European Journal of Combinatorics 31 (2010) 177–192

Contents lists available at ScienceDirect

European Journal of Combinatorics

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

Everywhere α-repetitive sequences and Sturmian words

Kalle Saari

Department of Mathematics, University of Turku, 20014 Turku, Finland
Turku Centre for Computer Science, University of Turku, 20014 Turku, Finland

ARTICLE INFO

Article history:
Received 2 October 2008
Accepted 28 January 2009
Available online 10 March 2009

ABSTRACT

Local constraints on an infinite sequence that imply global regularity are of general interest in combinatorics on words. We consider this topic by studying *everywhere α-repetitive sequences*. Such a sequence is defined by the property that there exists an integer $N \geq 2$ such that every length-N factor has a repetition of order α as a prefix. If each repetition is of order strictly larger than α, then the sequence is called *everywhere α^+-repetitive*. In both cases, the number of distinct minimal α-repetitions (or α^+-repetitions) occurring in the sequence is finite.

A natural question regarding global regularity is to determine the least number, denoted by $M(\alpha)$, of distinct minimal α-repetitions such that an α-repetitive sequence is not necessarily ultimately periodic. We call the everywhere α-repetitive sequences witnessing this property *optimal*. In this paper, we study optimal 2-repetitive sequences and optimal 2^+-repetitive sequences, and show that Sturmian words belong to both classes. We also give a characterization of 2-repetitive sequences and solve the values of $M(\alpha)$ for $1 \leq \alpha \leq 15/7$.

1. Introduction

Let us start with the following observation: Each position in the Fibonacci word

$$\mathbf{f} = 0100101001001010010100100101001001010010100100101001010\cdots$$

starts a square. More precisely, each position starts a square of period at most 5. The squares are

$$00, \quad 0101, \quad 010010, \quad 1010, \quad 100100, \quad 1001010010. \tag{1}$$

E-mail addresses: kasaar@utu.fi, kasaar2@gmail.com.

0195-6698/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejc.2009.01.004

Notice that there are 6 of these squares. Furthermore, each square is minimal in the sense that none of their proper prefixes is a square. In fact, all Sturmian words contain squares starting at each position, and the number of minimal squares always equals 6, which we will show in this paper. But what happens if a sequence with squares in every position has at most five minimal squares? We will show that the sequence is then ultimately periodic. Therefore Sturmian words are, in a way, *optimal* squareful sequences (see the exact definition in Section 2).

Relations between local regularity and global regularity in infinite sequences have been considered from various perspectives, see the survey [1, Chap. 8]. A fundamental result conjectured by J. Shallit and proved by Mignosi, Restivo, and Salemi [2] characterizes ultimately periodic sequences by considering left repetitions at each position: A sequence is ultimately periodic if and only if each sufficiently long prefix of the sequence contains a repetition of order $\phi+1$ as a suffix, where ϕ denotes the golden ratio $(1+\sqrt{5})/2$. Karhumäki, Lepistö, and Plandowski [3] considered the same situation, but with an additional condition imposed on the length of the repetitions in question. They call a sequence (ρ, l)-*repetitive*, where $\rho > 1$ is real and $l \geq 1$ an integer, if all sufficiently long prefixes of the sequence have a suffix of the form v^{σ} with $|v| \leq l$ and $\sigma \geq \rho$. They showed that a $(2, 4)$-repetitive sequence is ultimately periodic, while there exist aperiodic $(2, 5)$-repetitive sequences. Furthermore, a $(\rho, 5)$-repetitive sequence is ultimately periodic for all $\rho > 2$. For a thorough discussion of (ρ, l)-repetitive sequences, see [4].

In this work, we investigate local and global regularity not by restricting the length of the shortest repetitions, but instead by restricting the number of distinct such repetitions. Also, as with the example in the beginning of this section, we consider *right repetitions*, i.e., repetitions that extend to the right from its starting position, instead of *left repetitions*, i.e, repetitions that occur as a suffix of a prefix of the sequence. Unlike with left repetitions, the mere existence of right repetitions of large order occurring at each position does not guarantee ultimate periodicity. Indeed, one can easily construct a sequence with, say, cubes starting from every position that is not ultimately periodic. However, if the number of distinct minimal repetitions is bounded, global regularity emerges. For example, if a sequence has squares occurring at every position and there are only five different minimal squares, then the sequence is ultimately periodic, see Theorem 11. This is the observation that gave rise to the notion of an *everywhere α-repetitive* sequence, a sequence whose every position starts an α-repetition and the number of distinct minimal α-repetitions occurring in the sequence is finite. A little modification of the fundamental result by Mignosi et al. [2] reveals that any everywhere α-repetitive sequence with $\alpha \geq \phi + 1$ is ultimately periodic, see Theorem 1. For the sake of brevity, we will sometimes write "α-repetitive" instead of "everywhere α-repetitive".

The notion of an everywhere α-repetitive sequence can be approached from at least two perspectives. One perspective is to fix an infinite sequence, and ask whether it is α-repetitive. It is trivial that any sequence is α-repetitive for $\alpha = 1$. Therefore we may consider the supremum of all real numbers $\alpha \geq 1$ such that the sequence is everywhere α-repetitive. If the sequence is aperiodic, then the supremum of such numbers is at most $\phi + 1$. Also, if α is strictly less than the supremum, the set of minimal α-repetitions that are factors of the sequence is finite. Then it is natural to ask how many distinct minimal α-repetitions occur in the sequence.

The other perspective is to fix a real number $\alpha \geq 1$, and consider the sequences that are everywhere α-repetitive. In particular, what is the smallest number of distinct minimal α-repetitions that an *aperiodic* everywhere α-repetitive sequence can have? Denote this number by $M(\alpha)$. For instance, by what was said in the beginning of this section, we have $M(2) = 6$. Another question then arises: What is the structure of aperiodic everywhere α-repetitive sequences with precisely $M(\alpha)$ minimal α-repetitions? We call such sequences *optimal α-repetitive*.

In this paper, we study these questions as follows. In Section 2, we present some auxiliary definitions and results. In Section 3, we show that the Thue–Morse word is everywhere $5/3$-repetitive, and that the bound $5/3$ is optimal. In Section 4, we modify a result by Mignosi et al. [2] and show that the Fibonacci word is α-repetitive for all $\alpha < \phi + 1$. In Section 5, we first show that $M(2) = 6$. Then we characterize the structure of optimal 2-repetitive sequences, and show that these sequences are optimal also in an abelian sense. Finally, we show that Sturmian words are optimal 2-repetitive sequences. In Section 6, we show that $M(\alpha) > 12$ for $\alpha > 2$ and that Sturmian words are optimal

2^+-repetitive sequences. In Section 7, we determine the values $M(\alpha)$ for $1 \le \alpha \le 15/7$. Finally, in Section 8, we present several open problems.

An earlier version of this paper was presented as an extended abstract [5] in the Second International Computer Science Symposium in Russia 2007 (CSR'07).

2. Preliminaries

Here we present the notation, necessary definitions, and some auxiliary results. For any undefined notions, the reader should consult either [1] or [6].

Let A denote an *alphabet*, that is, a finite set of *letters*, or *symbols*. The set of all *words* obtained by concatenation of letters of A is denoted by A^*. The *length* of a word w is the number of letters it is composed of, and it is denoted by $|w|$. The number of times a letter $b \in A$ occurs in w is denoted by $|w|_b$.

Let $w = a_1a_2 \cdots a_n$ be a word with $n \ge 1$ and $a_i \in A$. If there exists an integer $p \ge 1$ such that $a_i = a_{i+p}$ for all $i = 1, \ldots, n-p$, then p is called a *period* of w. Let $\alpha \ge 1$ be a real number. The word w is called an *α-repetition* if

$$\frac{|w|}{p} \ge \alpha. \tag{2}$$

We also say that w is a repetition of *order* α. If the inequality in (2) is strict, w is an *α^+-repetition*.

If n is an even integer and $p = n/2$, then w is of the form $w = uu$, where $u = a_1a_2 \cdots a_{n/2}$, and w is called a *square*. If we can write $w = uv$ with $|u|_b = |v|_b$ for every letter $b \in A$, then w is called an *abelian square*. An *overlap* is a word of the form $axaxa$, where a is a letter and x is a possibly empty word.

The word w is a *minimal* α-repetition (resp. *minimal* α^+-repetition) if w itself is an α-repetition (resp. α^+-repetition) and no proper prefix of w is an α-repetition (resp. α^+-repetition). Observe that minimal 2^+-repetitions are overlaps with the property that no proper prefix is an overlap.

A sequence, or *infinite word*, is a mapping $\mathbf{z}\colon \mathbb{Z}^+ \to A$, where $\mathbb{Z}^+$ is the set of positive integers. It is customary to denote $\mathbf{z} = a_1a_2a_3\cdots$, where $a_n = \mathbf{z}(n)$. The sequence $\mathbf{z}$ is *ultimately periodic* if $\mathbf{z} = uvvv\cdots$ for some finite words u and v. Then we call the length of v a *period* of $\mathbf{z}$. A sequence that is not ultimately periodic is called *aperiodic*. We denote by A^ω the set of infinite words over the alphabet A.

Let w be a finite or infinite word. If we have $w = uxv$, where u and x are finite words, then x is called a *factor* of w. Furthermore, if u is empty, then x is termed a *prefix* of w.

We say that $\mathbf{z}$ is an *everywhere α-repetitive* sequence if the following condition holds: there exists an integer $N \ge 1$ such that any factor of $\mathbf{z}$ of length N has a prefix that is an α-repetition. Equivalently, $\mathbf{z}$ is everywhere α-repetitive if and only if each position in $\mathbf{z}$ begins an α-repetition, and the set of minimal α-repetitions that occur in $\mathbf{z}$ is finite.

The sequence $\mathbf{z}$ is called *everywhere α^+-repetitive* if it is everywhere $(\alpha+\varepsilon)$-repetitive for some $\varepsilon > 0$. Two special cases deserve more succinct names: We call everywhere 2-repetitive sequences *squareful*, and everywhere 2^+-repetitive sequences *overlapful*. Finally, a sequence $\mathbf{z}$ is called *abelian squareful* if every position in $\mathbf{z}$ begins an abelian square, and only finitely many minimal abelian squares occur in $\mathbf{z}$.

Having established the basic terminology, we are ready to present a fundamental theorem by Mignosi et al. [2] mentioned in the introduction. The original formulation of the theorem in question is rephrased in Theorem 1. It will suffice to give only a sketch of the proof; a complete version can be found in [7].

Theorem 1 (*Mignosi, Restivo, Salemi*). *If $\mathbf{z}$ is an everywhere α-repetitive sequence with $\alpha \ge \phi + 1$ then it is ultimately periodic.*

Proof (*Sketch*). The key property is the following. Let $i \ge 0$ denote a position in $\mathbf{z}$, and let $r_\alpha(\mathbf{z}, i)$ be the least period of the minimal α-repetition at position i. Then the least period of any α-repetition starting at position $i + r_\alpha(\mathbf{z}, i)$ is at least $r_\alpha(\mathbf{z}, i)$; in particular this holds for the minimal α-repetition at that position (cf., [1, Lemma 8.2.12]). Since $\mathbf{z}$ has only finitely many distinct minimal α-repetitions, it follows that $\mathbf{z}$ is ultimately periodic. □

A mapping $h: A^* \to A^*$ is called a *morphism* if $h(uv) = h(u)h(v)$ for all words $u, v \in A^*$. We recall two famous infinite fixed points of morphisms. The *Fibonacci word*, denoted by $\mathbf{f}$, is the fixed point of the morphism $\varphi: 0 \mapsto 01, 1 \mapsto 0$. The *Thue–Morse word*, denoted by $\mathbf{t}$, is the infinite fixed point of the morphism $\mu: 0 \mapsto 01, 1 \mapsto 10$ that begins with the letter 0. Hence,

$$\mathbf{f} = \lim_{n\to\infty} \varphi^n(0) = 0100101001001010010100100101001001010100\cdots;$$

$$\mathbf{t} = \lim_{n\to\infty} \mu^n(0) = 0110100110010110100101100110100110010 1\cdots.$$

Here, an important characteristic of the Fibonacci word is that it is everywhere α-repetitive for all $\alpha < \phi + 1$, where $\phi = (1 + \sqrt{5})/2$, see Theorem 8. According to Theorem 1, the quantity $\phi + 1$ here is optimal.

For a real number $1 \le \alpha < \phi + 1$, we denote by $M(\alpha)$ the smallest positive integer k such that there exists an aperiodic everywhere α-repetitive sequence with k distinct minimal α-repetitions, and any everywhere α-repetitive sequence with less than k minimal α-repetitions is ultimately periodic. The Fibonacci word shows that $M(\alpha)$ exists. An aperiodic α-repetitive sequence with $M(\alpha)$ distinct minimal α-repetitions is called *optimal*.

The following lemma is adapted from [3].

Lemma 2. *If α is a real number with $1 \le \alpha < \phi + 1$, then there exists an optimal α-repetitive sequence over the alphabet* $\{0, 1\}$.

Proof. Let $\mathbf{z}$ be an aperiodic everywhere α-repetitive sequence over an alphabet A. Without loss of generality, assume that $A \cap \{0, 1\} = \emptyset$. For each letter $a \in A$, define a morphism $\tau_a: A^* \to \{0, 1\}^*$ by

$$\tau_a(b) = \begin{cases} 1 & \text{if } b = a; \\ 0 & \text{if } b \in A \setminus \{a\}. \end{cases}$$

Now, the sequence $\tau_a(\mathbf{z})$ is over the alphabet $\{0, 1\}$, and since τ_a is a letter-to-letter morphism, it is everywhere α-repetitive. Furthermore, since τ_a maps α-repetitions to α-repetitions, the number of minimal α-repetitions in $\tau_a(\mathbf{z})$ is at most the number of minimal α-repetitions in $\mathbf{z}$.

The claim follows from the observation that there must exist a letter $a \in A$ such that $\tau_a(\mathbf{z})$ is aperiodic. Indeed, if $\tau_a(\mathbf{z})$ is ultimately periodic for all $a \in A$, then the least common multiple of their periods is a period of $\mathbf{z}$, contradicting the assumption that $\mathbf{z}$ is aperiodic. □

Let us introduce one more concept concerning α-repetitive words. The supremum of all real numbers $\alpha \ge 1$ for which the sequence $\mathbf{z}$ is α-repetitive is denoted by $P(\mathbf{z})$. Theorem 1 implies that $1 \le P(\mathbf{z}) \le \phi + 1$ for all aperiodic sequences $\mathbf{z}$. Furthermore, since the Fibonacci word $\mathbf{f}$ is α-repetitive for all $\alpha < \phi + 1$, we have $P(\mathbf{f}) = \phi + 1$.

A *Sturmian word*, see e.g. [1, Chap. 2], is an infinite word $\mathbf{z}$ such that, for any integer $n \ge 0$, the word $\mathbf{z}$ has exactly $n + 1$ factors of length n. By the definition, Sturmian words are over a two-letter alphabet; we fix that alphabet to be $\{0, 1\}$. The Fibonacci word is an archetype of Sturmian words. These words have the following characteristic, the so-called *balance* property: An aperiodic infinite word $\mathbf{z}$ over the alphabet $\{0, 1\}$ is *Sturmian* if and only if, for all factors x and y of $\mathbf{z}$ of the same length, we have

$$||x|_0 - |y|_0| \le 1.$$

The frequencies of both letters in $\mathbf{z}$ exist, and the frequency of 1 is referred to as the *slope* of $\mathbf{z}$.

Let $(d_n)_{n\ge 1}$ be a sequence of integers with $d_1 \ge 0$ and $d_n \ge 1$ for $n \ge 2$. We define words s_n by

$$s_{-1} = 1, \qquad s_0 = 0, \qquad s_n = s_{n-1}^{d_n} s_{n-2} \quad (n \ge 1). \tag{3}$$

Words obtained in this way are called *standard*. It is clear that s_n is a prefix of s_{n+1} for every $n \ge 1$, and therefore there exists a unique infinite word $\mathbf{c}$ such that s_n is a prefix of $\mathbf{c}$ for every $n \ge 1$. The word $\mathbf{c}$

is called a *characteristic word*; it is a very special kind of Sturmian word, see for example [8, Chap. 9]. For any Sturmian word, there exists a unique characteristic word with the same set of factors. The sequence $(d_n)_{n\geq 1}$ is called the *directive* sequence of $\mathbf{c}$. If α is the slope of $\mathbf{c}$, and its continued fraction expansion is

$$\alpha = [\,0, a_1, a_2, a_3, \ldots\,],$$

then we have $d_1 = a_1 - 1$ and $d_n = a_n$ for all $n \geq 2$.

3. The Thue–Morse word is 5/3-repetitive

To familiarize and motivate the notion of an α-repetitive sequence, we show here that the Thue–Morse word is everywhere 5/3-repetitive, and that the quantity 5/3 is optimal.

Theorem 3. *The Thue–Morse word* $\mathbf{t}$ *is everywhere* 5/3*-repetitive. Its minimal* 5/3*-repetitions are listed below.*

$$\begin{matrix} 00 & 01001 & 0101 & 0110010110 & 0110011 & 01101 \\ 11 & 10110 & 1010 & 1001101001 & 1001100 & 10010. \end{matrix} \qquad (4)$$

Proof. It is readily verified that each of the words in (4) is a repetition of order 5/3.

Conversely, let $R'_{\mathbf{t}}(n)$ denote the function that gives the length of the shortest prefix of $\mathbf{t}$ that contains an occurrence of each factor of length n of $\mathbf{t}$. This function is well-known, see [8, Section 10.10]. If $n \geq 3$, we have

$$R'_{\mathbf{t}}(n) = 3 \cdot 2^{k+1} + n - 1 \qquad (2^k + 2 \leq n \leq 2^{k+1} + 1).$$

Therefore it suffices to check that all factors of length 10 of the prefix of $\mathbf{t}$ of length $R'_{\mathbf{t}}(10) = 57$ has a prefix in (4). □

The next theorem shows that the bound 5/3 in the previous theorem is optimal.

Theorem 4. *The Thue–Morse word is not everywhere* $(5/3)^+$*-repetitive. Therefore,*

$$P(\mathbf{t}) = \frac{5}{3}.$$

Proof. It suffices to show that the Thue–Morse word $\mathbf{t}$ does not have a $(5/3)^+$-repetition as a prefix. Suppose the contrary. Let uu' be the shortest prefix of $\mathbf{t}$ such that u' is a prefix of u and

$$\frac{|uu'|}{|u|} > \frac{5}{3}.$$

Since the word 011 is a prefix of $\mathbf{t}$, we have $|u| \geq 3$. Since uu' is the shortest $(5/3)^+$-repetition that is a prefix of $\mathbf{t}$, it follows that u can be written in the form $u = \mu(x)1$, and u' can be written in the form $u' = 0\mu(y)$, where x and y are finite words. Indeed, otherwise we have $uu' = \mu(w)$ or $uu' = \mu(w)a$ for some word w and a letter a, and then w would be a shorter $(5/3)^+$-repetition that is a prefix of $\mathbf{t}$.

The word u' is a prefix of u, so we either have that 011 is a prefix of u', or u' is a prefix of 01. In the first case, the word 11 is a prefix of $\mu(y)$, which is impossible for any word y. In the second case, we have

$$\frac{|uu'|}{|u|} \leq \frac{|u| + 2}{|u|} \leq \frac{5}{3}.$$

This contradiction completes the proof. □

4. The Fibonacci word is $(\phi + 1 - \varepsilon)$-repetitive

Mignosi et al. [2] showed that every sufficiently long prefix of **f** has a suffix that is an α-repetition for all $\alpha < \phi + 1$. In this section, we prove a very similar result, which is that **f** is everywhere α-repetitive for all $\alpha < \phi + 1$.

Let us begin by denoting $f_n = \varphi^n(0)$ and $F_n = |\varphi^n(0)|$ for $n \geq 0$. Then it is easy to see that $f_n = f_{n-1}f_{n-2}$ for $n \geq 2$, and consequently, the numbers F_n are the Fibonacci numbers.

Since **f** has a prefix 0100, it follows that the square of $f_n = \varphi^n(0)$ occurs in **f** for all $n \geq 0$. As a Sturmian word, **f** has precisely $F_n + 1$ factors of length F_n. These two observations and the fact that f_n is primitive imply that there exists a unique factor of length F_n that is not a conjugate of f_n. We call this factor a *singular* factor, and denote it by c_n. These factors were studied by Wen and Wen [9]. The first singular factors are

$$c_0 = 1, \qquad c_1 = 00, \qquad c_2 = 101, \qquad c_3 = 00100.$$

Lemma 5 (*Wen, Wen*). *For $n \geq 3$, we have*

$$c_n = c_{n-2}c_{n-3}c_{n-2},$$

and for $n \geq 2$, we have

$$c_n = \begin{cases} c_{n-2}c_{n-3}c_{n-4}\cdots c_0 01, & \text{if } n \text{ is even;} \\ c_{n-2}c_{n-3}c_{n-4}\cdots c_0 00, & \text{if } n \text{ is odd.} \end{cases}$$

In the next lemma, we denote by c'_n the word obtained from c_n by erasing the last letter.

Lemma 6. *For $n \geq 7$, the singular word c_n has a prefix that is a repetition of order*

$$2 + \frac{F_{n-4} - 1}{F_{n-3}}.$$

Proof. By Lemma 5, we get

$$\begin{aligned} c_n &= c_{n-2}c_{n-3}c_{n-2} \\ &= (c_{n-4}c_{n-5}c_{n-4})(c_{n-5}c_{n-6}c_{n-5})c_{n-2} \\ &= (c_{n-4}c_{n-5})^2 c_{n-6}c_{n-5}c_{n-2} \\ &= (c_{n-4}c_{n-5})^2 c_{n-6}(c_{n-7}c_{n-8}c_{n-9}\cdots c_0 0a)c_{n-2} \\ &= (c_{n-4}c_{n-5})^2 c'_{n-4} a c_{n-2}, \end{aligned}$$

where $a \in \{0, 1\}$. Since $|c_n| = F_n$, the claim follows. □

In what follows, we denote by f''_n the word obtained from f_n by deleting its last two letters. If $n \geq 2$, then $f_n = f''_n ab$, where $ab \in \{01, 10\}$. A well-known near-commutative property of the finite Fibonacci words states that

$$f_n f''_{n-1} = f_{n-1} f''_n$$

for $n \geq 2$, see e.g. [10].

The next lemma is the crucial property of the Fibonacci word **f** that will explain why **f** is α-repetitive for all $\alpha < \phi + 1$.

Lemma 7. *Let n and k be integers with $n \geq 5$ and $2 \leq k \leq n - 3$. For all $m = k, k + 1, \ldots, n - 2$, the position*

$$F_{k+1} + \sum_{i=k}^{m-1} F_i$$

in the word $f_n f_n$ has an occurrence of $f_m^3 f''_{m-1}$.

Proof. First, it is an easy proof by induction to show that

$$f_n = f_{k+1} f_k f_{k+1} f_{k+2} \cdots f_{n-3} f_{n-2},$$

and so

$$f_n f_n = f_{k+1} f_k f_{k+1} f_{k+2} \cdots f_{n-3} f_{n-2} f_{n-1} f_{n-2}.$$

Since $m \geq 2$, we have

$$f_m f_{m+1} f_{m+2} = f_m f_m f_{m-1} f_m f_{m-1} f_m = f_m f_m f_m f''_{m-1} ab f_{m-1} f_m,$$

where $ab \in \{01, 10\}$. This proves the claim for $m = k, k+1, \ldots, n-3$. To prove the claim for $m = n-2$, we note that

$$f_{n-2} f_{n-1} f_{n-2} = f_{n-2} f_{n-2} f_{n-3} f_{n-2} = f_{n-2}^3 f''_{n-3} ab,$$

where $ab \in \{01, 10\}$. □

We are ready for the main theorem of this section.

Theorem 8. *The Fibonacci word is everywhere α-repetitive for all $\alpha < \phi + 1$.*

Proof. First we have to fix three parameters. Since $F_{n+1}/F_n \to \phi$ when $n \to \infty$, there exists an integer $k \geq 2$ such that

$$1 + \frac{F_{n+1} - 2}{F_n} > \alpha \tag{5}$$

for all $n \geq k$. Next, choose an integer $n \geq k+3$ such that

$$1 + \frac{F_{n+1} - 2 - F_{k+1}}{F_n} > \alpha.$$

Finally, let $N = n + 2$.

Now we will show that all factors of length F_N of the Fibonacci word have a prefix that is a repetition of order α. If w is a factor of length F_N, then it is either a conjugate of f_N or the singular factor c_N.

First, suppose that w is a conjugate of f_N. Then w occurs in the word $f_N f_N$ in a position i with $0 \leq i < F_N$. We have two cases to consider:

Case $0 \leq i < F_{k+1}$. Since $N = n + 2$ and $n \geq 5$, we have

$$f_N = f_{n+1} f_n = f_n f_{n-1} f_n = f_n f_n f''_{n-1} ab$$

for some $ab \in \{0, 1\}$. Consequently, w has a prefix that is a repetition of order

$$\frac{2F_n + F_{n-1} - 2 - F_{k+1}}{F_n} = 1 + \frac{F_{n+1} - 2 - F_{k+1}}{F_n} > \alpha.$$

Case $F_{k+1} \leq i < F_N$. Then Lemma 7 implies that w has a prefix that is a repetition of order

$$\frac{|f_m^2 f''_{m-1}|}{|f_m|} = 1 + \frac{F_{m+1} - 2}{F_m} > \alpha$$

for some $m = k, k+1, \ldots, N-2$.

Second, suppose that w is the singular word c_N. Then Lemma 6 and inequality (5) imply that the word c_N has a prefix that is a repetition of order

$$1 + \frac{F_{N-2} - 1}{F_{N-3}} > \alpha.$$

The proof is complete. □

5. Squareful sequences

In this section, we first consider some general properties of aperiodic squareful sequences. Then we characterize the optimal squareful sequences, and show that they are optimal also in the abelian sense. Finally, we show that Sturmian words are optimal squareful sequences.

5.1. The number of minimal squares

In the next two lemmas we assume that $\mathbf{z}$ is an aperiodic squareful sequence.

Lemma 9. *If the word uu is a minimal square in $\mathbf{z}$, then there exists a minimal square vv, different from uu, in $\mathbf{z}$ such that u is a prefix of vv.*

Proof. Consider an occurrence of the word uu in $\mathbf{z}$. The latter u in uu starts a minimal square in $\mathbf{z}$, say xx. The minimality of uu implies that xx is not a prefix of u. Hence either $x = u$ or u is a prefix of xx. The sequence $\mathbf{z}$ is aperiodic, so for some occurrence of uu, the second case must hold. □

Lemma 10. *For each letter b occurring in $\mathbf{z}$, there exist at least three distinct minimal squares in $\mathbf{z}$ that start with the letter b.*

Proof. Let uu denote a shortest minimal square that starts with the letter b. Lemma 9 implies that there exists another minimal square vv such that u, and thus the letter b, is a prefix of vv. The choice of uu implies that u is actually a prefix of v, and hence we can write $v = ut$ for some nonempty word t.

To derive a contradiction, suppose that the two minimal squares uu and vv are the only ones starting with the letter b. Then there exists a position in $\mathbf{z}$ that has an occurrence of both words v and uu. Since vv is a minimal square, it follows that v is a prefix of uu, and further that t is a proper prefix of u. Hence we can write $v = ts$ for some nonempty word s. But now $vv = utts$, and we see that the word tt is a square in $\mathbf{z}$. Since t is a prefix of u, it starts with the letter b. Therefore either the square tt or one of its prefixes is a minimal square starting with the letter b, and it is strictly shorter than uu, a contradiction. □

Theorem 11. *Any aperiodic squareful sequence has at least six minimal squares, and six is the optimal lower bound. Therefore $M(2) = 6$.*

Proof. If a squareful sequence is aperiodic, it has at least two distinct letters, so Lemma 10 implies that it must have at least six distinct minimal squares. That the quantity six is optimal follows from the fact that the Fibonacci word is squareful, as we already mentioned in the introduction; its minimal squares are listed in Eq. (1). □

Note that, by the proof of the previous theorem, any optimal squareful sequence is over a two-letter alphabet.

5.2. Characterization of optimal squareful sequences

Here we characterize the optimal squareful sequences by constructing their minimal squares.

In what follows, the symbol $\mathbf{z}$ denotes an optimal squareful sequence over the alphabet $\{0, 1\}$. Since all optimal squareful sequences are over a two-letter alphabet, this does not restrict the generality. Since $\mathbf{z}$ is aperiodic, it follows that one of the letters, say 0, occurs in blocks of at least two distinct lengths. That is, there exist integers i and j with $j > i > 0$ such that the words 10^i1 and 10^j1 are factors of $\mathbf{z}$. Suppose further that i and j are the least integers with these properties.

We start with a series of lemmas. We will often use Lemma 10 implying that both letters 0 and 1 correspond to exactly three distinct minimal squares in $\mathbf{z}$.

Lemma 12. *The word* 11 *does not occur in* $\mathbf{z}$.

Proof. There exists at least one minimal square with a prefix 10^i1. Since $j > i$, Lemma 9 implies that there exist two distinct minimal squares with a prefix 10^j. Therefore, the word 11 cannot occur in $\mathbf{z}$ because otherwise 11 would be a fourth minimal square in $\mathbf{z}$ with a prefix 1. $\square$

Lemma 13. *Let $n \geq 1$ be an integer. The word 10^n1 is a factor of $\mathbf{z}$ if and only if n equals either i or j.*

Proof. Suppose that, contrary to what we want to prove, the sequence $\mathbf{z}$ has a factor 10^n1 with $n \geq 1$ distinct from i and j. Then $\mathbf{z}$ has three distinct minimal squares with prefixes 10^i1, 10^j1, and 10^n1. In addition, since $n > j > i$, Lemma 9 implies that there exists another minimal square with prefix 10^n. Hence we have altogether at least four minimal squares starting with the letter 1, a contradiction. $\square$

Lemma 14. *We have $j = i + 1$.*

Proof. The words 00 and $010^{i-1}010^{i-1}$ are minimal squares in $\mathbf{z}$. There is also a third minimal square, denote it by uu, in $\mathbf{z}$ such that 010^{j-1} is a prefix of u. It follows from Lemma 9 that u, and hence also 010^{j-1}, is a prefix of $010^{i-1}010^{i-1}$. Since $j > i$, this is possible only if $j = i + 1$. $\square$

Now, we can write $i + 1$ in place of j.

Lemma 15. *There exists an integer $k \geq 0$ such that the word*

$$10^{i+1}(10^i)^n10^{i+1} \tag{6}$$

is a factor of $\mathbf{z}$ if and only if n equals either k or $k + 1$.

Proof. By the previous three lemmas, the suffix of $\mathbf{z}$ that starts from the first occurrence of the letter 1 in $\mathbf{z}$ can be factorized into words 10^i and 10^{i+1}. The aperiodicity of $\mathbf{z}$ implies, therefore, that the word of the form (6) occurs in $\mathbf{z}$ for at least two distinct values of $n \geq 0$. Let k and m denote two such distinct integers with k the smallest possible such integer. We will show that $m = k + 1$.

Observe first that the minimal squares

$$10^i10^i \qquad \text{and} \qquad 10^{i+1}(10^i)^k10^{i+1}(10^i)^k$$

are factors of $\mathbf{z}$. Since the word

$$10^{i+1}(10^i)^m10^{i+1}$$

is a factor of $\mathbf{z}$, it follows from Lemma 9 that there exist at least two minimal squares with a prefix $10^{i+1}(10^i)^m$. Since there exist only three minimal squares with prefix 1, the word $10^{i+1}(10^i)^m$ must be a prefix of

$$10^{i+1}(10^i)^k10^{i+1}(10^i)^k.$$

Since $m > k$, this is possible only if $m = k + 1$. The proof is complete. $\square$

In the following theorem, the *root* of a square uu refers to the word u.

Theorem 16. *If a squareful sequence is optimal, then there exist integers $i \geq 1$ and $k \geq 0$, and letters 0 and 1 such that the roots of the minimal squares of the sequence are*

$$0, \quad 010^{i-1}, \quad 010^i, \quad 10^i, \quad 10^{i+1}(10^i)^k, \quad \textit{and} \quad 10^{i+1}(10^i)^{k+1}. \tag{7}$$

Proof. Without loss of generality, it suffices to consider $\mathbf{z}$, the optimal squareful sequence we have been considering in this subsection. That the first four words listed in (7) are roots of minimal squares in $\mathbf{z}$ follows from the observation that the two words 010^i10^i and $010^{i+1}10^i$ are factors of $\mathbf{z}$. Since the suffix of $\mathbf{z}$ starting from the first occurrence of the letter 1 in $\mathbf{z}$ can be factorized into words 10^i and 10^{i+1}, it follows from Lemma 15 that the minimal squares

$$10^{i+1}(10^i)^k10^{i+1}(10^i)^k \qquad \text{and} \qquad 10^{i+1}(10^i)^{k+1}10^{i+1}(10^i)^{k+1}$$

both occur in $\mathbf{z}$ (the latter minimal square may be obtained in two ways). Therefore, the last two words listed in (7) are roots of minimal squares in $\mathbf{z}$. Since $\mathbf{z}$ is optimal, there exist precisely six minimal squares in $\mathbf{z}$, and so we have found all of them. The proof is complete. $\square$

Theorem 17. *Suppose that* $\mathbf{z}$ *is an aperiodic sequence. Then* $\mathbf{z}$ *is squareful and optimal if and only if, up to renaming letters, there exist integers* $i \geq 1$ *and* $k \geq 0$ *such that* $\mathbf{z}$ *is an element of the language*

$$0^*(10^i)^* \left(10^{i+1}(10^i)^k + 10^{i+1}(10^i)^{k+1}\right)^\omega . \tag{8}$$

Proof. Again, without loss of generality, it suffices to verify the claim for the optimal squareful sequence $\mathbf{z}$ we have been considering in this subsection. Hence the roots of the minimal squares are the ones listed in (7). Write $\mathbf{z} = u\mathbf{v}$, where u and $\mathbf{v}$ are chosen so that the block 10^{i+1} occurs for the first time in $\mathbf{z}$ as a prefix of $\mathbf{v}$. Then u is in the language $0^*(10^i)^*$. Since the word 10^{i+1} is a prefix of $\mathbf{v}$, Lemma 15 implies that $\mathbf{v}$ factorizes over the words $10^{i+1}(10^i)^k$ and $10^{i+1}(10^i)^{k+1}$. Therefore, $\mathbf{z}$ is in the language (8).

Conversely, if $\mathbf{z}$ is an infinite word in the language (8), it is readily verified that the sequence $\mathbf{z}$ is squareful. Furthermore, $\mathbf{z}$ has exactly six minimal squares—their roots are the ones listed in (7)—and so $\mathbf{z}$ is also optimal. This completes the proof. □

5.3. Optimal abelian squareful sequences

Any squareful sequence is trivially also abelian squareful. But the number of minimal squares can be different than the number of minimal abelian squares.

In this short subsection, we show that optimal minimal squareful sequences have exactly five minimal abelian square factors, and any abelian squareful sequence with at most four different abelian squares is ultimately periodic.

Theorem 18. *An optimal squareful sequence has exactly five minimal abelian square factors.*

Proof. Let $\mathbf{z}$ be an optimal squareful sequence. By Theorem 16, the roots of the minimal squares in $\mathbf{z}$ can be taken to be

$$0, \quad 010^{i-1}, \quad 010^i, \quad 10^i, \quad 10^{i+1}(10^i)^k, \quad \text{and} \quad 10^{i+1}(10^i)^{k+1}, \tag{9}$$

for some $i \geq 1$ and $k \geq 0$. Plainly, 00 and 10^i10^i are minimal abelian squares.

Let $a \geq 0$ be an integer. The square 010^a010^a has prefixes $010^{a+1}1$ and $010^{a+1}10$. The first one is a minimal abelian square if a is even; the second one is a minimal abelian square if a is odd. Therefore, the squares

$$010^{i-1}010^{i-1} \qquad \text{and} \qquad 010^i010^i$$

have minimal abelian squares in a prefix, and they are distinct from each other.

Finally, the squares

$$10^{i+1}(10^i)^k10^{i+1}(10^i)^k \qquad \text{and} \qquad 10^{i+1}(10^i)^{k+1}10^{i+1}(10^i)^{k+1}$$

both have prefixes $10^{i+1}1$ and $10^{i+1}10$. The first one is a minimal abelian square if i is odd; the latter one is a minimal abelian square if i is even. Consequently, $\mathbf{z}$ has precisely five minimal abelian squares. □

Theorem 19. *An aperiodic abelian squareful sequence must have at least five distinct minimal abelian squares.*

Proof. Let $\mathbf{z}$ be an aperiodic abelian squareful sequence. If $\mathbf{z}$ is composed of three or more distinct letters, it clearly must have at least six minimal abelian squares. Therefore we may assume that $\mathbf{z} \in \{0, 1\}^\omega$. Like before, one of the letters, say 0, must occur in blocks of at least two different lengths, and hence there exist positive integers i and j such that 10^i1 and 10^j1 occur in $\mathbf{z}$. This implies that 00 is a minimal abelian square, and there are two distinct minimal abelian squares with a prefix 10^i and 10^j, respectively.

If 11 occurs in $\mathbf{z}$, we have five minimal abelian squares because there is at least one abelian square with prefix 01. Suppose then that 11 does not occur in $\mathbf{z}$. We claim that there exist at least two distinct minimal abelian squares with a prefix 01. To shows this, let uv be a minimal abelian square with a prefix 01 and $|u| = |v|$. Then clearly 01 is a prefix of u. Now, 01 must occur at least twice in uv. Indeed, since 11 does not occur in $\mathbf{z}$, the word 010 must be a prefix of uv, and there is another occurrence of 1 after the prefix 010 in uv. Therefore if there is only one minimal abelian square with prefix 01, then $\mathbf{z}$ is ultimately periodic starting from the first occurrence of uv. □

5.4. *Sturmian words are optimal squareful sequences*

Now we will show that Sturmian words are optimal squareful sequences. Note, however, that this property does not characterize Sturmian words, as is immediately clear by characterization of optimal squareful sequences, Theorem 17.

Theorem 20. *Sturmian words are optimal squareful sequences.*

Proof. Let $\mathbf{z}$ denote a Sturmian word. Without loss of generality, we may suppose that the word 11 is not a factor of $\mathbf{z}$. Let i denote the least integer j such that the word 10^j1 is a factor of $\mathbf{z}$. Since 11 does not occur in $\mathbf{z}$, we have $i \geq 1$. Furthermore, since $\mathbf{z}$ is balanced, it follows that the maximal integer j such that 10^j1 is a factor of $\mathbf{z}$ equals $i + 1$. Therefore $\mathbf{z}$ is in the language

$$0^*(10^i + 10^{i+1})^\omega. \qquad (10)$$

Let k denote the least integer $n \geq 0$ such that the word

$$10^{i+1}(10^i)^n10^{i+1} \qquad (11)$$

is a factor of $\mathbf{z}$. Since $\mathbf{z}$ is balanced, the maximal integer n such that the word in (11) occurs in $\mathbf{z}$ equals $k + 1$. Indeed, otherwise both words

$$0^{i+1}(10^i)^k10^{i+1} \quad \text{and} \quad (10^i)^{k+2}1$$

would be factors of $\mathbf{z}$, have the same length, but differ in the number of letters 1 by 2, which contradicts the balance property of $\mathbf{z}$.

Now the claim follows from (10), from the fact that the word in (11) is a factor of $\mathbf{z}$ precisely when $n = k$ or $n = k + 1$, and from Theorem 17. □

6. Overlapful sequences

In this section, we first show that optimal overlapful sequences have exactly twelve minimal overlaps. Then we show that Sturmian words are among the optimal overlapful sequences.

6.1. *Optimal overlapful sequences*

We start by showing that $M(\alpha) \geq 12$ for $\alpha > 2$. To do so, we first need a few lemmas.

Let $\mathbf{z}$ denote an optimal overlapful sequence. We show that $\mathbf{z}$ has at least twelve minimal overlaps. By Lemma 2, we may suppose that $\mathbf{z} \in \{0, 1\}^\omega$. Since $\mathbf{z}$ is aperiodic, it follows that one of the letters, say 0, occurs in blocks of two distinct lengths. That is, there exist integers i and j with $j > i > 0$ such that the words 10^i1 and 10^j1 are factors of $\mathbf{z}$. Suppose further that i and j are the least integers with these properties. Note that each of the words 00, 01, 10^i1, and 10^j1 is a prefix of a minimal overlap occurring in $\mathbf{z}$.

Lemma 21. *The sequence* $\mathbf{z}$ *has at least three distinct minimal overlaps with a prefix* 01.

Proof. Let $vv0$ denote a longest minimal overlap with a prefix 01. Since the length of v is maximal and $\mathbf{z}$ is aperiodic, it follows that there must be another minimal overlap, say $uu0$, such that u is a proper prefix of v and 01 is a prefix of u. Furthermore, since $vv0$ is a minimal overlap, it follows that v is a proper prefix of uu. Therefore there exist nonempty words x and t such that $v = ux$ and $u = xt$. In addition, the word t starts with the letter 0, and so 01 is a prefix of x.

Now, we have $vv0 = uxxtx0$, and since t begins with 0, we find an overlap $xx0$. Since the word x has a prefix 01, it follows that either $xx0$ or one of its proper prefixes is a third minimal overlap with a prefix 01. □

Lemma 22. *The sequence* **z** *has at least three distinct minimal overlaps with a prefix* 00.

Proof. Let $vv0$ denote a longest minimal overlap with a prefix 001. Since the length of v is maximal, there must be another minimal overlap, say $uu0$, such that u is a prefix of v and 001 is a prefix of u. Furthermore, since $vv0$ is a minimal overlap, the word v is a proper prefix of uu. It follows that there exist nonempty words x and t such that $v = ux$ and $u = xt$. In addition, both x and t begin with the letter 0.

Since the overlap $vv0$ can be written as $uxxtx0$, and t begins with 0, we find an overlap $xx0$. Since x is a prefix of v, we see that either $x = 0$ or x has a prefix 00. In either case, $xx0$ has a prefix 00, and therefore either $xx0$ or one of its proper prefixes is a third minimal overlap. □

Lemma 23. *The sequence* **z** *has at least three distinct minimal overlaps with a prefix* 10^i1. *There are at least three distinct minimal overlaps with a prefix* 10^j1.

Proof. We give the proof only for 10^i1. The proof in the case 10^j1 is identical, up to writing j in place of i.

Let $vv1$ denote a longest minimal overlap with a prefix 10^i1. First we claim that the word 10^i1 is a prefix of v. For if not, we have $v = 10^i$, and since v is a longest word such that $vv1$ has a prefix 10^i1, it follows that **z** has a suffix $(10^i)^\omega$, a contradiction.

Since the length of v is maximal, there must be another minimal overlap, say $uu1$, such that u is a prefix of v and 10^i1 is a prefix of $uu1$. Furthermore, since $vv1$ is a minimal overlap, the word v is a proper prefix of uu. It follows that there exist nonempty words x and t such that $v = ux$, $u = xt$, and the word t starts with the letter 1. Hence the word $vv1$ can be written as $uxxtx1$. Since x starts with the letter 1, we see that $xx1$ is an overlap and a factor of **z**.

Next we claim that $xx1$ has a prefix 10^i1. Since the word 10^i1 is a prefix of $uu1$ and 1 is a prefix of t, the identity $u = xt$ implies that 10^i is a prefix of x. Therefore, either $x = 10^i$ or x has a prefix 10^i1. In both cases the overlap $xx1$ has a prefix 10^i1. So either $xx1$ or one of its proper prefixes is a third minimal overlap with prefix 10^i1 occurring in **z**. The proof is complete. □

The following lemma implies that twelve minimal overlaps are optimal.

Lemma 24. *The Fibonacci word is overlapful with twelve minimal overlaps. In fact, the Fibonacci word is everywhere* 17/8*-repetitive with twelve minimal* 17/8*-repetitions given below.*

0010010	1001001
00101001001010010	10010100100101001
00101001010	10010100101
0100100	10100100101001001
01001010010	10100100101001010010010100 10
01010	10100101001

Proof. The claim can be verified computationally: the Fibonacci word **f** has 29 factors of length 28, and each of these factors has a prefix in the list above. It can be verified that each word in the list above is a minimal 17/8-repetition. □

The following theorem is an immediate consequence of the previous four lemmas of this subsection.

Theorem 25. *Any aperiodic overlapful sequence has at least* 12 *minimal overlaps, and the* 12 *here is the smallest possible.*

6.2. Sturmian words are optimal overlapful sequences

In this subsection, we show that any Sturmian word is overlapful with exactly twelve minimal overlaps.

We start with a few auxiliary results. First we show that any characteristic word is a limit of infinite words obtained from the Fibonacci word $\mathbf{f}$ using the Fibonacci morphism $\varphi\colon 0 \mapsto 01,\ 1 \mapsto 0$ and the morphism $E\colon 0 \mapsto 1,\ 1 \mapsto 0$.

For all integers $m \geq 1$, define the morphisms

$$\theta_m\colon \begin{array}{ll} 0 & \mapsto 0^{m-1}1 \\ 1 & \mapsto 0^{m-1}10. \end{array}$$

Lemma 26. *If* $\mathbf{c}$ *is a characteristic word, it can be written as the limit of a sequence of infinite words*

$$\mathbf{c} = \lim_{n\to\infty} g_1 \circ g_2 \circ \cdots \circ g_n(\mathbf{f}),$$

where $g_i \in \{\varphi, E\}$ *for all* $i \geq 1$.

Proof. Let α denote the slope of $\mathbf{c}$, and denote its continued fraction expansion by

$$\alpha = [\,0, a_1, a_2, a_3, \ldots\,].$$

It is well-known (see the proof of [1, Prop. 2.2.24]) that the characteristic word $\mathbf{c}$ is obtained as the limit

$$\mathbf{c} = \lim_{n\to\infty} \theta_{a_1} \circ \theta_{a_2} \circ \cdots \circ \theta_{a_n}(0).$$

Now, an easy induction proof shows that $\theta_1 = E \circ \varphi \circ E$ and $\theta_m = (\varphi \circ E)^{m-1} \circ \theta_1$, and consequently $\theta_m \in \{\varphi, E\}^*$ for all $m \geq 1$. It follows that

$$\mathbf{c} = \lim_{n\to\infty} g_1 \circ g_2 \circ \cdots \circ g_n(0) \tag{12}$$

for some morphisms $g_i \in \{\varphi, E\}$. Since the Fibonacci word $\mathbf{f}$ starts with the letter 0, Eq. (12) implies that

$$\mathbf{c} = \lim_{n\to\infty} g_1 \circ g_2 \circ \cdots \circ g_n(\mathbf{f}). \quad \square$$

The morphisms E and φ preserve Sturmian words (see [1, Cor. 2.2.19]). The next lemma shows that E and φ also preserve the number of minimal squares and overlaps in Sturmian sequences.

Lemma 27. *If* $\mathbf{x}$ *is an optimal overlapful Sturmian word, then so are the sequences* $E(\mathbf{x})$ *and* $\varphi(\mathbf{x})$.

Proof. Since the morphism E only renames the letters, the claim is obvious for the sequence $E(\mathbf{x})$. Denote $\mathbf{y} = \varphi(\mathbf{x})$; then $\mathbf{y}$ is a Sturmian word. We show that $\mathbf{y}$ is overlapful with at most twelve minimal overlaps. To this end, denote $\mathbf{x} = x_0x_1x_2\cdots$ and $\mathbf{y} = y_0y_1y_2\cdots$, where $x_i, y_j \in \{0, 1\}$.

Let $j \geq 0$ be an integer. If the letter that occurs in $\mathbf{y}$ at position j is 0, there exists an integer $i \geq 0$ such that

$$y_jy_{j+1}\cdots = \varphi(x_ix_{i+1}\cdots).$$

Since $\mathbf{x}$ is Sturmian, it is squareful with six minimal squares by Theorem 20. Hence the sequence $x_ix_{i+1}\cdots$ has a prefix that is a minimal square; denote it by uu. It follows that the sequence $y_jy_{j+1}\cdots$ has a prefix $\varphi(uu)0$, which is an overlap. Consequently, any position in $\mathbf{y}$ that is occupied by the letter 0 has an occurrence of an overlap, which belongs to the set of six overlaps of the form $\varphi(uu)0$, where uu is a minimal square in $\mathbf{x}$.

Let us then assume that the letter of $\mathbf{y}$ at position j is 1. Since $\mathbf{y} = \varphi(\mathbf{x})$, we have $j \geq 1$, and furthermore, there exists an index $i \geq 0$ such that $x_i = 0$ and

$$0y_jy_{j+1}\cdots = \varphi(x_ix_{i+1}\cdots).$$

Since $\mathbf{x}$ is an optimal overlapful sequence, the sequence $x_i x_{i+1} \cdots$ has a prefix that is a minimal overlap; denote it by $0v0v0$. Lemmas 21 and 22 imply that there are exactly six distinct such minimal overlaps. Now, the sequence $0y_j y_{j+1} \cdots$ has a prefix $\varphi(0v0v0)$, whence the sequence $y_j y_{j+1} \cdots$ has a prefix $1\varphi(v0v0)$, which is an overlap. Consequently, any position in $\mathbf{y}$ that is occupied by the letter 1 has an occurrence of an overlap, which belongs to the set of six overlaps of the form $1\varphi(v0v0)$ such that $0v0v0$ is a minimal overlap in $\mathbf{x}$.

We have shown that each position of $\mathbf{y}$ is occupied by an overlap, not necessarily minimal, but there are at most twelve such overlaps. Since $\mathbf{y}$ is aperiodic, it now follows from Theorem 25 that $\mathbf{y} = \varphi(\mathbf{x})$ is optimal overlapful. □

We are ready for the main theorem of this subsection.

Theorem 28. *Sturmian words are optimal overlapful sequences.*

Proof. Let $\mathbf{z}$ be a Sturmian word, and let $\mathbf{c}$ be the characteristic sequence with the slope of $\mathbf{z}$. It suffices to prove the claim for $\mathbf{c}$ because the factors of $\mathbf{c}$ are exactly the factors of $\mathbf{z}$.

According to Lemma 26, there exist morphisms $g_i \in \{E, \varphi\}$ for $i \geq 1$ such that

$$\mathbf{c} = \lim_{n \to \infty} g_1 \circ g_2 \circ \cdots \circ g_n(\mathbf{f}).$$

By Lemma 24, the Fibonacci word is an optimal overlapful sequence. Therefore, so are the infinite words $g_1 \circ g_2 \circ \cdots \circ g_n(\mathbf{f})$ for $n \geq 1$ by Lemma 27. As a limit of optimal overlapful sequences, the word $\mathbf{c}$ is an optimal overlapful sequence. □

7. Values of $M(\alpha)$ for $1 \leq \alpha \leq 15/7$

Recall that $M(\alpha)$ denotes the least integer $k \geq 1$ such that there exists an aperiodic everywhere α-repetitive sequence with k minimal α-repetitions. In this section, we establish the values of $M(\alpha)$ for $1 \leq \alpha \leq 15/7$. To begin with, it is clear that $M(1) = 2$.

Theorem 29. *For $1 < \alpha \leq 3/2$, we have $M(\alpha) = 4$.*

Proof. Let $\mathbf{z}$ be an optimal α-repetitive sequence. According to Lemma 2, we may assume that $\mathbf{z} \in \{0, 1\}^\omega$. Since $\mathbf{z}$ is aperiodic and $\alpha > 1$, it has at least two distinct minimal α-repetitions that start with the letter 0. The same holds for the letter 1, and therefore $M(\alpha) \geq 4$.

Next we show that $M(\alpha) \leq 4$. Since any 3/2-repetitive sequence is also α-repetitive, it now suffices to find a 3/2-repetitive sequence with four minimal 3/2-repetitions. We do not have to look far: The Fibonacci word $\mathbf{f}$ has this property. Indeed, each factor of $\mathbf{f}$ of length 5 has a 3/2-repetition as a prefix, which is readily verified. The minimal 3/2-repetitions are 00, 010, 10010, 101. □

Theorem 30. *For $3/2 < \alpha < 2$, we have $M(\alpha) = 5$.*

Proof. Let $\mathbf{z}$ be an optimal α-repetitive sequence. As before, we may assume that $\mathbf{z} \in \{0, 1\}^\omega$. Since $\mathbf{z}$ is aperiodic, either 00 or 11 must occur in $\mathbf{z}$; say 00. Let w denote a minimal α-repetition in $\mathbf{z}$ with a prefix 01. Since $\alpha > 3/2$, we have $w \neq 010$, and hence w is of the form $w = 01u01v$, where u and v are finite, possibly empty, words. This implies that there exists another minimal α-repetition in $\mathbf{z}$ with a prefix 01. Noticing that 00 is a minimal α-repetition in $\mathbf{z}$, we have shown that $\mathbf{z}$ has at least three minimal α-repetitions with prefix 0. And because $\mathbf{z}$ must have at least two minimal α-repetitions starting with the letter 1, we deduce $M(\alpha) \geq 5$.

Next we show that $M(\alpha) \leq 5$. To this end, we construct an aperiodic α-repetitive sequence with exactly five minimal α-repetitions. Since $\alpha < 2$, there exists an integer $i \geq 1$ such that

$$\alpha \leq \frac{2i+3}{i+2}.$$

Now, let $\mathbf{c}$ be the characteristic sequence with the directive sequence $(d_n)_{n \geq 1}$, where $d_1 = i$ and $d_n = 1$ for $n \geq 2$. Since the sequence $\mathbf{c}$ is Sturmian, it has six minimal squares. The recursive formula (3) gives

$$s_5 = 0^i 1 0^{i+1} 1 0^i 1 0^{i+1} 1 0^{i+1} 1,$$

and from s_5 we can find the six minimal squares that occur in $\mathbf{x}$. They are

$$00, \quad 010^i10^{i-1}, \quad 010^{i+1}10^i, \quad 10^i10^i, \quad 10^{i+1}10^{i+1}, \quad 10^{i+1}10^i10^{i+1}10^i.$$

From this we see that each position of $\mathbf{x}$ has an occurrence of one of the five words

$$00, \quad 010^i10^{i-1}, \quad 010^{i+1}10^i, \quad 10^i10^i, \quad 10^{i+1}10^i.$$

The first four of these words are squares; the last one is a $(2i+3)/(i+2)$-repetition. Consequently, $\mathbf{c}$ is an aperiodic α-repetitive sequence with five minimal α-repetitions, implying that $M(\alpha) \leq 5$. □

For convenience, we repeat Theorem 11 here.

Theorem 31. *We have* $M(2) = 6$.

The last result of this paper is the following.

Theorem 32. *For* $2 < \alpha \leq 15/7$, *we have* $M(\alpha) = 12$.

Proof. By Theorem 25, we have $M(\alpha) \geq 12$. To show that $M(\alpha) \leq 12$, we construct the characteristic sequence $\mathbf{c}$ given by the directive sequence $(2, 1, 1, 1, \ldots)$. Now, we may verify by an exhaustive search that all factors of length 39 of $\mathbf{c}$ have a $15/7$-repetition as a prefix and there are exactly 12 minimal $15/7$-repetitions. □

Thus the values of $M(\alpha)$ for $1 \leq \alpha \leq 15/7$ are established.

8. Open problems and conjectures

To the knowledge of the author, everywhere α-repetitive sequences, or related concepts, have not been much studied. In addition to the works [2–4] already mentioned, the author is aware of only one other related paper by Berthé, Holton, and Zamboni [11], which deals with the so-called initial critical exponents of Sturmian words. Consequently, there is an abundance of intriguing open problems left. In this last section of this paper, we mention a few of them.

The first problem seems to be trivially true, but we have not been able to prove it.

Conjecture 33. *The optimal* α*-repetitive sequences are always over a two-letter alphabet.*

The next problem was posed by J. Karhumäki.

Conjecture 34. *We have* $M(\alpha) \to \infty$ *as* $\alpha \to \phi + 1$.

Some experimental computations we have made suggest the following.

Problem 35. Is it true that

$$M\left(2\frac{1}{3}\right) = 20 \quad \text{and} \quad M\left(2\frac{4}{9}\right) = 29?$$

Calculating the exact values of $M(\alpha)$ when $15/7 < \alpha < \phi + 1$ seems a formidable task. Therefore a better approach may be trying to find good upper bounds for them. The next problem is a step in this direction.

Problem 36. Given a real number α with $2 < \alpha < \phi + 1$, how many distinct minimal α-repetitions occur in the Fibonacci word?

The next problem comes from the observation that $M(\alpha)$ is an increasing step function. It was asked by J. Cassaigne.

Problem 37. Do the steps in the graph of M always occur at rational points?

The value of $P(\mathbf{z})$ may be considered as a variant of the classical critical exponent of a sequence. The following problem is inspired by a result of Krieger and Shallit [12].

Problem 38. What is the range of the values of $P(\mathbf{z})$ when $\mathbf{z}$ runs over the aperiodic binary sequences?

A formula for the critical exponent of a Sturmian word has been obtained several times, see for example [13,14]. Therefore it is natural to ask whether it is possible to obtain a formula for the value $P(\mathbf{z})$ as well. Computer-aided experiments indicate that $P(\mathbf{z})$ does not necessarily always equal $\phi + 1$.

Problem 39. For a Sturmian word $\mathbf{z}$, is there a formula for $P(\mathbf{z})$ using the continued fraction expansion of the slope of $\mathbf{z}$?

By Theorem 18, we know that every optimal squareful sequence is also an optimal abelian squareful sequence. The next question asks whether the converse holds.

Problem 40. Characterize optimal abelian squareful sequences.

The notion of an abelian squareful sequence extends naturally to abelian k-repetitive sequence for all *integers* $k \geq 2$. Then, for example, abelian squareful is the same as abelian 2-repetitive.

Currie and Visentin [15] showed that the Fibonacci word has abelian kth powers as a prefix for all integers $k \geq 2$. Recently it was shown that a stronger property holds true. Namely, the Fibonacci word is everywhere abelian k-repetitive for all integers $k \geq 2$, see [16].

It is easy to check that the Thue–Morse word $\mathbf{t}$ is abelian 2-repetitive. Also, since $\mathbf{t}$ can be factorized over 01 and 10, it is clear that each even position in $\mathbf{t}$ is occupied by abelian kth powers for every $k \geq 2$. However, this does not have to be true for odd positions. In fact, we believe that the following holds.

Conjecture 41. *The Thue–Morse word is abelian* 2*-repetitive, but not* 3*-repetitive.*

Acknowledgments

The author is grateful to J. Karhumäki for several discussions and for his encouragement to pursue investigating the topic of this paper. Thanks also to J. Cassaigne and D. Krieger for discussions about the values of $M(\alpha)$ and $P(\mathbf{t})$, respectively. The author would also like to thank the anonymous referee for many valuable comments. Finally, the support from the Finnish Academy under grant 8206039 is gratefully acknowledged.

References

[1] M. Lothaire, Algebraic Combinatorics on Words, in: Encyclopedia of Mathematics and its Applications, vol. 90, Cambrige University Press, Cambridge, 2002.
[2] F. Mignosi, A. Restivo, S. Salemi, Periodicity and the golden ratio, Theoret. Comput. Sci. 204 (1–2) (1998) 153–167. doi:10.1016/S0304-3975(98)00037-1.
[3] J. Karhumäki, A. Lepistö, W. Plandowski, Locally periodic versus globally periodic infinite words, J. Combin. Theory Ser. A 100 (2002) 250–264. doi:10.1006/jcta.2002.3292.
[4] A. Lepistö, On relations between local and global periodicity, Ph.D. Dissertation, Turku Centre for Computer Science, 2003.
[5] K. Saari, Everywhere α-repetitive sequences and Sturmian words, in: V. Diekert, M.V. Volkov, A. Voronkov (Eds.), Proc. of the Second International Symposium on Computer Science in Russia, in: Lecture Notes in Comput. Sci., vol. 4649, Springer, Berlin, 2007, pp. 363–372.
[6] C. Choffrut, J. Karhumäki, Combinatorics on words, in: A. Salomaa, G. Rozenberg (Eds.), Handbook of Formal Languages, Vol. 1, Springer, Berlin, 1997, pp. 329–438.
[7] K. Saari, On the periodicity and frequency of infinite words, Ph.D. Dissertation, Turku Centre for Computer Science, 2008.
[8] J.-P. Allouche, J. Shallit, Automatic Sequences. Theory, Applications, Generalizations, Cambrige University Press, Cambridge, 2003.
[9] Z.-X. Wen, Z.-Y. Wen, Some properties of the singular words of the Fibonacci word, European J. Combin. 15 (1994) 587–598. doi:10.1006/eujc.1994.1060.
[10] D.E. Knuth, J.H. Morris, V.R. Pratt, Fast pattern matching in strings, SIAM J. Comput. 6 (2) (1977) 323–350.
[11] V. Berthé, C. Holton, L.Q. Zamboni, Initial powers of Sturmian sequences, Acta Arith. 122 (4) (2006) 315–347.
[12] D. Krieger, J. Shallit, Every real number greater than 1 is a critical exponent, Theoret. Comput. Sci. 381 (1–3) (2007) 177–182. doi:10.1016/j.tcs.2007.04.037.
[13] A. Carpi, A. de Luca, Special factors, periodicity, and an application to Sturmian words, Acta Inform. 36 (2000) 983–1006.
[14] D. Damanik, D. Lenz, The index of Sturmian sequences, European J. Combin. 23 (2002) 23–29. doi:10.1006/eujc.2000.0496.
[15] J.D. Currie, T.I. Visentin, On Abelian 2-avoidable binary patterns, Acta Inform. 43 (2007) 521–533.
[16] G. Richomme, K. Saari, L.Q. Zamboni, Standard words and abelian powers in Sturmian words, in: Local Proc. of the 12èmes Journées Montoises d'Informatique Théorique, Institut d'Informatique Université de Mons-Hainaut and Académie Universitaire Wallonie-Bruxelles, 2008.