Characterization of repetitions in Sturmian words: A new proof

Jarkko Peltomäki
jspelt@utu.fi

Turku Centre for Computer Science TUCS, 20520 Turku, Finland
University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland

Abstract

We present a new, dynamical way to study powers (that is, repetitions) in Sturmian words based on results from Diophantine approximation theory. As a result, we provide an alternative and shorter proof of a result by Damanik and Lenz characterizing powers in Sturmian words [Powers in Sturmian sequences, Eur. J. Combin. 24 (2003), 377–390]. Further, as a consequence, we obtain a previously known formula for the fractional index of a Sturmian word based on the continued fraction expansion of its slope.

Keywords: sturmian word, standard word, power, combinatorics on words, continued fraction

1 Introduction

In 2003 Damanik and Lenz [6] completely described factors of length n of a Sturmian word which occur as p^{th} powers for every $n \geq 0$ and $p \geq 1$. Damanik and Lenz prove a series of results concerning how factors of a Sturmian word align to the corresponding (finite) standard words. By a careful analysis of the alignment, they obtain the complete description of powers thanks to known results on powers of standard words. Our method is based on the dynamical view of Sturmian words as codings of irrational rotations. Translating word-combinatorial concepts into corresponding dynamical concepts allows us to apply powerful results from Diophantine approximation theory (such as the Three Distance Theorem) providing a more geometric proof of the result of Damanik and Lenz. Our methods allow us to avoid tricky alignment arguments making the proof in our opinion easier to follow. Furthermore, the results allow us to infer a formula for the fractional index of a Sturmian word based on the continued fraction expansion of its slope. This formula and its proof appeared in an earlier paper by Damanik and Lenz [5] and was also established purely combinatorially using alignment arguments. The formula was independently obtained with different methods by Carpi and de Luca [3] and Justin and Pirillo [7]. For partial results and works related to powers in Sturmian words see e.g. the papers of Mignosi [11], Berstel [2], Vandeth [13], and Justin and Pirillo [7].

The paper is organized as follows: in Section 2 we briefly recall results concerning continued fractions and rational approximations and prove the purely number-theoretic and important Proposition 2.2 for later use in Section 4. In Section 3 we state needed facts about Sturmian words with appropriate references. Section 4 contains the main results and their proofs.

2 Continued Fractions and Rational Approximations

Every irrational real number α has a unique infinite continued fraction expansion

$$\alpha = [a_0; a_1, a_2, a_3, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} \tag{1}$$

with $a_0 \in \mathbb{Z}$ and $a_k \in \mathbb{N}$ for all $k \geq 1$. The numbers a_i are called the *partial quotients* of α. Good references on continued fractions are the books of Khinchin [8] and Cassels [4]. We focus here only on irrational numbers, but we note that with small tweaks much of what follows also holds for rational numbers, which have finite continued fraction expansions.

The *convergents* $c_k = \frac{p_k}{q_k}$ of α are defined by the recurrences

$$\begin{aligned} p_0 &= a_0, & p_1 &= a_1 a_0 + 1, & p_k &= a_k p_{k-1} + p_{k-2}, & k &\geq 2, \\ q_0 &= 1, & q_1 &= a_1, & q_k &= a_k q_{k-1} + q_{k-2}, & k &\geq 2. \end{aligned}$$

The sequence $(c_k)_{k \geq 0}$ converges to α. Moreover, the even convergents are less than α and form an increasing sequence and, on the other hand, the odd convergents are greater than α and form a decreasing sequence.

If $k \geq 2$ and $a_k > 1$, then between the convergents c_{k-2} and c_k there are *semiconvergents* (called intermediate fractions in Khinchin's book [8]) which are of the form

$$\frac{p_{k,l}}{q_{k,l}} = \frac{l p_{k-1} + p_{k-2}}{l q_{k-1} + q_{k-2}}$$

with $1 \leq l < a_k$. When the semiconvergents (if any) between c_{k-2} and c_k are ordered by the size of their denominators, the obtained sequence is increasing if k is even and decreasing if k is odd.

Note that we make a clear distinction between convergents and semiconvergents, i.e., convergents are not a specific subtype of semiconvergents.

For the rest of this paper we make the convention that α refers to an irrational number with a continued fraction expansion as in (1) having convergents $\frac{p_k}{q_k}$ and semiconvergents $\frac{p_{k,l}}{q_{k,l}}$ as above.

A rational number $\frac{a}{b}$ is a *best approximation* of the real number α if for every fraction $\frac{c}{d}$ such that $\frac{c}{d} \neq \frac{a}{b}$ and $d \leq b$ it holds that

$$|b\alpha - a| < |d\alpha - c| .$$

In other words, any other multiple of α with a coefficient at most b is further away from the nearest integer than is $b\alpha$. The next proposition shows that the best approximations of an irrational number are connected to its convergents (for a proof see Theorems 16 and 17 of [8]).

Proposition 2.1. *The best rational approximations of an irrational number are exactly its convergents.*

We identify the unit interval $[0, 1)$ with the unit circle $\mathbb{T}$. Let $\alpha \in (0, 1)$ be irrational. The map

$$R : [0, 1) \to [0, 1),\ x \mapsto \{x + \alpha\},$$

where $\{x\}$ stands for the fractional part of the number x, defines a rotation on $\mathbb{T}$. The circle partitions into the intervals $(0, \frac{1}{2})$ and $(\frac{1}{2}, 1)$. Points in the same interval of the partition are said to be on the same side of 0, and points in different intervals are said to be on the opposite sides of 0. (We are not interested in the location of the point $\frac{1}{2}$.) The points $\{q_k \alpha\}$ and $\{q_{k-1}\alpha\}$ are always on the opposite sides of 0. The points $\{q_{k,l}\alpha\}$ with $0 < l \leq a_k$ always lie between the points $\{q_{k-2}\alpha\}$ and $\{q_k \alpha\}$; see (3).

We measure the shortest distance to 0 on $\mathbb{T}$ by setting

$$\|x\| = \min\{\{x\}, 1 - \{x\}\}.$$

We have the following facts for $k \geq 2$ and for all l such that $0 < l \leq a_k$:

$$\|q_{k,l}\alpha\| = (-1)^k(q_{k,l}\alpha - p_{k,l}), \tag{2}$$
$$\|q_{k,l}\alpha\| = \|q_{k,l-1}\alpha\| - \|q_{k-1}\alpha\|. \tag{3}$$

We can now interpret Proposition 2.1 as

$$\min_{0<n<q_k} \|n\alpha\| = \|q_{k-1}\alpha\|, \quad \text{for } k \geq 1. \tag{4}$$

Note that rotating preserves distances; a fact we will often use without explicit mention. In particular, the distance between the points $\{n\alpha\}$ and $\{m\alpha\}$ is $\||n-m|\alpha\|$. Thus by (4) the minimum distance between the distinct points $\{n\alpha\}$ and $\{m\alpha\}$ with $0 \leq n, m < q_k$ is at least $\|q_{k-1}\alpha\|$. The formula (4) tells what is the point closest to 0 among the points $\{n\alpha\}$ for $1 \leq n \leq q_k - 1$. We are also interested to know the point closest to 0 on the side opposite to $\{q_{k-1}\alpha\}$. The next result is very important and concerns this.

Proposition 2.2. *Let α be an irrational number. Let n be an integer such that $0 < n < q_{k,l}$ with $k \geq 2$ and $0 < l \leq a_k$. If $\|n\alpha\| < \|q_{k,l-1}\alpha\|$, then $n = mq_{k-1}$ for some integer m such that $1 \leq m \leq \min\{l, a_k - l + 1\}$.*

Proof. Suppose that $\|n\alpha\| < \|q_{k,l-1}\alpha\|$, and assume for a contradiction that the point $\{n\alpha\}$ is on the same side of 0 as $\{q_{k-2}\alpha\}$. Since $n < q_{k,l}$, we conclude that $n \neq q_{k,r}$ for $r \geq l$. By (3) and our assumption that $\|n\alpha\| < \|q_{k,l-1}\|$, we see that $n \neq q_{k,r}$ with $0 \leq r \leq l-1$. As $\|n\alpha\| > \|q_k\alpha\|$ by (4), we infer that the point $\{n\alpha\}$ must lie between the points $\{q_{k,l'}\alpha\}$ and $\{q_{k,l'+1}\alpha\}$ for some l' such that $0 \leq l' < a_k$. The distance between the points $\{n\alpha\}$ and $\{q_{k,l'}\}$ is less than $\|q_{k-1}\alpha\|$. By (4), it must be that $q_{k,l'} \geq q_k$; a contradiction.

Suppose for a contradiction that n is not a multiple of q_{k-1}. Then the point $\{n\alpha\}$ lies between the points $\{tq_{k-1}\alpha\}$ and $\{(t+1)q_{k-1}\alpha\}$ for some t such that $0 < t < \lfloor 1/\|q_{k-1}\alpha\| \rfloor$. As $\{n\alpha\}$ is on the same side of 0 as the point $\{q_{k-1}\alpha\}$, it follows that $\|n\alpha\| > \|tq_{k-1}\alpha\|$ and $\|tq_{k-1}\alpha\| = t\|q_{k-1}\alpha\|$. The distance between the points $\{n\alpha\}$ and $\{tq_{k-1}\alpha\}$ is less than $\|q_{k-1}\alpha\|$, so by (4) it must be that $tq_{k-1} \geq q_k = a_k q_{k-1} + q_{k-2}$. Thus necessarily $t > a_k$. Using (3) we see that the distance between the points $\{q_k\alpha\}$ and $\{q_{k-2}\alpha\}$ is $a_k\|q_{k-1}\alpha\|$. Since $\|q_k\alpha\| < \|q_{k-1}\alpha\|$, we infer that

$$\|q_{k,l-1}\alpha\| \leq \|q_{k-2}\alpha\| = a_k\|q_{k-1}\alpha\| + \|q_k\alpha\| < (a_k+1)\|q_{k-1}\alpha\|. \tag{5}$$

Therefore by our assumption,

$$(a_k+1)\|q_{k-1}\alpha\| > \|q_{k,l-1}\alpha\| > \|n\alpha\| > t\|q_{k-1}\alpha\|,$$

so $a_k \geq t$; a contradiction. We have thus concluded that $n = mq_{k-1}$ for some $m \geq 1$.

Let us now analyze the upper bound on m. First of all, $mq_{k-1} < q_{k,l}$ exactly when $m \leq l \leq a_k$. It follows that $\|mq_{k-1}\alpha\| = m\|q_{k-1}\alpha\|$. By (3)

$$m\|q_{k-1}\alpha\| < \|q_{k,l-1}\alpha\| = (a_k - (l-1))\|q_{k-1}\alpha\| + \|q_k\alpha\|,$$

so $m \leq a_k - l + 1$. We conclude that $m \leq \min\{l, a_k - l + 1\}$. ∎

The inequalities (3) and (5) imply that $a_k\|q_{k-1}\alpha\| < \|q_{k-2}\alpha\| < (a_k+1)\|q_{k-1}\alpha\|$. We derive the following useful fact:

$$a_k = \left\lfloor \frac{\|q_{k-2}\alpha\|}{\|q_{k-1}\alpha\|} \right\rfloor. \tag{6}$$

We need the famous Three Distance Theorem (see e.g. [1] and the references therein).

Theorem 2.3 (The Three Distance Theorem). *Let α be an irrational number, and let $n > a_1$ be a positive integer uniquely expressed in the form $n = lq_{k-1} + q_{k-2} + r$ with $k \geq 2, 0 < l \leq a_k$, and $0 \leq r < q_{k-1}$. The points $0, \{\alpha\}, \{2\alpha\}, \ldots, \{n\alpha\}$ partition the circle $\mathbb{T}$ into $n+1$ intervals. There are exactly*

- $n + 1 - q_{k-1}$ *intervals of length* $\|q_{k-1}\alpha\|$,
- $r + 1$ *intervals of length* $\|q_{k,l}\alpha\|$, *and*
- $q_{k-1} - (r + 1)$ *intervals of length* $\|q_{k,l-1}\alpha\|$.

By (3) the intervals of the last type (if they exist) are the longest, and their length is the sum of the two other length types.

3 Word Combinatorics and Sturmian Words

We mention here only few key concepts from combinatorics on words; good background references are Lothaire's books [9, 10].

A word is *primitive* if it is not a non-trivial power of some word. A word w is primitive if and only if it occurs exactly twice in w^2. The *cyclic shift operator* C is defined by $C(a_1 \cdots a_{n-1}a_n) = a_n a_1 \cdots a_{n-1}$ where the a_i are letters. A word u is *conjugate* to v if $C^i(v) = u$ for some i such that $0 \leq i < |v|$. The *reversal* of the word $w = a_1 \cdots a_{n-1}a_n$ where the a_i are letters is the word $\widetilde{w} = a_n a_{n-1} \cdots a_1$.

Sturmian words are a well-known class of infinite, aperiodic binary words over $\{0, 1\}$ with minimal factor complexity. They are defined as the (right-)infinite words having $n + 1$ factors of length n for every $n \geq 0$. For our purposes it is more convenient to view Sturmian words equivalently as the infinite words obtained as codings of orbits of points in an irrational circle rotation with two intervals [12, 10]. Let us make this more precise. The frequency α of letter 1 (called the *slope*) in a Sturmian words exists, and it is irrational. Divide the circle $\mathbb{T}$ into two intervals I_0 and I_1 defined by the points 0 and $1 - \alpha$, and define the coding function ν by setting $\nu(x) = 0$ if $x \in I_0$ and $\nu(x) = 1$ if $x \in I_1$. The coding of the orbit of a point x is the infinite word $s_{x,\alpha}$ obtained by setting its $n^{\text{th}}, n \geq 0$, letter to equal $\nu(R^n(x))$ where R is the rotation by angle α. This word is Sturmian with slope α, and conversely every Sturmian word with slope α is obtained this way. To make the definition proper, we need to define how ν behaves in the endpoints 0 and $1 - \alpha$. We have two options: either take $I_0 = [0, 1 - \alpha)$ and $I_1 = [1 - \alpha, 1)$ or $I_0 = (0, 1 - \alpha]$ and $I_1 = (1 - \alpha, 1]$. The difference is seen in the codings of the orbits of the special points $\{-n\alpha\}$, and both options are needed to be able to obtain every Sturmian word of slope α as a coding of a rotation. However, in this paper we are not concerned about this choice. We make the convention that $I(x, y)$ with $x, y \neq 0$ is either of the half-open intervals of $\mathbb{T}$ separated by the points x and y (taken modulo 1 if necessary) not containing the point 0 as an interior point. The interval $I(x, 0) = I(0, x)$ is either of the half-open intervals separated by the points 0 and x having smallest length (the case $x = \frac{1}{2}$ is not important in this paper). Since the sequence $(\{n\alpha\})_{n \geq 0}$ is dense in $[0, 1)$—as is well-known—every Sturmian word of slope α has the same language (that is, the set of factors); this language is denoted by $\mathcal{L}(\alpha)$. Thus to study repetitions, it is sufficient to analyze $\mathcal{L}(\alpha)$. The *fractional index* $\operatorname{ind}_{\mathbb{Q}}(w)$ of a nonempty factor $w \in \mathcal{L}(\alpha)$ is defined as

$$\operatorname{ind}_{\mathbb{Q}}(w) = \sup\{k \in \mathbb{Q} \colon\ w^k \in \mathcal{L}(\alpha)\}$$

where the *fractional power* w^k is the word $(uv)^n u$ with $w = uv$ and $k = n + \frac{|u|}{|w|}$. The *index* $\operatorname{ind}(w)$ of a nonempty factor w is defined similarly by letting k take only integral values. The index of a factor in $\mathcal{L}(\alpha)$ is always finite. The *fractional index of a Sturmian word* with slope α is defined to be

$$\sup\{\operatorname{ind}_{\mathbb{Q}}(w) \colon\ w \in \mathcal{L}(\alpha)\}.$$

This quantity can be infinite.

For every factor $w = a_0 a_1 \cdots a_{n-1}$ of length n there exists a unique subinterval $[w]$ of $\mathbb{T}$ such that $s_{x,\alpha}$ begins with w if and only if $x \in [w]$. Clearly

$$[w] = I_{a_0} \cap R^{-1}(I_{a_1}) \cap \ldots \cap R^{-(n-1)}(I_{a_{n-1}}).$$

We denote the length of the interval $[w]$ by $|[w]|$. The points $0, \{-\alpha\}, \{-2\alpha\}, \ldots, \{-n\alpha\}$ partition the circle into $n+1$ intervals, which have one-to-one correspondence with the words of $\mathcal{L}(\alpha)$ of length n. Among these intervals the interval containing the point $\{-(n+1)\alpha\}$ corresponds to the right special factor of length n. A factor w is *right special* if both $w0, w1 \in \mathcal{L}(\alpha)$. Similarly a factor is *left special* if both $0w, 1w \in \mathcal{L}(\alpha)$. In a Sturmian word there exists a unique right special and a unique left special factor of length n for all $n \geq 0$. The language $\mathcal{L}(\alpha)$ is mirror-invariant, that is, for every $w \in \mathcal{L}(\alpha)$ also $\widetilde{w} \in \mathcal{L}(\alpha)$. It follows that the right special factor of length n is the reversal of the left special factor of length n.

Given the continued fraction expansion of an irrational $\alpha \in (0,1)$, we define the corresponding *standard sequence* $(s_k)_{k\geq 0}$ of words by

$$s_{-1} = 1, \qquad s_0 = 0, \qquad s_1 = s_0^{a_1-1}s_{-1}, \qquad s_k = s_{k-1}^{a_k}s_{k-2}, \qquad k \geq 2.$$

As s_k is a prefix of s_{k+1} for $k \geq 1$, the sequence (s_k) converges to a unique infinite word c_α called the infinite standard Sturmian word of slope α, and it equals $s_{\alpha,\alpha}$. Inspired by the notion of semiconvergents, we define *semistandard* words for $k \geq 2$ by

$$s_{k,l} = s_{k-1}^{l}s_{k-2}$$

with $1 \leq l < a_k$. Clearly $|s_k| = q_k$ and $|s_{k,l}| = q_{k,l}$. Every prefix of c_α is left special, so in particular, standard and semistandard words are left special. Every standard or semistandard word is primitive [10, Proposition 2.2.3]. An important property of standard words is that the words s_k and s_{k-1} almost commute; namely $s_k s_{k-1} = wab$ and $s_{k-1}s_k = wba$ for some word w and distinct letters a and b. For more information about standard words see Chapter 2 of [10] and Berstel's paper [2]. Here we see that the only difference between the words c_α and $c_{\overline{\alpha}}$ where $\alpha = [0; 1, a_2, a_3, \ldots]$ and $\overline{\alpha} = [0; a_2+1, a_3, \ldots]$ is that the roles of the letters 0 and 1 are reversed. Thus for the study of powers, we may assume without loss of generality that $a_1 \geq 2$.

For the rest of this paper we make the convention that the partial quotients of an irrational α satisfy $a_0 = 0$ and $a_1 \geq 2$, that is, $0 < \alpha < \frac{1}{2}$. Moreover, the words s_k and $s_{k,l}$ refer to the standard or semistandard words of slope α.

4 The Main Results

This section presents a complete description of powers occurring in a Sturmian word with slope α. As a side-product, in Theorem 4.3 we obtain a description of conjugacy classes of length $q_{k,l}$. Finally, as an easy consequence of the established results, we obtain a formula for the fractional index of a Sturmian word (Theorem 4.7).

The following important proposition shows the usefulness of Proposition 2.2 in the study of Sturmian words and plays a role similar to Theorem 1 of [6].

Proposition 4.1. *If $w^2 \in \mathcal{L}(\alpha)$ with w primitive, then $|w| = q_k$ for some $k \geq 0$ or $|w| = q_{k,l}$ for some $k \geq 2$ with $0 < l < a_k$.*

Proof. Let $n = |w|$. If $n < q_1 = a_1$, then the factors of length n are readily seen to be 0^n and the conjugates of $0^{n-1}1$. Since the minimum number of letters 0 between two occurrences of letter 1 in $\mathcal{L}(\alpha)$ is $a_1 - 1$ and the maximum number is a_1, the only way w^2 can be a factor is that $w = 0 = s_0$. Suppose then that $n \geq q_1$ and $[w] = I(-i\alpha, -j\alpha)$ with $0 \leq i, j \leq n$. We may assume without loss of generality that w is right special, so $\{-(n+1)\alpha\} \in [w]$. Further, since $[w^2] = [w] \cap R^{-n}([w]) \neq \emptyset$, then necessarily (depending on n) either $[w^2] = I(-i\alpha, -(j+n)\alpha)$ or $[w^2] = I(-j\alpha, -(i+n)\alpha)$. We assume that $[w^2] = I(-i\alpha, -(j+n)\alpha)$; the other case is symmetric. We wish to prove that the points $\{-(n+1)\alpha$ and $\{-(j+n)\alpha\}$ are actually the same point. This is equivalent to saying that $j = 1$. Assume on the contrary that $j \neq 1$. Let a be the first letter of w and b be a letter such that $b \neq a$. Note that $[w^2] \subset [wa]$. Now as w is right special, $[wa] = I(-j\alpha, -(n+1)\alpha)$ and $[wb] = I(-(n+1)\alpha, -i\alpha)$. Let $x \in [w^2]$ and $y \in [wa] \setminus [w^2]$, and let u be the longest common prefix of $s_{x,\alpha}$ and $s_{y,\alpha}$. Since $[w^2] \neq [wa]$, we have that $|u| < 2|w|$. Moreover, u is right special, so w is a suffix of u. However, w^2 is a prefix of $s_{x,\alpha}$ implying that u is a prefix of w^2. Thus w^2 contains at least three occurrences of w contradicting the primitivity of w. From this contradiction we conclude that $j = 1$. There are no points $\{-m\alpha\}$ in the interval $I(-(j+n)\alpha, -j\alpha) = I(-(n+1)\alpha, -\alpha)$ with $m \leq n$. Therefore the point $\{-n\alpha\}$ is the closest point to 0 from either side. If $q_1 \leq n < q_{2,1}$, then it must be that $n = q_1$. Otherwise let $k \geq 2$ be such that $q_{k,l} \leq n < q_{k,l+1}$ with $0 < l \leq a_k$. By Proposition 2.2 either $n = q_{k-1}$ or $n = q_{k,l}$ proving the claim. □

Indeed, for each length given in the statement of the previous proposition, there exists a factor occurring as a square.

Lemma 4.2. *We have that $s_0^2, s_1^2 \in \mathcal{L}(\alpha)$ and $s_{k,l}^2 \in \mathcal{L}(\alpha)$ for all $k \geq 2$ and l such that $0 < l \leq a_k$.*

Proof. As $s_0^2 = 0^2$ and $s_1^2 = (0^{a_1-1}1)^2$, clearly $s_0^2, s_1^2 \in \mathcal{L}(\alpha)$. Since the words $s_{k+1}s_k$ and $s_k s_{k+1}$ differ only by their last two letters, it follows that s_k^2 is a prefix of $s_{k+1}s_k$ if $k \geq 2$. As s_k is a prefix of s_{k+1} when $k \geq 0$, the word $s_{k,l} = s_{k-1}^l s_{k-2}$ is both a prefix and a suffix of $s_k = s_{k-1}^{a_k} s_{k-2}$ for all $k \geq 2$ and l such that $0 < l \leq a_k$. Thus s_k^2 contains $s_{k,l}^2$. The claim follows. □

As was seen in the proof of Proposition 4.1, the index of a factor w of length n depends only on the maximum $r \geq 0$ such that $R^{-tn}(x) \in [w]$ for $0 \leq t \leq r$ where x is either of the endpoints of $[w]$. That is, the index of a factor depends only on the length of its interval but not on its position. To put it more precisely, if w is a factor of length n, then

$$\operatorname{ind}(w) = \gamma + \left\lfloor \frac{|[w]|}{\| |w| \alpha \|} \right\rfloor \tag{7}$$

where γ is 1 if $|[w]| \neq \| |w| \alpha \|$ and 0 otherwise. Next we will carefully characterize the lengths of the intervals of factors of length $q_{k,l}$. After this it is easy to conclude the main results.

Theorem 4.3. *Let $n = q_{k,l}$ with $k \geq 2$ and $0 < l \leq a_k$. Then $C^i(\widetilde{s}_{k,l}) \in \mathcal{L}(\alpha)$ for $0 \leq i \leq n-1$. The intervals of the first $q_{k-1} - 1$ conjugates of $\widetilde{s}_{k,l}$ have length $\|q_{k,l-1}\alpha\|$, and the intervals of the latter $n + 1 - q_{k-1}$ conjugates have length $\|q_{k-1}\alpha\|$. The interval of the remaining factor has length $\|q_{k,l}\alpha\|$.*

Proof. The geometric ideas of this proof are illustrated in the example following this proof. The intervals of the factors of length n are called in this proof *level n intervals*. With the same effort we prove here more than what is claimed above; we give the exact positions of the intervals of the conjugates of $\widetilde{s}_{k,l}$ on the circle.

By Proposition 2.2 the interval $J = I(-q_{k,l-1}\alpha, 0)$ has exactly one point $\{-t\alpha\}$ with $0 < t \leq n$ as an interior point; namely the point $\{-n\alpha\}$. That is, the point $\{-n\alpha\}$ split the level $n-1$ interval J into the level n intervals $K = I(-q_{k,l-1}\alpha, -n\alpha)$ and $L = I(-n\alpha, 0)$. Observe that $\|q_{k,l-1}\alpha\| = |J| = |K| + |L| = \|q_{k-1}\alpha\| + \|q_{k,l}\alpha\|$. The Three Distance Theorem tells that level n intervals have lengths $\|q_{k,l-1}\alpha\|$, $\|q_{k,l}\alpha\|$, and $\|q_{k-1}\alpha\|$. In particular, interval L is the unique level n interval of length $\|q_{k,l}\alpha\|$. Let i be the smallest positive integer such that the interval $R^{-i}(J)$ is not any interval of level n. The interval $R^{-i}(J)$ must be a union of two level n intervals: one having length $|K|$ and the other having length $|L|$. This is true as by (3) it can be deduced that $|J|$ is never a multiple of $|K|$; further, $|K|$ is never a multiple of $|L|$. Since the interval of length $\|q_{k,l}\alpha\|$ is unique, we conclude that the other interval in the union is L. As α is irrational, $R^{-i}(J) \neq J$, so it must be that $R^{-i}(J) = M \cup L$ where $M = I(-q_{k-1}\alpha, 0)$. Therefore R^{-i} maps the endpoint 0 of L to the endpoint $\{-q_{k-1}\alpha\}$ of M, so $i = q_{k-1}$. As $k > 1$, also $i > 1$. We have shown that the level n intervals $R^{-1}(J), R^{-2}(J), \ldots, R^{-(i-1)}(J)$ have length $|J| = \|q_{k-1,l}\alpha\|$. By the Three Distance Theorem the remaining $n + 1 - q_{k-1}$ intervals excluding L have length $\|q_{k-1}\alpha\|$.

What remains is to analyze the connection between rotation and conjugation. Let u and v be factors of length n such that $[u] = M$ and $[v] = L$. Since the intervals M and L are on the opposite sides of 0, we have that $u = au'$ and $v = bv'$ for distinct letters a and b. Let $x \in M$ and

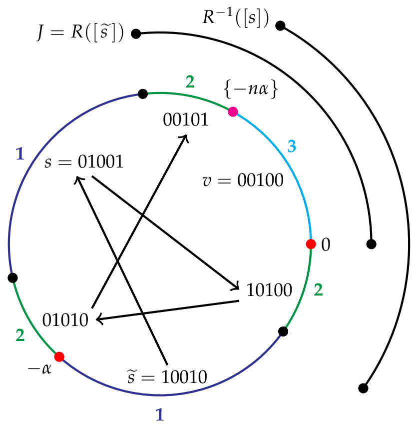

Figure 1: An example of the geometric ideas in the proof of Theorem 4.3.

$y \in L$. Since $i > 1$, the interval $R^{-(i-1)}(J) = R(M \cup L)$ is the interval of some factor w of length n. Therefore the Sturmian words $s_{x+\alpha,\alpha}$ and $s_{y+\alpha,\alpha}$ both have w as a prefix. Thus $s_{x,\alpha}$ begins with aw and $s_{y,\alpha}$ begins with bw. Hence w must be left special, that is, $w = s_{k,l}$. We will show next that v is not conjugate to $s_{k,l}$. Note that $k-1$ is odd if and only if $\{-q_{k-1}\alpha\} \in I_0$. Hence the first letter of v is 0 if and only if $k-1$ is odd. On the other hand, the last letter of $s_{k,l}$ is 0 if and only if $k-1$ is even. Thus we conclude that the first letter of v is distinct from the last letter of $s_{k,l}$. However, as the suffix of v of length $n-1$ is a prefix of $s_{k,l}$, we see that there are more letters b in v than there are in $s_{k,l}$, so v and $s_{k,l}$ cannot be conjugate.

Let then z be the factor of length n such that $[z] = R^{-1}(J)$. Since $\{-n\alpha\} \in J$, it must be that $\{-(n+1)\alpha\} \in R^{-1}(J) = [z]$. Thus z is right special, that is, $z = \widetilde{s}_{k,l}$. By Lemma 4.2 $s_{k,l}^2 \in \mathcal{L}(\alpha)$. Thus every conjugate of $s_{k,l}$ is a factor. Further, by the mirror-invariance of $\mathcal{L}(\alpha)$, we see that $s_{k,l}$ and $\widetilde{s}_{k,l}$ are conjugates. Moreover, every conjugate of $\widetilde{s}_{k,l}$ is extended to the left by its last letter.

Suppose that $\lambda \neq v$ is a factor of length n such that $R^{-1}([\lambda])$ is the interval of some factor μ of length n. As $R^{-1}([s_{k,l}])$ does not satisfy this condition, it follows that $\lambda \neq s_{k,l}$, so λ extends to the left uniquely. We will prove that $C(\lambda) = \mu$. Write $\lambda = \lambda' c$ for some letter c. Then obviously $\mu = d\lambda'$ for some letter d. By definition μ must be followed by the letter c, that is, $\mu c = d\lambda' c = d\lambda \in \mathcal{L}(\alpha)$. We have that $d = c$ because λ is uniquely extended to the left by its last letter. Therefore we conclude that $C(\lambda) = \mu$. In this way we see that the factors of length n having the intervals $R^{-1}(J), R^{-2}(J), \ldots, R^{-(i-1)}(J)$ correspond (in order) to the factors $\widetilde{s}_{k,l}, C(\widetilde{s}_{k,l}), \ldots, C^{q_{k-1}-2}(\widetilde{s}_{k,l}) = s_{k,l}$. We saw above that v is not conjugate to $s_{k,l}$, so it must be that $C(s_{k,l}) = u$. Thus the factors of length n having the intervals $[u] = L, R^{-1}(L), R^{-2}(L), \ldots, R^{-(n-q_{k-1})}(L)$ correspond (in order) to the factors $u, C(u), C^2(u), \ldots, C^{n-q_{k-1}}(u)$. As $u = C(s_{k,l}) = C^{q_{k-1}-1}(\widetilde{s}_{k,l})$, we have a complete description of the positions of the intervals of conjugates of $\widetilde{s}_{k,l}$ using the backward orbit of J under R. □

Example 4.4. Let $\alpha = [0; \overline{2,1}] = \frac{1}{2}(\sqrt{3}-1)$ (that is, the continued fraction expansion of α has period $2,1$). Consider the semiconvergent $\frac{p_{3,1}}{q_{3,1}} = \frac{1+1}{3+2} = \frac{2}{5}$ of α and factors of length $n = 5$. The factors of length n are $00100, 00101, 01001, 01010, 10010$, and 10100. Their intervals are depicted in Figure 1. There are intervals of type 1, 2, and 3 depending on their length. Intervals of type 1 have length $\|2\alpha\|$, intervals of type 2 have length $\|3\alpha\|$, and the interval of type 3 has length $\|2\alpha\| - \|3\alpha\| = \|5\alpha\|$. As in the proof of Theorem 4.3, the point $\{-n\alpha\}$ has split the type 1 interval $J = I(0, -2\alpha)$ into intervals of type 2 and 3. The interval $R^{-1}(J)$ corresponds to the right special factor $\widetilde{s} = \widetilde{s}_{3,1}$. The arrows in the figure indicate how conjugation acts on $\widetilde{s}$. The backward orbit of J corresponds to conjugates of $\widetilde{s}$ of type 1 until the interval of the left special factor s is encountered. As seen in the proof of Theorem 4.3, the interval $R^{-1}([s])$ no longer coincides with any interval of length n. Here $R^{-1}([s]) = L \cup M = I(0, -n\alpha) \cup I(0, -3\alpha)$ just as the proof requires. The factor having interval L is here seen to be not conjugate to $\widetilde{s}$ as it should be by the proof. The interval M must then correspond to the conjugate of s. As in the proof, the rest of the conjugates of $\widetilde{s}$ are obtained by rotating M backwards. The intervals obtained this way are of type 2.

We are now ready to prove the main result. The result was originally proven by Damanik and Lenz [6]. We present it here phrased in a different way.

Theorem 4.5. *Consider indices of factors of length $n > 0$ in $\mathcal{L}(\alpha)$, and let $k \geq 2$.*

(i) *If $n < q_1$, then the index of the conjugates of $0^{n-1}1$ is 1, and the index of the remaining factor 0^n is $\lfloor a_1/n \rfloor$.*

(ii) *If $n = q_1$, then the index of the conjugates of $\widetilde{s}_1$ is $a_2 + 1$, and the index of the remaining factor 0^{a_1} is 1.*

(iii) *If $n = q_k$, then the index of any of the first $q_{k-1} - 1$ conjugates of $\widetilde{s}_k$ is $a_{k+1} + 2$, the index of any of the remaining $n + 1 - q_{k-1}$ conjugates is $a_{k+1} + 1$, and the index of the remaining factor is 1.*

(iv) *If $n = q_{k,l}$ with $0 < l < a_k$, then the index of the first $q_{k-1} - 1$ conjugates of $\widetilde{s}_{k,l}$ is 2, and the index of the remaining factors is 1.*

(v) *If $n = mq_1$ with $1 < m < a_2 + 1$, then the index of any of the first q_1 conjugates of $\widetilde{s}_1^m$ is $\lfloor (a_2 + 1)/m \rfloor$, and the index of any remaining factor is 1.*

(vi) *If $n = mq_k$ with $1 < m < a_{k+1} + 2$, then the index of any of the first $q_{k-1} - 1$ conjugates of $\widetilde{s}_k^m$ is $\lfloor (a_{k+1} + 2)/m \rfloor$, the index of any of the next $q_k + 1 - q_{k-1}$ conjugates is $\lfloor (a_{k+1} + 1)/m \rfloor$, and the index of any remaining factor is 1.*

(vii) *If n does not fall into any of the above cases, then the index of every factor of length n is 1.*

Proof. First of all, observe that all the cases *(i)*–*(vii)* are mutually exclusive. Consider the cases *(i)* and *(ii)*. The factors of length $n \leq q_1$ are readily seen to be 0^n and the conjugates of $0^{n-1}1$. As the index of 0 is a_1, the index of the factor 0^n is $\lfloor a_1/n \rfloor$. The intervals of the conjugates of $0^{n-1}1$ have length α. If $n = 1$, then the index of the factor $0^{n-1}1 = 1$ is 1. If $n > 1$, the number $1 + \lfloor \alpha/\|n\alpha\| \rfloor$ equals 1 unless $n = q_1$ when it equals $a_2 + 1$ by (6). The claims of *(i)* and *(ii)* follow from (7).

Consider next a factor w of length $n = q_{k,l}$ for some $k \geq 2$ and l such that $0 < l \leq a_k$. By Theorem 4.3, the intervals of the first $q_{k-1} - 1$ conjugates of $\widetilde{s}_{k,l}$ have length $\|q_{k,l-1}\alpha\|$. Using (7) and (3) we see that their index equals to

$$1 + \left\lfloor \frac{\|q_{k,l-1}\alpha\|}{\|q_{k,l}\alpha\|} \right\rfloor = 1 + \left\lfloor \frac{\|q_{k,l}\alpha\| + \|q_{k-1}\alpha\|}{\|q_{k,l}\alpha\|} \right\rfloor = 2 + \left\lfloor \frac{\|q_{k-1}\alpha\|}{\|q_{k,l}\alpha\|} \right\rfloor.$$

If $l \neq a_k$, then by (4) $\|q_{k,l}\alpha\| > \|q_{k-1}\alpha\|$, so the index is 2. If $l = a_k$, then by (6), the index equals to $2 + a_{k+1}$. This proves the first claims in *(iii)* and *(iv)*. The latter cases are analogous, so *(iii)* and *(iv)* are proved.

Proposition 4.1 shows that the factors not covered by the cases *(i)*–*(iv)* having index higher than 1 must be nonprimitive. By *(i)* and *(iv)* they must have length mq_k for some $k \geq 1$, meaning that we are in either of the cases *(v)* or *(vi)*. It is a straightforward application of *(ii)* and *(iii)* to deduce *(v)* and *(vi)*. The theorem is proved. □

In particular, every Sturmian word contains infinitely many cubes, but fourth powers are avoidable. The theorem implies the following weaker version which is still useful (compare to [5, Lemma 3.6]):

Corollary 4.6. *Let $w \in \mathcal{L}(\alpha)$ be primitive. If $w^2 \in \mathcal{L}(\alpha)$, then w is conjugate to s_k for some $k \geq 0$ or to $s_{k,l}$ with $k \geq 2$ and $0 < l < a_k$. If $w^3 \in \mathcal{L}(\alpha)$, then either $w = 0$ and $a_1 > 2$ or w is conjugate to some s_k with $k \geq 1$.* □

We obtain the result of [5], [3], and [7] on the fractional index of Sturmian words as a direct consequence of the results so far:

Theorem 4.7. *The fractional index of a Sturmian word with slope α is*

$$\max\left\{a_1, 2+\sup_{k\geq 2}\{a_k+(q_{k-1}-2)/q_k\}\right\}.$$

Proof. The largest fractional power of a factor with length less than q_1 is clearly 0^{a_1}. Therefore by Theorem 4.5 it is sufficient to analyze the largest fractional power of a (primitive) factor of length q_k for $k \geq 1$. By Theorem 4.5 the index $a_{k+1}+2$ of the first $q_{k-1}-1$ conjugates of $\widetilde{s}_k$ dominates the index of the rest of the factors of length q_k. The fractional part of the fractional index of a factor w is determined by the shortest extension of w to a right special factor. Note that from the proof of Theorem 4.3 it is evident that $C^{q_{k-1}-2}(\widetilde{s}_k) = s_k$. Thus among the first $q_{k-1}-1$ conjugates of $\widetilde{s}_k$, the factor s_k has longest extension to a right special factor, and the length of the extension is $q_{k-1}-2$. Thus the fractional index of s_k is $a_{k+1}+2+(q_{k-1}-2)/q_k$. The claim follows. □

In particular, this theorem says that that a Sturmian word has bounded fractional index if and only if the partial quotients of its slope are bounded. This is a result of Mignosi [11]. An alternative proof was given by Berstel [2].

5 Acknowledgments

The author was supported by University of Turku Graduate School UTUGS Matti programme.

We thank Markus Whiteland, Tero Harju, and Luca Zamboni for giving fruitful feedback on the draft version of this paper.

References

[1] P. Alessandri and V. Berthé. Three distance theorems and combinatorics on words. *Enseign. Math.* 44 (1998), 103–132.

[2] J. Berstel. On the index of Sturmian words. In: *Jewels Are Forever*. Springer-Verlag, 1999, pp. 287–294.

[3] A. Carpi and A. de Luca. Special factors, periodicity, and an application to Sturmian words. *Acta Inf.* 36 (2000), 983–1006. DOI: 10.1007/PL00013299.

[4] J. W. S. Cassels. *An Introduction to Diophantine Approximation*. Cambridge Tracts in Mathematics and Mathematical Physics 45. Cambridge University Press, 1957.

[5] D. Damanik and D. Lenz. The index of Sturmian sequences. *Eur. J. Combin.* 23 (2002), 23–29. DOI: 10.1006/eujc.2000.0496.

[6] D. Damanik and D. Lenz. Powers in Sturmian sequences. *Eur. J. Combin.* 24 (2003), 377–390. DOI: 10.1016/S0195-6698(03)00026-X.

[7] J. Justin and G. Pirillo. Fractional powers in Sturmian words. *Theoret. Comput. Sci.* 255 (2001), 363–376. DOI: 10.1016/S0304-3975(99)90294-3.

[8] A. Y. Khinchin. *Continued Fractions*. Dover Publications, Mineola, New York, 1997.

[9] M. Lothaire. *Combinatorics on Words*. Encyclopedia of Mathematics and Its Applications 17. Addison-Wesley, 1983.

[10] M. Lothaire. *Algebraic Combinatorics on Words*. Encyclopedia of Mathematics and Its Applications 90. Cambridge University Press, 2002.

[11] F. Mignosi. Infinite words with linear subword complexity. *Theoret. Comput. Sci.* 65 (1989), 221–242. DOI: 10.1016/0304-3975(89)90046-7.

[12] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*. Lecture Notes in Mathematics 1794. Springer, 2002. DOI: 10.1007/b13861.

[13] D. Vandeth. Sturmian words and words with a critical exponent. *Theoret. Comput. Sci.* 242 (2000), 283–300. DOI: 10.1016/S0304-3975(98)00227-8.