Theoretical Computer Science 410 (2009) 3670–3684

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Overlap-free words and spectra of matrices

Raphaël M. Jungers [a,*], Vladimir Y. Protasov [b], Vincent D. Blondel [a]

[a] Division of Applied Mathematics, Université catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium

[b] Department of Mechanics and Mathematics, Moscow State University, Vorobyovy Gory, Moscow, 119992, Russia

ARTICLE INFO

Article history:
Received 12 September 2007
Received in revised form 9 February 2009
Accepted 28 April 2009
Communicated by D. Perrin

Keywords:
Overlap-free words
Combinatorics on words
Joint spectral radius
Joint spectral subradius
Lyapunov exponent

ABSTRACT

Overlap-free words are words over the binary alphabet $A = \{a, b\}$ that do not contain factors of the form $xvxvx$, where $x \in A$ and $v \in A^*$. We analyze the asymptotic growth of the number u_n of overlap-free words of length n as $n \to \infty$. We obtain explicit formulas for the minimal and maximal rates of growth of u_n in terms of spectral characteristics (the joint spectral subradius and the joint spectral radius) of certain sets of matrices of dimension 20×20. Using these descriptions we provide new estimates of the rates of growth that are within 0.4% and 0.03% of their exact values. The best previously known bounds were within 11% and 3%, respectively. We then prove that the value of u_n actually has the same rate of growth for "almost all" natural numbers n. This average growth is distinct from the maximal and minimal rates and can also be expressed in terms of a spectral quantity (the Lyapunov exponent). We use this expression to estimate it. In order to obtain our estimates, we introduce new algorithms to compute the spectral characteristics of sets of matrices. These algorithms can be used in other contexts and are of independent interest.

1. Introduction

Binary overlap-free words have been studied for more than a century. These are words over the binary alphabet $A = \{a, b\}$ that do not contain factors of the form $xvxvx$, where $x \in A$ and $v \in A^*$. For instance, the word *baabaa* is overlap free, but the word *baabaab* is not, since it can be written $xuxux$ with $x = b$ and $u = aa$. See [2] for a recent survey. Thue [27,28] proved in 1906 that there are infinitely many overlap-free words. Indeed, the well-known Thue–Morse sequence[1] 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.[2] The number of factors of length n in the Thue–Morse sequence is proved in [8] to be larger than $3n - 3$, thus providing a linear lower bound on u_n:

$$u_n \geq 3n - 3.$$

The next improvement was obtained by Restivo and Salemi [25]. By using a certain decomposition result, they showed that the number of overlap-free words grows at most polynomially:

$$u_n \leq C n^r,$$

where $r = \log(15) \approx 3.906$. This bound has been sharpened successively by Kfoury [16], Kobayashi [17], and finally by Lepistö [18] to the value $r = 1.37$. One could then suspect that the sequence u_n grows linearly. However, Kobayashi [17]

* Corresponding author. Tel.: +32 10 47 24 05.
E-mail addresses: raphael.jungers@uclouvain.be (R.M. Jungers), v-protassov@yandex.ru (V.Y. Protasov), vincent.blondel@uclouvain.be (V.D. Blondel).

1 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 [10].

2 The number of overlap-free words of length n is referenced in the On-Line Encyclopedia of Integer Sequences under the code A007777; see [26]. The sequence starts 1, 2, 4, 6, 10, 14, 20, 24, 30, 36, 44, 48, 60, 60, 62, 72, … .

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2009.04.022

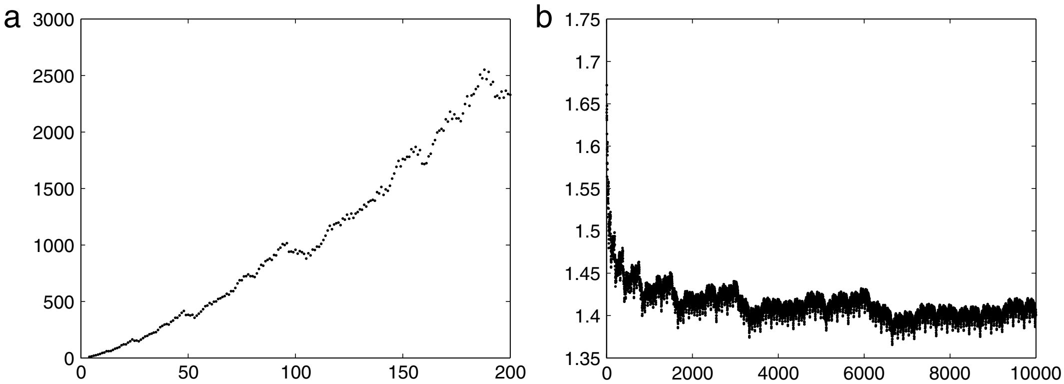

Fig. 1. The values of u_n for $1 \leq n \leq 200$ (a) and $\log u_n / \log n$ for $1 \leq n \leq 10\,000$ (b).

proved that this is not the case. By enumerating the subset of overlap-free words of length n that can be infinitely extended to the right he showed that $u_n \geq C\, n^{1.155}$ and so we have

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

In Fig. 1(a) we show the values of the sequence u_n for $1 \leq n \leq 200$ and in Fig. 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 [10] that the answer is negative. He introduced the lower and the upper exponents of growth:

$$\begin{aligned} \alpha &= \sup\left\{r \;\middle|\; \exists C > 0, u_n \geq Cn^r\right\}, \\ \beta &= \inf\left\{r \;\middle|\; \exists C > 0, u_n \leq Cn^r\right\}, \end{aligned} \tag{1}$$

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 linear recurrence relations. These relations are explicitly given in the next section and all numerical values can be found in Appendix A. 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 has 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:

$$1.155 < \alpha < 1.276 \quad \text{and} \quad 1.332 < \beta < 1.37. \tag{2}$$

Let us add that Carpi had already shown that the sequence u_n is 2-regular [9]. In this paper 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 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 *joint spectral characteristics* of these matrices. We prove that $\alpha = \log_2 \check{\rho}(A_0, A_1)$ and $\beta = \log_2 \hat{\rho}(A_0, A_1)$, where $\check{\rho}$ and $\hat{\rho}$ denote, respectively, the joint spectral subradius and the joint spectral radius of the matrices A_0, A_1. (We define these notions in the next section.) In Section 3, we estimate these values and we obtain the following improved bounds for α and β:

$$1.2690 < \alpha < 1.2736 \quad \text{and} \quad 1.3322 < \beta < 1.3326. \tag{3}$$

Our estimates are, respectively, within 0.4% and 0.03% of the exact values. In addition, we show in Theorem 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 $C_1\, n^{\alpha} \leq u_n \leq C_2\, n^{\beta}$ for all $n \in \mathbb{N}$.

Although the sequence u_n does not exhibit an asymptotic polynomial growth, we then show in Theorem 5 that for "almost all" values of n the rate of growth is actually the same and 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 grows neither 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 [10] to compute u_n are asymptotically negligible when considering the total number of overlap-free words.

The paper is organized as follows. In the next section we formulate and prove the main theorems (except for Theorem 2, whose proof is quite technical and is given in Appendix B). Then in Section 3 we present algorithms for estimating the joint spectral radius, the joint spectral subradius, and the Lyapunov exponent of linear operators. 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, we give a proof of Theorem 2 and present the results of our numerical algorithms.

2. The asymptotics of the overlap-free words

In what follows we use the following notation: $\mathbb{R}^d$ is the d-dimensional space, inequalities $x \geq 0$ and $A \geq 0$ mean that all the entries of the vector x (respectively, of the matrix A) are nonnegative. We write $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d, x \geq 0\}$, by $|x|$ we denote a norm of the vector $x \in \mathbb{R}^d$, and by $\|\cdot\|$ any matrix norm. In particular, $|x|_1 = \sum_{i=1}^d |x_i|$, $\|A\|_1 = \sup_{|x|_1=1} |Ax|_1 = \max_{j=1,\dots d} \sum_{i=1}^d |A_{ij}|$. We write $\mathbf{1}$ for the vector $(1, \dots, 1)^{\mathrm{T}} \in \mathbb{R}^d$, and $\rho(A)$ for the spectral radius of the matrix A, that is, the largest magnitude of its eigenvalues. If $A \geq 0$, then there is a vector $v \geq 0$ such that $Av = \rho(A)v$ (the so-called Perron–Frobenius eigenvector). For two functions f_1, f_2 from a set Y to $\mathbb{R}_+$ the relation $f_1(y) \asymp f_2(y)$ means that there are positive constants C_1, C_2 such that $C_1 f_1(y) \leq f_2(y) \leq C_2 f_1(y)$ for all $y \in Y$.

To compute the number u_n of overlap-free words of length n we use several results from [10] that we summarize in the following theorem:

Theorem 1. *Let $F_0, F_1 \in \mathbb{R}^{30\times 30}$, and let $w, y_8, \dots, y_{15} \in \mathbb{R}^{30}_+$ be as given in Appendix A. For $n \geq 16$, let y_n be the solution of the following recurrence equations:*

$$\begin{aligned} y_{2n} &= F_0 y_n \\ y_{2n+1} &= F_1 y_n. \end{aligned} \tag{4}$$

Then, for any $n \geq 9$, the number of overlap-free words of length n is equal to $w^{\mathrm{T}} y_{n-1}$.

It follows from this result that the number u_n of overlap-free words of length $n > 16$ can be obtained by first computing the binary expansion $d_k \cdots d_1$ of $n-1$, i.e., $n - 1 = \sum_{j=0}^{k-1} d_{j+1} 2^j$, and then defining

$$u_n = w^{\mathrm{T}} F_{d_1} \cdots F_{d_{k-4}} y_m, \tag{5}$$

where $m = d_{k-3} + d_{k-2}2 + d_{k-1}2^2 + d_k 2^3$ (and $d_k = 1$). To arrive at the results summarized in Theorem 1, 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 asymptotically given 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 [10] for more details).

Theorem 2. *Let $A_0, A_1 \in \mathbb{R}^{20\times 20}_+$ be the matrices defined in Appendix A (Eq. (A.3)), let $\|\cdot\|$ be a matrix norm, and let $A(n) : \mathbb{N} \to \mathbb{R}^{20\times 20}_+$ be defined as $A(n) = A_{d_1} \cdots A_{d_k}$ with $d_k \dots d_1$ the binary expansion of $n-1$. Then,*

$$u_n \asymp ||A(n)||. \tag{6}$$

Observe that the matrices F_0, F_1 in Theorem 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 [23].

Lemma 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*

$$\gamma(x)\|B\| \cdot |x| \leq |Bx| \leq \frac{1}{\gamma(x)}\|B\| \cdot |x|.$$

Corollary 1. *Let two matrices A_0, A_1 possess an invariant cone $K \subset \mathbb{R}^d$. Then for any $x \in$ int K we have $|A_{d_1} \cdots A_{d_k} x| \asymp \|A_{d_1} \cdots A_{d_k}\|$ for all k and for all indices $d_1, \dots, d_k \in \{0, 1\}$.*

In view of Corollary 1 and of Eq. (5), Theorem 2 may seem obvious, at least if we consider the matrices F_i instead of A_i. One cannot however directly apply Lemma 1 and Corollary 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 2 we need to construct a wider invariant cone of A_0 and A_1 by using special properties of these matrices. That construction is nontrivial, and we detail it in the proof given in Appendix B. Theorem 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 2 yields the following corollary:

Corollary 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*

$$\frac{\log_2 u_n}{\log_2 n} - \log_2 \|A(n)\|^{1/k} \to 0 \quad \text{as } n \to \infty. \tag{7}$$

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

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

and by Theorem 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 Eq. (1). For a given set of matrices $\Sigma = \{A_1, \dots, A_m\}$ we denote by $\check{\rho}$ and $\hat{\rho}$ its joint spectral subradius and its joint spectral radius:

$$\begin{aligned}
\check{\rho}(\Sigma) &= \lim_{k\to\infty} \min_{d_1,\dots,d_k \in \{1,\dots,m\}} \|A_{d_1} \cdots A_{d_k}\|^{1/k}, \\
\hat{\rho}(\Sigma) &= \lim_{k\to\infty} \max_{d_1,\dots,d_k \in \{1,\dots,m\}} \|A_{d_1} \cdots A_{d_k}\|^{1/k}.
\end{aligned} \tag{8}$$

Both limits are well defined and do not depend on the chosen norm. Moreover, for any product $A_{d_1} \cdots A_{d_k}$ we have

$$\check{\rho} \le \rho(A_{d_1} \cdots A_{d_k})^{1/k} \le \hat{\rho} \tag{9}$$

(see [7,12,21] for surveys on these notions).

Theorem 3. *For $k \ge 1$, let $\alpha_k = \min_{2^{k-1} < n \le 2^k} \frac{\log u_n}{\log n}$ and $\beta_k = \max_{2^{k-1} < n \le 2^k} \frac{\log u_n}{\log n}$. Then*

$$\alpha = \lim_{k\to\infty} \alpha_k = \log_2 \check{\rho}(A_0, A_1) \quad \text{and} \quad \beta = \lim_{k\to\infty} \beta_k = \log_2 \hat{\rho}(A_0, A_1), \tag{10}$$

where the matrices A_0, A_1 are defined in Appendix A. Moreover, there are positive constants C_1, C_2 such that

$$C_1 \le \min_{2^{k-1} < n \le 2^k} u_n n^{-\alpha} \quad \text{and} \quad C_1 \le \max_{2^{k-1} < n \le 2^k} u_n n^{-\beta} \le C_2 \tag{11}$$

for all $k \in \mathbb{N}$.

The proof of this theorem is based on the following auxiliary result taken from [23]. For a given set of indices $\{i_1, \dots, i_p\} \subsetneq \{1, \dots, d\}$, $1 \le p \le d-1$ we call the subspace $L_{i_1,\dots,i_p} = \{x \in \mathbb{R}^d, x_{i_1} = \cdots = x_{i_p} = 0\}$ a *coordinate plane*.

Proposition 1 ([23]). *Let A_0, A_1 be matrices with a common invariant cone. Then there is a positive constant c_1 such that*

$$\max_{d_1,\dots,d_k} \|A_{d_1} \cdots A_{d_k}\| \ge c_1 \hat{\rho}^k \quad \text{and} \quad \min_{d_1,\dots,d_k} \|A_{d_1} \cdots A_{d_k}\| \ge c_1 \check{\rho}^k, \quad k \in \mathbb{N}.$$

If, moreover, these matrices have no common invariant subspace among the coordinate planes, then there is a positive constant c_2 such that

$$\max_{d_1,\dots,d_k} \|A_{d_1} \cdots A_{d_k}\| \le c_2 \hat{\rho}^k, \qquad k \in \mathbb{N}.$$

Proof of Theorem 3. The equalities (10) follow immediately from Corollary 2 and the definitions (8). To prove the inequalities (11) we apply Proposition 1 to our matrices A_0, A_1, which leave $\mathbb{R}^{20}_+$ invariant. Theorem 2 yields

$$u_n n^{-\alpha} \asymp \|A_{d_1} \cdots A_{d_k}\| 2^{-\alpha k} = \|A_{d_1} \cdots A_{d_k}\| \check{\rho}^{-k}.$$

Taking the minimum over $n = 2^{k-1} + 1, \ldots, 2^k$ and invoking Proposition 1, we conclude that

$$\min_{2^{k-1} < n \leq 2^k} u_n n^{-\alpha} \geq C_1.$$

The same holds with the inequality

$$\max_{2^{k-1} < n \leq 2^k} u_n n^{-\beta} \geq C_1.$$

To prove the upper bound in Eq. (11) 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). $\square$

Corollary 3. *There are positive constants C_1, C_2 such that*

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

In the next section we will see 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 [10]. 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. 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 4 (*[19]*). *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 $\|A_{d_1} \cdots A_{d_k}\|^{1/k}$ converges to some number $\bar{\rho}$ with probability* 1. *This means that for any $\varepsilon > 0$ we have*

$$P\left(\left| \|A_{d_1} \cdots A_{d_k}\|^{1/k} - \bar{\rho}\right| > \varepsilon\right) \to 0 \quad \textit{as } k \to \infty.$$

The limit $\bar{\rho}$ in Theorem 4 is called the *Lyapunov exponent* of the set $\{A_0, A_1\}$. This value is given by the following formula:

$$\bar{\rho}(A_0, A_1) = \lim_{k \to \infty} \left(\prod_{d_1, \ldots, d_k} \|A_{d_1} \cdots A_{d_k}\|^{1/k} \right)^{1/2^k} \tag{12}$$

(for the proof see, for instance, [24]). To understand what this gives for the asymptotics of our sequence u_n we introduce some further notation. Let $\mathcal{P}$ be some property of natural numbers. For a given $k \in \mathbb{N}$ we denote

$$P_k(\mathcal{P}) = 2^{-(k-1)} \operatorname{Card}\left\{ n \in \{2^{k-1} + 1, \ldots, 2^k\}, n \text{ satisfies } \mathcal{P} \right\}.$$

Thus, P_k is the probability that the integer n uniformly distributed on the set $\{2^{k-1} + 1, \ldots, 2^k\}$ satisfies $\mathcal{P}$. Combining Corollary 2 and Theorem 4 we obtain

Theorem 5. *There is a number σ such that for any $\varepsilon > 0$ we have*

$$P_k\left(\left| \frac{\log u_n}{\log n} - \sigma \right| > \varepsilon \right) \to 0 \quad \textit{as } k \to \infty.$$

Moreover, $\sigma = \log_2 \bar{\rho}$, where $\bar{\rho}$ is the Lyapunov exponent of the matrices $\{A_0, A_1\}$ defined in Appendix A.

Thus, for almost all numbers $n \in \mathbb{N}$ the number of overlap-free words u_n has the same exponent of growth $\sigma = \log_2 \bar{\rho}$. For an arbitrary $q > 1$, if $a \in \mathbb{N}$ is large enough, then for a number n taken randomly from the segment $[a, qa]$ the value $\frac{\log u_n}{\log n}$ is close to σ with high probability. Let us recall that a subset $\mathcal{A} \subset \mathbb{N}$ is said to have density 1 if $\frac{1}{n}\operatorname{Card}\left\{ r \leq n, r \in \mathcal{A} \right\} \to 1$ as $n \to \infty$. We say that a sequence f_n converges to a number f along a set of density 1 if there is a set $\mathcal{A} \subset \mathbb{N}$ of density 1 such that $\lim_{n \to \infty, n \in \mathcal{A}} f_n = f$. Theorem 5 yields

Corollary 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 5 the values k_j are well defined for all j. Let a set $\mathcal{A}$ 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 $\mathcal{A}$. If, as usual, $2^{k-1} \leq n < 2^k$, then the total number of integers $r \leq n$ that do not belong to $\mathcal{A}$ 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 $\mathcal{A}$ 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, $\mathcal{A}$ has density 1. $\square$

3. Computation of the exponents

Theorems 3 and 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" [5]. For the joint spectral subradius and for the Lyapunov exponent we present new algorithms, which seem to be relatively efficient for nonnegative matrices. The results we obtain can be summarized in the following theorem:

Theorem 6.

$$\begin{aligned} &1.2690 < \alpha < 1.2736, \\ &1.3322 < \beta < 1.3326, \\ &1.3005 < \sigma < 1.3098. \end{aligned} \tag{13}$$

In this section we also make (and give arguments for) the following conjecture:

Conjecture 1.

$$\beta = \log_2 \sqrt{\rho(A_0 A_1)} = 1.3322\ldots.$$

3.1. Computation of β and the joint spectral radius

By Theorem 3, in order to estimate the exponent β one needs to estimate the joint spectral radius of the set $\{A_0, A_1\}$. A lower bound for $\hat{\rho}$ can be obtained by applying inequality (9). Taking $k = 2$ and $d_1 = 0$, $d_2 = 1$ we get

$$\hat{\rho} \geq \left[\rho(A_0 A_1)\right]^{1/2} = 2.5179\ldots, \tag{14}$$

and so $\beta > \log_2 2.5179 > 1.3322$ (this lower bound was already found in [10]).

Upper bounds for the joint spectral radius of sets of matrices $\Sigma = \{A_1, \ldots, A_m\}$ are usually derived from the following simple inequality:

$$\hat{\rho} \leq \max_{d_1, \ldots, d_k \in \{1, \ldots, m\}} \|A_{d_1} \cdots A_{d_k}\|^{1/k}, \tag{15}$$

which holds for every $k \geq 1$ and converges to $\hat{\rho}$ as $k \to \infty$. This, at least theoretically, gives arbitrarily sharp estimates for $\hat{\rho}$. 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 k. Faster convergence can be achieved by finding an appropriate norm. To do this we use the so-called ellipsoidal norm: $||A||_P = \max_{x \neq 0} \sqrt{\frac{x^T A^T P A x}{x^T P x}}$, where P is a positive definite matrix. This is the matrix norm induced by the vector norm $|x|_P = (x^T P x)^{1/2}$. The crucial idea is that the optimal P, for which the right-hand side in (15) for $k = 1$ is minimal, can be found by solving a simple semidefinite programming problem. This algorithm can be iterated using the relation $\rho(\Sigma^k) = \rho(\Sigma)^k$. In what follows we denote $\Sigma^k = \left\{A_{d_1} \cdots A_{d_k}, 1 \leq d_i \leq m, i = 1, \ldots, k\right\}$. Thus one can consider the set Σ^k as a new set of matrices, and approximate its joint spectral radius with the best possible ellipsoidal norm. In Appendix C we give an ellipsoidal norm such that each matrix in Σ^{14} has a norm smaller than 2.5186^{14}. This implies that $\hat{\rho} \leq 2.5186$, which gives $\beta < 1.3326$. Combining this with the inequality $\beta > 1.3322$ we complete the proof of the bounds for β in Theorem 6.

We have not been able to improve the lower bound of Eq. (14). However, the upper bound we obtain is very close to this lower bound, and the upper bounds obtained with an ellipsoidal norm for Σ^k get closer and closer to this value when k increases. Moreover, it has already been observed that for many sets of matrices, for which the joint spectral radius is known exactly, and, in particular, of matrices with nonnegative integer entries, there always is a product that achieves the joint spectral radius, i.e., a product $A \in \Sigma^t$ such that $\hat{\rho} = \rho(A)^{(1/t)}$ [3,13,14]. For these reasons, we conjecture that the exponent β is actually equal to the lower bound.

3.2. Computation of α and the joint spectral subradius

An upper bound for $\check{\rho}(A_0, A_1)$ can be obtained using inequalities (9) for $k = 1$ and $d_1 = 0$. We have

$$\alpha = \log_2(\check{\rho}) \le \log_2\left(\rho(A_0)\right) = 1.275\ldots. \tag{16}$$

This bound for α was first derived in [10]. It is, however, not optimal. Taking the product $A_1^{10}A_0$ (i.e., $k = 11$ in Inequality (9)), we get a better estimate:

$$\alpha \le \log_2\left[\rho(A_1^{10}A_0)^{1/11}\right] = 1.2735\ldots. \tag{17}$$

One can verify numerically that this product gives the best possible upper bound among all the matrix products of length $k \le 14$.

We now estimate α from below. The problem of approximating the joint spectral subradius is NP-hard even for nonnegative matrices [6]. Moreover, the undecidability result for the so-called "morality problem" [7] shows that there are no algorithms that would approximate the joint spectral subradius equally well for all matrices. This, however, does not mean that such algorithms cannot be constructed for special matrices, or for some classes of matrices. To the best of our knowledge, no algorithm has ever been proposed to compute $\check{\rho}$, even in particular cases. Here we propose two new algorithms that appear to work well with our matrices A_0, A_1. We first consider the particular case of nonnegative matrices. As we observed above, for any k we have $\check{\rho}(\Sigma^k) = \check{\rho}(\Sigma)^k$. 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.

Theorem 7. *Let Σ be a set of nonnegative matrices that do not have any common zero column. If, for some $r \in \mathbb{R}_+$, $s \le t \in \mathbb{N}$, there exists $x \in \mathbb{R}^d$ satisfying the following system of linear inequalities:*

$$\begin{aligned} B(Ax - rx) &\ge 0, \quad \forall B \in \Sigma^s,\ \forall A \in \Sigma^t, \\ x &\ge 0, \quad (x, \mathbf{1}) = 1, \end{aligned} \tag{18}$$

then $\check{\rho}(\Sigma) \ge r^{1/t}$.

Proof. Let x be a solution of (18). Let us consider a product of matrices $A_k \ldots A_1 \in \Sigma^{kt} : A_i \in \Sigma^t$. We show by induction on k that $A_k \ldots A_1 x \ge r^{k-1} A_k x$.

For $k = 2$ we have $A_2(A_1x - rx) = CB(A_1x - rx) \ge 0$, with $B \in \Sigma^s$, $C \in \Sigma^{t-s}$.

Let $k > 2$. By the inductive assumption, $A_{k-1} \ldots A_1 x \ge r^{k-2} A_{k-1} x$; then, multiplying by A_k, we obtain $A_k \ldots A_1 x \ge r^{k-2} A_k A_{k-1} x$. Now, as we have just seen, $A_k A_{k-1} x \ge r A_k x$. Thus,

$$||A_k \ldots A_1|| = \mathbf{1}^{\mathrm{T}} A_k \ldots A_1 \mathbf{1} \ge r^{k-1} \mathbf{1}^{\mathrm{T}} A_k x \ge r^k C,$$

where $C = (\min_k \mathbf{1}^{\mathrm{T}} A_k x)/r > 0$. The last inequality holds because $A_k x = 0$, together with the first inequality in (18), imply that $-rBx = 0$ for all $B \in \Sigma^s$, which means 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. $\square$

The linear programming problem (18) can be solved with optimization techniques for s, t not too large. We found a solution with the following values for the parameters: $r = 2.41^{16}$, $t = 16$, $s = 6$. As a result, we get the following lower bound: $\alpha \ge \frac{1}{t} \log_2 r > 1.2690$. The corresponding vector x is given in Appendix D. This completes the proof of Theorem 6.

Theorem 7 deals with nonnegative matrices, which suffices for our problem, because our matrices A_0, A_1 have their entries in $\{0, 1, 2\}$. However, we would like to extend our techniques to arbitrary matrices. This is possible, and the idea is to lift the matrices to a larger vector space, so that all the matrices share an invariant cone. This kind of lifting is rather classical and is known under several names in the literature as for instance semidefinite lifting or symmetric algebras [4,20,22]. The idea is to consider the matrices $A_i \in \Sigma$ as linear operators acting on the cone of positive semidefinite matrices S as $S \to A_i^{\mathrm{T}} S A_i$. For more on the semidefinite lifting and its application to joint spectral quantities computation, see [12]. It is not difficult to prove 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 $A \succeq 0 \Leftrightarrow \forall y, y^{\mathrm{T}} A y \ge 0$.

Theorem 8. *Let Σ be a set of matrices in $\mathbb{R}^{d \times d}$ and $s \le t \in \mathbb{N}$. Suppose that there are $r > 0$ and a symmetric matrix $S \succeq 0$ such that*

$$\begin{aligned} &B^{\mathrm{T}}(A^{\mathrm{T}} S A - rS)B \succeq 0 \quad \forall A \in \Sigma^t, B \in \Sigma^s \\ &S \succ 0 \end{aligned} \tag{19}$$

then $\check{\rho}(\Sigma) \ge r^{1/2t}$.

Proof. The proof is formally similar to the previous one. Let S be a solution of (19). We denote by M_k the product $A_1 \dots A_k, A_i \in \Sigma^t$. It is easy to show by induction that $M_k^{\mathrm{T}} S M_k \succeq r^{k-1}(A_k^{\mathrm{T}} S A_k)$. This is obvious for $k = 2$ for similar reasons as in the previous theorem, and for $k > 2$, if, by induction,

$$\forall y, \quad y^{\mathrm{T}} M_{k-1}^{\mathrm{T}} S M_{k-1} y \geq r^{k-2} y^{\mathrm{T}} A_{k-1}^{\mathrm{T}} S A_{k-1} y,$$

then, with $y = A_k x$, for all x,

$$x^{\mathrm{T}} M_k^{\mathrm{T}} S M_k x \geq r^{k-2} x^{\mathrm{T}} A_k^{\mathrm{T}} A_{k-1}^{\mathrm{T}} S A_{k-1} A_k x \geq r^{k-1} x^{\mathrm{T}} A_k^{\mathrm{T}} S A_k x.$$

Thus,

$$\sup \left\{ \frac{x^{\mathrm{T}} M_k^{\mathrm{T}} S M_k x}{x^{\mathrm{T}} S x} \right\} \geq r^{k-1} \sup \left\{ \frac{x^{\mathrm{T}} A_k^{\mathrm{T}} S A_k x}{x^{\mathrm{T}} S x} \right\}.$$

Finally, $||M_k||_S \geq r^{k/2} C$, where C is a constant. $\square$

For a given $r > 0$ the existence of a solution S can be established by solving the semidefinite programming problem (19), and the optimal r can be found by bisection in logarithmic time.

3.3. Computation 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:

$$\rho_1 = \lim_{k \to \infty} \left(2^{-k} \sum_{d_1, \dots, d_k} \|A_{d_1} \cdots A_{d_k}\| \right)^{1/k}.$$

For matrices with a common invariant cone we have $\rho_1 = \frac{1}{2}\rho(A_0 + A_1)$ [23]. Therefore, in our case $\rho_1 = \frac{1}{2}\rho(A_0 + A_1) = 2.479\ldots$. This exponent was first computed in [10], where it was shown that the value $\sum_{j=0}^{n-1} u_j$ is asymptotically equivalent to n^η, where $\eta = 1 + \log_2 \rho_1 = 2.310\ldots$. 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. Although the Lyapunov exponent has proved useful in many situations, very few algorithms exist to approximate it, to the best of our knowledge, except by application of Definition (12). See however [11] for an upper bound based on information theoretic techniques. We propose here an alternative way based on convex programming. It is easily seen that for any k the value $r_k = \left(\prod_{d_1, \dots, d_k} \|A_{d_1} \cdots A_{d_k}\|\right)^{\frac{1}{k2^k}}$ gives an upper bound for $\bar{\rho}$, that is $\bar{\rho} \leq r_k$ for any $k \in \mathbb{N}$. Since $r_k \to \bar{\rho}$ as $k \to \infty$, we see that this estimate can be arbitrarily sharp for large k. But for the dimension 20 this leads to extensive 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 k and $x \in \mathbb{R}^d$ we denote $p_k(x) = \left(\prod_{d_1, \dots, d_k} |A_{d_1} \cdots A_{d_k} x|\right)^{\frac{1}{2^k}}$ and $m_k = \sup_{x \geq 0, |x|=1} p_k(x)$.

Proposition 2. *Let A_0, A_1 be nonnegative matrices in $\mathbb{R}^d$. Then for any norm $|\cdot|$ and for any $k \geq 1$ we have $\bar{\rho} \leq (m_k)^{1/k}$.*

Proof. By Corollary 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, $p_{k+n}(x) \leq m_k p_n(x)$ for any $x \geq 0$ and for any $n, k \in \mathbb{N}$; therefore $p_{tk}(x) \leq (m_k)^t$. Thus, $\bar{\rho} \leq (m_k)^{1/k}$. $r_n \sim [p_n(x)]^{1/n}$. $\square$

Proposition 3. *Let A_0, A_1 be nonnegative matrices in $\mathbb{R}^d$ that do not have common invariant subspaces among the coordinate planes. If $\check{\rho} < \hat{\rho}$, then $\bar{\rho} < \rho_1$.*

Proof. Let v_* be the eigenvector of the matrix $\frac{1}{2}\left(A_0^{\mathrm{T}} + A_1^{\mathrm{T}}\right)$ corresponding to its largest eigenvalue ρ_1. Since the matrices have no common invariant coordinate planes, it follows that $v_* > 0$. Consider the norm $|x| = (x, v_*)$ on $\mathbb{R}_+^d$. Take some $k \geq 1$ and $y \in \mathbb{R}_+^d$, $|y| = (y, v_*) = 1$, such that $p_k(y) = m_k$. We have

$$\begin{aligned} m_k &= p_k(y) \leq 2^{-k} \sum_{d_1, \dots, d_k} |A_{d_1} \cdots A_{d_k} y| = 2^{-k} \sum_{d_1, \dots, d_k} \left(A_{d_1} \cdots A_{d_k} y, v_*\right) \\ &= \left(y, 2^{-k}\left(A_0^{\mathrm{T}} + A_1^{\mathrm{T}}\right)^k v_*\right) = \rho_1^k\left(y, v_*\right) = \rho_1^k. \end{aligned}$$

Thus, $m_k \leq \rho_1^k$, and the equality is possible only if all 2^k values $|A_{d_1} \cdots A_{d_k} y|$ are equal. Since $\check{\rho} < \hat{\rho}$, there must be k such that the inequality is strict. Hence, $m_k < \rho_1^k$ for some k, and by Proposition 2 we have $\bar{\rho} \leq (m_k)^{1/k} < \rho_1$. $\square$

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 3 the value $-\frac{1}{k}\log_2 m_k$ can be found as the solution of the following convex minimization problem with linear constraints:

$$\begin{aligned} \min \quad & -\frac{1}{k2^k \ln 2} \sum_{d_1,\ldots,d_k \in \{0,1\}} \ln\left(x, A_{d_1}^{\mathrm{T}} \cdots A_{d_k}^{\mathrm{T}} v_*\right) \\ \text{s.t.} \quad & x \geq 0, \quad (x, v_*) = 1. \end{aligned} \tag{20}$$

The optimal value of this optimization problem is equal to $-\frac{1}{k}\log_2 m_k$, which gives an upper bound for $\sigma = \log_2 \bar{\rho}$ (Proposition 2). Solving this problem for $k = 12$ we obtain $\sigma \leq 1.3098$.

To our knowledge, no nontrivial algorithm has been proposed to derive a nontrivial lower bound on $\bar{\rho}$. We now provide a theorem that allows us to derive such a lower bound. The idea is identical to the one used in Theorem 7, but transposed to the Lyapunov exponent.

Theorem 9. *Let Σ be a set of nonnegative matrices that do not have any common zero column, $s \leq t, s, t \in \mathbb{N}$. If for some numbers $r_i \in \mathbb{R}_+, 0 \leq i < 2^t$ there exists $x \in \mathbb{R}_+^d$ satisfying the following system of linear inequalities:*

$$\begin{aligned} B(A_i x - r_i x) &\geq 0, \quad \forall B \in \Sigma^s, A_i \in \Sigma^t, \\ x &\geq 0, \quad (x, \mathbf{1}) = 1, \end{aligned} \tag{21}$$

then $\bar{\rho}(\Sigma) \geq \prod_i {r_i}^{1/(t2^t)}$.

The proof is similar to the proof of Theorem 7 and is left to the reader. Also, a similar theorem can be stated for general matrices (not necessarily nonnegative), 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 D), with the values $s = 8, t = 16$ for the parameters, one gets a good lower bound for σ: $\sigma \geq 1.3005$.

4. Conclusions

The goal of this paper 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 us 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 the joint spectral subradius). The computational results we report in this paper 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 [15], and one could wonder how the joint spectral characteristics developed in this paper could represent both kinds of growth.

Acknowledgments

We would like to thank Prof. Stephen Boyd (Stanford University), Yuri Nesterov, and François Glineur (Université Catholique de Louvain) for their helpful suggestions on semidefinite programming techniques. This research was carried out during the visit of the second author to the Université Catholique de Louvain (Louvain-la-Neuve, Belgium). That author is grateful to the university for its hospitality.

We also would like to thank Prof. Wojciech Szpankowski for pointing out interesting references on the Lyapunov exponent.

We express our thanks to both anonymous referees for their attentive reading and many valuable remarks that helped us to improve the text. In particular, one of the referees found a mistake in the numerical part of the proof of Theorem 2 (Appendix B), after which we changed the construction of the proof.

The research of V.P. is supported by the grants MD-2195.2008.1, by the grant NSh-3233.2008.1 for leading scientific schools and by the RFBR grant No 08-01-00208. The research of R.J. and V.B. is partially supported by the "Communauté francaise de Belgique — Actions de Recherche Concertées" and by the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office. R.J. is an F.R.S.-FNRS Fellow.

Appendix A. Numerical values

We introduce the following auxiliary matrices. For the sake of simplicity our notations do not follow exactly those of [10].

$$D_1 = \begin{pmatrix} 0&0&0&0&0&0&0&1&2&1\\ 0&0&1&1&0&1&1&0&0&0\\ 0&0&0&0&0&0&0&1&1&0\\ 0&0&0&0&0&0&0&0&0&0\\ 1&2&0&0&1&0&0&0&0&0\\ 0&0&1&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0 \end{pmatrix}, \qquad B_1 = \begin{pmatrix} 0&0&0&0&0&0&0&1&2&1\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&1&0&0&0\\ 0&0&1&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0 \end{pmatrix},$$

$$C_1 = \begin{pmatrix} 0&0&0&0&0&0&0&2&4&2\\ 0&0&1&1&0&1&1&0&0&0\\ 0&0&0&0&0&1&1&1&1&0\\ 0&0&1&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&1&0&0&1&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&2&0&0&0&0\\ 0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0 \end{pmatrix},$$

$$B_2 = \begin{pmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&1&1\\ 0&1&1&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 1&0&0&0&0\\ 0&0&0&0&0 \end{pmatrix}, \qquad C_2 = \begin{pmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&2&0\\ 0&1&0&0&0\\ 0&0&0&0&0 \end{pmatrix}, \qquad C_4 = \begin{pmatrix} 0&1&1&1&1\\ 0&0&0&1&1\\ 0&1&1&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0 \end{pmatrix}.$$

Now, defining

$$F_0 = \left(\begin{array}{cc|cc} C_1 & 0_{10\times 10} & C_2 & 0_{10\times 5}\\ D_1 & B_1 & 0_{10\times 5} & B_2\\ \hline 0_{5\times 10} & 0_{5\times 10} & C_4 & 0_{5\times 5}\\ 0_{5\times 10} & 0_{5\times 10} & 0_{5\times 5} & 0_{5\times 5} \end{array}\right), \qquad F_1 = \left(\begin{array}{cc|cc} D_1 & B_1 & 0_{10\times 5} & B_2\\ 0_{10\times 10} & C_1 & 0_{10\times 5} & C_2\\ \hline 0_{5\times 10} & 0_{5\times 10} & 0_{5\times 5} & 0_{5\times 5}\\ 0_{5\times 10} & 0_{5\times 10} & 0_{5\times 5} & C_4 \end{array}\right), \tag{A.1}$$

$$\begin{aligned} w &= (1, 2, 2, 2, 1, 2, 2, 1, 2, 1, 0_{1\times 20})^{\mathrm{T}},\\ y_8 &= (4, 4, 4, 2, 0, 2, 2, 0, 2, 0, 6, 4, 4, 2, 4, 2, 0, 4, 2, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0)^{\mathrm{T}},\\ y_9 &= (6, 4, 4, 2, 4, 2, 0, 4, 2, 2, 8, 4, 4, 2, 0, 4, 4, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{\mathrm{T}},\\ y_{10} &= (8, 4, 4, 2, 0, 4, 4, 4, 0, 0, 8, 4, 6, 4, 8, 2, 0, 4, 2, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{\mathrm{T}},\\ y_{11} &= (8, 4, 6, 4, 8, 2, 0, 4, 2, 4, 8, 6, 6, 2, 0, 2, 6, 4, 2, 0, 0, 0, 0, 0, 0, 2, 0, 2, 2, 0)^{\mathrm{T}},\\ y_{12} &= (8, 6, 6, 2, 0, 2, 6, 4, 2, 0, 10, 6, 4, 4, 8, 2, 0, 4, 2, 4, 2, 0, 2, 2, 0, 0, 0, 0, 0, 0)^{\mathrm{T}},\\ y_{13} &= (10, 6, 4, 4, 8, 2, 0, 4, 2, 4, 12, 6, 4, 4, 0, 6, 6, 4, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{\mathrm{T}},\\ y_{14} &= (12, 6, 4, 4, 0, 6, 6, 4, 2, 0, 10, 6, 8, 6, 12, 4, 0, 0, 4, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{\mathrm{T}},\\ y_{15} &= (10, 6, 8, 6, 12, 4, 0, 0, 4, 4, 8, 10, 6, 6, 0, 4, 8, 4, 4, 0, 0, 0, 0, 0, 0, 2, 2, 0, 0, 0)^{\mathrm{T}}, \end{aligned}$$

and introducing the recurrence relation

$$y_{2n} = F_0 y_n, \qquad y_{2n+1} = F_1 y_n, \quad n \geq 8$$

one has the relation [10]

$$u_{n+1} = w^{\mathrm{T}} y_n. \tag{A.2}$$

We finally introduce two new matrices in $\mathbb{R}^{20\times 20}$ that rule the asymptotics of u_n:

$$A_0 = \begin{pmatrix} C_1 & 0_{10\times 10} \\ D_1 & B_1 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} D_1 & B_1 \\ 0_{10\times 10} & C_1 \end{pmatrix}. \tag{A.3}$$

Appendix B. Proof of Theorem 2

In this appendix we give a proof of Theorem 2.

Outline of the proof. We 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 $\mathbb{R}^{20}$) in its interior, to enable us to apply Lemma 1 and Corollary 1. That is why the positive orthant $\mathbb{R}^{20}_+$ is not appropriate, because all those vectors are on its boundary. So, we have to construct a wider invariant cone K. This is done in Lemma 4.

Then, invoking Lemma 1 and Corollary 1 we show that the products $F(n) = F_{d_1} \cdots F_{d_k}$ are asymptotically equivalent to their corresponding product $A(n) = A_{d_1} \cdots A_{d_k}$ (Lemma 5).

We then shed some light on the vectors z_n: their norms can be considered as norms of the products $A(n)$ (Lemma 7).

We finally show in Lemma 8 that $\|A_{d_1} \cdots A_{d_k}\|$ is equivalent to $\|A_{d_1} \cdots A_{d_{k-4}}\|$.

We end with the proof of Theorem 2 that puts all this together.

Let us first establish some special properties of the matrices A_0, A_1. Consider the following sets:

$$P = \left\{x \in \mathbb{R}^{20}, x \geq 0, \quad x_i > 0 \text{ for all } i \notin \{5, 10, 17, 18\}\right\},$$

$$Q = \left\{x \in \mathbb{R}^{20}, x \geq 0, \quad x_i > 0 \text{ for all } i \notin \{7, 8, 15, 20\}\right\}.$$

Let $S = P \cup Q$. For any $\varepsilon \geq 0$ let $p_\varepsilon \in \mathbb{R}^{20}$, $(p_\varepsilon)_i = -\varepsilon$ for $i \in \{5, 10, 17, 18\}$ and $(p_\varepsilon)_i = 1$ otherwise; also let $q_\varepsilon \in \mathbb{R}^{20}$, $(q_\varepsilon)_i = -\varepsilon$ for $i \in \{7, 8, 15, 20\}$ and $(q_\varepsilon)_i = 1$ otherwise. It is easy to verify by direct calculation that

$$\begin{array}{ll} A_0\, p_0 \in P, ; & A_0\, q_0 \in P, \\ A_1\, p_0 \in Q & A_1\, q_0 \in Q. \end{array} \tag{B.1}$$

and, moreover, for small $\varepsilon > 0$ (for instance, for $\varepsilon = 1/4$) we even have

$$\begin{array}{l} A_0\, p_\varepsilon \in P, \\ A_1\, p_\varepsilon \in Q. \end{array} \tag{B.2}$$

Unfortunately, the vector q_ε does not possess this property: $A_0 q_\varepsilon$ and $A_1 q_\varepsilon$ are both not in $\mathbb{R}^{20}_+$ for any $\varepsilon > 0$. That is why we make an extra construction. For this we need two technical lemmas. The first one is verified by computer calculation.

Lemma 2. *For $i = 0, 1$ the largest eigenvalue $\lambda_i = \rho(A_i)$ of the matrix A_i has multiplicity one. If v_0 and v_1 are the corresponding Perron–Frobenius eigenvectors of A_0 and A_1 respectively, then $v_0 \in P$ and $v_1 \in Q$. Moreover, $(v_0)_j = 0, j \in \{5, 10, 17\}$ and $(v_1)_k = 0$, $k \in \{7, 8, 15, 20\}$.*

The second fact is well known (see for instance [1]).

Lemma 3. *For any matrix A in $\mathbb{R}^d$ and for any $\mu > \rho(A)$ there is a convex compact set $M \subset \mathbb{R}^d$ such that $0 \in \operatorname{int} M$ and $AM \subset \mu M$.*

Let L be the invariant subspace of the operator A_1, which complements the Perron–Frobenius eigenvector v_1 to the entire space $\mathbb{R}^{20}$. That is, L is the linear subspace generated by the 19 last columns of the matrix T, with $J = T^{-1}A_1T$ the canonical Jordan form of A_1. Thus, $\dim L = 19$ and the spectral radius of the operator A_1 restricted to L is smaller than $\rho(A_1)$. By Lemma 3 there is a convex compact set $M \subset L$ that contains the origin as an interior point and such that $A_1 M \subset \rho(A_1)\, M$. For an arbitrary $\delta > 0$ we denote $R_\delta = \left\{ s(v_1 + \delta\, x), x \in M, s \in \mathbb{R}_+ \right\}$. For any $\delta > 0$ this is a convex closed pointed cone, $v_1 \in \operatorname{int} R_\delta$ and $A_1 R_\delta \subset R_\delta$.

Take now $\varepsilon = \frac{1}{4}$, and for any $\delta > 0$ consider the following convex closed cone: $K = \{u + y + tp_\varepsilon,\, u \in \mathbb{R}^{20}_+, y \in R_\delta,\; t \geq 0\}$. For small values of δ this cone is pointed. Indeed, since $v_1 \in Q$, it follows that if $\delta > 0$ is small enough, then for any $y \in R_\delta, y \neq 0$ we have $y_i > 0$ for all $i \notin \{7, 8, 15, 20\}$. Thus, any nonzero $y \in R_\delta$ has at least 16 positive entries, and the same holds for the vector tp_ε. Hence, any nonzero vector $x \in K$ has at least 12 positive entries; therefore $-x \notin K$. Thus, K is pointed.

Lemma 4. *We have $A_i S \subset S$ for $i = 0, 1$ and $S \subset \operatorname{int} K$ for any $\delta > 0$. Moreover, if δ, ϵ are small enough, then $A_i K \subset K$ for $i = 0, 1$.*

Proof. For any $x \in \mathbb{R}^d$ that has at least one positive coordinate we write $x_{\min} = \min_{x_i>0} x_i$ and $x_{\max} = \max x_i$. First we show that $A_i S \subset S,\ i = 0, 1$. Assume $x \in P$ (the proof for the case $x \in Q$ is literally the same). Since $x \ge x_{\min} p_0$ we have $A_0 x \ge x_{\min} A_0 p_0 \in P$ (assertion (B.1)); therefore $A_0 x \in P$. In the same way $A_1 x \ge x_{\min} A_1 p_0 \in Q$, and so $A_1 x \in Q$.

Now let us show that $S \subset \operatorname{int} K$ for any $\delta > 0$. For arbitrary $x \in P$ and $h \in \mathbb{R}^{20}$ such that $|h|_\infty < \frac{1}{8} x_{\min}$ we have $x + h \ge \frac{1}{2} x_{\min}\, p_\varepsilon$; therefore $x + h \in K$, which proves that $x \in \operatorname{int} K$. If $x \in Q$, then we take any element $z \in R_\delta$, for which $z_{\max} = 1$ and $z_i < -\alpha$, for $i \in \{7, 8, 15, 20\}$, where $\alpha \in (0, 1)$. It exists, because $v_1 \in \operatorname{int} R_\delta$ and $(v_1)_i = 0,\ i \in \{7, 8, 15, 20\}$ (Lemma 2). Then for any $h \in \mathbb{R}^{20}$ such that $|h|_\infty < \frac{1}{2} x_{\min}\, \alpha$ we have $x + h \ge \frac{1}{2} x_{\min}\, z$; therefore $x + h \in K$, which proves that $x \in \operatorname{int} K$.

It remains to prove that $A_i K \subset K,\ i = 0, 1$, whenever δ is small. Let $x \in K : x = u + y + t p_\varepsilon,\ u \in \mathbb{R}^{20}_+,\ y \in R_\delta,\ t \ge 0$. Since $A_i u \ge 0$ and $A_i p_\varepsilon \ge 0$ (this follows from (B.2)) we see that $A_i u$ and $A_i p_\varepsilon$ are both in K. So, we need to show that $A_i y \in R_\delta$. This is obvious for $i = 1$, because $A_1 R_\delta \subset R_\delta$. Let us now prove that $A_0 R_\delta \subset K$. If δ is sufficiently small, then for all $y \in R_\delta,\ y \neq 0$ one has $y_i > 0$ for $i \notin \{7, 8, 15, 20\}$. Without loss of generality it may be assumed that $y_{\max} = 1$; we also normalize v_1 by the same condition: $(v_1)_{\max} = 1$. By Lemma 2 we have $v_1 \in Q$, and therefore $A_0 v_1 \in P$. Since $y \to v_1$ as $\delta \to 0$, it follows that $(A_0 y)_i > 0$ for $i \notin \{5, 10, 17, 18\}$ and, moreover, negative entries of $A_0 y$ (if they exist) are less by modulo than $\frac{1}{4}(A_0 y)_{\min}$, whenever δ is small enough. This yields that $A_0 y \ge (A_0 y)_{\min}\, p_\varepsilon$ and hence $A_0 y \in K$, which completes the proof. □

Corollary 5. *For any $x \in S$ and for any sequence $d_1, \ldots, d_k$ we have*

$$A_{d_1} \cdots A_{d_k} x \in \operatorname{int} K.$$

Lemma 5. *Suppose $n \ge 1$ and consider the binary expansion $d_k \ldots d_1$ of the number $n - 1$. We define $A(n) = A_{d_1} \cdots A_{d_k}$ and similarly for $F(n)$. Then, for any matrix norm one has*

$$\|F(n)\| \asymp \|A(n)\|.$$

Proof. Since all matrix norms are equivalent, we can choose any norm. Obviously $\|F_{d_1} \cdots F_{d_k}\|_1 \ge \|A_{d_1} \cdots A_{d_k}\|_1$ (because A_i are submatrices of F_i), hence it remains to prove the opposite inequality: there is a positive constant C such that $\|A_{d_1} \cdots A_{d_k}\|_1 \ge C \|F_{d_1} \cdots F_{d_k}\|_1$ for all $k \in \mathbb{N}$ and $d_1, \ldots, d_k$. We consider the case $d_k = 1$; the proof for the other case is similar. Let $m \le k - 1$ be the largest number such that $d_m = 0$. If the sequence has no zero, we fix $m = 0$. Let A_i, H_i, and R_i denote respectively the upper left, the upper right and the lower right corners of the matrix F_i in block representation (A.1). Then the product $F_{d_1} \ldots F_{d_k}$ has the following form: the left upper block is $A_{d_1} \cdots A_{d_k}$, the left lower block is zero, the right lower block is $R_{d_1} \cdots R_{d_k}$, and finally, the right upper block is

$$\sum_{p=1}^{k} \left(\prod_{j=1}^{p-1} A_{d_j} \right) \cdot H_{d_p} \cdot \left(\prod_{j=p+1}^{k} R_{d_j} \right). \tag{B.3}$$

By convention, the product over an empty set is one. Since $R_0 R_1 = 0$, the right lower block is zero, except when $m = 0$, in which case it is R_1^k. Block (B.3) becomes $\sum_{p=m}^{k} \left(\prod_{j=1}^{p-1} A_{d_j} \right) H_{d_p} R_1^{k-p}$, whose norm can be estimated from above as

$$H \sum_{p=m}^{k} \left\| \prod_{j=1}^{p-1} A_{d_j} \right\|_1 \cdot \left\| R_1^{k-p} \right\|_1, \tag{B.4}$$

where $H = \max\{\|H_0\|_1, \|H_1\|_1\}$. It was shown in [10] that the sum of entries of the matrix R_i^l does not exceed $C2^l$ for any $i = 0, 1$ and $l \ge 1$, where $C > 0$ is a constant. Hence $\|R_1^{k-p}\|_1 \le C2^{k-p}$. Thus,

$$\|F_{d_1} \cdots F_{d_k}\|_1 \le \|A_{d_1} \cdots A_{d_k}\|_1 + HC \sum_{p=m}^{k} 2^{k-p} \left\| \prod_{j=1}^{p-1} A_{d_j} \right\|_1 + C\, 2^k. \tag{B.5}$$

We have seen that $\check{\rho}(\{A_0, A_1\}) > 2$, so that $C2^k \le C' \|A_{d_1} \cdots A_{d_k}\|_1$. On the other hand, for any $p \ge m$ we have $|A_{d_1} \cdots A_{d_k} \mathbf{1}| = |A_{d_1} \cdots A_{d_{p-1}} (A_{d_p} A_1^{k-p} \mathbf{1})|$, where $\mathbf{1}$ is the vector of ones. By Corollary 5 the vectors $A_{d_p} A_1^r \mathbf{1}$ belong to int$K$ for all $r \in \mathbb{N}$. Moreover, the vector $A_1^r \mathbf{1} / |A_1^r \mathbf{1}|$ converges to v_1 (the Perron–Frobenius eigenvector of A_1) as $r \to \infty$. Since $v_1 \in S$ (Lemma 2), Lemma 4 yields $v_1 \in \operatorname{int} K$. Therefore there is a constant $C_1 > 0$ such that $\gamma(A_{d_p} A_1^r \mathbf{1}) \ge C_1$ for all $r \in \mathbb{N}$. Let us recall that the value $\gamma(x)$ is defined in Lemma 1, and is continuous in x in the interior of the cone K. Applying now Lemma 1 for $x = A_{d_p} A_1^{k-p} \mathbf{1}$, we get

$$\begin{aligned} \|A_{d_1} \cdots A_{d_k}\| &\ge C_2 |A_{d_1} \cdots A_{d_{p-1}} (A_{d_p} A_1^{k-p} \mathbf{1})| \\ &\ge C_2 C_1 \|A_{d_1} \cdots A_{d_{p-1}}\| \cdot |A_{d_p} A_1^{k-p} \mathbf{1}| \ge C_3 \lambda^{k-p} \|A_{d_1} \cdots A_{d_{p-1}}\|, \end{aligned}$$

where $\lambda = \rho(A_1)$ (by the same reasoning we have $|A_{d_p} A_1^r \mathbf{1}| \ge \gamma(A_1^r \mathbf{1}) ||A_{d_p}|| \cdot |A_1^r \mathbf{1}| \ge C \lambda^r$; indeed, $A_1^r \mathbf{1} \in$ intK, and thus $||A^{d_p} A_1^r \mathbf{1}|| \asymp ||A^{d_p}|| \cdot |A_1^r \mathbf{1}|$). Thus, $\|A_{d_1} \cdots A_{d_{p-1}}\| \le C_3^{-1} \lambda^{p-k} \|A_{d_1} \cdots A_{d_k}\|$ for all $p \le k$. Substituting this in (B.5) and taking into account that $\frac{2}{\lambda} < 1$ (because $\lambda = \rho(A_1) > 2.42$) we take the sum of the geometrical progression and get $\|F_{d_1} \cdots F_{d_k}\|_1 \le C_4 \|A_{d_1} \cdots A_{d_k}\|_1$, where C_4 is some constant. This concludes the proof. □

Lemma 6. *For any n we have $u_{n+1} \leq 2u_n$.*

Proof. If a word of length $n+1$ is overlap free then so is its prefix of length n. On the other hand, at most two overlap-free words of length $n+1$ have the same prefix of length n. $\square$

Lemma 7. *Let the vectors $y_m \in \mathbb{R}^{30}$ be the solution of the recurrence equation (4), and $z_m \in \mathbb{R}^{20}$ be the vector with the first 20 entries of y_m. We have $z_m \in S$ for each $m = 64, \ldots, 127$.*

Proof. The proof is by direct calculation. $\square$

Lemma 8. *Suppose $n \in \mathbb{N}$ and $d_k \ldots d_1$ is the binary expansion of $n-1$; then $\|A(n)\| \asymp \|A'(n)\|$, where $A(n) = A_{d_1} \cdots A_{d_k}$ and $A'(n) = A_{d_1} \cdots A_{d_{k-4}}$.*

Proof. The inequality $\left\|A_{d_1} \cdots A_{d_k}\right\| \leq C \left\|A_{d_1} \cdots A_{d_{k-4}}\right\|$ is obvious by submultiplicativity of the norm. For the other direction, we have

$$\left\|A_{d_1} \cdots A_{d_k}\right\| \asymp \left|A_{d_1} \cdots A_{d_k}\mathbf{1}\right| \asymp \left|A_{d_1} \cdots A_{d_{k-4}}(A_{d_{k-3}} \ldots A_{d_k}\mathbf{1})\right|. \tag{B.6}$$

Corollary 5 yields $A_{d_{k-3}} \ldots A_{d_k}\mathbf{1} \in \operatorname{int} K$ for all $d_{k-3}, \ldots, d_k \in \{0, 1\}$. Applying now Lemma 1 we get

$$\delta = \min_{d_{k-3},\ldots,d_k \in \{0,1\}} \gamma(A_{d_{k-3}} \ldots A_{d_k}\mathbf{1}) > 0.$$

Therefore, for some $C_1 > 0$,

$$\begin{aligned} C_1\left\|A_{d_1} \cdots A_{d_{k-4}}\right\| &\leq \delta \left\|A_{d_1} \cdots A_{d_{k-4}}\right\| \cdot \left|A_{d_{k-3}} \cdots A_{d_k}\mathbf{1}\right| \\ &\leq \left|A_{d_1} \cdots A_{d_k}\mathbf{1}\right|. \end{aligned} \tag{B.7}$$

Combining this with (B.6) we get $\left\|A_{d_1} \cdots A_{d_{k-4}}\right\| \leq C_2 \left\|A_{d_1} \cdots A_{d_k}\right\|$. $\square$

We are now able to prove Theorem 2.

Proof of Theorem 2. Let g be the vector of $\mathbb{R}^{30}$, whose first 20 entries are ones and the last 10 entries are zeros. Let also $m = d_{k-3} + d_{k-2}2 + 2^2 d_{k-1} + 2^3 d_k$. Since $w \leq 2g$, we have

$$\begin{aligned} u_n &\leq 2\big(y_{n-1}, g\big) = 2\big(F_{d_1} \cdots F_{d_{k-4}} y_m, g\big) \\ &\leq C_0 \|F_{d_1} \cdots F_{d_{k-4}}\| \asymp \|A_{d_1} \cdots A_{d_{k-4}}\|, \end{aligned} \tag{B.8}$$

where C_0 does not depend on n (the first two relations are direct from fundamental assertions (A.2), the third relation comes from the fact that y_m and g are bounded, and the last equivalence is by Lemma 5). Combining Lemma 8 and (B.8) gives $u_n \leq C_3 \left\|A_{d_1} \cdots A_{d_k}\right\|$.

Let us now prove the opposite inequality. Lemma 6, together with the fact that, by construction, the first ten entries of y_n are equal to the entries $11, \ldots, 20$ of y_{n-1}, implies that $u_n \geq \frac{1}{3}\big(u_n + u_{n+1}\big) \geq \frac{1}{6}\big(y_{n-1}, g\big)$. Furthermore, for $n > 2^7$ we have $\big(y_{n-1}, g\big) = \big(F_{d_1} \cdots F_{d_{k-7}} y_l, g\big)$, where $l = \sum_{j=0}^{6} d_{k-6+j} 2^j$. Thus,

$$u_n \geq \frac{1}{6}\big(F_{d_1} \cdots F_{d_{k-7}} y_l, g\big). \tag{B.9}$$

On the other hand, defining $z_l \in \mathbb{R}^{20}$ as the vector with the first 20 entries of y_l: $\big(F_{d_1} \cdots F_{d_{k-7}} y_l, g\big) \geq \big(A_{d_1} \cdots A_{d_{k-7}} z_l, \mathbf{1}\big) = \left|A_{d_1} \cdots A_{d_{k-7}} z_l\right|_1$. By Lemma 7 we have $z_l \in \operatorname{int} K$ for all $l \in \{64, \ldots, 127\}$, and we can define $h = \min_{64 \leq l \leq 127} \gamma(z_l) > 0$ such that $\left|A_{d_1} \cdots A_{d_{k-7}} z_l\right|_1 \geq hC_4 \|A_{d_1} \cdots A_{d_{k-7}}\|$, where $C_4 = \min_{64 \leq l \leq 127} |z_l|_1$. Combining this with (B.9), we obtain

$$u_n \geq C_5 \|A_{d_1} \cdots A_{d_{k-7}}\|. \tag{B.10}$$

Now, by submultiplicativity of the norm,

$$u_n \geq C_6 \|A_{d_1} \cdots A_{d_k}\|. \quad \square$$

Appendix C. The ellipsoidal norm

Define

$$P_1 = \begin{pmatrix} 313 & 75 & 23 & 33 & -4 & -3 & 3 & 4 & 37 & 03 \\ 75 & 577 & 100 & 63 & 184 & 350 & 163 & -58 & 138 & 50 \\ 23 & 100 & 599 & 113 & 4 & 292 & 42 & 101 & 82 & 08 \\ 33 & 63 & 113 & 485 & 46 & 135 & 108 & 20 & 69 & 10 \\ -4 & 184 & 4 & 46 & 364 & 235 & 226 & 44 & 89 & -12 \\ -3 & 350 & 292 & 135 & 235 & 1059 & 384 & 95 & 337 & 61 \\ 3 & 163 & 42 & 108 & 226 & 384 & 590 & 27 & 174 & 92 \\ 4 & -58 & 101 & 20 & 44 & 95 & 27 & 386 & 148 & -17 \\ 37 & 138 & 82 & 69 & 89 & 337 & 174 & 148 & 575 & 86 \\ 3 & 50 & 8 & 10 & -12 & 61 & 92 & -17 & 86 & 423 \end{pmatrix},$$

$$P_2 = \begin{pmatrix} -104 & -17 & -181 & -4 & -58 & -51 & -49 & -8 & -27 & -9 \\ -111 & -224 & -82 & -147 & -99 & -303 & -167 & -113 & -169 & -66 \\ -22 & -164 & -158 & -50 & -85 & -72 & -54 & -185 & -35 & -34 \\ -2 & -136 & -52 & -90 & -107 & -146 & -92 & -16 & -113 & -11 \\ -46 & -170 & -130 & -91 & -6 & -112 & -239 & -70 & -121 & 3 \\ -59 & -264 & -274 & -174 & -310 & -376 & -280 & -44 & -273 & -74 \\ -14 & -193 & -116 & -108 & -223 & -179 & -117 & -113 & -120 & -98 \\ -63 & 21 & 17 & -34 & 32 & -76 & 2 & -52 & -31 & -14 \\ -74 & -159 & -47 & -67 & -122 & -173 & -116 & -53 & -68 & -16 \\ 13 & -57 & -36 & -32 & -4 & -61 & -90 & -14 & -69 & 4 \end{pmatrix},$$

$$P_4 = \begin{pmatrix} 291 & 83 & -16 & 48 & -13 & -44 & 6 & 17 & 75 & 11 \\ 83 & 473 & 136 & 28 & 117 & 198 & 174 & 6 & 100 & 37 \\ -16 & 136 & 466 & 104 & 65 & 249 & 118 & 65 & 125 & 14 \\ 48 & 28 & 104 & 476 & 51 & 80 & 76 & 51 & 37 & 18 \\ -13 & 117 & 65 & 51 & 328 & 195 & 194 & 76 & 67 & -2 \\ -44 & 198 & 249 & 80 & 195 & 648 & 162 & 114 & 138 & 68 \\ 6 & 174 & 118 & 76 & 194 & 162 & 567 & 76 & 122 & 65 \\ 17 & 6 & 65 & 51 & 76 & 114 & 76 & 387 & 112 & -10 \\ 75 & 100 & 125 & 37 & 67 & 138 & 122 & 112 & 556 & 42 \\ 11 & 37 & 14 & 18 & -2 & 68 & 65 & -10 & 42 & 438 \end{pmatrix},$$

$$P = \begin{pmatrix} P_1 & P_2 \\ P_2^{\mathrm{T}} & P_4 \end{pmatrix}.$$

One can check that $P \succ 0$, and that

$$A^t P A - (2.5186)^{28} P \prec 0, \quad \forall A \in \Sigma^{14}.$$

As explained in [5], this suffices to prove that $\rho(\Sigma) \leq 2.5186$.

Appendix D. The vector *x*

Define

$$x = \frac{1}{999}\left(153, 0, 60, 0, 50, 56, 99, 0, 58, 1, 157, 81, 0, 113, 0, 72, 0, 99, 0, 0\right)^{\mathrm{T}}.$$

Then one has the relation

$$\begin{aligned} B(Ax - rx) &\geq 0, \quad \forall B \in \Sigma^6, \forall A \in \Sigma^{16}, \\ x &\geq 0, \quad (x, \mathbf{1}) = 1 \end{aligned} \tag{D.1}$$

with $r = 2.41^{16}$. This proves the inequality $\check{\rho}(\Sigma) \geq 2.41$.

References

[1] M.A. Berger, Y. Wang, Bounded semigroups of matrices, Linear Algebra and its Applications 166 (1992) 21–27.
[2] J. Berstel, Growth of repetition-free words—a review, Theoretical Computer Science 340 (2) (2005) 280–290.
[3] V.D. Blondel, R. Jungers, V. Protasov, On the complexity of computing the capacity of codes that avoid forbidden difference patterns, IEEE Transactions on Information Theory 52 (11) (2006) 5122–5127.
[4] V.D. Blondel, Y. Nesterov, Computationally efficient approximations of the joint spectral radius, SIAM Journal of Matrix Analysis 27 (1) (2005) 256–272.

[5] V.D. Blondel, Y. Nesterov, J. Theys, On the accuracy of the ellipsoid norm approximation of the joint spectral radius, Linear Algebra and its Applications 394 (1) (2005) 91–107.
[6] V.D. Blondel, 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 (1997) 31–40.
[7] V.D. Blondel, J.N. Tsitsiklis, A survey of computational complexity results in systems and control, Automatica 36 (9) (2000) 1249–1274.
[8] S. Brlek, Enumeration of factors in the Thue–Morse word, Discrete Applied Mathematics 24 (1989) 83–96.
[9] A. Carpi, Overlap-free words and finite automata, Theoretical Computer Science 115 (2) (1993) 243–260.
[10] J. Cassaigne, Counting overlap-free binary words, in: STACS 93, Lecture Notes in Computer Science, 665, 1993, pp. 216–225.
[11] R. Gharavi, V. Anantharam, An upper bound for the largest Lyapunov exponent of a Markovian product of nonnegative matrices, Theoretical Computer Science 332 (2005) 543–557.
[12] R.M. Jungers, The joint spectral radius, theory and applications, in: Springer-Verlag (Ed.), Lecture Notes in Information and Computer Sciences, vol. 385, Berlin, 2009 (in press).
[13] R.M. Jungers, V.D. Blondel, On the finiteness property for rational matrices, Linear Algebra and its Applications 428 (10) (2008) 2283–2295.
[14] R.M. Jungers, V. Protasov, V.D. Blondel, Efficient algorithms for deciding the type of growth of products of integer matrices, Linear Algebra and its Applications 428 (10) (2008) 2296–2311.
[15] J. Karhumäki, J. Shallit, Polynomial versus exponential growth in repetition-free binary words, Journal of Combinatorial Theory Series A 105 (2) (2004) 335–347.
[16] A.J. Kfoury, A linear time algorithm to decide whether a binary word contains an overlap, Theoretical Informatics and Applications 22 (1988) 135–145.
[17] Y. Kobayashi, Enumeration of irreducible binary words, Discrete Applied Mathematics 20 (1988) 221–232.
[18] A. Lepistö, A characterization of 2+-free words over a binary alphabet, Master's Thesis, University of Turku, Finland, 1995.
[19] V.I. Oseledets, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems, Transactions of the Moscow Mathematical Society 19 (1968) 197–231.
[20] P. Parrilo, A. Jadbabaie, Approximation of the joint spectral radius using sum of squares, Linear Algebra and its Applications 428 (10) (2008) 2385–2402.
[21] V.Y. Protasov, The joint spectral radius and invariant sets of linear operators, Fundamentalnaya i prikladnaya matematika 2 (1) (1996) 205–231.
[22] V.Y. Protasov, The generalized spectral radius. A geometric approach, Izvestiya Mathematics 61 (5) (1997) 995–1030.
[23] V.Y. Protasov, On the asymptotics of the partition function, Sbornik Mathematics 191 (3–4) (2000) 381–414.
[24] V.Y. Protasov, On the regularity of de Rham curves, Izvestiya Mathematics 68 (3) (2004) 567–606.
[25] A. Restivo, S. Salemi, Overlap-free words on two symbols, in: Lecture Notes in Computer Science, Automata on Infinite Words, vol. 192, 1985, pp. 198–206.
[26] N.J.A. Sloane, On-line encyclopedia of integer sequences. Url: http://www.research.att.com/~njas/sequences.
[27] A. Thue, Uber unendliche zeichenreihen, Norske Videnskabers Selskabs Skrifter, I Mathematisch-Naturwissenschaftliche Klasse 7 (1906) 1–22.
[28] A. Thue, Uber die gegenseitige lage gleicher teile gewisser zeichenreihen, Norske Videnskabers Selskabs Skrifter, I Mathematisch-Naturwissenschaftliche Klasse 1 (1912) 1–67.