Available online at www.sciencedirect.com
SCIENCE DIRECT®
Discrete Mathematics 306 (2006) 1651–1668
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

Language structure of pattern Sturmian words

Teturo Kamae[a], Hui Rao[b], Bo Tan[c], Yu-Mei Xue[b]

[a]Matsuyama University, 790-8578, Japan
[b]Department of Mathematics, Tsinghua University, Beijing 100084, PR China
[c]Department of Mathematics, Huazhong University of Science and Technology, Wuhan 430074, PR China

Received 28 November 2005; received in revised form 14 March 2006; accepted 28 March 2006
Available online 19 June 2006

Abstract

Pattern Sturmian words introduced by Kamae and Zamboni [Sequence entropy and the maximal pattern complexity of infinite words, Ergodic Theory Dynamical Systems 22 (2002) 1191–1199; Maximal pattern complexity for discrete systems, Ergodic Theory Dynamical Systems 22 (2002) 1201–1214] are an analogy of Sturmian words for the maximal pattern complexity instead of the block complexity. So far, two kinds of recurrent pattern Sturmian words are known, namely, rotation words and Toeplitz words. But neither a structural characterization nor a reasonable classification of the recurrent pattern Sturmian words is known. In this paper, we introduce a new notion, pattern Sturmian sets, which are used to study the language structure of pattern Sturmian words. We prove that there are exactly two primitive structures for pattern Sturmian words. Consequently, we suggest a classification of pattern Sturmian words according to structures of pattern Sturmian sets and prove that there are at most three classes in this classification. Rotation words and Toeplitz words fall into two different classes, but no examples of words from the third class are known.

Keywords: Uniform complexity; Pattern Sturmian word; Language structure

1. Introduction

1.1. Pattern Sturmian words

Let A denote a nonempty finite set which is called an *alphabet*. Let $\alpha \in A^{\mathbb{N}}$ be an infinite word over A, where $\mathbb{N} = \{0, 1, 2, \ldots\}$ is the *index set*. Let k be a positive integer. By a *k-window* τ, we mean a subset of $\mathbb{N}$ with cardinality k. For a word $\alpha \in A^{\mathbb{N}}$ and a k-window $\tau = \{\tau_0 < \tau_1 < \cdots < \tau_{k-1}\}$, we denote

$$\alpha[n+\tau] := \alpha(n+\tau_0)\alpha(n+\tau_1)\cdots\alpha(n+\tau_{k-1}) \in A^{\tau},$$
$$F_\alpha(\tau) := \{\alpha[n+\tau]; n \in \mathbb{N}\},$$
$$p_\alpha(\tau) := \#F_\alpha(\tau),$$

where $\alpha[n+\tau]$ is considered as a word on the index set τ, and $\#E$ denotes the cardinality of a finite set E. An element in $F_\alpha(\tau)$ is called a *τ-factor* of α. The *maximal pattern complexity* p^*_α for a word α is introduced by Kamae and

E-mail addresses: kamae@apost.plala.or.jp (T. Kamae), hrao@math.tsinghua.edu.cn (H. Rao), bo_tan@163.com (B. Tan), yxue@math.tsinghua.edu.cn (Y-M. Xue).

0012-365X/$ - see front matter
doi:10.1016/j.disc.2006.03.043

Zamboni [10] as

$$p^*_\alpha(k) := \sup_\tau p_\alpha(\tau) \quad (k = 1, 2, 3, \ldots),$$

where the supremum is taken over all k-windows τ. The *block complexity* p_α is defined as

$$p_\alpha(k) = p_\alpha(\{0, 1, \ldots, k-1\}).$$

Morse and Hedlund [14] characterized the eventually periodicity in term of block complexity by showing that *a word α is eventually periodic if and only if $p_\alpha(k) < k+1$ for some $k \in \mathbb{Z}_+ := \{1, 2, \ldots\}$*. A word α with block complexity $p_\alpha(k) = k+1$ $(k \in \mathbb{Z}_+)$, which is of the minimal complexity among the nonperiodic words, is known as *a Sturmian word*. Excellent descriptions of Sturmian words can be found in Chapter 2 of [13] by J. Berstel and P. Séébold, and in Chapter 6 of [5] by P. Arnoux.

In a similar way, Kamae and Zamboni [10] characterized the eventually periodicity in term of maximal pattern complexity. They proved that *a word α is eventually periodic if and only if $p^*_\alpha(k) < 2k$ for some $k \in \mathbb{Z}_+$*. Accordingly, a word α with $p^*_\alpha(k) = 2k$ $(k \in \mathbb{Z}_+)$ is called a *pattern Sturmian word*.

It is shown that Sturmian words are pattern Sturmian. Indeed, the class of pattern Sturmian words is larger than that of Sturmian words. Till now, three classes of pattern Sturmian words are known: rotation words, Toeplitz words and a class of $\{0, 1\}$-words with rare 1, where the first two of them are recurrent, while the last ones are not (see [10,11]).

We do not know whether there are pattern Sturmian words other than of these kinds or not. We are also interested in what are the common points of the three known pattern Sturmian words, and what are the differences between them. In this paper, we analyze the language structure of recurrent pattern Sturmian words, and try to answer these questions.

1.2. Uniform set

Let A be an alphabet and Σ be a countable infinite set. An element $w \in A^\Sigma$ (which is a mapping from Σ to A) is called a *word* on the index set Σ over A, or a *Σ-word* over A. For a nonempty finite set $S \subset \Sigma$, define $\pi_S(w)$ to be the *S-word* which is the restriction of w to S. For $\Omega \subset A^\Sigma$, put $\pi_S(\Omega) := \{\pi_S(w);\ w \in \Omega\}$.

A subset $\Omega \subset A^\Sigma$ is called a *uniform set* if $\#\pi_S(\Omega)$ depends only on the size of S. Thus, we introduce the *uniform complexity* function $p_\Omega : \mathbb{Z}_+ \to \mathbb{Z}_+$ by $p_\Omega(k) = \#\pi_S(\Omega)$ with $\#S = k$. Special concern is paid to two classes of uniform sets, namely, *Sturmian sets* Ω with $p_\Omega(k) = k+1$ $(k \in \mathbb{Z}_+)$ and *pattern Sturmian sets* Ω with $p_\Omega(k) = 2k$ $(k \in \mathbb{Z}_+)$.

Example 1.1. Take $\Sigma = \mathbb{N}$. A word $\alpha \in \{0, 1\}^\mathbb{N}$ is called an *increasing word* (*a decreasing word*) if $\alpha(i) \leqslant \alpha(j)$ $(\alpha(i) \geqslant \alpha(j)$, resp.) whenever $i < j$. A word is monotone if it is increasing or decreasing. A word $\alpha \in \{0, 1\}^\mathbb{N}$ is called a *Dirac word* if there exists $i_0 \in \mathbb{N}$ such that $\alpha(i) = 0$ for any $i \neq i_0$. Define

$$\begin{aligned}
\Omega_0 &:= \{\alpha \in \{0, 1\}^\Sigma;\ \alpha \text{ is increasing}\},\\
\Omega'_0 &:= \{\alpha \in \{0, 1\}^\Sigma;\ \alpha \text{ is Dirac}\},\\
\Omega_1 &:= \{\alpha \in \{0, 1\}^\Sigma;\ \alpha \text{ is monotone}\},\\
\Omega'_1 &:= \{\alpha \in \{0, 1\}^\Sigma;\ \alpha \text{ is either decreasing or Dirac}\},\\
\Omega_2 &:= \{\alpha \in \{0, 1\}^\Sigma;\ \alpha \text{ is either increasing or Dirac}\}.
\end{aligned}$$

Then, it is easily seen that Ω_0 and Ω'_0 are Sturmian sets, while Ω_1, Ω'_1 and Ω_2 are pattern Sturmian sets. These sets will play an important role in our study and these notations will be used throughout the paper.

We will show that a uniform set Ω is finite if and only if $p_\Omega(k) \leqslant k$ holds for some k (Proposition 2.3). Hence, a Sturmian set is an infinite uniform set with the minimum uniform complexity.

As we will see, the pattern Sturmian sets are closely related to the pattern Sturmian words.

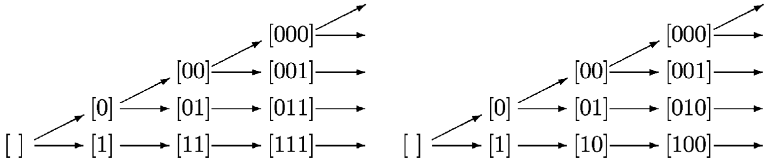

Fig. 1. Ω_0 and Ω'_0.

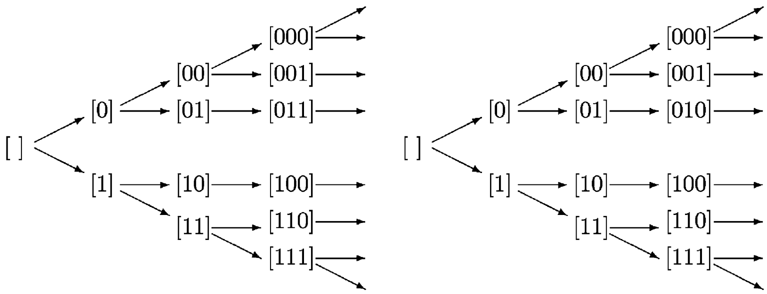

Fig. 2. Ω_1 and Ω'_1.

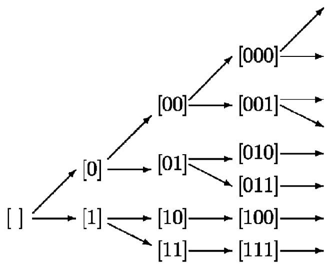

Fig. 3. Ω_2.

1.3. Classification of recurrent pattern Sturmian words

We study the language structure of the uniform sets on the index set $\mathbb{N}$.

We introduce in Section 3 the notion of isomorphism between uniform sets U and V on $\mathbb{N}$, so that U and V are isomorphic to each other if and only if the trees representing the extension schemes of the languages of them along the indices $0, 1, 2, \ldots$ are isomorphic. Then, the *structure* of a uniform set Ω on the index set $\mathbb{N}$ is defined to be the isomorphic class of this isomorphism containing Ω, which is denoted by $[\Omega]$.

It holds that in Example 1.1, Ω_0 and Ω'_0 are isomorphic to each other and Ω_1 and Ω'_1 are isomorphic to each other, while Ω_1 and Ω_2 are not isomorphic (see Figs. 1–3).

Let $\mathcal{N} = \{n_0 < n_1 < n_2 < \cdots\} \subset \mathbb{N}$ and $\psi_{\mathcal{N}} : \{0, 1\}^{\mathbb{N}} \to \{0, 1\}^{\mathbb{N}}$ be such that $\psi_{\mathcal{N}}(\omega)(k) = \omega(n_k)$ $(k \in \mathbb{N})$. The *induced set* $\Omega(\mathcal{N})$ of a set Ω on $\mathcal{N}$ is defined to be the set $\psi_{\mathcal{N}}(\Omega)$. It is a uniform set on $\mathbb{N}$ if Ω is so.

A uniform set Ω is called *primitive* if all *induced sets* of Ω are isomorphic to Ω itself. The structure $[\Omega]$ for a primitive uniform set Ω is called *primitive*. That is, $[\Omega]$ is primitive if there exists a primitive element in $[\Omega]$.

We prove that $[\Omega_0]$ is the unique primitive structure among the Sturmian sets, while there are exactly two different primitive structures among the pattern Sturmian sets, namely, $[\Omega_1]$ and $[\Omega_2]$.

The uniform sets are interesting subject to be studied in general. For example, how to characterize the uniform complexity functions is an interesting problem. Here, we only discuss finite uniform sets, Sturmian sets and pattern Sturmian sets in Sections 2 and 3. The results there except Theorem 3.5 are irrelevant to the arguments after Section 3.

1.4. Ultimate structure

Given a recurrent pattern Sturmian word $\alpha \in \{0,1\}^{\mathbb{N}}$. We prove in Theorem 4.1 that there exists an infinite subset $\mathcal{N}$ of $\mathbb{N}$, which is called an *optimal window*, such that for any nonempty finite set $\tau \subset \mathcal{N}$, we have $p_\alpha(\tau) = 2\#\tau$. Then, we have a pattern Sturmian set $\Omega(\alpha)(\mathcal{N})$, where $\Omega(\alpha)$ denotes the orbit closure of α with respect to the shift on $\{0,1\}^{\mathbb{N}}$.

We denote by $\mathrm{US}(\alpha)$ the set of structures $[\Omega(\alpha)(\mathcal{N})]$ for all optimal windows $\mathcal{N}$ of α such that $\Omega(\alpha)(\mathcal{N})$ is primitive. We prove that $\mathrm{US}(\alpha) = \{[\Omega_1]\}$ for all rotation words α, while $\mathrm{US}(\alpha) = \{[\Omega_2]\}$ for all Toeplitz words α (Theorems 4.3 and 4.8). Thus, we can classify the recurrent pattern Sturmian words in terms of the language structure.

Specially, for Toeplitz words, we give concrete constructions of optimal windows, which give an alternative proof of the fact that the simple Toeplitz words are pattern Sturmian, which is presented with a wrong proof in [11]. Also remark that a proof of this fact in a more general setting can be found in [7].

1.5. More references on the complexity in general

To survey the block complexity in general, see Ferenczi [4]. The block complexity of general Toeplitz words are discussed by Cassaigne and Karhumäki [3] and Koskas [12]. Other kinds of complexity are defined and discussed by Allouche et al. [1], Avgustinovich et al. [2], Frid [6], Nakashima et al. [15], Restivo and Salemi [16]. The notion of pattern Sturmian words is extended to the words over ℓ letters in [8], and to the two-dimensional words in [9].

1.6. Organization of the paper

This paper is organized as follows. Sections 2 and 3 are devoted to the study of uniform sets. In Section 2, the notion of uniform sets is introduced and some basic properties are investigated. We are specially interested in the pattern Sturmian sets which have the uniform complexity $2k$. In Section 3, we study the isomorphism between uniform sets. The isomorphism classes are called structures. We prove that there exist exactly two primitive structures among the pattern Sturmian sets. Section 4 is devoted to the study of language structure of pattern Sturmian words. In Section 4.1, we prove that all recurrent pattern Sturmian words admit optimal windows, which define the ultimate structure of them. In Section 4.2, we study the ultimate structure of the rotation words, while in Section 4.3, we study the ultimate structure of the Toeplitz words.

2. Uniform sets

Let A be an alphabet and Σ be an index set. Let $\mathscr{F}_k$ ($k \in \mathbb{Z}_+$) be the collection of subsets of Σ consisting of k elements, that is, $\mathscr{F}_k = \{S \subset \Sigma;\ \#S = k\}$. Set $\mathscr{F} = \cup_{k \geqslant 1} \mathscr{F}_k$.

For $S \subset \Sigma$, a S-word α over A is called a *constant* word if there exists $a \in A$ such that $\alpha(\sigma) = a$ for any $\sigma \in S$.

Let S and S' be two disjoint subsets of Σ, w and w' be an S-word and an S'-word, respectively, the *concatenation* of w and w' is defined to be the $S \cup S'$-word ww' with the property $ww'(\sigma) = w(\sigma)$ if $\sigma \in S$ and $ww'(\sigma) = w'(\sigma)$ if $\sigma \in S'$.

Given $w \in \pi_S(\Omega)$. If $w' \in \pi_{S'}(\Omega)$ satisfies that $ww' \in \pi_{S \cup S'}(\Omega)$, then we say that w' is an S'-*extension* of w in Ω. For $\sigma \in \Sigma \backslash S$, $w \in \pi_S(\Omega)$ is called σ-*special* if there are at least two different $\{\sigma\}$-extensions of w in Ω. The *complexity* of Ω is the function $p_\Omega : \mathscr{F} \to \mathbb{Z}_+$ defined by $p_\Omega(S) = \#\pi_S \Omega$.

Definition 2.1. A nonempty subset $\Omega \subset A^\Sigma$ is called a *uniform set* if the complexity $p_\Omega(S)$ depends only on $\#S$.

If Ω is a uniform set, we have a function $p_\Omega : \mathbb{Z}_+ \to \mathbb{Z}_+$ such that $p_\Omega(k) = p_\Omega(S)$ for any $S \in \mathscr{F}_k$. The function p_Ω is called the *uniform complexity function*.

From now on, we always take $A = \{0, 1\}$. We consider the finite uniform sets first.

Proposition 2.2. *Let $\Omega \in \{0,1\}^\Sigma$ be a finite uniform set, then either*

(i) $p_\Omega(k) \equiv 1$ *and* $\Omega = \{w\}$ *for some* $w \in \{0,1\}^\Sigma$ *or*
(ii) $p_\Omega(k) \equiv 2$ *and* $\Omega = \{w, \overline{w}\}$ *for some* $w \in \{0,1\}^\Sigma$,

where we put $\overline{0} = 1$, $\overline{1} = 0$ *and* $\overline{w}(\sigma) = \overline{w(\sigma)}$ *for any* $\sigma \in \Sigma$.

Proof. Assume that $\Omega = \{w_1, w_2, \ldots, w_n\}$ is a uniform set.

For $\sigma \in \Sigma$, define a vector $v(\sigma) = (w_1(\sigma), w_2(\sigma), \ldots, w_n(\sigma)) \in \{0,1\}^n$. Since there are only finite number of vectors in $\{0,1\}^n$, there exists a vector v, such that $v(\sigma) = v$ for infinitely many of σ. Denote $\Sigma' = \{\sigma;\ v(\sigma) = v\}$. Then for any $S \subset \Sigma'$, $\pi_S(\Omega)$ consists only of constant words, so that $\#\pi_S(\Omega) \leqslant 2$. Since $\#S$ as above can be any positive number and Ω is uniform, we have $p_\Omega(k) \leqslant 2$ $(k = 1, 2, \ldots)$. This implies that either $p_\Omega(k) \equiv 1$ or $p_\Omega(k) \equiv 2$ since $p_\Omega(k)$ is an increasing function of k and $p_\Omega(1) = 1$ implies $p_\Omega(k) = 1$ $(k = 1, 2, \ldots)$.

On the other hand, if $p_\Omega(k) \equiv 1$, then Ω is a singleton; and if $p_\Omega(k) \equiv 2$, then $\Omega = \{w, \overline{w}\}$. □

Proposition 2.3. *Let Ω be a uniform set. If there exists k such that $p_\Omega(k) \leqslant k$, then Ω is a finite set.*

Proof. If $p_\Omega(1) = 1$, then Ω is a singleton. If $p_\Omega(1) = 2$ and $p_\Omega(k) \leqslant k$, then there exists a $k' < k$ such that $p_\Omega(k'+1) = p_\Omega(k')$.

Take $S \subset \Sigma$ with $\#S = k'$. Then for any $\sigma \in \Sigma \setminus S$ and $w \in \Omega$, since $p_\Omega(k'+1) = p_\Omega(k')$, $w(\sigma)$ is determined by the S-word $\pi_S(w)$. Hence w is determined by $\pi_S(w)$, and Ω is a finite set. □

Definition 2.4. A uniform set $\Omega \subset \{0,1\}^\Sigma$ is called a *Sturmian set* if the uniform complexity satisfies $p_\Omega(k) = k+1$ for any $k \in \mathbb{Z}_+$; Ω is called a *pattern Sturmian set* if the uniform complexity satisfies $p_\Omega(k) = 2k$ for any $k \in \mathbb{Z}_+$.

Hence, Sturmian sets have the minimal uniform complexity among all the infinite uniform set by Proposition 2.3. Examples of Sturmian sets and pattern Sturmian sets are given in Example 1.1. We have the following characterizations.

Theorem 2.5. *If Ω is a uniform set. Then, Ω is a Sturmian set if and only if $p_\Omega(2) = 3$.*

Proof. Obviously, the condition $p_\Omega(2) = 3$ is necessary. To show the sufficiency, suppose that Ω is uniform and $p_\Omega(2) = 3$. Then by Propositions 2.2 and 2.3, Ω is an infinite set and $p_\Omega(k) \geqslant k+1$ for any $k \in \mathbb{Z}_+$. So we only need to show that $p_\Omega(k) \leqslant k+1$.

Otherwise, suppose that $p_\Omega(k) \geqslant k+2$ for some $k \geqslant 3$. Then there exists a $k' < k$ such that $p_\Omega(k'+1) \geqslant p_\Omega(k') + 2$.

Take $S \subset \Sigma$ with $\#S = k'$ and $\sigma \in \Sigma \setminus S$. Since $A = \{0,1\}$ and each S-word has at most two $\{\sigma\}$-extensions, there are two S-words w_1 and w_2 which are σ-special. Since w_1 and w_2 are distinct, there exists a $\tau \in S$ such that $w_1(\tau) \neq w_2(\tau)$. So $\pi_{\{\tau,\sigma\}}(\Omega) = \{00, 01, 10, 11\}$ and $p_\Omega(2) = 4$, a contradiction. □

A uniform set Ω is said to fulfill (k)-*Condition* if its uniform complexity satisfies

$$p_\Omega(m) = 2m \quad \text{for } m = 1, 2, \ldots, k; \quad \text{but } p_\Omega(k+1) = 2k+1.$$

Thus, by Theorem 2.5, Ω is a Sturmian set if and only if it fulfills (1)-Condition.

The following lemma will be used for the characterization of pattern Sturmian sets.

Lemma 2.6. *Let Ω be a uniform set.*

(1) *There exists an infinite subset $\Sigma' \subset \Sigma$ such that $\pi_{\Sigma'}(\Omega)$ contains at least one constant word.*
(2) *If Ω fulfills (k)-Condition for some $k \geqslant 2$, then for any $S \subset \Sigma$ with $\sharp S = k$, there exist $\omega_1 \in \pi_S(\Omega)$ and an infinite subset $\Sigma' \subset \Sigma \setminus S$ such that ω_1 is $\{\sigma\}$-special for any $\sigma \in \Sigma'$ and $\{\pi_{\Sigma'}(w);\ w \in \Omega$ and $\pi_S(w) \neq w_1\}$ consists of two constant words.*

Proof. (1) Take an arbitrary word $\omega \in \Omega$. Since Σ is an infinite set, there exists an infinite subset Σ' such that $\pi_{\Sigma'}(\omega)$ is a constant word. Thus, $\pi_{\Sigma'}(\Omega)$ contains at least one constant word.

(2) Assume that Ω fulfills (k)-Condition for some $k \geqslant 2$. Take $S \subset \Sigma$ with $\#S = k$ and write $\pi_S(\Omega) = \{w_1, w_2, \ldots, w_{2k}\}$. Since $p_\Omega(k+1) = 2k+1$, for any $\sigma \in \Sigma \backslash S$, just one element in $\{w_1, w_2, \ldots, w_{2k}\}$ is σ-special. Therefore, without loss of generality, we may assume that there exists an infinite set $\Sigma' \subset \Sigma \backslash S$ such that w_1 is the only element in $\pi_S(\Omega)$ which is σ-special for any $\sigma \in \Sigma'$ and

$$\{\pi_{\Sigma'}(w);\ w \in \Omega \text{ and } \pi_S(w) \neq w_1\}$$

consists only of constant words.

We conclude the proof by claiming that both constant words with 0 and 1 are included in the above set. Otherwise taking $\tau \in S$ and $\sigma \in \Sigma'$, one can find only three different $\{\tau, \sigma\}$-words, which contradicts the fact that $p_\Omega(2) = 4$. □

Theorem 2.7. *If Ω is a uniform set. Then Ω is a pattern Sturmian set if and only if $p_\Omega(2) = 4$ and $p_\Omega(3) = 6$.*

Proof. We only need to show that if Ω is uniform and $p_\Omega(2) = 4$, $p_\Omega(3) = 6$, then $p_\Omega(k) = 2k$ for any k.

- $p_\Omega(k) \leqslant 2k$ for any $k \geqslant 1$: otherwise, there exists a $k \geqslant 3$ such that $p_\Omega(k+1) \geqslant p_\Omega(k) + 3$.
  Take $S \subset \Sigma$ with $\#S = k$ and $\sigma \in \Sigma \backslash S$. There are three S-words w_1, w_2 and w_3 which are σ-special. Since w_1, w_2 and w_3 are distinct, there exists a $\tau_1, \tau_2 \in S$ such that $\pi_{\{\tau_1,\tau_2\}}(w_1)$, $\pi_{\{\tau_1,\tau_2\}}(w_2)$ and $\pi_{\{\tau_1,\tau_2\}}(w_3)$ are different from each other. Also all of them are σ-special. Since $p_\Omega(2) = 4$, there is another $\{\tau_1, \tau_2\}$-word besides $\pi_{\{\tau_1,\tau_2\}}(w_1)$, $\pi_{\{\tau_1,\tau_2\}}(w_2)$ and $\pi_{\{\tau_1,\tau_2\}}(w_3)$ which has at least one $\{\sigma\}$-extension. Then $\pi_{\{\tau_1,\tau_2,\sigma\}}(\Omega) \geqslant 7$, which contradicts the assumption $p_\Omega(3) = 6$.
- $p_\Omega(k) \geqslant 2k$ for any $k \geqslant 1$: otherwise, there exists a $k \geqslant 3$ such that $p_\Omega(k+1) \leqslant p_\Omega(k) + 1$.

If $p_\Omega(k+1) = p_\Omega(k)$, then as in the proof of Proposition 2.3, we can show that Ω is finite which is a contradiction.

It remains only one possibility: Ω fulfills (k)-Condition for some $k \geqslant 3$. In this case, fix $S \subset \Sigma$ with $\#S = k$. Then $\#\pi_S(\Omega) = 2k$, $\pi_S(\Omega) = \{w_1, w_2, \ldots, w_{2k}\}$. By Lemma 2.6(2), there exists an infinite subset $\Sigma' \subset \Sigma \backslash S$ and $w_1 \in \pi_S(\Omega)$ such that w_1 is σ-special for any $\sigma \in \Sigma'$ and $\{\pi_{\Sigma'}(w);\ w \in \Omega \text{ and } \pi_S(w) \neq w_1\}$ consists only of the two constant words.

Construct a set Ω' as follows:

$$\Omega' = \{\pi_{\Sigma'}(w);\ w \in \Omega \text{ and } \pi_S(w) = w_1\}.$$

We claim that Ω' is a uniform set. To see this, for a fixed finite subset $S' \subset \Sigma'$, consider the $S \cup S'$-words of Ω. Since any S-word but w_1 has a unique S'-extension, and w_1 has just $p_{\Omega'}(S')$ different S'-extensions, we have

$$\begin{aligned} p_{\Omega'}(S') &= p_\Omega(S \cup S') - (2k-1) \\ &= p_\Omega(k + \#S') - (2k-1) \end{aligned}$$

which implies Ω' is a uniform set such that

$$p_{\Omega'}(m) = p_\Omega(k+m) - (2k-1)$$

for any $m \geqslant 1$.

Obviously, $p_{\Omega'}(1) = 2$. We claim that $p_{\Omega'}(2) = 4$. Otherwise $p_{\Omega'}(2) = 2$ or 3. If $p_{\Omega'}(2) = 2$, then by Propositions 2.2 and 2.3, $\Omega' = \{\omega, \overline{\omega}\}$ for some $\omega \in \{0,1\}^{\Sigma'}$. Take $S' \subset \Sigma'$ with $\#S' = 3$. Consider the set $\pi_{S'}(\Omega)$ which is the union of $\pi_{S'}(\Omega')$ and the two constant words. Then, we have

$$p_\Omega(S') \leqslant 2 + 2 = 4,$$

contradicting to the fact that $p_\Omega(3) = 6$. If $p_{\Omega'}(2) = 3$, then Ω' is a Sturmian set. Then by Lemma 2.6(1), there exists a three-elements subset $S' \subset \Sigma'$ such that $\pi_{S'}(\Omega')$ consists of four elements and among them at least one is a constant word. Consider the set $\pi_{S'}(\Omega)$ which is the union of $\pi_{S'}(\Omega')$ and the two constant words. Then,

$$p_\Omega(S') \leqslant 4 + 2 - 1 = 5,$$

contradicting to the fact that $p_\Omega(3) = 6$.

Moreover, since Ω' is defined on $\Sigma' \subset \Sigma$ and $\Omega' \subset \pi_{\Sigma'}(\Omega)$, we have

$$p_{\Omega'}(m) \leqslant p_\Omega(m) \quad \text{for any } m \geqslant 1.$$

Therefore, Ω' fulfills (k')-Condition for some $2 \leqslant k' \leqslant k$.

Due to Lemma 2.6(2) applied to Ω', there exists an infinite set $\Sigma'' \subset \Sigma'$ such that $\Omega'' := \pi_{\Sigma''}(\Omega')$ contains the two constant words. Replacing Σ' and Ω' by these Σ'' and Ω'', we may assume that the above Ω' contains the two constant words. Since for any $S' \subset \Sigma'$, $\pi_{S'}(\Omega)$ is the union of $\pi_{S'}(\Omega')$ and two constant words, we have $\pi_{S'}(\Omega) = \pi_{S'}(\Omega')$ and $p_{\Omega'} \equiv p_\Omega$.

On the other hand, fix $\sigma \in S$ and $S' \subset \Sigma'$ with $\#S' = m$. Without loss of generality, assume that $w_1(\sigma) = 0$. Since $p_\Omega(2) = 4$, for any $\tau \in S'$, $\pi_{\{\sigma,\tau\}} = \{00, 01, 10, 11\}$. Since $\pi_{S'}(\Omega')$ contains both the constant words,

$$\pi_{\{\sigma\}\cup S'}(\Omega) = \{0w;\ w \in \pi_{S'}(\Omega')\} \cup \{10^m, 11^m\},$$

where, for example, 10^m denotes the $\{\sigma\} \cup S'$-word w with $w(\sigma) = 1$ and $w(s) = 0$ for $s \in S'$. Since the union is disjoint,

$$p_\Omega(m+1) = p_{\Omega'}(m) + 2 \quad \text{for any } m \geqslant 1.$$

This is a contradiction against the facts $p_{\Omega'} \equiv p_\Omega$, $p_\Omega(k) = 2k$ and $p_\Omega(k+1) = 2k+1$, which completes the proof of $p_\Omega(k) \geqslant 2k$. $\square$

3. Isomorphism between uniform sets

In this section, we consider only the uniform sets on the index set $\Sigma := \mathbb{N}$ equipped with the natural total ordering. Recall that the alphabet A is always $\{0, 1\}$.

The product topology defined on $\{0, 1\}^{\mathbb{N}}$ is consistent with the following metric: for $x = x(0)x(1)x(2)\ldots$, $y = y(0)y(1)y(2)\ldots \in \{0, 1\}^{\mathbb{N}}$,

$$d(x, y) = 2^{-\inf\{k \in \mathbb{N};\ x(k) \neq y(k)\}}.$$

Thus two points are closer to each other if they share a longer prefix. The cylinder $[\xi]$, where $\xi = \xi_1\xi_2\cdots\xi_n \in \{0, 1\}^n$, is the set of words of the form

$$[\xi] = \{x \in \{0, 1\}^{\mathbb{N}};\ x(0) = \xi_1,\ x(1) = \xi_2, \ldots, x(n-1) = \xi_n\}.$$

The *order* of a cylinder $[\xi]$ is defined to be the length n of ξ, denoted by $|\xi|$. Note that for the empty word $\emptyset$, $[] := [\emptyset] = \{0, 1\}^{\mathbb{N}}$.

Definition 3.1. Two uniform sets $\Omega, \Omega' \subset \{0, 1\}^{\mathbb{N}}$ are said to be *isomorphic* to each other, written $\Omega \approx \Omega'$, if there is an isometry between their closures $\overline{\Omega}$ and $\overline{\Omega'}$, that is, there is a bijection $\varphi : \overline{\Omega} \to \overline{\Omega'}$ such that for any $x, y \in \overline{\Omega}$,

$$d(\varphi(x), \varphi(y)) = d(x, y).$$

An equivalence class of uniform sets with respect to this isomorphism is called a *structure*. The structure containing Ω is denoted by $[\Omega]$.

Note that Ω and its closure $\overline{\Omega}$ always have the same language, that is, the set of finite words appearing in Ω and $\overline{\Omega}$ coincide. Also, note that two uniform sets which are isomorphic to each other have the same uniform complexity.

For a uniform set $\Omega \subset \{0, 1\}^{\mathbb{N}}$, we define the *prefix tree* $G(\Omega)$ as follows: $G(\Omega) = (V, E)$ is a directed graph. The set V of vertices is the set of the cylinders which meet Ω, and the set E of (directed) edges is the set of the ordered pairs $([u], [v])$ of cylinders in V such that v is an immediate extension of u, that is, $|v| = |u| + 1$ and $u_1 = v_1$, $u_2 = v_2, \ldots, u_{|u|} = v_{|u|}$.

Recall that two directed graphs $G = (V, E)$ and $G' = (V', E')$ are isomorphic, written $G \cong G'$, if there is a bijection $\phi : V \to V'$ between their vertices, such that there is an edge in E from u to v if and only if there is an edge in E' from $\phi(u)$ to $\phi(v)$.

Theorem 3.2. *Let Ω and Ω' be two uniform sets. Then $\Omega \approx \Omega'$ if and only if $G(\Omega) \cong G(\Omega')$.*

Proof. If $\Omega \approx \Omega'$, then there is an isometry $\varphi : \overline{\Omega} \to \overline{\Omega'}$. Thus, $x, y \in \overline{\Omega}$ are in an identical cylinder of order n if and only if $\varphi(x)$ and $\varphi(y)$ are also in an identical cylinder of order n in $\overline{\Omega'}$. Hence, φ induces a bijection ϕ between the cylinders intersecting with Ω and the cylinders intersecting with Ω' keeping the orders. Thus, ϕ is a bijection between the vertices of $G(\Omega)$ and $G(\Omega')$ which preserves the edges, and is an isomorphism between $G(\Omega)$ and $G(\Omega')$.

Conversely, assume $G(\Omega) \cong G(\Omega')$. Noticing that there is a natural correspondence between the words in $\overline{\Omega}$ and the infinite paths from the root in $G(\Omega)$, the isomorphism between the prefix trees induces a map $\varphi : \overline{\Omega} \to \overline{\Omega'}$, which is an isometry. □

Let $\mathcal{N} = \{n_0 < n_1 < n_2 < \cdots\}$ be an infinite subset of $\mathbb{N}$. Let Ω be a uniform set on $\mathbb{N}$. Recall the notion of *induced set* $\Omega(\mathcal{N})$ of Ω on $\mathcal{N}$ in Section 1.3.

Definition 3.3. A uniform set $\Omega \subset \{0, 1\}^{\mathbb{N}}$ or a structure $[\Omega]$ is said to be *primitive* if for any infinite subset $\mathcal{N} \subset \mathbb{N}$, the induced set $\Omega(\mathcal{N})$ is isomorphic to the original set Ω.

All the uniform sets in Example 1.1 are easily seen to be primitive. Their prefix trees are depicted in Figs. 1–3.

Theorem 3.4. *For any Sturmian set Ω, there exists an infinite subset $\mathcal{N} \subset \mathbb{N}$ such that $\Omega(\mathcal{N}) \approx \Omega_0$. In particular, if Ω is primitive, then $\Omega \approx \Omega_0$. Hence, $[\Omega_0]$ is the unique primitive structure among the Sturmian sets.*

Proof. Let $\Omega \subset \{0, 1\}^{\mathbb{N}}$ be a Sturmian set.

Put $n_0 = 0$. Just as in the proof of Lemma 2.6, we can take an infinite set $\mathcal{N}_1 \subset \mathbb{N}\backslash\{n_0\}$ such that one $\{n_0\}$-word, say a ($\in \{0, 1\}$), is σ-special for $\sigma \in \mathcal{N}_1$, while the other $\{n_0\}$-word $\overline{a}$ has only one $\{\sigma\}$-extension for any $\sigma \in \mathcal{N}_1$. Put $\Omega' = \{\pi_{\mathcal{N}_1}(\omega);\ \omega \in \Omega \text{ with } \omega(n_0) = a\}$.

Then, Ω' is again a Sturmian set since for any set $S \subset \mathcal{N}_1$ with $\#S = k$,

$$p_{\Omega'}(S) = p_\Omega(S \cup \{n_0\}) - 1 = k + 2 - 1 = k + 1.$$

Put $n_1 = \min \mathcal{N}_1$, and we continue the above process: find $\mathcal{N}_2 \subset \mathcal{N}_1\backslash\{n_1\}$ such that in Ω' one $\{n_1\}$-word has a unique $\mathcal{N}_2$-extension while the $\mathcal{N}_2$-extensions of the other $\{n_1\}$-word form a Sturmian set. Put $n_2 = \min \mathcal{N}_2$, and so on.

At last, setting $\mathcal{N} = \{n_0, n_1, n_2, \ldots\}$, by the construction of $\mathcal{N}$ we have $\Omega(\mathcal{N}) \approx \Omega_0$ (see Fig. 1). Moreover, if Ω is primitive, then $\Omega \approx \Omega(\mathcal{N}) \approx \Omega_0$. □

The next theorem characterizes the primitive pattern Sturmian sets.

Theorem 3.5. *Let Ω be a pattern Sturmian set. Then either $\Omega(\mathcal{N}) \approx \Omega_1$ for some $\mathcal{N} \subset \mathbb{N}$ or $\Omega(\mathcal{N}) \approx \Omega_2$ for some $\mathcal{N} \subset \mathbb{N}$. In particular, if Ω is primitive, then either $\Omega \approx \Omega_1$ or $\Omega \approx \Omega_2$. Hence, $[\Omega_1]$ and $[\Omega_2]$ are the only primitive structures among the pattern Sturmian sets.*

Proof. Let $\Omega \subset \{0, 1\}^{\mathbb{N}}$ be a pattern Sturmian set.

For any $\sigma, \tau \in \mathbb{N}$ with $\sigma < \tau$, $\pi_{\{\sigma,\tau\}}(\Omega) = \{00, 01, 10, 11\}$ because $p_\Omega(2) = 4$. For any $\zeta > \tau$, since $p_\Omega(3) = 6$, there are two $\{\sigma, \tau\}$-words which are ζ-special. If one of the special words comes from the set $\{00, 01\}$ and the other comes from $\{10, 11\}$, we call ζ a $\{\sigma, \tau\}$-*balanced place*. If all but finite number of places are $\{\sigma, \tau\}$-balanced places, we say that Ω is $\{\sigma, \tau\}$-*balanced*. If Ω is $\{\sigma, \tau\}$-balanced for any $\sigma, \tau \in \mathbb{N}$ with $\sigma < \tau$, we say that Ω is *balanced*.

We consider two cases according to the balance property.

Case 1: Ω is balanced.

Take $n_0 = 0$, $n_1 = 1$. Since Ω is $\{n_0, n_1\}$-balanced, we can take an infinite subset $\mathcal{N}_1 \subset \{n_1 + 1, n_1 + 2, \ldots\}$ such that one $\{n_0, n_1\}$-word in $\{00, 01\}$ is σ-special for any $\sigma \in \mathcal{N}_1$, and one $\{n_0, n_1\}$-word in $\{10, 11\}$ is also σ-special for any $\sigma \in \mathcal{N}_1$. Any of the other two words in $\{00, 01, 10, 11\}$ has a unique $\mathcal{N}_1$-extension.

Take $n_2 = \min \mathcal{N}_1$. For any $\{n_0, n_1\}$-word w which is $\{n_2\}$-special, $w0, w1 \in \pi_{\{n_0,n_1,n_2\}}(\Omega)$. Since Ω is $\{n_0, n_2\}$-balanced, for all but finite number of σ's with $\sigma > n_2$, one word in $\{w0, w1\}$ is not σ-special, we can take an infinite subset $\mathcal{N}_2 \subset \mathcal{N}_1\backslash\{n_2\}$ such that, for all $\{n_0, n_1\}$-words w which are $\{n_2\}$-special, one word in $\{w0, w1\}$ has a unique $\mathcal{N}_2$-extension.

Take $n_3 = \min \mathcal{N}_2$. Since 4 words in $\pi_{\{n_0,n_1,n_2\}}(\Omega)$ have a unique $\{n_3\}$-extension and $p_\Omega(4) = 8$, the other 2 words in $\pi_{\{n_0,n_1,n_2\}}(\Omega)$ are n_3-special. One of these 2 words starts by 0 and the other starts by 1.

We continue the above process. Finally, we get $\mathcal{N} = \{n_0, n_1, n_2, n_3, \ldots\}$ such that $\Omega(\mathcal{N}) \approx \Omega_1$ (see Fig. 2).

Case 2: Ω is not balanced.

There are places $\tau_0, \eta_0 \in \mathbb{N}$ with $\tau_0 < \eta_0$ such that Ω is not $\{\tau_0, \eta_0\}$-balanced, that is, there are infinite number of places $\sigma > \eta_0$ which are not $\{\tau_0, \eta_0\}$-balanced. Then, we find an infinite subset $\mathcal{N}'' \subset \{\eta_0+1, \eta_0+2, \ldots\}$ such that for any $\sigma \in \mathcal{N}''$, both $\{\tau_0, \eta_0\}$-words either in $\{00, 01\}$ or in $\{10, 11\}$ are σ-special. Without loss of generality, we assume that both $\{\tau_0, \eta_0\}$-words in $\{00, 01\}$ are σ-special for infinitely many $\sigma \in \mathcal{N}''$. Collecting all these σ, we define an infinite set $\mathcal{N}' \subset \mathcal{N}''$ such that both $\{\tau_0, \eta_0\}$-words in $\{00, 01\}$are σ-special for any $\sigma \in \mathcal{N}'$, while $\{\tau_0, \eta_0\}$-words in $\{10, 11\}$ are not σ-special for $\sigma \in \mathcal{N}'$. Denote by Ω' and Ω^* the $\mathcal{N}'$-extensions of $\{00, 01\}$ and $\{10, 11\}$, respectively. More precisely,

$$\Omega' = \{\pi_{\mathcal{N}'}(w);\ \ w \in \Omega \text{ such that } \pi_{\{\tau_0,\eta_1\}}(w) \in \{00, 01\}\},$$
$$\Omega^* = \{\pi_{\mathcal{N}'}(w);\ \ w \in \Omega \text{ such that } \pi_{\{\tau_0,\eta_1\}}(w) \in \{10, 11\}\}.$$

We claim that Ω' is again a pattern Sturmian set.

To see this, we study the set Ω^* first. Since any $\{\tau_0, \tau_1\}$-word in $\{10, 11\}$ has a unique $\mathcal{N}'$-extension, $\#\Omega^* \leqslant 2$. Moreover, since $\pi_{\{\tau_0,\sigma\}}(\Omega) = \{00, 01, 10, 11\}$ for any $\sigma \in \mathcal{N}'$, the σ-extensions of 10 and 11 are different, thus $\Omega^* = \{x, \overline{x}\}$ for some $\mathcal{N}'$-word x.

Hence, for any finite subset $S \subset \mathcal{N}'$,

$$\pi_{\{\tau_0\}\cup S}(\Omega) = \{0w;\, w \in \pi_S(\Omega')\} \cup \{1u, 1\overline{u}\}$$

for some S-word u. From this, we have

$$p_{\Omega'}(S) = p_\Omega(\{\tau_0\} \cup S) - 2 = 2(\#S + 1) - 2 = 2\#S,$$

Ω' being a pattern Sturmian set.

Subcase 2.1: If Ω' is balanced, then by Case 1, there is an $\mathcal{N}$ such that $\Omega'(\mathcal{N}) \approx \Omega_1$.

But for any $S \subset \mathcal{N}$, $\pi_S(\Omega'(\mathcal{N})) \subset \pi_S(\Omega(\mathcal{N}))$, and since both $\Omega'(\mathcal{N})$ and $\Omega(\mathcal{N})$ are pattern Sturmian, by comparing the cardinality, $\pi_S(\Omega'(\mathcal{N})) = \pi_S(\Omega(\mathcal{N}))$. Therefore, the prefix trees of $\Omega'(\mathcal{N})$ and $\Omega(\mathcal{N})$ are just the same, and $\Omega(\mathcal{N}) \approx \Omega'(\mathcal{N}) \approx \Omega_1$.

Subcase 2.2: If Ω' is not balanced, then there exist $\tau_1, \eta_1 \in \mathcal{N}'$ with $\tau_1 < \eta_1$ such that Ω' is not $\{\tau_1, \eta_1\}$-balanced, and we get a new pattern Sturmian set Ω'', and continue the above discussion.

In case of the new pattern Sturmian set constructed in some step is balanced, then just as shown in Subcase 2.1, we find $\mathcal{N} \subset \mathbb{N}$ such that $\Omega(\mathcal{N}) \approx \Omega_1$.

Otherwise, if all the pattern Sturmian sets constructed in the process are balanced, we get a sequence $\tau_0 < \tau_1 < \tau_2 < \cdots$. Putting

$$\mathcal{N} = \{\tau_0, \tau_1, \tau_2, \ldots\},$$

we have $\Omega(\mathcal{N}) \approx \Omega_2$ (see Fig. 3). □

Example 3.6. Let $\tilde{\Omega} \subset \{0, 1\}^{\mathbb{Z}}$ be the set of words $\tilde{\alpha}$ on the index set $\mathbb{Z}$ such that either $\tilde{\alpha}$ is increasing or $\tilde{\alpha}$ is Dirac. Define a word $\phi(\tilde{\alpha}) \in \{0, 1\}^{\mathbb{N}}$ by

$$\phi(\tilde{\alpha})(k) = \begin{cases} \tilde{\alpha}(i) & \text{if } k = 2i \text{ is even,} \\ \tilde{\alpha}(-i) & \text{if } k = 2i-1 \text{ is odd.} \end{cases}$$

Let $\Omega = \phi(\tilde{\Omega})$. Then, Ω is a pattern Sturmian set such that $\Omega(\mathcal{N}) = \Omega_2$ if $\mathcal{N} = \{0, 2, 4, \ldots\}$ and $\Omega(\mathcal{N}) = \Omega'_1 \approx \Omega_1$ if $\mathcal{N} = \{1, 3, 5, \ldots\}$. Hence, Ω is not primitive. Moreover, as shown in Fig. 4, Ω is isomorphic neither to Ω_1 nor to Ω_2, so that $[\Omega]$ is not primitive.

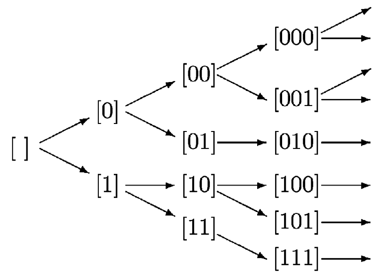

Fig. 4. Ω in Example 3.6.

4. Structures of pattern Sturmian words

4.1. From pattern Sturmian words to pattern Sturmian sets

Recall that a word α on $\mathbb{N}$ is called *recurrent* if for any $L \geqslant 1$, there exists $M \geqslant 1$ such that

$$\alpha(i) = \alpha(i + M) \quad \text{for } i = 0, 1, \ldots, L - 1. \tag{1}$$

Note that if α is recurrent, then there exist infinitely many M's satisfying (1).

In this subsection, we will show how to construct pattern Sturmian sets from a recurrent pattern Sturmian word. Let α be a pattern Sturmian word. For a finite or infinite subset $\mathscr{N} \subset \mathbb{N}$, consider the following property (we will call it the *optimal property*): for any nonempty finite subset $\tau \subset \mathscr{N}$, it holds that

$$p_\alpha(\tau) = 2\#\tau. \tag{2}$$

If an infinite subset $\mathscr{N}$ of $\mathbb{N}$ has the optimal property, then $\mathscr{N}$ is called an *optimal window* for α. We note that an infinite subset of an optimal window is again an optimal window.

Theorem 4.1. *Let α be a recurrent pattern Sturmian word. Then, there exists an optimal window for α.*

Proof. Suppose that α is a recurrent pattern Sturmian word. We construct an increasing family of sets satisfying the optimal property.

Put $n_0 = 0$.

Assume that $n_0 < n_1 < \cdots < n_{k-1}$ have been picked out such that the optimal property holds for the set $\eta := \{n_0, n_1, \ldots, n_{k-1}\}$.

Take $L \in \mathbb{N}$ such that

$$F_\alpha(\eta) = \{\alpha[n + \eta];\ n = 0, 1, \ldots, L - 1\}.$$

Since α is recurrent, there exists M with $M > n_{k-1}$ such that (1) holds for this L. Put $n_k = M$.

Now we show that (2) holds for any nonempty $\tau \subset \{n_0, n_1, \ldots, n_k\}$.

- If $n_k \notin \tau$, (2) holds for τ by the hypothesis of induction.
- If $n_k \in \tau$ but $n_0 \notin \tau$. Write $\tau' = \{n_0\} \cup (\tau \backslash \{n_k\})$. Then $\#\tau' = \#\tau$.

Note that (2) holds for τ' by the hypothesis of induction, actually

$$\#\{\alpha[n + \tau'];\ n = 0, 1, \ldots, L - 1\} = \#F_\alpha(\tau') = 2\#\tau'.$$

On the other hand, by (1), $\alpha(n_k) = \alpha(n_0)$ for $n = 0, 1, \ldots, L - 1$. Therefore, there is an one-to-one correspondence between the sets

$$\{\alpha[n + \tau];\ n = 0, 1, \ldots, L - 1\}$$

and

$$\{\alpha[n+\tau'];\ \ n=0,1,\ldots,L-1\}.$$

Hence

$$p_\alpha(\tau)\geqslant \#\{\alpha[n+\tau];\ \ n=0,1,\ldots,L-1\}=2\#\tau$$

holds, which implies $p_\alpha(\tau)=2\#\tau$ since α is pattern Sturmian. Thus, (2) holds for τ.

- If both $n_k\in\tau$ and $n_0\in\tau$. Denoting

$$F_\alpha^{(1)}(\tau)=\{\alpha[n+\tau];\ \ n\in\mathbb{N} \text{ and } \alpha(n+n_0)=\alpha(n+n_k)\},$$
$$F_\alpha^{(2)}(\tau)=\{\alpha[n+\tau];\ \ n\in\mathbb{N} \text{ and } \alpha(n+n_0)\neq\alpha(n+n_k)\},$$

then $F_\alpha(\tau)=F_\alpha^{(1)}(\tau)\cup F_\alpha^{(2)}(\tau)$ and the union is disjoint.

For $F_\alpha^{(1)}(\tau)$: denoting $\tau'=\tau\backslash\{n_k\}$, then $\#F_\alpha^{(1)}(\tau)=\#F_\alpha(\tau')$. By the hypothesis of induction, $\#F_\alpha(\tau')=2\#\tau'=2\#\tau-2$. Thus

$$\#F_\alpha^{(1)}(\tau)=2\#\tau-2.$$

For $F_\alpha^{(2)}(\tau)$: considering the set

$$F_\alpha^{(2)}(\{n_0,n_k\})=\{\alpha[n+\{n_0,n_k\}];\ \ n\in\mathbb{N} \text{ and } \alpha(n+n_0)\neq\alpha(n+n_k)\}.$$

Now if $\#F_\alpha^{(2)}(\{n_0,n_k\})\leqslant 1$, then α is eventually periodic (for the proof of this fact, see Section 3 in [10, pp. 1194–1195]). Since α is pattern Sturmian, $\#F_\alpha^{(2)}(\{n_0,n_k\})\geqslant 2$, and

$$\#F_\alpha^{(2)}(\tau)\geqslant\#F_\alpha^{(2)}(\{n_0,n_k\})\geqslant 2.$$

Therefore

$$\#F_\alpha(\tau)=\#F_\alpha^{(1)}(\tau)+\#F_\alpha^{(2)}(\tau)\geqslant 2\#\tau,$$

and (2) holds for τ.

Put $\mathscr{N}=\{n_0,n_1,n_2,\ldots\}$. Then, $\mathscr{N}$ is an optimal window for α. □

Let $T:\{0,1\}^{\mathbb{N}}\to\{0,1\}^{\mathbb{N}}$ be the *shift*, that is, $(T\alpha)(n)=\alpha(n+1)\ (\forall n\in\mathbb{N})$. For $\alpha\in\{0,1\}^{\mathbb{N}}$, let

$$O(\alpha):=\overline{\{T^n\alpha;\ \ n\in\mathbb{N}\}}$$

be the *orbit closure* of α with respect to T.

Let $\alpha\in\{0,1\}^{\mathbb{N}}$ be a recurrent pattern Sturmian word and $\mathscr{N}$ be an infinite subset of $\mathbb{N}$. Then, it is easy to see that $\mathscr{N}$ is an optimal window for α if and only if the induced set $O(\alpha)(\mathscr{N})$ is a pattern Sturmian set.

Definition 4.2. For a recurrent pattern Sturmian word α, the structure $[O(\alpha)(\mathscr{N})]$ is called an *ultimate structure* of α if $\mathscr{N}$ is an optimal window such that $O(\alpha)(\mathscr{N})$ is primitive. Let $\mathrm{US}(\alpha)$ be the set of ultimate structures of α.

In virtue of Theorem 3.5, we can classify recurrent pattern Sturmian words α into three classes:
Class 1: $\mathrm{US}(\alpha)=\{[\Omega_1]\}$,
Class 2: $\mathrm{US}(\alpha)=\{[\Omega_2]\}$, and
Class 3: $\mathrm{US}(\alpha)=\{[\Omega_1],[\Omega_2]\}$.

In the next two subsections, we will show that the rotation words belong to Class 1, while the Toeplitz words belong to Class 2. We do not know whether a recurrent pattern Sturmian word α belonging to Class 3 exists or not.

4.2. Rotation words

Let θ be an irrational number with $0<\theta<1$, S be a subset of the torus $\mathbb{T}=\mathbb{R}/\mathbb{Z}$, $x \in [0,1)$. The rotation word $\mathscr{R}(\theta, S, x) \in \{0,1\}^{\mathbb{N}}$ is defined as

$$\mathscr{R}(\theta, S, x)(n) = \begin{cases} 1 & \text{if } x+n\theta \in S \,(\mathrm{mod}\, \mathbb{Z}), \\ 0 & \text{otherwise.} \end{cases}$$

In [10], Kamae and Zamboni showed that when $S \subset \mathbb{T}$ is an interval with length <1, then $\mathscr{R}(\theta, S, x)$ is a pattern Sturmian word for any $x \in [0,1)$. While in [11], they constructed a compact set S such that $\mathscr{R}(\theta, S, x)$ has full maximal pattern complexity for almost every x. Remark that it is not so difficult to improve their construction to obtain a set S such that $\mathscr{R}(\theta, S, x)$ has full maximal pattern complexity for every x.

In the following, let $S=I=[a,b)$ be an interval of length $b-a<1$, so that $\alpha=\mathscr{R}(\theta, I, x)$ is a pattern Sturmian word. Let $I_1=I$ and I_0 be its complement in the torus $\mathbb{T}$. Thus $\mathscr{I}=\{I_0, I_1\}$ is a partition of $\mathbb{T}$. Take an arbitrary window $\tau: 0=\tau_0<\tau_1<\cdots<\tau_{k-1}$ of size $k \geqslant 1$. As shown in Example 2 of [10], $p_\alpha(\tau)$ is just the number of nonempty elements in the partition

$$(\mathscr{I}-\tau_0\theta) \vee (\mathscr{I}-\tau_1\theta) \vee \cdots \vee (\mathscr{I}-\tau_{k-1}\theta),$$

which "$\vee$" means the least common refinement of the partitions. Moreover, that $x+n\theta$ is in an element of the partition, say

$$x+n\theta \in (I_{c_0}-\tau_0\theta) \cap (I_{c_1}-\tau_1\theta) \cap \cdots \cap (I_{c_{k-1}}-\tau_{k-1}\theta) \,(\mathrm{mod}\, \mathbb{Z}),$$

is equivalent to that $\alpha(n+\tau_i)=c_i$ $(i=0,1,\ldots,k-1)$.

Now we are ready to show that $[\Omega_1]$ is the unique ultimate structure of a rotation word.

Theorem 4.3. *Let $\alpha=\mathscr{R}(\theta, I, x)$ be a rotation word. Then for any infinite subset $\Sigma \subset \mathbb{N}$, there exists an optimal window $\mathscr{N} \subset \Sigma$ such that $O(\alpha)(\mathscr{N})=\Omega_1$. Hence,* $\mathrm{US}(\alpha)=\{[\Omega_1]\}$.

Proof. Let $I=[a,b)$ be an interval in $\mathbb{T}$ with $0<b-a<1$. Let Σ be any infinite subset of $\mathbb{N}$. Then, there exists an accumulating point $d \in \mathbb{T}$ of $\{n\theta \,(\mathrm{mod}\, \mathbb{Z});\ n \in \Sigma\}$ and a sequence $n_0<n_1<n_2<\cdots$ in Σ such that the sequence $(n_k\theta)_{k=0,1,2,\ldots}$ $(\mathrm{mod}\, \mathbb{Z})$ converges to d either from the left or from the right. We represent d as $-1<d\leqslant 0$, and without loss of generality, we assume that $(n_k\theta)_{k=0,1,2,\ldots}$ $(\mathrm{mod}\, \mathbb{Z})$ converges to d from the left, so that

$$d-\varepsilon<-\xi_0<-\xi_1<-\xi_2<\cdots<d$$

with $\xi_k := \lceil n_k\theta \rceil - n_k\theta$ $(k=0,1,2,\ldots)$ and $\varepsilon>0$ satisfying that $\varepsilon<\min\{b-a, a+1-b\}$.

Let $\mathscr{N}=\{n_0, n_1, n_2, \ldots\}$. We prove that $O(\alpha)(\mathscr{N}) \approx \Omega_1$.

Assume that $\alpha(n+n_h)=1$ and $\alpha(n+n_k)=1$ for some $h<k$ and $n \in \mathbb{N}$. Then from the above argument, there exist integers M and N such that

$$a+M+\xi_h \leqslant x+n\theta<b+M+\xi_h$$

and

$$a+N+\xi_k \leqslant x+n\theta<b+N+\xi_k.$$

Since $0<\xi_h-\xi_k<\varepsilon$, we have $M=N$ and

$$a+M+\xi_h \leqslant x+n\theta<b+M+\xi_k.$$

Take any ℓ with $h<\ell<k$, then we have

$$a+M+\xi_\ell<a+M+\xi_h \leqslant x+n\theta<b+M+\xi_k<b+M+\xi_\ell,$$

and hence, $\alpha(n+n_\ell)=1$.

Assume that $\alpha(n+n_h)=0$ and $\alpha(n+n_k)=0$ for some $h<k$ and $n\in\mathbb{N}$. Then from the above argument, there exist integers M and N such that

$$b+M+\xi_h\leqslant x+n\theta<a+M+1+\xi_h$$

and

$$b+N+\xi_k\leqslant x+n\theta<a+N+1+\xi_k.$$

Since $0<\xi_h-\xi_k<\varepsilon$, we have $M=N$ and

$$b+M+\xi_h\leqslant x+n\theta<a+M+1+\xi_k.$$

Take any ℓ with $h<\ell<k$, then we have

$$b+M+\xi_\ell<b+M+\xi_h\leqslant x+n\theta<a+M+1+\xi_k<a+M+1+\xi_\ell,$$

and hence, $\alpha(n+n_\ell)=0$.

These two arguments imply that the word $(\alpha(n+n_k))_{k=0,1,2,\ldots}$ on $k\in\mathbb{N}$ is monotone for any $n\in\mathbb{N}$. Hence, $O(\alpha)(\mathcal{N})\subset\Omega_1$. On the other hand, since $O(\alpha)(\mathcal{N})$ and Ω_1 have the same uniform complexity function and they are compact, they must coincide. Thus, $O(\alpha)(\mathcal{N})=\Omega_1$.

Suppose that $[\Omega_2]\in\mathrm{US}(\alpha)$. Then, there exists an infinite set of $\Sigma\subset\mathbb{N}$ such that $O(\alpha)(\Sigma)$ is a primitive set satisfying that $O(\alpha)(\Sigma)\approx\Omega_2$. Take an infinite set $\mathcal{N}\subset\Sigma$ such that $O(\alpha)(\mathcal{N})=\Omega_1$ as above. Since $O(\alpha)(\Sigma)$ is primitive, we have $O(\alpha)(\mathcal{N})\approx\Omega_2$, contradicting to the fact that $O(\alpha)(\mathcal{N})=\Omega_1$. □

4.3. Toeplitz words

For $a\in\{0,1\}$ and integers l,r with $l\geqslant 2$ and $0\leqslant r\leqslant l-1$, 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 two letters $\{a,?\}$. We define $\beta^{(a,l,r)}\lhd\beta^{(b,m,s)}\in\{0,1,?\}^{\mathbb{N}}$ as the word obtained by replacing each occurrence of "?" in $\beta^{(a,l,r)}$ by the letters in $\beta^{(b,m,s)}$ one by one in the order.

An infinite word α over $\{0,1\}$ is called a *Toeplitz word* if there exists an infinite sequence

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

satisfying the conditions that

(1) $a_{i+1}\neq a_i$ for any $i\in\mathbb{N}$,
(2) $l_i\geqslant 2$ for any $i\in\mathbb{N}$, and
(3) $0\leqslant r_i\leqslant l_i-1$ $(i\in\mathbb{N})$ with $r_i\geqslant 1$ infinitely often, so that

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

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

Remark 4.4. The above Toeplitz words are called *simple* Toeplitz words in [11]. Since in this paper, only the simple Toeplitz words are considered, we simply call them Toeplitz words.

First, let us introduce a graph representation of a Toeplitz word. Let α be the Toeplitz word with the coding sequence $(a_0,l_0,r_0),(a_1,l_1,r_1),(a_2,l_2,r_2),\ldots$. We construct an infinite tree T_α according to the coding sequence as follows:

The root of the tree is [], which is the only vertex at level 0. The vertices at level 1 are $[0],[1],\ldots,[l_0-1]$ and there is an edge from [] to each $[i]$ $(0\leqslant i<l_0)$. The edge from [] to $[r_0]$ is labeled by "?", while the other edges are not labeled.

In general, the vertices at level n are $[i_0i_1\cdots i_{n-1}]$ $(0\leqslant i_j<l_j;\ \ j=0,1,\ldots,n-1)$. There are $l_0l_1\cdots l_{n-1}$ number of them. There is an edge from $[i_0i_1\cdots i_{n-1}]$ to each $[i_0i_1\cdots i_{n-1}i_n]$ $(0\leqslant i_n<l_n)$. The edge from $[r_0r_1\cdots r_{n-1}]$ to

$[r_0 r_1 \cdots r_{n-1} r_n]$ is labeled by "?", while the other edges between level n and $n+1$ are not labeled. We call each of $[i_0 i_1 \cdots i_{n-1} i_n]$ $(0 \leqslant i_n < l_n)$ an *offspring* of $[i_0 i_1 \cdots i_{n-1}]$.

The infinite path in T_α starting from the root and passing the vertices $[c_0], [c_0 c_1], [c_0 c_1 c_2] \cdots$ will be denoted by the sequence $(c_0, c_1, c_2, \ldots)$. We consider them as elements in $\prod_{n \in \mathbb{N}} \{0, 1, \ldots, l_n - 1\}$. The infinite paths in T_α will be simply called *paths*.

The path with code $(r_0, r_1, r_2, \ldots)$ is the unique infinite path passing only the edges with label "?". We denote this path by $\xi(\alpha)$.

The elements in $\prod_{n \in \mathbb{N}} \{0, 1, \ldots, l_n - 1\}$ having only finitely many nonzero coordinates correspond to the natural numbers by the mapping

$$\Theta : (c_0, c_1, \ldots, c_k, 0, 0, \ldots) \mapsto c_0 + c_1 \times l_0 + \cdots + c_k \times \prod_{i=0}^{k-1} l_i.$$

It is easy to see that Θ is a bijection. Put $\Psi = \Theta^{-1}$. The set of paths is a compact set with the metric $d(\zeta, \eta) = 2^{-n}$, where n is the level at which the paths ζ and η *part* (defined below).

Now we define the addition of two paths. For two paths $c=(c_0, c_1, c_2, \ldots), d=(d_0, d_1, d_2, \ldots) \in \prod_{n \in \mathbb{N}} \{0, 1, \ldots, l_n - 1\}$, the summation (with carry) $e = c \oplus d = (e_0, e_1, e_2, \ldots)$ is defined by induction as follows: $c_0 + d_0 = e_0 + \delta_0 l_0$, where $0 \leqslant e_0 \leqslant l_0 - 1$ and $\delta_0 \in \{0, 1\}$ is the 0th carry. Suppose that e_{n-1} and the $(n-1)$th carry δ_{n-1} are defined, then $c_n + d_n + \delta_{n-1} = e_n + \delta_n l_n$, where $0 \leqslant e_n \leqslant l_n - 1$ and $\delta_n \in \{0, 1\}$.

Under the addition $\oplus$, the set of paths $\prod_{n \in \mathbb{N}} \{0, 1, \ldots, l_n - 1\}$ becomes an abelian group. If $e = c \oplus d$, we also write $c = e \ominus d$. Remark that the addition between paths is compatible with the addition between natural numbers:

$$\Psi(m+n) = \Psi(m) \oplus \Psi(n) \quad \text{for any } m, n \in \mathbb{N}.$$

Given two paths $c = (c_0, c_1, c_2, \ldots)$ and $d = (d_0, d_1, d_2, \ldots)$, we say that c and d *part* at level k if $c_0 = d_0, c_1 = d_1, \ldots, c_{k-1} = d_{k-1}$, while $c_k \neq d_k$. In this case, we also say that c and d *part* at v, where v is the vertex at level k in T_α such that c and d pass.

We have the following characterization of the Toeplitz word.

Proposition 4.5. *For any $n \in \mathbb{N}$, if in T_α, the paths $\Psi(n)$ and $\xi(\alpha)$ part at level k, then $\alpha(n) = a_k$, or equivalently, $\alpha(n) \equiv a_0 + k \pmod 2$.*

Proof. Since $\xi(\alpha) = (r_0, r_1, r_2, \ldots)$ and $r_i \neq 0$ infinitely often and $\Psi(n)$ ends with 0's, $\Psi(n)$ and $\xi(\alpha)$ will part at some level.

If $\Psi(n)$ parts from $\xi(\alpha)$ at level 0, then $\beta^{(a_0, l_0, r_0)}(n) = a_0$. Thus $\alpha(n) = a_0$. If $\Psi(n)$ parts from $\xi(\alpha)$ at level 1, then $\beta^{(a_0, l_0, r_0)}(n)=?$, but this "?" will be filled by a_1 from the word $\beta^{(a_1, l_1, r_1)}$. Thus $\alpha(n) = a_1$.

The proof proceeds in this way. □

We illustrate in Fig. 5 the Toeplitz word with coding sequence

$$(0, 4, 1), (1, 5, 3), (0, 3, 2), (1, 3, 2), (0, 5, 0), \ldots$$

so that $\xi(\alpha) = (1, 3, 2, 2, 0, \ldots)$. The thick branch is coded by $(1, 3, 1, 0, 0, \ldots)$ and represents the natural number $33 = 1 + 3 \times 4 + 1 \times 4 \times 5$. Since the thick branch and $\xi(\alpha)$ part at level 2, $\alpha(33) = a_2 = 0$.

The following lemma is easy to check.

Lemma 4.6. *Let c and d be two paths in T_α. If c and d part at level k. Then for any path e, the paths $c \oplus e$ and $d \oplus e$ also part at level k.*

For any $n \in \mathbb{N}$, put $\xi(\alpha)_n = \xi(\alpha) \ominus \Psi(n)$. Let $\tau = \{\tau_0, \ldots, \tau_{k-1}\}$ be a k-window. We denote by $T_\alpha(\tau)$ the subtree of T_α generated by the paths $\xi(\alpha)_{\tau_0}, \xi(\alpha)_{\tau_1}, \ldots, \xi(\alpha)_{\tau_{k-1}}$. Then, the τ-factor $\alpha[n+\tau]$ of α is determined by $\Psi(n)$ and $T_\alpha(\tau)$ so that $\alpha(n+\tau_i)$ is the sum modulo 2 of a_0 and the level at which $\Psi(n)$ parts from $\xi(\alpha)_{\tau_i}$ in $T_\alpha(\tau)$.

For a vertex v of $T_\alpha(\tau)$ other than the root, denote

$$\deg_\tau(v) = \#\{w;\ \text{there is a (directed) edge from } v \text{ to } w \text{ in } T_\alpha(\tau)\}.$$

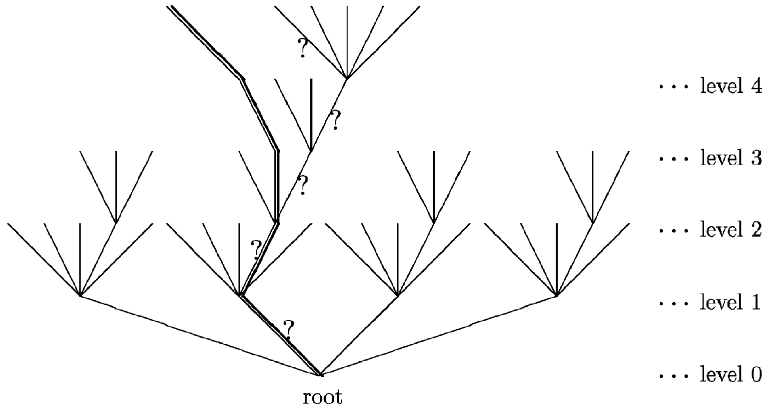

Fig. 5. Graph representation of a Toeplitz word.

For the root [], we denote

$$\deg_\tau([]) = \#\{w;\ \text{there is a (directed) edge from [] to } w \text{ in } T_\alpha(\tau)\} + 1.$$

A vertex v is called an *m-multiple* vertex of $T_\alpha(\tau)$ if $\deg_\tau(v) = m$ for $m = 1, 2, 3, \ldots$. We denote the set of m-multiple vertices of $T_\alpha(\tau)$ by $\mathbb{B}_m(\tau)$ and

$$\mathbb{B}(\tau) := \bigcup_{m=2}^{\infty} \mathbb{B}_m(\tau).$$

Note that $\#\mathbb{B}(\tau) \leqslant \#\tau$.

Theorem 4.7. *For any window* τ, *we have*

$$p_\alpha(\tau) \leqslant \#\mathbb{B}(\tau) + \#\tau \leqslant 2\#\tau.$$

Moreover, $p_\alpha(\tau) = 2\#\tau$ *holds if for all edges* (u, w) *in* $T_\alpha(\tau)$,

$$\deg_\tau(u) + \deg_\tau(w) \leqslant 3 \tag{3}$$

holds.

Proof. Use the induction on the size of τ. In the case $\#\tau = 1$, our theorem is trivial. Assume that the conclusion holds for windows of size $k \geqslant 1$. We prove that our theorem holds for any window of size $k + 1$.

Given a window τ of size $k + 1$ and consider the tree $T_\alpha(\tau)$. Choose a vertex $v \in \mathbb{B}(\tau)$ located at the maximal level among $\mathbb{B}(\tau)$.

Case 1: v is the root.

In this case, we have $\#\mathbb{B}(\tau) = 1$. Since for any $n \in \mathbb{N}$, $\Psi(n)$ parts from all but one paths among $\xi(\alpha)_i$ $(i \in \tau)$ at level 0, $\alpha[n + \tau]$ contains at most one $\overline{a_0}$. Hence, $p_\alpha(\tau) \leqslant \#\tau + 1$. Moreover, since $\deg_\tau(v) + \deg_\tau(w) \geqslant \#\tau + 1 + 1 \geqslant 4$ holds for any w such (v, w) is an edge in $T_\alpha(\tau)$, we get Theorem 4.7.

Consider the case where v is not the root.

Take i and j in τ such that in $T_\alpha(\tau)$, $\xi(\alpha)_i$ and $\xi(\alpha)_j$ part at v. Set $\tau' = \tau\backslash\{i\}$. There are two cases to be considered (Fig. 6).

Case 2.1: v is not the root and $\deg_\tau(v) \geqslant 3$.

In this case, $\deg_{\tau'}(v) \geqslant 2$ and $v \in \mathbb{B}(\tau')$. Let

$$J' := \{j' \in \tau'\backslash\{j\};\ \ j \text{ and } j' \text{ part at } v\} \neq \emptyset.$$

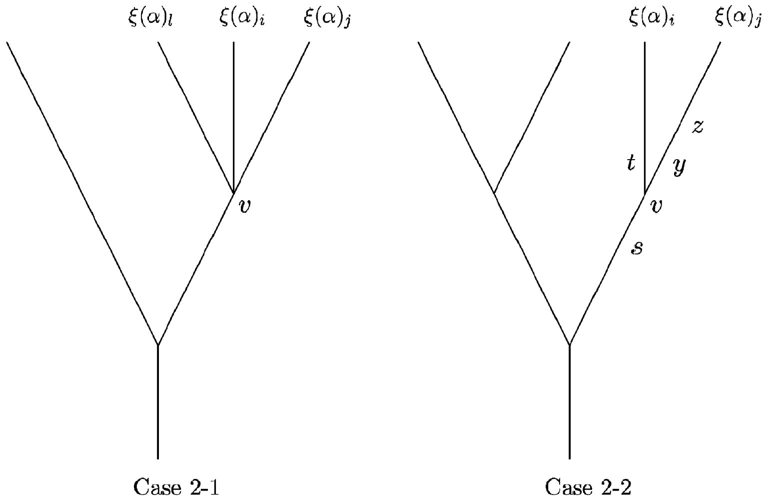

Fig. 6. τ-trees.

There is a partition of $\mathbb{N} = N_0 \cup N_1 \cup N_2$ with

$$N_0 = \{n \in \mathbb{N};\ \Psi(n) \text{ parts from } T_\alpha(\tau') \text{ at } v\},$$
$$N_1 = \{n \in \mathbb{N};\ \Psi(n) \text{ passes } v \text{ parting from } T_\alpha(\tau') \text{ at a higher level than } v\},$$
$$N_2 = \{n \in \mathbb{N};\ \Psi(n) \text{ doesn't pass } v\}.$$

It holds that $\{\alpha[n+\tau'];\ n \in N_0\}$ consists of one element, say ζ, which has two $\{i\}$-extensions in $F_\alpha(\tau)$, since there are two kinds of elements, say n_0 and n_1 in N_0, such that the parity of the level at which $\Psi(n_h)$ and $\xi(\alpha)_i$ part coincides with that of h for $h = 0, 1$.

It holds that

$$F_1 := \{\alpha[n+\tau'];\ n \in N_1\} \backslash \{\zeta\}$$

and

$$F_2 := \{\alpha[n+\tau'];\ n \in N_2\} \backslash \{\zeta\}$$

are disjoint, since all $\delta \in F_1$ satisfy that $\delta(j) \neq \delta(j')$ for some $j' \in J'$, while all $\delta \in F_2$ satisfy that $\delta(j) = \delta(j')$ for any $j' \in J'$. Moreover, all $\delta \in F_1$ have the unique $\{i\}$-extension whose value is the sum of the level of v and a_0 modulo 2, and all $\delta \in F_2$ have the unique $\{i\}$-extension whose value coincides with $\delta(j)$.

Therefore, $p_\alpha(\tau) = p_\alpha(\tau') + 1$. On the other hand,

$$\#\mathbb{B}(\tau) + \#\tau = \#\mathbb{B}(\tau') + \#\tau' + 1,$$

so that we have $p_\alpha(\tau) \leqslant \#\mathbb{B}(\tau) + \#\tau$ by the induction hypothesis.

In this case, (3) is not satisfied for any edge (v, w) in $T_\alpha(\tau)$ since

$$\deg_\tau(v) + \deg_\tau(w) \geqslant 3 + 1 = 4.$$

Case 2.2: v is not the root and $\deg_\tau(v) = 2$.

In this case, $\deg_{\tau'}(v) = 1$ and $v \notin \mathbb{B}(\tau')$.

Assume that (3) is satisfied for all edges (u, w) in $T_\alpha(\tau)$. Then, (3) is satisfied for all edges (u, w) in $T_\alpha(\tau')$ and by the induction hypothesis, we have $p_\alpha(\tau') = 2\#\tau'$.

We denote by s the vertex just before v in $T_\alpha(\tau)$, that is, v is an offspring of s. (In the following, "before" and "after" are taken in this sense.) Denote by t the vertex on $\xi(\alpha)_i$ just after v, by y the vertex on $\xi(\alpha)_j$ just after v and by z the vertex on $\xi(\alpha)_j$ just after y. Then, any of s, t, y, z is an 1-multiple vertex of $T_\alpha(\tau)$ by our assumptions. Let

n_1 (or n_2) $\in \mathbb{N}$ be such that $\Psi(n_1)$ (or $\Psi(n_2)$) parts from $\xi(\alpha)_i$ at s (or t, respectively). Let n_3 (or n_4) $\in \mathbb{N}$ be such that $\Psi(n_3)$ (or $\Psi(n_4)$) parts from $\xi(\alpha)_j$ at y (or z, respectively). Then it holds that

$$\alpha[n_1+\tau'] = \alpha[n_3+\tau'] \neq \alpha[n_2+\tau'] = \alpha[n_4+\tau']$$

and

$$\alpha[n_1+\tau] \neq \alpha[n_3+\tau], \quad \alpha[n_2+\tau] \neq \alpha[n_4+\tau].$$

On the other hand, if $\alpha[n+\tau'] \notin \{\alpha[n_1+\tau'],\ \alpha[n_2+\tau']\}$ for some $n \in \mathbb{N}$, then $\Psi(n)$ should part from $\xi(\alpha)_i$ before s. Then, $\alpha[n+\tau]$ is determined by $\alpha[n+\tau']$ since $\alpha[n+\tau](i) = \alpha[n+\tau'](j)$.

Thus, we have

$$p_\alpha(\tau) = p_\alpha(\tau') + 2 = 2\#\tau' + 2 = 2\#\tau.$$

If (3) does not hold, the only equality which is not necessarily fulfilled is $\alpha[n_1+\tau'] = \alpha[n_3+\tau']$. Still, we have

$$p_\alpha(\tau') + 1 \leqslant p_\alpha(\tau) \leqslant p_\alpha(\tau') + 2.$$

Since

$$\#\mathbb{B}(\tau) + \#\tau = \#\mathbb{B}(\tau') + \#\tau' + 2,$$

we have $p_\alpha(\tau) \leqslant \#\mathbb{B}(\tau) + \#\tau$ by the induction hypothesis. □

Now we are ready to show that $\mathrm{US}(\alpha) = \{[\Omega_2]\}$ for any Toeplitz word α.

Theorem 4.8. *Let α be a Toeplitz word. Then for any infinite subset $\Sigma \subset \mathbb{N}$, there exists an optimal window $\mathcal{N} \subset \Sigma$ such that $O(\alpha)(\mathcal{N})$ is primitive and $O(\alpha)(\mathcal{N}) \approx \Omega_2$. Hence, $\mathrm{US}(\alpha) = \{[\Omega_2]\}$.*

Proof. Take any infinite subset $\Sigma \subset \mathbb{N}$. Since the space Σ with the metric $d(\Psi(n), \Psi(m))$ between $n, m \in \Sigma$ is relatively compact, there exists an infinite path η in $T(\alpha)$ such that $\liminf_{n\in\Sigma, n\to\infty} d(\eta, \Psi(n)) = 0$. Denote by $D(\eta, n)$ the level at which η and $\Psi(n)$ part. Then, there exists $b \in \{0, 1\}$ and an infinite subset $\mathcal{N} = \{n_0 < n_1 < n_2 < \cdots\}$ of Σ such that $D(\eta, n_0) \geqslant 2$, $D(\eta, n_{i+1}) \geqslant D(\eta, n_i) + 2$ and $D(\eta, n_i) \equiv b \pmod 2$ for any $i = 0, 1, 2, \ldots$. Let $a \in \{0, 1\}$ be such that $a \equiv a_0 + b \pmod 2$.

Let $T_\alpha(\mathcal{N})$ be the subtree of T_α generated by the paths $\xi(\alpha)_{n_i}$'s ($i = 0, 1, 2, \ldots$). Then, condition (3) is always satisfied for any edge (u, w) in $T_\alpha(\mathcal{N})$. Also, $\xi(\alpha) \ominus \eta$ is a path in $T_\alpha(\mathcal{N})$.

For $n \in \mathbb{N}$, we define $\alpha[n, \mathcal{N}] \in \{0, 1\}^{\mathbb{N}}$ by $\alpha[n, \mathcal{N}](i) = \alpha(n + n_i)$ ($i \in \mathbb{N}$). Then, it is easy to check that

(i) if $\Psi(n)$ parts from $\xi(\alpha) \ominus \eta$ at level m such that $m < D(\eta, n_0)$, then either $\alpha[n, \mathcal{N}] = aaa\cdots$ or $\alpha[n, \mathcal{N}] = \overline{aaa}\cdots$ depending on whether $m \equiv b \pmod 2$ or not,

(ii) if $\Psi(n)$ parts from $\xi(\alpha) \ominus \eta$ at level m such that $\rho_\eta(n_i) < m < \rho_\eta(n_{i+1})$ for some $i \in \mathbb{N}$, then either $\alpha[n, \mathcal{N}] = aaa\cdots$ or $\alpha[n, \mathcal{N}] = aa\cdots \overset{i}{a}\,\overline{aa}\cdots$ depending on whether $m \equiv b \pmod 2$ or not,

(iii) if $\Psi(n)$ parts from $\xi(\alpha) \ominus \eta$ at level $\rho_\eta(n_i)$, then either $\alpha[n, \mathcal{N}] = aaa\cdots$ or $\alpha[n, \mathcal{N}] = a\cdots a\,\overset{i}{\overline{a}}\,a\cdots$ depending on whether the level m at which $\Psi(n)$ parts from $\xi(\alpha)_{n_i}$ satisfies $m \equiv b \pmod 2$ or not.

Thus, if $a = 0$, then $O(\alpha)(\mathcal{N}) = \Omega_2$, if else, then $O(\alpha)(\mathcal{N}) = \{\overline{\omega};\ \omega \in \Omega_2\}$. In any case, $O(\alpha)(\mathcal{N})$ is a primitive set satisfying that $O(\alpha)(\mathcal{N}) \approx \Omega_2$. Thus, $[\Omega_2] \in \mathrm{US}(\alpha)$.

Suppose that $[\Omega_1] \in \mathrm{US}(\alpha)$. Then, there exists an infinite set of $\Sigma \subset \mathbb{N}$ such that $O(\alpha)(\Sigma)$ is a primitive set satisfying that $O(\alpha)(\Sigma) \approx \Omega_1$. Take an infinite set $\mathcal{N} \subset \Sigma$ such that $O(\alpha)(\mathcal{N}) = \Omega_2$ as above. On the other hand, since $O(\alpha)(\Sigma)$ is primitive, we have $O(\alpha)(\mathcal{N}) \approx \Omega_1$, contradicting to the fact that $O(\alpha)(\mathcal{N}) = \Omega_2$. □

Acknowledgements

The authors thank Prof. Wen Zhi-Ying (Tsinghua University) for inviting them to the Morningside Center of Mathematics (CAS) where the paper is conceived and completed. The authors thank also the anonymous referees for a lot of useful suggestions.

References

[1] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, Theoret. Comput. Sci. 292 (2003) 9–31.
[2] S. Avgustinovich, D. Fon-Der-Flaass, A. Frid, Arithmetical complexity of infinite words, in: M. Ito, T. Imaoka (Eds.), Words, Languages and Combinatorics III, Kyoto, 2000, World Scientific, Singapore, 2003, pp. 51–62.
[3] J. Cassaigne, J. Karhumäki, Toeplitz words, generalized periodicity and periodically iterated morphisms, European J. Combin. 18 (1997) 497–510.
[4] S. Ferenczi, Complexity of sequences and dynamical systems, Discrete Math. 206 (1999) 145–154.
[5] N.P. Fogg, Substitutions in Dynamics, Arithmetics, and Combinatorics, in: V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel (Eds.), Lecture Notes in Mathematics, vol. 1794, Springer, Berlin, 2002.
[6] A. Frid, Sequences of linear arithmetical complexity, Theoret. Comput. Sci. 339 (2005) 68–87.
[7] N. Gjini, T. Kamae, B. Tan, Y-M. Xue, Maximal pattern complexity for Toeplitz words, preprint.
[8] T. Kamae, H. Rao, Maximal pattern complexity over ℓ letters, European J. Combinatorics 27 (2006) 125–137.
[9] T. Kamae, H. Rao, Y.-M. Xue, Maximal pattern complexity for 2-dimensional words, Theoret. Comput. Sci., to appear.
[10] T. Kamae, L. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words, Ergodic Theory Dynamical Systems 22 (2002) 1191–1199.
[11] T. Kamae, L. Zamboni, Maximal pattern complexity for discrete systems, Ergodic Theory Dynamical Systems 22 (2002) 1201–1214.
[12] M. Koskas, Complexités de suites de Toeplitz, Discrete Math. 183 (1998) 161–183.
[13] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.
[14] M. Morse, G.A. Hedlund, Symbolic dynamics II: Sturmian sequences, Amer. J. Math. 62 (1940) 1–42.
[15] I. Nakashima, J.-I. Tamura, S.-I. Yasutomi, *-Sturmian words and complexity, J. Théorie Nombres Bordeaux 15 (2003) 767–804.
[16] A. Restivo, S. Salemi, Binary patterns in infinite binary words, in: X. Brauer et al. (Eds.), Formal and Natural Computing, Lecture Notes in Computer Science, vol. 2300, Springer, Berlin, 2002, pp. 107–116.