Modified complexity and $*$-Sturmian word

By Izumi Nakashima,*) Jun-ichi Tamura,**) and Shin-ichi Yasutomi***)

We give analogies of the complexity $p(n)$ and Sturmian words which are called the $*$-complexity $p_*(n)$ and $*$-Sturmian words. We announce theorems about $*$-Sturmian words in this paper. The proofs and details will be published elsewhere. We consider words over an alphabet $L = \{0, 1\}$. Let L^n be the set of all words of length $n \geq 0$, $L^0 = \{\lambda\}$, λ is the empty word. Let L^* be the set of all finite words and $L^{\boldsymbol{N}}$ (resp. $L^{-\boldsymbol{N}}$) be the set of right-sided (resp. left-sided) infinite words. A two-sided infinite words $W \in L^{\boldsymbol{Z}}$ is defined to be a map $W : \boldsymbol{Z} \to L$. We identify two words $V, W \in L^{\boldsymbol{Z}}$ if $V(x+y) = W(x)$ for all $x \in \boldsymbol{Z}$ for some fixed $y \in \boldsymbol{Z}$. We put $L^\wedge = L^* \cup L^{\boldsymbol{N}} \cup L^{-\boldsymbol{N}} \cup L^{\boldsymbol{Z}}$. We denote the set of all subwords of W by $D(W)$. We put $D(n; W) := D(W) \cap L^n$ $(n \geq 0)$. The complexity of a word W is a function defined by

$$p(n) = p(n; W) := \sharp D(n; W).$$

A $*$-subword w of W is a word $w \in D(W)$ which occurs infinitely many times in W. We put $D_*(n; W) := D_*(W) \cap L^n$, where $D_*(W)$ is the set of $*$-subwords of W. We define $*$-complexity

$$p_*(n) = p_*(n; W) := \sharp D_*(n; W).$$

A Sturmian word is defined to be a word $W \in L^{\boldsymbol{N}} \cup L^{-\boldsymbol{N}} \cup L^{\boldsymbol{Z}}$ satisfying

$$|\xi(A) - \xi(B)| \leq 1$$

for any $A, B \in D(n; W)$ for all $n \geq 0$, where $\xi(w)$ denotes the number of occurrences of a symbol 1 appearing in a word $w \in L^*$, cf. [2]. We define a $*$-Sturmian word to be a word $W \in L^{\boldsymbol{N}} \cup L^{-\boldsymbol{N}} \cup L^{\boldsymbol{Z}}$ satisfying

$$|\xi(A) - \xi(B)| \leq 1$$

for any $A, B \in D_*(n; W)$ for all $n \geq 0$. Let $\sigma(n; W) = \max_{A \in D(n;W)} \xi(A)$ and $\sigma'(n; W) = \min_{A \in D(n;W)} \xi(A)$.

Theorem 1 (Morse and Hedlund [2]). *If W is a Sturmian word, then $p(n; W) \leq n+1$, and there is the density* $\alpha = \lim_{n\to\infty} \dfrac{\sigma(n, W)}{n} = \lim_{n\to\infty} \dfrac{\sigma'(n, W)}{n}$.

We can classify one-sided or two-sided infinite Sturmian words as follows:

(Type I) α is irrational,

(Type II) α is rational and W is purely periodic,

(Type III) α is rational and W is not purely periodic.

It is known that each case can occur. The words of Type III will be referred to as skew Sturmian words. Let $0 \leq \alpha \leq 1$ and β be real numbers. We define $G(n, \alpha, \beta) = \lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\alpha + \beta \rfloor$ and $G'(n, \alpha, \beta) = \lceil (n+1)\alpha + \beta \rceil - \lceil n\alpha + \beta \rceil$, where $\lfloor x \rfloor$ is the greatest integer which does not exceed x and $\lceil x \rceil$ is the least integer which is not smaller than x. A word $G(\alpha, \beta) \in L^{\boldsymbol{N}}$ is defined by

$$G(\alpha, \beta) = G(0, \alpha, \beta) G(1, \alpha, \beta) \cdots G(n, \alpha, \beta) \cdots.$$

$G'(\alpha, \beta)$ is defined similarly by using $G'(n, \alpha, \beta)$. We set $G(\alpha) = G(\alpha, 0)$, $G'(\alpha) = G'(\alpha, 0)$, $G(n, \alpha) = G(n, \alpha, 0)$ and $G'(n, \alpha) = G'(n, \alpha, 0)$.

Theorem 2 (Morse and Hedlund [2]). *If α is irrational* (*resp. rational*), *then $G(\alpha, \beta)$ and $G'(\alpha, \beta)$ are Sturmian words of Type I* (*resp. TypeII*). *Conversely, if $W \in L^{\boldsymbol{N}}$ is a Sturmian word of type I with density* $\alpha = \lim_{n\to\infty} \dfrac{\sigma(n, W)}{n}$, *there exists a real number β such that $W = G(\alpha, \beta)$ or $W = G'(\alpha, \beta)$.*

For A, $B \in L^*$ we denote by $\{A, B\}^*$ the set

$$\{A, B\}^* := \{w_1 \cdots w_n; w_i = A \text{ or } B\ n \geq 0\}.$$

We say a word $W \in \{a, b\}^*$ is strictly over $\{a, b\}$ if both a and b eventually occur in W. w^* (resp. ${}^*w$) $(\lambda \neq w \in L^*)$ denote the words $w^* := www\cdots \in L^{\boldsymbol{N}}$ (resp. ${}^*w := \cdots www \in L^{-\boldsymbol{N}}$), w^n $(n \in \boldsymbol{N} \cup \{0\},\ w \in L^*)$ is the word $w^n := v_1 v_2 \cdots v_n$ $(v_i = w)$. We mean by ${}^*vw$ (resp. vw^*) the word $({}^*v)w$ (resp. $v(w^*)$).

Theorem 3 (Morse and Hedlund [2]). *Let $W \in L^{\boldsymbol{N}}$ be a purely periodic Sturmian word with*

*) General Education, Gifu National College of Technology, 2236-2 Kamimakuwa, Shinsei-cho, Motosu-gun, Gifu 501–0495.

**) Faculty of General Education, International Junior College, 4-15-1 Ekoda, Nakano-ku, Tokyo 165–0022.

***) General Education, Suzuka National College of Technology, Shiroko, Suzuka, Mie 510–0294.

density $\alpha = p/q$ ($p \in \boldsymbol{N}$, $q > 1$, and $(p,q) = 1$). Then W can be extended in two ways to a two-sided infinite skew Sturmian word which is represented by ${}^*ACB^*$ ($A, B, C \in L^q$ with $\xi(A) = \xi(B) = p$, and $\xi(C) = p-1$ or $p+1$). If the density of a one-sided infinite Sturmian word W is 0 or 1, then W can be uniquly extended to a two-sided infinite skew Sturmian word.

If $x \neq 0, 1$ is rational, then $G(x)$ is purely periodic and there are two extensions to a two-sided infinite skew Sturmian word which is denoted by $\overline{G}(x)$ (resp. $\underline{G}(x)$) if $\xi(C) = p+1$ (resp. $\xi(C) = p-1$). If $x = 0$ (resp. $x = 1$), then $G(x)$ can be extended to a two-sided infinite skew Sturmian word which is denoted by $\overline{G}(x)$ (resp. $\underline{G}(x)$).

Definition 1 (super Bernoulli word, cf. [3]). If $W \in L^{\boldsymbol{N}} \cup L^{-\boldsymbol{N}} \cup L^{\boldsymbol{Z}}$ satisfies one of the following conditions (C1)–(C4), we call W a super Bernoulli word related to (x, y), $0 \leq x \leq y \leq 1$:

(C1) $D_*(W) = \bigcup_{z \in [x,y]} D(G(z))$.

(C2) $D_*(W) = \bigcup_{z \in [x,y]} D(G(z)) \bigcup D(\underline{G}(x))$ with $x \in \mathbf{Q}$.

(C3) $D_*(W) = \bigcup_{z \in [x,y]} D(G(z)) \bigcup D(\overline{G}(y))$ with $y \in \mathbf{Q}$.

(C4) $D_*(W) = \bigcup_{z \in [x,y]} D(G(z)) \bigcup D(\underline{G}(x)) \bigcup D(\overline{G}(y))$ with $x, y \in \mathbf{Q}$.

The converse of the assertion given in Theorem 1 dose not hold, but the words $W \in L^\wedge$ satisfying $p(n; W) \leq n+1$ for all $n \in \boldsymbol{N}$ are characterized by Coven and Hedlund [1].

We need some definitions.

Definition 2. We define substitutions δ_0, δ_1 by

$$\delta_0 : \begin{cases} 0 \to 0 \\ 1 \to 01 \end{cases}, \quad \delta_1 : \begin{cases} 0 \to 01 \\ 1 \to 1 \end{cases}.$$

δ_k can be extended to $L^\wedge$ by

$$\delta_k(W) := \cdots \delta_k(w_i) \cdots$$

for $W = \cdots w_i \cdots \in L^\wedge$. The map $\delta_k : L^\wedge \to L^\wedge$ is injective. Hence we can write $B = \delta_k^{-1}(A)$ if $A = \delta_k(B)$, $(A, B \in L^\wedge)$.

Definition 3. For $k_1, \ldots, k_i \in \{0, 1\}$, we define $A_i = A(k_1, \ldots, k_i) := \delta_{k_1} \circ \cdots \circ \delta_{k_i}(0)$, $B_i = B(k_1, \ldots, k_i) := \delta_{k_1} \circ \cdots \circ \delta_{k_i}(1)$ ($A_0 := 0$, $B_0 := 1$).

Theorem 4. *Let $W \in L^{\mathbf{N}}$. Then the following four conditions are equivalent*:

(i) *W is $*$-Sturmian.*

(ii) *$p_*(n; W) \leq n+1$ for all $n \geq 0$.*

(iii) *There exists a finite or infinite sequence $\kappa = \{k_1, k_2, \ldots, k_i \ldots\}$ $k_i \in \{0, 1\}$ such that*

$$W = u_0 u_1 \cdots u_i \cdots, \quad u_0 A_i^*, \quad or \quad u_0 B_i^*,$$

where $A_i = A(k_1, \cdots, k_i)$, $B_i = B(k_1, \cdots, k_i)$ are words given in Definition 3, $u_0 \in L^*$, and each u_i is a certain finite word strictly over $\{A_i, B_i\}$ for all $i > 0$.

(iv) *W is a super Bernoulli words which satisfies one of the conditions* (C1), (C2) *or* (C3) *in Definition 1 with $x = y$.*

Remark 1. In the condition (iii), if $p_*(m; W) = m+1$ for any m, then $W = u_0 u_1 \cdots u_i \cdots$. If $p_*(m; W) < m+1$ for some m, then $W = u_0 A_i^*$ or $u_0 B_i^*$ and $p_*(n; W)$ is bounded. In the condition (iv), if $x (= y)$ is an irrational number, or W satisfies the conditions (C2) or (C3) in Definition 1, then $p_*(n; W) = n+1$ for all n. If x is a rational number and W satisfies the condition (C1) in Definition 1, then $p_*(n; W)$ is bounded.

Theorem 5. *Let $W \in L^{\mathbf{Z}}$. Then the following three conditions are equivalent*:

(i) *W is $*$-Sturmian.*

(ii) *There exist a finite or infinite sequence $\kappa = \{k_1, k_2, \ldots, k_i, \ldots\}$, $k_i \in \{0, 1\}$ such that W has one of the following representations,*

1) *$W = \cdots u_{-i} \cdots u_{-1} u_0 u_1 \cdots u_i \cdots$, κ is an infinite sequence,*
2) *$W = \cdots u_{-i} \cdots u_{-1} u_0 A_j^*$, κ is infinite and $k_i = 0$ for all $i > j$,*
3) *$W = {}^*A_j u_0 u_1 \cdots u_i \cdots$, κ is infinite and $k_i = 0$ for all $i > j$,*
4) *$W = \cdots u_{-i} \cdots u_{-1} u_0 B_j^*$, κ is infinite and $k_i = 1$ for all $i > j$,*
5) *$W = {}^*B_j u_0 u_1 \cdots u_i \cdots$, κ is infinite and $k_i = 1$ for all $i > j$,*
6) *$W = {}^*A_j u_0 A_j^*$, κ is finite and k_j is its final term and*
7) *$W = {}^*B_j u_0 B_j^*$, κ is finite and k_j is its final term,*

where $A_i = A(k_1, \cdots, k_i)$, $B_i = B(k_1, \cdots, k_i)$ are words given in definition 3, $u_0 \in L^*$, and u_i and u_{-i} are certain finite words strictly over $\{A_i, B_i\}$ for $i > 0$.

(iii) *W is a super Bernoulli word which satisfies one of the conditions* (C1), (C2) *or* (C3) *in Definition 1 with $x = y$.*

Theorem 6. *Let $W \in L^{\mathbf{Z}}$ be a $*$-Sturmian* *word. Then,* $p_*(n;W) \le n+1$ *for all* $n \ge 0$.

Theorem 7. *Let* $W \in L^{\mathbf{Z}}$. *Suppose that* $p_*(n;W) \le n+1$ *for all* $n \ge 0$ *and* W *is not a* $*$-*Sturmian word. Then, there exists a finite sequence* $\{k_i\}_{i=1}^{j}$ *such that*

$$W = {}^*A_j u_0 B_j^*, \quad or \quad {}^*B_j u_0 A_j^*,$$

where $u_0 \in L^*$, $A_j = A(k_1, \cdots, k_j)$, $B_j = B(k_1, \cdots, k_j)$ *are words given in Definition* 3.

Let us consider the complexity of an infinite word W written by

(1) $W = 10^{a_1}10^{a_2}10^{a_3}\cdots, 0 \le a_1 \le a_2 \le a_3 \cdots.$

It is clear that W is a $*$-Sturmian word. We get following Theorems on W.

Theorem 8. *Let* W *be a word given by* (1) *with* $(a_0 :=)0 \le a_1 < a_2 < a_3 \cdots$. *Then* $p(n;W) = n+1+\sharp\{(i,j) \in N^2; j \le a_{i-1}+1, a_i+j \le n-1\}$, $n \ge 0$.

Theorem 9. *Let* W *be as in Theorem* 8. *Then,* $p(n;W) \le \frac{n^2}{4} + \frac{n}{2} + \frac{17}{8} + \frac{(-1)^{n+1}}{8} - \lfloor (\frac{3}{4} + \frac{n}{4})^{-1} \rfloor$ $(n \ge 0)$. *The above estimate is best possible; the equality is attained by* $W = W_0 := 11010^2 10^3 10^4 \cdots$.

We write $f(n) \sim g(n)$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

Theorem 10. *Let* W *be a word given by* (1) *with* $0 \le a_1 < a_2 < \cdots$ *and* $a_n \sim n^\alpha$ $(\alpha \ge 1)$. *Then* $p(n;W) \sim n^{1+1/\alpha}$.

Theorem 11. *Let* $k \ge 2$ *be an integer, and* $\{b_n\}_{n=1}^{\infty}$ *a linear recurrence sequence with* $x^k - x - 1$ *as its characteristic polynomial defined by the initial condition*:

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_k \end{pmatrix} = \begin{pmatrix} 1\,0\,1\,1 & \cdots & 1 \\ 1\,1\,1\,1 & \cdots & 1 \\ 1\,1\,2\,1 & \cdots & 1 \\ \vdots\,\vdots\,\vdots\,\ddots & \ddots & 1 \\ \vdots\,\vdots\,\vdots & 2 & 1 \\ 1\,1\,2\cdots & 2 & 2 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ \vdots \\ \vdots \\ t_k \end{pmatrix},$$

$$(t_1, t_2, \ldots, t_k) \in \boldsymbol{N}^k.$$

Let W *be a word defined by*

$$W := 10^{a_1}10^{a_2}10^{a_3}\cdots, \quad a_n := b_n - 1.$$

Then $p(n;W)$ *is given by the following, so that*

$$p(n;W) = kn + c \quad for\ all\ n \ge b_k + 1,\ c \le 0,$$

where c *is a non-positive constant, and* $c = 0$ *only if* $k = 2$, $t_1 = t_2 = 1$.

$$p(n;W) = \begin{cases} n+1 & (0 \le n \le b_1) \\ n+2 & (b_1+1 \le n \le b_2) \\ 2n - b_2 + 2 & (b_2+1 \le n \le b_3) \\ 3n - b_2 - b_3 + 2 & (b_3+1 \le n \le b_4) \\ \cdots & \cdots \\ jn - b_2 - \cdots - b_j + 2 & (b_j+1 \le n \le b_{j+1}) \\ \cdots & \cdots \\ kn - b_2 - \cdots - b_k + 2 & (n \ge b_k + 1) \end{cases}.$$

If a_n is unbounded in (1), then without loss of generality, we can rewrite (1):

(2) $\qquad W = (10^{a_1})^{e_1}(10^{a_2})^{e_2}(10^{a_3})^{e_3}\cdots,$

with $(a_0 := 0) \le a_1 < a_2 < \cdots$, $e_n \ge 1$.

Theorem 12. *Let* W *be a word given by* (2). *Then* $p(n;W) = n+1+\sharp\{(i,j,k) \in \mathbf{N}^3;\ j \le a_i+1, k \le e_i - 1, k(a_i+1)+j \le n\} + \sharp\{(i,j) \in \mathbf{N}^2;\ j \le a_{i-1}+1, e_i(a_i+1)+j \le n\}$ $(n \ge 0)$.

Related to the magnitude of the usual complexity of $*$-Sturmian words, we can show the following Theorems 13, 14.

Theorem 13. *Any* $*$-*Sturmian word* $W \in L^{\boldsymbol{N}}$ *is deterministic, i.e.,*

$$\lim_{n\to\infty} \frac{\log(p(n;W))}{n} = 0.$$

Theorem 14. *For any small positive number* ϵ *there exists a* $*$-*Sturmian word* $W \in L^{\boldsymbol{N}}$ *such that* $p(W;n) > 2^{n^{1-\epsilon}}$ *holds for all sufficiently large integer* n.

References

[1] E.M. Coven and G.A. Hedlund: Sequences with minimal block growth. Math. Systems Theory, **7**, 138–153 (1973).

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

[3] S. Yasutomi: The continued fraction expansion of α with $\mu(\alpha) = 3$. Acta Arith., **84**, no.4, 337–374 (1998).