Uniform Sets and Complexity

Teturo KAMAE Matsuyama University, 790-8578 Japan kamae@apost.plala.or.jp

An element $\omega \in \{0,1\}^{\mathbb{N}}$ is called an *infinite* 0-1-word which is a mapping from N to $\{0,1\}$, while it is also considered as an infinite sequence $\omega(0)\omega(1)\omega(2)\cdots$ of 0 and 1. On the other hand, an element u in $\{0,1\}^*:=\bigcup_{k=0}^{\infty}\{0,1\}^k$ is called a *finite* 0-1-word and represented as a finite sequence $u_1u_2\cdots u_k$ of 0 and 1, where k is such that $u\in\{0,1\}^k$, which is called the length of u and is denoted by |u|. We also denote $\{0,1\}^+=\bigcup_{k=1}^{\infty}\{0,1\}^k$.

The concatenation uv or $u\omega$ of $u \in \{0,1\}^*$ with $v \in \{0,1\}^*$ or $\omega \in \{0,1\}^N$ is defined as the word $u_1u_2 \cdots u_k v_1v_2 \cdots v_l$ or $u_1u_2 \cdots u_k \omega(0)\omega(1)\omega(2)\cdots$, where k = |u| and l = |v|, respectively. In this case, u is called a *prefix* of uv or $u\omega$, or equivalently, uv or $u\omega$ is called an *extension* of u.

For $u \in \{0,1\}^*$, the cylinder set [u] determined by u is defined by

$$[u] = \{\omega \in \{0,1\}^{\mathbb{N}}; u \text{ is a prefix of } \omega\}.$$

The prefix tree $G(\Omega) = (V, E)$ of a nonempty closed set $\Omega \subset \{0, 1\}^{\mathbb{N}}$ is defined to be a directed graph such that the set V of vertices is the set of cylinder sets [u] which meet Ω , and the set E of edges is the set of the ordered pairs $([u], [v]) \in V \times V$ such that v is an immediate extension of u, that is, u is the prefix of v such that |v| = |u| + 1.

Two nonempty closed sets Ω , $\Lambda \subset \{0,1\}^N$ are said to be *isomorphic* to each other if their prefix trees are isomorphic to each other. The class of all closed subsets of $\{0,1\}^N$ isomorphic to Ω is denoted by $[\Omega]$ and is called the *language structure* of (or determined by) Ω .

Define

$$\begin{split} \Theta_0 := \{0^{\infty}\} \;, \quad & \Theta_1 := \{1^{\infty}\}, \\ \Theta_{\delta} := \{\omega \in \{0,1\}^{\mathbb{N}}; \; \sum_{n \in \mathbb{N}} \omega(n) \leq 1\}, \\ \Theta_{1-\delta} := \{\omega \in \{0,1\}^{\mathbb{N}}; \; \sum_{n \in \mathbb{N}} (1-\omega(n)) \leq 1\}, \\ \Theta_+ := \{\omega \in \{0,1\}^{\mathbb{N}}; \; \omega \; \text{is increasing}\}, \\ \Theta_- := \{\omega \in \{0,1\}^{\mathbb{N}}; \; \omega \; \text{is decreasing}\}, \end{split}$$

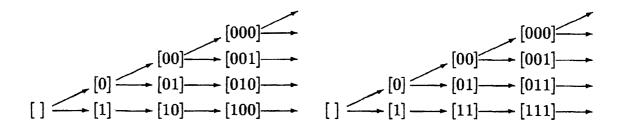


Figure 1: $G(\Theta_{\delta})$ (left) and $G(\Theta_{+})$ (right)

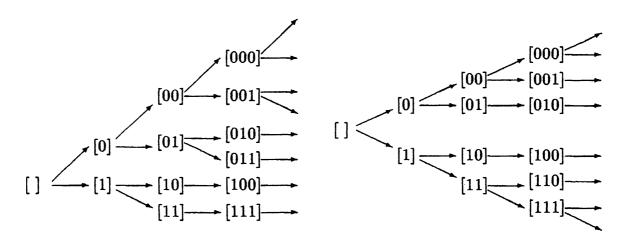


Figure 2: $G(\Theta_{\delta} \cup \Theta_{+})$ (left) and $G(\Theta_{\delta} \cup \Theta_{-})$ (right)

where $a^{\infty} = aaa \cdots$ for $a \in \{0,1\}$ and $\omega \in \{0,1\}^{\mathbb{N}}$ is called *increasing* (decreasing) if $\omega(n) \leq \omega(m)$ ($\omega(n) \geq \omega(m)$, respectively) for any n < m.

All of Θ_{δ} , $\Theta_{1-\delta}$, Θ_{+} , Θ_{-} are isomorphic to each other since for example, $G(\Theta_{\delta})$ and $G(\Theta_{+})$ are isomorphic (Figure 1). It also holds that $\Theta_{\delta} \cup \Theta_{-}$ and $\Theta_{+} \cup \Theta_{-}$ are isomorphic, while $\Theta_{\delta} \cup \Theta_{+}$ is not isomorphic to $\Theta_{\delta} \cup \Theta_{-}$ (Figure 2).

Definition 1. For a nonempty closed set $\Omega \subset \{0,1\}^{\Sigma}$, define the complexity function $p_{\Omega}(S) := \#\pi_S \Omega$, which is a function of finite sets $S \subset \Sigma$, where # denotes the number of elements in a set and $\pi_S : \{0,1\}^{\Sigma} \to \{0,1\}^{S}$ is the projection. We call Ω a uniform set if $p_{\Omega}(S)$ depends only on #S. In this case, the function $p_{\Omega}(k) := p_{\Omega}(S)$ of $k = 1, 2, \cdots$, where #S = k, is called the uniform complexity function of Ω . We also define the maximal pattern complexity function of Ω as $p_{\Omega}^*(k) := \sup_{S; \#S = k} p_{\Omega}(S)$ $(k = 1, 2, \cdots)$. Note that $p_{\Omega}(k) = p_{\Omega}^*(k)$ $(k = 1, 2, \cdots)$ if Ω is a uniform set.

Let $\mathcal{N} = \{N_0 < N_1 < N_2 < \cdots\}$ be an infinite subset of N. For $\omega \in$

 $\{0,1\}^{\mathbb{N}}$ and $\Omega \subset \{0,1\}^{\mathbb{N}}$, define $\omega[\mathcal{N}] \in \{0,1\}^{\mathbb{N}}$ and $\Omega[\mathcal{N}] \subset \{0,1\}^{\mathbb{N}}$ by

$$\omega[\mathcal{N}](n) := \omega(N_n) \quad (n \in \mathbb{N})
\Omega[\mathcal{N}] := \{\omega[\mathcal{N}] \in \{0, 1\}^{\mathbb{N}}; \ \omega \in \Omega\}.$$

We sometimes identify the infinite subset $\mathcal{N} \subset \mathbb{N}$ with an increasing injection $n \mapsto N_n$ from \mathbb{N} into itself. We use the same notation $\Omega[S]$ for a finite set $S = \{s_1 < s_2 < \dots < s_k\} \subset \mathbb{N}$ in the sense that

$$\Omega[S] = \{\omega(s_1)\omega(s_2)\cdots\omega(s_k)\in\{0,1\}^k;\ \omega\in\Omega\}.$$

For $\Omega \subset \{0,1\}^{\Sigma}$, where Σ is a countably infinite set, and an injection $\psi: \mathbb{N} \to \Sigma$, denote

$$\Omega \circ \psi := \{ \omega \circ \psi \in \{0,1\}^{\mathbb{N}}; \ \omega \in \Omega \}.$$

Note that if Ω is a uniform set, then $\Omega \circ \psi$ is also a uniform set with the same complexity function.

Definition 2. A nonempty closed set $\Omega \subset \{0,1\}^{\mathbb{N}}$ is called a *super-stationary* set if $\Omega[\mathcal{N}] = \Omega$ holds for any infinite subset \mathcal{N} of \mathbb{N} . Note that a super-stationary set is a uniform set and all of Θ_0 , Θ_1 , Θ_δ , $\Theta_{1-\delta}$, Θ_+ , Θ_- are super-stationary sets.

Definition 3. A nonempty closed set $\Omega \subset \{0,1\}^{\Sigma}$ is said to have a *primitive* factor $[\Omega \circ \phi]$ if $\Omega \circ \phi$ is a super-stationary set, where $\phi : \mathbb{N} \to \Sigma$ is an injection and $[\Omega \circ \phi]$ is the language structure determined by $\Omega \circ \phi$.

Definition 4. Let $\Omega \subset \{0,1\}^{\mathbb{N}}$ be a nonempty closed set. For $\omega \in \Omega$ and $k \in \mathbb{N}$, we denote $\omega|_k = \omega(0)\omega(1)\cdots\omega(k-1) \in \{0,1\}^k$. Let Ω' be the set of accumulating points of Ω , that is,

$$\Omega' = \{ \omega \in \Omega; \ \#([\omega|_k] \cap \Omega) = \infty \text{ for any } k \in \mathbb{N} \}.$$

We call Ω' the *derived set* of Ω . Clearly, Ω' is a closed set (possibly, the empty set). We denote $\Omega^{(0)} = \Omega$ and $\Omega^{(i)} = (\Omega^{(i-1)})'$ for $i = 1, 2, \cdots$. The degree of Ω is defined to be $d = 0, 1, 2, \cdots$ such that $\Omega^{(d)} \neq \emptyset$ and $\Omega^{(d+1)} = \emptyset$, if such d exists, otherwise, ∞ . The degree of Ω is denoted by deg Ω . For completeness, we define $\emptyset' = \emptyset$ and deg $\emptyset = -1$.

We have the following results.

Theorem 5. (Kamae [1]) Let Ω be a nonempty closed subset of $\{0,1\}^{\Sigma}$, where Σ is a countably infinite set.

- (1) If there exists an injection $\psi : \mathbb{N} \to \Sigma$ such that $\deg(\Omega \circ \rho) < \infty$, then there exists an increasing injection $\phi : \mathbb{N} \to \mathbb{N}$ such that $\Omega \circ \phi$ is a superstationary set.
- (2) If $deg(\Omega \circ \rho) = \infty$ for any injection $\rho : \mathbb{N} \to \Sigma$, then $p_{\Omega}^*(k) = 2^k$ $(k = 1, 2, \cdots)$.

Hence, any uniform set has a primitive factor and any uniform complexity function is realized by a super-stationary set.

Remark 6. (1) of the Main Theorem can be generalized easily to the case of general finite alphabet.

For $\xi = \xi_1 \xi_2 \cdots \xi_k \in \{0,1\}^k$ and $\eta = \eta_1 \eta_2 \cdots \eta_l \in \{0,1\}^l$ with $k \leq l$, we say that ξ is a *super-subword* of η , if $\xi = \eta_{s_1} \eta_{s_2} \cdots \eta_{s_k}$ holds for some $1 \leq s_1 < s_2 < \cdots < s_k \leq l$. For this ξ and $\omega \in \{0,1\}^N$, we say that ξ is a *super-subword* of ω , if $\xi = \omega(s_1)\omega(s_2)\cdots\omega(s_k)$ holds for some $0 \leq s_1 < s_2 < \cdots < s_k < \infty$. In these cases, we denote $\xi \ll \eta$ or $\xi \ll \omega$.

For $\xi \in \{0,1\}^*$, denote

$$\mathcal{P}(\xi) := \{ \omega \in \{0,1\}^{\mathbb{N}}; \ \xi \ll \omega \text{ does not hold} \},$$

that is, $\mathcal{P}(\xi)$ is the set of infinite 0-1-words with the prohibited word ξ as its super-subword. Denote for $\Xi \subset \{0,1\}^*$,

$$Q(\Xi) := \bigcup_{\xi \in \Xi} \mathcal{P}(\xi) \text{ and } \mathcal{P}(\Xi) := \bigcap_{\xi \in \Xi} \mathcal{P}(\xi).$$

We call $\eta \in \{0,1\}^* \cup \{0,1\}^{\mathbb{N}}$ a cover of Ξ if $\xi \ll \eta$ holds for any $\xi \in \Xi$. It is called a *minimal cover* if in addition, any $\zeta \leq \eta$ is not a cover of Ξ . Let $L(\Xi)$ be the set of minimal covers of Ξ .

Theorem 7. (T. Kamae, H. Rao, B. Tan and Y-M. Xue [2]) (1) The class of super-stationary sets other than $\{0,1\}^{\mathbb{N}}$ coincides with the class of sets $\mathcal{Q}(\Xi)$ with nonempty finite sets $\Xi \subset \{0,1\}^+$. It also coincides with the class of sets $\mathcal{P}(L(\Xi))$ with nonempty finite sets $\Xi \subset \{0,1\}^+$.

(2) The complexity function $p_{\Omega}(k)$ of a super-stationary set Ω other than $\{0,1\}^{\mathbb{N}}$ is a polynomial function of k for large k.

The following Corollary follows from abov 2 threorems.

Corollary 8. The complexity function $p_{\Omega}(k)$ of a uniform set Ω is either 2^k $(k = 1, 2, \cdots)$ or a polynomial function of k for large k.

Let X be a metrizable space with a continuous group or semi-group action G. For a family of subsets A_1, A_2, \dots, A_k of X, let $\mathbb{P}(\{A_i; i=1,2,\dots,k\})$ denote the *partition* of X generated by these subsets, that is, the family of nonempty sets of the form

$$A_1^{i_1} \cap A_2^{i_2} \cap \cdots \cap A_k^{i_k} \quad (i_1, i_2, \cdots, i_k \in \{0, 1\}),$$

where for a set $A \subset X$, we denote $A^1 = A$ and $A^0 = X \setminus A$.

Let D be a nonempty subset of X. Define the maximal pattern complexity function $p_{X,G,D}^*$ of the triple (X,G,D) by

$$p_{X,G,D}^*(k) = \sup_{\tau \subset G, \ \#\tau = k} \#\mathbb{P}(\{\sigma^{-1}D; \ \sigma \in \tau\}) \ (k = 1, 2, \cdots). \tag{1}$$

Definition 9. For a set U and $k \in \mathbb{N}$, $\mathcal{F}_k(U)$ denotes the family of sets $S \subset U$ with #S = k. A countably infinite subset Σ of G is called an *optimal position* of the triple (X, G, D) if

$$\#\mathbb{P}(\{\sigma^{-1}D; \ \sigma \in \tau\}) = p_{X,G,D}^*(k), \tag{2}$$

holds for any $k = 1, 2, \dots$ and $\tau \in \mathcal{F}_k(\Sigma)$.

Let $\Sigma \subset G$ be a countably infinite set. We call $\omega \in \{0,1\}^{\Sigma}$ a name of the partition $\mathbb{P}(\{\sigma^{-1}D; \sigma \in \Sigma\})$ if there exists $x \in X$ such that

$$\omega(\sigma) = \begin{cases} 1 & x \in \sigma^{-1}D \\ 0 & x \notin \sigma^{-1}D. \end{cases}$$
 (3)

The closure of the set of names of the partition $\mathbb{P}(\{\sigma^{-1}D; \sigma \in \Sigma\})$ is called the *name set* of Σ with respect to the triple (X, G, D).

The following theorem is clear from the definitions.

Theorem 10. The name set of any optimal position Σ of a triple (X, G, D) is a uniform set with the uniform complexity function $p_{X,G,D}^*$.

Example 11. Let $X = G = \mathbb{R}/\mathbb{Z}$. The action of $g \in G$ maps $x \in X$ to $x+g \in X$. Let D be an interval [a,b) in X such that a < b < a+1. Then, we have $p_{D,G}^*(k) = 2k$ $(k = 1, 2, \cdots)$. In this case, a countably infinite subset Σ of G is an optimal position of (X, G, D) if and only if for any $\sigma, \sigma' \in \Sigma$ with $\sigma \neq \sigma'$, $D - \sigma$ and $D - \sigma'$ intersect as well as their complements. This is also equivalent to that $\#\mathbb{P}(\{\sigma^{-1}D, \sigma'^{-1}D\}) = 4$ for any $\{\sigma, \sigma'\} \in \mathcal{F}_2(\Sigma)$.

Let Ω be the name set of an optimal position Σ . Then, Ω is known to have the unique primitive factor $[\Theta_{\delta} \cup \Theta_{-}] = [\mathcal{Q}(11,01)]$ ([?]).

Example 12. Let $X = \mathbb{R}^2$ and $G = (\mathbb{R}/2\pi\mathbb{Z}) \times \mathbb{R}^2$. The action of $(\theta, (u, v))$ in G maps $(x, y) \in X$ to the following $(x', y') \in X$:

$$\begin{cases} x' = x \cos \theta - y \sin \theta + u \\ y' = x \sin \theta + y \cos \theta + v. \end{cases}$$

Let D be a line in X. Then, $g^{-1}D$ is also a line for any $g \in G$ and we have $p_{X,G,D}^*(k) = (1/2)k^2 + (1/2)k + 1$ $(k = 1, 2, \cdots)$. In this case, Σ is an optimal position if and only if Σ is a countably infinite subset of G such that

- (1) for any $\sigma, \sigma' \in \Sigma$ with $\sigma \neq \sigma', \sigma^{-1}D \cap \sigma'^{-1}D \neq \emptyset$, and
- (2) for any $\sigma, \sigma', \sigma'' \in \Sigma$ which are different each other,

$$\sigma^{-1}D \cap \sigma'^{-1}D \cap \sigma''^{-1}D = \emptyset.$$

Let Ω be the name set of an optimal position Σ . Then,

$$\Omega = \{\omega \in \{0,1\}^{\Sigma}; \sum_{\sigma \in \Sigma} \omega(\sigma) \le 2\}.$$

Hence, Ω has the unique primitive factor [Q(111)].

Example 13. (Y-M. Xue [5]) Let $X = G = \mathbb{R}^2$. The action of $g = (g_1, g_2) \in G$ maps $(x, y) \in X$ to $(x + g_1, y + g_2) \in X$. Let $D := \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \le 1\}$ be the unit disk. Then, we have $p_{X,G,D}^*(k) = k^2 - k + 2 \ (k = 1, 2, \cdots)$. In this case, a countably infinite subset Σ of G is an optimal position if and only if $\#\mathbb{P}(\{\sigma^{-1}D; \sigma \in \tau\}) = p_{X,G,D}^*(3) = 8$ for any $\tau \in \mathcal{F}_3(\Sigma)$. Moreover, Σ satisfies this condition if $\Sigma \subset \{g \in G; g_1^2 + g_2^2 = r^2\}$ with 0 < r < 1. Moreover, the name set Ω has a unique primitive factor $[\mathcal{Q}(101,010)]$.

All the examples so far admit a finitely determined optimal position. The following example does not admit an optimal position.

Example 14. Let $X = \mathbf{T}_1 \cup \mathbf{T}_2$ and $G = \mathbf{T}_1 \times \mathbf{T}_2$, where $\mathbf{T}_i \cong \mathbb{R}/\mathbb{Z}$ (i = 1, 2) and $\mathbf{T}_1, \mathbf{T}_2$ are disjoint each other. The action of $g = (g_1, g_2) \in G$ maps $x \in \mathbf{T}_i$ to $x + g_i \in \mathbf{T}_i$ for i = 1, 2. Let $D = [a_1, b_1) \cup [a_2, b_2)$, where $[a_i, b_i) \subset \mathbf{T}_i$ and $a_i < b_i < a_i + 1$ for i = 1, 2.

Then, we have $p_{X,G,D}^*(k) = 4k - 4$ $(k = 2, 3, \cdots)$. In this case, there is no optimal position since for any infinite subset Σ of G, there exists a sequence $g_n = (g_{n,1}, g_{n,2}) \in \Sigma$ for $n = 1, 2, \cdots$ such that $g_{n,i}$ converges monotonously to, say $c_i \in \mathbb{T}_i$, for i = 1, 2. Then, for any sufficiently large n_0 , $\#\mathbb{P}(\{g_n^{-1}D; n = n_0 + 1, n_0 + 2, n_0 + 3\}) = 6$ but not 8.

Definition 15. A nonempty closed set $\Omega \subset \{0,1\}^{\mathbb{N}}$ is called a *stationary* set if $T\Omega = \Omega$, where $T: \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ is the shift. Note that a super-stationary set is always stationary since $T\Omega = \Omega[\{1,2,\cdots\}]$. We call $\mathcal{N} = \{N_0 < N_1 < N_2 < \cdots\} \subset \mathbb{N}$ an optimal window of Ω if $p_{\Omega}(S) = p_{\Omega}^*(k)$ for any $k = 1, 2, \cdots$ and $S \subset \mathcal{N}$ with #S = k.

Take a stationary set $\Omega \subset \{0,1\}^{\mathbb{N}}$ as X and the additive semi-group \mathbb{N} as G. Let the action of $n \in \mathbb{N}$ to $\omega \in \Omega$ be $T^n\omega$. Let $D = \{\omega \in \Omega; \ \omega(0) = 1\}$. In this case, it is easy to see that

Theorem 16. For an infinite subset \mathcal{N} of \mathbb{N} , \mathcal{N} is an optimal position of (Ω, \mathbb{N}, D) if and only if \mathcal{N} is an optimal window of Ω .

Hence, the following theorem follows from Theorem 4.1 of T. Kamae, H. Rao, B. Tan, Y-M. Xue [3].

Theorem 17. Let $\alpha \in \{0,1\}^{\mathbb{N}}$ be a recurrent pattern Sturmian word. Let $X = \overline{O}(\alpha)$, $G = \{T^n; n \in \mathbb{N}\}$ and $D = \{\omega \in \Omega; \omega(0) = 1\}$. Then, an optimal position of the triple (X, G, D) exists.

Example 18. Let $\Omega = \overline{O}(\alpha)$ with the non-simple Toeplitz word $\alpha \in \{0, 1\}^N$ defined in Example 3 in N. Gjini, T. Kamae, B. Tan, and Y.-M. Xue [4]. Then, $p_{\Omega}^*(k) = 2^k$ $(k = 1, 2, \cdots)$ holds. In this case, an optimal window does not exist. Take an arbitrary $\mathcal{N} = \{N_0 < N_1 < N_2 < \cdots\} \subset \mathbb{N}$. For any $k \in \mathbb{N}$, there exists $K \in \mathbb{N}$ with $K \geq k$ and $\xi \in \{0, 1\}^K$ such that $\alpha = (\xi a_0)(\xi a_1)(\xi a_2)\cdots$ holds with $a_0, a_1, a_2 \cdots \in \{0, 1\}$. There exists such

K with the property that there exist 3 elements in \mathcal{N} , say $N_u < N_v < N_w$ with $N_u \not\equiv N_v \equiv N_w$ modulo K+1. Then, either 001 or 101 is not in $\Omega[\{N_u, N_v, N_w\}]$. Hence, \mathcal{N} is not an optimal window.

The following is the list of the language structures and the complexity functions with degree ≤ 1 .

- (1) $[\Theta_0] = [Q(1)], p_{\Omega}(k) = 1,$
- (2) $[\Theta_0 \cup \Theta_1] = [\mathcal{Q}(0,1)], \ p_{\Omega}(k) = 2,$
- (3) $[\Theta_{\delta}] = [\mathcal{Q}(11)], \ p_{\Omega}(k) = k+1,$
- (4) $[\Theta_{\delta} \cup \Theta_1] = [\mathcal{Q}(11,0)], \ p_{\Omega}(k) = k+2-1_{k=1}$
- (5) $[\Theta_{\delta} \cup \Theta_{+}] = [\mathcal{Q}(11, 10)], \ p_{\Omega}(k) = 2k,$
- (6) $[\Theta_{\delta} \cup \Theta_{-}] = [\mathcal{Q}(11,01)], \ p_{\Omega}(k) = 2k,$
- (7) $[\Theta_{\delta} \cup \Theta_{1-\delta}] = [\mathcal{Q}(11,00)], \ p_{\Omega}(k) = 2k + 2 2 \cdot 1_{k=1},$
- (8) $[\Theta_{\delta} \cup \Theta_{+} \cup \Theta_{-}] = [\mathcal{Q}(11, 10, 01)], \ p_{\Omega}(k) = 3k 2 + 1_{k=1}$
- $(9) \left[\Theta_{\delta} \cup \Theta_{1-\delta} \cup \Theta_{+}\right] = [\mathcal{Q}(11,00,10)], \ p_{\Omega}(k) = 3k 1 1_{k=2},$
- $(10) \left[\Theta_{\delta} \cup \Theta_{1-\delta} \cup \Theta_{+} \cup \Theta_{-} \right] = \left[\mathcal{Q}(11, 00, 10, 01) \right], \ p_{\Omega}(k) = 4k 4 + 2 \cdot 1_{k=1}.$

References

- [1] T. Kamae, Uniform sets and complexity, Discrete Math. (to appear).
- [2] T. Kamae, H. Rao, B. Tan, Y.-M. Xue, Super-stationary set, subword problem and the complexity (preprint).
- [3] T. Kamae, H. Rao, B. Tan, Y.-M. Xue, Language Structure of Pattern Sturmian Word, *Discrete Math.* 306 (2006), pp. 1651-1668.
- [4] N. Gjini, T. Kamae, B. Tan, Y.-M. Xue, Maximal pattern complexity for Toeplitz words, *Ergod. Th. & Dynam. Sys.* **26** (2006), pp. 1-14.
- [5] Y.-M. Xue, Transformations with discrete spectrum and sequenceentropy, Master Thesis at Osaka City University, 2000 (in Japanese).