Maximal pattern complexity of words over ℓ letters

(European J. Combinatorics 27 (2006) pp.125-137)

Teturo KAMAE* and RAO Hui†

Abstract: The maximal pattern complexity function $p^*_\alpha(k)$ of an infinite word $\alpha = \alpha_0\alpha_1\alpha_2\cdots$ over ℓ letters, is introduced and studied by [3],[4].

In the present paper we introduce two new techniques, *the ascending chain of alphabets* and *the singular decomposition*, to study the maximal pattern complexity. It is shown that if $p^*_\alpha(k) < \ell k$ holds for some $k \geq 1$, then α is *periodic by projection.* Accordingly we define a *pattern Sturmian word* over ℓ letters to be a word which is not periodic by projection and has maximal pattern complexity function $p^*_\alpha(k) = \ell k$. Two classes of pattern Sturmian words are given. This generalizes the definition and results of [3] where $\ell = 2$.

1 Introduction

Let A be a finite alphabet. An element of A is called a *letter*. An element $\alpha = \alpha_0\alpha_1\alpha_2\cdots \in A^{\mathbb{N}}$, where $\mathbb{N} := \{0, 1, 2, \cdots\}$, is called a *word over* A, and in particular, it is called a *word over all* A if every letter of A appears in α. We denote $A^* = \cup_{n=0}^{\infty} A^n$ the set of *finite words* over A.

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

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

The k-window τ with $\tau(i) = i$ $(i = 0, 1, \cdots, k-1)$ is called the k*-block window*. For a k-window $\tau :\ 0 = \tau(0) < \tau(1) < \cdots < \tau(k-1)$ and a word α, the word

$$\alpha[n+\tau] \ := \ \alpha_{n+\tau(0)}\alpha_{n+\tau(1)}\cdots\alpha_{n+\tau(k-1)}$$

is the pattern of α through the window τ at position n. We denote by $F_\alpha(\tau)$ the set of all patterns of α through the window τ, i.e.,

$$F_\alpha(\tau) \ := \ \{\alpha[n+\tau];\ n = 0, 1, 2, \cdots\}.$$

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

†Department of Mathematics, Tsinghua University, Beijing, PR China (HRao@math.tsinghua.edu.cn)

In particular, we denote $F_\alpha(k) := F_\alpha(\tau)$ for the k-block window τ.

The *maximal pattern complexity function* p^*_α for a word α is introduced by the first author together with Zamboni [3] as

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

where the supremum is taken over all k-windows τ, while the *block complexity function* p_α is defined as $p_\alpha(k) = \sharp F_\alpha(k)$.

It is known (Morse and Hedlund [5]) that for a word α, the following statements are equivalent:

(i) α is eventually periodic,
(ii) $p_\alpha(k)$ is bounded in k,
(iii) $p_\alpha(k) < k + 1$ for some $k = 1, 2, \cdots$.

The following parallel statements with respect to the maximal pattern complexity function are equivalent ([3]):

(i) α is eventually periodic,
(ii′) $p^*_\alpha(k)$ is bounded in k,
(iii′) $p^*_\alpha(k) < 2k$ for some $k = 1, 2, \cdots$.

A word α with block complexity $p_\alpha(k) = k + 1$ $(k = 1, 2, 3, \cdots)$ is known as a *Sturmian word* and is studied extensively (see for example Berthé [1] and the references therein). A word α with maximal pattern complexity $p^*_\alpha(k) = 2k$ $(k = 1, 2, 3, \cdots)$ is called a *pattern Sturmian word* and is studied in [3].

Let 1_S be the indicator function. A word α over all A is called *periodic by projection* if there exists S with $\emptyset \neq S \subsetneq A$ such that the word

$$1_S(\alpha_0)1_S(\alpha_1)1_S(\alpha_2)\cdots \in \{0, 1\}^{\mathbb{N}}$$

is eventually periodic. Let A_α denote the set of letters occurred in α. If $\sharp A_\alpha = \ell$, we call α a *word over ℓ letters*. The main result of this paper is

Theorem 1.1. *Let α be a word over ℓ letters with $\ell \geq 2$. If $p^*_\alpha(k) < \ell k$ holds for some $k = 1, 2, \cdots$, then α is periodic by projection.*

Note that if there is a letter in A_α which appears in α only finitely often, then α is periodic by projection and Theorem 1.1 holds trivially. So, we may and do always assume that **any letter appearing in α appears infinitely often**.

Theorem 1.1 says that low pattern complexity implies periodic by projection. If a word over ℓ letters is not periodic by projection, then the maximal pattern complexity is at least ℓk. Hence, according to Theorem 1.1, we generalize the definition of pattern Sturmian word in [3]. A word over ℓ letters is called a *pattern Sturmian word* if it is not periodic by projection and has maximal pattern complexity function $p^*_\alpha(k) = \ell k$.

In Section 6, we give two classes of pattern Sturmian words over ℓ letters. One class (generated by an irrational rotation on torus) is recurrent and another class is not. When $\ell = 2$, there is another class which is called Topelitz words, but they are not pattern Sturmian when $\ell > 2$ ([3]). We are interested to know some new examples of pattern Sturmian words.

Note that if $\ell = 2$, then α is periodic by projection if and only if α is eventually periodic. Hence the essential part of the above equivalence that (iii') implies (i) follows from Theorem 1.1.

Let $\alpha = \beta_0 0 \beta_1 0 \beta_2 0 \dots$, where $\beta_0\beta_1\beta_2\dots$ is a classical Sturmian word over all $\{1, 2\}$. Then $A_\alpha = \{0, 1, 2\}$. Let τ be a k-window with $k \geq 2$. Let k_0 (or k_1) be the number of i such that $\tau(i)$ is even (or odd). Note that $p^*_\beta(k) = 2k$. If $k_1 k_2 \neq 0$, then $\sharp F_\alpha(\tau) \leq 2k_0 + 2k_1 = 2k$; otherwise $\sharp F_\alpha(\tau) \leq 2k + 1$. Therefore the maximal pattern complexity of α is $p^*_\alpha(k) \leq 2k + 1 < 3k$. Clearly α is periodic by projection.

In the above example, if the complexity of β is high, then the complexity of α is also high. Hence the inverse of Theorem 1.1 is not true.

The outline of the paper is following.

Recurrent property In this paper we will see that, one of the striking features of the maximal pattern complexity is that it has very strong relation with the recurrent property of the word in consideration. A word $\beta = \beta_0\beta_1\beta_2\cdots$ is called *recurrent* if for any $L = 1, 2, \cdots$, there exists $M \geq 1$ such that

$$\beta_0\beta_1\cdots\beta_{L-1} = \beta_M\beta_{M+1}\cdots\beta_{M+L-1}. \tag{1.1}$$

Note that if β is recurrent, then for any L there exist infinitely many M's which makes (1.1) hold. Moreover, β is called *uniformly recurrent* if for any L, the set of M as above is relatively dense in $\mathbb{N}$ (that is, the distance between two consecutive M is bounded by some constant).

When α is recurrent, the proof of Theorem 1.1 is easy, and this is done in Section 2 (in the proof of Lemma 2.3). There we construct our first graph in this paper.

Singular decomposition When α is not recurrent, the situation is much more complicated. Let T be the *shift* on the space $A^{\mathbb{N}}$ such that $(T\alpha)_n = \alpha_{n+1}$ for $\alpha \in A^{\mathbb{N}}$. The *orbit closure* of α is defined by

$$\overline{O}(\alpha) := \overline{\{T^n\alpha;\ n = 0, 1, 2, \dots\}},$$

where the topology is the product topology on $A^{\mathbb{N}}$. It is well known (K. Petersen [6]) that α is uniformly recurrent if and only if $\overline{O}(\alpha)$ is minimal, where a nonempty T-invariant closed set $\Omega \subset A^{\mathbb{N}}$ is called *minimal* if Ω has no nonempty T-invariant closed proper subset. For a word α, there always exists a recurrent word $\beta \in \overline{O}(\alpha)$, since $\overline{O}(\alpha)$ contains at least one minimal set and any element in a minimal set is recurrent.

A word $\beta \in \overline{O}(\alpha)$ is called an *auxiliary word* of α, if β is recurrent and satisfies an additional technical condition (ii) in Section 4. Our strategy is to make use of β to study the maximal pattern complexity of α. We introduce two new techniques to study a word, *ascending chain of alphabet* and *singular decomposition*.

Let β be an auxiliary word of α. A letter b which appears in β is called a *singular letter* of β if the word

$$1_{\{b\}}(\beta_0)1_{\{b\}}(\beta_1)1_{\{b\}}(\beta_2)\cdots \in \{0,1\}^{\mathbb{N}}$$

is periodic. The minimum m which is a period of $1_{\{b\}}(\beta_0)1_{\{b\}}(\beta_1)1_{\{b\}}(\beta_2)\cdots$ for any singular letter b of β, is called the *decomposition cycle* of β. We set the decomposition cycle of β to be 1 if β has no singular letter. Let

$$\beta^{(i)} = \beta_i\beta_{i+m}\beta_{i+2m}\cdots, \quad i = 0,1,\cdots,m-1. \tag{1.2}$$

Note that b is a singular letter of β if and only if $\beta^{(i)} = b^\infty$ for some $i \in \{0,1,\ldots,m\}$, and b does not appear in any nonconstant word $\beta^{(i)}$. Accordingly, let

$$\alpha^{(i)} = \alpha_i\alpha_{i+m}\alpha_{i+2m}\cdots, \quad i = 0,1,\cdots,m-1. \tag{1.3}$$

The decomposition (1.3) and (1.2) is called the *singular decomposition* of the pair α and β.

Ascending chain of alphabet Usually A_β is a proper subset of A_α. A letter $a \in A_\alpha \setminus A_\beta$ is a neighbor of A_β if and only if a occurs in a bounded distance (of the right side) from arbitrarily large block of α consisting of letters in A_β. In Section 3, we construct a chain

$$A_\beta = A_0 \subsetneq A_1 \subsetneq \cdots \subsetneq A_h = A_\alpha,$$

where $A_j \setminus A_{j-1}$ is the set of neighbors of A_{j-1}. The ascending chain $\{A_j\}$ illustrates the distribution of the letters in α and helps us to find the patterns of a given window in α.

Graph on alphabets Let m be the decomposition cycle of β and let $V = \{0,1,\ldots,m-1\}$. For $u \in V$, let $A_{\alpha^{(u)}}$ be the alphabet of the word $\alpha^{(u)}$. Considering the relations among these alphabets, we define a non-directed graph $\Gamma(V,H)$, where there is an edge in H connecting u and v ($u \neq v$) if and only if $A_{\alpha^{(u)}} \cap A_{\alpha^{(v)}} \neq \emptyset$. It is shown that Γ is connected if α is not periodic by projection. We actually prove a result which is stronger than Theorem 1.1.

Theorem 1.2. *If the graph Γ is connected and $\ell \geq 2$, then $p^*_\alpha(k) \geq \ell k$ holds for any $k = 1,2,\cdots$, where $\ell = \sharp A_\alpha$.*

The paper is organized as follows: In Section 2, we proved Theorem 1.1 in case α is recurrent. In Section 3, we introduce the ascending chain of alphabets. In Section 4, we introduce the singular decomposition of α and establish several lemmas. Theorem 1.1 and Theorem 1.2 are proved in Section 5. In Section 6, two classes of pattern Sturmian words are given.

For the block complexity function p_α, refer [2], [7] and [8].

2 Recurrent words

In this section, we prove Theorem 1.1 in case α is a recurrent word. Actually we will prove the following stronger result.

Theorem 2.1. *Let β be a recurrent word over ℓ letters with $\ell \geq 2$. If $p^*_\beta(k) < \ell k$ holds for some $k = 1, 2, \cdots$, then β contains at least one singular letter (and thus is periodic by projection).*

For $s \geq 1$ and $m \geq 1$, a window $\tau = \{0 = \tau(0) < \tau(1) < \cdots < \tau(k-1)\}$ is called *s-separated* if $\tau(i) - \tau(i-1) \geq s$ $(i = 1, 2, \cdots, k-1)$; is *m-divisible* if m divides every $\tau(i)$. For $m \geq 1$, we define subsequences $\beta^{(i)}$ $(0 \leq i \leq m-1)$ to be

$$\beta^{(i)} = \beta_i \beta_{i+m} \beta_{i+2m} \cdots . \tag{2.1}$$

If τ is a m-divisible window, then

$$F_\beta(\tau) = \cup_{i=0}^{m-1} F_{\beta^{(i)}}(\tau/m),$$

where τ/m is the window $\{0 = \tau(0)/m < \tau(1)/m < \cdots < \tau(k-1)/m\}$.

A family of words $\beta^{(i)}$ $(i = 0, 1, \cdots, m-1)$ are called *simultaneously recurrent* if for any $L = 1, 2, \cdots$, there exists $M > 0$ such that

$$\beta_0^{(i)} \beta_1^{(i)} \cdots \beta_{L-1}^{(i)} = \beta_M^{(i)} \beta_{M+1}^{(i)} \cdots \beta_{M+L-1}^{(i)} \tag{2.2}$$

holds for $i = 0, 1, \cdots, m-1$ simultaneously.

Lemma 2.2. *If β is recurrent, then for any $m \geq 1$, the family of words $\beta^{(i)}$ $(i = 0, 1, \cdots, m-1)$ in* (2.1) *are simultaneously recurrent.*

Proof. To prove the lemma, we need only show that for any $L \geq 1$, there exists a $M \geq 1$ with $m|M$ such that (1.1) holds. Then for $mL \geq 1$, there exists $M \geq 1$ such that

$$\beta_0 \beta_1 \cdots \beta_{mL-1} = \beta_{mM} \beta_{mM+1} \cdots \beta_{mM+mL-1}.$$

So the family of words $\beta^{(i)}$ $(i = 0, 1, \cdots, m-1)$ are simultaneously recurrent by taking M corresponding to L.

Take an arbitrary L_1. Take $M > 0$ satisfying (1.1) for $L = L_1$ and denote this M by M_1. Let $L_2 = L_1 + M_1$. Take $M > 0$ satisfying (1.1) for $L = L_2$ and denote this M byM_2. Then for $L = L_1$, (1.1) holds for

$$M \in \{M_1, M_2, M_1 + M_2\}.$$

In general, let $L_n = L_{n-1} + M_{n-1}$, take $M > 0$ satisfying(1.1) for $L = L_n$ and denote this M by M_n. We obtain a sequence of positive integers $M_1, M_2, M_3, \cdots$. For $L = L_1$, (1.1) holds for $M = M_{i_1} + M_{i_2} + \cdots + M_{i_j}$ for any $1 \leq i_1 < i_2 < \cdots < i_j$.

So there exists $i < j$ such that $M := M_i + M_{i+1} + \cdots + M_j \equiv 0 \pmod{m}$. Hence, (1.1) holds for $L = L_1$ and this M satisfying $m|M$. □

We will prove the following lemma, which we need later. Theorem 2.1 follows immediately from this lemma.

Lemma 2.3. *Let β be a recurrent word containing no singular letter. Then for any $s \geq 1$, $m \geq 1$ and $k \geq 1$, there exists an s-separated, m-divisible k-window τ such that*

$$\sharp F_\beta(\tau) \geq \sharp A_\beta k. \tag{2.3}$$

Proof. We prove the lemma by the induction on k. For $k = 1$, the lemma is clear. Assume that the lemma holds for k, and let $\tau = \{0 = \tau(0) < \tau(1) < \cdots < \tau(k-1)\}$ be an s-separated, m-divisible k-window satisfying (2.3).

Let $\beta^{(i)}$ $(i = 0, 1, \cdots, m-1)$ be the family of words defined by (2.1). Suppose that all the possible patterns of $\beta^{(i)}$ through τ/m appear in the first L positions, i.e.,

$$F_{\beta^{(i)}}(\tau/m) = \{\beta^{(i)}[n + (\tau/m)];\ n = 0, 1, \cdots, L-1\}, \quad i = 0, \ldots, m-1.$$

Since the family $\beta^{(i)}$ are simultaneously recurrent, there exists an integer M with $mM \geq \tau(k-1) + s$ such that

$$\beta_0^{(i)}\beta_1^{(i)} \cdots \beta_{L-1}^{(i)} = \beta_M^{(i)}\beta_{M+1}^{(i)} \cdots \beta_{M+L-1}^{(i)} \quad i = 0, 1, \cdots, m-1.$$

Define a window

$$\tau' = \tau \cup \{mM\}.$$

Then, τ' is a s-separated, m-divisible $(k+1)$-window which is an extension of τ by $\tau'(k) = mM$. We will prove that

$$\sharp F_\beta(\tau') \geq \sharp A_\beta(k+1).$$

We define a directed graph $G = (A_\beta, E)$ on the vertex set A_β, where the set of directed edges is given by

$$E = \{ab \in A_\beta^2;\ a = \beta_n \neq \beta_{n+mM} = b \text{ for some } n \geq 0\}.$$

We prove first that any connected component of G is strongly connected. We decompose β into mM parts as follows:

$$\beta^{<j>} = \beta_j\beta_{j+mM}\beta_{j+2mM}\cdots \quad (0 \leq j \leq mM-1).$$

Note that every $\beta^{<j>}$ can be realized as an infinite path in the graph $G \cup H$, where $H = (A_\beta, \Delta)$ with $\Delta = \{aa;\ a \in A_\beta\}$. Since any $\beta^{<j>}$ $(0 \leq j \leq mM-1)$ is recurrent (Lemma 2.2), any connected component of G is strongly connected.

Second, any connected component of G contains at least two vertices. Suppose that $\{b\}$ is a connected component of G. Then for any $j = 0, 1, \cdots, mM-1$, either $\beta^{<j>} = bbb\cdots$ is a constant word or b does not appear in $\beta^{<j>}$. This implies that $1_{\{b\}}(\beta_0)1_{\{b\}}(\beta_1)1_{\{b\}}(\beta_2)\cdots$ is periodic with period mM. Hence b is a singular letter of β, which contradicts our assumption.

Therefore, for any $b \in A_\beta$, there exists a circle in G which contains b. This implies that $\sharp E$, the number of the edges of G, is not less than $\sharp A_\beta$.

Now by the construction of τ', we have

$$\{\xi_0\xi_1\cdots\xi_k;\ \xi_0\xi_1\cdots\xi_{k-1} \in F_\beta(\tau) \text{ and } \xi_0 = \xi_k\} \subseteq F_\beta(\tau').$$

So

$$\begin{aligned} & \sharp F_\beta(\tau') \\ = \ & \sharp\{\xi_0\xi_1\cdots\xi_k \in F_\beta(\tau');\ \xi_0 = \xi_k\} + \sharp\{\xi_0\xi_1\cdots\xi_k \in F_\beta(\tau');\ \xi_0 \neq \xi_k\} \\ \geq \ & \sharp F_\beta(\tau) + \sharp E \geq \sharp A_\beta k + \sharp A_\beta \geq \sharp A_\beta(k+1), \end{aligned}$$

which completes the proof. □

3 Ascending chain of alphabet

Let $\beta \in \overline{O(\alpha)}$. Denote $A_0 = A_\beta,\ A = A_\alpha$. Then there exists an arbitrary long block in α consisting of letters in A_0.

It is clear that $A_0 \subset A$. Let us assume that $A_0 \neq A$. For any set B with $A_0 \subset B$ and $A \setminus B \neq \emptyset$, we say a letter $a \in A \setminus B$ is a *neighbor* of B in α, if there exists a finite word $\eta \in A^*$ such that for any $n \in \mathbb{N}$, there exists a word $\xi \in B^n$ such that the word $\xi\eta a$ occurs in α. In another word, a is a neighbor of B if and only if a occurs in a bounded distance after an arbitrarily long block consisting of letters in B in α. The set consisting of elements of B together with neighbors of B in α is denoted by $\tilde{B}$ and called the *neighbor set* of B in α. By the assumption that any element in A appears in α infinitely often, we have $B \subsetneq \tilde{B}$.

Define A_1 by $A_1 = \tilde{A}_0$. If $A_1 \subsetneq A$, then define A_2 by $A_2 = \tilde{A}_1$. In this way, we get a chain

$$A_0 \subsetneq A_1 \subsetneq \cdots \subsetneq A_h = A$$

with some integer $h \geq 1$. To be complete, we define $h = 1$ if $A = A_0$.

For $a \in A \setminus A_0$, let $\sigma(a)$ be the minimum value of the length of ηa as above with respect to A_{j-1} and α such that $a \in A_j \setminus A_{j-1}$. Let σ be the maximal value of $\sigma(a)$, where a runs over $A_h \setminus A_0$. We call $\sigma(a)$ the *distance bound* of the letter a and σ the *distance bound* of the chain.

Examples Let $\alpha = 123112311123\cdots = \prod 1^n 23$ and $\beta = 1^\infty$. Then $A_0 = \{1\}, A_1 = \{1,2,3\}$; $\sigma(2) = 1, \sigma(3) = 2$, and $\sigma = 2$.

Let $\alpha = 123112231112223\cdots = \prod 1^n 2^n 3$. If $\beta = 1^\infty$, then $A_0 = \{1\}, A_1 = \{1,2\}, A_2 = \{1,2,3\}$ and $\sigma = 1$. If $\beta = 2^\infty$, then $A_0 = \{2\}, A_1 = \{1,2,3\}$ and $\sigma = 2$.

The ascending chain $\{A_j\}$ illustrates the distribution of the letters and helps us to find the patterns of a given window in α. Lemma 3.1 follows from the definition of A_i directly.

Lemma 3.1. *Let $1 \le j \le h$ and $a \in A_j$. Then, there exists a word η with length less than σ, such that for any $n \ge 0$, there exist $\xi \in A_{j-1}^n$, and $\gamma \in A_j^n$ such that $\xi\eta a\gamma$ occurrs in α.*

Lemma 3.2. *Let $1 \le j \le h$ and $a \in A_j$. Let $a \in A_j \setminus A_{j-1}$, and let τ be an s-separated k-window with $s \ge \sigma$. Then for each $i = 0, 1, \cdots, k-1$, there exists a pattern of the form $\lambda a\zeta \in F_\alpha(\tau)$, where $\lambda \in A_{j-1}^i$ and $\zeta \in A_j^{k-i-1}$.*

All these patterns, where $a \in A_h \setminus A_0$ and $i = 0, 1, \cdots, k-1$, are different from each other and the total number is $(\sharp A_h - \sharp A_0)k$.

Proof. Take $a \in A_j \setminus A_{j-1}$ and let $n \ge \tau(k-1)$. For this a and n, there exist ξ, η and γ satisfy the conditions in Lemma 3.1. We move the window τ on the word $\xi\eta a\gamma$. When the letter a is in position $\tau(i)$, we get a pattern in $F_\alpha(\tau)$ with the expecting form. The pattern consists of letters in A_j, and a is the first letter which does not belongs to A_{j-1}. We say a is the *critical letter* of the pattern. So we get k different patterns with critical letter a.

When a runs over $A_h \setminus A_0$, we obtain $(\sharp A_h - \sharp A_0)k$ different patterns because two patterns either have different critical letters, or have the same critical letters in different positions. □

As an application of Lemma 3.2, we will show that Theorem 1.1 holds in case that α is a nonconstant word containing arbitrarily large blocks of a letter b. It is obvious that α is not periodic by projection in this case.

Theorem 3.3. *If α is nonconstant and contains arbitrarily large blocks of a letter, then $p_\alpha^*(k) \ge \sharp A_\alpha k$ holds for all $k = 1, 2, \cdots$.*

Theorem 3.3 follows immediately from Lemma 3.4, which is needed in Section 5.

Lemma 3.4. *If α is nonconstant and contains arbitrarily large blocks of a letter b, then for any s and $k = 1, 2, \cdots$, there exists an s-separated k-window τ such that*

$$\sharp F_\alpha(\tau) \ge \sharp A_\alpha k.$$

Proof. Note that α is not eventually periodic (recall the assumption that any letter in A_α appears in α infinitely often). The lemma is true for $k = 1$, so we assume $k \ge 2$.

Clearly $\beta := b^\infty \in \overline{O(\alpha)}$. Let

$$\{b\} = A_0 \subsetneq A_1 \subsetneq \cdots \subsetneq A_h = A_\alpha$$

be the ascending chain and σ the corresponding distance bound. We may assume that $s \geq \sigma$. For any $a \in A_h \setminus A_{h-1}$, we claim that

Claim 1. For $k \geq 2$, there exists an s-separated k-window τ such that for each $i = 0, 1, \cdots, k-2$, there are two different patterns $\lambda a \zeta, \lambda a \zeta' \in F_\alpha(\tau)$ with critical letter a in position i. Namely, $\lambda \in A_{h-1}^i$ and $\zeta, \zeta' \in A_h^{k-i-1}$ with $\zeta \neq \zeta'$.

We prove the claim by induction on $k \geq 2$. Assume that the claim holds for $k \geq 2$ with an s-separated k-window τ. We are going to show that Claim 1 holds for $k+1$ as well as.

Set $n = \tau(k-1)$. Note that $n \geq s \geq \sigma$. By Lemma 3.1, there exist η with $|\eta| < \sigma$ and $\xi \in A_{h-1}^{n-|\eta|}$ such that $\xi\eta a$ appears in α infinitely often.

Take integers $n_1 \neq n_2$ such that $\xi\eta a$ is a prefix of both $T^{n_1}\alpha$ and $T^{n_2}\alpha$.

Since α is not eventually periodic, we can find an integer $j > n + s$ such that $\alpha(n_1 + j) \neq \alpha(n_2 + j)$. Define $\tau' = \tau \cup \{j\}$. Then the claim holds for $(k+1)$-window τ' and $i = k-1$ since $\alpha[n_1 + \tau']$ and $\alpha[n_2 + \tau']$ are the required elements $\lambda a \zeta$ and $\lambda a \zeta'$; the claim hold for window τ' and $i = 0, 1, \cdots, k-2$ by the induction hypothesis. Hence the claim holds for $k+1$. To complete the proof of Claim 1, we prove it when $k = 2$. This is done by taking $\xi\eta = \emptyset$ in the above.

Let τ be the window in Claim 1. First there are $(\sharp A_h - \sharp A_0)k$ patterns in $F_\alpha(\tau)$ as described in Lemma 3.2, we denote the set of these patterns by $\mathcal{P}$.

Pick $a \in A_h \setminus A_{h-1}$. For each $i = 0, 1, \cdots, k-2$, by Claim 1, we have at least 2 patterns of the form $\lambda a \zeta$ ($\lambda \in A_{h-1}^i$, $\zeta \in A_h^{k-i-1}$) and at most one of them belongs to $\mathcal{P}$. Therefore, we get $k-1$ additional patterns of τ.

We also have $bbb\cdots$ in $F_\alpha(\tau)$ which is not in the above list. Thus, we have

$$\sharp F_\alpha(\tau) \geq (\sharp A_h - \sharp A_0)k + k = (\sharp A_\alpha - 1)k + k = \sharp A_\alpha k,$$

which completes the proof. □

4 Singular decomposition

Take a recurrent word $\beta \in \overline{O}(\alpha)$. Let m be the decomposition cycle of β, that is, $m \geq 1$ is the smallest integer such that for any singular letter b of β, m is a period of $1_{\{b\}}(\beta_0)1_{\{b\}}(\beta_1)1_{\{b\}}(\beta_2)\cdots \in \{0,1\}^{\mathbb{N}}$. Let

$$\begin{aligned} \alpha^{(i)} &= \alpha_i \alpha_{i+m} \alpha_{i+2m} \cdots, \\ \beta^{(i)} &= \beta_i \beta_{i+m} \beta_{i+2m} \cdots, \end{aligned} \tag{4.1}$$

for $i = 0, 1, \cdots, m-1$. For some technical reason, we wish that $\beta^{(i)} \in \overline{O}(\alpha^{(i)})$ for $i = 0, 1, \cdots, m-1$.

Take $r \in \{0, 1, \cdots, m-1\}$ such that $\beta = \lim_{n\to\infty} T^{mk_n+r}\alpha$ holds for some $k_1 \le k_2 \le \cdots$. Set $\beta' = T^{m-r}\beta$, then it is obvious that

(i) $\beta' \in \overline{O}(\alpha)$ and β' is recurrent;

(ii) $\beta'^{(i)} \in \overline{O}(\alpha^{(i)})$, where m is the decomposition cycle of β'.

A word satisfying (i) and (ii) is called a *auxiliary word* of α. The decomposition (4.1) is called a *singular decomposition* of α.

From now on, we will always use β to denote an auxiliary word of α. Clearly every finite word appearing in $\beta^{(i)}$ appears in $\alpha^{(i)}$. Our strategy is to make use of β to study the maximal pattern complexity of α.

We use the following notations for a set $D \subset \{0, 1, \cdots, m-1\}$ and a window τ:

$$\begin{aligned} A_{\alpha,D} &:= \cup_{i\in D} A_{\alpha^{(i)}} \\ A_{\beta,D} &:= \cup_{i\in D} A_{\beta^{(i)}} \\ F_{\alpha,D}(\tau) &:= \cup_{i\in D} F_{\alpha^{(i)}}(\tau) \\ F_{\beta,D}(\tau) &:= \cup_{i\in D} F_{\beta^{(i)}}(\tau) \end{aligned}$$

In the rest of this section, we will extend the construction of the ascending chain of alphabet in Section 3, and prove several technical lemmas which are needed in next section. The notations are complicated, but the ideas are very simple: we extend the discussion of one word (in Section 3) to a set of finite words.

Let D be a nonempty subset of $\{0, 1, \cdots, m-1\}$, and denote $A(D) = A_{\alpha,D}$, $A_0(D) = A_{\beta,D}$. It is clear that $A_0(D) \subset A(D)$.

Assume that $A_0(D) \ne A(D)$. Let $A_1^{(i)}$ be the neighbor set of $A_0(D)$ in the word $\alpha^{(i)}$, and $\sigma^{(i)}(a)$ be the distance bound for a letter $a \in A_1^{(i)} \setminus A_0(D)$ with respect to $A_0(D)$ and $\alpha^{(i)}$. Let

$$A_1(D) = \cup_{i\in D} A_1^{(i)} \quad \text{and} \quad \sigma(a) = \min_{i\in D} \sigma^{(i)}(a) \ (a \in A_1(D)),$$

where we set $\sigma^{(i)}(a) = \infty$ if $a \notin A_1^{(i)} \setminus A_0(D)$. Note that $A_1(D) \supsetneq A_0(D)$ since any $a \in A(D) \setminus A_0(D)$ appears in some of $\alpha^{(i)}$ infinitely often.

If $A_1(D) \ne A(D)$, then define $A_2(D)$ and the distance bound in the same manner. We can continue this process until we get a chain

$$A_0(D) \subsetneq A_1(D) \subsetneq \cdots \subsetneq A_h(D) = A(D).$$

We denote this h by $h(D)$, and we define the distance bound to be $\sigma(D) = \max\{\sigma(a); a \in A_h(D) \setminus A_0(D)\}$. We define $h(D) = 1$ if $A_0(D) = A(D)$.

Lemma 4.1 is a parallel one to Lemma 3.2, and the proof is also the same.

Lemma 4.1. *Let D be a nonempty subset of $\{0, 1, \cdots, m-1\}$. Let $a \in A_j(D) \setminus A_{j-1}(D)$ for some j with $j \ge 1$, and let τ be an s-separated k-window with $s \ge \sigma(D)$. Then for each $i = 0, 1, \cdots, k-1$, there exists a pattern of the form $\lambda a \zeta \in F_{\alpha,D}(\tau)$, where $\lambda \in A_{j-1}^i(D)$ and $\zeta \in A_j^{k-i-1}(D)$.*

All these elements for $a \in A_h(D) \setminus A_0(D)$ and $i = 0, 1, \cdots, k-1$ are different from each other, and the total number is $(\sharp A_h(D) - \sharp A_0(D))k$, where $h = h(D)$.

Let

$$s_0 = \max\{\sigma(D);\ \emptyset \neq D \subset \{0, 1, \cdots, m-1\}\} \tag{4.2}$$

to be the total distance bound for the pair α and β.

Lemma 4.2. *Let β be an auxiliary word of α, and m the deomposition cycle of β. Then for any $k = 1, 2, \cdots$, any s_0-separated k-window τ, any nonempty $D \subset \{0, 1, \cdots, m-1\}$:*

(i) If B satisfies $A_{\beta,D} \subset B \subset A_{\alpha,D}$, then it holds that

$$\sharp(F_{\alpha,D}(\tau) \setminus B^k) \geq (\sharp A_{\alpha,D} - \sharp B)k. \tag{4.3}$$

(ii) If B satisfies $\emptyset \neq B \subset A_{\alpha,D} \setminus A_{\beta,D}$, then

$$\sharp(F_{\alpha,D}(\tau) \setminus B^k) \geq (\sharp A_{\alpha,D} - \sharp A_{\beta,D} - \sharp B + 1)k. \tag{4.4}$$

Proof. Let $A_i = A_i(D)$ $(i = 0, 1, \cdots, h)$ be the ascending chain for the set D and $h = h(D)$.

Suppose $A_{\beta,D} \subset B \subset A_{\alpha,D}$. Collecting all patterns $\lambda a \zeta$ in the list of Lemma 4.1 with a running over $A_{\alpha,D} \setminus B$, we obtain $(\sharp A_h - \sharp B)k$ patterns in $F_{\alpha,D}(\tau)$. They do not belong to B^k since the critical letters are not in B. This proves (4.3).

Suppose $\emptyset \neq B \subset A_{\alpha,D} \setminus A_{\beta,D}$. Take the smallest j such that $A_j \cap B \neq \emptyset$, then $j \geq 1$. Pick any $b \in A_j \cap B$. From the list of Lemma 4.1, collect all the patterns $\lambda a \zeta$ with $a \notin B$ together with patterns $\lambda b \zeta$ with $\lambda \neq \emptyset$. They are in $F_{\alpha,D}(\tau) \setminus B^k$ and there are $(\sharp A_{\alpha,D} - \sharp A_{\beta,D} - \sharp B)k + k - 1$ of them. Finally there is at least one pattern consisting only of letters in $A_{\beta,D}$; it is in $F_{\alpha,D}(\tau) \setminus B^k$ and it is not in the above list. Hence, we have at least $(\sharp A_{\alpha,D} - \sharp A_{\beta,D} - \sharp B)k + k$ elements in $F_{\alpha,D}(\tau) \setminus B^k$, which proves (4.4). □

We call $i \in \{0, 1, \cdots, m-1\}$ a *singular residue* of β if $\beta^{(i)}$ is a constant word, otherwise we call i a *regular residue* of β. The set of singular residues and the set of regular residues are denoted by D_S and D_R respectively.

Lemma 4.3. *Let β be an auxiliary word of α with $D_R \neq \emptyset$. Then for any $k = 0, 1, \ldots$, there is an s_0-separated k-window τ such that*

$$\sharp F_{\alpha,D_R}(\tau) \geq \sharp A_{\alpha,D_R} k.$$

Proof. First we show that there is an s_0-separated k-window τ such that

$$\sharp F_{\beta,D_R}(\tau) \geq \sharp A_{\beta,D_R} k.$$

Let $D_R = \{j_0 < j_1 < \cdots < j_{p-1}\}$. We construct a new word γ by

$$\gamma_{pk+i} = \beta_k^{(j_i)}, \quad i = 0, 1, \cdots, p-1, \ k = 0, 1, \ldots.$$

Then, it is not difficult to see that γ is a recurrent word containing no singular letter. So by Lemma 2.3, there is an s_0-separated, p-divisible window τ' such that $\sharp F_\gamma(\tau') \geq \sharp A_\gamma k$. Dividing each elements of τ' by p, we obtain a new window $\tau = \tau'/p$, and clearly

$$\sharp F_{\beta,D_R}(\tau) = \sharp F_\gamma(\tau') \geq \sharp A_\gamma k = \sharp A_{\beta,D_R} k.$$

Setting $B = A_{\beta,D_R}$ in Lemma 4.2(i), we have that

$$\sharp(F_{\alpha,D_R}(\tau) \setminus F_{\beta,D_R}(\tau)) \geq (\sharp A_{\alpha,D_R} - \sharp A_{\beta,D_R})k.$$

We obtain the required result by adding the above formulas. □

Corollary 4.4. *If there exists a recurrent $\beta \in \overline{O}(\alpha)$ such that β contains no singular letter, then $p^*_\alpha(k) \geq \sharp A_\alpha k$ holds for any $k = 1, 2, \cdots$.*

5 Graph Γ and the main results

Let β be an auxiliary word of α, and m be the decomposition cycle of β. We define a non-directed graph $\Gamma = \Gamma(V, H)$ with the vertex set $V = \{0, 1, \cdots, m-1\}$ and the edge set H such that $\{u, v\} \in H$ if and only if $u \neq v$ and $A_{\alpha^{(u)}} \cap A_{\alpha^{(v)}} \neq \emptyset$.

Lemma 5.1. *If α is not periodic by projection, then the graph Γ is connected.*

Proof. Suppose that the graph Γ is not connected. Take a connected component U of Γ and denote $S = A_{\alpha,U}$. Then Γ is not connected implies that $\emptyset \neq U \subsetneq V$ and $\emptyset \neq S \subsetneq A_\alpha$. Therefore, the word $1_S(\alpha_0)1_S(\alpha_1)1_S(\alpha_2)\cdots$ is eventually periodic with period m. The lemma is proved. □

Proof of Theorem 1.2 Assume that β is an auxiliary word of α, and the graph Γ defined by α and β is connected.

Let U be a subset of V, and $u \notin U$ be a singular residue such that $\{u, u'\} \in H$ for some $u' \in U$. We assert that if

$$\sharp F_{\alpha,U}(\tau) \geq \sharp A_{\alpha,U} k \tag{5.1}$$

holds for an s_0-separated k-window τ, then $\sharp F_{\alpha,U'}(\tau) \geq \sharp A_{\alpha,U'} k$ holds for $U' = U \cup \{u\}$ with the same window τ.

Since

$$F_{\alpha,U'}(\tau) = F_{\alpha,U}(\tau) \cup F_{\alpha^{(u)}}(\tau),$$

we have that

$$\sharp F_{\alpha,U'}(\tau) \geq \sharp F_{\alpha,U}(\tau) + \sharp(F_{\alpha^{(u)}}(\tau) \setminus B^k), \tag{5.2}$$

where $B = A_{\alpha,U} \cap A_{\alpha^{(u)}}$. Note that B is not empty.

Let $\beta^{(u)} = b^\infty$. By Lemma 4.2 with $D = \{u\}$ and $B = A_{\alpha,U} \cap A_{\alpha^{(u)}}$, we have

$$\sharp(F_{\alpha^{(u)}}(\tau) \setminus B^k) \geq (\sharp A_{\alpha^{(u)}} - \sharp B)k. \tag{5.3}$$

(Use Lemma 4.2 (i) when $b \in B$, use (ii) when $b \notin B$). Since

$$\sharp A_{\alpha,U} + \sharp A_{\alpha^{(u)}} - \sharp B = \sharp(A_{\alpha,U} \cup A_{\alpha^{(u)}}) = \sharp A_{\alpha,U'},$$

we have that $F_{\alpha,U'}(\tau) \geq \sharp A_{\alpha,U'}k$ by (5.1), (5.2) and (5.3). Our assertion is proved.

Recall that D_R is the regular residues of β. If $D_R = \emptyset$, then take $v \in V$ such that $\sharp A_{\alpha^{(v)}} \geq 2$, which exists since $\ell \geq 2$ and the graph Γ is connected. We set $U_0 = \{v\}$ when $D_R = \emptyset$, and $U_0 = D_R$ otherwise. We claim that U_0 satisfies (5.1). If $U_0 = \{v\}$, our claim follows from Lemma 3.4; if $U_0 = D_R$, our claim follows from Lemma 4.3.

Recall that $V = \{0, 1, \cdots, m-1\}$. Since $V \setminus U_0$ contains only singular residues, by adding them to U_0 one by one, we conclude that there exists a k-window τ such that $F_{\alpha,V}(\tau) \geq \sharp A_{\alpha,V}k$.

Define a k-window $m\tau$ to be $\{0 = m\tau(0) < m\tau(1) < \cdots < m\tau(k-1)\}$, then

$$\sharp F_\alpha(m\tau) = \sharp F_{\alpha,V}(\tau) \geq \sharp A_{\alpha,V}k = \sharp A_\alpha k,$$

which implies $p^*_\alpha(k) \geq \sharp A_\alpha k$. □

Theorem 1.1 follows immediately from Lemma 5.1 and Theorem 1.2.

6 Pattern Sturimian words

In this section, we give two classes of pattern Sturmian words.

Example 1. Let θ be an irrational number and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Let $\ell \geq 2$ and $\mathcal{P} = \{I_0, I_1, \cdots, I_{\ell-1}\}$ be a partition of $\mathbb{T}$ into ℓ intervals with nonempty interiors. Define $\alpha = \alpha_0\alpha_1 \cdots \in \{0, 1, \cdots, \ell-1\}^{\mathbb{N}}$ by

$$\alpha_n = i, \quad \text{if } n\theta \in I_i.$$

Then α is not periodic by projection since θ is irrational.

Let $\tau = \{\tau(0) < \tau(1) < \cdots < \tau(k-1)\} = \{0, N_1, \ldots, N_{k-1}\}$ be a k-window. Denote $S - x = \{s - x;\ s \in S\}$. Since $\{n\theta;\ n \in \mathbb{N}\}$ is dense in $\mathbb{T}$, $\xi_0\xi_1 \cdots \xi_{k-1} \in F_\alpha(\tau)$ if and only if

$$I_{\xi_0}\ \cap\ (I_{\xi_1} - N_1\theta)\ \cap \cdots \cap\ (I_{\xi_{k-1}} - N_{k-1}\theta) \neq \emptyset.$$

Therefore,

$$\sharp F_\alpha(\tau) \leq \sharp\ (\mathcal{P}\ \vee\ (\mathcal{P} - N_1\theta)\ \vee \cdots \vee\ (\mathcal{P} - N_{k-1}\theta)), \tag{6.1}$$

where"$\vee$" is the common refinement of partitions. Since the right side of (6.1) is no greater than the number of the end points of the intervals

$$I_i - N_j\theta \quad (i = 0, 1, \cdots, \ell - 1;\ j = 0, 1, \cdots, k - 1),$$

we have $p^*_\alpha(k) \leq \ell k$ $(k = 1, 2, \cdots)$. Since α is not periodic by projection, $p^*_\alpha(k) = \ell k$ $(k = 1, 2, \cdots)$ holds by Theorem 1.1.

Note that the same result holds for any orientation preserving homeomorphism on $\mathbb{T}$ with an irrational rotation number instead of the irrational rotation.

Example 2. Let $\ell \geq 2$. Let $C = \{c_0 < c_1 < c_2 < \cdots\}$ be a set of nonnegative integers such that $2c_j < c_{j+1}$ $(j = 0, 1, 2, \cdots)$. Let $\{C_1, C_2 \cdots C_{\ell-1}\}$ be a partition of C into $\ell - 1$ infinite sets. Let $C_0 = \mathbb{N} \setminus C$. Define $\alpha = \alpha_0\alpha_1 \cdots \in \{0, 1, \cdots, \ell - 1\}^{\mathbb{N}}$ by

$$\alpha_n = i, \quad \text{if } n \in C_i.$$

Then α is not periodic by projection since it contains arbitrarily long block of 0. It is nor recurrent.

For any k-window τ, $F_\alpha(\tau)$ contains 0^k, $0^i a 0^{k-i-1}$ $(i = 0, 1, \cdots, k - 1;\ a = 1, 2, \cdots, \ell - 1)$. Moreover, for any $i = 1, 2, \ldots, k - 1$, $F_\alpha(\tau)$ can contain at most one finite word of the form $\eta a 0^{k-i-1}$, where $a \in \{1, 2, \cdots, \ell - 1\}$ and $\eta \in \{0, 1, \cdots, \ell - 1\}^i \setminus \{0^i\}$. Therefore, $\sharp F_\alpha(\tau) \leq \ell k$, which implies that $p^*_\alpha(k) = \ell k$ $(k = 1, 2, \cdots)$ by Theorem 1.1.

References

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

[2] E. Coven and G.A. Hedlund, Sequences with minimal block growth, *Math. Syst. Theor.,* **7:3** (1973), 138-153.

[3] T. Kamae and L. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words, *Ergodic Theory and Dynamical Systems,* **22:4** (2002), 1191-1199.

[4] T. Kamae and L. Zamboni, Maximal pattern complexity for discrete systems, *Ergodic Theory and Dynamical Systems,* **22:4** (2002), 1201-1214.

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

[6] K. Petersen, Ergodic Theory, Cambridge University Press, Cambridge, 1983.

[7] R. Risley and L. Zamboni, A generalization of Sturmian sequences: combinatorial structure and transcendence, *Acta Arith.*, **95:2** (2000), 167-184.

[8] G. Rote, Sequences with subword complexity $2n$, *J. Number Theory,* **46** (1994), 196-213.