Journées de numération, Prague, May 26th, 2008

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

Julien Cassaigne
Institut de mathématiques de Luminy - CNRS, Marseille, France
cassaigne@iml.univ-mrs.fr

Raphaël Jungers
CESAME, Université catholique de Louvain, Belgique
Raphaël Jungers is an FNRS fellow

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

- Words avoiding fractional powers

- The threshold between polynomial and exponential growth

- Known results on overlap-free binary words

- The number of $7/3$-power-free binary words

- General case

Words avoiding fractional powers

Let $w = w_1 \ldots w_n \in A^*$ be a non-empty finite word, and $n = |w|$.

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 exponent of w is the rational number $e(w) = n/p$.

Example: $e(abacabacab) = 10/4 = 5/2$.

Let $\alpha \in \mathbb{R}$. A word $v \in A^* \cup A^\omega$ is α-power-free if every finite factor w of v satisfies $e(w) < \alpha$.

v is α^+-power-free if every finite factor w of v satisfies $e(w) \leq \alpha$.

Repetition thresholds (1)

It is well known (Thue 1906) that there are only finitely many binary square-free (i.e., 2-power-free) words, and that the infinite Thue-Morse word is overlap-free (i.e., 2^+-power-free).

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 $k \leq 14$ (Thue 1906, Dejean 1972, Pansiot 1984, Moulin-Ollagnier 1992, Currie and Mohammad-Noori 2005) and for $k \geq 33$ (Carpi 2007).

Repetition thresholds (2)

How fast does the number $u_{k,\alpha}(n)$ of α-power-free words of length n on a given alphabet with k letters grow as a function of n ?

At least two behaviours are possible: the number of overlap-free binary words grows polynomially (Restivo and Salemi 1985), whereas the number of cube-free binary words grows exponentially (Brandenburg 1983).

We can define (Kobayashi 1983) a new threshold $RT'(k)$ such that the growth is polynomial when $\alpha < RT'(k)$, and not polynomial when $\alpha > RT'(k)$ (no example is known with intermediate growth, neither polynomial nor exponential, but this case cannot be a priori excluded).

Karhumäki and Shallit's result

Karhumäki and Shallit computed $RT'(2)$ in 2004: $RT'(2) = 7/3$.

More precisely, there are polynomially many $7/3$-power-free binary words, and exponentially many $7/3^+$-power-free binary words.

For $k > 2$, not much is known about $RT'(k)$, only that $RT'(3) \leq 2$ (Brandenburg 1983).

Structure lemma

The main ingredient in the proof is the following structure lemma (generalizing a result of Restivo and Salemi 1985 for overlaps):

Lemma. Let $A = \{a, b\}$, $2 < \alpha \leq 7/3$, 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 $|w| \geq 7$.

Example: $bbabbaabbaa = bb.\theta(abab).a$ (note that this word contains an overlap, and is a maximal $7/3$-power-free word).

This lemma fails for $\alpha > 7/3$, for instance with
$w = abbabaabbaabaabbabaab$.

How to compute the growth exactly ?

Our goal is to compute exactly, or at least asymptotically, the numbers $u_{k,\alpha}(n)$.

Namely, we are interested in:

- if $u_{k,\alpha}(n)$ is polynomial in n, find its degree;
- if $u_{k,\alpha}(n)$ is exponential in n, find its basis;
- find recurrence relations to compute $u_{k,\alpha}(n)$ efficiently.

Iterating the structure lemma

Iterating the structure lemma produces 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 $(x_1, y_1), \ldots, (x_n, y_n)$ are enough to describe w.

Unfortunately, not all sequences $((x_i, y_i))$ are admissible. For overlap-free words, Carpi proved in 1993 that admissible sequences form a regular language. As a consequence, $u_{2,2^+}(n)$ is a 2-regular sequence in the sense of Allouche and Shallit 1992. However it is not easy to compute an automaton explicitely.

An alternative structure lemma for overlap-free words

Computation becomes easier if the additive structure lemma is replaced by a subtractive variant (Cassaigne 1993).

Let U be the set of overlap-free binary words, V the set of minimal overlaps, S a finite set of short words, and $E = \{\kappa, \delta, \iota\}$ a set of transformations acting on either end of a word and defined as follows:

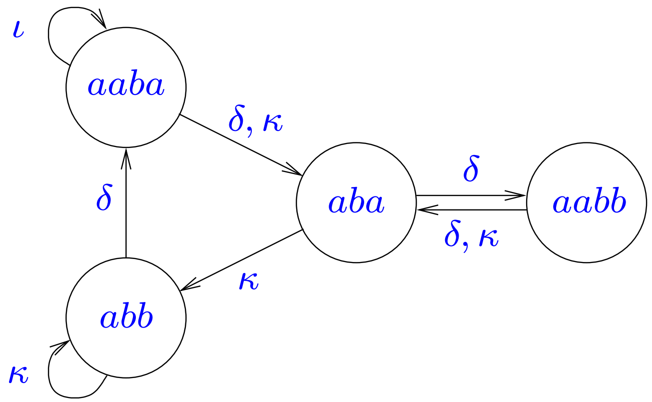
- κ does nothing;
- δ deletes the first (or last) letter;
- ι inverts the first (or last) letter.

Lemma. 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 $w = \gamma_1.\theta(v).\gamma_2$.

One-sided control for overlap-free words

The prefix type of a word w with $|w| \geq 4$ is $t_1 \in \{aaa, aaba, aabb, aba, abb\}$ such that w starts with t_1 or $\overline{t_1}$.

If $w = \gamma.\theta(v)$ with $v, w \in (U \cup V) \setminus S$, then their types are related like this:



Note that ι is not always permitted, and type aaa is excluded.

Results on overlap-free binary words

Using the subtractive structure lemma, we obtained in 1993:

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

This theorem allows to compute $u(n) = u_{2,2^+}(n)$ very efficiently, using the binary expansion of n to construct a product of the matrices F_0 and F_1.

Asymptotic behaviour

A surprising corollary is that, although $u(n)$ grows polynomially, it does not have a fixed degree.

Let $r^- = \liminf \frac{\log u(n)}{\log n}$ and $r^+ = \limsup \frac{\log u(n)}{\log n}$.

Then, considering subsequences $u(2^m)$ and $u(\frac{4^m-1}{3})$, we get

$r^- \leq \log_2 \rho(F_0) < \log_4 \rho(F_0 F_1) \leq r^+$.

Using joint spectral radius techniques, Jungers, Blondel and Protasov obtained in 2007:

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

Asymptotic behaviour of the sum

The function $s(n) = \sum_{m<n} u(m)$ is tamer and easier to compute. It counts the number of overlap-free binary words of length at most n.

We proved in 1993 that $s(n) = \Theta(n^r)$ with

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

Actually, $r = 2\log_2 \rho(M)$ where M is the 4×4 incidence matrix of the one-sided control automaton:

$$M = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Subtractive structure lemma for α-power-free words

Let $2 < \alpha \leq 7/3$ (α may also be a "plussed" real). We want to adapt our technique to α-power-free words.

Let U be the set of α-power-free binary words, $V = AUA \setminus U$ the set of almost α-power-free words, S a finite set of short words, and $E = \{\kappa, \delta, \iota, \sigma, \tau\}$ the set of transformations: $\kappa.xw = xw$, $\delta.xw = w$, $\iota.xw = \bar{x}w$, $\sigma.xw = x\bar{x}w$, $\tau.xw = \bar{x}\bar{x}w$, and $w.\gamma$ is the mirror image of $\gamma.\tilde{w}$, for any $\gamma \in E$.

Lemma. 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 $w = \gamma_1.\theta(v).\gamma_2$.

One-sided control for 7/3-power-free words (1)

Unlike with overlap-free words, V contains words that start or end with $7/3$-powers, but are not $7/3$-powers themselves.

Let the prefix excess of w be $f_1 \in \{1, 2, 3\}$ if w starts with xxy where $|y| = \frac{|x|+f_1-1}{3}$, $f_1 = 0$ otherwise.

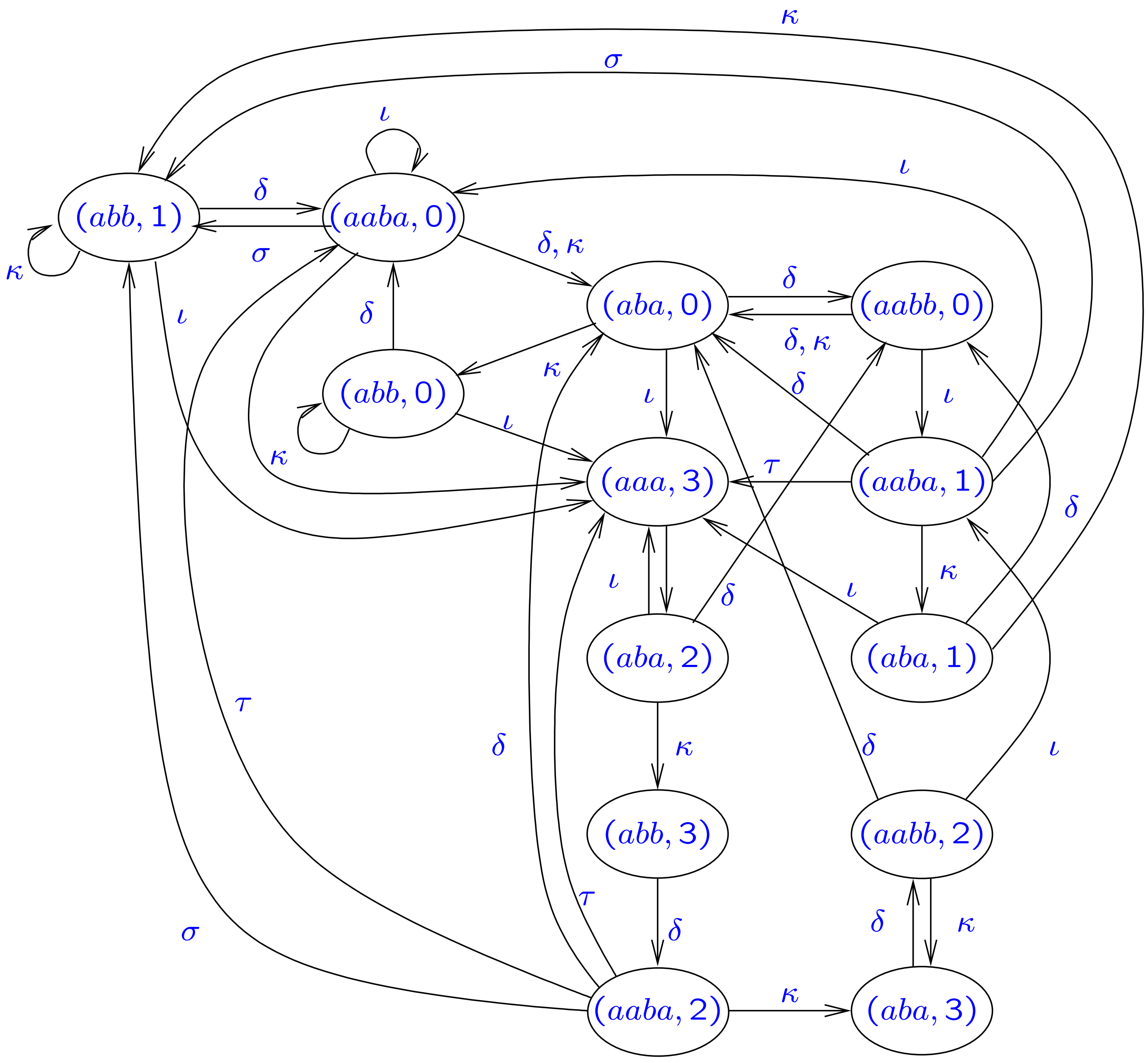
If $w = \gamma.\theta(v)$ with $v, w \in (U \cup V) \setminus S$, then their types and excesses are related by an automaton with 13 states: $(aaba, 0)$, $(aabb, 0)$, $(aba, 0)$, $(abb, 0)$, $(aaba, 1)$, $(aba, 1)$, $(abb, 1)$, $(aaba, 2)$, $(aabb, 2)$, $(aba, 2)$, $(aaa, 3)$, $(aba, 3)$, $(abb, 3)$. Only the first four correspond to words in U.

One-sided control for 7/3-power-free words (2)

For instance, starting with $(aaba, 0)$: $\theta(v) = ababbaab...$ so κ and δ lead to $(aba, 0)$ and ι to $(aaba, 0)$ as before; $\sigma.\theta(v) = abbabbaab...$ leads to $(abb, 1)$ because of the initial repetition $abbabba$, and $\tau.\theta(v) = bbbabbaab...$ leads to $(aaa, 3)$.

Starting with $(aaa, 3)$, the only possibility is δ that leads to $(aba, 2)$. Other transformations lead to words outside $U \cup V$, for instance $\iota.\theta(v) = bbabab...$ ($5/2 \geq 7/3$).

Starting with $(aba, 2)$, δ leads back to $(aabb, 0)$ in U, κ leads to $(abb, 3)$ and ι to $(aaa, 3)$. Other transformations are not permitted.



(abb, 1)
(aaba, 0)
(aba, 0)
(aabb, 0)
(abb, 0)
(aaa, 3)
(aaba, 1)
(aba, 2)
(aba, 1)
(abb, 3)
(aabb, 2)
(aaba, 2)
(aba, 3)
κ
σ
ι
δ
δ, κ
τ

Results on 7/3-power-free binary words

Transforming this automaton to take into account two-sided control and variations in the lengths of words, we finally get a matrix expression for $u_{2,7/3}(n)$, involving two matrices F_0 and F_1 of dimension 548.

For asymptotic study, the matrices can be reduced to dimension 227 (one strongly connected component in the associated automaton). We get: $1.8874 < r^-_{2,7/3} < 2.0035 < 2.0121 < r^+_{2,7/3} < 2.1050$.

For the sum, we get $s_{2,7/3}(n) = \sum_{m<n} u_{2,7/3}(m) = \Theta(n^{r_{2,7/3}})$ with $r_{2,7/3} = 2\log_2 \rho(M) \simeq 3.0053$ where M is the 13×13 incidence matrix of the one-sided control automaton.

Generalization (1)

Let $\alpha = p/q$, $2 < \alpha \leq 7/3$. Define the prefix excess of a word w as the maximal value f_1 of $\max(0, q|z| - p|x| + 1)$, where $z = xxy$ is a proper prefix of w of period $|x|$. Note that $f_1 = 0$ if $w \in U$, $0 \leq f_1 \leq q$ if $w \in V$. Define the suffix excess f_2 similarly. Define the global excess g as $\max(0, q|w| - p|x| + 1)$, where $|x|$ is a period of $|w|$. Note that $0 \leq g \leq \min(f_1, f_2) + q$.

We construct an automaton with states (t_1, f_1, t_2, f_2, g), that is at most $25.(q+1)^2.(2q+1)$ states, and transitions labelled by $E \times E$. For instance, (κ, δ) maps (aba, f_1, aaa, f_2, g) to:
$(abb, \max(0, 2f_1 - 1), aba, \max(0, 2f_2 - 1 - q), \max(0, 2g - 1 - q))$, provided these values remain in the allowed range.

Generalization (2)

We first get recurrence relations $X_{2n} = G_{-2}X_{n+1} + G_0 X_n + G_2 X_{n-1}$ and $X_{2n+1} = G_{-1}X_{n+1} + G_1 X_n$, where X_n is the vector that counts the number of words of length n that fall in each state, and the G_j are the matrices corresponding to transitions (γ_1, γ_2) such that $|\gamma_1.\theta(v).\gamma_2| = 2|v| + j$.

Multiplying the dimension by four, we can turn these relations into $Y_{2n} = F_0 Y_n$, $Y_{2n+1} = F_1 Y_n$ where

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

Perspectives

We can now easily compute $r_{2,\alpha}$ and r_{2,α^+} when $\alpha \in \mathbb{Q}$. The next step is to study how these depend on α: are they strictly increasing? What are the discontinuities? We expect a devil-staircase-like behaviour.

The study of the limits $r^-_{2,\alpha}$ and $r^+_{2,\alpha}$ seems much more difficult, as even for overlaps only approximations are known.

Also very challenging is to adapt this 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 at all possible: maybe $RT'(k) = RT(k)$ and there is no more polynomial growth.