# Relation between powers of factors and the recurrence function characterizing Sturmian words 

Z. Masáková*, E. Pelantová<br>Doppler Institute \& Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Trojanova 13, 12000 Praha 2, Czech Republic

## A RTICLE INFO

## Article history:

Received 4 September 2008
Received in revised form 31 March 2009
Accepted 8 April 2009
Communicated by D. Perrin

## Keywords:

Recurrence function
Index of infinite words
Sturmian words


#### Abstract

In this paper we use the relation of the index of an infinite aperiodic word and its recurrence function to give another characterization of Sturmian words. As a by-product, we give a new proof of the theorem describing the index of a Sturmian word in terms of the continued fraction expansion of its slope. This theorem was independently proved in [A. Carpi, A. de Luca, Special factors, periodicity, and an application to Sturmian words, Acta Inform. 36 (2000) 983-1006] and [D. Damanik, D. Lenz, The index of Sturmian sequences, European J. Combin. 23 (2002) 23-29].


© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

Sturmian words constitute the most studied example of aperiodic infinite words. They appeared for the first time in the paper of Morse and Hedlund in 1938 [17]. But even after 70 years of extensive research, Sturmian words continue to attract the attention of numerous mathematicians and recently also computer scientists. The appeal of Sturmian words stems from the fact that they appear in various contexts. This is also why Sturmian words are often hidden under different titles: cutting sequences, Beatty sequences, mechanical words, etc. The beauty of Sturmian words consists in the abundance of equivalent definitions. Morse and Hedlund in [18] show that Sturmian words can be characterized by the so-called balance property. Ref. [14] contains a nice exposition on diverse definitions of Sturmian words. The most recent ones, which [14] does not mention, are characterization of Sturmian words using return words given by Vuillon [21] (for a less technical proof see [2]), characterization using the number of palindromes of given length given in [10] and yet another characterization by Richomme [19].

The aim of this paper is to give another equivalent definition of Sturmian words. Our characteristic gives a relation between the recurrence function and the index of an infinite word $u$. Obvious similarity between formulas for the recurrence quotient and the index of a Sturmian word was already noted in [1,5,7].

The recurrence function $R$ associates to every $n \in \mathbb{N}$ the minimal length $R(n) \in \mathbb{N}$ such that an arbitrary segment of the infinite word $u$ of length $R(n)$ contains all factors of $u$ of length $n$. This function has been studied already by Hedlund and Morse, who gave an explicit formula for $R(n)$ for an arbitrary Sturmian word $u$ and determined the so-called recurrence quotient, $\lim \sup _{n \rightarrow \infty} R(n) / n$. On the other hand, the index of an infinite word $u$ describes the maximal repetition of a factor of $u$. The study of the index of infinite words is considerably younger, though, in the last decade, very intense, especially due to applications in spectral theory for corresponding Schrödinger operators [8].

Repetitions in the most prominent Sturmian word, namely the Fibonacci word, were studied in [13]. More general results about the index of Sturmian words can be found in $[3,5,6,12,15,16,20]$. The complete solution to the problem was given independently by Carpi and de Luca in [7] and by Damanik and Lenz in [9].

[^0]The paper is organized as follows. In Section 2 we introduce all necessary notions and notations. Section 3 contains the proof of the main result of the paper, namely the following theorem.

Theorem 1. A uniformly recurrent infinite word $u$ is Sturmian if and only if there exist infinitely many factors $w$ of $u$ such that

$$
R(|w|)=|w| \operatorname{ind}(w)+1
$$

Notation $|w|$ stands for the length of the factor $w$, and ind $(w)$ is the maximal rational exponent $r$ such that $w^{r}$ is a factor of $u$.

It was pointed to us that already from [7] one can extract that Sturmian words satisfy the above equality for infinitely many factors. The proof of Carpi and de Luca uses the explicit formula for the recurrence function from [18]. Our proof relies on Vuillon's description of Sturmian words by return words and avoids manipulation with the continued fraction of the slope of the Sturmian word. Our theorem moreover states that Sturmian words are the only words having the above property.

With the help of Theorem 1, one can derive the upper bound on the index of $u$ (Section 4). In Section 5 we prove that the bound is in fact reached. For the construction of factors of $u$ with large repetition we use the knowledge of Sturmian morphisms, i.e. morphisms preserving the family of Sturmian words, as described in [4]. Sections 4 and 5 thus represent an alternative proof of the result of [7] and [9].

## 2. Preliminaries

An alphabet $\mathcal{A}$ is a finite set of symbols, called letters. A word $w$ of length $|w|=n$ is a concatenation of $n$ letters. The number of letters $X \in \mathcal{A}$ occurring in the word $w$ is denoted by $|w|_{X}$. The symbol $\mathcal{A}^{*}$ denotes the set of all finite words over the alphabet $\mathcal{A}$ including the empty word $\epsilon$. Equipped with the operation of concatenation, it is a monoid. We also define infinite words $u=\left(u_{n}\right)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$.

A finite word $v \in \mathcal{A}^{*}$ is called a factor of a word $w$ (finite or infinite) if there exist words $w^{(1)}$, $w^{(2)}$ such that $w=w^{(1)} v w^{(2)}$. If $w^{(1)}=\epsilon$, then $v$ is said to be a prefix of $w$; if $w^{(2)}=\epsilon$, then $v$ is a suffix of $w$. The set of all factors of length $n$ of an infinite word $u$ is denoted by $\mathscr{L}_{n}(u)$; the set of all factors of $u$ is called the language of $u$, and is denoted by $\mathcal{L}(u)$.

The mapping $\mathcal{C}: n \mapsto \# \mathcal{L}_{n}(u)$ is called the complexity of the infinite word $u$. For determining the complexity of an infinite word one uses the so-called special factors. A factor $w \in \mathcal{L}(u)$ is called left special if there exist letters $A, B \in \mathcal{A}$, $A \neq B$, such that both $A w$ and $B w$ belong to $\mathcal{L}(u)$. Similarly, one defines right special factors. A factor of $u$ is called bispecial if it is at the same time right special and left special. Every eventually periodic word has bounded complexity. For aperiodic words, one has for all $n \in \mathbb{N}$ that $\mathcal{C}(n) \geq n+1$. Infinite words, for which equality holds for all $n \in \mathbb{N}$, i.e. aperiodic words with minimal complexity, are called Sturmian words. Directly from the definition one can derive that in the language of a Sturmian word $u$ one has exactly one left special and exactly one right special factor of each length, and Sturmian words are characterized by this property.

Sturmian words are obviously defined over a binary alphabet, say $\{A, B\}$. The densities of letters $A, B$ in a Sturmian word $u=\left(u_{i}\right)_{i \in \mathbb{N}}$ are well defined and satisfy

$$
\varrho(A)=\lim _{n \rightarrow \infty} \frac{\left|u_{0} \cdots u_{n-1}\right|_{A}}{n}=\alpha, \quad \varrho(B)=\lim _{n \rightarrow \infty} \frac{\left|u_{0} \cdots u_{n-1}\right|_{B}}{n}=1-\alpha
$$

for some $\alpha \in(0,1)$. In fact, the language of a Sturmian word $u$ depends only on the parameter $\alpha$, which is also called the slope of $u$. For a given $\alpha$, one can construct all Sturmian words with the slope $\alpha$, for example, as codings of different orbits under an exchange of two intervals. Let $\alpha \in(0,1)$ be an irrational number. Denote $I=[0,1)(\operatorname{resp} . I=(0,1])$ and $I_{A}=[0, \alpha)$, $I_{B}=[\alpha, 1)\left(\operatorname{resp} . I_{A}=(0, \alpha], I_{B}=(\alpha, 1]\right)$. The mapping $T: I \mapsto I$ given by the prescription

$$
T(x)= \begin{cases}x+1-\alpha & \text { for } x \in I_{A}, \\ x-\alpha & \text { for } x \in I_{B}\end{cases}
$$

is called an exchange of two intervals with slope $\alpha$. For an arbitrary $x_{0} \in I$ we define an infinite word $\left(u_{n}\right)_{n \in \mathbb{N}}$ by

$$
\begin{equation*}
u_{n}=X \in\{A, B\} \quad \text { if } T^{n}\left(x_{0}\right) \in I_{X} \tag{1}
\end{equation*}
$$

It is known that the set of Sturmian words coincides with the set of infinite words given by the prescription (1). Since we assume that the slope is irrational, the language of a Sturmian word does not depend on the choice of the initial point $x_{0}$, but only on $\alpha$. Due to the symmetry $\alpha \leftrightarrow 1-\alpha$, studying the language of a Sturmian word, one can consider without loss of generality only parameters $\alpha>\frac{1}{2}$. From the exchange of intervals is not difficult to see that, with such an assumption, $\varrho(A)>\varrho(B)$ and, in fact, the Sturmian word can be viewed as composed by blocks of the form $A^{k}, A^{k+1}$, with $k=\left\lfloor\frac{\alpha}{1-\alpha}\right\rfloor$, separated by single letters $B$.

In this paper we study the repetition of factors in Sturmian words. We say that a word $v$ is a power of a word $w$ if $|v| \geq|w|$ and $v$ is a prefix of the periodic word $w w w \cdots$. We write $v=w^{r}$, where $r=|v| /|w|$. The index of a word $w$ in an infinite word $u$ is defined by

$$
\begin{equation*}
\operatorname{ind}(w)=\sup \left\{r \in \mathbb{Q} \mid w^{r} \in \mathscr{L}(u)\right\} \tag{2}
\end{equation*}
$$

A power $v$ of $w$ with maximal $r$ is called a maximal repetition of $w$. We have thus $v=w^{\text {ind }(w)}$. From what was said above, it is clear that, in a Sturmian word with slope $\alpha>\frac{1}{2}$, one has

$$
\begin{equation*}
\operatorname{ind}(B)=1 \quad \text { and } \quad \operatorname{ind}(A)=\left\lfloor\frac{\alpha}{1-\alpha}\right\rfloor+1 \tag{3}
\end{equation*}
$$

Taking the supremum of indices over all factors of an infinite word $u$, one obtains an important characteristic of $u$, the so-called index of $u$. Formally,

$$
\begin{equation*}
\operatorname{ind}(u)=\sup \{\operatorname{ind}(w) \mid w \in \mathscr{L}(u)\} \tag{4}
\end{equation*}
$$

It turns out that for the study of the index of Sturmian words, the notion of return words and recurrence function is important. A return word of a factor $w$ of an infinite word $u$ is a factor $v \in \mathscr{L}(u)$ such that $v w \in \mathcal{L}(u), w$ is a prefix of $v w$ and the factor $w$ occurs in $v w$ exactly twice. The factor $v w$ is often called a complete return word of $w$. The set of return words of a factor $w$ is denoted by $\operatorname{Ret}(w)$. If the set $\operatorname{Ret}(w)$ is finite for any factor $w$ of an infinite word $u$, then $u$ is said to be uniformly recurrent. In fact, this means that distances between consecutive occurrences of a given factor are bounded. Let us mention that for a uniformly recurrent word $u$ the supremum in (2) is always reached, as will be explained later, and therefore the notion of the index of $u$ in (4) has sense. For a uniformly recurrent infinite word $u$ we define a mapping $R: \mathbb{N} \mapsto \mathbb{N}$ by the prescription

$$
\begin{equation*}
R(n):=-1+\max \left\{|v w| \mid v \in \operatorname{Ret}(w), w \in \mathscr{L}_{n}(u)\right\}, \tag{5}
\end{equation*}
$$

i.e. $R(n)+1$ is equal to the maximum of lengths of a complete return word over all factors of length $n$. It is not difficult to see that an arbitrary segment of the infinite word $u$ of length $R(n)$ contains all factors of the word $u$ of length $n$. Formally, we have

$$
\begin{equation*}
\mathscr{L}_{n}(u)=\left\{u_{i} u_{i+1} \cdots u_{i+n-1} \mid k \leq i \leq k+R(n)-n+1\right\}, \quad \text { for all } k \in \mathbb{N} . \tag{6}
\end{equation*}
$$

Moreover, the number $R(n)$ is the smallest possible so that (6) remains valid. The mapping $R(n)$ is called the recurrence function of the infinite word $u$.

## 3. Recurrence function and index

Our aim is to find the relation between the recurrence function (well defined for uniformly recurrent words) and the index of aperiodic words. We first show that the index of every factor in an aperiodic uniformly recurrent word is finite, and we then determine a lower bound on the recurrence function.

Proposition 2. Let $u$ be an aperiodic uniformly recurrent word. Then for every factor $w \in \mathscr{L}(u)$ we have ind $(w)<+\infty$ and

$$
\begin{equation*}
R(|w|) \geq|w| \operatorname{ind}(w)+\mathcal{C}(|w|)-|w| \tag{7}
\end{equation*}
$$

Proof. Let $w=w_{1} \cdots w_{n}$ be a factor of $u$. We first show that $\operatorname{ind}(w)$ is finite. Without loss of generality, let ind $(w) \geq 2$. Obviously, all factors of the form $w_{i} \cdots w_{n} w_{1} \cdots w_{i-1}$ for any $1 \leq i \leq n$ belong to $\mathcal{L}(u)$. (Such factors are called conjugates of $w$.) Since $\mathcal{C}(n) \geq n+1$, there exists a factor $w^{\prime}$ which is not a conjugate of $w$. If $\mathcal{L}(u)$ contained factors $w^{k}$ for all $k \in \mathbb{N}$, then the distances between consecutive occurrences of $w^{\prime}$ would be unbounded, which would contradict the uniform recurrence of $u$. Therefore ind $(w)<+\infty$.

Now let $v$ be a maximal repetition of $w$. We prolong $v$ to a factor $v v^{\prime} \in \mathcal{L}(u)$ so that $v v^{\prime}$ contains all $\mathcal{C}(|w|)$ factors of $u$ of length $|w|$, but none of the prefixes of $v v^{\prime}$ does. Since $v$ has at most $|w|$ factors of length $|w|$ (namely the conjugates of $w)$, we must have $\left|v^{\prime}\right| \geq \mathcal{C}(|w|)-|w|$. From the definition of the recurrence function, we have

$$
R(|w|) \geq\left|v v^{\prime}\right| \geq|v|+\mathcal{C}(|w|)-|w|
$$

As $v=|w| \operatorname{ind}(w)$, the proof is complete.
Note that, in particular, for a Sturmian word $u$ one has $R(|w|) \geq|w| \operatorname{ind}(w)+1$ for every factor $w$ of $u$. The following proposition states that if equality is reached for infinitely many factors $w$ of an aperiodic word $u$, then $u$ is Sturmian.
Proposition 3. Let $u$ be an aperiodic uniformly recurrent infinite word. If there exist infinitely many factors $w \in \mathcal{L}(u)$ such that $R(|w|)=|w| \operatorname{ind}(w)+1$, then $u$ is a Sturmian word.

Proof. Using the assumption of the proposition and (7), there exist infinitely many factors $w$ of $u$ such that $\mathcal{C}(|w|) \leq|w|+1$, i.e. for infinitely many $n \in \mathbb{N}$ we have $\mathcal{C}(n) \leq n+1$. The complexity of an aperiodic word is a strictly increasing function and $\mathcal{C}(1) \geq 2$. This implies that $\mathcal{C}(n)=n+1$ for all $n$, and $u$ is therefore Sturmian.

In order to show the opposite implication to that of Proposition 3, we need to cite a nice result of Vuillon [21] which characterizes Sturmian words using return words. He shows that a binary infinite word $u$ is Sturmian if and only if every factor of $u$ has exactly two return words. For every factor $w$ of a Sturmian word $u$ thus there exist two finite words $r_{0}(w)$,
$r_{1}(w)$ such that the suffix of $u$ starting with the first occurrence of $w$ can be written as an infinite concatenation of blocks $r_{0}(w)$ and $r_{1}(w)$, i.e.

$$
u=p r_{i_{0}}(w) r_{i_{1}}(w) r_{i_{2}}(w) r_{i_{3}}(w) \cdots
$$

where $p$ is a prefix of $u$ and $i_{0}, i_{1}, i_{2}, i_{3}, \ldots \in\{0,1\}$. We can therefore define the so-called derivated word $v=\left(v_{n}\right)_{n \in \mathbb{N}}$ over the alphabet $\{0,1\}$ by the prescription $v_{n}=i_{n}$, coding the order of the blocks $r_{0}(w), r_{1}(w)$ in the infinite concatenation. We could now study return words of factors of the newly defined infinite word $v$. However, since return words of factors of the derivated word are in one-to-one correspondence with return words of factors in the original infinite word (see [11]), we deduce that every factor of $v$ has again exactly two return words, and thus is itself Sturmian.

It is obvious that for finding factors $w$ with the maximal index in the infinite word, we can limit our consideration to primitive factors $w$, i.e. such that $w \neq z^{k}$ for any $z \in \mathcal{L}(u)$ and any $k \in \mathbb{N}, k \geq 2$.
Proposition 4. Let $u$ be a Sturmian word and let $w \in \mathscr{L}_{n}(u)$ be a primitive factor such that $w w \in \mathscr{L}(u)$, and, moreover, let it have the maximal index among all factors of $u$ of length $n$ with the above properties. Then

$$
R(n)=n \operatorname{ind}(w)+1
$$

Proof. Let $k:=\lfloor\operatorname{ind}(w)\rfloor$ and $\theta:=\operatorname{ind}(w)-\lfloor\operatorname{ind}(w)\rfloor$. Then $w$ can be written as $w=w_{1} w_{2}$, where $\left|w_{1}\right|=\theta n$, and the maximal repetition of $w$ is the word

$$
\underbrace{\left(w_{1} w_{2}\right)\left(w_{1} w_{2}\right) \cdots\left(w_{1} w_{2}\right)}_{k \text { times }} w_{1} \in \mathscr{L}(u)
$$

Let us find $X, Y \in\{A, B\}$ such that

$$
\begin{equation*}
X w_{1} w_{2} \cdots w_{1} w_{2} w_{1} Y \in \mathscr{L}(u) \tag{8}
\end{equation*}
$$

Since ind $(w)=k+\theta$ is the greatest power such that $w^{k+\theta} \in \mathcal{L}(u)$, the letter $Y$ is not a prefix of $w_{2}$. Since $w$ is a primitive word with the greatest index in $\mathcal{L}_{n}(u)$, the letter $X$ is not a suffix of $w_{2}$. This, together with the fact that $k \geq 2$, means that $w_{1} w_{2}=w$ is a left special factor and $w_{2} w_{1}=: w^{\prime}$ is a right special factor. A Sturmian word has exactly one left special and one right special factor of each length.

Let us consider the Rauzy graph $\Gamma_{n}$ of $u$. The set of vertices of $\Gamma_{n}$ is equal to $\mathscr{L}_{n}(u)$ and the set of its edges to $\mathscr{L}_{n+1}(u)$. The Rauzy graph $\Gamma_{n}$ of a Sturmian word thus has $n+1$ vertices and $n+2$ edges. An edge $e \in \mathcal{L}_{n+1}(u)$ starts in a vertex $v \in \mathscr{L}_{n}(u)$ and ends in $v^{\prime} \in \mathscr{L}_{n}(u)$ if $v$ is a prefix and $v^{\prime}$ a suffix of $e$. An arbitrary factor $u$ of length $m \geq n$ in the language of the infinite word $u$ can be viewed as a path of length $m-n$ in the graph $\Gamma_{n}$ starting in the vertex corresponding to the prefix and ending in the vertex corresponding to the suffix of $u$ of length $n$.

Since $w \in \mathcal{L}_{n}(u), w w \in \mathscr{L}(u)$ and $w$ is primitive, there exists a cycle $C$ of length $n$ in the graph $\Gamma_{n}$ containing the factor $w$. Let us denote the vertices of the cycle $C$ by $v^{(0)}=w, v^{(1)}, \ldots, v^{(n-1)}$. Since $\Gamma_{n}$ has $n+1$ vertices, only one of them is missing in $C$. Let us denote it by $v^{(n)}$. Recall that $w$ is the only left special factor in $\mathscr{L}_{n}(u)$, and thus the only vertex in $\Gamma_{n}$ with indegree 2 . Similarly, $w^{\prime}$ is the only right special factor in $\mathcal{L}_{n}(u)$, and thus the only vertex in $\Gamma_{n}$ with outdegree 2 . Since $\Gamma_{n}$ is a strongly connected graph, an edge must go from the vertex $v^{(n)}$ to the cycle $C$ and an edge from the cycle $C$ to the vertex $v^{(n)}$. Thus $w^{\prime}=v^{(s)}$ for some $0 \leq s \leq n-1$. Relation (8) implies that the edge from $v^{(s)}$ to $v^{(n)}$ is $w_{2} w_{1} Y$ and the edge from $v^{(n)}$ to $v^{(0)}$ is $X w_{1} w_{2}$. The Rauzy graph $\Gamma_{n}$ is thus of the following form.


Let us consider the return words of $w$. Since $w w \in \mathcal{L}(u)$, one of the return words of $w$ is $r_{0}(w)=w$, the complete return word is $w w$, and the corresponding path in the Rauzy graph is the cycle $C$. We denote the other return word of $w$ by $r_{1}(w)$. From the structure of the graph $\Gamma_{n}$ it follows that the complete return word $r_{1}(w) w$ corresponds to the cycle $C^{\prime}$ given by vertices $v^{(0)}, v^{(1)}, \ldots, v^{(s)}, v^{(n)}$.

As we have already mentioned, the order of the blocks $r_{0}(w), r_{1}(w)$ is given by the derivated word over the alphabet $\{0,1\}$, which is Sturmian. Since $\left(r_{0}(w)\right)^{k}=w^{k} \in \mathcal{L}(u)$, for $k=\lfloor\operatorname{ind}(w)\rfloor \geq 2$, the derivated word has blocks $0^{k}, 0^{k-1}$ separated by single letters 1 . As a consequence, among all factors of length $n$, it is $v^{(n)}$ which has the longest complete return word, namely of the form

$$
X \underbrace{w w \cdots w}_{k \text { times }} w_{1} Y
$$

From the definition (5) it follows that

$$
R(n)=-1+\left|w^{k+\theta}\right|+2=1+(k+\theta) n
$$

which completes the proof.
Proof of Theorem 1. In order to complete the proof of Theorem 1, we have to show that there exist infinitely many primitive factors $w$ with index at least 2 . For the construction of such factors we make use of bispecial factors. Let $b$ be a bispecial factor in $\mathcal{L}(u)$. Denote by $n$ its length, $n:=|b|$, and by $r_{0}(b), r_{1}(b)$ its return words. From the Rauzy graph $\Gamma_{n}$ it follows that the two return words of $b$ are given by the two cycles in $\Gamma_{n}$, which have $b$ as the only common vertex. Therefore $\left|r_{0}(b)\right|+\left|r_{1}(b)\right|=n+2$. Without loss of generality, let $b$ contain both letters. Then $\left|r_{i}(b)\right| \geq 2$. At least for one of the return words, say $r_{0}(b)$, it holds that $n / 2<\left|r_{0}(b)\right| \leq n$, and therefore $r_{0}(b)$ is a prefix of $b$. It follows that the complete return word $r_{0}(b) b \in \mathcal{L}(u)$ has as its prefix $r_{0}(b) r_{0}(b)$. Moreover, a return word of an arbitrary factor of any uniformly recurrent word is primitive. Thus we can take $r_{0}(b)$ for the desired factor $w$. Since there are infinitely many bispecial factors $b$, we can construct infinitely many primitive factors with index $\geq 2$ and length $\geq \frac{|b|}{2}$.

## 4. Upper bound on the index of Sturmian words

In this section we mention the consequences of Proposition 4, which gives a relation between the recurrence function and index of factors of a Sturmian word. In particular, we can very easily derive the upper bound on the index of a Sturmian word, which constitutes an alternative proof for the result of Damanik and Lenz [9]. The bound depends on the continued fraction expansion of the slope of the Sturmian word.

Recall the notion of a continued fraction. To every irrational $\beta \in(0,1)$ one associates the continued fraction $\beta=$ [ $\left.0, b_{1}, b_{2}, \ldots\right]$, where $b_{i} \in \mathbb{Z}, b_{i} \geq 1$. Obviously, if $\beta>\frac{1}{2}$, then $b_{1}=1$. The convergents of $\beta$ form a sequence of fractions $\left(\frac{p_{n}}{q_{n}}\right)$,

$$
\frac{p_{1}}{q_{1}}=\frac{1}{b_{1}}, \quad \frac{p_{2}}{q_{2}}=\frac{1}{b_{1}+\frac{1}{b_{2}}}, \quad \frac{p_{3}}{q_{3}}=\frac{1}{b_{1}+\frac{1}{b_{2}+\frac{1}{b_{3}}}}
$$

We have $p_{n}$ coprime to $q_{n}$ and $\lim _{n \rightarrow \infty} \frac{p_{n}}{q_{n}}=\beta$.
It is known that the denominators $q_{n}$ of convergents of $\beta$ satisfy the recurrence

$$
q_{N}=b_{N} q_{N-1}+q_{N-2}
$$

with initial values $q_{-1}=0, q_{0}=1$. Denoting the matrix $M_{c}:=\left(\begin{array}{ll}c & 1 \\ 1 & 0\end{array}\right)$, then the recurrence can be rewritten as

$$
\left(q_{N}, q_{N-1}\right)=\left(q_{N-1}, q_{N-2}\right) M_{b_{N}}
$$

and by repetition, we obtain

$$
\left(q_{N}, q_{N-1}\right)=(1,0) M_{b_{1}} M_{b_{2}} \cdots M_{b_{N}}
$$

In order to extract the component $q_{N}$, it suffices to multiply the latter from the right by the vector $\binom{1}{0}$. We obtain

$$
\begin{equation*}
q_{N}=(1,0) M_{b_{1}} M_{b_{2}} \cdots M_{b_{N}}\binom{1}{0}=(1,0) M_{b_{N}} \cdots M_{b_{2}} M_{b_{1}}\binom{1}{0} \tag{9}
\end{equation*}
$$

where we have used that equality must hold also for the transpose $q_{N}^{T}=q_{N}$ and $M_{c}^{T}=M_{c}$ for all $c \in \mathbb{N}$.
For the derivation of the lower bound on the index of Sturmian words we use an old result on recurrence function of Sturmian words given in [18].
Theorem 5 ([18]). Let $u$ be a Sturmian word with slope $\alpha$. Denote by $q_{0}, q_{1}, q_{2}, \ldots$ the denominators of the convergents of $\alpha$. Then, for every $n \in \mathbb{N}$,

$$
R(n)=q_{N+1}+q_{N}+n-1, \quad \text { where } N \text { is such that } q_{N} \leq n<q_{N+1}
$$

Substituting into Proposition 4, one obtains an easy proof of the following result. One can find a similar derivation in [7].
Corollary 6. The index of every factor of a Sturmian word $u$ with slope $\alpha=\left[0,1, a_{2}, a_{3}, \ldots\right]$ is bounded by

$$
\sup \left\{\left.2+a_{N+1}+\frac{q_{N-1}-2}{q_{N}} \right\rvert\, N \geq 1\right\}
$$

where $q_{N}$ are the denominators of the convergents of $\alpha$.
Proof. Obviously, it suffices to consider only factors $w$ satisfying the assumptions of Proposition 4. Let $|w|=n$ and let $q_{N} \leq n<q_{N+1}$. Using Proposition 4 and Theorem 5, we have

$$
n \operatorname{ind}(w)+1=R(n)=q_{N+1}+q_{N}+n-1
$$

Therefore

$$
q_{N}(\operatorname{ind}(w)-1) \leq n(\operatorname{ind}(w)-1)=q_{N+1}+q_{N}-2=\left(a_{N+1}+1\right) q_{N}+q_{N-1}-2
$$

and consequently

$$
\operatorname{ind}(w) \leq 2+a_{N+1}+\frac{q_{N-1}-2}{q_{N}}
$$

## 5. Sturmian morphisms and factors with maximal index

In this section we provide a lower bound on the index of a Sturmian word $u$ of slope $\alpha$. Obviously, ind $(u) \geq a_{2}+1$, since $\left\lfloor\frac{\alpha}{1-\alpha}\right\rfloor$ in the formula (3) for the index of the letter $A$ is equal to the coefficient $a_{2}$ of the continued fraction of $\alpha$. The idea for the construction of factors with large index in a Sturmian word $u$ is based on the application of specific Sturmian morphisms. Since the application of a morphism preserves repetitions, it suffices to know how the chosen morphism changes the slope of the Sturmian word. Let us recall the necessary facts.

A morphism over the alphabet $\{A, B\}$ is a mapping $\varphi:\{A, B\}^{*} \mapsto\{A, B\}^{*}$ satisfying $\varphi\left(w_{1} w_{2}\right)=\varphi\left(w_{1}\right) \varphi\left(w_{2}\right)$. Obviously, a morphism is uniquely determined by $\varphi(A), \varphi(B)$. The incidence matrix of a morphism $\varphi$ is given by

$$
M_{\varphi}=\left(\begin{array}{ll}
|\varphi(A)|_{A} & |\varphi(A)|_{B} \\
|\varphi(B)|_{A} & |\varphi(B)|_{B}
\end{array}\right)
$$

The action of a morphism can be naturally extended to infinite words by

$$
\varphi\left(u_{0} u_{1} u_{2} \cdots\right)=\varphi\left(u_{0}\right) \varphi\left(u_{1}\right) \varphi\left(u_{2}\right) \cdots
$$

It is easy to show that, for the number of letters in the image of a word $w$, one has

$$
\begin{equation*}
\left(|\varphi(w)|_{A},|\varphi(w)|_{B}\right)=\left(|w|_{A},|w|_{B}\right) M_{\varphi} \tag{10}
\end{equation*}
$$

From that, we can deduce the following fact for the densities of letters in an infinite word $u$. If $\varrho(A), \varrho(B)$ are the densities in $u$, than the densities in the word $u^{\prime}=\varphi(u)$ are $\varrho^{\prime}(A), \varrho^{\prime}(B)$, where

$$
\begin{equation*}
\left(\varrho^{\prime}(A), \varrho^{\prime}(B)\right)=\text { const. }(\varrho(A), \varrho(B)) M_{\varphi} \tag{11}
\end{equation*}
$$

and const. is chosen so that $\varrho^{\prime}(A)+\varrho^{\prime}(B)=1$.
A morphism $\varphi$ is called Sturmian if $\varphi(u)$ is a Sturmian word for every Sturmian word $u$. Obviously, the set of Sturmian morphisms equipped with the operation of composition is a monoid, denoted by St. It is known [4] that the monoid St has three generators, namely

$$
\psi_{1}: \begin{array}{lll}
A & \mapsto & A B  \tag{12}\\
B & \mapsto & B
\end{array} \quad \psi_{2}: \begin{array}{lll}
A & \mapsto & B A \\
B & \mapsto & B
\end{array} \quad E: \begin{array}{lll}
A & \mapsto & B \\
B & \mapsto & A
\end{array} .
$$

Consider a Sturmian word with slope $\beta \in\left(\frac{1}{2}, 1\right)$ whose continued fraction is of the form $\beta=\left[0,1, b_{2}, b_{3}, \ldots\right]$. For $c \in \mathbb{N}$, we shall study the action of the morphism

$$
\varphi: \begin{array}{lll}
A & \mapsto & A^{c} B  \tag{13}\\
B & \mapsto & A
\end{array}
$$

on the Sturmian word $u$ with slope $\beta$. The morphism $\varphi$ is a Sturmian morphism; it is a composition of the generators (12) of the Sturmian monoid, namely $\varphi=E \psi_{2}^{c}$. The corresponding incidence matrix is $M_{\varphi}=M_{c}=\left(\begin{array}{ll}c & 1 \\ 1 & 0\end{array}\right)$, as defined at the beginning of Section 4. Consequently, the infinite word $\varphi(u)$ is also Sturmian, i.e. there exists an irrational $\beta^{\prime}$ such that $u^{\prime}:=\varphi(u)$ is a Sturmian word with slope $\beta^{\prime}$. According to (11), the densities of letters $a, b$ in the word $u^{\prime}$ satisfy

$$
\left(\beta^{\prime}, 1-\beta^{\prime}\right)=\text { const. }(\beta, 1-\beta)\left(\begin{array}{ll}
c & 1 \\
1 & 0
\end{array}\right)
$$

Therefore $\beta^{\prime}=\frac{c \beta+1-\beta}{c \beta+1}$. It is not difficult to show that the continued fraction of $\beta^{\prime}$ is equal to

$$
\begin{equation*}
\beta^{\prime}=\left[0,1, c, b_{2}, b_{3}, \ldots\right] \tag{14}
\end{equation*}
$$

The following lemma is crucial for the construction of factors of a Sturmian word with maximal index.
Lemma 7. Let $u$ be a Sturmian word with slope $\beta$ having the continued fraction $\beta=\left[0,1, b_{2}, b_{3}, \ldots\right]$. Let $w \in \mathcal{L}(u)$, and let $r \in \mathbb{Q}, r \geq 2$ be such that $v=w^{r} \in \mathcal{L}(u)$. Denote

$$
w^{\prime}=\varphi(w) \quad \text { and } \quad v^{\prime}=\varphi(v) A^{c}
$$

where $\varphi$ is the morphism given by (13). Then $v^{\prime}$ is a rational power of $w^{\prime}$ in a Sturmian word $u^{\prime}$ with slope $\beta^{\prime}=\left[0,1, c, b_{2}, b_{3}, \ldots\right]$.

Proof. If $|w|=1$, then necessarily $w=A, v=A^{r}$ for $2<r \leq b_{2}+1, \varphi(w)=A^{c} B$, and $\varphi(v) A^{c}=\left(A^{c} B\right)^{r} A^{c}$ is a factor of $u^{\prime}$, since a Sturmian word with slope $\beta^{\prime}=\left[0,1, c, b_{2}, b_{3}, \ldots\right]$ has blocks $A^{c}, A^{c+1}$ separated by single letters $B$.

If $|w| \geq 2$, let us write $w=w_{1} w_{2}$ so that $w_{2} \neq \epsilon$ and $v=\left(w_{1} w_{2}\right)^{\lfloor r\rfloor} w_{1}$. Then $\varphi(v) A^{c}=\varphi\left(w^{\lfloor r\rfloor}\right) \varphi\left(w_{1}\right) A^{c}$. In order to show that $\varphi(v) A^{c}$ is a power of $\varphi(w)$, it suffices to show that $\varphi\left(w_{1}\right) A^{c}$ is a prefix of $\varphi(w)$ or $\varphi(w) \varphi(w)$. If $w_{2}$ starts with $A$ or $B A$, then $\varphi\left(w_{2}\right)$ has prefix $A^{c}$, and thus $\varphi\left(w_{1}\right) A^{c}$ is a prefix of $\varphi(w)=\varphi\left(w_{1}\right) \varphi\left(w_{2}\right)$. Since $B B \notin \mathcal{L}(u)$, it remains to discuss the special case when $w_{2}=B$. As $|w| \geq 2$, we have $w_{1} \neq \epsilon$. Since $w_{2} w_{1} \in \mathcal{L}(u)$, the word $w_{1}$ must start with the letter $A$ and therefore $\varphi\left(w_{1}\right) A^{c}$ is a prefix of $\varphi\left(w_{1} B\right) \varphi\left(w_{1} B\right)=(\varphi(w))^{2}$.
Theorem 8. Let $u$ be a Sturmian word with slope $\alpha=\left[0,1, a_{2}, a_{3}, \ldots\right]$. Then for every $N \in \mathbb{N}$ there exists a factor $w \in \mathscr{L}(u)$ with index at least equal to $2+a_{N+1}+\frac{q_{N-1}-2}{q_{N}}$, where $q_{N}$ is the denominator of the $N$-th convergent of $\alpha$.
Proof. For $N=1$ it follows from the continued fraction of $\alpha$ that $q_{1}=1, q_{0}=1$, and therefore we have to find a factor with index $2+a_{2}-1=a_{2}+1$. It suffices to put $w=A$. Therefore we consider $N \geq 2$. We shall construct the desired factor $w$ and its power $v$ by $(N-1)$-fold application of Lemma 7. Consider the irrational number $\alpha_{0}$ with the continued fraction $\alpha_{0}=\left[0,1, a_{N+1}, a_{N+2, \ldots}\right]$. Take a Sturmian word $u^{(0)}$ with slope $\alpha_{0}$ and its factors $w^{(0)}:=A, v^{(0)}:=A^{1+a_{N+1}}$ for initial values of the construction. For $1 \leq i \leq N-1$, define

$$
w^{(i)}:=\varphi_{i}\left(w^{(i-1)}\right), \quad v^{(i)}:=\varphi_{i}\left(v^{(i-1)}\right) A^{a_{N-i+1}}, \quad \text { where } \varphi_{i}: \begin{array}{rll}
A & \mapsto & A^{a_{N-i+1}} B \\
B & \mapsto & A .
\end{array}
$$

By Lemma 7, the word $w^{(i)}$ is a factor of a Sturmian word $u^{(i)}$ with slope $\alpha_{i}$, where $\alpha_{i}$ has the continued fraction $\alpha_{i}=$ [ $\left.0,1, a_{N+1-i}, a_{N+2-i}, \ldots\right]$ and $v^{(i)}$ is a power of $w^{(i)}$ in the word $u^{(i)}$. In particular, $w^{(N-1)}$ is a factor of a Sturmian word $u$ with slope $\alpha=\left[0,1, a_{2}, a_{3}, \ldots\right]$ and $v^{(N-1)}$ is its power in $u$.

It suffices now to show that the length of $w^{(N-1)}$ is $q_{N}$ and the length of $v^{(N-1)}$ is $\left(2+a_{N+1}\right) q_{N}+q_{N-1}-2$. For the recurrent expression of lengths of factors $w^{(i)}, v^{(i)}$ we use formula (10). We have

$$
\left(\left|w^{(i)}\right|_{A},\left|w^{(i)}\right|_{B}\right)=\left(\left|w^{(i-1)}\right|_{A},\left|w^{(i-1)}\right|_{B}\right) M_{a_{N-i+1}},
$$

for all $i=1,2, \ldots, N-1$, with $\left(\left|w^{(0)}\right|_{A},\left|w^{(0)}\right|_{B}\right)=(1,0)$. It can be easily seen that

$$
\left(\left|w^{(N-1)}\right|_{A},\left|w^{(N-1)}\right|_{B}\right)=(1,0) M_{a_{N}} M_{a_{N-1}} \cdots M_{a_{2}}
$$

In order to obtain $\left|w^{(N-1)}\right|=\left|w^{(N-1)}\right|_{A}+\left|w^{(N-1)}\right|_{B}$, we multiply the latter from the right by the vector $\binom{1}{1}$, which can be also written as $\binom{1}{1}=\left(\begin{array}{ll}1 & 1 \\ 1 & 0\end{array}\right)\binom{1}{0}$. Since in the continued fraction of $\alpha$ we have $a_{1}=1$, we can use (9) to obtain

$$
\left|w^{(N-1)}\right|=(1,0) M_{a_{N}} M_{a_{N-1}} \cdots M_{a_{2}} M_{a_{1}}\binom{1}{0}=q_{N} .
$$

From the definition of words $v^{(i)}$ we have for their lengths

$$
\begin{equation*}
\left(\left|v^{(i)}\right|_{A},\left|v^{(i)}\right|_{B}\right)=\left(\left|v^{(i-1)}\right|_{A},\left|v^{(i-1)}\right|_{B}\right) M_{a_{N-i+1}}+\left(a_{N-i+1}, 0\right), \tag{15}
\end{equation*}
$$

with $\left(\left|v^{(0)}\right|_{A},\left|v^{(0)}\right|_{B}\right)=\left(1+a_{N+1}, 0\right)$. Let us compute the lengths for $N=1$,

$$
\left(\left|v^{(1)}\right|_{A},\left|v^{(1)}\right|_{B}\right)=\left(1+a_{N+1}, 0\right)\left(\begin{array}{cc}
a_{N} & 1 \\
1 & 0
\end{array}\right)+\left(a_{N}, 0\right)=\left(2+a_{N+1}\right)(1,0) M_{a_{N}}+(1,0)-(1,1)
$$

Since for every $c$ we have $-(1,1) M_{c}+(c, 0)=-(1,1)$, by repeated application of the recurrence (15) we obtain

$$
\left(\left|v^{(N-1)}\right|_{A},\left|v^{(N-1)}\right|_{B}\right)=\left(2+a_{N+1}\right)(1,0) M_{a_{N}} M_{a_{N-1}} \cdots M_{a_{2}}+(1,0) M_{a_{N-1}} \cdots M_{a_{2}}-(1,1)
$$

Again, multiplying the latter from the right by the vector $\binom{1}{1}=M_{a_{1}}\binom{1}{0}$ and using (9), we obtain

$$
\left|v^{(N-1)}\right|=\left(2+a_{N+1}\right) q_{N}+q_{N-1}-2
$$

## Acknowledgements

We are grateful to J.-P. Allouche for pointing out the reference [7]. We acknowledge financial support by the Czech Science Foundation grant 201/09/0584 and by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic.

## References

[1] B. Adamczewski, J.-P. Allouche, Reversals and palindromes in continued fractions, Theoret. Comput. Sci. 380 (2007) $220-237$.
[2] L. Balková, E. Pelantová, W. Steiner, Sequences with constant number of return words, Monatsh. Math. 155 (2008) 251-263.
[3] J. Berstel, On the index of Sturmian words, in: Jewels are Forever, Springer, 1999, pp. 287-294.
[4] J. Berstel, P. Séébold, Morphismes de Sturm, Bull. Belg. Math. Soc. Simon Stevin 1 (1994) 175-189.
[5] V. Berthé, C. Holton, L.Q. Zamboni, Initial powers of Sturmian sequences, Acta Arith. 122 (2006) 315-347.
[6] W.-T. Cao, Z.-Y. Wen, Some properties of the factors of Sturmian sequences, Theoret. Comput. Sci. 304 (2003) 365-385.
[7] A. Carpi, A. de Luca, Special factors, periodicity, and an application to Sturmian words, Acta Inform. 36 (2000) 983-1006.
[8] D. Damanik, Singular continuous spectrum for a class of substitution Hamiltonians. II, Lett. Math. Phys. 54 (2000) 25-31.
[9] D. Damanik, D. Lenz, The index of Sturmian sequences, European J. Combin. 23 (2002) 23-29.
[10] X. Droubay, G. Pirillo, Palindromes and Sturmian words, Theoret. Comput. Sci. 223 (1999) 73-85.
[11] F. Durand, A characterization of substitutive sequences using return words, Discrete Math. 179 (1998) 89-101.
[12] J. Justin, G. Pirillo, Fractional powers in Sturmian words, Theoret. Comput. Sci. 255 (2001) 363-376.
[13] J. Karhumäki, On cube-free $\omega$-words generated by binary morphisms, Discrete Appl. Math. 5 (1983) 279-297.
[14] M. Lothaire, Algebraic combinatorics on words, in: Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.
[15] F. Mignosi, G. Pirillo, Repetitions in the Fibonacci infinite word, RAIRO Inform. Théor. Appl. 26 (1992) 199-204.
[16] F. Mignosi, Infinite words with linear subword complexity, Theoret. Comput. Sci. 65 (1989) 221-242.
[17] M. Morse, G.A. Hedlund, Symbolic dynamics, Amer. J. Math. 60 (1938) 815-866.
[18] M. Morse, G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories, Amer. J. Math. 62 (1940) 1-42.
[19] G. Richomme, Another characterization of Sturmian words (one more), Bull. Eur. Assoc. Theor. Comput. Sci. EATCS 67 (1999) 173-175.
[20] D. Vandeth, Sturmian words and words with a critical exponent, Theoret. Comput. Sci. 242 (2000) 283-300.
[21] L. Vuillon, A characterization of Sturmian words by return words, European J. Combin. 22 (2001) 263-275.


[^0]:    * Corresponding author. Tel.: +420 224358544.

    E-mail addresses: zuzana.masakova@fjfi.cvut.cz (Z. Masáková), edita.pelantova@fjfi.cvut.cz (E. Pelantová).

