INFORMATIQUE THÉORIQUE ET APPLICATIONS

T. TAPSOBA

Minimum complexity of automatic non sturmian sequences

Informatique théorique et applications, tome 29, nº 4 (1995), p. 285-291

<http://www.numdam.org/item?id=ITA_1995__29_4_285_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 29, n° 4, 1995, p. 285 à 291)

MINIMUM COMPLEXITY OF AUTOMATIC NON STURMIAN SEQUENCES (*)

by T. TAPSOBA ([1])

Communicated by C. CHOFFRUT

Abstract. – *We determine the minimum complexity of automatic non Sturmian sequences (fixed points of injective constant-length substitutions) on sets of two elements.*

Résumé. – *Nous déterminons la complexité minimum des suites automatiques non Sturmiennes (points fixes de substitutions injectives uniformes) sur des alphabets de cardinal deux.*

1. INTRODUCTION

The study of factors of infinite sequences goes back at least to THUE ([11], [12]) and one of the questions which have been addressed is the problem of computing the complexity function P, where $P(n)$ is the number of distinct factors of length n.

If a sequence is not periodic, one can easily prove that the complexity function is strictly increasing. In this case, one has for any $n \in N$, $P(n) \geq n + 1$. So, sequences such that $P(n) = n + 1$, called Sturmian sequences [3], have minimum complexity. The topological entropy of the dynamical systems associated to those sequences is nul.

After the investigation of an example, we determine the minimum complexity of automatic non Sturmian sequences (fixed points of injective constant-length substitutions) on sets of two elements.

(*) Received October 1993, revised June 1994, accepted December 1994.

([1]) Département de Mathématiques, Faculté des Sciences et Techniques, Université de Ouagadougou, 03 BP. 7021, Ouagadougou 03, Burkina Faso.

2. PRELIMINARIES

Let A^* be the free monoid generated by a non-empty finite set A called alphabet. The elements of A are called letters and those of A^* words. For any word v in A^*, $|v|$ denotes the length of v, namely the numbers of its letters. The identity element of A^* denoted by ε is the empty word. It is a word of length 0. A word v is said to be a factor of w if $w = xvy$ for some x, y in A^*. We then write $v|w$. If $x = \varepsilon$ (resp. $y = \varepsilon$), v is called a prefix (resp. suffix) of w. A prefix or a suffix of w is said to be strict if it is different from w.

We call substitution, a morphism $f : A \to A^*$. It can be naturally extended to a morphism from A^* to A^*. A substitution is said to be a constant-length σ substitution if $\sigma = |f(i)|$ for any letter i of A. If there exists a letter $a \in A$ such that $f(a) = am$ with $|m| > 0$, then the set of the words with prefix a has a fixed point $u = amf(m) f^2(m) \ldots f^k(m) \ldots$

When an infinite word u is a fixed point of a constant-length substitution on an alphabet, it is called automatic. In fact it is well known ([2], [1]) that such a sequence is recognizable by a k-automaton. (The interested reader can find in [1], [4] or [9] the definition of recognizability).

We denote by F the set of the finite factors of u and by $F(n)$ its subset consisting of the factors of length n. It is trivial to verify that every factor of a word v of F is a word of F and that there exists a letter a such that va is in the set F. The factor v of u is said to be special if for any letter i of A, wi is a factor of u. Denote by FS the set of the special factors of u and by $FS(n)$ the set of the special factors of length n.

Let S be the shift defined by $S(a_0 a_1 a_2 \ldots) = a_1 a_2 \ldots$ and let Ω be the closure of the set $\{S^k(u);\ k \in N\}$ where the distance d is given by $d(v, w) = \exp(-\inf\{n \in N;\ v_n \neq w_n\})$. The sequence u is associated to the dynamical system (Ω, T) (where T is the restriction of S to Ω) and it is said to be minimal when the empty set and Ω are the only closed subsets of Ω invariant under T.

As noticed in [5], the behavior of $P(n)$ is directly linked to the topological entropy h of the associated dynamical system: $h = \lim \dfrac{\operatorname{Log} P(n)}{n}$.

We give here a simple criterium for minimality:

PROPOSITION 1: *Let u be a fixed point of the substitution f on the alphabet A. If a is a prefix of u with $|f(a)| \geq 2$ and if every letter of A is a factor of u, then the following are equivalent:*

(i) *u is minimal and* $\lim f^k(b) = +\infty$ *for every letter* $b \in A$.

(ii) *There exists* $L \leq \operatorname{Card}(A)$ *such that for any* $b \in A$, $a \mid f^L(b)$.

(iii) *For any* $b \in A$, *there exists* $k(b) \in N$ *such that* $a \mid f^{k(b)}(b)$.

3. FACTORS AND SPECIAL FACTORS

Let u be a minimal sequence, fixed point of an injective constant-length σ substitution f and let w be a factor of u. It can be decomposed as follows:

(1):$w = x\, f(v)\, y$. In (1) x is a strict suffix of a word $f(v_1)$, y is a strict prefix of a word $f(v_2)$ and $v_1 \vee v_2$ is a factor of u.

A factor w of u is said to be rythmical if it has a unique decomposition with condition (1).

The following results, noticed in [6], have been proved in [7] and [8].

PROPOSITION 2: *There exists* $L_0\ (= L_0(\sigma, \operatorname{Card}(A)))$ *such that every factor of* u *of length* $> L_0$ *is rythmic.*

We recall here some useful properties (*see* [9]) of factors and special factors.

PROPERTY 1: *If there exists a rythmical factor R of u with* $R \geq \sigma$, *then every factor of u which has R as a factor is rythmic.*

PROPERTY 2: *Every suffix of a special factor is special.*

PROPERTY 3: *If* $FS(p)$ *is empty, then for any* $n \geq p$, $FS(n)$ *is empty.*

Let $n > L_0$ be an integer, where L_0 is the constant of Proposition 2. For two different letters i and j, $PR_{i,j}$ (resp. *PR*) will denote the greatest common prefix of $f(i)$ and $f(j)$ (resp. the greatest common prefix of all the $f(i)$). In the same way, $SU_{i,j}$ (resp. *SU*) will denote the greatest common suffix of $f(i)$ and $f(j)$ (resp. the greatest common suffix of all the $f(i)$). Let us set $\alpha_{i,j} = |PR_{i,j}|$, $\alpha = |PR|$, $\mu_{i,j} = |SU_{i,j}|$ and $\mu = |SU|$.

The following results describe an inductive method to determine the rythmical special factors and have been proved in [9]:

PROPOSITION 3: *If there exist two different letters i and j such that* $|PR_{i,j}| \neq |PR|$, *then there is no special factor of length n.*

PROPOSITION 4: *Let k be the least integer such that* $\sigma k + \alpha \geq n$ *and let us suppose that for any two different letters i and j,* $|PR_{i,j}| = |PR|$. *Then* *the special factors of length n are suffix of length $(n - \alpha)$ of the images of the special factors of length k to which one has concatenated PR on the right hand side.*

Under the same hypothesis as in Proposition 3 and 4, one has (*see* [10]) for substitutions on sets of two elements:

PROPOSITION 5: *Two distincts special factors of length k give the same factor of length $n \in [\sigma(k-1) + \alpha + 1, \sigma k + \alpha]$ if and only if $\mu > 0$ and $n \leq \sigma(k-1) + \alpha + \mu$.*

The study of an example will be of some help to understand the general case.

4. EXAMPLE

Let us consider the sequence u, fixed point of the substitution $1 \to 112$, $2 \to 111$. A direct counting gives $P(1) = 2$, $P(2) = 3$, $P(3) = 4$, $P(4) = 6$ and $P(5) = 8$. Thus, Card $(FS(1)) = \text{Card}(FS(2)) = 1$ and Card $(FS(3)) = \text{Card}(FS(4)) = 2$. The 8 words of $F(5)$ are decomposed as follow: $\underline{111}11$; $11\underline{112}$; $1\underline{112}1$; $\underline{112}11$; $12\underline{111}$; $12\underline{112}$; $2\underline{111}1$; $2\underline{112}1$. These decompositions are unique so that every factor of length 5 is rythmic. (We can remark that among the six words of F(4) the only non rythmic factor is 1111 which can be decomposed in two ways: $\underline{111}1$ and $\underline{11}11$). Hence, every factor of u of length ≥ 5 is rythmic.

As $SU = \varepsilon$ and $PR = 11$, one has $\mu = 0$ and $\alpha = 2$. It follows from Propositions 3, 4 and 5 that for any $q \geq 5$, $\text{Card}(FS(q)) = 2$ if and only if $\text{Card}(FS(k)) = 2$ where k is the unique integer such that $\frac{q-2}{3} \leq k < \frac{q-2}{3} + 1$. Thus, the integers q such that Card $(FS(q)) = 2$ are 3, 4, 9, 10, 11, 12, 13, 14, 27, 28,... A straightforward computation gives $\text{Card}(FS(q)) = 2$ if and only if there exists an integer $r \geq 1$ such that $3^r \leq q < 3^r + 2.3^{r-1}$.

Since $P(n+1) = P(n) + \text{Card}(FS(n))$ for any integer $n \geq 1$, one has

$$\begin{aligned} P(n+1) &= P(1) + \sum_{q=1}^{n} \text{Card}(FS(q)) \\ &= 2 + n + \text{Card}\{q \in [1, n];\ \text{Card}(FS(q)) = 2\} \\ &= n + 2 + \text{Card}\{q \in [3, n];\ \text{Card}(FS(q)) = 2\} \end{aligned}$$

Let us recall that the number of integers $q \in [3^r, 3^r + 2.3^{r-1}]$ is equal to 2. 3^{r-1} and that for every integer n, there exists a unique integer k such that $3^k \leq n < 3^{k+1}$.

Thus if $3^k \leq n < 3^k + 2.3^{k-1}$, one has

$$\begin{aligned} P(n+1) &= n + 2 + \left(\sum_{r=1}^{k-1} 2.3^{r-1}\right) + n - 3^k + 1 \\ &= n + 2 + 2 \,.\, \frac{3^{k-1} - 1}{3 - 1} + n - 3^k + 1 \\ &= 2\left(n + 1 - 3^{k-1}\right) \end{aligned}$$

and if $3^k + 2.3^{k-1} \leq n < 3^{k+1}$, one has

$$\begin{aligned} P(n+1) &= n + 2 + \sum_{r=1}^{k} 2.\, 3^{r-1} \\ &= n + 2 + 2.\, \frac{3^k - 1}{3 - 1} \\ &= n + 1 + 3^k \end{aligned}$$

Let us remark that if $3^k \leq n < 3^{k+1}$, then $k \leq \log_3(n) < k + 1$ so that $k = [\log_3(n)]$ where $[x]$ denote the largest integer contained in the real x.

Moreover $3^k + 2.\, 3^{k-1} = 3^{k-1}(3 + 2) = 5.\, 3^{k-1}$;

Therefore we finally have

$$\begin{aligned} P(n+1) &= 2\left(n + 1 - 3^{[\log_3(n)]-1}\right) &\quad \text{if} \quad n < 5.\, 3^{[\log_3(n)]-1} \\ P(n+1) &= n + 1 + 3^{[\log_3(n)]} &\quad \text{if} \quad n \geq 5.\, 3^{[\log_3(n)]-1} \end{aligned}$$

Let $n \in [3^k, 3^{k+1}[$. One has $[\log_3(n)] = [\log_3(n-1)]$ if $n \neq 3^k$ and $[\log_3(3^k)] = [\log_3(3^k - 1)] + 1$. Hence, after replacing n's by $(n-1)$'s in the formula above, we obtain a new formula that allows us to compute $P(n)$ for every $n \geq 2$.

$$\begin{aligned} P(n) &= 2\left(n - 3^{[\log_3(n)]-1}\right) &\quad \text{if} \quad n \leq 5.\, 3^{[\log_3(n)]-1} \\ P(n) &= n + 3^{[\log_3(n)]} &\quad \text{if} \quad n > 5.\, 3^{[\log_3(n)]-1} \end{aligned}$$

We can now tackle the general case.

5. THE MINIMUM COMPLEXITY

Let us consider a sequence u, fixed point of an injective constant-length σ substitution on a set of two elements. If for any $k \leq \sigma$ one has Card $(FS(k)) = 1$, one can see by inspection that u is a Sturmian sequence. To have a non-Sturmian sequence with a minimum complexity we may have Card $(FS(k)) = 1$ for any $k < \sigma$, Card $(FS(\sigma)) = 2$ and clearly the number of integers q such that Card $(FS(q)) = 2$ must be lower. Moreover, one derives from the inductive method to determine the rythmical special factors that Card $(FS(q)) = 2$ if and only if Card $(FS(k)) = 2$ where k is the unique integer such that $\frac{q-\alpha}{\sigma} \leq k < \frac{q-\alpha}{\sigma} + 1$.

Hence, it is not very difficult to see that Card $(FS(q)) = 2$ if and only if there exists an integer $r \geq 1$ such that $\sigma^r \leq q < \sigma^r + 2.\sigma^{r-1}$.

Since $P(n+1) = n + 2 + \text{Card}\ \{q \in [\sigma, n];\ FS(q) = 2\}$, counting the number of integers $q \in [\sigma^r, \sigma^r + 2.\sigma^{r-1}]$ and arguing as in the preceding paragraph gives:

$$\text{If} \quad \sigma^k \leq n < \sigma^k + 2.\sigma^{k-1},$$

$$\begin{aligned} P(n+1) &= n + 2 + \left(\sum_{r=1}^{k-1} 2.\sigma^{r-1}\right) + n - \sigma^k + 1 \\ &= n + 2 + 2.\frac{\sigma^{k-1} - 1}{\sigma - 1} + n - \sigma^k + 1 \\ &= 2n + 3 + 2.\frac{\sigma^{k-1} - 1}{\sigma - 1} - \sigma^k \end{aligned}$$

and if $\sigma^k + 2.\sigma^{k-1} \leq n < \sigma^{k+1}$,

$$\begin{aligned} P(n+1) &= n + 2 + \sum_{r=1}^{k} 2.\sigma^{r-1} \\ &= n + 2 + 2.\frac{\sigma^k - 1}{\sigma - 1} \end{aligned}$$

Moreover, as $k = [\log_\sigma(n)]$, one finally has:

$$P(n+1) = 2(n+1) + 1 + 2.\frac{\sigma^{[\log_\sigma(\mathrm{n})]-1} - 1}{\sigma - 1} - \sigma^{[\log_\sigma(n)]}$$

$$\text{if} \quad n < (\sigma + 2)\,\sigma^{[\log_\sigma(n)]-1}$$

$$P(n+1) = (n+1) + 1 + 2.\frac{\sigma^{[\log_\sigma(\mathrm{n})]} - 1}{\sigma - 1} \qquad \text{if} \quad n \geq (\sigma + 2)\,\sigma^{[\log_\sigma(n)]-1}$$

As for the example, one can easily see that for $n \in [\sigma^k, \sigma^{k+1}[$ if $[\log_\sigma(n)] \neq [\log_\sigma(n-1)]$ then $n = \sigma^k$ and $[\log_\sigma(n)] = [\log_\sigma(n-1)] + 1$. Hence after replacing n's by $(n-1)$'s in the formula above, we have thus proved the following theorem on minimum complexity of automatic non-Sturmian sequences on sets of two elements:

THEOREM: *For $n \geq 2$, the complexity function P is given by*

$$P(n) = 2n + 1 + 2 \, . \frac{\sigma^{[\log_\sigma(n)]-1} - 1}{\sigma - 1} - \sigma^{[\log_\sigma(n)]}$$

$$\text{if} \quad n \leq (\sigma + 2)\, \sigma^{[\log_\sigma(n)]-1}$$

$$P(n) = n + 1 + 2 \, . \frac{\sigma^{[\log_\sigma(n)]} - 1}{\sigma - 1} \qquad \text{if} \quad n > (\sigma + 2)\, \sigma^{[\log_\sigma(n)]-1}$$

REFERENCES

1. G. Christol, T. Kamae, M. Mendès-France, G. Rauzy, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France*, 1980, *108*, pp. 401-419.
2. A. Cobham, Uniform Tag Sequences, *Math. Syst. Theo,* 1972, *6*, pp. 164-192.
3. E. M. Coven, G. A. Hedlung, Sequences with minimal block groups, *Math. Syst. Theo.,* 1973, *7*, pp. 138-153.
4. C. Mauduit, Automates finis et ensembles normaux, *Ann. Inst. Fourier, Grenoble*, 1986, *36*, 2, pp. 1-25.
5. G. Rauzy, Suites à termes dans un alphabet fini, *Sém. Théorie des nombres Bordeaux,* 1982-1983, *Exp. 35*, pp. 1-16.
6. G. Rauzy, Rotation sur les groupes, nombres algébriques et substitutions, *Sem. Théorie des nombres Bordeaux*, 1987-1988, *Exp. 21*, pp. 1-12.
7. T. Tapsoba, Complexité de suites automatiques, Thèse de troisième Cycle, Université d'Aix-Marseille II, 1987.
8. T. Tapsoba, Sur la factorisation des mots, *Annales de l'Université de Ouagadougou*, 1992, Série B, pp. 41-56.
9. T. Tapsoba, Special factors of automatic sequences, To appear in *Journal of Pure and Applied Algebra.*
10. T. Tapsoba, Automates calculant la complexité de suites automatiques, *Journal de Théorie des nombres Bordeaux*, 1994, *6*, pp. 127-134.
11. A. Thue, Uber unendliche zeichenreihen, *Norske Vid. Selsk. Skr., I. Mat. Nat. Kl, Christiana,* 1906, *7*, pp. 1-22.
12. A. Thue, Uber die gegenseitige Lage gleicher Teile genvisser zeichenreihen, *Norske Vid. Selsk. Skr., I. Math. Nat. Kl, Christiana,* 1912, *1*, pp. 1-67.