Finite-Repetition threshold for infinite ternary words

Golnaz Badkobeh
King's College London, UK
golnaz.badkobeh@kcl.ac.uk

Maxime Crochemore
King's College London, UK
Université Paris-Est, France
maxime.crochemore@kcl.ac.uk

The exponent of a word is the ratio of its length over its smallest period. The repetitive threshold $r(a)$ of an a-letter alphabet is the smallest rational number for which there exists an infinite word whose finite factors have exponent at most $r(a)$. This notion was introduced in 1972 by Dejean who gave the exact values of $r(a)$ for every alphabet size a as it has been eventually proved in 2009.

The finite-repetition threshold for an a-letter alphabet refines the above notion. It is the smallest rational number $\mathrm{FRt}(a)$ for which there exists an infinite word whose finite factors have exponent at most $\mathrm{FRt}(a)$ and that contains a finite number of factors with exponent $r(a)$. It is known from Shallit (2008) that $\mathrm{FRt}(2) = 7/3$.

With each finite-repetition threshold is associated the smallest number of $r(a)$-exponent factors that can be found in the corresponding infinite word. It has been proved by Badkobeh and Crochemore (2010) that this number is 12 for infinite binary words whose maximal exponent is $7/3$.

We show that $\mathrm{FRt}(3) = r(3) = 7/4$ and that the bound is achieved with an infinite word containing only two $7/4$-exponent words, the smallest number.

Based on deep experiments we conjecture that $\mathrm{FRt}(4) = r(4) = 7/5$. The question remains open for alphabets with more than four letters.

Keywords: combinatorics on words, repetition, repeat, word powers, word exponent, repetition threshold, pattern avoidability, word morphisms.

MSC: 68R15 Combinatorics on words.

1 Introduction

The article deals with repetitions in strings and their avoidability. The question is grounded on the notion of the exponent of a word: it is the ratio of its length over its smallest period. A word of exponent e is also called an e-power.

An infinite word is said to avoid e-powers or to be e-power free if the exponents of its finite factors are smaller than e.

The repetitive threshold $r(a)$ of an a-letter alphabet is the smallest rational number for which there exists an infinite word whose finite factors have exponent at most $r(a)$. The word is said to be $r(a)^+$-power free. It is known from Thue [14] that $r(2) = 2$. Indeed, the notion was introduced in 1972 by Dejean [5] who proved that $r(3) = 7/4$ and gave the exact values of $r(a)$ for every alphabet size $a > 3$. Her conjecture was eventually proved in 2009 after partial proofs given by several authors (see [12, 4] and references therein).

A generalised version of the repetitive threshold is by Ilie et al. [8]. The authors introduce the notion of (β, p)-freeness: a word is (β, p)-free if it contains no factor that is a (β', p')-repetition (i.e. a word of period p' and exponent β') for $\beta' \geq \beta$ and $p' \geq p$; it is (β^+, p)-free if $\beta' > \beta$ instead. Their generalized repetition threshold $R(a, p)$ defined for an a-letter alphabet as the real number α for which either

P. Ambrož, Š. Holub and Z. Masáková (Eds.):
8th International Conference WORDS 2011
EPTCS 63, 2011, pp. 37–43, doi:10.4204/EPTCS.63.7

(a) there exists an (α^+, p)-free infinite word and all (α, p)-free words are finite,

(b) or there exists an (α, p)-free infinite word and for all $\varepsilon > 0$, $(\alpha - \varepsilon, p)$-free words are finite.

where p is the minimum avoided period. The proof of boundary of this threshold for all alphabet sizes is presented in [8], and $R(k,1)$ is essentially Dejean's repetition threshold.

For infinite words whose maximal exponent of factors is bounded, it is legitimate to ask whether they can contain only a finite number of $r(a)$-powers. This is an extra constraint on the word. When such words exist, it is as legitimate to exhibit the minimal number of $r(a)$-powers they can contain, which adds another measure of the word complexity. The first result of this type is by Fraenkel and Simpson [6] for the binary alphabet. They showed that an infinite binary word can contain only 3 squares, not less. Two simple proofs of the result are by Harju and Nowotka [7] and the present authors [2].

The above consideration leads to the notion of finite-repetition threshold associated with an a-letter alphabet. It is the smallest rational number, noted $\mathrm{FRt}(a)$, for which there exists an infinite word whose finite factors have exponent at most $\mathrm{FRt}(a)$ and that contains a finite number of $r(a)$-powers. It is known from Shallit [13] that $\mathrm{FRt}(2) = 7/3$ (see also [11]). The present authors [3] proved that the associated minimal number of squares is 12 if the infinite word contains two $7/3$-powers. Badkobeh [1] even refined the results by showing the number is 8 if the infinite word admits two $5/2$-powers, extending the result of Fraenkel and Simpson [6] recalled above for which it is 3 if two cubes are allowed.

In this article, we consider the finite-repetition threshold of the ternary alphabet. We show that $\mathrm{FRt}(3) = r(3) = 7/4$. We provide a direct proof of the result and another proof based on a previous result on word morphisms by Ochem [10].

The experiments reported in the conclusion show that the finite-repetition threshold of the 4-letter alphabet $\mathrm{FRt}(4)$ is likely to be $r(4) = 7/5$, which we conjecture. The hypothetical property $\mathrm{FRt}(a) = r(a)$ (for $a > 2$) would be equivalent to say that infinite words whose maximal exponent of factors is Dejean's repetition threshold can be constrained to containing a finite number of factors with that exponent.

2 Repetitions in ternary words

Let A be a finite alphabet. A word w in A^* of length $|w| = n$ is a sequence of letters $w[0]w[1]\dots w[n-1]$ also noted $w[0\,.\,.\,n-1]$. The period of w is the smallest positive integer $\mathit{period}(w) = p$ for which $w[i] = w[i+p]$ whenever both sides of the equality are defined. The exponent of w is the rational ratio $|w|/\mathit{period}(w)$. Thus, the exponent of a word is a rational number that is at least 1. For example, a square is a nonempty word with an even integer exponent and `1020102` of exponent $7/4$ can be written $(\mathtt{1020})^{7/4}$. A word of exponent e is also called an e-power.

An infinite word is a function from the natural number to the alphabet A. An infinite word is said to avoid e-powers (resp. e^+-powers) if the exponents of its finite factors are smaller than e (resp. not more than e). In this case we also say that the word is e-power free (resp. e^+-power free).

The repetitive threshold $r(a)$ of an a-letter alphabet is the smallest rational number for which there exists an infinite word whose finite factors have exponent at most $r(a)$. The word is then $r(a)^+$-power free.

The *finite-repetition threshold* for the alphabet of a letters is defined as the smallest rational number $\mathrm{FRt}(a)$ for which there exists an infinite word that both avoids $\mathrm{FRt}(a)^+$-powers and contains a finite number of r-powers, where r is Dejean's repetitive threshold.

The above notion is inspired by the following results of Karhumäki and Shallit. [13]

Theorem 1 (Karhumäki and Shallit [9]) *For all $t \geq 1$, there are no infinite binary words that simultaneously avoid all squares yy with $|y| \geq t$ and $7/3$-powers.*

Theorem 2 (Shallit [13]) *There is an infinite binary word that simultaneously avoids all squares yy with* $|y| \geq 7$ *and* $7/3^+$*-powers.*

When an infinite binary word avoids $7/3^+$-powers and contains a finite number of squares it is natural to ask more on these few squares. The previous theorem shows that their period can be bounded by 7. The next result goes slightly beyond by showing that their number is at least 12. But the two properties cannot be satisfied simultaneously.

Theorem 3 (Crochemore and Badkobeh [3]) *The smallest number of squares occurring in a* $7/3$*-power free infinite binary word is* 12*.*

Showing that no infinite word satisfying the conditions can contain less squares is done by mere computation. The second part is done by producing an infinite word satisfying the condition and containing exactly 12 squares. Following Shallit's hierarchy of infinite binary words in [13], the previous result was refined by Badkobeh [1] according to the next table.

Maximal exponent e	Allowed number of e-powers	Smallest number of squares
7/3	2	12
	1	14
5/2	2	8
	1	11
3	2	3
	1	4

The main result of the present article is the following theorem.

Theorem 4 *The finite-repetition threshold of the 3-letter alphabet is its Dejean's repetition threshold, that is,* $7/4$*.*
The smallest number of $7/4$*-powers occurring in a* $7/4^+$*-power free infinite ternary word is* 2*.*

On the alphabet $\{0,1,2\}$, the two unavoidable $7/4$-powers occurring in the word below are, up to a permutation of letters, $(0121)^{7/4} = 0121012$ and $(2010)^{7/4} = 2010201$.

Computation shows that the longest ternary words with only one $7/4$-power are 102 letters long. However we may think of having a larger threshold as in the binary case. But even if we increase the threshold to $e < 2$, the maximal length of words stays at 102 with only one e-power.

If we relax further the maximal exponent condition, it can be shown that there exists an infinite ternary word in which occur only one square, namely 00 up to a permutation of letters, and no e-power with $7/4 \leq e < 2$.

Since the repetition threshold for a 3-letter alphabet is $7/4$, to prove this ratio is also its finite-repetition threshold it is sufficient to show (contrary to the binary case) that there exists a $7/4^+$-free infinite ternary word with finitely many $7/4$-powers. To do it, we use the fact that the repetition threshold of 4-letter alphabets is $7/5$ and provide a translation morphism from 4 letters to 3 letters with suitable conditions.

We consider the morphism g from $\{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}^*$ to $\{\mathtt{0},\mathtt{1},\mathtt{2}\}^*$ defined by:

$$\begin{cases} g(\mathtt{a}) = & \mathtt{01021012021020102101210201202101201021201210201202101 21} \\ & \mathtt{0212010210121020102120121020120210121020102101202102012} \\ & \mathtt{102120102101210201202101201021201210201021012 10212}, \\ g(\mathtt{b}) = & \mathtt{0102101202102010210121020120210120102120121020120210121} \\ & \mathtt{0201021012021020121021201021012102010212012102012021012} \\ & \mathtt{10212010210121020120210120102120121020102101210212}, \\ g(\mathtt{c}) = & \mathtt{0102101202102010210121020120210120102120121020102101202} \\ & \mathtt{1020121021201021012102010212012102012021012102010210120} \\ & \mathtt{21020102120121020120210120102120121020102101210212}, \\ g(\mathtt{d}) = & \mathtt{0102101202102010210121020120210120102120121020102101202} \\ & \mathtt{1020102120121020120210121020102101202102012102120102101} \\ & \mathtt{21020102120121020120210120102120121020102101210212}. \end{cases}$$

The morphism is uniform with codeword length 160. Another presentation of the morphism g is:

$$\begin{cases} g(\mathtt{a}) = uv\mathtt{0212012102012021012102010210120210201210212010 2101}yz, \\ g(\mathtt{b}) = uv\mathtt{2102120102101210201021201210201202101210201021 0120}yz, \\ g(\mathtt{c}) = uw\mathtt{0102101202102012102120102101210201021201210201 2021}xz, \\ g(\mathtt{d}) = uw\mathtt{1201021012102010212012102012021012102010210120 2102}xz, \end{cases}$$

where u, v, w, x, y and z are:
$u = \mathtt{01021012021020102101210201202101201021201210201}$,
$v = \mathtt{2021012102}$, $w = \mathtt{0210120210201}$, $x = \mathtt{2102010212}$, $y = \mathtt{0121021201021}$,
$z = \mathtt{0121020120210120102120121020102101210212}$.
The word u is the longest common prefix of the codewords, $|u| = 47$, and z is their longest common suffix, $|z| = 40$.

Theorem 4 is a direct consequence of the next proposition.

Proposition 1 *The morphism g translates any infinite $7/5^+$-free word on the alphabet $\{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}$ into a $7/4^+$-free ternary word containing only two $7/4$-powers, the fewest possible.*

We present two proofs of Proposition 1. The first one is a direct proof that involves the longest common prefix and the longest common suffix of the codewords to derive a contradiction from the existence of any $7/4$-power other than $\mathtt{0121012}$ and $\mathtt{2010201}$ in the image by g of a $7/5^+$-free word. The second proof is derived from a lemma on morphisms stated by Ochem in [10].

Direct proof of Proposition 1

Let us assume that $g(s)$ contains a non-extensible repetition, excluding the two $7/4$-powers $\mathtt{0121012}$ and $\mathtt{2010201}$, with exponent at least $7/4$. The repetition can be written pq where $|p|$ is its period. Then $|pq|/|p| \geq 7/4$. A simple computation verifies that no image of a $7/5^+$-free word with length at most 3 contains the repetition. Therefore the repetition is long and occurs in the image by g of a word of length at least 4.

We consider two cases.

- Case $|p| \leq |q|$. The word pq is of the form

$$pq = \overbrace{u_1 \underbrace{\cdots}_{} v_1}\overbrace{u_1 \underbrace{\cdots}_{} v_1}\cdots$$

where u_1v_1 is codeword. Indeed it starts with the square pp of the form

$$\overbrace{u_1g(s')v_1}\overbrace{u_1g(s')v_1}$$

where $s' \in \{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}^*$.

Note that s' cannot be the empty word because pq would occur in the image of a triplet.

Let $\alpha \in \{a,b,c,d\}$ be such that $g(\alpha) = v_1u_1$. Therefore $s'\alpha s'$ is a factor of s. The letter occurring before s' in s and the letter occurring after it must differ from α to avoid the squares $\alpha s'\alpha s'$ or $s'\alpha s'\alpha$ since s is $7/5^+$-power free).

Then u_1 is not longer than the longest common prefix between two different codewords, that is, $|u_1| \le |uw| = 60$. Symmetrically, v_1 is not longer than the longest common suffix of two different codewords, that is, $|v_1| \le |yz| = 53$. But then $|v_1u_1| \le 113$ and cannot be a complete codeword, a contradiction.

- Case $|p| > |q|$. The word pq is of the form

$$pq = \overbrace{u_1\underbrace{\cdots}v_1}\cdots\overbrace{u_1\underbrace{\cdots}v_1}$$

More precisely a_0pqb_1 is of the form

$$a_0\underbrace{u_1g(s')v_1}a_1\cdots b_0\underbrace{u_1g(s')v_1}b_1$$

where $s' \in \{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}^*$, $a_0,a_1,b_0,b_1 \in \{\mathtt{0},\mathtt{1},\mathtt{2}\}$ and $a_0 \neq b_0$ and $a_1 \neq b_1$, because pq is inextensible. It rewrites as

$$a_0u_1g(s')g(s'')g(s')v_1b_1$$

where $g(s'') = v_1q^{-1}pu_1$ because the morphism is synchronizing (no codeword occurs in the concatenation of two codewords). Therefore $g(s')g(s'')g(s')$ is a factor of $g(s)$ thus $s's''s'$ is a factor of s and since s is $7/5^+$-free we get

$$\frac{|s's''s'|}{|s's''|} \le \frac{7}{5}$$

and

$$3|s'| \le 2|s''|$$

and eventually

$$3|g(s')| \le 2|g(s'')| \tag{1}$$

because the morphism g is uniform.

Furthermore $pq = u_1g(s')g(s'')g(s')v_1$ and $p = u_1g(s')g(s'')u_1{}^{-1}$ so its exponent satisfies

$$\frac{|u_1g(s')g(s'')g(s')v_1|}{|g(s')g(s'')|} \quad \ge \quad \frac{7}{4}$$

which rewrites as

$$|g(s')| + 4|u_1v_1| \quad \ge \quad 3|g(s'')| \tag{2}$$

Using Equations 1 and 2 we get

$$\begin{aligned} 9|g(s')| &\leq 6|g(s'')| \\ &\leq 2(|g(s')| + 4|u_1v_1|) \end{aligned}$$

and then

$$|g(s')| \leq \frac{8}{7}|u_1v_1|.$$

But since $|u_1v_1| \leq 113$ as in the first case, this implies that s' is empty. Therefore the repetition pq is a factor of the image of a triplet, a contradiction.

This completes the direct proof of Proposition 1.

Proof based on Ochem's result

Here we split the proof in two parts: first we show that $g(s)$ is $7/4^+$-free, second we show the only $7/4$-powers are the ones mentioned above. The proof depends on the following result. In the statement, Σ_s (resp. Σ_e) is an alphabet with s (resp. e) letters; and the morphism $h : \Sigma_s^* \to \Sigma_e^*$ is synchronizing if for any $a,b,c \in \Sigma_s$ and $v,w \in \Sigma_e^*$, $h(ab) = vh(c)w$ implies either $v = \varepsilon$ and $a = c$ or $w = \varepsilon$ and $b = c$.

Lemma 1 (Ochem [10]) *Let $\alpha,\beta \in \mathbb{Q}$, $1 < \alpha < \beta < 2$, and $p \in \mathbb{N}^*$. Let $h : \Sigma_s^* \to \Sigma_e^*$ be a synchronizing q-uniform morphism (with $q \geq 1$). If $h(w)$ is (β^+, p)-free for every α^+-free word w such that $|w| < \max\{\frac{2\beta}{\beta-\alpha}, \frac{2(q-1)(2\beta-1)}{q(\beta-1)}\}$, then $h(t)$ is (β^+, p)-free for every (finite or infinite) α^+-free word t.*

To apply the lemma to the morphism g above, we have $\alpha = 7/5$ and $q = 160$. We choose $\beta = 17/10$ and $p = 5$. Then we can show that the morphism is $(17/10^+, 5)$-free if $g(w)$ is $(17/10^+, 5)$-power free for all words w for which

$$|w| < \max\{\frac{2\beta}{\beta-\alpha}, \frac{2(q-1)(2\beta-1)}{q(\beta-1)}\},$$

which implies $|w| < 12$. This set is finite and a simple computation can verify the claim.

Since every $(7/4^+, 5)$-power is also a $(17/10^+, 5)$-power then we can claim the morphism is $(7/4^+, 5)$-power free.

So the only possible $7/4$-powers with period less that 5 are:
$(0121)^{7/4}$, $(0212)^{7/4}$, $(1020)^{7/4}$, $(1202)^{7/4}$, $(2010)^{7/4}$, and $(2101)^{7/4}$. Any of those strings must be either a factor of a codeword or a factor of the image of a doublet. We immediately conclude the words $(1020)^{7/4}$ and $(2101)^{7/4}$ are the only factors of $g(s)$.

This concludes the whole proof of Theorem 4.

3 Repetitions for larger alphabets

Experiments show that a word on a 4-letter alphabet for which the maximal exponent of factors is $7/5$ and that contains at most one $7/5$-power has maximal length 230. However if the constraint on the number of $7/5$-powers is relaxed to 2 the length grows to at least 100000. This experiment intrigued us to study the string further and to state the following conjecture.

Conjecture 1 *The finite-repetition threshold of 4-letter alphabets is $7/5$ and their exist an infinite $7/5^+$-power free word containing only two $7/5$-powers.*
The two $7/5$-powers are $(0231203213)^{\frac{7}{5}}$ and $(1230213203)^{\frac{7}{5}}$ up to a permutation of the letters.

Note the above words have period 10, the smallest possible period, since there is no $7/5$-power with period 5 that is $7/5^+$-free.

The experiments are done with a mere backtracking technique to generate the suitable words. It implements efficient algorithms for testing the properties.

Acknowledgments

We warmly thank Pascal Ochem for pointing his powerful result on word morphisms and for profitable discussions.

References

[1] Golnaz Badkobeh (2011): *Fewest repetitions vs maximal-exponent powers in infinite binary words*. *Theoret. Comput. Sci.* In press.

[2] Golnaz Badkobeh & Maxime Crochemore (2010): *An infinite binary word containing only three distinct squares* Submitted.

[3] Golnaz Badkobeh & Maxime Crochemore (2011): *Fewest repetitions in infinite binary words*. *RAIRO - Theoretical Informatics and Applications* DOI: 10.1051/ita/2011109. In press.

[4] James D. Currie & Narad Rampersad (2011): *A proof of Dejean's conjecture*. *Math. Comput.* 80(274), pp. 1063–1070. Available at `http://dx.doi.org/10.1090/S0025-5718-2010-02407-X`.

[5] Françoise Dejean (1972): *Sur un Théorème de Thue*. *J. Comb. Theory, Ser. A* 13(1), pp. 90–99.

[6] Aviezri S. Fraenkel & Jamie Simpson (1995): *How Many Squares Must a Binary Sequence Contain?* *Electr. J. Comb.* 2.

[7] Tero Harju & Dirk Nowotka (2006): *Binary Words with Few Squares*. *Bulletin of the EATCS* 89, pp. 164–166.

[8] Lucian Ilie, Pascal Ochem & Jeffrey Shallit (2005): *A generalization of repetition threshold*. *Theor. Comput. Sci.* 345(2-3), pp. 359–369. Available at `http://dx.doi.org/10.1016/j.tcs.2005.07.016`.

[9] Juhani Karhumäki & Jeffrey Shallit (2004): *Polynomial versus exponential growth in repetition-free binary words*. *J. Comb. Theory, Ser. A* 105(2), pp. 335–347. Available at `http://dx.doi.org/10.1016/j.jcta.2003.12.004`.

[10] Pascal Ochem (2006): *A generator of morphisms for infinite words*. *Theor. Inform. Appl.* 40(3), pp. 427–441.

[11] Narad Rampersad, Jeffrey Shallit & Ming Wei Wang (2005): *Avoiding large squares in infinite binary words*. *Theor. Comput. Sci.* 339(1), pp. 19–34. Available at `http://dx.doi.org/10.1016/j.tcs.2005.01.005`.

[12] Michaël Rao (2011): *Last cases of Dejean's conjecture*. *Theor. Comput. Sci.* 412(27), pp. 3010–3018. Available at `http://dx.doi.org/10.1016/j.tcs.2010.06.020`.

[13] Jeffrey Shallit (2004): *Simultaneous avoidance of large squares and fractional powers in infinite binary words*. *Intl. J. Found. Comput. Sci* 15(2), pp. 317–327.

[14] Axel Thue (1906): *Über unendliche Zeichenreihen*. *Norske vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana* 7, pp. 1–22.