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.

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

[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].

[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].

R. Jungers: The Joint Spectral Radius: Theory and Applications, LNCIS 385, pp. 101–114.
springerlink.com © Springer-Verlag Berlin Heidelberg 2009

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

$$C_1 n^{1.155} \leq u_n \leq C_2 n^{1.37}.$$

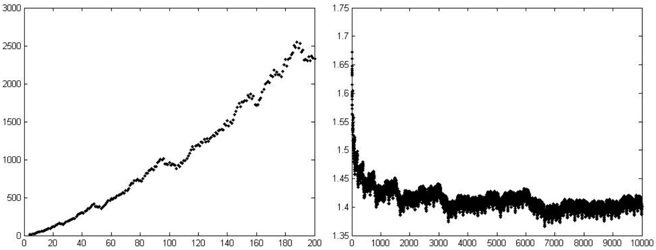
Carpi showed that there is a finite automaton allowing to compute u_n (the sequence u_n is 2-regular [25]). In Figure 7.1(a) we show the values of the sequence u_n for $1 \leq n \leq 200$ and in Figure 7.1(b) we show the behavior of $\log u_n / \log n$ for larger values of n. One can see that the sequence 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 numerical 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:

$$1.155 < \alpha < 1.276 \qquad \text{and} \qquad 1.332 < \beta < 1.37. \tag{7.2}$$

[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,...



(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 $d_t \dots d_1$ 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 $\mathrm{int}K$ such that for any $x \in \mathrm{int}K$ 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 $x \in \mathrm{int}K$, 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 $d_k \dots d_1$ 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

$$\lim_{n\to\infty} \left(\frac{\log_2 u_n}{\log_2 n} - \frac{\log_2 \|A_{d_1}\cdots A_{d_k}\|}{k} \right) =$$
$$\lim_{n\to\infty} \left(\frac{\log_2 u_n - \log_2 \|A_{d_1}\cdots A_{d_k}\|}{k} + \left(\frac{k}{\log_2 n} - 1\right) \frac{\log_2 u_n}{k} \right) =$$
$$\lim_{n\to\infty} \left(\frac{\log_2 u_n - \log_2 \|A_{d_1}\cdots A_{d_k}\|}{k} \right) = \lim_{n\to\infty} \frac{\log_2 \left(u_n \cdot \|A_{d_1}\cdots A_{d_k}\|^{-1}\right)}{k},$$

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 $\alpha_t = \min_{2^{t-1}<n\le 2^t} \frac{\log u_n}{\log n}$ and $\beta_t = \max_{2^{t-1}<n\le 2^t} \frac{\log u_n}{\log n}$. 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 $\hat{\rho}_t \ge \rho^t$ and $\check{\rho}_t \ge \check{\rho}^t$ 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,

$$\hat{\rho}_t \le C_2 \rho^t.$$

Corollary 7.3. *There are positive constants C_1, C_2 such that*

$$C_1 n^\alpha \le u_n \le C_2 n^\beta, \; n \in \mathbb{N}.$$

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 $\frac{1}{n}\mathrm{Card}\{r \le n,\ r \in \mathscr{A}\} \to\ 1$ 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 $d_1, d_2, \dots$ 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:

$$\bar{\rho}(A_0, A_1) = \lim_{t \to \infty} \Big(\prod_{d_1, \dots, d_t} \|A_{d_1} \cdots A_{d_t}\|^{1/t} \Big)^{1/2^t} \tag{7.10}$$

(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

$$P_t(\mathscr{P}) =\ 2^{-(t-1)}\mathrm{Card}\big\{n \in \{2^{t-1}+1, \dots, 2^t\} : n \,\text{satisfies}\, \mathscr{P}\ \big\}.$$

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 $\lim_{n \to \infty, n \in \mathscr{A}} f_n =\ f$. 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

$$P_k\left(\left|\frac{\log u_n}{\log n} - \sigma\right| > \frac{1}{j}\right) \leq \frac{1}{j} \qquad \text{for all} \quad k \geq k_j\,.$$

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 $\left|\frac{\log u_n}{\log n} - \sigma\right| \leq \frac{1}{j}$, 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

$$\frac{2^k}{j} + \frac{2^{k_j}}{j-1} + \cdots + \frac{2^{k_2}}{1} \leq \sum_{s=1}^{j} \frac{2^{k-j+s}}{s} = 2^{k-j} \sum_{s=1}^{j} \frac{2^s}{s}\,.$$

Observe that $\sum_{s=1}^{j} \frac{2^s}{s} \leq \frac{3 \cdot 2^j}{j}$, 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

$$\rho \quad \geq \quad \left[\rho(A_0A_1)\right]^{1/2} = 2.5179\ldots, \tag{7.12}$$

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,

$$\beta = \sqrt{\rho(A_0A_1)}.$$

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 $A_2(A_1 x - rx) = CB(A_1 x - rx) \geq 0$, with $B \in \Sigma^s, C \in \Sigma^{t-s}$. For $k > 2$ we have $A_k \dots A_1 x = A_k A_{k-1} \dots A_1 x \geq r^{k-2} A_k A_{k-1} x \geq r^{k-1} A_k x$. In the last inequality the case for $k = 2$ was reused.

Hence,

$$||A_k \dots A_1|| = \mathbf{1}^T A_k \dots A_1 \mathbf{1} \geq r^{k-1} \mathbf{1}^T A_k x \geq K r^k,$$

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*

$$\begin{array}{l} B^T(A^T S A - rS)B \succeq 0 \ \forall A \in \Sigma^t, B \in \Sigma^s, \\ S \succ 0, \end{array} \tag{7.17}$$

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 $A_1 \dots A_k$, 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:

$$\rho_1 = \lim_{t \to \infty} \Big(2^{-t} \sum_{d_1, \dots, d_t} \|A_{d_1} \cdots A_{d_t}\|\Big)^{1/t}.$$

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 $r_t = \Big(\prod_{d_1, \dots, d_t} \|A_{d_1} \cdots A_{d_t}\|\Big)^{\frac{1}{t 2^t}}$ 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 $p_t(x) = \Big(\prod_{d_1,\dots,d_t} |A_{d_1}\cdots A_{d_t}x|\Big)^{\frac{1}{2^t}}$ 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

$$\lim_{t\to\infty} \big[p_{tk}(x)\big]^{\;1/tk} \to \bar{\rho}$$

as $t \to \infty$. On the other hand, $p_{k+n}(x) \leq m_k p_n(x)$ for any $x \geq 0$ and for any $n, k \in \mathbb{N}$, therefore $p_{tk}(x) \leq (m_k)^t$. Thus, $\bar{\rho} \leq (m_k)^{1/k}$.

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 $\frac{1}{2}\big(A_0^T + A_1^T\big)$ 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 $t \geq 1$ and $y \in \mathbb{R}^d_+$, $|y| = (y, v_*) = 1$, such that $p_t(y) = m_t$. We have

$$m_t = p_t(y) \leq 2^{-t} \sum_{d_1,\dots,d_t} |A_{d_1}\cdots A_{d_t}y| = \; 2^{-t} \sum_{d_1,\dots,d_t} \big(A_{d_1}\cdots A_{d_t}y, v_*\big)$$

$$= \; \Big(y, 2^{-t}\big(A_0^T + A_1^T\big)^t v_*\Big) = \; \rho_1^t\big(y, v_*\big) = \rho_1^t.$$

Thus, $m_t \leq \rho_1^t$, 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 $\bar{\rho} \leq (m_t)^{1/t} < \rho_1$.

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:

$$\begin{array}{ll} \min \; -\frac{1}{t2^t \ln 2} \sum\limits_{d_1,\dots,d_t \in \{0,1\}} \ln\Big(x, A_{d_1}^T \cdots A_{d_t}^T v_*\Big) & \\ \text{s.t.} \qquad\qquad x \geq 0, \quad (x, v_*) = 1. & \end{array} \tag{7.18}$$

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 $\sigma \leq 1.3098$. 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*

$$\begin{array}{lll} B(A_i x - r_i x) & \geq 0, & \forall B \in \Sigma^s, A_i \in \Sigma^t, \\ x & \geq 0, & (x, \mathbf{1}) = 1, \end{array} \tag{7.19}$$

then $\bar{\rho}(\Sigma) \geq \prod_i r_i^{1/(tm^t)}$.

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.