Journal of Discrete Algorithms 20 (2013) 38–42

Contents lists available at SciVerse ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

Infinite words containing the minimal number of repetitions

Golnaz Badkobeh

King's College London, London, UK

ARTICLE INFO

Article history:
Available online 15 March 2013

Keywords:
Combinatorics on words
Repetitions
Word morphisms

ABSTRACT

A square in a word is composed of two adjacent occurrences of a nonempty word. This note gives a simple proof and a straight construction of the existence of an infinite binary word that contains only three squares. No infinite binary word can contain fewer squares. The only factors of exponent larger than two that our infinite binary word contains are two cubes. Furthermore, we provide two additional results on alphabets of size 3 and 4. We prove that there exists an infinite overlap-free ternary word containing only one square. On a 4-letter alphabet we show there exists an infinite $3/2^+$-free 4-ary word containing only one $3/2$-power.

1. Introduction

The number of repetitions in infinite words is a classic problem in combinatorics on words which, for the last century, has been studied in depth. Let $g(n)$ be the length of a longest binary word containing at most n squares. Then $g(0) = 3$ (e.g. `010`), $g(1) = 7$ (e.g. `0001000`) and $g(2) = 18$ (e.g. `010011000111001101`).

The question of behaviour of this function was posed by Erdös [6]. Entringer, Jackson, and Schatz [5] showed in 1974 that there exists an infinite word with 5 different squares, yielding $g(5) = \infty$. Later Fraenkel and Simpson [7] showed that there exists an infinite binary word that has only three squares `00`, `11`, and `0101`, and thus $g(3) = \infty$. A somewhat simplified proof of this result was given by Rampersad, Shallit and Wang [13], using two uniform morphisms. Later, in 2006, Harju and Nowotka [8] gave an even simpler proof of this result. We give here a new proof that the maximal length is infinite if 3 squares are allowed to appear in a binary word. Our proof is simpler than the original proof in [7] and uses a morphism simpler than the proofs of [13] and [8].

The exponent of a word is the quotient of its length over its smallest period. For example `alfalfa` has period 3 and exponent $7/3$. A string with exponent e is also called an e-power. The notion of maximal exponent is central in questions related to the avoidability of patterns in infinite words. An infinite word is said to avoid e-powers (resp. e^+-powers) if the exponents of its finite factors are smaller than e (resp. no greater than e).

The constraints on repetitions in infinite words have been raised to optimality after Dejean's conjecture [4] on the repetitive threshold associated with the alphabet size. *The repetitive threshold* (Dejean's repetitive threshold) of order k is the infimum of maximal exponents of factors of all (infinite) words over a k-letter alphabet.

Looking at the maximal exponent of factors in words containing a bounded number of r_k-powers introduces a new type of threshold (by this Badkobeh and Crochemore [2]). This constraint is called the *finite-repetition threshold*. For the alphabet of k letters, $\mathrm{FRt}(k)$ is defined as the smallest rational number for which there exists an infinite word avoiding $\mathrm{FRt}(k)^+$-powers and containing a finite number of r_k-powers, where r_k is Dejean's repetitive threshold. Associated with the *finite-repetition threshold* is the smallest number of r_k-powers (limit repetitions), $Rn(k)$, that an infinite Dejean's word can accommodate.

E-mail address: golnaz.badkobeh@kcl.ac.uk.

1570-8667/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jda.2013.03.001

Table 1
The gaps between consecutive occurrences of $\mathbf{z} = 000$ are 1101, 11010011, 1110100111, and 11100110100111.

$g_1(ac) = 01001110001101\ 000111$	4
$g_1(abc) = 01001110001101\ 0011\ 000111$	8
$g_1(ca) = 000111\ 01001110001101$	10
$g_1(cba) = 000111\ 0011\ 01001110001101$	14

The results by Karhumäki and Shallit appearing in [9] can then be restated as $\mathrm{FRt}(2) = 7/3$. Deepening this result, in [2] a new proof for $\mathrm{FRt}(2) = 7/3$ is demonstrated and shows that the associated number of squares is 12 ($Rn(2) = 12$). This idea was extended to ternary words in [1] where it is shown that $\mathrm{FRt}(3) = r_3 = 7/4$ and the minimum number of associated r_k-powers is 2.

Moreover, in [3] Badkobeh et al. show that there exists an infinite word on 4 letters containing only 2 7/5-powers and no factor of exponent more than 7/5, proving $\mathrm{FRt}(4) = 7/5$ and $Rn(4) = 2$.

In this note, we provide two additional results of the ... type for alphabets of size 3 and 4. We show that there exists an infinite overlap-free ternary word containing only one square and no e-power where $r_3 \leqslant e \leqslant 2$ ($r_3 = 7/4$).

On a 4-letter alphabet we show that there exists an infinite $3/2^+$-free 4-ary word containing only one 3/2-power and no e-power where $r_4 \leqslant e \leqslant 3/2$ ($r_4 = 7/5$).

2. Squares in binary words

A word is a sequence of letters drawn from a finite alphabet. We consider the ternary alphabet $A = \{a, b, c\}$ and the binary alphabet $B = \{0, 1\}$. A square is a word of the form uu where u is a nonempty (finite) word. This section is dedicated to a new proof of the Fraenkel and Simpson result [7] stated as the following:

Theorem 1. *(See [7].) There exists an infinite binary word containing only three squares.*

Our proof relies on two morphisms f and g_1 defined as follows. The morphism f is defined from A to itself by

$$f(a) = abc,$$
$$f(b) = ac,$$
$$f(c) = b.$$

Since the letter a is a prefix of $f(a)$, the infinite word $\mathbf{f} = f^\infty(a)$ is well defined. It is known that this word is square-free (see [10, Chapter 2]). It can additionally be checked that all square-free words of length 3 occur in $\mathbf{f}$ except aba and cbc.

The morphism g_1 is from A to B and defined by

$$g_1(a) = 01001110001101,$$
$$g_1(b) = 0011,$$
$$g_1(c) = 000111.$$

Proposition 1. *The infinite word $\mathbf{g}_1 = g_1(\mathbf{f})$ contains the 3 squares 00, 11, and 1010 only. The cubes 000 and 111 are the only factors of exponent larger than 2 occurring in $\mathbf{g}_1$.*

The codewords of the morphism g_1 form a prefix code, which implies that the morphism itself is an injective function. Therefore the word $\mathbf{g}_1$ can be parsed uniquely to recover the square-free word $\mathbf{f}$. Proof of Proposition 1 relies on parsing $\mathbf{g}_1$ by locating the occurrences of the triplet $\mathbf{z} = 000$ in $\mathbf{g}_1$. Indeed, any occurrence of $\mathbf{z}$ preceded by 111 determines an occurrence of a in $\mathbf{f}$, and otherwise it is followed by 111 and determines an occurrence of c.

Gaps between occurrences of specific factors. For the purpose of the proof we define the gap function *gap* related to $\mathbf{g}_1$ as follows. For any factors u and v of $\mathbf{g}_1$:

$$gap(u, v) = \{|w| \mid uwv \text{ factor of } \mathbf{g}_1 \text{ and only one occurrence of } v \text{ in } wv\}.$$

Although *gap* is only used in the proof in a very restricted way, note that the gap between any two factors of $\mathbf{g}_1$ is well defined. Table 1 shows that $gap(\mathbf{z}, \mathbf{z}) = \{4, 8, 10, 14\}$.

Proof. We assume that w^2 occurs in $\mathbf{g}_1$ for some non-empty word w and distinguish three cases: where w^2 contains at most one occurrence of $\mathbf{z} = 000$, an even number of occurrences of it, or an odd number.