$ possess property 3.2. Indeed, a path of length upq along the first cycle, beginning at node i_{p-s+1} terminates in the same node, hence $A^{upq}_{i_{p-s+1}, i_{p-s+1}} \geq 1$. Similarly $(A^{upq})_{j_1, j_1} \geq 1$. On the other hand, the path going from $i_{p-s+1} \to \ldots \to i_1$, then going x times around the first cycle from i_1 to itself, and then going from i_1 to j_1, has a total length $s + xp + l = vp + xp = upq$, therefore $A^{upq}_{i_{p-s+1}, j_1} \geq 1$.

The fact that there must be two nontrivial sets connected by a path had already been proved by Mandel and Simon [78, Lemma 2.6]. We now provide a generalization of this result to the case of several matrices.

Proposition 3.2 *Let Σ be a finite set of integer nonnegative matrices with $\rho(\Sigma) = 1$. The set of norms $\{\|A\| : A \in \Sigma^*\}$ is unbounded if and only if there exists a product $A \in \Sigma^*$, and indices i and j ($i \neq j$) such that*

$$A_{i,i}, A_{i,j}, A_{j,j} \geq 1. \tag{3.2}$$

Proof. The sufficiency is obvious by the previous proposition. Let us prove the necessity. We have a set Σ of nonnegative integer matrices, and their products in Σ^* are unbounded. Consider again the partition of the nodes in strongly connected sets $V_1, \ldots, V_I$ for Σ. Our proof proceeds by induction on I. For $I = 1$ the products are bounded by corollary 3.4, and there is nothing to prove. Let $I \geq 2$ and the theorem holds for any smaller number of sets in the partition. If the value $\hat{\rho}_t^t(\Sigma, U)$ is unbounded on the set $U = \cup_{s=2}^{I} V_s$, then the theorem follows by induction. Suppose then that the products are bounded on this subset of nodes, by some constant M. Let us consider a product of t matrices, and count the paths from any leftmost node to any rightmost node. First, there are less than n^2 paths beginning in V_1 and ending in V_1, since the corresponding adjacency matrix must have $\{0,1\}$ entries (recall that n is the total number of nodes). Second, there are at most Mn^2 paths beginning and ending in U, since each entry is bounded by M. Let us count the paths beginning in V_1 and ending in U : Let $i_0 \to \cdots \to i_t$ be one of these paths. The nodes $i_0, \ldots i_{r-1}$ are in V_1, the nodes $i_r, \ldots, i_t$ are in U and $i_{r-1} i_r$ is an edge connecting V_1 and U. The number r will be called a switching level. For any switching level there are at most KMn^2 different paths connecting V_1 with U, where K is the maximum number of edges jumping from V_1 to U at the same level, counting with weights. Indeed for one switching edge $i_{r-1} i_r$, the total number of paths from i_r to any node at the last level is bounded by M, and there are less than n nodes in U. By the same way of thinking, there is maximum one path from each node in V_1 to i_{r-1}, and there are less than n nodes in V_1. The number of switching levels is thus not bounded, because so would be the number of paths. To a given switching level r we associate a triple (A', A'', d), where $A' = A_{d_1} \ldots A_{d_{r-1}|V_1}$ and $A'' = A_{d_{r+1}} \ldots A_{d_t|U}$ are matrices and $d = d_r$ is the index of the rth matrix. The notation $A_{|V_1}$ means the square submatrix of A corresponding to the nodes in V_1. Since A' is a binary matrix (Corollary 3.4), A'' is an integer matrix with entries less than M, and d can take finitely many values, it follows that there exist finitely many, say N, different triples (A', A'', d). Taking t large enough, it can be assumed that the number of switching levels $r \in \{2, \ldots, t-1\}$ exceeds N, since for any switching level there are at most KMn^2 different paths. Thus, there are two switching levels r and $r+s$, $s \geq 1$ with the same triple. Define $d = d_r = d_{r+s}$ and

$$B = A_1 \ldots A_{d_{r-1}}, \quad D = A_{d_{r+1}} \ldots A_{d_{r+s-1}}, \quad E = A_{d_{r+s+1}} \ldots A_{d_t} \tag{3.3}$$

(if $s = 1$, then D is the identity matrix). Thus, $A_{d_1} \ldots A_{d_t} = BA_dDA_dE$. Since $A' = B_{|V_1} = BA_dD_{|V_1}$ it follows that $B_{|V_1} = B(A_dD)^k_{|V_1}$ for any k. Similarly $A'' = E_{|U} = DA_dE_{|U}$ implies that $E_{|U} = (DA_d)^k E_{|U}$. Therefore for any k the cascade graph corre-

sponding to the product $B(A_dD)^kA_dE$ has at least $k+1$ paths of length $t+(k-1)s$ starting at i_0. Those paths have switching levels $r, r+s, \ldots, r+(k-1)s$ respectively. Indeed, for any $l \in \{0,\ldots,k\}$ there is a path from i_0 to $i_{r-1+ls} = i_{r-1}$, because $B(A_dD)^l_{|V_1} = B_{|V_1}$; there is an edge from i_{r-1+ls} to $i_{r+ls} = i_r$, because $A_{d_{r+ls}} = A_{d_r} = A_d$; finally there is a path from $i_{r+ls} = i_r$ to $i_{t+(k-1)s} = i_t$, because $(DA_d)^{k-l}E_{|U} = E_{|U}$. Therefore, $\|B(A_dD)^kA_dE\| \geq k+1$ for any k, hence $\|B(A_dD)^kA_dE\| \to \infty$ as $k \to \infty$, and so $\|(A_dD)^k\| \to \infty$. Now we apply Proposition 3.1 for the matrix A_dD; since the powers of this matrix are unbounded it follows that some power $A = (A_dD)^k$, which is $(A_{d_r} \ldots A_{d_{r+s-1}})^k$ possesses the property $A_{i,i}, A_{j,j}, A_{i,j} \geq 1$ for suitable i and j.

In the last proof, we find a matrix $A_dD \in \Sigma^*$ such that $||(A_dD)^k|| \to \infty$. There is a different way to prove the existence of such a matrix that is based on the generic theorem of McNaughton and Zalcstein, which states that every torsion semigroup of matrices over a field is locally finite [84]. We have given here a self-contained proof that uses the combinatorics for nonnegative integer matrices.

The meaning of the condition (3.2) in terms of cascade graphs can be seen from the following simple example. If one matrix in Σ has those three entries (and no other) equal to one, then we have two infinite and separate paths: one is a circuit passing through the node i, the other is a circuit passing through the node j. Those cycles are linked in a unique direction, so that the first one is a source and the second one is a sink, that eventually collects all these paths, as shown on Figure 3.3.

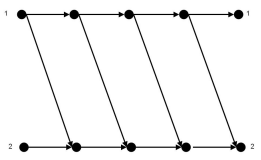


Fig. 3.3 A cascade graph with linear growth

We now prove that the criterion of Proposition 3.2 can be checked in polynomial time.

Theorem 3.2 *There is a polynomial time algorithm for verifying product boundedness of families of nonnegative integer matrices.*

Proof. Assume we are given a finite set of nonnegative integer matrices Σ. First, we decide between the cases $\rho = 0, \rho = 1$ and $\rho > 1$ with the algorithm provided in the previous section. In the first case $\hat{\rho}^l_t(\Sigma)$ is bounded, in the latter it is not. The main problem is to check boundedness for the case $\rho = 1$. By Proposition 3.2 it suffices to check if there exists a product $A \in \Sigma^*$ possessing the property of Equation (3.2) for some indices i, j. Consider the product graph G^3 with n^3 nodes defined as follows.

The nodes of G^3 are ordered triples (i,j,k), where $i,j,k \in \{1,\ldots,n\}$. There is an edge from a vertex (i,j,k) to a vertex (i',j',k') if and only if there is a matrix $A \in \Sigma$, for which $(A)_{i,i'}, (A)_{j,j'}, (A)_{k,k'} \geq 1$. (The adjacency matrix of G^3 is obtained by taking the 3-th Kronecker power of each matrix in Σ, and by taking the maximum of these matrices componentwise.) The above condition means that there are indices $i \neq j$ such that there is a path in G^3 from the node (i,i,j) to the node (i,j,j). The algorithm involves checking $n(n-1)$ pairs, and for each pair at most $O(n^3)$ operations to verify the existence of a path from (i,i,j) to (i,j,j). In total one needs to perform $O(n^5)$ operations to check boundedness.

3.5 The rate of polynomial growth

We have provided in the previous section a polynomial time algorithm for checking product boundedness of sets of nonnegative integer matrices. In this section we consider sets of matrices that are not product bounded and we analyze the *rate of growth* of the value $\hat{\rho}_t^t(\Sigma)$ when t grows. When the set consists of only one matrix A with spectral radius equal to one, the norm of A^k increases polynomially with k and the degree of the polynomial is given by the size of the largest Jordan block of eigenvalue one. A generalization of this for several matrices is given in the following theorem.

Theorem 3.3 *For any finite set of nonnegative integer matrices with joint spectral radius equal to one, there are positive constants C_1 and C_2 and an integer $k \geq 0$ (the rate of growth) such that*

$$C_1 t^k \;\leq\; \hat{\rho}_t^t(\Sigma) \;\leq\; C_2 t^k \tag{3.4}$$

for all t. The rate of growth k is the largest integer possessing the following property: there exist k different ordered pairs of indices $(i_1,j_1),\ldots,(i_k,j_k)$ such that for every pair (i_s,j_s) there is a product $A \in \Sigma^*$, for which

$$A_{i_s,i_s},\, A_{i_s,j_s},\, A_{j_s,j_s} \;\geq\; 1, \tag{3.5}$$

and for each $1 \leq s \leq k-1$, there exists $B \in \Sigma^*$ such that $B_{j_s,i_{s+1}} \geq 1$.

The idea behind this theorem is the following: if we have a polynomial growth of degree k, we must have a combination of k linear growths that combine themselves successively to create a growth of degree k. This can be illustrated by the cascade graph in Figure 3.4.

Before we give a proof of Theorem 3.3 let us observe one of its corollaries. Consider the ordered chain of maximal strongly connected subsets $V_1,\ldots,V_I$ provided by Lemma 3.1. By Corollary 3.4 the elements i_s, j_s of each pair (i_s,j_s) belong to different sets and are such that

$$i_s \in V_k,\; j_s \in V_l \Rightarrow l > k.$$

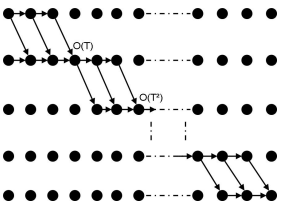


Fig. 3.4 A cascade graph with polynomial growth

This implies that there are fewer such couples than strongly connected subsets, and then:

Corollary 3.5 *The rate of growth k does not exceed $I-1$, where I is the number of strongly connected sets of the family Σ. In particular, $k \leq n-1$.*

We now provide the proof of Theorem 3.3.

Proof. We shall say that a node i is $O(t^k)$ if there is a constant $C > 0$ such that $\max_{A\in\Sigma^t, 1\leq j\leq n} A_{i,j} \leq Ct^k$ for all t. Suppose that for some k we have k pairs $(i_1, j_1), \ldots, (i_k, j_k)$ satisfying the assumption of the theorem. We construct a cascade graph similar to the one represented in Figure 3.4: Let $A_s,\ s = 1,\ldots,k$ and $B_s,\ s = 1,\ldots,k-1$ be the corresponding products and m be their maximal length. Then for any s and any $p \in \mathbb{N}$ one has $(A_s^p)_{i_s j_s} \geq p$, and therefore $\left(A_1^p B_1 A_2^p B_2 \ldots A_k^p\right)_{i_1, j_k} \geq p^k$ for any p. Denote this product by D_p and its length by l_p. Obviously $l_p \leq (pk+k-1)m$. For an arbitrary $t > (2k-1)m$ take the largest p such that $l_p < t$. It follows that $l_p \geq t - km$, and therefore $p \geq \frac{l_p}{km} - 1 + \frac{1}{k} \geq \frac{t}{km} - 2 + \frac{1}{k}$. In order to complete the product, take for instance $A_k^{t-l_p}$. Then the product $D_p A_k^{t-l_p}$ has length t and its $(i_1 j_k)$-entry is bigger than $p^k \geq \left(\frac{t}{km} - 2 + \frac{1}{k}\right)^k$, which is bigger than Ct^k for some positive constant C. This proves sufficiency.

It remains to establish the converse: if for some k there is a node that is not $O(t^{k-1})$, then there exist k required pairs of indices. We prove this by induction on the dimension n (number of nodes). For $n = 2$ and $k = 1$ it follows from Proposition 3.2. For $n = 2$ and $k > 2$ this is impossible, since one node (say, node 1) is an invariant by Corollary 3.3, then the edge $(1,2)$ is forbidden, and there are at most $t+2$ paths of length t (if all other edges occur at each level).

Suppose the theorem holds for all $n' \leq n-1$. Let a node i_0 be not $O(t^{k-1})$. Assume first that there are two nodes i, j of the graph $G(\Sigma)$ that are not connected by any path. Therefore there are no paths containing these nodes. Hence one can remove one of these nodes (with all corresponding edges) so that i_0 is still not $O(t^{k-1})$. Now by induction the theorem follows. It remains to consider the case when any pair of nodes is (weakly) connected. Take the decomposition in strongly connected subsets $V_1, \ldots, V_I$ for Σ. The nodes are ordered so that all the matrices in Σ are in block upper triangular

form. Let p be the smallest integer such that all nodes in $G_p = \cup_{s=p}^{I} V_s$ are $O(1)$, *i.e.*, G_p is the biggest invariant on which the number of paths is bounded. By Corollary 3.4 such p does exist. On the other hand, by the assumption we have $p \geq 2$. Since the products in Σ^* restricted to the subspace corresponding to $G_{p-1} = G_p \cup V_{p-1}$ are unbounded, it follows from Proposition 3.2 that there is a pair $(i_k, j_k) \in G_{p-1}$ realizing Equation (3.2). Observe that $i_k \in V_{p-1}$ and $j_k \in G_p$. Otherwise these nodes are either in V_{p-1} (hence the restriction of Σ^* to V_{p-1} is unbounded, which violates Corollary 3.4) or in G_p (contradicts the boundedness of Σ^* on G_p). Now consider the products restricted on the set $\cup_{s=1}^{p-1} V_s$. We claim that at least one node is not $O(t^{k-2})$ in this restriction: For any product in Σ^* of length t consider the corresponding cascade graph. Any path of length t starting at a node $i \in \cup_{s=1}^{p-1} V_s$ consists of 3 parts (some of them may be empty): a path $i \to v \in \cup_{s=1}^{p-1} V_s$ of some length l, an edge $v \to u \in G_p$, and a path from u inside G_p of length $t-l-1$. Suppose that each entry in the restriction of the products to $\cup_{s=1}^{p-1} V_s$ is $O(t^{k-2})$, then for a given l there are at most Cl^{k-2} paths for the first part ($C > 0$ is a constant), for each of them the number of different edges $v \to u$ (counting with edges) is bounded by a constant K, and the number of paths from u to the end is bounded by C_0 by the assumption. Taking the sum over all l we obtain at most $\sum_{l=0}^{t} CKC_0 l^{k-2} = O(t^{k-1})$ paths, which contradicts our assumption.
Hence there is a node in $\cup_{s=1}^{p-1} V_s$ that is not $O(t^{k-2})$. Applying now the inductive assumption to this set of nodes we obtain $k-1$ pairs $(i_s, j_s),\ s = 1, \ldots, k-1$ with the required properties. Note that they are different from (i_k, j_k), because $j_k \in G_p$. It remains to show that there is a path in $G(\Sigma)$ from j_{k-1} to i_k. Let us remember that $i_k \in V_{p-1}$. If $j_{k-1} \in V_{p-1}$ as well, then such a path exists, because V_{p-1} is strongly connected. Otherwise, if $j_{k-1} \in V_j$ for some $j < p-1$, then there is no path from i_k to j_{k-1}, which yields that there is a path from j_{k-1} to i_k, since each pair of nodes is weakly connected.

Remark 3.1. Let us note that the products of maximal growth constructed in the proof of Theorem 3.3 are not periodic, that is, the optimal asymptotic product is not the power of one product. Indeed, we multiply the first matrix A_0 p times, and then the second one p times, etc. This leads to a family of products of length t that are not the repetition of a period. In general, those aperiodic products can be the optimal ones, as illustrated by the following simple example.

$$\Sigma = \left\{ \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \right\}.$$

Any finite product of these matrices has spectral radius equal to one and has at most linear growth. Indeed, every $A \in \Sigma$ has rank at most two, therefore the condition of Theorem 3.3 for any $k \geq 2$ is not satisfied for the product A. Nevertheless, the aperiodic sequence of products of the type $A_0^{t/2} A_1^{t/2}$ gives a quadratic growth in t. It is interesting to compare this phenomenon with the finiteness property (see Section 2.4

and Chapter 4): for this set of matrices, the maximal behavior is a quadratic growth, which is possible only for aperiodic products.
On the other hand, considering the boundedness of the products, such phenomenon is impossible: by Proposition 3.2 if $\hat{\rho}_t^t(\Sigma)$ is unbounded, this unbounded growth can always be obtained by a periodic sequence. This fact is not true for general matrices, since the following example gives a set of complex matrices for which the products are unbounded while all periodic products are bounded:

$$\Sigma = \left\{ \begin{pmatrix} e^{i\theta 2\pi} & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} e^{i\theta 2\pi} & 0 \\ 0 & 1 \end{pmatrix} \right\}.$$

If $0 \leq \theta \leq 1$ is irrational, then the powers of any $A \in \Sigma^*$ are bounded, while $\hat{\rho}_t^t(\Sigma)$ grows linearly in t.

Proposition 3.3 *The rate of growth of a set of nonnegative integer matrices with joint spectral radius equal to one can be found in polynomial time.*

Proof. For each pair (i, j) of vertices one can check in polynomial time whether there is a product A such that $A_{i,j}, A_{i,i}, A_{j,j} \geq 1$. For each couple of those pairs $(i_1, j_1), (i_2, j_2)$, we can check in polynomial time whether there is a path from j_1 to i_2, or from j_2 to i_1. Finally we are left with a directed graph whose nodes are the couples (i, j) satisfying Equation (3.2) and with an edge between the nodes $(i_1, j_1), (i_2, j_2)$ if there is a path from j_1 to i_2. This graph is acyclic (because if there is also a path from j_2 to i_1 then there are two paths from i_1 to itself, and $\rho > 1$ by Lemma 3.3), and it is known that the problem of finding a longest path in a directed acyclic graph can be solved in linear time.

3.6 Polynomial growth for arbitrary matrices

Theorem 3.3 shows that for a finite family Σ of nonnegative integer matrices with joint spectral radius equal to one the value $\hat{\rho}_t^t(\Sigma)$ is asymptotically equivalent to t^k, where k is an integer. Moreover, we have shown that the exponent k can be computed in polynomial time. A natural question arises: do these properties hold for all sets of matrices (without the constraint of nonnegative integer entries)?

Open question 3 *Is this true that for any set of real matrices Σ with $\rho(\Sigma) = 1$ one has $\hat{\rho}_t^t(\Sigma) \asymp t^k$ for some integer k ?*

In other words, is the asymptotic behavior of the value $\hat{\rho}_t^t(\Sigma)$ really polynomial with an integer rate of growth? This property can obviously be reformulated without the restriction $\rho(\Sigma) = 1$ as follows: is it true that for any family of matrices Σ we have

$$\hat{\rho}_t^t(\Sigma) \asymp \rho^t t^k, \tag{3.6}$$

where $\rho = \rho(\Sigma)$ and k is an integer ? A more general problem arises if we remove the strict requirements of asymptotic equivalence up to a positive constant:

Open question 4 *Is this true that for any family of matrices Σ the following limit*

$$\lim_{t\to\infty} \frac{\ln \rho^{-t}\hat{\rho}_t^t(\Sigma)}{\ln t}, \tag{3.7}$$

exists and is always an integer?

In particular, does property (3.6) or, more generally, property (3.7) hold for nonnegative integer matrices ? If the answer is positive, can the rate of growth be computed? We have solved these problems only for the case $\rho = 1$. Thus, is it possible to obtain a sharper information on the asymptotic behavior of the value $\hat{\rho}_t^t(\Sigma)$ as $t \to \infty$ than the well-known relation $\lim_{t\to\infty} \ln \hat{\rho}_t^t(\Sigma)/t \;=\; \ln \rho(\Sigma)$? The question is reduced to the study of the value $r(t) = \rho^{-t}\hat{\rho}_t^t(\Sigma)$. For some special families of matrices this question has appeared in the literature many times. S. Dubuc in 1986 studied it for a special pair of 2×2 matrices in connection with the rate of convergence of some approximation algorithm [40]. In 1991 I. Daubechies and J. Lagarias [34] estimated the value $r(t)$ for special pairs of $n\times n$ matrices to get a sharp information on the continuity of wavelets and refinable functions, and their technique was developed in many later works (see [118] for references). In 1990 B. Reznik [102] formulated several open problems on the asymptotics of binary partition functions (combinatorial number theory) that were actually reduced to computing the value $r(t)$ for special binary matrices [96]. This value also appeared in other works, in the study of various problems [29, 39, 103]. For general families of matrices very little is known about the asymptotic behavior of $r(t)$, although some estimates are available. First, if the matrices in Σ do not have a nontrivial common invariant subspace, then $r(t) \asymp 1$, *i.e.*, the set is nondefective (see Chapter 2 Section 2.1). So, in this case the answer to Open Question 3 is positive with $k = 0$. This assumption was relaxed for nonnegative matrices in [96]. It was shown that if a family of nonnegative matrices is irreducible (has no common invariant subspaces among the coordinate planes), then we still have $r(t) \asymp 1$. For all other cases, if the matrices are arbitrary and may have common invariant subspaces, we have only rough estimates. For the lower bound we always have $r(t) \geq C$ by the three members inequality. For the upper bound, as it is shown in [34], we have $r(t) \leq Ct^{n-1}$. This upper bound was sharpened in the following way [29]. Let l be the maximal integer such that there is a basis in $\mathbb{R}^n$, in which all the matrices from Σ get a block upper-triangular form with l blocks. Then $r(t) \leq Ct^{l-1}$. The next improvement was obtained in [98] (see also Proposition 1.5). Let $\Sigma = \{A_1, \ldots, A_N\}$ and each matrix $A_d \in \Sigma$ are in upper triangular form, with diagonal blocks $B_{d,1}, \ldots, B_{d,l}$. Let s be the total number of indices $j \in \{1, \ldots, l\}$ such that $\rho(B_{1,j}, \ldots, B_{N,j}) = \rho(\Sigma)$. Then $r(t) \leq Ct^{s-1}$. Thus, for an arbitrary family of matrices we have $C_1 \leq \rho^{-t}\hat{\rho}_t^t(\Sigma) \leq C_2 t^{s-1}$. To the best of our knowledge this is the sharpest information about the asymptotic behavior of $r(t)$ available thus far.

3.7 Conclusion and remarks

The results presented in this chapter completely characterize finite sets of nonnegative integer matrices with bounded products and with polynomially growing products. Without any change the results can be applied to general sets of nonnegative matrices that have no entries between zero and one. Unlike the proofs, which are quite technical, the results are easily implementable in algorithms. One question we have not addressed here is that of the exact computation of the joint spectral radius when $\rho > 1$, but we know this problem is NP-hard even for binary matrices (see Chapter 2). We also provide an example of two matrices whose joint spectral radius is equal to one but for which the optimal asymptotic behavior (quadratic growth) is not periodic. All periodic products have indeed a linear growth. This example may possibly help for the analysis of the finiteness property (see Chapter 4). Finally, in the last section we leave several open problems on possible generalizations of these results for more general sets of matrices.

Chapter 4
On the finiteness property for rational matrices

Abstract In this chapter we analyze a recent conjecture stating that the finiteness property holds for pairs of binary matrices. We show that the finiteness property holds for all pairs of binary matrices if and only if it holds for all sets of nonnegative rational matrices. We provide a similar result for matrices with positive and negative entries. We finally prove the conjecture for 2×2 matrices.

4.1 Introduction

Let us recall from previous chapters the definition of the finiteness property[1]:

Definition 4.1 *A set Σ of matrices is said to have the* finiteness property *if there exists some product $A = A_1 \dots A_t$ with $A_i \in \Sigma$ for which $\rho(\Sigma) = \rho^{1/t}(A)$.*

This property is of importance in practice, because of the following proposition, proved in Section 2.4:

Proposition 4.1 *Stability is algorithmically decidable for sets of matrices that have the finiteness property.*

In that section we have shown that the finiteness property does not hold in general, but its definition leads to a number of natural questions: When does the finiteness property hold? Is it decidable to determine if a given set of matrices satisfies the finiteness property? Do matrices with rational entries satisfy the finiteness property? Do matrices with binary entries satisfy the finiteness property? These questions have a natural justification. First, we are interested in rational matrices because for engineering purposes, the matrices that one handles (or enters in a computer) are rational-valued. So, in some sense, a proof of the finiteness property for rational matrices would be satisfactory in practice. Moreover, the case of binary matrices appears to be important

[1] The chapter presents research work that has been published in [59, 60].

in a number of applications. For instance, the rate of growth of the binary partition function in combinatorial number theory is expressed in terms of the joint spectral radius of binary matrices, that is, matrices whose entries are zeros and ones [96, 102]. Moision et al. [81–83] have shown how to compute the capacity of a code under certain constraints (caused by the noise in a channel) by using the joint spectral radius of binary matrices. Recently the joint spectral radius of binary matrices has also been used to express trackability of mobiles in a sensor network [30]. These applications (some of which are presented in Part II) have led to a number of numerical computations [57, 81, 82]. The results obtained so far seem to indicate that for binary matrices the finiteness property holds. When the matrices have binary entries they can be interpreted as adjacency matrices of graphs on an identical set of nodes and in this context it seems natural to expect optimality to be obtained for periodic products. Motivated by these observations, the following conjecture appears in [11]:

Conjecture 4.1 *Pairs of binary matrices have the finiteness property.*

In the first theorem in this chapter we prove a connection between rational and binary matrices:

Theorem 4.1 *The finiteness property holds for all sets of nonnegative rational matrices if and only if it holds for all* pairs *of* binary *matrices.*

If Conjecture 4.1 is correct then nonnegative rational matrices also satisfy the finiteness property and this in turn implies that stability, that is, the question $\rho < 1$, is decidable for sets of matrices with nonnegative rational entries. From a decidability perspective this last result would be somewhat surprising since it is known that the closely related question $\rho \leq 1$ is not algorithmically decidable for such sets of matrices (see Section 2.2).

Motivated by the relation between binary and rational matrices, we prove in a subsequent theorem that sets of 2×2 binary matrices satisfy the finiteness property. We have not been able to find a unique argument for all possible pairs and we therefore proceed by enumerating a number of cases and by providing separate proofs for each of them. This somewhat unsatisfactory proof is nevertheless encouraging in that it could possibly be representative of the difficulties arising for pairs of binary matrices of arbitrary dimension. In particular, some of the techniques we use for the 2×2 case can be applied to matrices of arbitrary dimension.

4.2 Rational vs. binary matrices

In this section, we prove that the finiteness property holds for nonnegative rational matrices if and only if it holds for *pairs* of *binary* matrices. The proof proceeds in three steps. First we reduce the nonnegative rational case to the nonnegative integer case, we then reduce this case to the binary case, and finally we show how to reduce

the number of matrices to two. In the last theorem we give an analogous result for matrices with arbitrary rational entries: the finiteness property holds for matrices with rational entries if and only if it holds for matrices with entries in $\{-1,0,1\}$.

Proposition 4.2 *The finiteness property holds for finite sets of nonnegative rational matrices if and only if it holds for finite sets of nonnegative integer matrices.*

Proof. Recall that for any $\alpha > 0$, $\rho(\Sigma) = (1/\alpha)\rho(\alpha\Sigma)$. Now, for any set Σ of matrices with nonnegative rational entries, let us pick an $\alpha \neq 0 \in \mathbb{N}$ such that $\alpha\Sigma \subseteq \mathbb{N}^{n\times n}$. If there exists a positive integer t and a matrix $A \in (\alpha\Sigma)^t$ such that $\rho(\alpha\Sigma) = \rho^{1/t}(A)$, then $\rho(\Sigma) = (1/\alpha)\rho^{1/t}(A) = \rho^{1/t}(A/\alpha^t)$, where $A/\alpha^t \in \Sigma^t$.

We now turn to the reduction from the integer to the binary case. Figure 4.1 represents the reduction for a particular example.

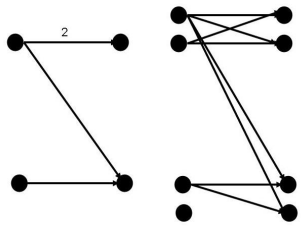


Fig. 4.1 The cascade graph (see Chapter 3) of a nonnegative integer matrix A (a) and its binary representation $\tilde{A}$ (b).

Theorem 4.2 *The finiteness property holds for finite sets of nonnegative integer matrices if and only if it holds for finite sets of binary matrices.*

Proof. Consider a finite set of nonnegative integer matrices $\Sigma \subset \mathbb{N}^{n\times n}$. We think of the matrices in Σ as adjacency matrices of weighted graphs on a set of n nodes and we construct auxiliary graphs such that paths of weight w in the original weighted graphs are replaced by w paths of weight one in the auxiliary graphs. For every matrix $A \in \Sigma \subset \mathbb{N}^{n\times n}$, we introduce a new matrix $\tilde{A} \in \{0,1\}^{nm\times nm}$ as follows (see Figure 4.1). We define m as the largest entry of the matrices in Σ ($m = 2$ in Figure 4.1). Then, for every node v_i ($i = 1,\ldots,n$) in the original graphs, we introduce m nodes $\tilde{v}_{i,1},\ldots,\tilde{v}_{i,m}$ in the auxiliary graphs. The auxiliary graphs have nm nodes; we now define their edges. For all $A \in \Sigma$ and $A_{i,j} = k \neq 0$, we define km edges in $\tilde{A}$ from nodes $\tilde{v}_{i,s} : 1 \leq s \leq k$ to the nodes $\tilde{v}_{j,t} : 1 \leq t \leq m$.
Now, we claim that for all t, and for all $A \in \Sigma^t$, the corresponding product $\tilde{A} \in \tilde{\Sigma}^t$ is such that $||A||_1 = ||\tilde{A}||_1$, where $||\cdot||_1$ represents the maximum sum of the absolute values of all entries of any column in a matrix. This implies that $\rho(\tilde{\Sigma}) = \rho(\Sigma)$, and $\tilde{\Sigma}$

has the finiteness property if and only if so does Σ. This proves the theorem, since $\tilde{\Sigma}$ has binary entries. The following reasoning leads to this claim:

1. For any product $\tilde{A} \in \tilde{\Sigma}^t$, and for any indices i, r, j, s, s', $\tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{j,s}} = \tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{j,s'}}$. This is due to the fact that for every matrix in $\tilde{\Sigma}$, the columns corresponding to $\tilde{v}_{i,s}$ and $\tilde{v}_{i,s'}$ are equal.
2. For any product $A \in \Sigma^t$, and any couple of indices (i, j), the corresponding product $\tilde{A} \in \tilde{\Sigma}^t$ has the following property: for all s, $A_{i,j} = \sum_r \tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{j,s}}$. We show this by induction on the length of the product: First, this is true by construction for every matrix in $\tilde{\Sigma}$. Now suppose that it is true for every product of length t, and consider a product of length $t+1 : AB \in \Sigma^{t+1}$ and its corresponding product $\tilde{A}\tilde{B} \in \tilde{\Sigma}^{t+1}$, with $\tilde{A} \in \tilde{\Sigma}^t$ and $\tilde{B} \in \tilde{\Sigma}$. We have the following implications:

$$\begin{aligned}
(AB)_{i,j} &= \sum_{1\le k\le n} A_{i,k}B_{k,j} \\
&= \sum_{1\le k\le n} \left(\sum_{1\le r\le m} \tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{k,s}} \right) B_{k,j} \\
&= \sum_{1\le r\le m} \left(\sum_{1\le k\le n} \tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{k,s}} B_{k,j} \right) \\
&= \sum_{1\le r\le m} \left(\sum_{1\le k\le n, 1\le s\le m} \tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{k,s}} \tilde{B}_{\tilde{v}_{k,s},\tilde{v}_{j,s'}} \right) \\
&= \sum_{1\le r\le m} \left(\tilde{A}\tilde{B}\right)_{\tilde{v}_{i,r},\tilde{v}_{j,s'}}.
\end{aligned}$$

   In the first implication we used the induction hypothesis for products of length t, in the second implication we reverse the order of summation, while for the third implication we use both the induction hypothesis for products of length 1 to transform $B_{k,j}$, and moreover we use the item 1 of this proof in order to let the index s of the matrix $\tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{k,s}}$ vary. Since s' can be chosen arbitrarily between 1 and m, the proof is done.
3. For all t, and for all $A \in \Sigma^t$, the corresponding product $\tilde{A} \in \tilde{\Sigma}^t$ is such that $||A||_1 = ||\tilde{A}||_1$, where $||\cdot||_1$ represents the maximum sum of the absolute values of all entries of any column in a matrix.
4. We have that $\rho(\Sigma) = \rho(\tilde{\Sigma})$, and if $\rho(\tilde{\Sigma}) = \rho^{1/T}(\tilde{A}) : \tilde{A} \in \tilde{\Sigma}^T$, then $\rho(\Sigma) = \rho^{1/T}(A)$, where A is the product in Σ^T corresponding to $\tilde{A}$.

We finally consider the last reduction: we are given a set of matrices and we reformulate the finiteness property for this set into the finiteness property for two particular matrices constructed from the set. The construction is such that all the entries of the two matrices have values identical to those of the original matrices, except for some entries that are equal to zero or one.

More specifically, assume that we are given m matrices $A_1,\dots,A_m$ of dimension n. From these m matrices we construct two matrices $\tilde{A}_0,\tilde{A}_1$ of dimension $(2m-1)n$. The matrices $\tilde{A}_0,\tilde{A}_1$ consist of $(2m-1)\times(2m-1)$ square blocks of dimension n that are either equal to 0, I or to one of the matrices A_i. The explicit construction of these two matrices is best illustrated with a graph.

Consider the graph G_0 on a set of $2m-1$ nodes s_i $(i=1,\dots,2m-1)$ and whose edges are given by $(i,i+1)$ for $i=1,\dots,2m-2$. We also consider a graph G_1 defined on the same set of nodes and whose edges of weight a_i are given by $(m+i-1,i)$ for $i=1,\dots,m$. These two graphs are represented on Figure 4.2 for $m=5$. In this construction a directed path that leaves the node m returns there after m steps and whenever it does so, the path passes exactly once through an edge of graph G_1. Let us now describe how to construct the matrices $\tilde{A}_0,\tilde{A}_1$. The matrices are obtained by constructing the adjacency matrices of the graphs G_0 and G_1 and by replacing the entries 1 and 0 by the matrices I and 0 of dimension n, and the weights a_i by the matrices A_i. For $m=5$ the matrices $\tilde{A}_0,\tilde{A}_1$ are thus given by:

$$\tilde{A}_0=\begin{pmatrix}
0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix},$$

$$\tilde{A}_1=\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & A_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & A_3 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & A_4 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & A_5 & 0 & 0 & 0 & 0
\end{pmatrix}.$$

The two matrices so constructed inherit some of the properties of the graphs G_0 and G_1 : loosely speaking, in a product of length m, one is forced to use $m-1$ times the matrix $\tilde{A}_0$ and one time $\tilde{A}_1$, and this product represents the use of one particular matrix in Σ. The moment one uses $\tilde{A}_1$ determines uniquely the matrix in Σ. These ideas are formalized in the following theorem.

Theorem 4.3 *Consider a set of $m\geq 1$ matrices*

$$\Sigma=\{A_1,\dots,A_m : A_i\in\mathbb{R}^{n\times n}\},$$

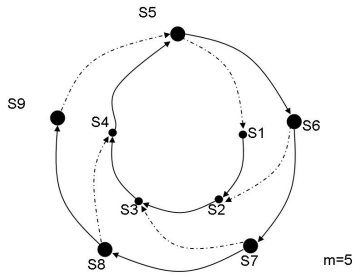


Fig. 4.2 Schematic representation of the macro transitions between subspaces. The full edges represent transitions in $\tilde{A}_0$ and the dashed edges transitions in $\tilde{A}_1$.

and $\tilde{\Sigma} = \{\tilde{A}_0, \tilde{A}_1\}$ *with the matrices* $\tilde{A}_0$ *and* $\tilde{A}_1$ *as defined above. Then* $\rho(\tilde{\Sigma}) = \rho(\Sigma)^{1/m}$. *Moreover, the finiteness property holds for* Σ *if and only if it holds for* $\tilde{\Sigma}$.

Proof. The crucial observation for the proof is the following. Consider a path in G_0 and G_1. Edges in G_0 and G_1 have outdegree at most equal to one. So if a sequence of graphs among G_0 and G_1 is given, there is only one path leaving i that follows that particular sequence. This fact ensures that any block in any product of matrices in $\tilde{\Sigma}$ is a pure product of blocks of the matrices in $\tilde{\Sigma}$, and not a sum of such products. Moreover, any path leaving from i and of length km either returns to i after passing through k edges of G_1, or ends at node $i+m$ after passing through $k-1$ edges of G_1, or ends at node $i+m \pmod{2m}$ after passing through $k+1$ edges of G_1. From this it follows that in a product of length km of the matrices $\tilde{A}_0$ and $\tilde{A}_1$ there is exactly one nonzero block in every line of blocks, and this block is a product of length $k-1$, k, or $k+1$ of matrices from Σ.

We now show that $\rho(\tilde{\Sigma}) \geq \rho(\Sigma)^{1/m}$ by proving that for any matrix $A \in \Sigma^t$, there is a matrix $\tilde{A} \in \tilde{\Sigma}^{tm}$ such that $||\tilde{A}|| \geq ||A||$. Define $\tilde{B}_i = \tilde{A}_0^{i-1}\tilde{A}_1\tilde{A}_0^{m-i} \in \tilde{\Sigma}^m$ for $i = 1, \ldots, m$ so that the block in position (m,m) in $\tilde{B}_i$ is simply equal to A_i. Consider now some product of length t, $A = A_{i_1} \cdots A_{i_t} \in \Sigma^t$ and construct the corresponding matrix product $\tilde{A} = \tilde{B}_{i_1} \ldots \tilde{B}_{i_t} \in \tilde{\Sigma}^{tm}$. The block in position (m,m) in $\tilde{A}$ is equal to $A_{i_1} \ldots A_{i_t}$ and so $||\tilde{A}|| \geq ||A||$ and $\rho(\tilde{\Sigma}) \geq \rho(\Sigma)^{1/m}$.

Let us now show that $\rho(\tilde{\Sigma}) \leq \rho(\Sigma)^{1/m}$. Consider therefore an arbitrary product $\tilde{A} \in \tilde{\Sigma}^l$ and decompose $\tilde{A} = \tilde{C}\tilde{A}'$ with $\tilde{C}$ a product of at most m factors and $\tilde{A}' \in \Sigma^{km}$. By the observation above we know that there is at most one nonzero block in every line of blocks of $\tilde{A}'$, and this block is a product of length $k-1$, k, or $k+1$ of matrices from Σ. Therefore, if the norm is chosen to be the maximum line sum, we have $||\tilde{A}|| \leq K_1 K_2 ||A||$ where A is some product of length $k-1$ of matrices from Σ, K_1 is the maximal norm of a product of at most m matrices in $\tilde{\Sigma}$, and K_2 is the maximal norm of a product of at most 2 matrices in Σ. Using this inequality, we arrive at

$$||\tilde{A}||^{1/(k-1)} \leq (K_1 K_2)^{1/(k-1)} ||A||^{1/(k-1)}.$$

The initial product $\tilde{A}$ is an arbitrary product of length $l = km + r$ and so by letting k tend to infinity and using the definition of the joint spectral radius we conclude that $\rho(\tilde{\Sigma}) \leq \rho(\Sigma)^{1/m}$.

We have thus proved that $\rho(\tilde{\Sigma}) = \rho(\Sigma)^{1/m}$. It remains to prove the equivalence of the finiteness property. If Σ satisfies the finiteness property then $\rho(\Sigma) = \rho(A_1 \dots A_t)^{1/t}$, then $\rho(\tilde{\Sigma}) = \rho(\Sigma)^{1/m} = \rho(\tilde{B}_1 \dots \tilde{B}_t)^{1/(tm)}$ and so $\tilde{\Sigma}$ also satisfies the finiteness property. In the opposite direction, if the finiteness property holds for $\tilde{\Sigma}$, then we must have $\rho(\tilde{\Sigma}) = \rho(\tilde{B}_1 \dots \tilde{B}_t)^{1/t}$ because every other product of matrices in $\tilde{\Sigma}$ has its spectral radius equal to zero, and then $\rho(\Sigma) = \rho(\tilde{\Sigma})^m = \rho(A_1 \dots A_t)^{1/t}$.

Combining the results obtained so far we now state the main result of this section.

Theorem 4.4 *The finiteness property holds for all sets of nonnegative rational matrices if and only if it holds for all pairs of binary matrices.*

The finiteness property holds for all sets of rational matrices if and only if it holds for all pairs of matrices with entries in $\{0, 1, -1\}$.

Proof. The proof for the nonnegative case is a direct consequence of Proposition 4.2, Theorem 4.2 and Theorem 4.3.
For the case of arbitrary rational entries, the statements and proofs of Proposition 4.2 and Theorem 4.3 can easily be adapted. We now show how to modify Theorem 4.2 so as to prove that the finiteness property holds for all sets of integer matrices if and only if it holds for matrices with entries in $\{0, -1, 1\}$. The value m in the proof of Theorem 4.2 is now given by the largest magnitude of the entries of the matrices in Σ, and we weight the edges of the auxiliary graphs by -1 whenever they correspond to a negative entry. Then, the arguments for proving items 1 and 2 in the above proof do not need any modification since they rely on equalities that are also valid for matrices with entries in $\{0, 1, -1\}$. From this we deduce that

$$||A|| = \sum_i |A_{i,j}| \leq \sum_{i,r} |\tilde{A}_{\tilde{v}_{i,r},\tilde{v}_{j,s}}| \leq ||\tilde{A}||,$$

and so $\rho(\tilde{\Sigma}) \geq \rho(\Sigma)$. Now, let us decompose any product $\tilde{A} = \tilde{B}\tilde{C} : \tilde{B} \in \tilde{\Sigma}, \tilde{C} \in \tilde{\Sigma}^{t-1}$, and consider the corresponding product $A = BC \in \Sigma^t$. Remark that

$$|\tilde{A}_{v_{i,r},v_{j,s}}| = \left| \sum_{k:|B_{i,k}| \geq r} sign(B_{i,k}) \sum_q \tilde{C}_{v_{k,q},v_{j,s}} \right| \leq \sum_k |C_{k,j}|.$$

So, we have $||\tilde{A}|| \leq mn||C||$.
Finally, if $\tilde{\Sigma}$ has the finiteness property, there exists $\tilde{A} \in \tilde{\Sigma}^t : \rho(\tilde{\Sigma}) = \rho(\tilde{A})^{1/t}$, and, taking the same decomposition $A = BC$ as above, we have the following relations:

$$\begin{aligned} \rho(\Sigma) \leq \rho(\tilde{\Sigma}) &= \lim_{k\to\infty} ||(\tilde{B}\tilde{C})^k||^{1/(kt)} \\ &\leq \lim_{k\to\infty} (mn||C(BC)^{k-1}||)^{1/(kt)} \leq \rho(A)^{1/t} \leq \rho(\Sigma), \end{aligned}$$

and $\rho(\Sigma) = \rho(A)^{1/t}$.

Let us finally remark that for the purpose of reducing the finiteness property of rational matrices to pairs of binary matrices, we have provided a construction that, given a set Σ of m matrices with nonnegative rational entries, produces a pair of matrices $\tilde{\Sigma}$ with binary entries and an integer $k \geq 0$ such that $\rho(\Sigma) = \rho(\tilde{\Sigma})^k$. The joint spectral radius of a set of nonnegative rational matrices can thus be captured as the power of the joint spectral radius of two binary matrices. In the same way of thinking, the joint spectral radius of a set of arbitrary rational matrices can be captured as the power of the joint spectral radius of two matrices with entries in $\{0,1,-1\}$.

4.3 The finiteness property for pairs of 2×2 binary matrices

In this section, we prove that the finiteness property holds for pairs of binary matrices of size 2×2. We will see that, even for this 2×2 case, nontrivial behaviors occur. As an illustration, the set of matrices

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\},$$

whose behavior could at first sight seem very simple, happens to have a joint spectral radius equal to $((3+\sqrt{13})/2)^{1/4}$, and this value is only reached by products of length at least four. Another interest of this section is to introduce techniques that may prove useful to establish the finiteness property for matrices of larger dimension.

There are 256 ordered pairs of binary matrices. Since we are only interested in unordered sets we can lower this number to $(2^4(2^4-1))/2 = 120$. We first present or recall a series of simple properties that allow us to handle most of these cases and we then give a complete analysis of the few remaining cases. In the following, we note $A \leq B$ if the matrix $B-A$ has nonnegative entries.

Proposition 4.3 *For any set of matrices $\Sigma = \{A_0, A_1\} \subset \mathbb{R}^{2\times 2}$, we have*

- $\rho(\{A_0, A_1\}) = \rho(\{A_0^T, A_1^T\})$, *where A^T is the transpose of A,*
- $\rho(\{A_0, A_1\}) = \rho(\{SA_0S, SA_1S\})$, *where* $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Moreover, in both cases the finiteness property holds for one of the sets if and only if it holds for the other.

Proposition 4.4 *The finiteness property holds for sets of symmetric matrices.*

Proof. See Corollary 2.3 in Chapter 2.

Proposition 4.5 *Let $\Sigma = \{A_0, A_1\} \in \mathbb{N}^{n\times n}$. The finiteness property holds in any of the following situations:*

1. $\rho(\Sigma) \leq 1$,
2. $A_0 \leq I$ *(or* $A_1 \leq I$*).*

Proof. 1. We know that for sets of nonnegative integer matrices, if $\rho \leq 1$, then either $\rho = 0$ and the finiteness property holds, or $\rho = 1$, and there is a product of matrices in Σ with a diagonal entry equal to one (see Chapter 3). Such a product $A \in \Sigma^t$ satisfies $\rho(\Sigma) = \rho(A)^{1/t} = 1$ and so the finiteness property holds when $\rho(\Sigma) \leq 1$.

2. Suppose first $\rho(A_1) \leq 1$; then $\rho(A) \leq 1$ for all $A \in \Sigma^t$ because $A_0 \leq I$ and thus $\rho(\Sigma) \leq 1$ and the result follows from item 1. Now if $\rho(A_1) > 1$ then $\rho(\Sigma) = \rho(A_1)$ and so the finiteness property holds.

Proposition 4.6 *Let* $\Sigma = \{A_0, A_1\} \in \mathbb{N}^{n \times n}$*. The finiteness property holds in the following situations:*

1. $A_0 \leq A_1$ *(or* $A_1 \leq A_0$*),*
2. $A_0 A_1 \leq A_1^2$ *(or* $A_1 A_0 \leq A_1^2$*),*
3. $A_0 A_1 \leq A_1 A_0$*.*

Proof. 1. Any product of length t is bounded by A_1^t. Hence the joint spectral radius of Σ is given by $\lim_{t\to\infty} ||A_1^t||^{1/t} = \rho(A_1)$.

2. and 3. Let $A \in \Sigma^t$ be some product of length t. If $A_0 A_1 \leq A_1^2$ or $A_0 A_1 \leq A_1 A_0$ we have $A \leq A_1^{t_1} A_0^{t_0}$ for some $t_0 + t_1 = t$. The joint spectral radius is thus given by

$$\begin{aligned} \rho = \lim_{t\to\infty} \max_{t_1+t_0=t} ||A_1^{t_1} A_0^{t_0}||^{1/t} &\leq \lim_{t\to\infty} \max_{t_1+t_0=t} ||A_1^{t_1}||^{1/t} ||A_0^{t_0}||^{1/t} \\ &\leq \max(\rho(A_0), \rho(A_1)). \end{aligned}$$

Hence the joint spectral radius is given by $\max(\rho(A_0), \rho(A_1))$.

In order to analyze all possible sets of matrices, we consider all possible couples (n_0, n_1), where n_i is the number of nonzero entries in A_i. From Proposition 4.6, we can suppose $n_i = 1, 2$, or 3 and without loss of generality we take $n_0 \leq n_1$.

- $n_0 = 1$:
  - If $n_1 = 1$ or $n_1 = 2$, the maximum row sum or the maximum column sum is equal to one for both matrices, and since these quantities are norms it follows from the three members inequality (1.6) that the joint spectral radius is less than one and from Proposition 4.5 that the finiteness property holds.
  - If $n_1 = 3$, it follows from Proposition 4.6 that the only interesting cases are:

$$\Sigma = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \right\} \text{ and } \Sigma_0 = \left\{ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\}.$$

    In the first case the matrices are symmetric and so the finiteness property holds. We keep Σ_0 for later analysis.

- $n_0 = 2$:
  - $n_1 = 2$: The only interesting cases are:

$$\Sigma = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\} \text{ and } \Sigma_1 = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \right\}.$$

    Indeed in all the other cases either the maximum row sum or the maximum column sum is equal to one and the finiteness property follows from Proposition 4.5. The joint spectral radius of the first set is equal to one. Indeed, the matrices are upper triangular. We keep Σ_1 for further analysis.
  - $n_1 = 3$: If the zero entry of A_1 is on the diagonal (say, the second diagonal entry), then, by Proposition 4.5 we only need to consider the following case:

$$\left\{ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \right\}.$$

    These matrices are such that $A_0A_1 \leq A_1^2$ and so the finiteness property follows from Proposition 4.6.
    If the zero entry of A_1 is not a diagonal entry, we have to consider the following cases:

$$\Sigma_2 = \left\{ \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\} \text{ and } \Sigma_3 = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\}.$$

    We will handle Σ_2 and Σ_3 later on.
- $n_0, n_1 = 3$: It has already been noticed by several authors (see, e.g., [111, Proposition 5.17]) that

$$\rho\left(\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\} \right) = \rho\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right)^{1/2} = \sqrt{\frac{1+\sqrt{5}}{2}}.$$

  After excluding the case of symmetric matrices and using the symmetry argument of Proposition 4.3, the only remaining case is:

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \right\},$$

  but again these matrices are such that $A_0A_1 \leq A_1^2$ and so the finiteness property follows from Proposition 4.6.

We now analyze the cases $\Sigma_0, \Sigma_1, \Sigma_2, \Sigma_3$ that we have identified above. For Σ_0, notice that $A_0^2 \leq A_0A_1$. Therefore, any product of length t is dominated (that is, is entrywise smaller) by a product of the form $A_1^{t_1}A_0A_1^{t_2}A_0 \ldots A_1^{t_l}$ for some $t_1, t_l \geq 0$ and $t_i \geq 1$ $(i = 2, \ldots, l-1)$. The norm of such a product is equal to $(t_1+1)(t_l+1)t_2 \ldots t_{l-1}$. It is not difficult to see that the maximal rate of growth of this quantity with the product

length is given by $\sqrt[5]{4}$ and so the joint spectral radius is equal to $\sqrt[5]{4} = \rho(A_1^4 A_0)^{1/5}$, and the finiteness property holds.

For Σ_1, simply notice that $\max_{A\in\Sigma^2} \rho(A) = \max_{A\in\Sigma^2} ||A||_\infty = 2$, where $||\cdot||_\infty$ denotes the maximum row sum norm. Hence by the three members inequality we have $\rho(\Sigma) = \rho(A_0A_1)^{1/2} = \sqrt{2}$.

Consider now Σ_2. These matrices are such that $A_0^2 \leq A_0A_1$ and so any product of length t is dominated by a product of the form $A_1^{t_1}A_0A_1^{t_2}A_0\ldots A_1^{t_l}$ for some $t_1, t_l \geq 0$ and $t_i \geq 1$ $(i = 2,\ldots,l-1)$. We have

$$A_1^{t_1}A_0\ldots A_1^{t_l}A_0 = \begin{pmatrix} (t_1+1)\ldots(t_l+1) & 0 \\ (t_2+1)\ldots(t_l+1) & 0 \end{pmatrix}.$$

Again it is not difficult to show that the maximum rate of growth of the norm of such a product is equal to $\sqrt{2}$. This rate is obtained for $t_i = 3$ and $\rho = \rho(A_1^3A_0)^{1/4} = \sqrt{2}$.

The last case, Σ_3, is more complex and we give an independent proof for it.

Proposition 4.7 *The finiteness property holds for the set*

$$\left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\}.$$

Proof. Because $A_0^2 = I$ we can assume the existence of a sequence of maximal-normed products

$$\Pi_i$$

of length L_i, of the form $B_{t_1}\ldots B_{t_l}$ with $B_{t_i} = A_1^{t_i}A_0$, $\sum t_k + l = L_i$, and $\lim ||\Pi_i||^{1/L_i} = \rho(\Sigma)$. We show that actually any maximal-normed product only has factors B_3, except a bounded number of factors that are equal to B_1, B_2, or B_4 and so the finiteness property holds.

Let us analyze one of these products Π. We suppose without loss of generality that Π begins with a factor B_3 (premultiplying by B_3 does not change the asymptotic growth). First, it does not contain any factor $B_t : t > 4$ because for such t, $B_{t-3}B_2 \geq B_t$ and we can replace these factors without changing the length.

Now, our product Π has less than 8 factors B_4, because replacing the first seven factors B_4 with B_3, and the eighth one with $(B_3)^3$ we get a product of the same length but with larger norm (this is because $B_3 \geq (3/4)B_4$, $(B_3)^3 \geq (33/4)B_4$, and $(3/4)^7(33/4) > 1$). We remove these (at most) seven factors B_4 and by doing this, we just divide the norm by at most a constant K_0.

We now construct a product Π' of larger norm by replacing the left hand sides of the following inequalities by the respective right hand sides, which are products of the same length:

$$B_iB_1B_1B_j \leq B_iB_3B_j,$$
$$B_2B_1B_2 \leq B_3B_3,$$
$$B_3B_1B_2 \leq B_2B_2B_2,$$
$$B_2B_1B_3 \leq B_2B_2B_2.$$

If the factor $B_3B_1B_3$ appears eight times, we replace it seven times with $B_2^3 \geq (4/5)B_3B_1B_3$ and the last time with $B_2^3B_3^2$ which is greater than $7B_2^3$. By repeating this we get a new product $\Pi'' \geq 7(4/5)^8\Pi'(1/K_0) > \Pi'(1/K_0)$ that has a bounded number of factors B_1. We remove these factors from the product and by doing this we only divide by at most a constant K_1.
If there are more than four factors B_2 in the product, we replace the first three ones with B_3, and remove the fourth one. It appears that for any $X \in \{B_2, B_3\}, B_3^2X > 1.35B_3B_2X$, and on the other hand, $B_3^2X \geq B_3^2B_2X\frac{1}{2.4349}$. Then each time we replace four factors B_2 we get a new product: $\Pi''' \geq \frac{1.35^3}{2.4348}\Pi''(1/K_1) > \Pi''(1/K_1)$. Finally we can remove the (at most) last three factors B_2 and by doing this, we only divide the product by at most a constant K_2. By doing these operations to every Π_i, we get a sequence of products Π_i''', of length at most L_i. Now, introducing $K = K_0K_1K_2$, we compute

$$\rho \geq \lim ||\Pi_i'''||^{1/(L_i)} \geq \lim ||(1/K)\Pi_i||^{1/(L_i)} = \rho.$$

Hence $\rho = \lim ||(A_1^3A_0)^t||^{1/(4t)} = \rho(A_1^3A_0)^{1/4} = ((3+\sqrt{13})/2)^{1/4}$, and the finiteness property holds.

This concludes the proof of the main theorem of this section:

Theorem 4.5 *The finiteness property holds for any pair of* 2×2 *binary matrices.*

4.4 Conclusion

In this Chapter we have analyzed the finiteness property for matrices that have rational entries. We have shown that the finiteness property holds for matrices with nonnegative rational entries if and only if it holds for pairs of matrices with binary entries. For pairs of binary matrices of dimension 2×2 we have shown that the property holds true. It has been conjectured [11] that it holds for pairs of binary matrices of arbitrary dimension:

Open question 5 *Does the finiteness property hold for pairs of binary matrices?*

We also ask the equivalent question for matrices with negative entries:

Open question 6 *Does the finiteness property hold for pairs of matrices with entries in* $\{0, 1, -1\}$?

To the author's knowledge, in all the numerical computations that have been performed on binary matrices not only the finiteness property always seemed to hold but the period length of optimal products was always very short. The computation of the joint spectral radius is NP-hard even for binary matrices but this does not exclude the possibility of a bound on the period length that is linear, or polynomial, in the dimensions of the matrices. In the case of matrices characterizing the capacity of codes avoiding forbidden difference patterns, the length of the period is even suspected to be sublinear (see Conjecture 1 in [57] or Chapter 6).

A natural way to prove the conjecture for pairs of binary matrices would be to use induction on the size of the matrices, but this does not seem to be easy. If the conjecture is true, it follows that the stability question for matrices with nonnegative rational entries is algorithmically decidable. If the conjecture is false, then the results and techniques developed in this chapter could possibly help for constructing a counterexample.

Another way of proving the finiteness property for a set of matrices is to prove the existence of a complex polytope extremal norm. Preliminary results in this direction seem promising [47, 51].

Another problem, related to the finiteness property, seems interesting: the algebraicity of the joint spectral radius for rational matrices. Clearly, sets of rational matrices for which the finiteness property holds have a joint spectral radius which is an algebraic number. Indeed, it is the root of the characteristic polynomial of a finite product of rational matrices. The question is: is it always the case? Can this fact be put in relation with Kozyakin's theorem (Theorem 2.5 in Section 2.2), which tells us about non algebraicity of stable sets of matrices? Could it lead to a constructive counterexample to the finiteness property?

Open question 7 *Is the joint spectral radius of rational matrices always an algebraic number? Can the algebraicity of the joint spectral radius be put in relation with Kozyakin's non algebraicity result?*

Finally let us add that the constructions provided in this chapter have the additional interest that they can be used to transform the computation of the joint spectral radius of matrices with nonnegative rational entries into the computation of the joint spectral radius of two binary matrices.

Part II
Applications

Chapter 5
Continuity of wavelet functions

Abstract This chapter presents a brief survey of an important application of the joint spectral radius: the continuity of wavelet functions. Historically, this application seems to have motivated the interest of mathematicians for the joint spectral radius. This, and the fact that this application of the joint spectral radius is perhaps the one that has the greatest impact on the industry, motivates the existence of this small chapter. Our goal here is not to provide a survey of wavelet theory. We will limit ourself to present how the joint spectral radius allows to characterize the regularity of certain wavelets (Sections 5.1 and 5.2). In Section 5.3 we present two examples of such wavelets.

5.1 From two-scale difference equations to matrices

Wavelet transform is a tool of tremendous importance nowadays. For a survey, see [37, 54, 109]. The basic idea of this concept is to decompose a scalar function in $\mathbb{L}^2$ in an orthonormal base, just like with the Fourier Transform. However, Wavelets try to avoid some problems that one encounters with the Fourier Transform. For this purpose, we are looking here for compactly supported wavelets, that is

$$\exists N : \forall x \notin [0,N], \psi(x) = 0.$$

However, we want to keep the nice property that dilations of the wavelets creates an orthogonal family of functions. So, just like for the fourier transform, the different functions in the basis will be dilations of $\psi(x)$. However, since ψ has compact support, we will also need translates to represent arbitrary functions, so that the basic functions will have the following form:

$$\psi(2^j x - k).$$

Remark that here we restrict ourselves to dilations with ratios that are powers of two.

It is not obvious at first time whether the requirements for compact support and orthogonality of dilations are compatible. Actually, it was doubted that a continuous function with both of these nice properties could exist. It is the contribution of Ingrid Daubechies to have constructed such a family of functions. She shows in [36] how to obtain a wavelet: First, choose $N+1$ coefficients $c_k : 0 \leq k \leq N$. Then solve the following *two-scale difference (functional) equation:*

$$\phi(x) = \sum_0^N c_k \phi(2x-k), \tag{5.1}$$

where $\phi(x)$ is a function with compact support $[0,N]$.

The following theorem, initially inspired from [36], appears under different forms in [29, 34]:

Theorem 5.1 *Let $N \in \mathbb{N}$ and $c_0, \ldots, c_N \in \mathbb{C}$. Let us suppose that*

- $\sum_k c_{2k} = \sum_k c_{2k+1} = 1$,
- $\sum_k c_k \bar{c}_{k-2m} = \delta_{0,m}$,

where we suppose $c_k = 0$ for $k < 0$ or $k > N$.

Then, there exists a nontrivial compactly supported and square-integrable solution $\phi(x)$ that satisfies the two-scale difference equation (5.1). Moreover, the function

$$\psi(x) = \sum (-1)^k c_{N-k} \phi(2x-k)$$

is a compactly supported function such that the family

$$\psi_{j,k}(x) = \psi(2^j x - k)$$

forms an orthogonal basis for $\mathbb{L}^2$.

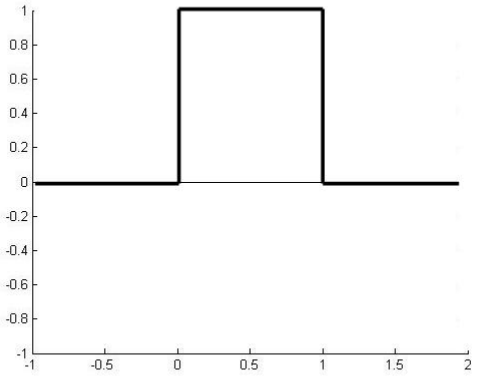
The function ϕ in the above theorem is called the *scaling function* and the function ψ is called the *mother function*. The simplest example of such a function is the well known *Haar wavelet* (see Figure 5.1), obtained from Theorem 5.1 with $c_0 = 1, c_1 = 1$:

$$\phi(x) = \phi(2x) + \phi(2x-1), \tag{5.2}$$

and thus

$$\psi(x) = \phi(2x) - \phi(2x-1). \tag{5.3}$$

This proves the existence of compactly supported wavelets, but for obvious reasons, one would like these basis functions to be continuous, and here comes linear algebra and the joint spectral radius. Let us have another look at Equation (5.1): since this equation is linear, it suggests to define the vector



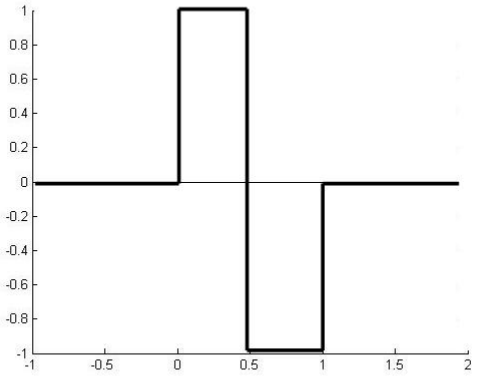


Fig. 5.1 The Haar wavelet. The scaling function ϕ and the mother wavelet ψ.

$$v(x) = \begin{pmatrix} \phi(x) \\ \phi(x+1) \\ \vdots \\ \phi(x+N-1) \end{pmatrix} \quad 0 \leq x \leq 1.$$

Thus, the vectorial function $v(x)$ satisfies the fundamental equations

$$\begin{array}{ll} v(x) = T_O v(2x) & 0 \quad \leq x \leq 1/2, \\ v(x) = T_1 v(2x-1) & 1/2 \leq x \leq 1. \end{array} \tag{5.4}$$

We denote $\Sigma = \{T_0, T_1\}$. The nonzero entries of T_0, T_1 are defined as follows:

$$(T_0)_{i,j} = c_{2i-j-1},$$

$$(T_1)_{i,j} = c_{2i-j}.$$

From now on, we write real numbers $x \in [0,1]$ in base 2. For instance, $3/4$ writes $.1100\ldots$. We say that a number is *dyadic* if its development in base 2 is not unique, that is, if it can end with an infinite sequence of zeros, or an infinite sequence of ones. These numbers are exactly the nonnegative integers divided by a power of two. Thus, introducing the notation

$$\tau(x) = 2x \mod 1,$$

that is, $\tau(x)$ is the real number obtained by erasing the first bit in the binary expansion of x, we get:

$$v(.b_1 b_2 \ldots) = T_{b_1} v(.b_2 \ldots) = T_{b_1} v(\tau(x)). \tag{5.5}$$

We can now write:

$$\begin{aligned} v(1/2) &= v(.10\ldots), \\ &= T_1 v(0), \\ v(.110\ldots) &= T_1 T_1 v(0), \\ v(.010\ldots) &= T_0 T_1 v(0), \end{aligned}$$

and so on. Finally we are able to compute $v(x)$ at each dyadic point.

5.2 Continuity and joint spectral radius

In the following we focus on dyadic numbers. Indeed, if ϕ is continuous on dyadic numbers, then it extends to a continuous function over the real numbers. Let us take two dyadic points that are close to each other:

$$x = .b_1 \dots b_k b_{k+1} \dots, \tag{5.6}$$
$$y = .b_1 \dots b_k b'_{k+1} \dots. \tag{5.7}$$

We have $|x-y| \leq 2^{-k}$. Now,

$$v(x) - v(y) = T_{b_1} \dots T_{b_k} (T_x v(0) - T_y v(0)), \tag{5.8}$$

where T_x, T_y are arbitrary products in Σ^*. Let us still define

$$W = \text{span}\{v(x) - v(y) : x, y \text{ dyadic}\}.$$

The following properties hold:

Lemma 5.1 *[29, 33] Let $T_0, T_1, W, v(x)$ be defined as above. We have the following properties*

1. $v(0) = T_0 v(0), \quad v(1) = T_1 v(1),$
2. $T_1 v(0) = T_0 v(1),$
3. $W = span\{Tv(0) - v(0) : T \in \Sigma^*\}.$

Proof. 1. This is obvious from Equation (5.4).
2. Apply again Equation (5.4) with $x = 1/2$.
3. Just write $v(x) - v(y) = (v(x) - v(0)) - (v(y) - v(0))$.

So, the value of $\phi(x)$ at integer points are actually imposed to be the entries of an eigenvector of T_0. Moreover, we will now see that the space W allows for a simple characterization: it is the smallest linear subspace that contains the vector $v(1) - v(0)$ and that is invariant under T_0 and T_1.

Proposition 5.1 *[29]*

$$W = span\{T(v(1) - v(0)) : T \in \Sigma^*\}.$$

Proof. The nontrivial part is to prove that $W \subset \text{span}\{T(v(1) - v(0)) : T \in \Sigma^*\}$.

So let us take a vector $Tv(0) - v(0)$ and let us prove that it is equal to

$$\sum \alpha_i T_i (v(1) - v(0)),$$

for some $\alpha_i \in \mathbb{C}$ and some $T_i \in \Sigma^*$. We show it by induction on the length t of the product T.

It is true if T is T_0 or T_1. Indeed, $T_0 v(0) - v(0) = 0$, and $T_1 v(0) - v(0) = T_0 v(1) - v(0) = T_0(v(1) - v(0))$.

If $T = T_{b_1} T'$, then by induction

$$\begin{aligned} Tv(0) - v(0) &= T_{b_1}(T'v(0) - v(0)) + T_{b_1} v(0) - v(0) \\ &= T_{b_1}(\sum \alpha_i T_i (v(1) - v(0))) + \sum \alpha'_i T'_i (v(1) - v(0)). \end{aligned}$$

Actually, it is possible to derive a slightly stronger result, that will be useful for the characterization of the continuity in terms of a joint spectral radius:

Corollary 5.1 *Let $T_{0|W}, T_{1|W}$ be the corresponding matrices restricted to the linear space W, and let $\rho(T_{0|W}, T_{1|W}) < 1$.*
Then, there is a constant K such that for all dyadic x, y, $|v(x) - v(y)| < K$.

Proof. From the proof of Proposition 5.1, we get that if $T \in \Sigma^t$,

$$Tv(0) - v(0) = \sum_1^t \alpha_i T_i (v(1) - v(0)),$$

where $T_i \in \Sigma^i$, and α_i is equal to zero or one. Then, using Proposition 1.4

$$|Tv(0) - v(0)| \leq \sum_1^t \gamma^i |v(1) - v(0)| \leq K_1,$$

for some particular norm and any γ: $\rho < \gamma < 1$. Finally, $|v(x) - v(y)| = |(v(x) - v(0)) - (v(y) - v(0))| \leq 2K_1$.

The above reasoning allows us to prove the two main results of this section, that explicits the link between the joint spectral radius and the continuity of wavelets:

Theorem 5.2 *If ϕ is continuous,*

$$\rho(T_{0|W}, T_{1|W}) < 1.$$

Proof. We prove it by contraposition. If $\rho \geq 1$, there is a constant K and there are arbitrarily long products $T \in \Sigma^t$ such that $|T(v(1) - v(0))| > K$. From (5.8) we get $|v(.b_1 \dots b_t 1) - v(.b_1 \dots b_t 0)| > K$, while $|.b_1 \dots b_t 1 - .b_1 \dots b_t 0| < 2^{-t}$. Since this holds for arbitrary t, we reach a contradiction.

The converse also holds:

Theorem 5.3 *If*

$$\rho(T_{0|W}, T_{1|W}) < 1,$$

then ϕ is continuous on the dyadic numbers.

Proof. We take two dyadic numbers x and y that are close enough: $|x - y| < 2^{-t}$, for a given $t \in \mathbb{N}$. Thus, $x = .b_1 \dots b_t b_{t+1} \dots$, $y = .b_1 \dots b_t b'_{t+1} \dots$. Now, we invoke Corollary 5.1, and we obtain some constant K_2 such that

$$|v(x) - v(y)| = T_{b_1} \dots T_{b_t} (v(\tau^t(x)) - v(\tau^t(y))) < K_2 ||T_{b_1} \dots T_{b_t}||.$$

So we obtain $|v(x) - v(y)| < K_1 \rho^t K_2 < K \rho^t$ for some K. This proves that ϕ is continuous on the dyadic numbers.

A direct refinement of the above results is possible when looking at Equation (5.8): the joint spectral radius characterizes the *Holder exponent of continuity* of the scaling function.

Definition 5.1 *A scalar function is said* Holder continuous *with coefficient* α *if there exists a constant K such that*

$$|f(x)-f(y)| \leq K|x-y|^{\alpha}$$

for all x, y *in the domain of* f.

We have the following theorem:

Theorem 5.4 *[29] If*

$$\rho(T_{0|W}, T_{1|W}) < 1,$$

then ϕ *is Holder-continuous for all coefficient* $\alpha < -\log_2(\rho)$. *If moreover* $\{T_{0|W}, T_{1|W}\}$ *is nondefective,* $\alpha = -\log_2(\rho)$ *is a Holder exponent.*

Further details on this can be found in [29]. Let us add finally that with these techniques it is also possible to prove that no scaling function is in $\mathscr{C}^{\infty}$ [32].

5.3 Example

In this section we give an example of a compactly supported continuous square integrable wavelet that was proposed by Daubechies [36]. For this, take the coefficients

$$c_0 = \frac{1}{4}(1+\sqrt{3}),\ c_1 = \frac{1}{4}(3+\sqrt{3}),\ c_2 = \frac{1}{4}(3-\sqrt{3}),\ c_3 = \frac{1}{4}(1-\sqrt{3}).$$

This example is sometimes referred to as D_4 in the literature [108]. A simple application of the formulas gives:

$$T_0 = \begin{pmatrix} c_0 & 0 & 0 \\ c_2 & c_1 & c_0 \\ 0 & c_3 & c_2 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1+\sqrt{3} & 0 & 0 \\ 3-\sqrt{3} & 3+\sqrt{3} & 1+\sqrt{3} \\ 0 & 1-\sqrt{3} & 3-\sqrt{3} \end{pmatrix},$$

$$T_1 = \begin{pmatrix} c_1 & c_0 & 0 \\ c_3 & c_2 & c_1 \\ 0 & 0 & c_3 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3+\sqrt{3} & 1+\sqrt{3} & 0 \\ 1-\sqrt{3} & 3-\sqrt{3} & 3+\sqrt{3} \\ 0 & 0 & 1-\sqrt{3} \end{pmatrix}.$$

Thus,

$$v(0) = \begin{pmatrix} 0 \\ 1+\sqrt{3} \\ 1-\sqrt{3} \end{pmatrix},$$

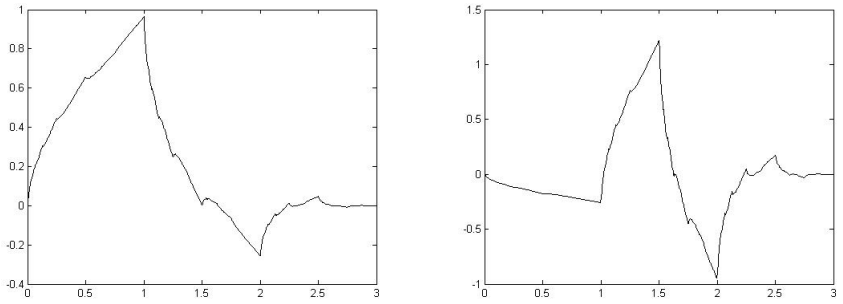
$$v(1) = \begin{pmatrix} 1+\sqrt{3} \\ 1-\sqrt{3} \\ 0 \end{pmatrix}.$$

One can show that $W = \{(x,y,z) : x+y+z = 0\}$. So, choosing $(1,-1,0)$ and $(1,1,-2)$ as basis vectors, we get

$$T_{0|W} = \frac{1}{8} \begin{pmatrix} 1+3\sqrt{3} & 3-\sqrt{3} \\ \sqrt{3}-3 & 5-\sqrt{3} \end{pmatrix},$$

$$T_{1|W} = \frac{1}{4} \begin{pmatrix} 2 & 3+\sqrt{3} \\ 0 & 1-\sqrt{3} \end{pmatrix}.$$

The joint spectral radius of the set $\{T_{0|W}, T_{1|W}\}$ is equal to $2^{-0.55\ldots}$, so the wavelet is Holder continuous with exponent 0.55. Figure 5.2 represents the corresponding scaling function and the mother wavelet.



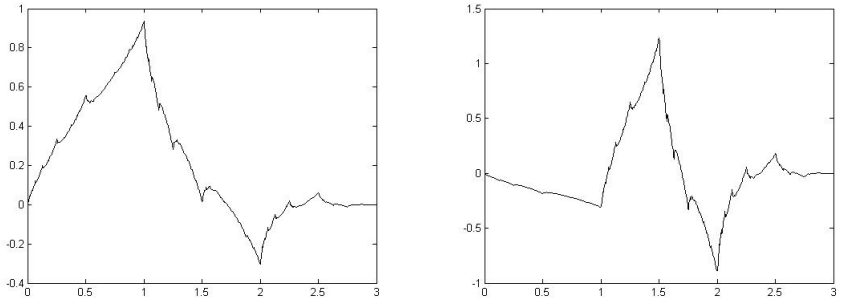
(a) (b)

Fig. 5.2 Scaling function (a) and mother wavelet (b) for the Daubechies wavelet D_4 (the notation D_4 comes from [108]).

Figure 5.3 represents the scaling function and the mother wavelet for another choice of coefficients c_k.

5.4 Conclusion

The goal of this chapter was not to give an introduction to the wavelet theory, but rather to present how the continuity of functions satisfying the two-scale difference Equation (5.1) is ruled by a joint spectral radius. We end with an obvious question:



(a) (b)

Fig. 5.3 Scaling function (a) and mother wavelet (b) obtained with $c_0 = 3/5$, $c_1 = 6/5$, $c_2 = 2/5$, $c_3 = -1/5$.

Open question 8 *Is the joint spectral radius of matrices arising in the context of wavelets easier to compute than in general?*

To the author's knowledge, no particular algorithm is known, that would be dedicated to this specific family of sets of matrices.

Chapter 6
Capacity of codes

Abstract This chapter presents personal research on an application of the joint spectral radius to a problem in constrained coding: the computation of the capacity of codes submitted to forbidden differences constraints. We first present how the joint spectral radius appears to be the good tool to compute the capacity in this particular problem. We show how the quantity $\hat{\rho}_t$ provides bounds on the joint spectral radius that are tighter than in the general case. We show how the situation is even better in some particular situations. We then provide a polynomial time algorithm that decides if the capacity is positive. We introduce a closely related problem that we prove to be NP-hard. We then prove the existence of extremal norms for sets of matrices arising in this coding problem.

6.1 Introduction

In certain coding applications one is interested in binary codes whose elements avoid a set of forbidden patterns[1]. This problem is rather classical and has been widely studied in the past century [75]. In order to minimize the error probability of some particular magnetic-recording systems (see for instance [81]), a more complicated problem arises when it is desirable to find code words whose *differences* avoid forbidden patterns. We now describe this problem formally.

Let $\{0,1\}^t$ denote the set of words of length t over $\{0,1\}$ and let $u,v \in \{0,1\}^t$. The difference $u-v$ is a word of length t over $\{-1,0,+1\}$ (as a shorthand we shall use $\{-,0,+\}$ instead of $\{-1,0,+1\}$). The difference $u-v$ is obtained from u and v by symbol-by-symbol subtraction so that, for example, $0110-1011 = -+0-$. Consider now a finite set D of words over $\{-,0,+\}$; we think of D as a set of *forbidden difference patterns*. A set (or *code*) $C \subseteq \{0,1\}^t$ is said to *avoid* the set D if none of the

[1] The chapter presents research work that has been published in [11, 12].

differences of words in C contain a word from D as subword, that is, none of the differences $u-v$ with $u,v \in C$ can be written as $u-v=xdy$ for $d \in D$ and some (possibly empty) words x and y over $\{-,0,+\}$.

We are interested in the largest cardinality, which we denote by $\delta_t(D)$, of sets of words of length t whose differences avoid the forbidden patterns in D.

$$\delta_t(D) = \max_{W \subset \{0,1\}^t : W \text{ avoids } D} |W|.$$

If the set D is empty, then there are no forbidden patterns and $\delta_t(D) = 2^t$. We will see that when D is nonempty, $\delta_t(D)$ grows exponentially with the word length t and is asymptotically equal to $2^{\text{cap}(D)t}$ where the scalar $0 \leq \text{cap}(D) \leq 1$ is the *capacity* of the set D. The capacity is thus a measure of how constraining a set D is; the smaller the capacity, the more constraining the forbidden difference patterns are.

As an illustration consider the set of forbidden patterns $D = \{+-,++\}$. Differences between two words in $C = \{u_1 0 u_2 0 \cdots 0 u_k : u_i \in \{0,1\}\}$ will have a "0" in any succession of two characters and will therefore not contain any of the forbidden patterns. From this it follows that $\delta_t \geq 2^{\lceil t/2 \rceil}$ and so $\text{cap}(D) \geq 1/2$. One can show that in fact $\text{cap}(D) = 1/2$. This follows from the next proposition combined with the simple observation that the capacity of the set $D = \{+-,++\}$ is identical to the capacity of the set $D = \{+-,++,-+,--\}$, that we denote $D = \{+,-\}^2$ as usual.

Proposition 6.1 *The capacity of the set $\{+,-\}^m$ is given by $(m-1)/m$.*

Proof. Let C_{km} be a code of length km avoiding D. In any given window of length m, the set of words appearing cannot contain both u and $\bar{u}$ (we use $\bar{u}$ to denote the word obtained by inverting the ones and the zeros in u). This implies that there are at most 2^{m-1} different words in any given window of size m. Let us now consider words in C_{km} as a concatenation of k words of length m. There are at most $2^{(m-1)k}$ words in C_{km} and so $\text{cap}(D) \leq (m-1)/m$.
Now consider the code

$$C_{km} = \{z_1 0 z_2 0 \cdots 0 z_k : z_i \in \{0,1\}^{m-1}\}. \tag{6.1}$$

This code satisfies the constraints, and the bound $(m-1)/m$ is reached.

The computation of the capacity is not always that easy. As an example it is proved in [82] that the capacity of $\{+++\}$ is given by $\log_2((1+(19+3\sqrt{33})^{1/3}+(19-3\sqrt{33})^{1/3})/3) = .8791\ldots$ and the same reference provides numerical bounds for the capacity of $\{0+-+\}$ for which no explicit expression is known.

The capacity of codes that avoid forbidden difference patterns was first introduced and studied by Moision, Orlitsky and Siegel. In [82], these authors provide explicit values for the capacity of particular sets of forbidden patterns and they prove that, in general, the capacity of a forbidden set D can be obtained as the logarithm of the joint spectral radius of a set of matrices that have binary entries. The size of the matrices

constructed in [82] for computing the capacity is not polynomial in the size of the forbidden set D and so even the construction of the set of matrices is an operation that cannot be performed in polynomial time. Since moreover the computation of the joint spectral radius is NP-hard even if the matrices have binary entries, computing the capacity of codes seems at first sight to be a challenging task. However, as pointed out in [82], the matrices that arise in the context of capacity computation have a particular structure and so the capacity could very well be computable in polynomial time.

In this chapter we first present this in details. We then provide several results ; all are related to the capacity computation and its complexity.

We first provide new bounds that relate the capacity of a set of forbidden patterns D with the values $\delta_t(D)$, the maximum size of a code of length t avoiding D. These bounds depend on parameters that express the number and positions of zeros in the patterns of D. These new bounds allow us to compute the capacity of any set to any given degree of accuracy by numerically evaluating $\delta_t(D)$ for some value of t. The approximation algorithm resulting from these bounds has exponential complexity but provides an a-priori guaranteed precision, and so the computational effort required to compute the capacity to a given degree of accuracy can be evaluated before the calculations are actually performed. As an example, it follows from the bounds we provide that the capacity of a set of forbidden patterns that does not contain any 0s can be computed with an accuracy of 90% by evaluating $\delta_t(D)$ for $t = 10$ (see Corollary 6.3 below).

In a subsequent section, we provide explicit necessary and sufficient conditions for a set to have positive capacity and we use this condition for producing a polynomial time algorithm that decides whether or not the capacity of a set is positive. These conditions are directly based on theoretical results presented in Chapter 3.

We then consider the situation where in addition to the forbidden symbols $-, 0$ and $+$ the forbidden patterns in D may also include the symbol $\pm$, where $\pm$ stands for both the symbols $+$ and $-$. We prove that in this case the problem of computing the capacity, or even determining if this capacity is positive, becomes NP-hard.

Finally, we show that sets of matrices constructed in order to compute the capacity always have an extremal norm.

These results allow us to better delineate the capacity computation problems that are polynomial time solvable from those that are not. We do however not provide an answer to the question, which was the original motivation for the research reported here, as to whether or not one can compute the capacity of sets of forbidden patterns over $\{-, 0, +\}$ in polynomial time. This interesting question that was already raised in [82], remains unsettled.

6.2 Capacity and joint spectral radius

Let D be a set of forbidden patterns over the alphabet $\{-,0,+\}$ and consider for any $t \geq 1$ the largest cardinality, denoted by $\delta_t(D)$, of sets of words of length t whose pairwise differences avoid the forbidden patterns in D. The capacity of D is defined by

$$\mathrm{cap}(D) = \lim_{t\to\infty} \frac{\log_2 \delta_t(D)}{t}. \tag{6.2}$$

The existence of this limit is a simple consequence of Fekete's Lemma (Lemma 1.1). We skip the formal proof, since it will be clear after the formulation of the problem with a joint spectral radius.
Moision et al. show in [82] how to represent codes submitted to a set of constraints D as products of matrices taken in a finite set $\Sigma(D)$. The idea of the proof is to make use of De Bruijn graphs. De Bruijn graphs were introduced in [38]; for an introduction, see for instance [75]. Let us construct the De Bruijn graph of binary words of length T equal to the lengths of the forbidden patterns. Edges in these graphs represent words of length T, and since some pairs of words cannot appear together, a subgraph of the De Bruijn graph is said *admissible* if it does not contain two edges that represent words of length T whose difference is forbidden. Figure 6.1 (a) represents a De Bruijn graph that is admissible for the forbidden pattern $D = \{++-\}$. An efficient way of drawing these graphs is to represent them as cascade graphs (see Chapter 3) as in Figure 6.1 (b). In order to construct longer codes, one just has to juxtapose admissible cascade graphs, such that each path from left to right represents an admissible word.

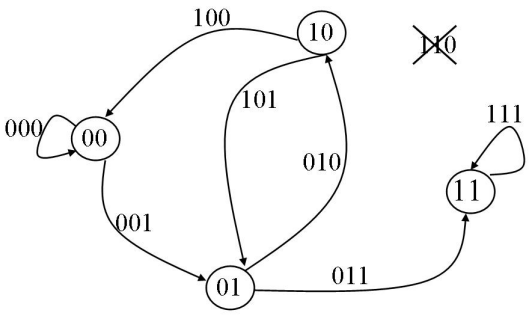
In such a construction, the edges in the leftmost cascade graph represent words of length T, and each subsequent edge represents the addition of one letter to the word. A cascade graph for words of length 5 that are admissible for $D = \{++-\}$ is represented in Figure 6.2. Since we have a bijection between the paths of length t in an admissible cascade graph and the words in an admissible code of length $T+t-1$, the maximal size of a code of length $T+t-1$ is given by the cascade graph of length T that maximizes the number of paths from left to right. We have seen in Chapter 3 how the joint spectral radius of binary matrices represents the asymptotics of the maximum number of paths in long cascade graphs. This reasoning leads to the following theorem:

Theorem 6.1 *Associated to any set D of forbidden patterns of length at most m, there exists a finite set $\Sigma(D)$ of binary matrices for which*

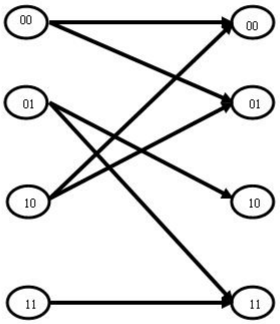
$$\delta_{m-1+t} = \hat{\rho}_t^t(\Sigma(D)) = \max\{\|A_1 \dots A_t\| : A_i \in \Sigma(D)\}. \tag{6.3}$$

In this expression, the matrix norm used is the sum of the absolute values of the matrix entries. The main result of this section is then a direct consequence of the definition of the joint spectral radius:

Corollary 6.1 *Let D be a set of forbidden patterns and $\Sigma(D)$ be the set of binary matrices constructed as described above, then*



(a)



(b)

Fig. 6.1 An admissible De Bruijn graph for $D = \{++-\}$ (a), and the same graph under its cascade graph form (b)

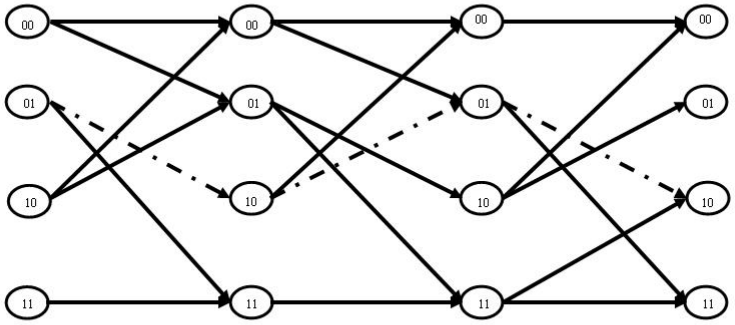


Fig. 6.2 An admissible cascade graph that represents a maximal set of admissible words of length 5 for $D = \{++-\}$. For example, the path on the top represents the word 00000 and the dashed path represents the word 01010. Such graphs are maximal in the sense that no word can be added to the corresponding code, but perhaps another choice of elementary cascade graphs would generate more paths.

$$cap(D) = \log_2(\rho(\Sigma(D))).$$

Example 6.1 *Let $D = \{++-\}$. The set $\Sigma(D)$ contains two matrices :*

$$A_0 = \begin{pmatrix} 1\,1\,0\,0 \\ 0\,0\,1\,1 \\ 1\,1\,0\,0 \\ 0\,0\,0\,1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1\,0\,0\,0 \\ 0\,0\,1\,1 \\ 1\,1\,0\,0 \\ 0\,0\,1\,1 \end{pmatrix}.$$

One can check that the cascade graph in Figure 6.2 represents the product $A_0A_0A_1$ (the sum of the entries equals the number of paths).

The joint spectral radius of the set Σ is $\rho(\Sigma) = 1.75\ldots$ [82], and the product that ensures this value is $A_0A_0A_1A_1$, that is, $\rho(\Sigma) = \rho(A_0^2A_1^2)^{1/4}$, and $cap(D) = \log_2 1.75\ldots = 0.8113\ldots$.

Example 6.2 *Let $D = \{+++-\}$. The set $\Sigma(D)$ contains two matrices :*

$$A_0 = \begin{pmatrix} 1&1&0&0&0&0&0&0\\ 0&0&1&1&0&0&0&0\\ 0&0&0&0&1&1&0&0\\ 0&0&0&0&0&0&1&1\\ 1&1&0&0&0&0&0&0\\ 0&0&1&1&0&0&0&0\\ 0&0&0&0&1&1&0&0\\ 0&0&0&0&0&0&0&1 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&0&1&1&0&0&0&0\\ 0&0&0&0&1&1&0&0\\ 0&0&0&0&0&0&1&1\\ 1&1&0&0&0&0&0&0\\ 0&0&1&1&0&0&0&0\\ 0&0&0&0&1&1&0&0\\ 0&0&0&0&0&0&1&1 \end{pmatrix}.$$

We will see that $cap(D) = 0.9005\ldots$ and that the product that ensures this value is A_0A_1 (see Example 6.5);

Let us comment here on the number and size of the matrices in $\Sigma(D)$; these issues are relevant for the questions raised hereafter: If the forbidden patterns in D have identical length m, then the number of matrices in $\Sigma(D)$ can be doubly exponential in m and all matrices in $\Sigma(D)$ have dimension $2^{m-1} \times 2^{m-1}$. If the forbidden patterns in D have different lengths, then one can construct a set D' whose forbidden patterns have equal length and for which $\mathrm{cap}(D) = \mathrm{cap}(D')$. Unfortunately, the number of patterns in D' can grow exponentially with the size of D so that the number of matrices in the set $\Sigma(D)$ is in fact even worse than in the former case. Capacity approximation algorithms based on the direct computation of the set $\Sigma(D)$ will therefore not be tractable even for small sets D.

6.3 Upper and lower bounds

In this section, we derive bounds that relate the capacity of a set D with $\delta_t(D)$. Consider some set D of forbidden patterns and denote by r_1 (respectively r_2) the maximal k for which 0^k is the prefix (respectively suffix) of some pattern in D: No pattern in D begins with more than r_1 zeros and no pattern in D ends with more than r_2 zeros. We also denote by r the maximal number of consecutive zeros in any pattern in D;

obviously, $r \geq \max(r_1, r_2)$. In the next theorem we provide upper and lower bounds on the capacity $\mathrm{cap}(D)$ in terms of $\delta_t(D)$.

Theorem 6.2 *For any $t \geq r_1 + r_2$ we have*

$$\frac{\log_2 \delta_t(D) - (r_1 + r_2)}{t + r + 1 - (r_1 + r_2)} \leq \mathrm{cap}(D) \leq \frac{\log_2 \delta_t(D)}{t}. \tag{6.4}$$

Proof. Let us first consider the upper bound. The following equation is straightforward, given any positive integers k, t, and any set of forbidden patterns D :

$$\delta_{kt} \leq \delta_t^k.$$

Indeed, considering any word of length kt as the concatenation of k subwords of length t, for each of these subwords we have at most δ_t possibilities. Taking the $\frac{1}{kt}$th power of both sides of this inequality and taking the limit $k \to \infty$, we obtain :

$$\rho(\Sigma) = 2^{\mathrm{cap(D)}} \leq \delta_t^{1/t}.$$

Now let us consider the lower bound. The optimal code of length t contains at least $\lceil 2^{-r_1 - r_2} \delta_t(D) \rceil$ words that coincide in the first r_1 bits and in the last r_2 bits (because there are in total $2^{r_1 + r_2}$ different words of length $r_1 + r_2$). Denote the set of strings of all these words from $(r_1 + 1)$st bit to $(t - r_2)$th bit by C'. This set contains at least $\lceil 2^{-r_1 - r_2} \delta_t(D) \rceil$ different words of length $t - r_1 - r_2$. Then for any $l \geq 1$ the code

$$C = \left\{ u_1 0^{r+1} u_2 0^{r+1} \cdots 0^{r+1} u_l 0^{r+1},\ u_k \in C',\ k = 1, \ldots, l \right\} \tag{6.5}$$

avoids D. The cardinality of this code is at least $\lceil 2^{-r_1 - r_2} \delta_t(D) \rceil^l$ and the length of its words is $T = l(t - r_1 - r_2 + r + 1)$. Therefore, for any l we have

$$\delta_T(D) \ \geq \ \lceil 2^{-r_1 - r_2} \delta_t(D) \rceil^l.$$

Taking the power $1/T$ of both sides of this inequality, we get

$$\left[\delta_T(D) \right]^{1/T} \ \geq \ \lceil 2^{-r_1 - r_2} \delta_t(D) \rceil^{1/(t - r_1 - r_2 + r + 1)},$$

which as $T \to \infty$ yields

$$\rho \ \geq \ \lceil 2^{-r_1 - r_2} \delta_t(D) \rceil^{1/(t - r_1 - r_2 + r + 1)}.$$

Now after elementary simplifications we arrive at the lower bound on $\mathrm{cap}(D)$.

Both bounds in Theorem 6.2 are sharp in the sense that they are both attained for particular sets D. The upper bound is attained for the set $D = \emptyset$ and the lower bound is attained, for instance, for the set $D = \{0^{m-1}+\}$. Indeed, in this case $r = r_1 = m -$

$1, r_2 = 0$ and $\text{cap}(D) = 0$, because $\delta_t = 2^{m-1}$ for $t \geq m-1$. Here is a direct proof of this equality, drawn from [82]: Clearly, for all $t > m-1$, we can construct a code of size $\delta_t = 2^{m-1}$. It happens that for any given length t this size is maximum. Otherwise, there must be two *different* words u and v whose prefixes of length k coincide. In order to avoid the forbidden pattern, the $k+1$-th symbols must also be equal, and so on. But then both words are equal, and we have reached a contradiction.

Corollary 6.2 *Let D be given and let r, r_1 and r_2 be defined as above. Then* $\frac{\log_2 \delta_t(D)}{t} - \frac{1}{t}\max(r_1 + r_2, r+1) \leq \text{cap}(D) \leq \frac{\log_2 \delta_t(D)}{t}$.

Proof. If $r_1 + r_2 \geq r+1$ this follows from Theorem 6.2 and from simple calculations. If $r_1 + r_2 < r+1$ simply use the fact that the capacity is always less than one in Theorem 6.2, and

$$\frac{\log_2 \delta_t(D)}{t} - (r_1 + r_2) \leq (t + (r+1) - (r_1 + r_2))\text{cap}(D) \leq t\,\text{cap}(D) + (r+1) - (r_1 + r_2).$$

These bounds can be used to design an approximation algorithm that computes the capacity to any desired accuracy by evaluating δ_t for sufficiently large values of t. In contrast to previously known algorithms this algorithm has guaranteed computational cost: once the desired accuracy is given, the corresponding computational cost can easily be computed. As an illustration, consider the case of a set D for which $r_1 = r_2 = 2$ and $r = 4$. Then, by Corollary 6.2,

$$\frac{\log_2 \delta_t(D)}{t} - \frac{5}{t} \leq \text{cap}(D) \leq \frac{\log_2 \delta_t(D)}{t} \tag{6.6}$$

and we can use $\log_2 \delta_t(D)/t$ as an estimate for $\text{cap}(D)$ and choose a value of t for which (6.6) provides satisfactory accuracy bounds.
The easiest way of computing δ_t is to apply Equation (6.3), by evaluating the maximum-normed product of length $t-m+1$ of matrices taken in the set Σ. Moision et al. mention in [83] an improvement of this brute force method, similar to the ones proposed in Chapter 2: The main idea is to compute successively some sets of matrices $\bar{\Sigma}_l$, $l = 1, 2\ldots$, with $\bar{\Sigma}_1 = \Sigma$. These are sets of products of length l, obtained by computing iteratively all products of a matrix in $\bar{\Sigma}_{l-1}$ with a matrix in Σ, and then removing from the set $\bar{\Sigma}_l$ a matrix A if it is dominated by another matrix B in this set, that is, if each entry of A is less or equal than the corresponding entry of B. For more information about this algorithm, we refer the reader to [83]. We propose here an improvement of this method: given the set $\bar{\Sigma}_l$, one can directly compute a set $\bar{\Sigma}_{2l}$ by computing the set $\bar{\Sigma}_l^2$ and then removing from this set all matrices that are dominated. This small modification of the algorithm has dramatically improved the computational time for all the examples on which we have used it.

We may specialize the general bounds of Theorem 6.2 to sets of particular interest.

Corollary 6.3 *Let D be given and let r, r_1 and r_2 be defined as above. Then*

1. *If* $\mathrm{cap}(D) = 0$ *the size of any code avoiding* D *is bounded above by the constant* $2^{r_1+r_2}$.
2. *If the patterns in* D *contain no zero, then*

$$t\,\mathrm{cap}(D) \leq \log_2 \delta_t(D) \leq (t+1)\,\mathrm{cap}(D).$$

3. *If none of the patterns in* D *starts nor ends with a zero, then*

$$t\,\mathrm{cap}(D) \leq \log_2 \delta_t(D) \leq (t+r+1)\,\mathrm{cap}(D).$$

6.4 Positive capacity can be decided in polynomial time

As previously seen by a direct argument, the capacity of the set $\{0^{m-1}+\}$ is equal to zero. In this section we provide a systematic way of deciding when the capacity of a set is equal to zero. We first provide a simple positivity criterion that can be verified in finite time and then exploit this criterion for producing a positivity checking algorithm that runs in polynomial time. In the sequel we shall use the notation $-D$ to denote the set of elements that are the opposites to the elements of D, for example if $D = \{-+0, 0--\}$ then $-D = \{+-0, 0++\}$.

Theorem 6.3 *Let* D *be a set of forbidden patterns of lengths at most* m. *Then* $\mathrm{cap}(D) > 0$ *if and only if there exists a word on the alphabet* $\{+,-,0\}$ *that does not contain any word of* $D \cup -D$ *as subword and that has a prefix* 0^m *and a suffix* $+0^{m-1}$.

Proof. Let us first suppose $0^m \notin D$. The capacity is positive iff $\rho(\Sigma(D)) > 1$. We know (see Chapter 3) that for binary matrices this is equivalent to the fact that there is a product in Σ^* that has a diagonal entry larger than one. In turn, by construction of the set $\Sigma(D)$, this is equivalent to the existence of two words with the same $m-1$ first characters, and the same $m-1$ last characters, whose difference avoids the forbidden patterns. Now, this latter fact is possible iff there is a nontrivial sequence on $\{0,+,-\}$ of the shape $0^{m-1}d0^{m-1}$ that avoids $D \cup -D$.
Now in order to handle the case $0^m \in D$, which implies $\mathrm{cap}(D) = 0$, we add a zero at the beginning and by doing this, we do not change anything to the admissibility of this word, except that we remove the possibility $0^m \in D$.

Corollary 6.4 *If every word in* D *contains at least two nonzero symbols, then*

$$cap(D) > 0.$$

Proof. For any such set the word $d = 0^m + 0^{m-1}$ is admissible, and by Theorem 6.3 the capacity is positive.

Corollary 6.5 *If* D *consists of one forbidden pattern* p *of length* m, *then its capacity is zero if and only if* p *has at least* $m-1$ *consecutive zeros.*

Proof. If a pattern p is 0^m or $+0^{m-1}$, then obviously there are no admissible strings, and by Theorem 6.3 the capacity is zero. The same holds for -0^{m-1}, since this is the negation of $+0^{m-1}$ and for $0^{m-1}\pm$ because of the symmetry. In all the other cases the admissible string exists and so $\mathrm{cap(D)} > 0$. Indeed, if p has a unique nonzero character, then the word $d = 0^m + +0^{m-1}$ is admissible, if it has at least two nonzero characters, then the proof follows from Corollary 6.4.

We now prove the polynomial-time solvability of the problem of determining whether the capacity of a set D is positive. The proof is constructive and is based on the so-called *Aho-Corasick* automaton that checks whether a given text contains as a subsequence a pattern taken from a given set [1]. Let P be a set of patterns, that do not have to be of the same length. The transition graph of the Aho-Corasick automaton for the set P is defined as follows (see Figure 6.3 for an example). First, construct the *retrieval tree*, or trie, of the set P. The trie of P is the directed tree of which each vertex has a label representing a prefix of a pattern in P, and all prefixes are represented, including the patterns themselves. The label of the root of the tree is the empty string. Edges have a label too, which is a symbol of the used alphabet. There is an edge labeled with the symbol a from a vertex s to a vertex t if t is the concatenation sa.
In order to have an automaton, we complete the trie by adding edges so that for each vertex s, and each symbol a, there is an edge labeled a leaving s. This edge points to the vertex of the trie of which the label is the longest suffix of the concatenation sa. Note that this vertex can be the root (that is, the empty string) if no vertex in the trie is a suffix of sa. Finally, the accepting states of the automaton are the vertices whose labels are patterns of P. This automaton accepts words that contain a pattern in P and halts whenever this pattern is a suffix of the entered text.

If $0^k \in D$ or $+0^k \in D$, then, by Theorem 6.3, $\mathrm{cap}(D) = 0$. If this is not the case, we construct the graph of the automaton of Aho-Corasick for the set $P = D \cup (-D) \cup \{+0^{m-1}\}$. We then remove any vertex whose label is a complete pattern in P (i.e., a state reached when a suffix of the text entered is in the set P) except the vertex labeled $\{+0^{m-1}\}$. The size of the constructed graph is polynomial in the size and the number of the forbidden patterns. Moreover, since we have removed vertices corresponding to forbidden patterns, any path in the remaining graph is an admissible word. Let us now denote q_{0^m} the state reached after entering the word 0^m. This state is well defined since 0^m does not contain any forbidden pattern, and hence no state reached after entering any prefix of the string 0^m was removed from the primary automaton. We also denote $q_{+0^{m-1}}$ the state corresponding to the suffix $+0^{m-1}$ for the entered text (i.e. the accepting state corresponding to the pattern $+0^{m-1}$ in the Aho-Corasick automaton). Figure 6.3 presents the graph for $D = \{0+0\}$ that is obtained from the Aho-Corasick automaton of the set $P = \{0+0, 0-0, +00\}$.

We have the following criterion for zero-capacity:

Theorem 6.4 *The capacity of a set D is positive if and only if there is a path from q_{0^m} to $q_{+0^{m-1}}$ in the graph constructed above.*

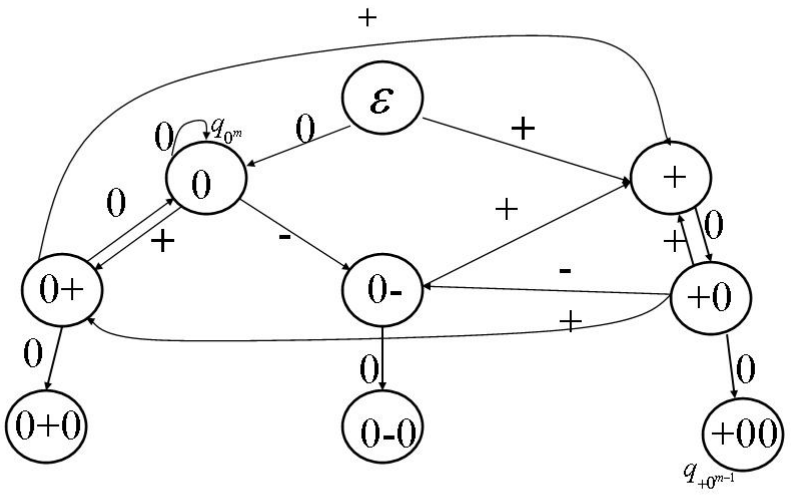


Fig. 6.3 The graph for $D = \{0+0\}$. We have constructed the Aho-Corasick automaton for $P = \{0+0, 0-0, +00\}$, and then removed the states $0+0$ and $0-0$ that are forbidden. The empty word is represented by ε. The path $0 \to 0- \to + \to +0 \to +00$ provides the admissible word $000-+00$.

Proof. If cap(D) > 0, by Theorem 6.3, there exists a word d, beginning with m zeros, and ending with $+0^{m-1}$, that avoids $D \cup -D$. Hence, entering this word in the automaton, the finite state will be (well defined and will be) the vertex labeled $+0^{m-1}$, because the vertices removed from the original automaton of Aho-Corasick do not make any problem, since we do not reach the vertices labeled with forbidden patterns.
On the other hand, a path in the constructed graph represents an acceptable word, since it does not pass through any removed vertex, and hence no suffix of any prefix of this word will be in the forbidden set.
Moreover, a shortest path will give the shortest acceptable word, since the length of the path is equal to the length of the represented word.

Corollary 6.6 *The problem of determining whether or not the capacity of a given set of forbidden patterns is positive can be solved in polynomial time.*

Proof. Aho shows in [1] that the automaton is constructible in polynomial time. The determination of the state q_{0^m} and the computation of the shortest path are obviously polynomially feasible.

Corollary 6.7 *If for a set D of forbidden patterns there are admissible words, then the length of a shortest admissible word does not exceed $2M+2m$, where m is the maximal length of all patterns in D and M is the sum of the lengths of each forbidden pattern.*

Proof. The number of vertices of the graph does not exceed $2M+m+1$. Indeed, for each pattern of length l in $D \cup -D$ we add to the automaton at most l states, since there are no more than l prefixes of this pattern. We still add the pattern $\{+0^{m-1}\}$ (maximum m new states), and the root. If there is a path connecting two given vertices, this path can be chosen so that its length (in terms of number of vertices) will not exceed the total number of vertices (if it does not pass through the same vertex twice). Every edge of this path adds one bit to the admissible string. The initial length of the string is m (we start from 0^m), therefore the total length of the admissible word is at most $2M+2m$.

Proposition 6.2 *If the capacity is positive, then* $\mathrm{cap}(D) > 1/(2M+m)$*, where m is the maximal length of all patterns in D and M is the sum of the lengths of each forbidden pattern.*

Proof. If $\mathrm{cap(D)} > 0$, then there is an admissible string of length $t \leq 2M+2m$ (Corollary 6.7). Consider a code as given by Equation (6.5). Its size is 2^l and the length of its words is at most

$$T_l \;=\; l(2M+2m-m) = l\big(2M+m\big).$$

Therefore

$$\begin{aligned} \mathrm{cap(D)} \;&=\; \lim_{\mathrm{l}\to\infty} \frac{\log_2 \delta_{\mathrm{T_l}}}{\mathrm{T_l}} \\ \geq\; \lim_{l\to\infty} \frac{\log_2 2^l}{l\big(2M+m\big)} \;&=\; \frac{1}{2M+m}. \end{aligned}$$

6.5 Positive capacity is NP-hard for extended sets

We now consider the situation where forbidden patterns are allowed to contain the $\pm$ symbol. The symbol $\pm$ is to be understood in the following sense: whenever it occurs in a forbidden pattern, both the occurrences of $+$ and of $-$ are forbidden at that particular location. So, for example, avoiding the forbidden set $\{0\pm+\pm\}$ is equivalent to avoiding the set $\{0{+}{+}{+},0{+}{+}{-},0{-}{+}{+},0{-}{+}{-}\}$. All results obtained for forbidden patterns over $\{-,0,+\}$ have therefore their natural counterparts in the situation where the forbidden patterns are defined over the alphabet $\{-,0,+,\pm\}$. In particular, the results of Section 6.3 do transfer *verbatim* and the bounds derived in Theorem 6.2 are valid exactly as stated there. However, the symbol $\pm$ allows us to compress the number of forbidden patterns so that the new instance is exponentially smaller. Thus, the polynomial time algorithm described above for normal sets could well not be polynomial

in the size of the compressed instance. We now prove that unless $P = NP$, there is no polynomial time algorithm to decide zero capacity when the symbol $\pm$ is allowed.

Theorem 6.5 *The problem of determining if the capacity of a set of forbidden patterns over $\{0,+,-,\pm\}$ is equal to zero is NP-hard.*

Proof. The proof proceeds by reduction from the Not-All-Equal 3SAT problem that is known to be NP-complete (see [44]). In the Not-All-Equal 3SAT problem, we are given m binary variables $x_1,\ldots,x_m$ and t clauses that each contain three literals (a literal can be a variable or its negation), and we search a truth assignment for the variables such that each clause has at least one true literal and one false literal.
Suppose that we are given a set of clauses. We construct a set of forbidden patterns D such that $\text{cap}(D) > 0$ if and only if the instance of Not-All-Equal 3SAT has a solution. The first part of D is given by:

$$\{(0\pm 0),(0\pm\pm 0),\ldots,(0\pm^{m-1}0)\}. \tag{6.7}$$

Words over $\{-,0,+\}$ that avoid these patterns are exactly those words for which any two consecutive zeros are either adjacent or have at least m symbols on $\{+,-\}$ between them. We use these m symbols as a way of encoding possible truth assignments for the variables (the first one is "+" if $x_1 = 1$, etc...).

We then add to D two patterns for every clause: they will force a sequence of m nonzero symbols to encode a satisfying assignement for the instance of Not-All-Equal 3SAT. These patterns are of length m and are entirely composed of symbols $\pm$, except for the positions corresponding to the three variables of the clause, which we set to $+$ if the clause contains the variable itself, or to $-$ if the clause contains the negation of the variable. We also add the opposite of this pattern; this last pattern is not necessary for the proof but preserves the symmetry and simplifies the construction.

For example, if the instance of Not-All-Equal 3SAT consists of the two clauses $(x_1,\bar{x}_3,x_4)$ and $(\bar{x}_2,x_4,x_5)$, the corresponding set D will be $D = \{(0\pm 0),(0\pm\pm 0),(0\pm\pm\pm 0),(0\pm\pm\pm\pm 0),(+\pm-+\pm),(-\pm+-\pm),(\pm-\pm++),(\pm+\pm--)\}$.
Such a set D has always a length polynomial in the number of clauses and the number of variables.
We now prove that there is a solution to the instance of Not-All-Equal 3SAT if and only if $\text{cap}(D) > 0$. First, suppose that there exists a satisfying truth assignment for x and denote it by $(\omega_1,\ldots,\omega_m) \in \{0,1\}^m$. Associated to any $k \geq 1$ we construct a code of length $k(m+1)$ containing 2^k words as follows:

$$C_{k(m+1)} = \{0\omega 0\omega 0\omega 0\cdots 0\omega 0\omega, 0\omega 0\omega 0\omega 0\cdots 0\omega 0\bar{\omega},$$

$$0\omega 0\omega 0\omega 0\cdots 0\bar{\omega}0\omega,\ldots,0\bar{\omega}0\bar{\omega}0\bar{\omega}0\cdots 0\bar{\omega}0\bar{\omega}\},$$

where $\omega = \omega_1\cdots\omega_m$.

Any difference between two words in this code is a word of the form

$$0z_1 0z_2 0\cdots 0z_k,$$

where for every $1 \leq i \leq k$, z_i is either a sequence of m 0's or a word of length m over $\{-,+\}$. Because ω satisfies the instance of Not-All-Equal 3SAT, these words avoid the set D constructed above. Moreover, the cardinality of $C_{k(m+1)}$ is 2^k and hence

$$\operatorname{cap}(D) \geq \lim_{k\to\infty} \log_2 2^{\frac{k}{k(m+1)}} = \frac{1}{m+1} > 0. \tag{6.8}$$

For the converse implication, assume now that $\operatorname{cap}(D) > 0$. The capacity is positive, and so one can find two words whose differences contain a 0 and a $+$. But then since this difference must avoid the first part of the forbidden pattern, for a code C large enough, there must exist two words in the code whose difference contains a word over $\{-,+\}$ of length m. But this sequence avoids also the second part of D, and thus it represents an acceptable solution to our instance of Not-All-Equal 3SAT.

Note that a similar proof can be given if we replace the symbol "$\pm$" in the statement of the theorem by a symbol that represents either $+$, $-$, or 0.

6.6 Extremal norms and computing the capacity

As we have seen in previous chapters, the existence of an extremal norm can simplify many problems related to the joint spectral radius: it allows for instance to apply the geometrical algorithm exposed in Section 2.3. Recall that an extremal norm is a norm $||\cdot||$ such that

$$\max_{A\in\Sigma} ||A|| = \rho(\Sigma).$$

It turns out that in the case of capacity computation, the matrices do in fact always possess an extremal norm:

Theorem 6.6 *For any set D of forbidden patterns the set $\Sigma(D)$ possesses an extremal norm.*

Proof. Corollary 6.2 implies that $\Sigma(D)$ is not defective. To see this, replace $\operatorname{cap}(D)$ by $\log_2 \rho$ in Corollary 6.2 and recall that δ_t is, by definition of the set Σ, the maximal norm of products of length $t-(m-1)$ of matrices taken in Σ. We have seen in Section 2.1 that the nondefectiveness implies the existence of an extremal norm.

The existence of an extremal norm for a set of matrices makes it possible to apply the geometric algorithm described in Section 2.3 for computing the capacity with a given relative accuracy.

The complexity of this algorithm is exponential with respect to m, as the one proposed in Section 6.3 that approximates the capacity by successive estimations of δ_t. The advantages of one algorithm over the other appear in numerical computation of the capacity. Moreover, in many cases the approximation of invariant bodies by polytopes can lead to the exact value of the joint spectral radius, as mentioned in Section

2.3. Let us illustrate this method by computing the exact values of the capacity for several codes. In Examples 6.3 and 6.4 we find the values of capacities that were approximated in [82]. Example 6.5 deals with a code with $m = 4$.

Example 6.3 *$cap(\{0++\}) = \log_2 \rho(A_0) = \log_2\left(\frac{\sqrt{5}+1}{2}\right) = 0.69424191\ldots$. The eigenvector is $v = (2, \sqrt{5}-1, 2, \sqrt{5}-1)^T$. The algorithm terminates after five steps, the polytope $P = P_5$ has 32 vertices.*

Example 6.4 *$cap(\{0+-\}) = \log_2 \rho(A_0) = \log_2\left(\frac{\sqrt{5}+1}{2}\right)$. The algorithm terminates after four steps, $v = (2, \sqrt{5}-1, \sqrt{5}-1, 2)^T$, $P = P_4$, the polytope has 40 vertices.*

Example 6.5 *$cap(\{+++-\}) = \log_2\left(\frac{\sqrt{3+2\sqrt{5}}+1}{2}\right) = \log_2 \sqrt{\rho(A_0A_1)} = 0.90\ldots$. The algorithm terminates after eleven steps, the polytope $P = P_{11}$ has 528 vertices.*

6.7 Conclusion

One way to compute the capacity of a set of forbidden patterns is to compute the joint spectral radius of a set of matrices. In practice, this leads to a number of difficulties: first, the size of the matrices is exponential in the size of the set of forbidden patterns. Second, their number can also be exponential in the size of the instance. Third, the joint spectral radius is in general NP-hard to compute.
We have shown here that, in spite of these discouraging results, the simpler problem of checking the positivity of the capacity of a set defined on $\{+,-,0\}$ is polynomially decidable. However the same problem becomes NP-hard when defined over the alphabet $\{+,-,0,\pm\}$, so that we see a threshold between polynomial time and exponential time feasibility. We have also provided bounds that allow faster computation of the capacity. Finally we have proved the existence of extremal norms for the sets of matrices arising in the capacity computation, which is the only "good news" that we see concerning the possible feasibility of the capacity computation. To the best of our knowledge the problem remains open for the moment:

Open question 9 *Is the capacity computation/approximation NP-hard?*

For instance, one has to keep in mind that the approach that consists in computing a joint spectral radius cannot lead to a polynomial algorithm because of the exponential size of the sets of matrices. Nevertheless, it is conjectured in [11] that the sets of matrices with binary entries, and, in particular, those constructed in order to compute a capacity do always possess the finiteness property:

Open question 10 *Do matrices that arise in the context of capacity computation satisfy the finiteness property?*

Numerical results in [82], [57], and in this chapter seem to support this conjecture, and moreover the length of the period seems to be very short: it seems to be of the order

of the size of the forbidden patterns, which would be surprising, because this length would be logarithmic in the size of the matrices.

We end this chapter by mentionning another question that has not been solved yet. We have seen that if the capacity is positive, one is able to exhibit an admissible word of the shape $0^m d 0^{m-1}$. This word has moreover a size which is polynomial in the size of D since it is represented by a path in the auxiliary graph constructed from the Aho-Corasick automaton. Now if we allow the use of "$\pm$" characters, since the problem can be translated in a classical instance D' with characters in $\{0,+,-\}$, a positive capacity also implies the existence of a certificate of the shape $0^m d 0^{m-1}$. But what about the length of this word? Since this length is only polynomial in the new instance D', we cannot conclude that there exists a certificate whose size is polynomial in the former instance. If this was the case, we would have that the problem with "$\pm$" characters would be in NP. This motivates our last open question:

Open question 11 *Is the problem of determining if the capacity of a set of forbidden patterns D over $\{0,+,-,\pm\}$ is equal to zero in NP? Is there, for any set $D \in \{0,+,-,\pm\}^*$, an admissible word of the shape $0^m d 0^{m-1}$ whose length is polynomial in the size of D?*

Chapter 7
Overlap-free words

Abstract In this chapter we present the notion of overlap-free words and show how the number u_n of overlap-free words of length n is ruled by joint spectral characteristics. We use these results to provide tight estimates on the asymptotic growth of u_n. We provide new algorithms to estimate the joint spectral subradius and the Lyapunov exponent, that appear to be very efficient in practice.

7.1 Introduction

Binary overlap-free words have been studied for more than a century[1]. These are words over the binary alphabet $A = \{a, b\}$ that do not contain factors of the form $xvxvx$, where $x \in A$ and $v \in A^*$ (A^* is the set of all words on the alphabet A)[2]. Such factors are called *overlaps*, because the word xvx is written twice, with the two instances of this word overlapping at the middle x.

Perhaps the simplest way to understand overlap-free words is the following: In combinatorics on words, a *square* is the repetition of twice the same word, as for instance the french word *bobo*. A *cube* is the repetition of three times the same word, like *bobobo*. Now, an overlap is any repetition that is more than a square. For instance, the word *baabaa* is overlap-free (it is a square), but the word *baabaab* is an overlap, because *baa* is repeated "more than twice" (one could say that it is repeated $7/3$ times). This word satisfies the definition of an overlap, since it can be written $xuxux$ with $x = b$ and $u = aa$. See [6] for a recent survey.

[1] The chapter presents research work that has been published in [63, 64].

[2] This chapter uses classical results from combinatorics on words. For a survey on this branch of theoretical computer science, we refer the reader to [76].

Thue [112, 113] proved in 1906 that there are infinitely many overlap-free words. Indeed, the well-known Thue-Morse sequence[3] is overlap-free, and so the set of its factors provides an infinite number of different overlap-free words. The asymptotics of the number u_n of such words of a given length n was analyzed in a number of subsequent contributions[4]. The number of factors of length n in the Thue-Morse sequence is proved in [23] to be larger or equal to $3n-3$, thus providing a linear lower bound on u_n:

$$u_n \geq Cn.$$

The next improvement was obtained by Restivo and Salemi [101]. By using a certain decomposition result, they showed that the number of overlap-free words grows at most polynomially:

$$u_n \leq Cn^r,$$

where $r = \log(15) \approx 3.906$. This bound has been sharpened successively by Kfoury [67], Kobayashi [68], and finally by Lepistö [73] to the value $r = 1.37$. One could then suspect that the sequence u_n grows linearly. However, Kobayashi proved that this is not the case [68]. By enumerating the subset of overlap-free words of length n that can be infinitely extended to the right he showed that $u_n \geq Cn^{1.155}$ and so we have

$$C_1 n^{1.155} \leq u_n \leq C_2 n^{1.37}.$$

Carpi showed that there is a finite automaton allowing to compute u_n (the sequence u_n is 2-regular [25]). In Figure 7.1(a) we show the values of the sequence u_n for $1 \leq n \leq 200$ and in Figure 7.1(b) we show the behavior of $\log u_n / \log n$ for larger values of n. One can see that the sequence u_n is not monotonic, but is globally increasing with n. Moreover the sequence does not appear to have a polynomial growth since the value $\log u_n / \log n$ does not seem to converge. In view of this, a natural question arises: is the sequence u_n asymptotically equivalent to n^r for some r ? Cassaigne proved in [26] that the answer is negative. He introduced the lower and the upper exponents of growth:

$$\begin{aligned} \alpha &= \sup\{r \mid \exists C > 0, u_n \geq Cn^r\}, \\ \beta &= \inf\{r \mid \exists C > 0, u_n \leq Cn^r\}, \end{aligned} \tag{7.1}$$

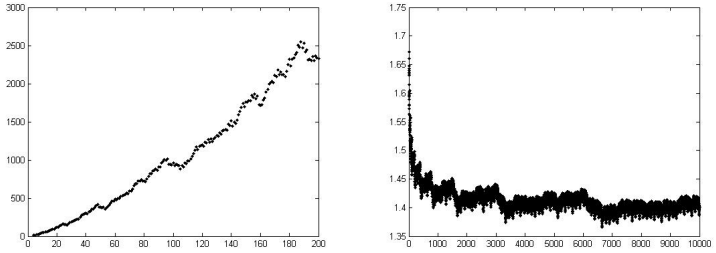
and showed that $\alpha < \beta$. Cassaigne made a real breakthrough in the study of overlap-free words by characterizing in a constructive way the whole set of overlap-free words. By improving the decomposition theorem of Restivo and Salemi he showed that the numbers u_n can be computed as sums of variables that are obtained by certain recurrence relations. These relations are explicitly given in the next section and all numeri-

[3] The Thue-Morse sequence is the infinite word obtained as the limit of $\theta^n(a)$ as $n \to \infty$ with $\theta(a) = ab,\ \theta(b) = ba$; see [26].

[4] The number of overlap-free words of length n is referenced in the On-Line Encyclopedia of Integer Sequences under the code A007777; see [107]. The sequence starts 1, 2, 4, 6, 10, 14, 20, 24, 30, 36, 44, 48, 60, 60, 62, 72,...

cal values can be found in Appendix A.1. As a result of this description, the number of overlap-free words of length n can be computed in logarithmic time. For the exponents of growth Cassaigne also obtained the following bounds: $\alpha < 1.276$ and $\beta > 1.332$. Thus, combining this with the earlier results described above, one has the following inequalities:

$$1.155 < \alpha < 1.276 \qquad \text{and} \qquad 1.332 < \beta < 1.37. \tag{7.2}$$



(a) (b)

Fig. 7.1 The values of u_n for $1 \le n \le 200$ (a) and $\log u_n / \log n$ for $1 \le n \le 10000$ (b).

In this chapter we develop a linear algebraic approach to study the asymptotic behavior of the number of overlap-free words of length n. Using the results of Cassaigne we show in Theorem 7.2 that u_n is asymptotically equivalent to the norm of a long product of two particular matrices A_0 and A_1 of dimension 20×20. This product corresponds to the binary expansion of the number $n-1$. Using this result we express the values of α and β by means of certain joint spectral characteristics of these matrices. We prove that $\alpha = \log_2 \check{\rho}(A_0, A_1)$ and $\beta = \log_2 \rho(A_0, A_1)$. In Section 7.3, we estimate these values and we obtain the following improved bounds for α and β:

$$1.2690 < \alpha < 1.2736 \qquad \text{and} \qquad 1.3322 < \beta < 1.3326. \tag{7.3}$$

Our estimates are, respectively, within 0.4% and 0.03% of the exact values. In addition, we show in Theorem 7.3 that the smallest and the largest rates of growth of u_n are effectively attained, and there exist positive constants C_1, C_2 such that $C_1 n^{\alpha} \le u_n \le C_2 n^{\beta}$ for all $n \in \mathbb{N}$.

Although the sequence u_n does not exhibit an asymptotic polynomial growth, we then show in Theorem 7.5 that for "almost all" values of n the rate of growth is actually equal to $\sigma = \log_2 \bar{\rho}(A_0, A_1)$, where $\bar{\rho}$ is the Lyapunov exponent of the matrices. For almost all values of n the number of overlap-free words does not grow as n^{α}, nor as n^{β}, but in an intermediary way, as n^{σ}. This means in particular that the value $\frac{\log u_n}{\log n}$ converges to σ as $n \to \infty$ along a subset of density 1. We obtain the following bounds

for the limit σ, which provides an estimate within 0.8% of the exact value:

$$1.3005 < \sigma < 1.3098.$$

These bounds clearly show that $\alpha < \sigma < \beta$.

To compute the exponents α and σ we introduce new efficient algorithms for estimating the joint spectral subradius $\check{\rho}$ and the Lyapunov exponent $\bar{\rho}$ of matrices. These algorithms are both of independent interest as they can be applied to arbitrary matrices.

Our linear algebraic approach not only allows us to improve the estimates of the asymptotics of the number of overlap-free words, but also clarifies some aspects of the nature of these words. For instance, we show that the "non purely overlap-free words" used in [26] to compute u_n are asymptotically negligible when considering the total number of overlap-free words.

The chapter is organized as follows. In the next section we formulate and prove the main theorems (except for Theorem 7.2, whose proof is quite long and technical). Then in Section 7.3 we present new algorithms for estimating the joint spectral subradius and the Lyapunov exponent of a given set of matrices. Applying them to those special matrices we obtain the estimates for α, β and σ. In the appendices we write explicit forms of the matrices and initial vectors used to compute u_n and present the results of our numerical algorithms.

7.2 The asymptotics of overlap-free words

To compute the number u_n of overlap-free words of length n we use several results from [26] that we summarize in the following theorem:

Theorem 7.1 *Let $F_0, F_1 \in \mathbb{R}^{30\times 30}$, $w, y_8, \ldots, y_{15} \in \mathbb{R}^{30}_+$ be as given in Appendix A.1. For $n \geq 16$, let y_n be the solution of the following recurrence equations*

$$\begin{aligned} y_{2n} &= F_0 y_n, \\ y_{2n+1} &= F_1 y_n. \end{aligned} \tag{7.4}$$

Then, for any $n \geq 9$, the number of overlap-free words of length n is equal to $w^T y_{n-1}$.

It follows from this result that the number u_n of overlap-free words of length $n \geq 16$ can be obtained by first computing the binary expansion $d_t \cdots d_1$ of $n-1$, i.e., $n-1 = \sum_{j=0}^{t-1} d_{j+1} 2^j$, and then computing

$$u_n = w^T F_{d_1} \cdots F_{d_{t-4}} y_m, \tag{7.5}$$

where $m = d_{t-3} + d_{t-2} 2 + d_{t-1} 2^2 + d_t 2^3$ (and $d_t = 1$). To arrive at the results summarized in Theorem 7.2, Cassaigne builds a system of recurrence equations allowing the computation of a vector U_n whose entries are the number of overlap-free words

of certain types (there are 16 different types). These recurrence equations also involve the recursive computation of a vector V_n that counts other words of length n, the so-called "single overlaps". The single overlap words are not overlap-free, but have to be computed, as they generate overlap-free words of larger lengths.

We now present the main result of this section which improves the above theorem in two directions. First we reduce the dimension of the matrices from 30 to 20, and second we prove that u_n is given asymptotically by the norm of a matrix product. The reduction of the dimension to 20 has a straightforward interpretation: when computing the asymptotic growth of the number of overlap-free words, one can neglect the number of "single overlaps" V_n defined by Cassaigne. We call the remaining words *purely overlap-free words,* as they can be entirely decomposed in a sequence of overlap-free words via Cassaigne's decomposition (see [26] for more details). In the following Theorem, the notation $f(n) \asymp g(n)$ means that there are two positive constants K_1, K_2 such that for all n, $K_1 f(n) < g(n) < K_2 f(n)$.

Theorem 7.2 *Let $A_0, A_1 \in \mathbb{R}_+^{20\times 20}$ be the matrices defined in Appendix A.1 (Equation (A.3)), let $\|\cdot\|$ be a matrix norm, and let $A(n) : \mathbb{N} \to \mathbb{R}_+^{20\times 20}$ be defined as $A(n) = A_{d_1} \cdots A_{d_t}$ with $d_t \ldots d_1$ the binary expansion of $n-1$. Then,*

$$u_n \asymp ||A(n)||. \tag{7.6}$$

Observe that the matrices F_0, F_1 in Theorem 7.1 are both nonnegative and hence possess a common invariant cone $K = \mathbb{R}_+^{30}$. We say that a cone K is *invariant* for a linear operator B if $BK \subset K$. All cones are assumed to be solid, convex, closed, and pointed. We start with the following simple result proved in [96].

Lemma 7.1 *For any cone $K \subset \mathbb{R}^d$, for any norm $|\cdot|$ in $\mathbb{R}^d$ and any matrix norm $\|\cdot\|$ there is a homogeneous continuous function $\gamma : K \to \mathbb{R}_+$ positive on* int$K$ *such that for any $x \in$ int$K$ and for any matrix B that leaves K invariant one has*

$$\gamma(x)\|B\| \cdot |x| \le |Bx| \le \frac{1}{\gamma(x)}\|B\| \cdot |x|.$$

Corollary 7.1 *Let two matrices A_0, A_1 possess an invariant cone $K \subset \mathbb{R}^d$. Then for any $x \in$ int$K$, with the notation $A(n)$ of Theorem 7.2, we have*

$$|A(n)x| \asymp \|A(n)\|.$$

In view of Corollary 7.1 and of Equation (7.5), Theorem 7.2 may seem obvious, at least if we consider the matrices F_i instead of A_i. One can however not directly apply Lemma 7.1 and Corollary 7.1 to the matrices A_0, A_1 or to the matrices F_0, F_1 because the vector corresponding to x is not in the interior of the positive orthant, which is an invariant cone of these matrices.

To prove Theorem 7.2 one has to first construct a common invariant cone K for the matrices A_0, A_1. This cone has to contain all the vectors z_n, $n \in \mathbb{N}$ (the restriction

of y_n to $\mathbb{R}^{20}$, see Theorem 7.1) in its interior, to enable us to apply Lemma 7.1 and Corollary 7.1.
Then, invoking Lemma 7.1 and Corollary 7.1 it is possible to show that the products $F(n) = F_{d_1} \cdots F_{d_k}$ are asymptotically equivalent to their corresponding product $A(n) = A_{d_1} \cdots A_{d_k}$.
Finally one shows that $\|A_{d_1} \cdots A_{d_k}\|$ is equivalent to $\|A_{d_1} \cdots A_{d_{k-4}}\|$.
Putting all this together, one proves Theorem 7.2. Details of the proof can be found in [63].

Theorem 7.2 allows us to express the rates of growth of the sequence u_n in terms of norms of products of the matrices A_0, A_1 and then to use joint spectral characteristics of these matrices to estimate the rates of growth. More explicitly, Theorem 7.2 yields the following corollary:

Corollary 7.2 *Let $A_0, A_1 \in \mathbb{R}_+^{20\times 20}$ be the matrices defined in Appendix A and let $A(n) : \mathbb{N} \to \mathbb{R}_+^{20\times 20}$ be defined as $A(n) = A_{d_1} \cdots A_{d_k}$ with $d_k \ldots d_1$ the binary expansion of $n-1$. Then*

$$\frac{\log_2 u_n}{\log_2 n} \;-\; \log_2 \|A(n)\|^{1/k} \;\to\; 0 \qquad as \qquad n \to \infty. \tag{7.7}$$

Proof. Observe first that $\left(\frac{k}{\log_2 n} - 1\right) \frac{\log_2 u_n}{k} \to 0$ as $n \to \infty$. Indeed, the first factor tends to zero, and the second one is uniformly bounded, because, as we have seen, $u_n \le Cn^r$. Hence

$$\begin{aligned}
&\lim_{n\to\infty} \left(\frac{\log_2 u_n}{\log_2 n} - \frac{\log_2 \|A_{d_1} \cdots A_{d_k}\|}{k} \right) = \\
&\lim_{n\to\infty} \left(\frac{\log_2 u_n - \log_2 \|A_{d_1} \cdots A_{d_k}\|}{k} + \left(\frac{k}{\log_2 n} - 1\right) \frac{\log_2 u_n}{k} \right) = \\
&\lim_{n\to\infty} \left(\frac{\log_2 u_n - \log_2 \|A_{d_1} \cdots A_{d_k}\|}{k} \right) = \lim_{n\to\infty} \frac{\log_2 \left(u_n \cdot \|A_{d_1} \cdots A_{d_k}\|^{-1} \right)}{k},
\end{aligned}$$

and by Theorem 7.2 the value $\log_2 \left(u_n \cdot \|A_{d_1} \cdots A_{d_k}\|^{-1} \right)$ is bounded uniformly over $n \in \mathbb{N}$.

We first analyze the smallest and the largest exponents of growth α and β defined in Equation (7.1).

Theorem 7.3 *For $t \ge 1$, let $\alpha_t = \min\limits_{2^{t-1}<n\le 2^t} \frac{\log u_n}{\log n}$ and $\beta_t = \max\limits_{2^{t-1}<n\le 2^t} \frac{\log u_n}{\log n}$. Then*

$$\alpha = \lim_{t\to\infty} \alpha_t = \log_2 \check{\rho}(A_0, A_1) \quad and \quad \beta = \lim_{t\to\infty} \beta_t = \log_2 \rho(A_0, A_1), \tag{7.8}$$

where the matrices A_0, A_1 are defined in Appendix A.1. Moreover, there are positive constants C_1, C_2 such that

$$C_1 \le \min_{2^{t-1}<n\le 2^t} u_n n^{-\alpha} \quad and \quad C_1 \le \max_{2^{t-1}<n\le 2^t} u_n n^{-\beta} \le C_2 \tag{7.9}$$

for all $t \in \mathbb{N}$.

Proof. The equalities in Equation (7.8) follow immediately from Corollary 7.2 and the definitions.
The lower bounds in Equation (7.9) are a consequence of Theorem 7.2 and the fact that $\hat{\rho}_t \geq \rho^t$ and $\check{\rho}_t \geq \check{\rho}^t$ always hold (see Chapter 1).
For the upper bound in Equation (7.9) we note that the matrices A_0, A_1 have no common invariant subspaces among the coordinate planes (to see this observe, for instance, that $(A_0 + A_1)^5$ has no zero entry). As shown in Chapter 3, this proves that the set is nondefective, that is,

$$\hat{\rho}_t \leq C_2 \rho^t.$$

Corollary 7.3 *There are positive constants* C_1, C_2 *such that*

$$C_1 n^{\alpha} \leq u_n \leq C_2 n^{\beta}, \ n \in \mathbb{N}.$$

In the next section we show that $\alpha < \beta$. In particular, the sequence u_n does not have a constant rate of growth, and the value $\frac{\log u_n}{\log n}$ does not converge as $n \to \infty$. This was already noted by Cassaigne in [26]. Nevertheless, it appears that the value $\frac{\log u_n}{\log n}$ actually has a limit as $n \to \infty$, not along all the natural numbers $n \in \mathbb{N}$, but along a subsequence of $\mathbb{N}$ of density 1. A subset $\mathscr{A} \subset \mathbb{N}$ is said to have density 1 if $\frac{1}{n}\mathrm{Card}\{r \leq n, \ r \in \mathscr{A}\} \to 1$ as $n \to \infty$. In other terms, the sequence converges with probability 1. The limit, which differs from both α and β can be expressed by the so-called Lyapunov exponent $\bar{\rho}$ of the matrices A_0, A_1. To show this we apply the following result proved by Oseledets in 1968. For the sake of simplicity we formulate it for two matrices, although it can be easily generalized to any finite set of matrices.

Theorem 7.4 *[88] Let* A_0, A_1 *be arbitrary matrices and* $d_1, d_2, \ldots$ *be a sequence of independent random variables that take values* 0 *and* 1 *with equal probabilities* $1/2$. *Then the value* $\|A_{d_1} \cdots A_{d_t}\|^{1/t}$ *converges to some number* $\bar{\rho}$ *with probability* 1. *This means that for any* $\varepsilon > 0$ *we have* $P\big(\big|\|A_{d_1} \cdots A_{d_t}\|^{1/t} - \bar{\rho}\big| > \varepsilon\big) \to 0$ *as* $t \to \infty$.

The limit $\bar{\rho}$ in Theorem 7.4 is called the *Lyapunov exponent* of the set $\{A_0, A_1\}$. This value is given by the following formula:

$$\bar{\rho}(A_0, A_1) = \lim_{t \to \infty} \Big(\prod_{d_1, \ldots, d_t} \|A_{d_1} \cdots A_{d_t}\|^{1/t} \Big)^{1/2^t} \tag{7.10}$$

(for a proof see, for instance, [97]). To understand what this gives for the asymptotics of our sequence u_n we introduce some further notation. Let $\mathscr{P}$ be some property of natural numbers. For a given $t \in \mathbb{N}$ we denote

$$P_t(\mathscr{P}) = 2^{-(t-1)}\mathrm{Card}\{n \in \{2^{t-1}+1, \ldots, 2^t\} : n \,\text{satisfies}\, \mathscr{P}\}.$$

Thus, P_t is the probability that the integer n uniformly distributed on the set

$$\{2^{t-1}+1,\dots,2^t\}$$

satisfies $\mathscr{P}$. Combining Corollary 7.2 and Theorem 7.4 we obtain

Theorem 7.5 *There is a number σ such that for any $\varepsilon>0$ we have*

$$P_t\Big(\Big|\frac{\log u_n}{\log n}-\sigma\Big|>\varepsilon\Big)\to 0 \qquad \text{as } t\to\infty.$$

Moreover, $\sigma=\log_2\bar{\rho}$, where $\bar{\rho}$ is the Lyapunov exponent of the matrices $\{A_0,A_1\}$ defined in Appendix A.1.

Thus, for almost all $n\in\mathbb{N}$ the number of overlap-free words u_n has the same exponent of growth $\sigma=\log_2\bar{\rho}$. If positive a and b are large enough and $a<b$, then for a number n taken randomly from the segment $[a,b]$ the value $\log u_n/\log n$ is close to σ. We say that a sequence f_n converges to a number f along a set of density 1 if there is a set $\mathscr{A}\subset\mathbb{N}$ of density 1 such that $\lim\limits_{n\to\infty,n\in\mathscr{A}} f_n=f$. Theorem 7.5 yields

Corollary 7.4 *The value $\frac{\log u_n}{\log n}$ converges to σ along a set of density* 1.

Proof. Let us define a sequence $\{k_j\}$ inductively: $k_1=1$, and for each $j\geq 2$ let k_j be the smallest integer such that $k_j>k_{j-1}$ and

$$P_k\Big(\Big|\frac{\log u_n}{\log n}-\sigma\Big|>\frac{1}{j}\Big)\leq\frac{1}{j} \qquad \text{for all} \quad k\geq k_j.$$

By Theorem 7.5 the values k_j are well-defined for all j. Let a set $\mathscr{A}$ consist of numbers n, for which $\Big|\frac{\log u_n}{\log n}-\sigma\Big|\leq\frac{1}{j}$, where j is the largest integer such that $n\geq 2^{k_j-1}$. Clearly, $\frac{\log u_n}{\log n}\to\sigma$ as $n\to\infty$ along $\mathscr{A}$. If, as usual, $2^{k-1}\leq n<2^k$, then the total number of integers $r\leq n$ that do not belong to $\mathscr{A}$ is less than

$$\frac{2^k}{j}+\frac{2^{k_j}}{j-1}+\dots+\frac{2^{k_2}}{1}\leq\sum_{s=1}^{j}\frac{2^{k-j+s}}{s}=2^{k-j}\sum_{s=1}^{j}\frac{2^s}{s}.$$

Observe that $\sum\limits_{s=1}^{j}\frac{2^s}{s}\leq\frac{3\cdot 2^j}{j}$, hence the number of integers $r\leq n$ that do not belong to $\mathscr{A}$ is less than $\frac{3\cdot 2^k}{j}\leq\frac{6n}{j}$, which tends to zero being divided by n as $n\to\infty$. Thus, $\mathscr{A}$ has density 1.

7.3 Estimation of the exponents

Theorems 7.2 and 7.5 reduce the problem of estimating the exponents of growth of u_n to computing joint spectral characteristics of the matrices A_0 and A_1. In order to

estimate the joint spectral radius we use a modified version of the "ellipsoidal norm algorithm" presented in Chapter 2. For the joint spectral subradius and for the Lyapunov exponent we present new algorithms, which seem to be relatively efficient, at least for nonnegative matrices. The results we obtain can be summarized in the following theorem:

Theorem 7.6

$$\begin{aligned} 1.2690 < \alpha < 1.2736, \\ 1.3322 < \beta < 1.3326, \\ 1.3005 < \sigma < 1.3098. \end{aligned} \tag{7.11}$$

In this section we also make (and give arguments for) the following conjecture:

Conjecture 7.1

$$\beta = \log_2 \sqrt{\rho(A_0 A_1)} = 1.3322\ldots.$$

7.3.1 Estimation of β and the joint spectral radius

By Theorem 7.3 to estimate the exponent β one needs to estimate the joint spectral radius of the set $\{A_0, A_1\}$. A lower bound for ρ can be obtained by applying the three members inequality (1.6). Taking $t = 2$ and $d_1 = 0, d_2 = 1$ we get

$$\rho \quad \geq \quad \left[\rho(A_0 A_1)\right]^{1/2} = 2.5179\ldots, \tag{7.12}$$

and so $\beta > \log_2 2.5179 > 1.3322$ (this lower bound was already found in [26]).
One could also try to derive an upper bound on ρ with the three members inequality, that is:

$$\rho \quad \leq \quad \max_{d_1,\ldots,d_t \in \{1,\ldots,m\}} \|A_{d_1} \cdots A_{d_t}\|^{1/t}. \tag{7.13}$$

This, at least theoretically, gives arbitrarily sharp estimates for ρ. However, in our case, due to the size of the matrices A_0, A_1, this method leads to computations that are too expensive even for relatively small values of t. As we have seen in Chapter 2, faster convergence can be achieved by finding an appropriate norm. The ellipsoidal norms are good candidates, because the optimum among these norms can be found via a simple SDP program. In Appendix A.2 we give an ellipsoidal norm such that each matrix in Σ^{14} has a norm smaller than 2.5186^{14}. This implies that $\rho \leq 2.5186$, which gives $\beta < 1.3326$. Combining this with the inequality $\beta > 1.3322$ we complete the proof of the bounds for β in Theorem 7.6.
We have not been able to improve the lower bound of Equation (7.12). However, the upper bound we obtain is very close to this lower bound, and the upper bounds obtained with an ellipsoidal norm for Σ^t get closer and closer to this value when t increases. Moreover, as mentioned in Chapter 4, it has already been observed that for

many sets of matrices for which the joint spectral radius is known exactly, and in particular matrices with nonnegative integer entries, the finiteness property holds, i.e., there is a product $A \in \Sigma^t$ such that $\rho = \rho(A)^{1/t}$ [61]. For these reasons, we conjecture that the exponent β is actually equal to the lower bound, that is,

$$\beta = \sqrt{\rho(A_0 A_1)}.$$

7.3.2 Estimation of α and the joint spectral subradius

An upper bound for $\check{\rho}(A_0, A_1)$ can be obtained using the three members inequality for $t = 1$ and $d_1 = 0$. We have

$$\alpha = \log_2(\check{\rho}) \leq \log_2(\rho(A_0)) = 1.276... \tag{7.14}$$

This bound for α was first derived in [26]. It is however not optimal. Taking the product $A_1^{10} A_0$ (i.e., $t = 11$), we get a better estimate:

$$\alpha \leq \log_2 \left[(\rho(A_1^{10} A_0)^{1/11} \right] = 1.2735... \tag{7.15}$$

One can verify numerically that this product gives the best possible upper bound among all the matrix products of length $t \leq 14$.

We now estimate α from below. As we know, the problem of approximating the joint spectral subradius is NP-hard [17] and to the best of our knowledge, no algorithm is known to compute this quantity. Here we propose two new algorithms. We first consider nonnegative matrices. As proved in Chapter 1, for any t and any set of matrices Σ, we have $\check{\rho}(\Sigma^t) = \check{\rho}^t(\Sigma)$. Without loss of generality it can be assumed that the matrices of the set Σ do not have a common zero column. Otherwise, by suppressing this column and the corresponding row we obtain a set of matrices of smaller dimension with the same joint spectral subradius. The vector of ones is denoted by $\mathbf{1}$.

Theorem 7.7 *Let Σ be a set of nonnegative matrices that do not have any common zero column. If for some $r \in \mathbb{R}^+, s \leq t \in \mathbb{N}$, there exists $x \in \mathbb{R}^d$ satisfying the following system of linear inequalities*

$$\begin{array}{lll} B(Ax - rx) & \geq 0, & \forall B \in \Sigma^s, A \in \Sigma^t, \\ x & \geq 0, & (x, \mathbf{1}) = 1, \end{array} \tag{7.16}$$

then $\check{\rho}(\Sigma) \geq r^{1/t}$.

Proof. Let x be a solution of (7.16). Let us consider a product of matrices $A_k \dots A_1 \in \Sigma^{kt} : A_i \in \Sigma^t$. We show by induction on k that $A_k \dots A_1 x \geq r^{k-1} A_k x$: For $k = 2$, we

have $A_2(A_1x - rx) = CB(A_1x - rx) \geq 0$, with $B \in \Sigma^s, C \in \Sigma^{t-s}$. For $k > 2$ we have $A_k \ldots A_1 x = A_k A_{k-1} \ldots A_1 x \geq r^{k-2} A_k A_{k-1} x \geq r^{k-1} A_k x$. In the last inequality the case for $k = 2$ was reused.

Hence,

$$||A_k \ldots A_1|| = \mathbf{1}^T A_k \ldots A_1 \mathbf{1} \geq r^{k-1} \mathbf{1}^T A_k x \geq K r^k,$$

where $K = (\min_k \mathbf{1}^T A_k x)/r > 0$. The last inequality holds because $A_k x = 0$, together with the first inequality in (7.16), imply that $-rBx = 0$ for all $B \in \Sigma^s$, which implies that all $B \in \Sigma^s$ have a common zero column. This is in contradiction with our assumption because the matrices in Σ^s share a common zero column if and only if the matrices in Σ do.

Clearly, the size of the instance of the linear program 7.16 grows exponentially with t and s. We were able to find a solution to the linear programming problem (7.16) with $r = 2.41^{16}$, $t = 16, s = 6$. Hence we get the following lower bound: $\alpha \geq \log_2 r/16 > 1.2690$. The corresponding vector x is given in Appendix A.3. This completes the proof of Theorem 7.6.

Theorem 7.7 handles nonnegative matrices, and we propose now a way to generalize this result to arbitrary real matrices. For this purpose, we use the semidefinite lifting presented in Chapter 2, and we consider the set of linear operators acting on the cone of positive semidefinite symmetric matrices S as $S \to A_i^T S A_i$. We know that the joint spectral subradius of this new set of linear operators is equal to $\check{\rho}(\Sigma)^2$. We use the notation $A \succeq B$ to denote that the matrix $A - B$ is positive semidefinite. Recall that $A \succeq 0 \Leftrightarrow \forall y, y^T A y \geq 0$.

Theorem 7.8 *Let Σ be a set of matrices in $\mathbb{R}^{d \times d}$ and $s \leq t \in \mathbb{N}$. Suppose that there are $r > 0$ and a symmetric matrix $S \succeq 0$ such that*

$$\begin{aligned} &B^T(A^T S A - rS)B \succeq 0 \ \forall A \in \Sigma^t, B \in \Sigma^s, \\ &S \succ 0, \end{aligned} \tag{7.17}$$

then $\check{\rho}(\Sigma) \geq r^{1/2t}$.

Proof. The proof is formally similar to the previous one. Let S be a solution of (7.17). We denote by M_k the product $A_1 \ldots A_k$, where $A_i \in \Sigma^t$. It is easy to show by induction that $M_k^T S M_k \succeq r^{k-1}(A_k^T S A_k)$. This is obvious for $k = 2$ for similar reasons as in the previous theorem, and for $k > 2$, if, by induction,

$$\forall y, \quad y^T M_{k-1}^T S M_{k-1} y \geq r^{k-2} y^T A_{k-1}^T S A_{k-1} y,$$

then, with $y = A_k x$, for all x,

$$x^T M_k^T S M_k x \geq r^{k-2} x^T A_k^T A_{k-1}^T S A_{k-1} A_k x \geq r^{k-1} x^T A_k^T S A_k x.$$

Thus,

$$\sup\left\{\frac{x^T M_k^T S M_k x}{x^T S x}\right\} \geq r^{k-1} \sup\left\{\frac{x^T A_k^T S A_k x}{x^T S x}\right\}.$$

Finally, $||M_k||_S \geq r^{k/2} C$, where C is a constant.

For a given $r > 0$ the existence of a solution S can be established by solving the semidefinite programming problem (7.17), and the optimal r can be found by bisection in logarithmic time.

7.3.3 Estimation of σ and the Lyapunov exponent

The exponent of the average growth σ is obviously between α and β, so $1.2690 < \sigma < 1.3326$. To get better bounds we need to estimate the Lyapunov exponent $\bar{\rho}$ of the matrices A_0, A_1. The first upper bound can be given by the so-called 1-radius ρ_1:

$$\rho_1 = \lim_{t\to\infty}\Big(2^{-t}\sum_{d_1,\ldots,d_t} \|A_{d_1}\cdots A_{d_t}\|\Big)^{1/t}.$$

For matrices with a common invariant cone we have $\rho_1 = \frac{1}{2}\rho(A_0+A_1)$ [96]. Therefore, in our case $\rho_1 = \frac{1}{2}\rho(A_0+A_1) = 2.479....$ This exponent was first computed in [26], where it was shown that the value $\sum_{j=0}^{n-1} u_j$ is asymptotically equivalent to n^η, where $\eta = 1+\log_2\rho_1 = 2.310....$ It follows immediately from the inequality between the arithmetic mean and the geometric mean that $\bar{\rho} \leq \rho_1$. Thus, $\sigma \leq \eta$. In fact, as we show below, σ is strictly smaller than η. We are not aware of any approximation algorithm for the Lyapunov exponent, except by application of Definition (7.10). It follows from submultiplicativity of the norm that for any t the value $r_t = \Big(\prod_{d_1,\ldots,d_t} \|A_{d_1}\cdots A_{d_t}\|\Big)^{\frac{1}{t2^t}}$ gives an upper bound for $\bar{\rho}$, that is $\bar{\rho} \leq r_t$ for any $t \in \mathbb{N}$. Since $r_t \to \bar{\rho}$ as $t \to \infty$, we see that this estimate can be arbitrarily sharp for large t. But for the dimension 20 this leads quickly to prohibitive numerical computations. For example, for the norm $\|\cdot\|_1$ we have $r_{20} = 2.4865$, which is even larger than ρ_1. In order to obtain a better bound for $\bar{\rho}$ we state the following results. For any t and $x \in \mathbb{R}^d$ we denote $p_t(x) = \Big(\prod_{d_1,\ldots,d_t} |A_{d_1}\cdots A_{d_t}x|\Big)^{\frac{1}{2^t}}$ and $m_t = \sup_{x\geq 0,|x|=1} p_t(x)$. In general, this expression is hard to evaluate, but in the following we will use a particular norm for which m_t is easy to handle.

Proposition 7.1 *Let A_0, A_1 be nonnegative matrices in $\mathbb{R}^d$. Then for any norm $|\cdot|$ and for any $t \geq 1$ we have $\bar{\rho} \leq (m_t)^{1/t}$.*

Proof. By Corollary 7.1, for $x > 0$ we have $r_n \asymp [p_n(x)]^{1/n}$, and consequently

$$\lim_{t\to\infty}\left[p_{tk}(x)\right]^{1/tk}\to\bar{\rho}$$

as $t\to\infty$. On the other hand, $p_{k+n}(x)\le m_k p_n(x)$ for any $x\ge 0$ and for any $n,k\in\mathbb{N}$, therefore $p_{tk}(x)\le(m_k)^t$. Thus, $\bar{\rho}\le(m_k)^{1/k}$.

Proposition 7.2 *Let A_0,A_1 be nonnegative matrices in $\mathbb{R}^d$ that do not have common invariant subspaces among the coordinate planes. If $\check{\rho}<\rho$, then $\bar{\rho}<\rho_1$.*

Proof. Let v_* be the eigenvector of the matrix $\frac{1}{2}\left(A_0^T+A_1^T\right)$ corresponding to its largest eigenvalue ρ_1. Since the matrices have no common invariant coordinate planes, it follows from the Perron-Frobenius theorem that $v_*>0$. Consider the norm $|x|=(x,v_*)$ on $\mathbb{R}^d_+$. Take some $t\ge 1$ and $y\in\mathbb{R}^d_+, |y|=(y,v_*)=1$, such that $p_t(y)=m_t$. We have

$$m_t=p_t(y)\le 2^{-t}\sum_{d_1,\dots,d_t}|A_{d_1}\cdots A_{d_t}y|=\ 2^{-t}\sum_{d_1,\dots,d_t}\left(A_{d_1}\cdots A_{d_t}y,v_*\right)$$

$$=\ \left(y,2^{-t}\left(A_0^T+A_1^T\right)^t v_*\right)=\ \rho_1^t\left(y,v_*\right)=\rho_1^t.$$

Thus, $m_t\le\rho_1^t$, and the equality is possible only if all 2^t values $|A_{d_1}\cdots A_{d_t}y|$ are equal. Since $\check{\rho}<\rho$, there must be a t such that the inequality is strict. Thus, $m_t<\rho_1^t$ for some t, and by Proposition 7.1 we have $\bar{\rho}\le(m_t)^{1/t}<\rho_1$.

We are now able to estimate $\bar{\rho}$ for the matrices A_0,A_1. For the norm $|x|=(x,v_*)$ used in the proof of Proposition 7.2 the value $-\frac{1}{t}\log_2 m_t$ can be found as the solution of the following convex minimization problem with linear constraints:

$$\begin{array}{ll}\min & -\frac{1}{t2^t\ln 2}\displaystyle\sum_{d_1,\dots,d_t\in\{0,1\}}\ln\left(x,A_{d_1}^T\cdots A_{d_t}^T v_*\right)\\ \text{s.t.} & x\ge 0,\quad (x,v_*)=1.\end{array}\tag{7.18}$$

The optimal value of this optimization problem is equal to $-(1/t)\log_2 m_t$, which gives an upper bound for $\sigma=\log_2\bar{\rho}$ (Proposition 7.1). Solving this problem for $t=12$ we obtain $\sigma\le 1.3098$. We finally provide a theorem that allows us to derive a lower bound on σ. The idea is identical to the one used in Theorem 7.7, but transposed to the Lyapunov exponent.

Theorem 7.9 *Let Σ be a set of m nonnegative matrices that do not have any common zero column. If for some $s\le t\in\mathbb{N}$, $r_i\in\mathbb{R}_+ : 0\le i<m^t$, there exists $x\in\mathbb{R}^d_+$ satisfying the following system of linear inequalities*

$$\begin{array}{lll}B(A_ix-r_ix)\ge 0, & \forall B\in\Sigma^s, A_i\in\Sigma^t,\\ x\ge 0, & (x,\mathbf{1})=1,\end{array}\tag{7.19}$$

then $\bar{\rho}(\Sigma)\ge\prod_i r_i^{1/(tm^t)}$.

The proof is similar to the proof of Theorem 7.7 and is left to the reader. Also, a similar theorem can be stated for general matrices (with negative entries), but involving linear matrix inequalities. Due to the number of different variables r_i, one cannot hope to find the optimal x with SDP and bisection techniques. However, by using the vector x computed for approximating the joint spectral subradius (given in Appendix A.3), with the values $s = 8, t = 16$ for the parameters, one gets a good lower bound for $\sigma : \sigma \geq 1.3005$.

7.4 Conclusion

The goal of this chapter is to precisely characterize the asymptotic rate of growth of the number of overlap-free words. Based on Cassaigne's description of these words with products of matrices, we first prove that these matrices can be simplified, by decreasing the state space dimension from 30 to 20. This improvement is not only useful for numerical computations, but allows to characterize the overlap-free words that "count" for the asymptotics: we call these words *purely overlap-free,* as they can be expressed iteratively as the image of shorter purely overlap free words.
We have then proved that the lower and upper exponents α and β defined by Cassaigne are effectively reached for an infinite number of lengths, and we have characterized them respectively as the logarithms of the *joint spectral subradius* and the *joint spectral radius* of the simplified matrices that we constructed. This characterization, combined with new algorithms that we propose to approximate the joint spectral subradius, allow us to compute them within 0.4%. The algorithms we propose can of course be used to reach any degree of accuracy for β (this seems also to be the case for α and σ, but no theoretical result is known for the approximation of these quantities). The computational results we report in this chapter have all been obtained in a few minutes of computation time on a standard PC desktop and can therefore easily be improved.
Finally we have shown that for almost all values of n, the number of overlap-free words of length n does not grow as n^{α}, nor as n^{β}, but in an intermediary way as n^{σ}, and we have provided sharp bounds for this value of σ.
This work opens obvious questions: Can joint spectral characteristics be used to describe the rate of growth of other languages, such as for instance the more general repetition-free languages ? The generalization does not seem to be straightforward for several reasons: first, the somewhat technical proofs of the links between u_n and the norm of a corresponding matrix product take into account the very structure of these particular matrices, and second, it is known that a bifurcation occurs for the growth of repetition-free words: for some members of this class of languages the growth is polynomial, as for overlap-free words, but for some others the growth is exponential, as shown by Karhumaki and Shallit [66]. See [10] for more on repetition-free words and joint spectral characteristics.

Chapter 8
Trackable graphs

Abstract In this chapter we present the notion of trackable graph. We show how results presented in this monograph allow to efficiently recognize trackable graphs.

Imagine you are responsible for a network on which an agent is moving. The network can be modeled as a (directed) graph, and at each step the agent choses a node among all neighbors of its current node, and jumps to it. A number of recent contributions deal with the problem of "tracking" such an agent, that is, in some sense localize the agent on the network [24, 30, 87, 119, 120]. This network is endowed with sensors that give you information about the node in which the agent is. In practical applications however, the information is rarely sufficient to determine uniquely the current node of the agent: for instance, the network can be subject to noise, or two nodes that are too close can be activated together. Also, the sensors data can be transmitted in real time through a channel that only allows you to receive at each step a limited information about the sensors activations. Clearly, in general, the longer the experience lasts, the more trajectories will be possible. How to compute the set of all possible trajectories, given a sequence of observations? What are the possible growths of the number of trajectories when the observation length increases? How to determine the worst growth for a particular network? In Section 8.1 we formalize this problem and present the notion of *trackable graphs* recently introduced by Crespi et al. [30], and we give practical motivations for it. In Section 8.2 we answer to the above questions. We then conclude and raise some possible future work.

8.1 What is a trackable graph?

Initially motivated by tracking vehicles in noisy sensor networks, the concept of *trackable network* has recently been introduced [30] in the framework of Hidden Markov Models (HMM) (see [42, 100] for a survey on HMM's). We introduce here trackable graphs[1] in a self-contained and general framework, in terms of an underlying directed graph with colors on edges, but the mathematical reality behind these two definitions is exactly the same.

Let $G = (V,E)$ be a graph and C a set of colors. To every edge $e \in E$ we associate one (or more) color from C. A word w on C is the concatenation $w_0 \dots w_T$ of symbols taken from C; the length $|w|$ of w is the number of its symbols. A subword $w_{[i,j]} : 1 \leq i \leq j \leq |w|$ of $w = w_1 \dots w_T$ is the concatenation of the symbols $w_i \dots w_j$. We say that a path is allowed by a word w if for all i the ith edge of the path has color w_i. Finally, for a word w and a set $S \subset V$, we denote by $\mathscr{T}_w(S)$ the set of paths allowed by w beginning in a node in S; $\mathscr{T}_w(V)$ is the set of paths in G with the color sequence w. Since we are interested in the worst case, we introduce the *complexity function* $N(t)$ that counts the maximal number of trajectories compatible with an observation of length t:

$$N(t) : \mathbb{N} \to \mathbb{N} \triangleq \max \{ |\mathscr{T}_w(V)| : |w| = t \}.$$

We will say that a graph is *trackable* if its complexity function grows at most polynomially:

Definition 8.1 *A graph is* trackable *if there exists a polynomial $p(t)$ such that for any color sequence of length T, the number of possible trajectories compatible with this observation is bounded by $p(T)$.*

Figure 8.1 presents two similar graphs. The first one (a) is trackable, because the worst possible observation is $DDD\dots$ for which the number of compatible paths is $N_a(t) \approx t$. The second graph (b) is not trackable, because a possible observation is $w = DSDSDS\dots$ for which the number of possible trajectories is asymptotically equal to $2^{\lfloor \frac{t}{2} \rfloor}$.

In the seminal article [30], the colors are put on the nodes, rather than on the edges. This latter situation is a particular case of the one considered here, since this is equivalent to attribute the same color to each edge pointing to a same node. This fact is illustrated in Figure 8.2: the first graph (a) has its colors on the nodes, but one could analyze dynamics on this graph by constructing the edge-colored graph (b). For any path on the "actual" graph (a) leading to the color sequence w, the same path on the graph (b) leads to the color sequence $w_{[1,|w|]}$. So for all $t > 0$ the complexity functions $N_a(t)$, $N_b(t)$ of the graphs (a) and (b) satisfy $N_a(t+1) \leq N_b(t) \leq mN_a(t+1)$, with m

[1] The property of being trackable is a property of directed graphs that have their edges colored. For simplicity, we will talk in the sequel about trackable *graphs* rather than trackable *edge-colored directed graphs*.

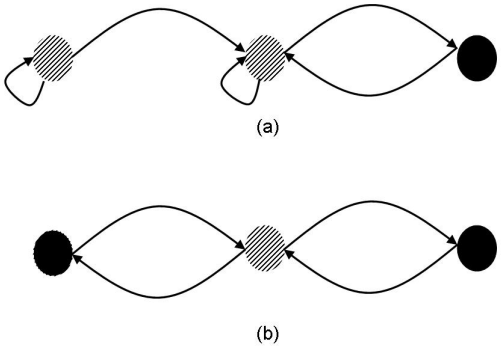


Fig. 8.1 Two node-colored graphs. The "colors" are dashed (D) or solid (S). The graph (a) is trackable but (b) is not.

the number of colors in the graph. So (a) is trackable if and only if (b) is, and all analysis results presented here are valid for node-colored graphs. Note however that this is only valid for analysis purposes. In Section 8.3 we will briefly address the design question. For this sort of questions, the node-colored case and the edge-colored case are not equivalent, since the set of feasible solutions is not the same. Indeed, when designing a node-colored graph, if one puts colors on edges in the corresponding edge-colored graph, he is restricted in that all edges going to a same node must have the same color.

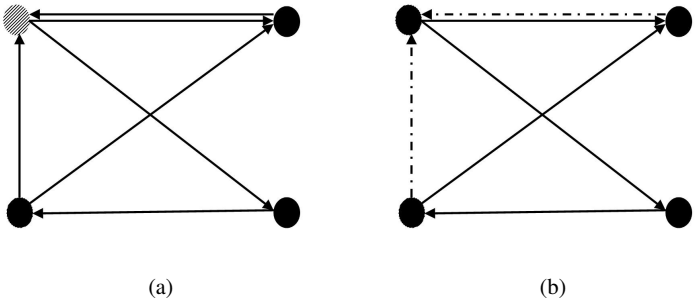


Fig. 8.2 A node-colored graphs (a) and the equivalent edge-colored graph (b).

8.2 How to recognize a trackable graph?

In this section, we consider two algorithmic problems. The first problem is that of counting the possible paths in a graph for a given sequence of color observations. That problem is easy and we describe a simple solution to it. The second problem is that of deciding trackability.

Let us consider the first problem: we are given a color sequence and we would like to count the paths that are compatible with the observed sequence of colors. A simple algebraic solution is as follows. For every color c, there is an associated graph G_c for which we can construct the corresponding adjacency matrix A_c. This graph is simply the graph on the initial set of vertices, but keeping only the edges colored in c. To a color sequence $w = w_1, \ldots, w_{|w|}$ we then associate A_w, the corresponding product of matrices $A_w = A_{w_1} \ldots A_{w_{|w|}}$. It is easy to verify that the (i, j)th entry of A_w is equal to the number of paths from i to j allowed by w. The total number of compatible paths $|\mathscr{T}_w|$ is therefore obtained by taking the sum of all entries of the matrix A_w.

We now turn to the problem of recognizing trackable graphs. We have the following theorem:

Theorem 8.1 *[30] Let G be a colored graph and $\Sigma = \{A_c\}$ be the set of adjacency matrices corresponding to each color. G is trackable if and only if $\rho(\Sigma) \leq 1$.*

Proof. The proof essentially uses the fact that the number of paths compatible with the color sequence w is the sum of the entries of A_w. Moreover, since A_w is nonnegative, the sum of its entries is actually a norm:

$$||A_w||_1 = \sum_{i,j} (A_w)_{(i,j)}.$$

Now, applying the definition of the joint spectral radius:

$$\rho(\Sigma) = \lim_{t \to \infty} \max \{||A||_1^{1/t} : A \in \Sigma^t\}, \tag{8.1}$$

$$= \lim_{t \to \infty} \max \{\sum_{i,j} (A_w)_{(i,j)} : |w| = t\}^{(1/t)}, \tag{8.2}$$

$$= \lim_{t \to \infty} N(t)^{1/t}, \tag{8.3}$$

and this latter quantity is less or equal to one if and only if $N(t)$ grows less than exponentially.

We can now apply all the machinery of Chapter 3 to the adjacency matrices of a colored graph:

Theorem 8.2 *There is a polynomial time algorithm that recognizes trackable graphs. This algorithm uses at most $O(n^5)$ operations, where n is the number of nodes in the graph. Moreover, only the following cases are possible:*

- *the length of the paths is bounded, i.e., there exists a T such that for all $t \geq T$, $N(t) = 0$,*
- *the function $N(t)$ is bounded,*
- *the function $N(t)$ grows polynomially with a degree k between 1 and $n-1$,*
- *the function $N(t)$ grows exponentially,*

and the algorithm recognizes these cases (and in case of polynomial growth can decide the degree of the polynomial).

Proof. This follows from Theorems 3.1 and 3.2 together with Theorem 8.1.

This theorem answers questions raised in the seminal paper on trackability [31]. We end this section by providing an example inspired from practical applications presented in [30].

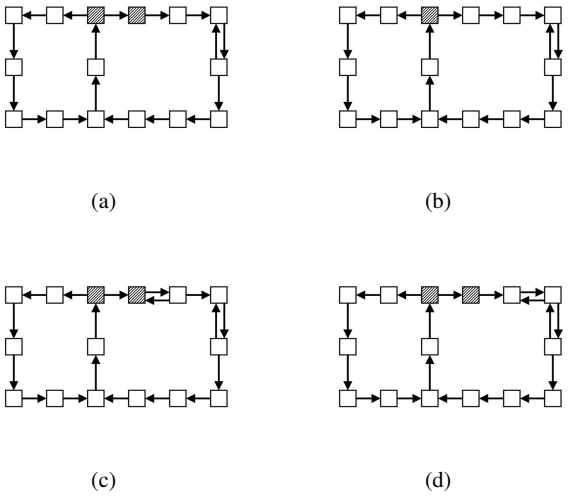


Fig. 8.3 Four sensor networks on which a ground vehicle is moving. Which ones are trackable?

Example 8.1 *Figure 8.3 shows a series of abstracted sensor networks. There is a ground vehicle moving from cell to cell according to a specified kinematics. There is a sensor which is activated if and only if the vehicle is in one of the grey cells. The white cells are not equipped with a sensor (but this is equivalent to consider that they are all equipped with a sensor that sends the signal NULL).*
The network (a) is trackable because in any (long) observation, a grey flash means that the vehicle has turned left, and two consecutive flashes mean that the vehicle

has turned right. Now if the vehicle turns right, the number of NULL flashes before the next grey flash tells how long it stayed in the right loop. Note however that the function $N(t)$ is not asymptotically equal to one, because a long sequence of NULL signals at the end of an observation does not determine exactly the final node (it could be any node lying after the loop and before the grey flash).
The network (b) is actually also trackable, even though a right turn is not indicated by a series of two grey flashes. Indeed, it is indicated by a sequence of more than seven consecutive NULL signals.
The network (c) is also trackable: the loop which is added is advertised by a sequence $GREY-NULL-GREY$.
The network (d) is not trackable, because even the subgraph at the upper right corner is not (it is a by-product of this chapter that if a subgraph is not trackable, the whole graph is not trackable either).

8.3 Conclusion and future work

In this chapter we have studied the concept of trackable graphs. We have shown how it relates to the joint spectral radius of a set of nonnegative integer matrices, and how to recognize them in polynomial time. We have briefly described some applications of this concept, in tracking vehicles in a noisy (or not fully observed) environment, or in remote control of a network via a size-constrained bandwith. An interesting related question is the design question:

Open question 12 *Given a directed graph G, how many colors are necessary in order to have a trackable graph? Is this minimal number computable in polynomial time? If one is given a number of colors, how to arrange these colors in the best way, so as to minimize the asymptotics of the maximal number of compatible trajectories $N(t)$?*

Another way of asking this question is the following: given a binary matrix A, what is the minimal number c of binary matrices $A_i : 1 \leq i \leq c$ such that $A = \sum A_i$ and $\rho(\{A_i\}) \leq 1$. We have a straightforward upper bound on this problem: $c \leq ||A_i \mathbf{1}||_\infty \leq n$. Indeed, one can decompose a matrix A in at most n matrices such that each line has at most one nonzero entry, and we have $\rho(\{A_i\}) \leq \max ||A_i||_\infty$.
One could also derive a lower bound on the minimal number of colors c by using techniques from symbolic dynamics: the entropy of the edge-shift of the graph must be smaller than the entropy of the free shift on the set of colors (for comprehensive surveys on symbolic dynamics, see [75,79]). Let us present this idea in a self-contained way: the number of paths of length t on the graph G is $\mathscr{T}_t(G) \geq \rho(A_G)^t$, with $\rho(A_G)$ the spectral radius of the adjacency matrix of G. If there exists a coloration of G with c colors such that the obtained graph is trackable, then the number of words of length t on C has to be large enough so that at most a polynomial number of paths share the same word, and we have the simple lower bound $\rho \leq c$.

This kind of techniques, though allowing one to quickly derive simple bounds, can sometimes be relatively inefficient (see [60]).

Open question 13 *Are there other (tighter) simple bounds for the minimal number of colors?*

These questions, though interesting, have not been investigated yet, to the best of the author's knowledge.

Conclusion

At the time of writing these lines, *Linear Algebra and its Applications* was editing a special issue devoted to the joint spectral radius. This is another indication of the growing interest and the increasing number of applications of the joint spectral radius. The goal of this work was twofold: to present a survey on the joint spectral radius, and to report the research that had been done during our Ph.D. on this topic. In this conclusion, we quickly recall some points developed in this thesis. We then try to put this work in perspective. We end with a personal conclusion.

Brief summary

Chapters 1 and 2 constitute a survey on the joint spectral radius.
In Chapter 1 we present elementary or fundamental results. Since it was possible to derive their counterpart concerning the joint spectral subradius, we have decided to present them.
In Chapter 2 we present more advanced results and try to understand the very nature of the joint spectral radius. In a first section, we have seen that the whole behavior is simple at first sight: The joint spectral radius is simply reached by commonly irreducible components, and for these components there exists an extremal norm, that is, a common norm that bounds individually the norm of each matrix with the exact value of the joint spectral radius. Moreover, these irreducible components can effectively be computed by quantifier elimination. In Section 2.2 we have seen that the reality is more complex: It is impossible to compute exactly the joint spectral radius. In Section 2.3 we show that despite these infeasibility results, it is possible to approximate the joint spectral radius up to an arbitrary accuracy, and that several algorithms exist, which often appear to be complementary. We end by saying a word on the finiteness property.

Concerning our own research work, two theoretical points were more deeply analyzed: First, the case of nonnegative integer matrices, for which we have delineated the polynomial time feasible questions, versus the infeasible ones. Second, the fascinating finiteness property: in the course of trying to prove that it holds for nonnegative rational (resp. rational) matrices, we have shown that it suffices to prove it for pairs of binary (resp. signed binary) matrices. In addition, we have shown that the property holds for 2×2 binary matrices.

We have also studied a number of applications of the joint spectral radius: We start with a classical one: the continuity of wavelet functions. We then turn to the capacity of codes, for which we have proved some convergence results that are more accurate than for general matrices. We have shown that the question of zero capacity is solvable in polynomial time, but that this is at the border of polynomial time feasibility, since adding don't care characters makes the problem NP-hard. We have then presented a new application of the joint spectral radius to the computation of the asymptotics of overlap-free words, a longstanding question that arises in combinatorics on words. It has been shown recently that our results can be generalized to wider applications in this area, but this still needs further investigations. We finally studied trackable sensor networks, and showed that they are recognizable in polynomial time.

What is next?

To our knowledge, the theoretical questions analyzed in Chapter 2 have not been studied for the joint spectral subradius. Some of them are perhaps not as deep as for the joint spectral radius. Indeed for instance, it is not difficult to show that the finiteness property does not hold for the joint spectral subradius: simple counterexamples exist for which the joint spectral subradius is not reached by a finite product. Nevertheless, we have the feeling that the joint spectral subradius has not been studied as much as it deserves, for instance for what concerns approximation algorithms. Perhaps the negative results mentioned in Chapter 1 are responsible for this situation, but they should not put an end to the analysis of this quantity. In this way of thinking, we present in Chapter 7 new algorithms for estimating the joint spectral subradius, that exhibit good performance in practice, at least on the particular matrices that we studied. We think that future research should analyze these algorithms and their convergence properties.

Research on the joint spectral radius is certainly not an ended story, and we have tried all along this book to emphasize questions that remain unsolved today. Some of them have been studied by several researchers from different communities, like for instance the finiteness conjecture for binary matrices (see Chapter 4). Some others have (to our knowledge) been less studied, like for instance the maximal growth of the products when the joint spectral radius is equal to one. In both cases, we felt it was worth to enlighten them, because they would have important implications in practice. These questions are summarized at the end of each chapter.

An important work that remains to be done, according to us, is a deeper understanding of the algorithms existing to approximate the joint spectral radius. One should most probably try to classify these algorithms, looking closely at their differences and similarities. The presentation of several approximation algorithms in Chapter 2 is intended to be a first step in this direction, but is definitely not a completed work. As mentioned in that chapter, it seems that a fair amount of both theoretical and numerical work is still needed in order to properly understand the different ways of approximating the joint spectral radius.
Finally, from the point of view of applications, we are wondering whether or not the joint spectral radius could be useful for more applied fields of mathematics. Indeed, as soon as a linear dynamical system is involved, and if the generalization to a switched dynamical system makes sense, the use of a joint spectral radius (and related quantities) is very natural. We have the feeling that some applications could benefit from the theoretical advances that researchers have done these last decades on such complex systems.

Personal conclusion

Before ending this book, and to summarize this work, we would like to stress one point: At first sight, and in view of the profusion of negative results on the joint spectral characteristics (undecidability, NP-hardness, non algebraicity,... see Chapter 2), one could have the impression that studying the joint spectral radius is useless. He or she could think that hoping to get an information on a system via a joint spectral radius computation is an utopia.

This is not the case at all.

On the one hand, despite all the infeasibility results, recent contributions have provided several approximation algorithms that appear to be very efficient in practice. Clearly, they require a certain amount of time in order to reach a high precision, but their flexibility often allows one to reach the precision needed. Indeed, a number of facts are of great help in practice and allow computations up to a reasonable accuracy. For instance, some algorithms allow to compute a priori the time needed to reach a given accuracy; also, algorithms of very different nature exist; finally, some algorithms can be tuned depending on algebraic properties of the particular set of matrices under study (non-negative matrices, cone-preserving matrices, commonly irreducible matrices,...). Let us mention for example the case of overlap-free words: even though the size of the matrices was relatively large (twenty by twenty), we have been able to reach a very satisfactory accuracy for the bounds on the joint spectral radius and the other related quantities. What is more, the bounds we have derived significantly outperform preexisting bounds in the literature, that had been derived with other tools.

On the other hand, the theoretical study of joint spectral characteristics is indispensable to understand the intrinsic behavior of complex systems such as switching

linear systems. In this more theoretical point of view, the joint spectral radius can be seen as a first step in the understanding of these complex dynamical systems, leading to a number of questions that remain a source of beautiful results nowadays.

Part III
Appendices

Appendix A
Numerical values for overlap-free words

A.1 Numerical values of Chapter 7

We introduce the following auxiliary matrices. For the sake of simplicity our notations do not follow exactly those of [26].

$$D_1 = \begin{pmatrix}
0&0&0&0&0&0&0&1&2&1\\
0&0&1&1&0&1&1&0&0&0\\
0&0&0&0&0&0&0&1&1&0\\
0&0&0&0&0&0&0&0&0&0\\
1&2&0&0&1&0&0&0&0&0\\
0&0&1&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0
\end{pmatrix},$$

$$B_1 = \begin{pmatrix}
0&0&0&0&0&0&0&1&2&1\\
0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&1&0&0&0\\
0&0&1&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0\\
0&1&0&0&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0&0&0
\end{pmatrix},$$

$$C_1 = \begin{pmatrix} 0&0&0&0&0&0&0&2&4&2 \\ 0&0&1&1&0&1&1&0&0&0 \\ 0&0&0&0&0&1&1&1&1&0 \\ 0&0&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&1&0&0&0&0&0 \\ 1&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&2&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0 \end{pmatrix},$$

$$B_2 = \begin{pmatrix} 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&1&1 \\ 0&1&1&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 1&0&0&0&0 \\ 0&0&0&0&0 \end{pmatrix}, C_2 = \begin{pmatrix} 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&2&0 \\ 0&1&0&0&0 \\ 0&0&0&0&0 \end{pmatrix},$$

$$C_4 = \begin{pmatrix} 0&1&1&1&1 \\ 0&0&0&1&1 \\ 0&1&1&0&0 \\ 1&0&0&0&0 \\ 1&0&0&0&0 \end{pmatrix}.$$

Now, defining

$$F_0 = \left(\begin{array}{cc|cc} C_1 & 0_{10\times 10} & C_2 & 0_{10\times 5} \\ D_1 & B_1 & 0_{10\times 5} & B_2 \\ \hline 0_{5\times 10} & 0_{5\times 10} & C_4 & 0_{5\times 5} \\ 0_{5\times 10} & 0_{5\times 10} & 0_{5\times 5} & 0_{5\times 5} \end{array}\right),$$

$$F_1 = \left(\begin{array}{cc|cc} D_1 & B_1 & 0_{10\times 5} & B_2 \\ 0_{10\times 10} & C_1 & 0_{10\times 5} & C_2 \\ \hline 0_{5\times 10} & 0_{5\times 10} & 0_{5\times 5} & 0_{5\times 5} \\ 0_{5\times 10} & 0_{5\times 10} & 0_{5\times 5} & C_4 \end{array}\right), \tag{A.1}$$

$$w = (1,2,2,2,1,2,2,1,2,1,0_{1\times 20})^T,$$

$$y_8 = (4,4,4,2,0,2,2,0,2,0,0,0,0,0,2,6,4,4,2,4,2,0,4,2,2,0,0,0,0,0)^T,$$

$$y_9 = (6,4,4,2,4,2,0,4,2,2,0,0,0,0,0,8,4,4,2,0,4,4,4,0,0,0,0,0,0,0)^T,$$

$$y_{10} = (8,4,4,2,0,4,4,4,0,0,0,0,0,0,0,8,4,6,4,8,2,0,4,2,4,0,0,0,0,0)^T,$$

$$y_{11} = (8,4,6,4,8,2,0,4,2,4,0,0,0,0,0,8,6,6,2,0,2,6,4,2,0,2,0,2,2,0)^T,$$
$$y_{12} = (8,6,6,2,0,2,6,4,2,0,2,0,2,2,0,10,6,4,4,8,2,0,4,2,4,0,0,0,0,0)^T,$$
$$y_{13} = (10,6,4,4,8,2,0,4,2,4,0,0,0,0,0,12,6,4,4,0,6,6,4,2,0,0,0,0,0,0)^T,$$
$$y_{14} = (12,6,4,4,0,6,6,4,2,0,0,0,0,0,0,10,6,8,6,12,4,0,0,4,4,0,0,0,0,0)^T,$$
$$y_{15} = (10,6,8,6,12,4,0,0,4,4,0,0,0,0,0,8,10,6,6,0,4,8,4,4,0,2,2,0,0,0)^T,$$

and introducing the recurrence relation

$$y_{2n} = F_0 y_n, \quad y_{2n+1} = F_1 y_n, \quad n \geq 8,$$

one has the relation [26]:

$$u_{n+1} = w^T y_n. \tag{A.2}$$

We finally introduce two new matrices in $\mathbb{R}^{20\times 20}$ that rule the asymptotics of u_n :

$$A_0 = \begin{pmatrix} C_1 & 0_{10\times 10} \\ D_1 & B_1 \end{pmatrix}, A_1 = \begin{pmatrix} D_1 & B_1 \\ 0_{10\times 10} & C_1 \end{pmatrix}. \tag{A.3}$$

A.2 The ellipsoidal norm

Define

$$P_1 = \begin{pmatrix} 313 & 75 & 23 & 33 & -4 & -3 & 3 & 4 & 37 & 03 \\ 75 & 577 & 100 & 63 & 184 & 350 & 163 & -58 & 138 & 50 \\ 23 & 100 & 599 & 113 & 4 & 292 & 42 & 101 & 82 & 08 \\ 33 & 63 & 113 & 485 & 46 & 135 & 108 & 20 & 69 & 10 \\ -4 & 184 & 4 & 46 & 364 & 235 & 226 & 44 & 89 & -12 \\ -3 & 350 & 292 & 135 & 235 & 1059 & 384 & 95 & 337 & 61 \\ 3 & 163 & 42 & 108 & 226 & 384 & 590 & 27 & 174 & 92 \\ 4 & -58 & 101 & 20 & 44 & 95 & 27 & 386 & 148 & -17 \\ 37 & 138 & 82 & 69 & 89 & 337 & 174 & 148 & 575 & 86 \\ 3 & 50 & 8 & 10 & -12 & 61 & 92 & -17 & 86 & 423 \end{pmatrix},$$

$$P_2 = \begin{pmatrix} -104 & -17 & -181 & -4 & -58 & -51 & -49 & -8 & -27 & -9 \\ -111 & -224 & -82 & -147 & -99 & -303 & -167 & -113 & -169 & -66 \\ -22 & -164 & -158 & -50 & -85 & -72 & -54 & -185 & -35 & -34 \\ -2 & -136 & -52 & -90 & -107 & -146 & -92 & -16 & -113 & -11 \\ -46 & -170 & -130 & -91 & -6 & -112 & -239 & -70 & -121 & 3 \\ -59 & -264 & -274 & -174 & -310 & -376 & -280 & -44 & -273 & -74 \\ -14 & -193 & -116 & -108 & -223 & -179 & -117 & -113 & -120 & -98 \\ -63 & 21 & 17 & -34 & 32 & -76 & 2 & -52 & -31 & -14 \\ -74 & -159 & -47 & -67 & -122 & -173 & -116 & -53 & -68 & -16 \\ 13 & -57 & -36 & -32 & -4 & -61 & -90 & -14 & -69 & 4 \end{pmatrix},$$

$$P_4 = \begin{pmatrix} 291 & 83 & -16 & 48 & -13 & -44 & 6 & 17 & 75 & 11 \\ 83 & 473 & 136 & 28 & 117 & 198 & 174 & 6 & 100 & 37 \\ -16 & 136 & 466 & 104 & 65 & 249 & 118 & 65 & 125 & 14 \\ 48 & 28 & 104 & 476 & 51 & 80 & 76 & 51 & 37 & 18 \\ -13 & 117 & 65 & 51 & 328 & 195 & 194 & 76 & 67 & -2 \\ -44 & 198 & 249 & 80 & 195 & 648 & 162 & 114 & 138 & 68 \\ 6 & 174 & 118 & 76 & 194 & 162 & 567 & 76 & 122 & 65 \\ 17 & 6 & 65 & 51 & 76 & 114 & 76 & 387 & 112 & -10 \\ 75 & 100 & 125 & 37 & 67 & 138 & 122 & 112 & 556 & 42 \\ 11 & 37 & 14 & 18 & -2 & 68 & 65 & -10 & 42 & 438 \end{pmatrix},$$

$$P = \begin{pmatrix} P_1 & P_2 \\ P_2^T & P_4 \end{pmatrix}.$$

Then one has the relations:

$$A^t P A - (2.5186)^{28} P \prec 0, \quad \forall A \in \Sigma^{14}.$$

As explained in Chapter 2 Section 2.3, this suffices to prove that $\rho(\Sigma) \leq 2.5186$.

A.3 The vector x

Define

$$x = (153,0,60,0,50,56,99,0,58,1,157,81,0,113,0,72,0,99,0,0)^T.$$

Then, for all $B \in \Sigma^6$ and $A \in \Sigma^{16}$, one has the relation

$$\begin{array}{ll} B(Ax - rx) & \geq 0, \\ x & \geq 0, \end{array} \tag{A.4}$$

with $r = 2.41^{16}$. This proves that $\check{\rho}(\Sigma) \geq 2.41$.

List of Figures

Index

References

1. A. V. Aho. Algorithms for finding patterns in strings. In *Handbook of theoretical computer science (vol. A): algorithms and complexity*, pages 255–300. MIT Press, Cambridge, MA, 1990.
2. T. Ando and M.-H. Shih. Simultaneous contractibility. *SIAM Journal on Matrix Analysis and Applications*, 19(2):487–498, 1998.
3. N. Barabanov. Lyapunov indicators of discrete inclusions i-iii. *Automation and Remote Control*, 49:152–157, 283–287, 558–565, 1988.
4. A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2001.
5. M. A. Berger and Y. Wang. Bounded semigroups of matrices. *Linear Algebra and its Applications*, 166:21–27, 1992.
6. J. Berstel. Growth of repetition-free words–a review. *Theoretical Computer Science*, 340(2):280–290, 2005.
7. D. P. Bertsekas, A. Nedic, and A.E. Ozdaglar. *Convex analysis and optimization*. Athena Scientific, Belmont, MA, 2003.
8. D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Belmont, MA, 1997.
9. V. D. Blondel and V. Canterini. Undecidable problems for probabilistic automata of fixed dimension. *Theory of Computing Systems*, 36(3):231–245, 2003.
10. V. D. Blondel, J. Cassaigne, and R. M. Jungers. On the number of α-power-free words for $2 < \alpha < 7/3$. *Theoretical Computer Science*, 2009. to appear.
11. V. D. Blondel, R. Jungers, and V. Protasov. On the complexity of computing the capacity of codes that avoid forbidden difference patterns. *IEEE Transactions on Information Theory*, 52(11):5122–5127, 2006.
12. V. D. Blondel, R. M. Jungers, and V. Protasov. On the complexity of computing the capacity of codes that avoid forbidden difference patterns. In *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, pages 207–212, Kyoto, 2006.
13. V. D. Blondel and Y. Nesterov. Computationally efficient approximations of the joint spectral radius. *SIAM Journal of Matrix Analysis*, 27(1):256–272, 2005.
14. V. D. Blondel and Y. Nesterov. Polynomial-time computation of the joint spectral radius for some sets of nonnegative matrices. 2008. submitted.
15. V. D. Blondel, Y. Nesterov, and J. Theys. On the accuracy of the ellipsoid norm approximation of the joint spectral radius. *Linear Algebra and its Applications*, 394(1):91–107, 2005.
16. V. D. Blondel, J. Theys, and A. Vladimirov. An elementary counterexample to the finiteness conjecture. *SIAM Journal on Matrix Analysis*, 24(4):963–970, 2003.

17. V. D. Blondel and J. N. Tsitsiklis. The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible - to compute and to approximate. *Mathematics of Control, Signals, and Systems*, 10:31–40, 1997.
18. V. D Blondel and J. N. Tsitsiklis. When is a pair of matrices mortal? *Information Processing Letters*, 63:283–286, 1997.
19. V. D. Blondel and J. N. Tsitsiklis. The boundedness of all products of a pair of matrices is undecidable. *Systems & Control Letters*, 41(2):135–140, 2000.
20. V. D. Blondel and J. N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000.
21. T. Bousch and J. Mairesse. Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture. *Journal of the Mathematical American Society*, 15(1):77–111, 2002.
22. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, 2004.
23. S. Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Applied Mathematics*, 24:83–96, 1989.
24. R. Brooks, D. Friedlander, J. Koch, and S Phoha. Tracking multiple targets with self-organizing distributed ground sensors. *Journal of Parallel and Distributed Computing*, 64(7):874–884, 2004.
25. A. Carpi. Overlap-free words and finite automata. *Theoretical Computer Science*, 115(2):243–260, 1993.
26. J. Cassaigne. Counting overlap-free binary words. *STACS 93, Lecture Notes in Computer Science*, 665:216–225, 1993.
27. B. F. Caviness and J. R. Johnson, editors. *Quantifier elimination and cylindrical algebraic decomposition*. Texts and Monographs in Symbolic Computation. Springer-Verlag, Vienna, 1998.
28. M.-D. Choi, T.-Y. Lam, and B. Reznick. Sums of squares of real polynomials. *Proceedings of Symposia in Pure Mathematics*, 58(2):103–126, 1995.
29. D. Collela and D. Heil. Characterization of scaling functions: Continuous solutions. *SIAM Journal on Matrix Analysis and Applications*, 15:496–518, 1994.
30. V. Crespi, G. Cybenko, and G. Jiang. The theory of trackability with applications to sensor networks. *ACM Transactions on Sensor Networks*, 4(3):1–42, 2008.
31. V. Crespi, G. V. Cybenko, and G. Jiang. The Theory of Trackability with Applications to Sensor Networks. Technical Report TR2005-555, Dartmouth College, Computer Science, Hanover, NH, August 2005.
32. I. Daubechies and J. C. Lagarias. Two-scale difference equations: I. existence and global regularity of solutions. *SIAM Journal of Mathematical Analysis*, 22:1388–1410, 1991.
33. I. Daubechies and J. C. Lagarias. Sets of matrices all infinite products of which converge. *Linear Algebra and its Applications*, 161:227–263, 1992.
34. I. Daubechies and J. C. Lagarias. Two-scale difference equations. ii. local regularity, infinite products of matrices and fractals. *SIAM Journal of Mathematical Analysis*, 23:1031–1079, 1992.
35. I. Daubechies and J. C. Lagarias. Corrigendum/addendum to: "Sets of matrices all infinite products of which converge" [Linear Algebra and its Applications 161 (1992), 227–263]. *Linear Algebra and its Applications*, 327(1-3):69–83, 2001.
36. Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications On Pure & Applied Mathematics*, 41:909–996, 1988.
37. Ingrid Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.
38. N.G. de Bruijn. A combinatorial problem. *Nederlandse Akademische Wetenschappen*, 49:758–764, 1946.
39. G. de Rham. Sur les courbes limites de polygones obtenus par trisection. *Enseignement Mathématique*, II(5):29–43, 1959.

40. S. Dubuc. Interpolation through an iterative scheme. *Journal of Mathematical Analysis and Applications*, 114(1):185–204, 1986.
41. L. Elsner. The generalized spectral-radius theorem: An analytic-geometric proof. *Linear Algebra and its Applications*, 220:151–159, 1995.
42. Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569, 2002.
43. M. Fekete. Uber die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit. ganzzahligen Koeffizienten. *Mathematische Zeitschrift*, 17:228–249, 1923.
44. M. R. Garey and D. S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, 1990.
45. G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, second edition, 1989.
46. G. Gripenberg. Computing the joint spectral radius. *Linear Algebra and its Applications*, 234:43–60, 1996.
47. N. Guglielmi, A. Cicone, S. Serra Capizzano, and M. Zennaro. Finiteness properties of pairs of 22 sign-matrices via polytope norms. *in preparation*.
48. N. Guglielmi, F. Wirth, and M. Zennaro. Complex polytope extremality results for families of matrices. *SIAM Journal on Matrix Analysis and Applications*, 27(3):721–743, 2005.
49. N. Guglielmi and M. Zennaro. On the asymptotic properties of a family of matrices. *Linear Algebra and its Applications*, 322:169–192, 2001.
50. N. Guglielmi and M. Zennaro. On the limit products of a family of matrices. *Linear Algebra and its Applications*, 362:11–27, 2003.
51. N. Guglielmi and M. Zennaro. Finding extremal complex polytope norms for families of real matrices. *Submitted to the SIAM Journal of Matrix Analysis and Applications*, 2008.
52. L. Gurvits. Stability of discrete linear inclusions. *Linear Algebra and its Applications*, 231:47–85, 1995.
53. L. Gurvits. Stability of linear inclusions - part 2. *NECI technical report TR*, pages 96–173, 1996.
54. C. Heil. Some stability properties of wavelets and scaling functions. In *Wavelets and Their Applications*, pages 19–38, Dordrecht, 1994. Kluwer.
55. G. Jacob. Un algorithme calculant le cardinal, fini ou infini, des demi-groupes de matrices. *Theoretical Computer Science*, 5:183–204, 1977.
56. F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays, in Honor of R. Courant*, pages 187–204. Interscience, New York, NY, 1948.
57. R. Jungers. On the growth of codes whose differences avoid forbidden patterns. Master's thesis, Université Catholique de Louvain, Belgium, 2005.
58. R. M. Jungers. *Infinite matrix products, from the joint spectral radius to combinatorics*. PhD thesis, Université Catholique de Louvain, Belgium, 2008.
59. R. M. Jungers and V. D. Blondel. Is the joint spectral radius of rational matrices reachable by a finite product? In *Proceedings of the satellite workshops of DLT2007*, pages 25–37, Turku, 2007.
60. R. M. Jungers and V. D. Blondel. Observable graphs. *Submitted*, 2007.
61. R. M. Jungers and V. D. Blondel. On the finiteness property for rational matrices. *Linear Algebra and its Applications*, 428(10):2283–2295, 2008.
62. R. M. Jungers, V. Protasov, and V. D. Blondel. Efficient algorithms for deciding the type of growth of products of integer matrices (extended abstract). In *Proceedings of the 13th ILAS Conference*, volume 1, pages 556–563, Amsterdam, 2006.
63. R. M. Jungers, V. Protasov, and V. D. Blondel. Overlap-free words and spectra of matrices. *Submitted*, 2007. Preprint: `http://arxiv.org/abs/0709.1794`.
64. R. M. Jungers, V. Protasov, and V. D. Blondel. Computing the growth of the number of overlap-free words with spectra of matrices. In *Lecture Notes in Computer Science, Proceedings of LATIN08*, volume 4957, pages 84–93, Buzios, Rio de Janeiro, Brazil, 2008.

65. R. M. Jungers, V. Protasov, and V. D. Blondel. Efficient algorithms for deciding the type of growth of products of integer matrices. *Linear Algebra and its Applications*, 428(10):2296–2311, 2008.
66. J. Karhumäki and J. Shallit. Polynomial versus exponential growth in repetition-free binary words. *Journal of Combinatorial Theory Series A*, 105(2):335–347, 2004.
67. A. J. Kfoury. A linear time algorithm to decide whether a binary word contains an overlap. *Theoretical Informatics and Applications*, 22:135–145, 1988.
68. Y. Kobayashi. Enumeration of irreducible binary words. *Discrete Applied Mathematics*, 20:221–232, 1988.
69. V. Kozyakin. A dynamical systems construction of a counterexample to the finiteness conjecture. *Proceedings of the 44th IEEE Conference on Decision and Control and ECC 2005*, pages 2338–2343, 2005.
70. V. S. Kozyakin. Algebraic unsolvability of problem of absolute stability of desynchronized systems. *Automation and Remote Control*, 51(6):754–759, 1990.
71. J. C. Lagarias and Y. Wang. The finiteness conjecture for the generalized spectral radius of a set of matrices. *Linear Algebra and its Applications*, 214:17–42, 1995.
72. P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, New York, NY, second edition, 1985.
73. A. Lepistö. A characterization of 2+-free words over a binary alphabet. Master's thesis, University of Turku, Finland, 1995.
74. D. Liberzon, J. P. Hespanha, and A. S. Morse. Stability of switched systems: a Lie-algebraic condition. *Systems and Control Letters*, 37(3):117–122, 1999.
75. D. A. Lind and B. H. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.
76. M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics*. Cambridge University Press, 1982.
77. M. Maesumi. An efficient lower bound for the generalized spectral radius of a set of matrices. *Linear Algebra and its Applications*, 240:1–7, 1996.
78. A. Mandel and I. Simon. On finite semigroups of matrices. *Journal of Theoretical Computer Science*, 5(2):101–111, October 1977.
79. B. H. Marcus, R. M. Roth, and P. H. Siegel. An introduction to coding for constrained systems. Draft version (2001) available on line at `http://vivaldi.ucsd.edu:8080/~psiegel/Book/Book_Draft.html`.
80. Y. Matiyasevich and G. Senizergues. Decision problems for Semi-Thue systems with a few rules. In *Logic in Computer Science*, pages 523–531, 1996.
81. B. E. Moision, A. Orlitsky, and P. H. Siegel. Bounds on the rate of codes which forbid specified difference sequences. In *Proceedings of the IEEE Global Telecommunication Conference (GLOBECOM'99)*, 1999.
82. B. E. Moision, A. Orlitsky, and P. H. Siegel. On codes that avoid specified differences. *IEEE Transactions on Information Theory*, 47:433–442, 2001.
83. B. E. Moision, A. Orlitsky, and P. H. Siegel. On codes with local joint constraints. *Linear Algebra and its Applications*, 422(2-3):442–454, 2007.
84. R. Mc Naughton and Y. Zalcstein. The Burnside problem for semigroups. *Journal of Algebra*, 34:292–299, 1975.
85. Y. Nesterov. Squared functional systems and optimization problems. In *High performance optimization, volume 33 of Applied Optimization*, pages 405–440. Kluwer, Dordrecht, 2000.
86. Y. Nesterov and A. Nemirovski. *Interior point polynomial methods in convex programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.
87. S. Oh, L. Schenato, P. Chen, and S. Sastry. A scalable real-time multiple-target tracking algorithm for sensor networks. Technical Report UCB//ERL M05/9, University of California, Berkeley, Hanover, NH, 2005.

88. V. I. Oseledets. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Transactions of the Moscow Mathematical Society*, 19:197–231, 1968.
89. P. Parrilo and A. Jadbabaie. Approximation of the joint spectral radius of a set of matrices using sum of squares. 2007. In Alberto Bemporad, Antonio Bicchi and Giorgio Buttazzo (Editors), *Hybrid Systems:Computation and Control* Springer Lecture Notes in Compter Science.
90. P. Parrilo and A. Jadbabaie. Approximation of the joint spectral radius using sum of squares. *Linear Algebra and its Applications*, 428(10):2385–2402, 2008.
91. P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, USA, 2000. `http://resolver.caltech.edu/CaltechETD:etd-05062004-055516`.
92. M. S. Paterson. Unsolvability in 3×3 matrices. *Studies in Applied Mathematics*, 49:105–107, 1970.
93. A. Paz. *Introduction to Probabilistic Automata*. Academic Press, New York, 1971.
94. V. Y. Protasov. The joint spectral radius and invariant sets of linear operators. *Fundamentalnaya i prikladnaya matematika*, 2(1):205–231, 1996.
95. V. Y. Protasov. The generalized spectral radius. A geometric approach. *Izvestiya Mathematics*, 61(5):995–1030, 1997.
96. V. Y. Protasov. Asymptotic behaviour of the partition function. *Sbornik Mathematics*, 191(3-4):381–414, 2000.
97. V. Y. Protasov. On the regularity of de Rham curves. *Izvestiya Mathematics*, 68(3):567–606, 2004.
98. V. Y. Protasov. Fractal curves and wavelets. *Izvestiya Mathematics*, 70(5):123–162, 2006.
99. V. Y. Protasov, R. M. Jungers, and V. D. Blondel. Joint spectral characteristics of matrices: a conic programming approach. *submitted*, 2009.
100. L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In A. Waibel and K.-F. Lee, editors, *Readings in Speech Recognition*, pages 267–296. Kaufmann, San Mateo, CA, 1990.
101. A. Restivo and S. Salemi. Overlap-free words on two symbols. *Lecture Notes in Computer Science, Automata on Infinite Words*, 192:198–206, 1985.
102. B. Reznick. Some binary partition functions. In B.C. Berndt, H.G. Diamond, H. Halberstam, and A. Hildebrand, editors, *Analytic Number Theory, proceedings of a Conference in Honor of Paul T. Bateman*, pages 451–477, Boston, 1990. Birkhäuser.
103. O. Rioul. Simple regularity criteria for subdivision schemes. *SIAM Journal of Mathematical Analysis*, 23:1544–1576, 1999.
104. G. C. Rota and G. Strang. A note on the joint spectral radius. *Proceedings of the Netherlands Academy*, 22:379–381, 1960.
105. H. Samelson. *Notes on Lie Algebras*. Van Nostrand Reinhold, New York,NY, 1969.
106. N. Z. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23(6):731–734, 1987. Russian orig.: Kibernetika, No. 6, (1987), 9-11.
107. N. J. A. Sloane. On-line encyclopedia of integer sequences. Url : `http://www.research.att.com/~njas/sequences`.
108. G. Strang. Wavelets and dilation equations: a brief introduction. *SIAM Review*, 31(4):614–627, 1989.
109. Gilbert Strang. Wavelet transforms versus fourier transforms. *Bulletin (New Series) of the American Mathematical Society*, 28:288–305, 1993.
110. R. E. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.
111. J. Theys. *Joint Spectral Radius : Theory and approximations*. PhD thesis, Université catholique de Louvain, Belgium, 2005.
112. A. Thue. über unendliche Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.*, 7:1–22, 1906.

113. A. Thue. über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.*, 1:1–67, 1912.
114. I. Vinogradov. *Elements of number theory. Translated by S. Kravetz*. Dover Publications, New York, NY, 1954.
115. F. Wirth. The generalized spectral radius and extremal norms. *Linear Algebra and its Applications*, 342:17–40, 2000.
116. F. Wirth. The generalized spectral radius is strictly increasing. *Linear Algebra and its Applications*, 395:141–153, 2005.
117. F. Wirth. On the structure of the set of extremal norms of a linear inclusion. *Proceedings of the 44th IEEE Conference on Decision and Control and ECC 2005*, pages 3019–3024, 2005.
118. K.-S. Lau X. G. He. Characterization of tile digit sets with prime determinants. *Applied and Computational Harmonic Analysis*, 16:159–173, 2004.
119. H. Yang and B Sikdar. A protocol for tracking mobile targets using sensor networks. In *Proceedings of the First IEEE Workshop on Sensor Network Protocols and Applications*, pages 71–81, Anchorage, AL, 2003.
120. F. Zhao, J. Shin, and J. Reich. Information-driven dynamic sensor collaboration. *IEEE Signal Processing Magazine*, 19(2):61–72, 2002.