Theoretical Computer Science 410 (2009) 2823–2833

Contents lists available at ScienceDirect

Theoretical Computer Science

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

On the number of α-power-free binary words for $2 < \alpha \leq 7/3$

Vincent D. Blondel [a], Julien Cassaigne [b], Raphaël M. Jungers [a]

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

[b] Institut de mathématiques de Luminy, case 907, 163 avenue de Luminy, F-13288 Marseille Cedex 9, France

ARTICLE INFO

Keywords:
Overlap-free words
Combinatorics on words
Joint spectral radius

ABSTRACT

We study the number $u_\alpha(n)$ of α-power-free binary words of length n, and the asymptotics of this number when n tends to infinity, for a fixed rational number α in $(2, 7/3]$. For any such α, we prove a structure result that allows us to describe constructively the sequence $u_\alpha(n)$ as a 2-regular sequence. This provides an algorithm that computes the number $u_\alpha(n)$ in logarithmic time, for fixed α. Then, generalizing recent results on 2^+-free words, we describe the asymptotic behaviour of $u_\alpha(n)$ in terms of *joint spectral quantities* of a pair of matrices that one can efficiently construct, given a rational number α.

For $\alpha = 7/3$, we compute the automaton and give sharp estimates for the asymptotic behaviour of $u_\alpha(n)$.

1. Introduction

In combinatorics on words, a square is the repetition of the same word twice, as for instance the word *baba*. Similarly, the kth power of a word ($k \in \mathbb{N}$) consists of the concatenation of k times this word. This notion is classically generalized as follows. Let $w = w_1 \dots w_n \in A^*$ be a non-empty finite word on a finite alphabet A, and $n = |w|$. (See [14] for the usual definitions and notations of combinatorics on words.) The *period* of w is the smallest positive integer p such that $w_i = w_{i+p}$ for all i such that $1 \leq i \leq i + p \leq n$. Note that $1 \leq p \leq n$. The *period cycle* of w is the prefix of w of length p. The *exponent* of w is the rational number $e(w) = n/p$. As an example, $e(abacabacab) = 10/4 = 5/2$.

Since the beginning of the twentieth century, much research effort has been devoted to the so-called α-power-free words. Let α be a real number. A word $v \in A^* \cup A^\omega$ is *α-power-free* if every finite factor w of v satisfies $e(w) < \alpha$. The word v is *α^+-power-free* if every finite factor w of v satisfies $e(w) \leq \alpha$. It is easily seen that there are only finitely many binary *square-free* (i.e., 2-power-free) words. Indeed, every word of length 4 has a square. On the other hand, the infinite Thue–Morse word is *overlap-free* [21,14] (i.e., 2^+-power-free), and so, there are an infinite number of overlap-free words.

More generally, on an alphabet with k letters, there is a threshold $RT(k)$ such that there are only finitely many α-power-free words for $\alpha < RT(k)$, and infinitely many for $\alpha > RT(k)$. The value of $RT(k)$ was conjectured by Dejean in 1972: $RT(2) = 2$, $RT(3) = 7/4$, $RT(4) = 7/5$, $RT(k) = k/(k-1)$ for $k \geq 5$. Currently it is proved for $k \leq 14$ [9,15] and for $k \geq 30$ [4,7].

So, for binary words, the distinction between a finite number of α-power-free words and an infinite number is well understood. But the question arises to know how fast the number $u_\alpha(n)$ of binary α-power-free words of length n grows as a function of n.

Karhumäki and Shallit proved [12] that there are polynomially many 7/3-power-free binary words, and exponentially many $7/3^+$-power-free binary words. The main ingredient in the proof is the following structure lemma (generalizing a

E-mail addresses: vincent.blondel@uclouvain.be (V.D. Blondel), cassaigne@iml.univ-mrs.fr (J. Cassaigne), raphael.jungers@uclouvain.be (R.M. Jungers).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2009.01.031

result of [18] for overlaps):

Lemma 1 ([12]). *Let $A = \{a, b\}$, $2 < \alpha \le 7/3$, and $w \in A^*$ be α-power-free. Then there exist $x, y \in \{\varepsilon, a, b, aa, bb\}$ and $v \in A^*$ such that v is α-power-free and $w = x\theta(v)y$, where θ is the Thue–Morse morphism: $a \mapsto ab$, $b \mapsto ba$. Moreover, (x, v, y) is unique provided that $|w| \ge 7$.*

This lemma fails for $\alpha > 7/3$, for instance with $w = abbabba$. Moreover, a similar lemma where more values are allowed for x and y does not hold either, as *abbabba* can occur deep inside a $7/3^+$-power-free word, for instance $w = abbabaabbabbaabbabaab$.

Our goal is to compute exactly, or more precisely, the numbers $u_\alpha(n)$ in the polynomial case, i.e., for $2 < \alpha \le 7/3$. Namely, we are interested in the following:

- since $u_\alpha(n)$ is polynomial in n, find or bound its degree;
- find recurrence relations to compute $u_\alpha(n)$ efficiently.

A first idea is to iterate Lemma 1, producing a sequence of words (w_i), such that $w_n = w$, $w_i = x_i\theta(w_{i-1})y_i$, and w_0 is short. The short word w_0 and the sequence $(x_1, y_1), \dots, (x_n, y_n)$ are enough to describe w. Unfortunately, not all sequences $((x_i, y_i))$ generate α-power-free words; a sequence $((x_i, y_i))$, together with a choice of w_0, is said to be *admissible* if the word w it generates is α-power-free. For overlap-free words, Carpi proved [3] that admissible sequences form a regular language. As a consequence, $u_{2^+}(n)$ is a 2-*regular sequence* in the sense of [1]. However it is not easy to compute an automaton explicitly.

In order to make this computation easier, a *subtractive* variant of Lemma 1 has been proposed for overlap-free words. In this variant, one has to take into account some words that are not overlap-free, but that are *almost overlap-free words* in the sense that they can be written as xy with $x \in A$ such that y is overlap-free. Let U be the set of overlap-free binary words, V the set of almost overlap-free words, and S the set of words in $U \cup V$ of length less than 8. We also define the set $E = \{\kappa, \delta, \iota\}$ of transformations acting on either end of a word and defined as follows:

- κ does nothing;
- δ deletes the first (or last) letter of a word;
- ι inverts the first (or last) letter of a word.

Lemma 2 ([5]). *Let $w \in (U \cup V) \setminus S$. Then there exists a unique pair $(\gamma_1, \gamma_2) \in E \times E$ and $v \in U \cup V$ such that $w = \gamma_1.\theta(v).\gamma_2$.*

The advantage of this lemma is that given the first few and last few characters of the word v, it is possible to determine which functions will produce 2^+-free words and which ones will produce almost 2^+-free words. Moreover, since it is also possible to compute the first few and last few characters of w, it is possible to iterate the procedure.

The subtractive structure lemma allows one to derive the following theorem:

Theorem 3 ([5]). *Let (Y_n) be the sequence of vectors in $\mathbb{N}^{30}$ defined by initial terms and $Y_{2n} = F_0Y_n$, $Y_{2n+1} = F_1Y_n$ for $n > 6$, where F_0 and F_1 are specific matrices. Then $u_{2^+}(n) = RY_n$ for some specific row vector R.*

Let us mention that, in the above theorem, the entries of R, F_0, F_1 are all in the set $\{0, 1, 2\}$. This theorem allows one to compute $u_{2^+}(n)$ very efficiently, using the binary expansion of n to construct a product of the matrices F_0 and F_1.

A surprising corollary is that, although $u_{2^+}(n)$ grows polynomially, it does not have a fixed degree. Let $r^- = \liminf_{n\to\infty} \frac{\log u_{2^+}(n)}{\log n}$ and $r^+ = \limsup_{n\to\infty} \frac{\log u_{2^+}(n)}{\log n}$. Then, considering subsequences $u_{2^+}(2^m)$ and $u_{2^+}(\frac{4^m-1}{3})$, we get

$$r^- \le \log_2 \rho(F_0) < \log_4 \rho(F_0F_1) \le r^+,$$

where $\rho(F)$ denotes the spectral radius of the matrix F. Also, one can be interested in the function $s(n) = \sum_{m<n} u_{2^+}(m)$, which is easier to compute. Indeed, it satisfies the relation $s(n) = \Theta(n^r)$ with

$$r = \log_2\left(\frac{3}{2} + \sqrt{3} + \sqrt{\frac{5}{4} + \sqrt{3}}\right) = 2\log_2 \rho(M) \simeq 2.3100,$$

where $M = F_0 + F_1$.

Based on Theorem 3, it is possible to show that the quantities r^+ and r^- can be expressed in terms of *joint spectral quantities* of two matrices of size 20×20. For a given set of matrices $\Sigma = \{A_1, \dots, A_m\}$ we denote by $\check{\rho}$ and $\hat{\rho}$ its joint spectral subradius (also called lower spectral radius) and its joint spectral radius:

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

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

$$\check{\rho} \le \rho(A_{d_1}\cdots A_{d_k})^{1/k} \le \hat{\rho}. \tag{2}$$

(See [10,17,2] for surveys on these notions.) We have the following result:

Theorem 4 ([11]). *There exist two matrices $A_0, A_1 \in \{0, 1, 2\}^{20\times 20}$ such that*

$$\begin{aligned} r^+ &= \log_2 \hat{\rho}(\{A_0, A_1\}). \\ r^- &= \log_2 \check{\rho}(\{A_0, A_1\}). \end{aligned}$$

The proof of this theorem is based on numerical properties of the matrices F_0, F_1 in Theorem 3. Thanks to this result, the following accurate estimates appear in [11]:

$$1.2690 < r^- < 1.2736 < 1.3322 < r^+ < 1.3326.$$

2. Construction of automata

In this section, we show how to adapt the above described techniques to α-power-free words, for arbitrary $\alpha \in (2, 7/3]$. Again, the idea is to provide a structure result that is subtractive rather than additive. It expresses any α-power-free word w as the image of a shorter word v that is "almost" α-power-free under a function taken from a particular set. When α is rational, this result enables us to construct an automaton that describes the construction of all α-power-free words.

Let $A = \{a, b\}$ be a binary alphabet. For a given $\alpha \in \mathbb{R}$, we consider the set $U \subset A^*$ of α-power-free binary words.

2.1. Some properties of α-powers

We start with a few useful lemmas.

Let $\alpha \in \mathbb{R}$. A word w will be called an *α-power* if $e(w) \geq \alpha$. Note that the exponent need not be exactly α (for instance, *ababa* is a 7/3-power, since $5/2 \geq 7/3$). This definition may not be the standard one, but it happens to be practical here, and it is consistent with the definition of α-power-free words: a word is α-power-free if and only if it contains no α-power.

The following lemma is essentially due to Shur [19]. A more detailed proof can be found in [12]. We recall that θ, the Thue–Morse morphism, is defined on A by $\theta(a) = ab$, $\theta(b) = ba$.

Lemma 5 ([19,12]). *Let $\alpha > 2$ and $v \in A^*$. If $\theta(v)$ contains an α-power z of period p, then p is even and v contains an α-power y of period $p/2$, such that $\theta(y)$ contains z. In particular, if $\theta(v)$ is an α-power, then so is v, with half period.*

Proof. If p is odd, then we find *aa* or *bb* at two positions of different parities in $\theta(v)$, which is impossible. So p is even. If $|z|$ is even and z occurs at an even position (counting from 0), then it can be decoded and we find y such that $\theta(y) = z$. If $|z|$ is odd, or z occurs at an odd position, or both, then z can be extended on one or both sides to get a longer α-power of period p to which the previous case can be applied. □

Note that Lemma 5 still holds if α is replaced with α^+.

An α-power is said to be *minimal* if it contains no shorter α-power. It turns out that minimal α-powers are very constrained.

Lemma 6. *Let $2 < \alpha \leq 7/3$, and x be the period cycle of a minimal α-power $z \in A^*$. Then x is either a letter (then $|z| = 3$), or conjugated to aba or bab (then $|z| = 7$), or has even length and is conjugated to $\theta(x')$, where x' is the period cycle of another minimal α-power. Therefore, x is conjugated to one of the words $\theta^k(a)$, $\theta^k(b)$, $\theta^k(aba)$, $\theta^k(bab)$, with $k \in \mathbb{N}$.*

Proof. Consider all positions in z where *aa* or *bb* occur. Assume first that these positions do not all have the same parity. We can then find two successive such positions of different parities, i.e., a factor $aa(ba)^k a$ or $bb(ab)^k b$, with $k \in \mathbb{N}$. If $k = 0$, then z contains *aaa* or *bbb*, contradicting the minimality (except if z itself is *aaa* or *bbb*). If $k \geq 2$, then z contains an internal factor *ababa* or *babab*, again contradicting the minimality. If $k = 1$, then z contains *aabaa* or *bbabb* (say the former). If it is an internal factor, then we get *aaa* or *baabaab*, depending on the surrounding letters, contradicting the minimality except if $z = baabaab$. If it is a prefix, then either $z = aabaaba$ or $|x| \geq 4$ and *aabaa* occurs again as an internal factor of z at position $|x|$, which we have seen is a contradiction. If it is a suffix, a similar argument gives $z = abaabaa$.

If $|x|$ is odd, then either *aa* or *bb* occurs in x, or x is $(ab)^k a$ or $(ba)^k b$, with $k \in \mathbb{N}$, and *aa* or *bb* occurs at position $|x| - 1$ in z. In both cases, it occurs again $|x|$ positions further, therefore the above argument applies.

If $|x|$ is even, and all occurrences of *aa* and *bb* are at odd positions, then x can be factored on $\{ab, ba\}$, i.e., $x = \theta(x')$. After possibly extending z by one letter y at the end to make its length even, we find an α-power z' with period cycle x' such that $z = \theta(z')$ or $zy = \theta(z')$. Then Lemma 5 and the minimality of z ensure that z' is minimal.

If $|x|$ is even, and all occurrences of *aa* and *bb* are at even positions, then let y be the last letter of x and y' the last letter of z. The word yzy'^{-1} is again a minimal α-power, with period cycle conjugated to x, but now *aa* and *bb* occur at odd positions and the previous case can be applied.

Finally, the last statement is obtained by iteration. □

Note that Lemma 6 generalizes the characterization of minimal binary overlaps by Thue [22, Satz 13]. It is also a consequence of the recent characterization of 7/3-power-free binary squares by Currie and Rampersad [6, Theorem 2]. The words xx, where x is one of $\theta^k(a)$, $\theta^k(b)$, $\theta^k(aba)$, $\theta^k(bab)$, with $k \in \mathbb{N}$, are exactly the squares that occur in the Thue–Morse word [16].

This lemma has an interesting corollary:

Corollary 7. *Let $2 < \alpha \leq 7/3$ be a real number. There exist α^+-power-free words that are not α-power-free if and only if $\alpha = r/2^k$ or $\alpha = r/(3.2^k)$, with integer r and k.*

Proof. Suppose that w is α^+-power-free but not α-power-free. Let z be the shortest α-power contained in w. By construction, it is a minimal α-power, so by Lemma 6 its period p is 2^k or 3.2^k. But as w is α^+-power-free, the exponent of z cannot exceed α, so it has to be exactly α. Hence $\alpha = e(z) = |z|/p$.

Conversely, if $\alpha = r/2^k \geq 2$, let $x = \theta^k(a)$; if $\alpha = r/(3.2^k) \geq 2$, let $x = \theta^k(aba)$. Let then w be any factor of length r of x^4. The word w is not α-power-free, as $e(w) \geq |w|/|x| = \alpha$. Assume that w contains an α^+-power z'. Then the period of z' is $p' = |z'|/e(z') < |w|/\alpha = |x|$. Applying k times Lemma 5, we find that a^4 or $(aba)^4$ contains an α^+-power of period $p'/2^k < |x|/2^k$. This is clearly impossible if $|x| = 2^k$; if $|x| = 3.2^k$, then it means that $(aba)^4$ contains an α^+-power of period 1 or 2, a contradiction. So w is α^+-power-free. $\square$

Note that the "if" part also works for $\alpha > 7/3$. As a consequence, the set of reals α such that there exist α^+-power-free binary words containing α-powers is dense in $[2, +\infty)$; a similar result can be found in [8, Theorem 14]. The exact structure of this set is, as far as we know, not known outside $[2, 7/3]$.

From now on, we assume that $2 < \alpha \leq 7/3$.

Lemma 8. *Let $w \in AU A$ be a word that does not contain α-powers as internal factors. Let z and z' be two distinct prefixes of w that are α-powers, with respective periods p and p'. Then $p = p'$, the longer word in $\{z, z'\}$ is w itself, and the other one is shorter by just one letter.*

Proof. Assume first that $p < p'$. Let x and x' be the respective period cycles of z and z', and write $z = xxy$, $z' = x'x'y'$. Recall that $\alpha > 2$, so y and y' are not empty. Note that $|z| > |z'|$ is impossible, as the internal factor of length $|z| - 2$ of z would then be an α-power (its exponent being at least $(|z'| - 1)/(p' - 1)$, which is larger than $|z'|/p' \geq \alpha$). Then $|z| < |z'|$, and z is a proper prefix of w. As a consequence, $|z| < \alpha p + 1$, otherwise z could be shortened to get an internal α-power. In turn, this implies that $|xy| < (\alpha - 1)p + 1 \leq (\alpha - 1)(p' - 1) + 1 < (\alpha - 1)p' \leq |x'y'|$.

If $|z| < |x'y'|$, then z is a proper prefix of $x'y'$ as $x'y'$ is also a prefix of w. Then z occurs as an internal factor of w, a contradiction.

If $|z| \geq |x'y'|$, then xy is a proper prefix of $x'y'$ and occurs at positions p and p' in w. We have $(\alpha - 1)(2p + 1) = (\alpha - 2)(p + 1) + \alpha p + 1$, where $(\alpha - 2)(p + 1) > 0$ and $\alpha p + 1 > |z| \geq |x'y'| = |z'| - p' \geq (\alpha - 1)p'$. Consequently $2p + 1 > p'$, i.e., $p' - p \leq p$. Let s be the suffix of length $p' - p$ of x', so that the word sxy occurs at position p in w and has period $p' - p$. Then sxy is an internal factor of w with exponent $e(sxy) \geq |sxy|/|s| = 1 + |xy|/(p' - p) \geq 1 + |xy|/p \geq \alpha$, again a contradiction.

Assume now that $p = p'$, and that $|z| < |z'|$. Any factor of length $|z|$ of z' is an α-power, and we can find one that is an internal factor of w except in one case, when $z' = w$ and $|z| = |z'| - 1$. $\square$

2.2. Subtractive structure lemma

Lemma 1 (the structure lemma of Karhumäki and Shallit) is an *additive structure lemma*, as letters are added on both sides of $\theta(v)$ to get w. Instead, we will use a *subtractive structure lemma* similar to Lemma 2, in which letters can be deleted from both sides of $\theta(v)$ (and then also added).

We also consider a larger set of words, $AU A$. Its elements are *almost α-power-free* words: they may contain an α-power, but only as a prefix or as a suffix. The number $w(n)$ of words of length n in $AU A$ satisfies $u_\alpha(n) \leq w(n) = 4u_\alpha(n - 2)$, and can therefore be used instead of $u_\alpha(n)$ for computing asymptotic quantities such as $r_\alpha^+ = \limsup_{n\to\infty} \log u_\alpha(n)/\log n$.

We define a set of five transformations $E = \{\delta, \kappa, \iota, \sigma, \tau\}$, extending the three transformations used in Lemma 2. Each element of E acts to the left of a non-empty word as follows:

- $\delta.xw = w$,
- $\kappa.xw = xw$,
- $\iota.xw = \bar{x}w$,
- $\sigma.xw = x\bar{x}w$,
- $\tau.xw = \bar{x}\bar{x}w$,

where $x \in A$, and $\bar{x}$ denotes the other letter (so that $A = \{x, \bar{x}\}$). Each $\gamma \in E$ also acts to the right: $w.\gamma$ is the mirror image of $\gamma.\tilde{w}$, where $\tilde{w}$ denotes the mirror image of w.

Lemma 9. *If $w \in AU A$ and $|w| \geq 9$, then there exists a unique triple $(\gamma_1, \gamma_2, v) \in E \times E \times (AU A)$ such that $w = \gamma_1.\theta(v).\gamma_2$.*

Proof. Let $w = x'w'y'$, with $x', y' \in A$ and $w' \in U$. Note that $|w'| \geq 7$. By Lemma 1, there exist unique $r_1, r_2 \in \{\varepsilon, a, b, aa, bb\}$ and $v' \in U$ such that $w' = r_1\theta(v')r_2$.

Let y be the first letter of r_2y'. Then, by construction, r_2y' is one of $y, y\bar{y}, yy, yy\bar{y}, yyy$. Defining γ_2 as respectively $\delta, \kappa, \iota, \sigma, \tau$, we get $r_2y' = \theta(y).\gamma_2$. Similarly, let x be such that $\bar{x}$ is the last letter of $x'r_1$. Then $x'r_1 \in \{\bar{x}, x\bar{x}, \bar{x}\bar{x}, x\bar{x}\bar{x}, \bar{x}\bar{x}\bar{x}\}$ so that $x'r_1 = \gamma_1.\theta(x)$ for an appropriate γ_1. Let $v = xv'y \in AU A$: we then have $w = \gamma_1.\theta(v).\gamma_2$.

To prove uniqueness, assume that $w = \gamma_1.\theta(v).\gamma_2$ with $v \in AU A$. Let $w = x'w'y'$ and $v = xv'y$ (note that $|v|$ must be at least 4, since $|w| \geq 9$, each transformation increases the length at most by 1, and θ doubles the length). Then $w = r_1\theta(v')r_2$ where $r_1 = (x')^{-1}\gamma_1.\theta(x)$ and $r_2 = \theta(y).\gamma_2(y')^{-1}$. Note that $v' \in U$ and $r_1, r_2 \in \{\varepsilon, a, b, aa, bb\}$, and that the map between (x', y', r_1, r_2) and $(x, y, \gamma_1, \gamma_2)$ is one-to-one. Therefore uniqueness of (r_1, r_2, v') guarantees uniqueness of (γ_1, γ_2, v). $\square$

The bound 9 in Lemma 9 is the best possible: indeed, the word $w = aaababba$, for instance, has two decompositions, $aaababba = \tau.\theta(bbb).\sigma = \iota.\theta(baab).\kappa$.

We define a map $\Phi : AA^+ \times (E\times E)^* \to AA^+$ inductively as follows: $\Phi(u, \varepsilon) = u$, and $\Phi(u, (\gamma_1, \gamma_2)\xi) = \Phi(\gamma_1.\theta(u).\gamma_2, \xi)$, for any $u \in AA^+$ (a word of length at least 2), $(\gamma_1, \gamma_2) \in E \times E$ (a pair of transformations), and $\xi \in (E \times E)^*$ (a sequence of

Table 1
Relations between the type of v and the type of $w = \gamma_1.\theta(v).\gamma_2$. X means that $w \notin AUA$.

	aaa	*aaba*	*aabb*	*aba*	*abb*
δ	*aba*	*aba*	*aba*	*aabb*	*aaba*
κ	X	*aba*	*aba*	*abb*	*abb*
ι	X	*aaba*	*aaba*	*aaa*	*aaa*
σ	X	*abb*	X	X	X
τ	X	*aaa*	X	X	X

pairs of transformations). This allows us to iterate Lemma 9 and to represent elements of AUA as images, under a repeated application of θ alternated with transformations from E on both sides, of an initial short word.

2.3. One-sided control

The converse of Lemma 9 does not hold: given $(\gamma_1, \gamma_2, v) \in E \times E \times (AUA)$, $w = \gamma_1.\theta(v).\gamma_2$ need not be in AUA, as internal α-powers may appear. Lemma 5 ensures that the morphism θ itself does not create α-powers, but the transformations γ_1 and γ_2 may do so.

Let $v = xv'y \in AUA$, where $v' \in U$ and $x, y \in A$. As $v' \in U$, by Lemma 5, $\theta(v') \in U$ as well. If w contains an internal α-power, then it has to intersect either the prefix $\gamma_1.\theta(x)$, or the suffix $\theta(y).\gamma_2$, or both. We first restrict to α-powers that do not intersect $\theta(y).\gamma_2$, and for this we may as well assume that $\gamma_2 = \delta$.

Lemma 10. *Let $\gamma_1 \in E$ and $v = xv'y \in AUA$, with $x, y \in A$ and $|v| \geq 5$. Then $w = \gamma_1.\theta(v).\delta$ contains an internal α-power z if and only if one of the following situations occur:*

- *γ_1 is σ or τ, and v starts with ab or ba (then z is aaa or bbb);*
- *γ_1 is σ or τ, and v starts with aabb or bbaa (then z is aabaaba or bbabbab);*
- *γ_1 is other than δ, and v has a proper prefix which is an α-power z' of period p, with $e(z') \geq \alpha + 1/2p$ (then $z = \delta.\theta(z')$).*

Proof. First note that, as $\alpha \leq 7/3 < 5/2 < 3$, the words *aaa*, *ababa*, *aabaaba*, *abaabaa*, *baabaab* and their complements are α-powers whatever the value of α. Moreover, they are the only minimal α-powers of period up to 3.

In each of the three situations, it is clear that w contains an internal α-power.

Conversely, assume that w contains an internal α-power z, which can be taken minimal. As we saw above, z has to intersect either the prefix $\gamma_1.\theta(x)$ or the suffix $\theta(y).\delta = y$, and the latter is impossible since z is internal. So $\gamma_1 \neq \delta$; if γ_1 is κ or ι, then z starts at position 1 in w (counting from 0), so it is a prefix of $\delta.\theta(xv') = \bar{x}\theta(v')$; if γ_1 is σ or τ, then either z starts at position 2 and is again a prefix of $\delta.\theta(xv')$, or z starts at position 1 and it is a prefix of $\iota.\theta(xv') = \bar{x}\bar{x}\theta(v')$.

If z has odd period, then by Lemma 6 its period is 1 or 3, and it is easily checked that the only possibilities are *aaa* and *aabaaba*, and complements, occurring as prefixes of $\iota.\theta(xv')$, which correspond to the first two cases.

If z has even period $2p$, then it cannot be a prefix of $\iota.\theta(xv')$, or $\bar{x}\bar{x}$ would occur at an even position in $\theta(v')$. So it is a prefix of $\delta.\theta(xv')$, and by Lemma 5 we get an α-power z' of period p in v (that must be a proper prefix) such that $\theta(z')$ contains z. Actually z is a prefix of $\delta.\theta(z')$, so we have $2|z'| - 1 \geq 2p\alpha$ and $e(z') \geq \alpha + 1/2p$. □

Note that, unlike for the first two cases, the conditions in which the third case occurs are rather sensitive to the value of α. For instance, if $v = ababaab$, then $w = \kappa.abbaabbaababba.\delta = abbaabbaababb$ contains *bbaabbaab* as an internal α-power if $\alpha \leq 9/4$.

To control the first two cases (as well as the subcase $p = 2$ of the third one), we define the *prefix type* of a word w of length at least 4 as $t_1 \in \{aaa, aaba, aabb, aba, abb\}$ such that w starts with t_1 or $\overline{t_1}$.

Lemma 11. *Let $v \in AUA$, $|v| \geq 4$, and $\gamma_1, \gamma_2 \in E$. Assume that $w = \gamma_1.\theta(v).\gamma_2 \in AUA$. Then the prefix type of w is determined by γ_1 and the prefix type of v, according to Table 1, where the columns correspond to the prefix type of v and the rows correspond to the function γ_1. An X in the table means that w cannot be in AUA.*

Proof. Since $|v| \geq 4$, the prefix of length 5 of $\theta(v).\gamma_2$ is not affected by γ_2, and depends only on the prefix type of v. Therefore the prefix type of w depends only on the prefix type of v and γ_1, and is easy to compute.

It remains to explain the X in the table. If the prefix type of v is *aabb*, *aba*, or *abb*, and $\gamma_1 \in \{\sigma, \tau\}$, then w contains an internal α-power by the first two cases of Lemma 10. If the prefix type of v is *aaa*, and $\gamma_1 \neq \delta$, then w contains an internal α-power (*ababa* or *babab*) by the third case of Lemma 10. □

We now turn to the third case. We define the *prefix excess* of a word w as the maximal value f_1 of $|z| - p\alpha$, where z is a proper prefix of w of period $p = |z|/e(z)$. If $w \in UA$, then its prefix excess is negative (and its actual value does not matter); if $w \in AUA \setminus UA$, then its prefix excess is in $[0, 1)$. For example, if $\alpha = 16/7$, the word *abbaabbaababb* has prefix excess $f_1 = |abbaabbaab| - |abba|\alpha = 6/7$. The condition $e(z) \geq \alpha + 1/2p$ in the third case of Lemma 10 translates to $f_1 \geq 1/2$.

Lemma 12. *Let $v \in AUA$, $|v| \geq 5$, and $\gamma_1, \gamma_2 \in E$. Assume that $w = \gamma_1.\theta(v).\gamma_2 \in AUA$. Let t_1 be the prefix type and f_1 the prefix excess of v. Assume also, if γ_1 is δ or κ, that v is not itself an α-power. Then the prefix excess f_1' of w is determined by γ_1,*

Table 2
Relations between the prefix excess f_1 of v and the prefix excess f_1' of $w = \gamma_1.\theta(v).\gamma_2$. X means that $w \notin AU\,A$.

γ_1	t_1	$f_1 < 0$	$0 \le f_1 < 1/2$	$1/2 \le f_1 < 1$
δ		$f_1' < 0$	$f_1' < 0$	$f_1' = 2f_1 - 1$
κ		$f_1' < 0$	$f_1' = 2f_1$	X
ι	*aaba*	$f_1' < 0$	$f_1' < 0$	X
ι	*aabb*	$f_1' = 7 - 3\alpha$	$f_1' = 7 - 3\alpha$	X
ι	*aba*	$f_1' = 3 - \alpha$	$f_1' = 3 - \alpha$	X
ι	*abb*	$f_1' = 3 - \alpha$	$f_1' = 3 - \alpha$	X
σ		$f_1' = 7 - 3\alpha$	$f_1' = 7 - 3\alpha$	X
τ		$f_1' = 3 - \alpha$	$f_1' = 3 - \alpha$	X

t_1, and f_1, according to Table 2, where the columns correspond to intervals of values of f_1 and the rows correspond to values of γ_1 and, when $\gamma_1 = \iota$, of t_1. When $\gamma_1 \neq \iota$, t_1 does not matter. An X in the table means that w cannot be in $AU\,A$.

Proof. First note that, by the third case of Lemma 10, $f_1 \geq 1/2$ is only possible when $\gamma_1 = \delta$.

Let z be a proper prefix of v such that $f_1 = |z|(1 - \alpha/e(z))$ (by Lemma 8, z is unique when $f_1 \geq 0$). If $\gamma_1 = \kappa$, then $\theta(z)$ is a proper prefix of w, and $e(\theta(z)) \geq e(z)$, so that $f_1' \geq |\theta(z)|(1 - \alpha/e(\theta(z))) \geq 2f_1$. If $\gamma_1 = \delta$, then $\delta.\theta(z)$ is a proper prefix of w, with the same period as $\theta(z)$, so that $f_1' \geq 2f_1 - 1$. If $\gamma_1 = \sigma$, then the only possibility is $t_1 = aaba$, and then *abbabba* or *baabaab* is a proper prefix of w, so that $f_1' \geq 7 - 3\alpha$. If $\gamma_1 = \tau$, then w is of prefix type *aaa*, so that $f_1' \geq 3 - \alpha$. If $\gamma_1 = \iota$ and $t_1 = aabb$, then *bbabbab* or *aabaaba* is a proper prefix of w, so that $f_1' \geq 7 - 3\alpha$. If $\gamma_1 = \iota$ and $t_1 = aba$ or $t_1 = abb$, then w is of prefix type *aaa*, so that $f_1' \geq 3 - \alpha$. Therefore the values given in the table are lower bounds.

By Lemma 8, in the cases where we have found an α-power of length 3 or 7 as a prefix of w, no other prefix can be an α-power (note that $\gamma_1 \neq \delta$ in those cases so that $|w| \geq 9$ as $|v| \geq 5$).

For the other cases, it remains to show that f_1' cannot be higher than the values given. Assume the contrary. Then an α-power z' such that $f_1' = |z'|(1 - \alpha/e(z'))$ occurs as a proper prefix of w. Let x' be its period cycle, so that $x'x'$ is a proper prefix of w. Note that z' cannot intersect the suffix of w affected by γ_2 (i.e., the second from last letter in w when γ_2 is σ or τ), because z' would then end in *aaa*, *bbb*, *aabaa*, *bbabb* which would then have another occurrence in the middle of $\theta(v)$.

If $\gamma_1 = \kappa$, then by Lemma 5 there is a prefix z of v of period $|x'|/2$ and exponent at least $e(z')$, and it is a proper prefix since we assumed that v is not an α-power in this case. Therefore $f_1 \geq f_1'/2$.

If $\gamma_1 = \delta$, then z can be extended to the left to get an α-power as a proper prefix of $\theta(v)$, and then by the previous argument we get $f_1 \geq (f_1' + 1)/2$.

The only remaining case is when $\gamma_1 = \iota$ and $t_1 = aaba$. Then w starts with *bbabbaa*. It is clear that $|x'|$ cannot be less than 7, but then *bbabbaa* has to occur a second time, a contradiction. □

Lemma 12 requires in some cases that v is not an α-power. The following lemma ensures that this property propagates to w.

Lemma 13. *Let $v \in AU\,A$, with $|v| \geq 5$, and $\gamma_1, \gamma_2 \in E$. Assume that $w = \gamma_1.\theta(v).\gamma_2 \in AU\,A$. If v is not an α-power, then w is not an α-power either.*

Proof. Assume that w is an α-power. If γ_1 is ι, σ or τ, then w contains at position 0 or 1 one of *aaa*, *bbb*, *aabaa*, *bbabb*, which must then have another occurrence in the middle of $\theta(v)$, a contradiction ($|v| \geq 5$ is needed here). The same argument applies to γ_2. So both γ_1 and γ_2 are κ or δ, and then Lemma 5 implies that v is an α-power. □

With Lemmas 11 and 12, we can keep track of t_1 and f_1 when Lemma 9 is iterated (excluding α-powers for the moment).

When α is rational, f_1 takes finitely many useful values, so that this can be done with a deterministic finite automaton $\mathcal{A}_\alpha$. Assume that $\alpha = r/q$, with r and q being coprime. The states of $\mathcal{A}_\alpha$ are labelled $(t_1, \hat{f}_1)$, where $t_1 \in \{aaa, aaba, aabb, aba, abb\}$ is the prefix type and $\hat{f}_1 \in \{-, 0, 1, \ldots, q-1\}$ is $-$ if the prefix excess is negative, and $\hat{f}_1 = qf_1$ if $f_1 \geq 0$ (indeed, this is an integer between 0 and $q-1$). Accepting states (for $U\,A$) are states with $\hat{f}_1 = -$. Transitions of $\mathcal{A}_\alpha$ are labelled by E, and $(t_1', \hat{f}_1') = \gamma_1.(t_1, \hat{f}_1)$ is defined as follows:

- t_1' is given by Table 1;
- $\hat{f}_1'$ is given by Table 2.

If either table has an X, there is no transition.

2.4. Two-sided control

We define the suffix type t_2 and the suffix excess f_2 of v as the prefix type and prefix excess of $\tilde{v}$. However, it is not enough to consider both ends of the word independently: we need a special treatment for α-powers that intersect both $\gamma_1.\theta(x)$ and $\theta(y).\gamma_2$.

Let the global excess g of v be $|v| - p\alpha$, where p is the period of v: it is non-negative when v is an α-power, and $g < 2$. By Lemma 8, if $0 \leq g < 1$ then f_1 and f_2 are both negative, and if $1 \leq g < 2$ then $f_1 = f_2 = g - 1$.

Table 3
Relations between the global excess g of v and the global excess g' of $w = \gamma_1.\theta(v).\gamma_2$. The symbol $*$ means any of ι, σ, τ, and X means that $w \notin AU A$.

(γ_1, γ_2)	$g < 0$	$0 \le g < 1/2$	$1/2 \le g < 1$	$1 \le g < 3/2$	$3/2 \le g < 2$
(δ, δ)	$g' < 0$	$g' < 0$ $f_1' < 0, f_2' < 0$	$g' < 0$ $f_1' < 0, f_2' < 0$	$g' = 2g - 2$ $f_1' < 0, f_2' < 0$	$g' = 2g - 2$ $f_1' = f_2' = 2g - 3$
(δ, κ) (κ, δ)	$g' < 0$	$g' < 0$ $f_1' < 0, f_2' < 0$	$g' = 2g - 1$ $f_1' < 0, f_2' < 0$	$g' = 2g - 1$ $f_1' = f_2' = 2g - 2$	X
(κ, κ)	$g' < 0$	$g' = 2g$ $f_1' < 0, f_2' < 0$	$g' = 2g$ $f_1' = f_2' = 2g - 1$	X	X
$(\delta, *)$	$g' < 0$	$g' < 0$ $f_1' < 0$	$g' < 0$ $f_1' < 0$	$g' < 0$ $f_1' = 2g - 2$	X
$(*, \delta)$	$g' < 0$	$g' < 0$ $f_2' < 0$	$g' < 0$ $f_2' < 0$	$g' < 0$ $f_2' = 2g - 2$	X
$(\kappa, *)$	$g' < 0$	$g' < 0$ $f_1' < 0$	$g' < 0$ $f_1' = 2g - 1$	X	X
$(*, \kappa)$	$g' < 0$	$g' < 0$ $f_2' < 0$	$g' < 0$ $f_2' = 2g - 1$	X	X
$(*, *)$	$g' < 0$	$g' < 0$	$g' < 0$	X	X

Lemma 14. *Let $\gamma_1, \gamma_2 \in E$ and $v = xv'y \in AU A$, with $x, y \in A$ and $|v| \ge 5$. Let t_1, t_2, f_1, f_2, and g be the types and excesses associated with v. Then $w = \gamma_1.\theta(v).\gamma_2$ contains an internal α-power z if and only if one of the following situations occur:*

- *(γ_1, v) satisfies one of the three conditions in Lemma 10;*
- *$(\gamma_2, \tilde{v})$ satisfies one of the three conditions in Lemma 10;*
- *$g \ge 1$, $\gamma_1 \ne \delta$, and $\gamma_2 \ne \delta$ (then $z = \delta.\theta(v).\delta$).*

Proof. In each of the three situations, it is clear that w contains an internal α-power.

Conversely, assume that w contains an internal α-power z, which can be taken minimal. We can also assume that Lemma 10 does not apply on either side. Then z intersects both $\gamma_1.\theta(x)$ and $\theta(y).\gamma_2$, so neither γ_1 nor γ_2 is equal to δ, and z contains $\delta.\theta(v).\delta$, which implies that $|z| \ge 2|v| - 2 \ge 8$. In particular, by Lemma 6, z has an even period $2p \ge 4$.

If γ_1 is σ or τ and z starts at position 1 in w, then z starts with $\bar{x}\bar{x}$, which then occurs at position $2p$ in $\theta(v)$, a contradiction. Similarly it is impossible to have γ_2 equal to σ or τ and z ending only one letter before the end of w. In all cases that remain, $z = \delta.\theta(v).\delta$, which implies by Lemma 5 that $xz\bar{y} = \theta(v)$ is an $(\alpha + 1/p)$-power, and so is v, with period p. Therefore $g \ge 1$ and we are in the third case. $\square$

The following lemma describes how global excess evolves, as well as how it influences prefix and suffix excesses in the cases not covered by Lemma 12.

Lemma 15. *Let $v \in AU A$, $|v| \ge 5$, and $\gamma_1, \gamma_2 \in E$. Let t_1, t_2, f_1, f_2, and g be the types and excesses associated with v. Assume that $w = \gamma_1.\theta(v).\gamma_2 \in AU A$. Then the global excess g' of w is determined by (γ_1, γ_2) and by g, according to Table 3, where the columns correspond to intervals of values of g and the rows correspond to pairs (γ_1, γ_2). A $*$ in a row label stands for any of ι, σ, τ. An X in the table means that w cannot be in $AU A$. Some values of the prefix excess f_1' and of the suffix excess f_2' of w are also given in the table, when Lemma 12 does not apply.*

We now have all the elements to state our main structure result.

Theorem 16. *Let $2 < \alpha \le 7/3$ be a rational number. There exist finite sets of words S and U_0, and a regular language $L \subset S \times (E \times E)^*$, recognized by an explicit automaton $\mathcal{B}_\alpha$, such that Φ induces a one-to-one map from L to $U \backslash U_0$, where U is the language of α-power-free binary words.*

Proof. Let U_0 be the set of α-power-free binary words of length up to 4. This set has 23 elements and does not depend on α.

Recall that $\alpha = q/r$. We construct an automaton $\mathcal{B}_\alpha$ with states $(t_1, \hat{f}_1, t_2, \hat{f}_2, \hat{g})$, where $\hat{f}_2 \in \{-, 0, 1, \ldots, q - 1\}$ and $\hat{g} \in \{-, 0, 1, \ldots, 2q - 1\}$ are defined as $\hat{f}_1$. There are $25(q + 1)^2(2q + 1)$ such tuples, but most of them are not used since, for instance, $t_1 = aaa$ implies $\hat{f}_1 = 3q - r$, or $\hat{g} \ne -$ determines $\hat{f}_1$ and $\hat{f}_2$. Transitions are labelled by $E \times E$, and $(t_1', \hat{f}_1', t_2', \hat{f}_2', \hat{g}') = (\gamma_1, \gamma_2).(t_1, \hat{f}_1, t_2, \hat{f}_2, \hat{g})$ is defined as follows:

- t_1' is given by Table 1 applied to γ_1 and t_1;
- t_2' is given by Table 1 applied to γ_2 and t_2;
- $\hat{f}_1'$ is given by Table 2, except when $g \ge 0$ and $\gamma_1 \in \{\delta, \kappa\}$ where it is given by Table 3;

- $\hat{f}_2{}'$ is given symmetrically;
- $\hat{g}'$ is given by Table 3.

If an impossibility occurs at any of these steps, then the transition does not exist.

Let S be the set of elements of $AU\,A$ of length 5 to 7, as well as those elements of $AU\,A$ of length 8 to 10 that cannot be obtained from a shorter element of S. We add to the automaton an extra state i, which will be the initial state, with for each v in S a transition labelled by v from i to the state describing v. Accepting states are states with $\hat{f}_1 = \hat{f}_2 = \hat{g} = -$. Finally, the automaton can be trimmed of all unreachable states.

Let $L = L(\mathcal{B}_\alpha)$. According to the previous lemmas, $\Phi(u, \xi) \in U \backslash U_0$ for any (u, ξ) in L. The choice of S and Lemma 9 ensure that the map is one-to-one. □

3. Counting α-power-free binary words

We now study the consequences of Theorem 16 on $u_\alpha(n)$, the number of α-power-free binary words of length n. In this section, α is always assumed to be rational.

3.1. A 2-regular sequence

Theorem 17. *Let $2 < \alpha \le 7/3$ be a rational number. The sequence $u_\alpha(n)$ is 2-regular in the sense of [1]: there exist integers m and d, matrices F_0 and F_1 in $\mathbb{N}^{d\times d}$, vectors $Y_0, \ldots, Y_{2^m-1}$ in $\mathbb{N}^d$, and a row vector R in $\mathbb{N}^{1\times d}$ such that, if the sequence of vectors (Y_n) is defined inductively by $Y_{2n} = F_0 Y_n$ and $Y_{2n+1} = F_1 Y_n$ for $n \ge 2^{m-1}$, then $u_\alpha(n) = RY_n$ for all $n \ge 0$.*

Proof. Let $\mathcal{B}_\alpha$ be the automaton constructed in Theorem 16. Number its states arbitrarily from 1 to some s. For each $n \ge 5$ and $j \in \{1, \ldots, s\}$, let $x_{n,j}$ be the number of words w of length n in $AU\,A$ such that, reading $\Phi^{-1}(w)$, the automaton $\mathcal{B}_\alpha$ ends in state j. This defines a sequence of vectors $X_n = (x_{n,j})_{1\le j\le s} \in \mathbb{N}^s$. Moreover, let $H = (h_j)_{1\le j\le s} \in \{0, 1\}^{1\times s}$ be the row vector defined by $h_j = 1$ if state j is accepting, $h_j = 0$ otherwise, so that $u_\alpha(n) = HX_n$.

Let $G_{(\gamma_1,\gamma_2)}$ be the transition matrix of $\mathcal{B}_\alpha$ associated with the symbol (γ_1, γ_2). Note that, if $w = \gamma_1.\theta(v).\gamma_2$, then $|w| - 2|v| \in \{-2, -1, 0, 1, 2\}$ depends only on (γ_1, γ_2). Accordingly, define:

$$
\begin{aligned}
G_{-2} &= G_{(\delta,\delta)};\\
G_{-1} &= G_{(\delta,\kappa)} + G_{(\delta,\iota)} + G_{(\kappa,\delta)} + G_{(\iota,\delta)};\\
G_0 &= G_{(\delta,\sigma)} + G_{(\delta,\tau)} + G_{(\kappa,\kappa)} + G_{(\kappa,\iota)} + G_{(\iota,\kappa)} + G_{(\iota,\iota)} + G_{(\sigma,\delta)} + G_{(\tau,\delta)};\\
G_1 &= G_{(\kappa,\sigma)} + G_{(\kappa,\tau)} + G_{(\iota,\sigma)} + G_{(\iota,\tau)} + G_{(\sigma,\kappa)} + G_{(\sigma,\iota)} + G_{(\tau,\kappa)} + G_{(\tau,\iota)};\\
G_2 &= G_{(\sigma,\sigma)} + G_{(\sigma,\tau)} + G_{(\tau,\sigma)} + G_{(\tau,\tau)}.
\end{aligned}
$$

We thus get recurrence relations

$$
\begin{aligned}
X_{2n+1} &= G_{-1}X_{n+1} + G_1 X_n,\\
X_{2n+2} &= G_{-2}X_{n+2} + G_0 X_{n+1} + G_2 X_n,
\end{aligned}
$$

valid for $n \ge 5$.

Multiplying the dimension by four, we can turn them into simpler equations. Let $d = 4s$ and $m = 4$. Define $Y_n \in \mathbb{N}^d$, for $n \ge 6$, $F_0, F_1 \in \mathbb{N}^{d\times d}$, and $R \in \mathbb{N}^{1\times d}$, by

$$
Y_n = \begin{pmatrix} X_{n-1}\\ X_n\\ X_{n+1}\\ X_{n+2}\end{pmatrix}, \qquad
F_0 = \begin{pmatrix} G_1 & G_{-1} & 0 & 0\\ G_2 & G_0 & G_{-2} & 0\\ 0 & G_1 & G_{-1} & 0\\ 0 & G_2 & G_0 & G_{-2}\end{pmatrix}, \qquad
F_1 = \begin{pmatrix} G_2 & G_0 & G_{-2} & 0\\ 0 & G_1 & G_{-1} & 0\\ 0 & G_2 & G_0 & G_{-2}\\ 0 & 0 & G_1 & G_{-1}\end{pmatrix},
$$

and $R = (0\ H\ 0\ 0)$. Then we have $Y_{2n} = F_0 Y_n$ and $Y_{2n+1} = F_1 Y_n$ for $n \ge 6$. These equations completely define $(Y_n)_{n\ge 6}$, provided initial values Y_6 to Y_{11} are given. It remains to choose $Y_0, \ldots, Y_5 \in \mathbb{N}^{d\times d}$ such that $RY_0 = 1$, $RY_1 = 2$, $RY_2 = 4$, $RY_3 = 6$, $RY_4 = 10$, and $RY_5 = 14$ in order to have $u_\alpha(n) = RY_n$ for all $n \in \mathbb{N}$. □

Theorem 17 provides a fast algorithm for computing values of $u_\alpha(n)$:

Corollary 18. *The number $u_\alpha(n)$ of α-power-free words of length $n \ge 2^m$ can be obtained by first computing the binary expansion $d_k \cdots d_0$ of n, i.e., $n = \sum_{j=0}^{k} d_j 2^j$, with $d_j \in \{0, 1\}$, $d_k = 1$, and then defining*

$$
u_\alpha(n) = RF_{d_0} \cdots F_{d_{k-m}} Y_{n_0}
$$

where $n_0 = \sum_{j=1}^{m} d_{k-m+j} 2^{j-1}$.

3.2. r_α^+ and r_α^- as joint spectral quantities

Let $r_\alpha^- = \liminf_{n\to\infty} \frac{\log u_\alpha(n)}{\log n}$ and $r_\alpha^+ = \limsup_{n\to\infty} \frac{\log u_\alpha(n)}{\log n}$.

In this section we prove that r_α^+ and r_α^- are related to the joint spectral radius and the joint spectral subradius of the matrices F_0 and F_1. This has already been proved in [11] for $\alpha = 2^+$, but this result came from a precise numerical analysis of the matrices, and could not be applied to an arbitrary pair $\{F_0, F_1\}$. Here we adopt a more abstract approach, based on the knowledge of the structure of the automaton.

Let $|.|$ denote some norm on $\mathbb{R}^d$, and $\|.\|$ some norm on $\mathbb{R}^{d\times d}$.

Lemma 19. *With the notations of Theorem 17, if there is a sequence of vectors Y_{n_i}, for increasing n_i such that $\lim_{i\to\infty} \log|Y_{n_i}|/\log n_i = c$, then*

$$r_\alpha^+ \geq c.$$

Proof. By construction, the sum of entries in Y_n is the total number of words of length $n-1, n, n+1$ or $n+2$ in $AU\,A$, that is $4u(n-3) + 4u(n-2) + 4u(n-1) + 4u(n)$, which is at most $60u(n-3)$, since $u(n-3+k) \leq 2^k u(n-3)$. As norms on $\mathbb{R}^d$ are equivalent, there exists a positive constant K such that $|Y_n| \leq Ku(n-3)$ for all n. If $\lim_{i\to\infty} \log|Y_{n_i}|/\log n_i = c$, this means that there is a subsequence (n_i') of $(n_i - 3)$ such that $\lim_{i\to\infty} \log u(n_i')/\log(n_i') \geq c$. Then $r_\alpha^+ \geq c$. □

Theorem 20. *Let F_0, F_1 be given by Theorem 17. Then,*

$$r_\alpha^+ = \log_2 \hat{\rho}(\{F_0, F_1\}).$$
$$r_\alpha^- \leq \log_2 \check{\rho}(\{F_0, F_1\}).$$

Proof. First, it is clear that $r_\alpha^+ \leq \log_2 \hat{\rho}(\{F_0, F_1\})$ and $r_\alpha^- \leq \log_2 \check{\rho}(\{F_0, F_1\})$. We have seen in Corollary 18 that $u_\alpha(n)$ can be written as

$$u_\alpha(n) = RF_{d_0}\cdots F_{d_{k-m}} Y_{n_0}$$

with $n_0 < 2^m$ and $k = \lfloor \log_2 n \rfloor$. It follows that

$$u_\alpha(n) = RY_n \leq |R|\, \|F_{d_0}\cdots F_{d_{k-m}}\|\, |Y_{n_0}| \leq K\|F_{d_0}\cdots F_{d_{k-m}}\|,$$

where K is a positive constant. So,

$$r_\alpha^- = \liminf_{n\to\infty} \frac{\log u_\alpha(n)}{\log n} \leq \liminf_{n\to\infty} \frac{\log_2 \|F_{d_0}\cdots F_{d_{k-m}}\|}{\log_2 n} = \log_2 \check{\rho}(\{F_0, F_1\})$$

and

$$r_\alpha^+ = \limsup_{n\to\infty} \frac{\log u_\alpha(n)}{\log n} \leq \limsup_{n\to\infty} \frac{\log_2 \|F_{d_0}\cdots F_{d_{k-m}}\|}{\log_2 n} = \log_2 \hat{\rho}(\{F_0, F_1\}).$$

We now prove the converse for the joint spectral radius. It is well known (see [10]) that there exists an infinite product $\ldots F_{d_2}F_{d_1}$ such that

$$\lim_{k\to\infty} \|F_{d_k}\ldots F_{d_1}\|^{1/k} = \hat{\rho}.$$

So, there is an index j such that

$$\limsup_{k\to\infty} |F_{d_k}\ldots F_{d_1} e_j|^{1/k} = \hat{\rho},$$

where e_j is the jth vector of the canonical basis. Moreover it is clear from the construction of $\mathcal{B}_\alpha$ that for any index j, there is an h such that the vector Y_h has the jth entry larger than zero. Defining $n_k = 2^k h + d_1 2^{k-1} + d_2 2^{k-2} + \cdots + d_{k-1}2 + d_k$, so that $Y_{n_k} = F_{d_k}\ldots F_{d_1} Y_h$, we have that

$$\limsup_{k\to\infty} |Y_{n_k}|^{1/k} = \hat{\rho},$$

and taking logarithms we can apply Lemma 19 to conclude that $r_\alpha^+ \geq \log_2 \hat{\rho}$. □

The joint spectral radius and subradius have been the subject of intense research during the last decade, and, even if they are notoriously difficult to compute, accurate techniques exist to estimate their value. See [10] for a survey. In the next section we apply some of these techniques to find good estimates in the case $\alpha = 7/3$.

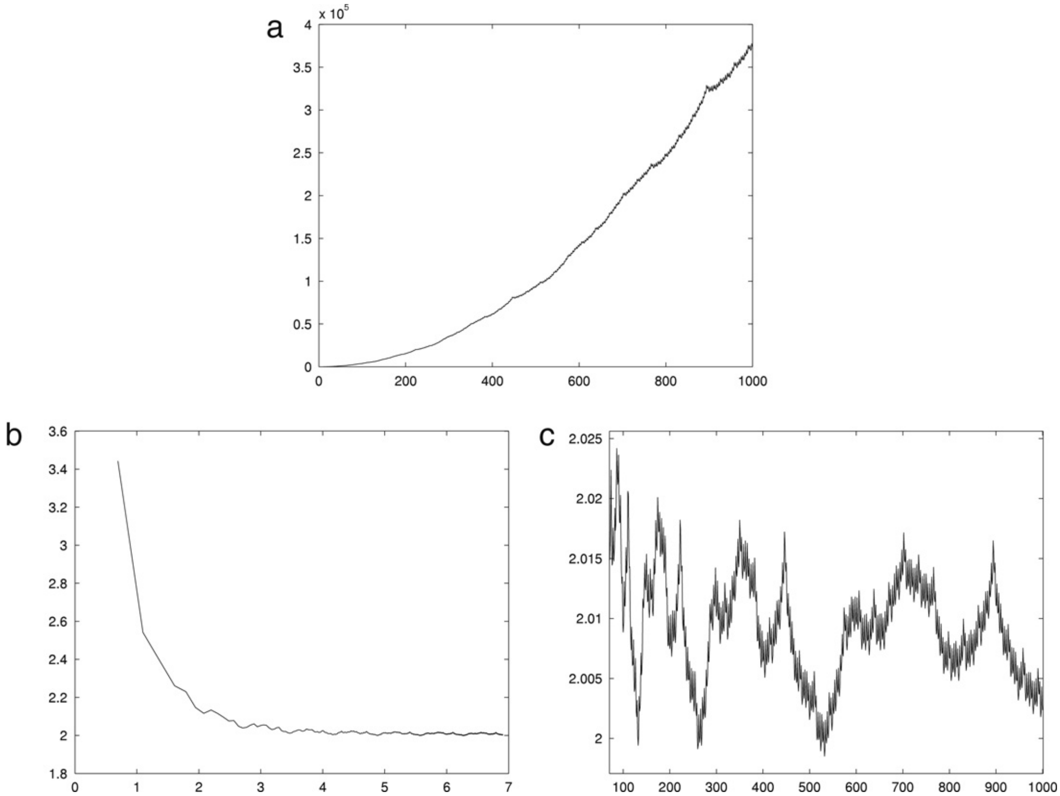

Fig. 1. (a) The values of $u_{7/3}(n)$ for $1 \le n \le 1000$; (b) $(1 + \log u_{7/3}(n))/\log n$ against $\log n$; (c) $(1 + \log u_{7/3}(n))/\log n$ against n, zoomed around the apparent limit.

4. The particular case $\alpha = 7/3$

4.1. The automata

In the particular case of $\alpha = 7/3$, we get an automaton $\mathcal{A}_{7/3}$ with 13 states: $(aaba, -)$, $(aabb, -)$, $(aba, -)$, $(abb, -)$, $(aaba, 0)$, $(aba, 0)$, $(abb, 0)$, $(aaba, 1)$, $(aabb, 1)$, $(aba, 1)$, $(aaa, 2)$, $(aba, 2)$, $(abb, 2)$.

The automaton $\mathcal{B}_{7/3}$ has 141 states, among which 10 are accepting. Four states are transient and can be ignored for asymptotic study, so we get incidence matrices of dimension 137.

4.2. Numerical analysis

Theorem 17 provides a matrix expression for $u_{7/3}(n)$, involving two matrices F_0 and F_1 of dimension 548. In this section we briefly describe the result of numerical analysis of this pair of matrices. Fig. 1 represents the evolution of (a) $u_{7/3}(n)$ and (b, c) $(1 + \log u_{7/3}(n))/\log n$ for the first few n, where $1/\log n$ has been added to accelerate convergence. The graph (b) seems to indicate convergence to 2, but zooming it (c) reveals that the exponent oscillates.

As for overlap-free words, one could wonder whether, for asymptotic study, the dimension of the matrices can be reduced. It appears that there is a permutation of the coordinates that puts F_0 and F_1 under block triangular form. This means that there are several separate strongly connected components in the automaton. The largest connected component has dimension 227 and moreover, it is the only connected component whose submatrices have a norm larger than 2. Since we know that for any increasing sequence of natural numbers n, $u_{7/3}(n)$ grows superlinearly, then the asymptotics are completely ruled by this component.

By analysing the joint spectral quantities of this component, we get: $r^-_{7/3} < 2.0035 < 2.0121 < r^+_{7/3} < 2.1050$. We were not able to find a better lower bound for $r^-_{7/3}$ than the bound for overlap-free words from [11], $1.2690 < r^-_{2+} \le r^-_{7/3}$. If we knew that $r^-_{7/3} = \log_2 \check{\rho}(\{F_0, F_1\})$, then standard techniques for approximating the joint spectral subradius would give $1.8874 < r^-_{7/3}$.

For the sum, we get $\sum_{m<n} u_{7/3}(n) = \Theta(n^{r_{7/3}})$ with $r_{7/3} = 2\log_2 \rho(F_0 + F_1) \simeq 3.0053$.

5. Conclusion and perspectives

In this paper we generalize recent results on overlap-free words to α-power-free words for arbitrary rational $\alpha \in (2, 7/3]$. The generalization is far from being straightforward. As an example, it was known that for overlap-free words, the quantities r^+ and r^- can be expressed in terms of joint spectral quantities of a set of matrices. However, the proof of this result involved a precise and numerical analysis of the matrices F_0 and F_1, so that it was not clear at all that the result could be generalized for arbitrary α. The construction proposed in this paper allows one to derive a proof for arbitrary α, at least for r_α^+; we only get an inequality for r_α^-.

For asymptotic behaviour (that is, the quantities r_α^+ and r_α^-) it is possible to simplify the matrices. For instance, for overlap-free words it was possible to lower the dimension from 30×30 to 20×20. For 7/3-free words, it was possible to lower it from 548×548 to 227×227. This large number for such a "simple" value for α (the denominator is small) is rather discouraging, and it seems that the number of states for the automaton grows very rapidly when the denominator of α increases. This is the minimal dimension that one can reach by applying permutations on the initial matrices. But we do not know whether it is possible to still decrease the dimension with more complex transformations. Also, is it possible to compute a priori the minimum number of states that one needs?

The asymptotic behaviour of the sum $s_\alpha(n) = \sum_{m<n} u_\alpha(n)$ can be described precisely when $\alpha \in (2, 7/3]$ is rational, as we can compute $r_\alpha = \log_2 \rho(F_0 + F_1)$. The next step is to study how this quantity depends on α: is it strictly increasing? What are the discontinuities? How to express r_α if α is not rational? We expect a "Devil's Staircase"-like behaviour.

Clearly, these questions also have interest for the limits r_α^- and r_α^+. However, they seem much more difficult, as even for overlap-free words only approximations are known.

Also very challenging is to adapt this study to a ternary or larger alphabet. To do this, one has to find a replacement for the Thue–Morse morphism and a new structure lemma. It is not obvious that this is even possible: indeed it might be that $RT'(k) = RT(k)$ for $k \geq 3$ (where the threshold $RT'(k)$, introduced by Kobayashi [13], is such that the growth is polynomial when $RT(k) < \alpha < RT'(k)$, and not polynomial when $\alpha > RT'(k)$; Lemma 1 states that $RT'(2) = 7/3$) and there is no more polynomial growth. This is conjectured by Shur [20], and supported by some numerical evidence.

Acknowledgements

Raphaël M. Jungers is a FNRS fellow. His research was partially supported by Communauté française de Belgique — Actions de recherche concertées, by the HYCON Network of Excellence (contract number FP6-IST-511368) and by the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office. The scientific responsibility rests with the authors.

References

[1] J.-P. Allouche, J. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge University Press, 2003.
[2] V.D. Blondel, J.N. Tsitsiklis, A survey of computational complexity results in systems and control, Automatica 36 (9) (2000) 1249–1274.
[3] A. Carpi, Overlap-free words and finite automata, Theoret. Comput. Sci. 115 (2) (1993) 243–260.
[4] A. Carpi, On Dejean's conjecture over large alphabets, Theoret. Comput. Sci. 385 (1–3) (2007) 137–151.
[5] J. Cassaigne, Counting overlap-free binary words. STACS 93, Lecture Notes in Computer Sci. 665 (1993) 216–225.
[6] J.D. Currie, N. Rampersad, Infinite words containing squares at every position, Preprint: http://arxiv.org/abs/0803.1189.
[7] J.D. Currie, N. Rampersad, Dejean's conjecture holds for $n \geq 30$, Theoret. Comput. Sci. 410 (30–32) (2009) 2885–2888. Preprint: http://arxiv.org/abs/0806.0043.
[8] J.D. Currie, N. Rampersad, J. Shallit, Binary words containing infinitely many overlaps, Electron. J. Combin. 13 (2006) #R82.
[9] F. Dejean, Sur un théorème de Thue, J. Combin. Theory Ser. A 13 (1972) 90–99.
[10] R.M. Jungers, Infinite matrix products, from the joint spectral radius to combinatorics, Ph.D. Thesis, Université Catholique de Louvain, 2008.
[11] R.M. Jungers, V. Protasov, V.D. Blondel, Overlap-free words and spectra of matrices, Preprint: http://arxiv.org/abs/0709.1794 submitted for publication.
[12] J. Karhumäki, J. Shallit, Polynomial versus exponential growth in repetition-free binary words, J. Combin. Theory Ser. A 105 (2) (2004) 335–347.
[13] Y. Kobayashi, Repetition-free words, Theoret. Comput. Sci. 44 (1986) 175–197.
[14] M. Lothaire, Combinatorics on Words, in: Encyclopedia of Mathematics, vol. 17, Addison-Wesley, New-York, 1983, Reprinted by Cambridge University Press, 1997.
[15] M. Mohammad-Noori, J.D. Currie, Dejean's conjecture and Sturmian words, Eur. J. Comb. 28 (3) (2007) 876–890.
[16] J.-J. Pansiot, The Morse sequence and iterated morphisms, Inform. Process. Lett. 12 (1981) 68–70.
[17] V.Y. Protasov, The joint spectral radius and invariant sets of linear operators, Fundam. Prikl. Mat. 2 (1) (1996) 205–231.
[18] A. Restivo, S. Salemi, Overlap-free words on two symbols. Automata on infinite words, Lecture Notes in Computer Sci. 192 (1985) 198–206.
[19] A.M. Shur, The structure of the set of cube-free $\mathbb{Z}$-words in a two-letter alphabet, Izv. Ross. Akad. Nauk Ser. Mat. 64 (4) (2000) 201–224 (in Russian), English translation: Izv. Math. 64 (2000) 847–871.
[20] A.M. Shur, I. Gorbunova, On the growth rates of complexity of threshold languages, Journées montoises d'informatique théorique, Université de Mons-Hainaut, 2008.
[21] A. Thue, Über unendliche Zeichenreihen, Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl. 7 (1906) 1–22.
[22] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl. 1 (1912) 1–67.