Maximal pattern complexity for discrete systems

Ergodic Theory and Dynamical Systems 22-4 (2002), pp.1201-1214

Teturo Kamae* and Luca Zamboni†

Abstract

For an infinite word $\alpha = \alpha_0\alpha_1\alpha_2\cdots$ over a finite alphabet, the authors introduced a new notion of complexity called maximal pattern complexity defined by

$$p^*_\alpha(k) := \sup_\tau \sharp\{\alpha_{n+\tau(0)}\alpha_{n+\tau(1)}\cdots\alpha_{n+\tau(k-1)};\ n = 0, 1, 2, \cdots\}$$

where the supremum is taken over all sequences of integers $0 = \tau(0) < \tau(1) < \cdots < \tau(k-1)$ of length k. The authors proved that α is aperiodic if and only if $p^*_\alpha(k) \geq 2k$ for every $k = 1, 2, \cdots$. A word α with $p^*_\alpha(k) = 2k$ for every $k \geq 1$ is called pattern Sturmian. In this paper, we give a simple criterion to be pattern Sturmian and exhibit a new class of recurrent pattern Sturmian words which do not arise from rotations. We also investigate the maximal pattern complexity of various discrete dynamical systems including irrational rotations on the circle, and self-similar systems generated by substitutions. We show that for each irrational rotation on the circle, there exists a twofold partition of the circle, with respect to which the system generated has full maximal pattern

*Department of Mathematics, Osaka City University, Osaka, 558-8585 Japan (kamae@sci.osaka-cu.ac.jp)

†Department of Mathematics, P.O. Box 311430, University of North Texas, Denton, TX 76203-1430, USA (luca@unt.edu)

complexity with probability 1. Using the arithmetic properties of the underlying numeration system associated to a substitution dynamical system, we prove that the maximal pattern complexity of the fixed point of the Rauzy substitution $1 \mapsto 12\ 2 \mapsto 13\ 3 \mapsto 1$ has exponential growth. It is well known that the system generated by the Rauzy substitution is isomorphic in measure to an irrational rotation on the 2-torus.

1 Introduction

Let $\alpha = \alpha_0\alpha_1\alpha_2 \ldots \in A^{\mathbb{N}}$ be an infinite word over a finite alphabet A with $\sharp A \geq 2$, where $\mathbb{N} = \{0, 1, 2, \cdots\}$ and $\sharp A$ denotes the number of elements in A.

For $k \geq 1$, let $F_\alpha(k)$ denote the set of all *factors* of α of length k, that is

$$F_\alpha(k) = \{\alpha_n\alpha_{n+1} \cdots \alpha_{n+k-1};\ n \in \mathbb{N}\},$$

and set $F_\alpha(0) = \{\varepsilon\}$ where ε denotes the *empty word*, the unique word of length zero. Set

$$F(\alpha) = \bigcup_{k \geq 0} F_\alpha(k).$$

The *block complexity function* $p_\alpha : \mathbb{N} \to \mathbb{N}$ is defined as $p_\alpha(k) := \sharp F_\alpha(k)$. A fundamental result due to Hedlund and Morse states that a word α is eventually periodic if and only if for some k the block complexity $p_\alpha(k) \leq k$. (See [8]).

Infinite words α such that $p_\alpha(k) = k+1$ $(k = 0, 1, 2, \cdots)$ are called *Sturmian sequences* or *Sturmian words.* The best known example is the so-called Fibonacci word

$$1211212112112121121211211212112112121121211211212112121121 \ldots$$

fixed by the substitution $1 \mapsto 12$ and $2 \mapsto 1$. It is well known that all Sturmian words can be realized geometrically by an irrational rotation on the circle (see [5, 8]). More precisely, every Sturmian word is obtained by coding the orbit of a point x on the circle (of circumference one) under a rotation by an irrational angle θ where the circle is partitioned into two complementary intervals, one of length θ and the other of length $1-\theta$. And conversely every such coding gives rise to a Sturmian word.

Let k be a positive integer. By a *k-window* τ, we mean a sequence of integers of length k with

$$0=\tau(0)<\tau(1)<\tau(2)<\cdots<\tau(k-1).$$

A $k+1$-window τ' is called an *immediate extension* of the above τ if $\tau'(i)=\tau(i)$ $(i=0,1,\cdots,k-1)$. In this case, we also call τ the *immediate restriction* of τ'. Let $\alpha=\alpha_0\alpha_1\alpha_2\cdots$ be an infinite word over a finite alphabet.

For each k-window $0=\tau(0)<\tau(1)<\cdots<\tau(k-1)$ put

$$F_\alpha(\tau)=\{\alpha_{n+\tau(0)}\alpha_{n+\tau(1)}\cdots\alpha_{n+\tau(k-1)};\ n=0,1,2,\cdots\}.$$

The *maximal pattern complexity function* $p_\alpha^*:\{1,2,3,\cdots\}\to\{1,2,3,\cdots\}$ introduced by the authors in a recent paper [6], is defined by

$$p_\alpha^*(k):=\sup_\tau \sharp F_\alpha(\tau)\quad(k=1,2,3,\cdots),$$

where the supremum is taken over all k-windows τ. A k-window τ is said to *attain* $p_\alpha^*(k)$ if $\sharp F_\alpha(\tau)=p_\alpha^*(k)$.

This notion of complexity is intimately related to sequence entropy (see [6]). In fact, in [6] it is shown that any infinite word α which induces a dynamical system with a partially continuous spectrum satisfies

$$\limsup_{k\to\infty}(1/k)\log p_\alpha^*(k)>0.$$

As in the case of block complexity, maximal pattern complexity also gives a characterization of eventually periodic words:

Theorem 1 ([6]). *An infinite word α is eventually periodic if and only if for some k the maximal pattern complexity $p_\alpha^*(k)\le 2k-1$.*

Infinite words α of maximal pattern complexity $p_\alpha^*(k)=2k$ $(k=1,2,3,\cdots)$ are called *pattern Sturmian* words. Thus amongst all aperiodic words, pattern Sturmian words are those of lowest maximal pattern complexity. If α is pattern Sturmian, then α is a binary word since $p_\alpha^*(1) = 2$.

It is proved in [6] that for any irrational θ with $0 < \theta < 1$, any interval $I \subset \mathbb{R}$ with $0 < |I| < 1$ and any $x \in [0,1)$, $p_\alpha^*(k) = 2k$ ($k = 1, 2, \cdots$) for $\alpha = \mathcal{R}(\theta, I, x, \mathbb{Z})$, where

$$\mathcal{R}(\theta, I, x, \mathbb{Z})_n = \begin{cases} 0 & \text{if } x + n\theta \in I \pmod{\mathbb{Z}} \\ 1 & \text{otherwise} \end{cases} \tag{1.1}$$

In particular, every Sturmian word is pattern Sturmian, but not conversely, since if in the above $|I| \notin \{\theta, 1-\theta\}$, then α is not Sturmian (see [11]).

In this paper, we begin by giving a new proof that every Sturmian word is pattern Sturmian. Our proof differs from the original proof given in [6] in that it is purely combinatorial in nature, and relies only on the *balanced* property of Sturmian words (Section 2).

In Section 3, we give a simple criterion to be pattern Sturmian and apply it in Section 4 to exhibit a new class of recurrent pattern Sturmian words which do not arise from a rotation. These words are special kinds of so called Toeplitz words and the dynamical systems induced by them have rational discrete spectrums.

Sections 5 and 6 are devoted to the study of the maximal pattern complexity of various discrete dynamical systems: In Section 5 we show that for each irrational rotation on the circle, there exists a twofold partition of the circle, with respect to which the system generated has full maximal pattern complexity with probability 1. In Section 6, we investigate the maximal pattern complexity of the self-similar dynamical system generated by the Rauzy substitution ([9], also called the Tribonacci substitution):

$$\begin{array}{ccc} 1 & \to & 12 \\ 2 & \to & 13 \\ 3 & \to & 1 \end{array} \tag{1.2}$$

Rauzy showed that the subshift (X, T) generated by this substitution is a natural coding of a rotation on the torus T^2, i.e., is measure-theoretically conjugate to an exchange of three fractal domains on a compact set in $\mathbb{R}^2$, each domain being translated by the same vector modulo a lattice. The fixed point β of the Rauzy substitution of (1.2) has block complexity $p_\beta(k) = 2k + 1$ and is just one example of a broader class of sequences of complexity $2k+1$ originally studied by Arnoux and Rauzy in [2] now called *Arnoux-Rauzy sequences.* From the point of view of block complexity, this β is the simplest natural generalization of Sturmian sequences to a 3-letter alphabet, with the Tribonacci substitution, the analogue of the Fibonacci substitution (see also [10, 7]), and is called the *tribonacci word.* However from the point of view of maximal pattern complexity, Sturmian words and the tribonacci word exhibit a drastically different behavior. While Sturmian words are pattern Sturmian, we show that the maximal pattern complexity of β has exponential growth (see Theorem 6). Our proof relies on the arithmetic properties of the underlying numeration system associated to a substitution dynamical system.

Acknowledgements

The authors wish to thank the Institut de Mathématiques de Luminy for their hospitality and support during the time that much of this work was carried out. We are also grateful to P. Arnoux, V. Berthé, J. Cassaigne and S. Ferenczi for numerousdiscussions and exchanges, and to the anonymous referee for his (or her) useful suggestions for the terminology and notation.

2 A new proof that Sturmian implies pattern Sturmian

In this section we give a combinatorial proof of the following result first proved in [6] using geometric arguments:

Theorem 2. *Let* $\alpha = \alpha_0\alpha_1\alpha_2\cdots \in \{0,1\}^{\mathbb{N}}$ *be a Sturmian word. Then* α *is pattern Sturmian.*

Proof. Since α is aperiodic, by Theorem 1 we have $p^*_\alpha(k) \geq 2k$ $(k = 1, 2, 3, \cdots)$. To prove that α is pattern Sturmian, it is sufficient to prove that $p^*_\alpha(k) \leq 2k$ $(k = 1, 2, 3, \cdots)$. Take an arbitrary integer $k \geq 1$ and a k-window $\tau : 0 = \tau(0) < \tau(1) < \cdots < \tau(k-1)$. We will show that $\sharp F(\tau) \leq 2k$.

Consider the set S of words $c = c_0 c_1 c_2 \cdots c_{2k-2}$ on $\mathbb{N}^{2k-1}$ such that there exists $n \in \mathbb{N}$ satisfying that

$$c_{2i} = \alpha_{n+\tau(i)} \quad (i = 0, 1, \cdots, k-1)$$

and

$$c_{2i+1} = \sum_{i=n+\tau(i)+1}^{n+\tau(i+1)-1} \alpha_i \quad (i = 0, 1, \cdots, k-2).$$

Let $\psi : S \to \mathbb{N}^k$ be the mapping defined by

$$\psi(c) = c_0 c_2 c_4 \cdots c_{2k-2}$$

for any $c \in S$. Then since $\psi(S) = F_\alpha(\tau)$, we have $\sharp F(\tau) \leq \sharp S$. Therefore, it is sufficient to prove that $\sharp S \leq 2k$.

Since α is balanced, it holds that

$$|(c_i + c_{i+1} + \cdots + c_j) - (d_i + d_{i+1} + \cdots + d_j)| \leq 1 \tag{2.1}$$

for any $c, d \in S$ and i, j with $0 \leq i \leq j \leq 2k-2$.

For $i = 0, 1, \cdots, 2k-1$, let $\pi_i : S \to \mathbb{N}^{i+1}$ be the mapping defined by

$$\pi_i(c) = c_0 c_1 c_2 \cdots c_i$$

for any $c \in S$. Then we have $\sharp \pi_0(S) = \sharp\{0, 1\} = 2$. Therefore to prove $\sharp S \leq 2k$, it is sufficient to prove that

$$\sharp \pi_{i+1}(S) \leq \sharp \pi_i(S) + 1 \tag{2.2}$$

for any $i = 0, 1, 2, \cdots, 2k-3$ since $S = \pi_{2k-2}(S)$.

Suppose to the contrary that

$$\sharp \pi_{i+1}(S) \geq \sharp \pi_i(S) + 2 \tag{2.3}$$

holds for some $i = 0, 1, 2, \cdots, 2k-3$. By (2.1), any $c_0 c_1 \cdots c_i$ in $\pi_i(S)$ has at most 2 extensions in $\pi_{i+1}(S)$, that is, there exists at most 2 elements $u \in \mathbb{N}$ such that $c_0c_1 \cdots c_iu \in \pi_{i+1}(S)$. Hence by (2.3), there exist at least 2 elements, say $c_0c_1 \cdots c_i$ and $d_0d_1 \cdots d_i$ having 2 extensions in $\pi_{i+1}(S)$. That is, there exist $u, v \in \mathbb{N}$ with $u \neq v$ such that all of $c_0c_1 \cdots c_iu$, $c_0c_1 \cdots c_iv$, $d_0d_1 \cdots d_iu$, $d_0d_1 \cdots d_iv$ are in $\pi_{i+1}(S)$. Let h be the maximum integer such that $h \leq i$ and $c_h \neq d_h$. Without loss of generality, we assume that $u - v \geq 1$ and $c_h - d_h \geq 1$. Then we have

$$(c_h + c_{h+1} + \cdots + c_i + u) - (d_h + d_{h+1} + \cdots + d_i + v) \geq 2,$$

which contradicts (2.1).

Thus we have (2.2) as required. □

3 A criterion for pattern Sturmian words

In this section we establish the following criterion for pattern Sturmian words :

Theorem 3. *An aperiodic binary infinite word α is pattern Sturmian if $\sharp F_\alpha(\tau') \leq \sharp F_\alpha(\tau) + 2$ holds for any 3-window τ', where τ is the immediate restriction of τ'.*

Proof. Assume that α satisfies the conditions in Theorem 3. Then, $p^*_\alpha(1) = 2$ and $p^*_\alpha(2) = 4$ holds by Theorem 1.

Suppose that there exists $k \geq 2$ such that $p^*_\alpha(k+1) \geq p^*_\alpha(k) + 3$. There exits a $k+1$-window τ' which attains $p^*_\alpha(k+1)$. Let τ be the immediate restriction of τ'. Then we have

$$\sharp F_\alpha(\tau) + 3 \leq \sharp F_\alpha(\tau') \tag{3.1}$$

since

$$\sharp F_\alpha(\tau) + 3 \leq p^*_\alpha(k) + 3 \leq p^*_\alpha(k+1) = \sharp F_\alpha(\tau').$$

Take the minimum k such that there exists a $k+1$-window τ' satifying (3.1), where τ denotes the immediate restriction of τ'. By the above argument, such k exists, while $k \geq 3$ holds by our assumptions.

Since α is a word over 2 letters, say over $\{0, 1\}$, and (3.1) holds, there exists at least 3 different words, say $u^1, u^2, u^3 \in F_\alpha(\tau)$ which have 2 extensions in $F_\alpha(\tau')$.

Any of 0 and 1 is in the set of last letters of u^1, u^2, u^3, since otherwise, we can write $u^1 = u^4 a, u^2 = u^5 a, u^3 = u^6 a$ with $a \in \{0, 1\}$ and 3 different words u^4, u^5, u^6 of length $k-1$, so that the k-window $\tau(0) < \tau(1) < \cdots < \tau(k-2) < \tau'(k)$ satisfies (3.1) contradidting with the minimality of k. Hence, without loss of generality we may assume that $u^1 = u^7 0, u^2 = u^8 0, u^3 = u^9 1$ with some words u^7, u^8, u^9 of length $k-1$. Since $u^1 \neq u^2$, we have $u^7 \neq u^8$. Therefore, there exists a subscript i with $0 \leq i \leq k-2$ such that $u_i^7 \neq u_i^8$. Then, the 3-window η' with $\eta'(0) = 0$, $\eta'(1) = \tau(k-1) - \tau(i)$, $\eta'(2) = \tau'(k) - \tau(i)$ satisfies (3.1), contradicting the assumption in Theorem 3.

Therefore, we have $p_\alpha^*(k+1) \leq p_\alpha^*(k) + 2$ for any $k \geq 2$. This implies that $p_\alpha^*(k) = 2k$ for $k = 1, 2, 3, \cdots$ by Theorem 1. Thus, α is pattern Sturmian as required. $\square$

Remark 1. *The converse of* Theorem 3 *is not true. Consider a pattern Sturmian word α such that $\alpha = \mathcal{R}(\theta, I, 0, \mathbb{Z})$ with an irrational θ and an interval I with $0 < |I| < 1/3$. Take an integer $n > 0$ such that the fractional part $\{n\theta\}$ satisfies that $|I| < \{n\theta\} < 2|I|$. Let τ be the 2-window with $\tau(0) = 0$ and $\tau(1) = n$. Since $x \in I$ implies that $x + n\theta \notin I \pmod{\mathbb{Z}}$, we have $F(\tau) = \{01, 10, 11\}$. Let m be an integer such that $m > n$ and $\{n\theta\} - |I| < \{m\theta\} < |I|$. Then, For the 3-window τ' which is an immediate extension of τ with $\tau'(2) = m$, we have $F(\tau') = \{010, 011, 100, 101, 110, 111\}$. Thus, we have $\sharp F_\alpha(\tau) + 3 = \sharp F_\alpha(\tau')$ for a pattern Sturmian word α.*

4 A class of recurrent pattern Sturmian words not arising from rotations

In this section, we apply Theorem 3 to obtain a new class of recurrent pattern Sturmian words which do not arise from rotations.

Definition 1. For $a \in \{0, 1\}$ and integers l, r with $l \geq 2$ and $0 \leq r \leq l-1$, let

$$\beta^{(a,l,r)} := (a^r ? a^{l-1-r})(a^r ? a^{l-1-r}) \cdots$$

be a periodic word with period l over 2 letters $\{a, ?\}$, where we denote $a^r = \overbrace{a \cdots a}^{r}$. We define $\beta^{(a,l,r)} \triangleleft \beta^{(b,m,s)} \in \{0, 1, ?\}^{\mathbb{N}}$ by replacing each occurrence of "?" in $\beta^{(a,l,r)}$ by the letters in $\beta^{(b,m,s)}$ one by one in the order. That is,

$$(a^r ba^{l-1-r})^s (a^r ? a^{l-1-r})(a^r ba^{l-1-r})^{m-1-s} \cdots \ .$$

An infinite word $\alpha = \alpha_0\alpha_1\alpha_2 \ldots$ over $\{0, 1\}$ is called a *simple Toeplitz word* if there exists an infinite word

$$(a_0, l_0, r_0)(a_1, l_1, r_1)(a_2, l_2, r_2) \cdots$$

satisfying that
(1) $a_i \in \{0, 1\}$ $(i = 0, 1, 2, \cdots)$ with both of $a_i = 0$ and $a_i = 1$ infinitely often,
(2) $l_i \geq 2$ $(i = 0, 1, 2, \cdots)$, and
(3) $0 \leq r_i \leq l_i - 1$ $(i = 0, 1, 2, \cdots)$ with $r_i \geq 1$ infinitely often
so that

$$\alpha = \beta^{(a_0,l_0,r_0)} \triangleleft \beta^{(a_1,l_1,r_1)} \triangleleft \beta^{(a_2,l_2,r_2)} \triangleleft \cdots \ .$$

We call $(a_0, l_0, r_0)(a_1, l_1, r_1)(a_2, l_2, r_2) \cdots$ the *coding word* of α.

It is not difficult to check that the simple Toeplitz word α with coding word $(a_0, l_0, r_0)(a_1, l_1, r_1)(a_2, l_2, r_2) \cdots$ can be written as an infinite composition of the form

$$\alpha = \lim_{n\to\infty} \rho_{a_0,l_0,r_0} \circ \rho_{a_1,l_1,r_1} \circ \cdots \circ \rho_{a_n,l_n,r_n}(0),$$

where $\rho_{a,l,r}$ denotes the substitution

$$\rho_{a,l,r} : \quad \begin{array}{lcl} 0 & \to & \overbrace{a \cdots a}^{r} 0 \overbrace{a \cdots a}^{l-1-r} \\ 1 & \to & \overbrace{a \cdots a}^{r} 1 \overbrace{a \cdots a}^{l-1-r} . \end{array} \tag{4.1}$$

For example, let α be the fixed point of the substitution

$$\phi : \quad \begin{array}{lcl} 0 & \to & 01 \\ 1 & \to & 00 \end{array}$$

for which $\phi \circ \phi = \rho_{0,2,1} \circ \rho_{1,2,1}$ holds. Then, α is the simple Toeplitz word with coding word $(0,2,1)(1,2,1)(0,2,1)(1,2,1)\cdots$.

Let $a \in \{0,1\}$ and set $\bar{a} = 1-a$, so that $\bar{0} = 1$ and $\bar{1} = 0$. A simple Toeplitz word α is called of type (a,l,r) or a if the coding word $(a_0,l_0,r_0)(a_1,l_1,r_1)(a_2,l_2,r_2)\cdots$ begin in (a,l,r), i.e. $(a_0,l_0,r_0) = (a,l,r)$. In this case, α can be written as an infinite concatenation of the factors $a^r a a^{l-1-r}$ and $a^r \bar{a} a^{l-1-r}$. In fact, we can write $\alpha = \rho_{a,l,r}(\beta)$ where β is the simple Toeplitz word whose coding word is $(a_1,l_1,r_1)(a_2,l_2,r_2)(a_3,l_3,r_3)\cdots$.

Lemma 1. *Let α be a simple Toeplitz word of type (a,l,r) and let $\alpha = \rho_{a,l,r}(\beta)$. Then, $\alpha_n = \bar{a}$ implies that $n \equiv r \pmod{l}$. Moreover, $\alpha_{ln+r} = \beta_n$ holds for any $n \in \mathbb{N}$.*

Proof. Clear from the definition (4.1) of the substitution $\rho_{a,l,r}$. □

Lemma 2. *Let α be a simple Toeplitz word. Then for any 3-window τ', we have $\sharp F_\alpha(\tau') \leq \sharp F_\alpha(\tau)+2$, where τ is the immediate restriction of τ'.*

Proof. Suppose to the contrary that there exists a simple Toeplitz word α and a 3-window τ' such that

$$\sharp F_\alpha(\tau') \geq \sharp F_\alpha(\tau) + 3, \tag{4.2}$$

where τ is the direct restriction of τ'. Let Θ be the set of all pairs (α, τ') of a simple Toeplitz word α and a 3-window τ' such that
(*) either $\overline{tt}t \in F_\alpha(\tau')$ and $\sharp(F_\alpha(\tau') \setminus \{\overline{tt}t, \overline{ttt}\}) \geq \sharp(F_\alpha(\tau) \setminus \{\overline{tt}\}) + 2$
or (4.2) holds,
where t is the type of α. Take a pair $(\alpha,\tau') \in \Theta$ such that $\tau'(2)$ is minimum among all the pairs in Θ, which exists by (4.2). Let (a,l,r) be the type of α. By (*), there exist 2 elements, say $u_0^1u_1^1u_2^1, u_0^2u_1^2u_2^2$ in $F_\alpha(\tau')$ such that $u_i^1 = u_j^1 = \bar{a}$ and $u_{i'}^2 = u_{j'}^2 = \bar{a}$ with $i<j$, $i'<j'$ and $\{i,j\} \cup \{i',j'\} = \{1,2,3\}$. Hence by Lemma 1, $\tau'(1) \equiv \tau'(2) \equiv 0 \pmod{l}$. Let β be such that $\rho_{a,l,r}(\beta) = \alpha$ and 3-window η' be such that $\eta'(k) = \tau'(k)/l$ $(k=0,1,2)$. Then by Lemma 1, we have $(\beta,\eta') \in \Theta$. Since $\eta'(2) < \tau'(2)$, this contradicts with the minimality of $\tau'(2)$, which completes the proof. □

Theorem 4. *A simple Toeplitz word is a recurrent pattern Sturmian word.*

Proof. By Theorem 3 and Lemma 2, it is sufficient to prove that a simple Toeplitz word is recurrent but not eventually periodic.

Let $(a_0, l_0, r_0)(a_1, l_1, r_1)(a_2, l_2, r_2)\cdots$ be the coding word for α. Let

$$\begin{aligned} L_i &:= l_0 \cdot l_1 \cdots l_i \quad (i = 0, 1, 2, \cdots) \\ R_0 &:= r_0 \\ R_i &:= R_{i-1} + L_{i-1} \cdot r_i \quad (i = 1, 2, 3, \cdots). \end{aligned}$$

Then for any $k \in \mathbb{N}$, we have

$$\beta^{(a_0,l_0,r_0)} \lhd \beta^{(a_1,l_1,r_1)} \lhd \cdots \lhd \beta^{(a_k,l_k,r_k)} =$$
$$(\alpha_0 \cdots \alpha_{R_k-1} ? \alpha_{R_k+1} \cdots \alpha_{L_k-1})(\alpha_0 \cdots \alpha_{R_k-1} ? \alpha_{R_k+1} \cdots \alpha_{L_k-1}) \cdots \ .$$

Therefore for any $k, n \in \mathbb{N}$, either

$$\alpha_{nL_k} \cdots \alpha_{nL_k+L_k-1} = \alpha_0 \cdots \alpha_{R_k-1} 0 \alpha_{R_k+1} \cdots \alpha_{L_k-1}$$

or

$$\alpha_{mL_k} \cdots \alpha_{mL_k+L_k-1} = \alpha_0 \cdots \alpha_{R_k-1} 1 \alpha_{R_k+1} \cdots \alpha_{L_k-1}$$

holds for any $n = 0, 1, 2, \cdots$. Moreover, both of these 2 cases occur infinitely often since $a_i' s$ contain infinitely many 0's and 1's.

This implies first that α is recurrent.

Secondly, this implies that α is aperiodic. Suppose to the contrary that α is eventually periodic with period $d > 0$. Then there exists $N \in \mathbb{N}$ such that $\alpha_{c+d} = \alpha_c$ holds for any $c \geq N$. Since $r_i \geq 1$ occurs infinitely often, we have $\lim_{k\to\infty} R_k = \infty$. Take k such that $d < R_k$. Take $n,\ m$ such that $nL_k \geq N,\ mL_k \geq N$ and

$$\begin{aligned} \alpha_{nL_k} \cdots \alpha_{nL_k+L_k-1} &= \alpha_0 \cdots \alpha_{R_k-1} 0 \alpha_{R_k+1} \cdots \alpha_{L_k-1} \\ \alpha_{nL_k} \cdots \alpha_{nL_k+L_k-1} &= \alpha_0 \cdots \alpha_{R_k-1} 1 \alpha_{R_k+1} \cdots \alpha_{L_k-1} \end{aligned}$$

hold. Let $c := nL_k + R_k - d$ and $c' := mL_k + R_k - d$. Then, we have $\alpha_c = \alpha_{c'} = \alpha_{R_k-d}$. By the eventual periodicity, this implies that

$$\begin{aligned} 0 = \alpha_{nL_k+R_k} &= \alpha_{c+d} = \alpha_c \\ = \quad \alpha_{c'} &= \alpha_{c'+d} = \alpha_{mL_k+R_k} = 1, \end{aligned}$$

which is a contradiction. Thus, α is not eventually periodic. □

5 Irrational rotations having full maximal pattern complexity

In this section we show:

Theorem 5. *For any irrational rotation θ, there exists a closed set S in $[0,1)$ such that $p^*_\alpha(k) = 2^k$ $(k = 1, 2, \cdots)$ for almost all $x \in [0,1)$ with $\alpha = \mathcal{R}(\theta, S, x, \mathbb{Z})$.*

Proof. Let q_i $(i = 1, 2, \cdots)$ be a sequence of positive integers such that $1 =: \rho_0 > \rho_1 > \rho_2 > \cdots > 0$ and $\sum_{i=1}^{\infty} \rho_i/\rho_{i-1} < 1/4$, where $\rho_i := \{q_i\theta\}$, the fractional part of $q_i\theta$ for $i = 1, 2, \cdots$. Let

$$S_0 := \left\{ \sum_{i=1}^{\infty} r_i\rho_i;\ r_i \in [0, \rho_{i-1}/\rho_i - 1) \cap \mathbb{Z} \text{ and } r_i \neq 1 \quad (i = 1, 2, \cdots) \right\}$$

and

$$S_1 := [0,1) \setminus S_0.$$

Then, it holds that S_0 is a closed set in $[0,1)$ with

$$\begin{aligned} \lambda(S_0) &\geq \prod_{i=1}^{\infty}(1 - 2\rho_i/\rho_{i-1}) \\ &\geq 1 - \sum_{i=1}^{\infty} 2\rho_i/\rho_{i-1} > 1/2, \end{aligned}$$

where λ is the Lebesgue measure on $[0,1)$.

Let $\alpha(x) := \mathcal{R}(\theta, S_0, x, \mathbb{Z}) \in \{0,1\}^{\mathbb{N}}$ for $x \in [0,1)$. Let $k = 1, 2, \cdots$. Take any $\xi = (\xi_1, \xi_2, \cdots, \xi_k) \in \{0,1\}^k$. Then, it holds that

$$\Omega_\xi := \bigcap_{i=1}^{k}(S_{\xi_i} - \{q_i\theta\}) \supset$$
$$\left\{ \sum_{i=1}^{\infty} r_i\rho_i;\ r_i \in [0, \frac{\rho_{i-1}}{\rho_i} - 1) \cap \mathbb{Z}\ (i = 1, 2, \cdots),\quad \begin{array}{ll} r_i = 2 - 2\xi_i & (i \leq k) \\ r_i \neq 1 & (i > k) \end{array} \right\}.$$

Therefore, it holds that

$$\begin{aligned}\lambda(\Omega_\xi) &\geq \prod_{i=1}^{k} \rho_i/\rho_{i-1} \prod_{i=k+1}^{\infty} (1-2\rho_i/\rho_{i-1}) \\ &\geq \rho_k(1-\sum_{i=k+1}^{\infty} 2\rho_i/\rho_{i-1}) > \rho_k/2 > 0.\end{aligned}$$

Hence, for almost all $x \in [0,1)$ and for all $\xi \in \bigcup_{k=1}^{\infty}\{0,1\}^k$, there exists $n \in \mathbb{N}$ such that $\{x+n\theta\} \in \Omega_\xi$ by the ergodicity of the irrational rotation. This implies that for almost all $x \in [0,1)$, any $k = 1,2,\cdots$ and $\xi \in \{0,1\}^k$, there exists $n \in \mathbb{N}$ such that

$$\alpha(x)_{n+q_1+\tau(i)} = \xi_{i+1} \quad (i = 0,1\cdots,k-1),$$

where $\tau(i) := q_{i+1} - q_1$ $(i = 0,1\cdots,k-1)$. Thus, we have $p^*_{\alpha(x)}(k) = 2^k$ $(k = 1,2,\cdots)$ for almost all $x \in [0,1)$. □

6 Tribonacci word

In this section we show:

Theorem 6. *Let $\beta = \beta_0\beta_1\beta_2\cdots$ be the tribonacci word, that is the fixed point of the Rauzy substitution (1.2) and $\alpha \in \{0,1\}^{\mathbb{N}}$ be such that*

$$\alpha_n = \begin{cases} 0 & \text{if } \beta_n = 1 \\ 1 & \text{if } \beta_n = 2 \text{ or } 3 \end{cases} \qquad (n \in \mathbb{N}). \tag{6.1}$$

Then for $k = 1,2,3,\cdots$, we have

$$p_\alpha(k) = \begin{cases} k+1 & \text{for } k \leq 8 \\ 2k-7 & \text{for } k \geq 8 \end{cases} \tag{6.2}$$

$$p^*_\alpha(k) = 2^k. \tag{6.3}$$

Remark 2. *The word α in Theorem 6 is also of the form $\alpha = \mathcal{R}(\vec{\theta}, S, \vec{0}, \mathbf{V})$ where $\vec{\theta}$ is a 2-dimensional irrational rotation, S a fractal compact domain in $\mathbb{R}^2$ (called the Rauzy fractal), and $\mathbf{V}$ a lattice in $\mathbb{R}^2$.*

Let us define a sequence $b = \{b_i;\ i = 0, 1, 2, \cdots\}$ by

$$\begin{aligned} b_0 = 1, \quad b_1 = 2, \quad b_2 = 4 \\ b_n = b_{n-3} + b_{n-2} + b_{n-1} \quad (n = 3, 4, \cdots). \end{aligned} \tag{6.4}$$

For any $n \in \mathbb{N}$, there exists a unique *admissible* representation of n in base b, that is

$$n = \sum_{i=0}^{\infty} n_i b_i$$

with the restriction that

$$n_i \in \{0, 1\}, \quad n_i \cdot n_{i+1} \cdot n_{i+2} = 0 \quad (i = 0, 1, 2, \cdots).$$

We thus define the i-th digit n_i $(i = 0, 1, 2, \cdots)$ of $n \in \mathbb{N}$ in base b.

Lemma 3. *For each $n \in \mathbb{N}$ we have $\alpha_n = n_0$, where n_0 is the 0-th digit of n in base b.*

Proof. Let $\psi(n) = n_0 n_1 n_2 \cdots \in \{0, 1\}^{\mathbb{N}}$ be the admissible representation of $n \in \mathbb{N}$ and $\Lambda := \psi(\mathbb{N}) \subset \{0, 1\}^{\mathbb{N}}$. Note that the admissible representation of $n + 1$ is the next element to that of n in the reverse lexicographical order for any $n \in \mathbb{N}$. We define a mapping σ from Λ to $\Lambda \cup \Lambda^2$ by

$$\sigma(n_0 n_1 n_2 \cdots) = \begin{cases} (0 n_0 n_1 n_2 \cdots,\ 1 n_0 n_1 n_2 \cdots) & n_0 n_1 \in \{00, 01, 10\} \\ 0 n_0 n_1 n_2 \cdots & n_0 n_1 = 11. \end{cases}$$

Then, it holds that the sequence

$$\sigma(\psi(0)),\ \sigma(\psi(1)),\ \sigma(\psi(2)),\ \cdots,$$

where if $\sigma(\psi(n)) \in \Lambda^2$, we put 2 elements $0 n_0 n_1 n_2 \cdots$, $1 n_0 n_1 n_2 \cdots$ in the place of $\sigma(\psi(n))$ in the above, coincides with the sequence

$$\psi(0),\ \psi(1),\ \psi(2),\ \cdots,$$

since both sequences begins in $\psi(0)$ and at any place of the sequences, the next element is determined as the next element in Λ in the reverse lexicographical order.

Let $\gamma \in \{1,2,3\}^{\mathbb{N}}$ be such that

$$\gamma_n = \begin{cases} 1 & n_0 = 0 \\ 2 & n_0 n_1 = 10 \\ 3 & n_0 n_1 = 11 \end{cases} \quad (n \in \mathbb{N}).$$

Then by the above argument, it follows that γ is invariant under the substitution (1.2). Since the fixed point of the substitution (1.2) is unique, we have $\gamma = \beta$. Thus, we have $\alpha_n = n_0$ $(n \in \mathbb{N})$. □

Proof of (6.3) in Theorem 6: Take any $k = 1, 2, \cdots$ and define a k-window τ by $\tau(i) = b_{4+6i} - b_4$ $(i = 0, 1, \cdots, k-1)$, where the sequence b is defined in (6.4). Take any $\xi = (\xi_0, \xi_1, \cdots \xi_{k-1}) \in \{0,1\}^k$. Define $n(\xi) \in \mathbb{N}$ in term of the digits in base b as follows:

$$\psi(n(\xi)) = 001\ 0\xi_0 1\ 001\ 0\xi_1 1\ \cdots\ 0\xi_{k-2} 1\ 001\ 0\xi_{k-1} 1\ 000 \cdots.$$

We prove that $\alpha_{n(\xi)+b_4+\tau(i)} = \xi_i$ $(i = 0, 1, \cdots, k-1)$ for any $\xi \in \{0,1\}^k$, which implies that $p^*_\alpha(k) = 2^k$. For this purpose, it is sufficient to prove that

$$(n(\xi) + b_{4+6i})_0 = \xi_i. \tag{6.5}$$

If $\xi_i = 0$, then since

$$\psi(n(\xi) + \xi_i) = 001\ 0\xi_0 1\ \cdots\ 001\ \overbrace{011}^{4+6i}\ \cdots\ 001\ 0\xi_{k-1} 1\ 000 \cdots,$$

we have $(n(\xi) + b_{4+6i})_0 = 0 = \xi_i$.

If $\xi_i = 1$, then we have non-admissible representation

$$001\ 0\xi_0 1\ \cdots\ 001\ \overbrace{021}^{4+6i}\ \cdots\ 001\ 0\xi_{k-1} 1\ 000 \cdots$$

for $n(\xi) + \xi_i$. Since $2b_j = b_{j-3} + b_{j+1}$ $(j = 3, 4, \cdots)$ and $2b_2 = b_0 + b_3$, we normalize the above non-admissible representation as follows:

$$\begin{aligned}
& 001\ 0\xi_0 1\ \cdots\ 0\xi_{i-1}1\ 001\ \overbrace{021}^{4+6i}\ 00\cdots \\
= \ & 001\ 0\xi_0 1\ \cdots\ 0\xi_{i-1}1\ 011\ \overbrace{002}^{4+6i}\ 00\cdots \\
= \ & 001\ 0\xi_0 1\ \cdots\ 0\xi_{i-1}1\ 012\ \overbrace{000}^{4+6i}\ 10\cdots \\
= \ & 001\ 0\xi_0 1\ \cdots\ 0\xi_{i-1}2\ 010\overbrace{100}^{4+6i}\ 10\cdots \\
= \ & \cdots \\
= \ & 002\ 0\xi_0 0\ \cdots\ 1\xi_{i-1}0\ 110\ \overbrace{100}^{4+6i}\ 10\cdots \\
= \ & 100\ 1\xi_0 0\ \cdots\ 1\xi_{i-1}0\ 110\ \overbrace{100}^{4+6i}\ 10\cdots .
\end{aligned}$$

Thus, we have $(n(\xi) + b_{4+6i})_0 = 1 = \xi_i$, which proves (6.5) and completes the proof.

Proof of (6.2) in Theorem 6: We introduce some notations and definitions used here. A factor $u \in F(\alpha)$ is called *right special* (resp. *left special*) if there exists distinct $a, b \in A$ such that ua and ub (resp. au and bu) are each in $F(\alpha)$. A word u which is both right special and left special is called *bispecial*.

Given a word $u = u_0u_1 \cdots u_k \in F(\alpha)$ set

$$\mathrm{Pref}(u) = \{u_0u_1 \cdots u_l | 0 \le l \le k\} \cup \{\varepsilon\}$$

and

$$\mathrm{Suff}(u) = \{u_lu_{l+1} \cdots u_k | 0 \le l \le k\} \cup \{\varepsilon\}.$$

Given a finite subset $\{u^{(1)}, u^{(2)}, \ldots, u^{(n)}\} \subset F(\alpha)$ put

$$\mathrm{Pref}(u^{(1)}, u^{(2)}, \ldots, u^{(n)}) = \bigcup_{1 \le k \le n} \mathrm{Pref}(u^{(k)})$$

and

$$\mathrm{Suff}(u^{(1)}, u^{(2)}, \ldots, u^{(n)}) = \bigcup_{1 \le k \le n} \mathrm{Suff}(u^{(k)}).$$

If u and v are non-empty words in $F(\alpha)$ we will write $u \vdash uv$ to mean that for each word $w \in F(\alpha)$ with $|w| = |u| + |v|$, if u is a prefix of w then $w = uv$. Similarly we will write $vu \dashv u$ to mean that for each word $w \in F(\alpha)$ with $|w| = |u| + |v|$, if u is a suffix of w then $w = vu$.

Let u, $U \in F(\alpha)$. We say that U is a *first return* to u (in α) if Uu is a factor of α having exactly two occurrences of u, one as a prefix and one as a suffix.

It is readily verified that the three first returns to 11 in β are

- $r_1 = 1121312$
- $r_2 = 11213121312$
- $r_3 = 1121312121312$

Thus in α, blocks between consecutive occurrences of 00 are uniquely decoded with respect to σ. For instance, the factor 001010100 of α is necessarily the image under σ of the factor 112131211 of β. Thus any word u in α containing 00 can be uniquely decoded except for possibly a short segment at the beginning and at the end of u; more precisely:

Lemma 4. *Let u be a factor of α containing* 00 *and let v and v' be factors of β with $\sigma(v) = \sigma(v') = u$. Then we can write $v = rxs$ and $v' = r'xs'$ where $x \in F(\beta)$, $|r| = |r'|$, $|s| = |s'|$, and $r, r' \in$ Suff*$(1213, 1312)$, *and* $s, s' \in$ *Pref*$(2131, 3121)$.

Proof. As mentioned above, blocks in u between consecutive occurrences of 00 are uniquely decoded with respect to σ. Thus the only possible ambiguities can occur preceding the first occurrence of 00 and following the last occurrence of 00. We can write $v = v_0 11 v_1$ and $v' = v'_0 11 v'_1$ such that v_1 and v'_1 have no occurrences of 11. Then $|v_1| = |v'_1|$. Since $r_1 \vdash r_1 1$ and $r_1 1$ is the longest common prefix of $r_1 1$, r_2 and r_3, if $|v_1| \leq 6$, then $v_1 = v'_1$ is a prefix (possibly empty) of 213121. If $|v_1| \geq 7$, then both v_1 and v'_1 begin in 213121, and we can write $v_1 = 213121s$ and $v'_1 = 213121s'$. Since $r_2 \vdash r_2 11$ and $r_3 \vdash r_3 11$, it follows that $s, s' \in \mathrm{Pref}(2131, 3121)$. A similar argument establishes that $r, r' \in \mathrm{Suff}(1213, 1312)$. ◻

Remark 3. *We remark that the assumption that* u *contain an occurrence of* 00 *is necessary as for example the factors* $v = 121213121$ *and* $v' = 121312131$ *of* β *map onto the same word* $u = 010101010$ *under* σ *but do not satisfy the conclusion of the lemma.*

Lemma 5. *Let* $u \in F(\alpha)$ *be a right special factor of* α. *Then either*

Type 1 $u = \sigma(v)$ *where* v *is the unique right special factor of* β *of length* $|u|$.

or

Type 2 $u = \sigma(v)1010$ *where* v *is the unique right special factor of* β *of length* $|u| - 4$.

Moreover, if $|u| \leq 3$, *then* u *is of Type 1, if* $4 \leq |u| \leq 7$, *then* u *is both of Type 1 and Type 2, while for each* $n \geq 8$, α *has exactly two right special factors of length* n, *one of Type 1 and one of Type 2.*

Proof. We first observe that if v is a right special factor of β, then $\sigma(v)$ is a right special factor of α. Moreover, since both $v21312$ and $v31211$ are factors of β applying σ it follows that $\sigma(v)10101$ and $\sigma(v)10100$ are both factors of α whence $\sigma(v)1010$ is also a right special factor of α. In other words, all factors of α of Type 1 and of Type 2 are right special.

It is readily verified, that for each $n \leq 7$, α has a unique right special factor of length n which is either of Type 1 or of Type 2, or both. On the other hand, for $n \geq 8$ a right special factor of Type 1 of length n will have 00101010 as a suffix, while a right special factor of Type 2 of length n will have 10101010 as a suffix. Hence for $n \geq 8$, α will have at least two right special factors of length n.

Thus it remains to show that if u is a right special factor of α of length $|u| \geq 8$, then u is either of Type 1 or of Type 2. The result is verified directly in case $8 \leq |u| \leq 13$. Now suppose $|u| \geq 14$. then u must contain an occurrence of 00. Since $u0, u1 \in F(\alpha)$ there exist $v, v' \in F(\beta)$ and $a \in \{2, 3\}$ such that $v1, v'a \in F(\beta)$ and $u = \sigma(v) = \sigma(v')$. Moreover we can write $v = rxs$ and $v' = r'xs'$

where $x \in F(\beta)$, $|r| = |r'|$, $|s| = |s'|$, and $r, r' \in \mathrm{Suff}(1213, 1312)$, and $s, s' \in \mathrm{Pref}(2131, 3121)$.

Case 1. If $v = v'$ then v is a right special factor of β and hence u is of Type 1.

Case 2. If $v \neq v'$ and $s = s'$, then $r \neq r'$ and hence $w = xs$ is a bispecial factor of β. We claim that $1w$ cannot be right special. In fact, if $1w$ is right special, then $w1$ is left special (since w is a palindrome, and $F(\beta)$ is closed under mirror image), whence $2w \vdash 2w1$ and $3w \vdash 3w1$ which gives a contradiction since $v'a$ either ends in $2wa$ or in $3wa$. If $2w$ is right special, then since $|w| \geq 2$ and w in particular is left special, it follows that $2w$ begins in 212. But as $131212 \dashv 212$, we have $1312w \dashv 2w$ and hence $1312w$ is also right special. Since one of v or v' is a suffix of $1312w$, v or v' is right special, and again u is of Type 1. Finally, if $3w$ is right special, then as $1213 \dashv 3$, it follows that $1213w \dashv 3w$ and hence $1213w$ is right special. Since one of v or v' is a suffix of $1213w$, v or v' is right special and so u is of Type 1.

Case 3. If $v \neq v'$ and $s \neq s'$, then x is right special. We begin by showing that in this case $s, s' \in \{3121, 2131\}$. Clearly either s or s' begin in 3. If s begins in 3, then as $s1 \in F(\beta)$ we have that $s \in \{3, 312, 3121\}$. But if $s = 3$ then $s' = 2$, and if $s = 312$ then $s' = 213$ so in either case $s' \vdash s'1$. But this is impossible since $s'a \in F(\beta)$. Hence if s begins in 3, then $s = 3121$ and $s' = 2131$. On the other hand, if s' begins in 3, then since $s'a \in F(\beta)$ it follows that $s' \in \{31, 3121\}$. But if $s' = 31$, then $s = 21$ and since $|x| \geq 2$, it follows that x ends in 21 and hence $v1$ ends in 21211 which is a contradiction since $21211 \notin F(\beta)$. Hence if s' begins in 3 then $s' = 3121$ and $s = 2131$.

Thus $v, v' \in \{rx2131, rx3121, r'x2131, r'x3121\}$, and hence $u = \sigma(z)1010$ for any choice of $z \in \{rx, r'x\}$. Now if $r = r'$, then rx is right special; if $r \neq r'$, then the same argument used in *Case 2* shows that either rx or $r'x$ is right special. In either case, u will be a right special factor of α of Type 2. This concludes the proof of the lemma.

□

We now return to the proof of the theorem. It follows from the previous lemma that for $n \leq 7$, α has a unique right special factor of length n, while for $n \geq 8$, α has exactly two right special factors of length n. Hence, $p_\alpha(8) = 9$, and $p_\alpha(n+1) - p_\alpha(n) = 2$ for $n \geq 8$. Hence, for $n \geq 8$, we have $p_\alpha(n) = 2n + C$. Setting $n = 8$ we find $C = -7$ as required.

References

[1] P. Arnoux, S. Ito, Pisot substitutions and Rauzy fractals, Journés Montoises d'Informatique Thorique (Marne-la-Valle, 2000). Bull. Belg. Math. Soc. Simon Stevin 8(2), 2001, p. 181–207.

[2] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n+1$, *Bull. Soc. Math. France,* **119** (1991), p. 199–215.

[3] V. Berthé, Sequences of low complexity: automatic and Sturmian sequences, London Math. Soc. Lecture Note Ser. 279, Cambridge Univ. Press, 2000 p. 1–34

[4] J. Cassaigne, S. Ferenczi, L.Q. Zamboni, Imbalances in Arnoux-Rauzy sequences, *Ann. Inst. Fourier (Grenoble),* **50** fasc. 4 (2000), p. 1265–1276.

[5] E. Coven, G.A. Hedlund, Sequences with minimal block growth, *Math. Syst. Theor.,* **7** no. 3 (1973), p. 138–153.

[6] T. Kamae and L. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words, *Ergodic Theory and Dynamical Systems* **22** no. 4 (2002), p. 1191–1199.

[7] F. Mignosi, L.Q. Zamboni, On the number of Arnoux-Rauzy words, *Acta Arith.* 101(2) (2002), p. 121–129.

[8] M. Morse, G.A. Hedlund, Symbolic dynamics II: Sturmian sequences, *Amer. J. Math.,* **62** (1940), p. 1–42.

[9] G. Rauzy, Nombres algébriques et substitutions, *Bull. Soc. Math. France*, **110** (1982), p. 147–178.

[10] R.N. Risley, L.Q. Zamboni, A generalization of Sturmian sequences; combinatorial structure and transcendence, *Acta Arith.*, **95** no. 2 (2000), p. 167–184.

[11] G. Rote, Sequences with subword complexity 2n J. Number Theory, 46 (1994), p. 196–213.