Acta Univ. Sapientiae, Mathematica, **2**, 2 (2010) 127–136

On a new notion of complexity on infinite words

Julien Cassaigne
Institut de Mathématiques de Luminy
163 avenue de Luminy, case 907
F-13288 Marseille Cedex 9
France
email: cassaigne@iml.univ-mrs.fr

Idrissa Kaboré
Institut des Sciences Exactes et Appliquées, Université polytechnique de Bobo-Dioulasso
01 BP 1091 Bobo-Dioulasso 01
Burkina Faso
email: ikaborei@yahoo.fr

Théodore Tapsoba
Ecole Supérieure d'Informatique
Université polytechnique de Bobo-Dioulasso
01 BP 1091 Bobo-Dioulasso 01
Burkina Faso
email: theo_tapsoba@univ-ouaga.bf

Abstract. The intuition according to which an infinite word is "complicated" all the more as it has many distinct factors can be translated into terms of "complexity function" of this word. In this paper, some properties of a new notion of complexity called "window complexity" are studied. A characterization of modulo-recurrent words via window complexity is given.

1 Introduction

The study of factors of infinite words goes back at least to the work of Thue [13, 14]. Among questions which have been addressed, is the problem of computing the complexity function P, where $\mathsf{P}(n)$ is the number of distinct factors of length n; it was introduced in 1975 by Ehrenfeucht et al. [6]. And since then,

2010 Mathematics Subject Classification: 68R15, 11B85
Key words and phrases: infinite word, complexity, periodicity, modulo-recurrent

it has been abundantly used to study infinite words; in particular it allowed the classification of certain families of infinite words (see for instance [1, 4, 10]).

As is shown in [12], this classical definition of complexity does not always show how complicated an infinite word is. That is why other notions of complexity were introduced by many authors, such as arithmetic complexity [3] and palindromic complexity [2].

Our aim in this paper is to give some properties of the window complexity. This new complexity was introduced by two of the authors in [8].

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 number of its letters. The identity element of A^* denoted by ε is the empty word; it is the word of length 0.

An infinite word is a sequence of letters in A indexed by $\mathbb{N}$. We denote by A^ω the set of infinite words in A and we set $A^\infty = A^* \cup A^\omega$.

An infinite word u is said τ-periodic if τ is the least positive integer such that $u_{i+\tau} = u_i$ for all $i \geq 0$.

A finite word u of length n formed by repeating a single letter x is typically denoted x^n. We define the nth power of a finite word v as being the concatenation of n copies of v; we denote it v^n. We say that an infinite word u is eventually periodic if there exist two finite words v and w such that $u = wvvv\cdots$; then u is simply denoted wv^ω.

Let $u \in A^\infty$ and $v \in A^*$. The word v is said to be a factor of u if there exist $u_1 \in A^*$ and $u_2 \in A^\infty$ such that $u = u_1 v u_2$.

For any infinite word u in A^ω, we shall write $u = u_0u_1u_2u_3\cdots$ where $u_i \in A$ for all $i \geq 0$. Let $u \in A^\omega$. The language of length n of u, denoted by $L_n(u)$, is the set of factors of u of length n:

$$L_n(u) = \{u_k u_{k+1} \cdots u_{k+n-1} : k \geq 0\}.$$

The set of all the factors of u is simply denoted by $L(u)$. A factor v of length n of a word $u = u_0u_1u_2\cdots$ appears in u at position k if $v = u_k u_{k+1} \cdots u_{k+n-1}$. A word u is said to be recurrent if every factor of u appears infinitely many times in u.

The complexity function of the infinite word u is the map from $\mathbb{N}$ to $\mathbb{N}^*$ defined by $P(u, n) = \#L_n(u)$, where $\#L_n(u)$ is the number of elements in $L_n(u)$.

A Sturmian word is an infinite word u such that $P(u, n) = n + 1$ for every integer $n \geq 0$. Sturmian words are non-eventually periodic infinite words of minimal complexity, for more details see for instance [9, 11].

Let us recall the definition of window complexity, which was introduced in [8].

Definition 1 *Let* $u = u_0u_1u_2\cdots$ *be an infinite word. The window complexity function of* u *is the map* $P_f(u, .) : \mathbb{N} \longrightarrow \mathbb{N}^*$ *defined by*[1]

$$P_f(u, n) = \#\left\{u_{kn}u_{kn+1}\cdots u_{(k+1)n-1} : k \geq 0\right\} \quad .$$

Factors of length n occurring in u at a position multiple of n, as above, will be called "window factors of length n of u". The decomposition of u into such factors will be called the "window decomposition of size n of u" or simply "n-window decomposition of u".

3 Properties of the window complexity

3.1 Comparison of $P_f(u, .)$ and $P(u, .)$

Let us first compare the window complexity function with the usual complexity function.

Proposition 1 *For any infinite word* u*, we have:*

$$\forall n \geq 0, \; P_f(u, n) \leq P(u, n) \quad .$$

Proof. For any infinite word u, we have

$$\left\{u_{kn}u_{kn+1}\cdots u_{(k+1)n-1} : k \geq 0\right\} \subseteq L_n(u) \quad .$$

Thus $P_f(u, n) \leq P(u, n)$. □

We shall see in the next section that this proposition is sharp, *i.e.*, there exist infinite words for which $P_f(u, n) = P(u, n)$ for all $n \in \mathbb{N}$.

Proposition 2 *For any infinite word* u*, we have:*

$$\forall n \geq 2, \; P(u, n) \leq (n-1)\left(P_f(u, n-1)\right)^2 \quad .$$

[1] f in P_f is from "fenêtre", window in French

Proof. For all $n \geq 2$, let v, w be two window factors of length $n-1$ in u such that vw appears in the n-decomposition of u. Then, there are at most $n-1$ factors of u of length n contained in vw, and all factors of length n are obtained this way. The result follows. □

We deduce from this proposition that if P_f is bounded, then P is at most linear. Such infinite words actually exist, as we shall see in Proposition 7.

3.2 Window complexity and modulo-recurrent words

Let us now study the window complexity of a particular class of infinite words, introduced in [7]: modulo-recurrent words.

Definition 2 *An infinite word* $u = u_0u_1u_2\cdots$ *is said to be modulo-recurrent if, for any* $k \geq 1$*, every factor* w *of* u *appears in* u *at every position modulo* k*, i.e.,*

$$\forall i \in \{0, 1, \ldots, k-1\}\,,\ \exists l_i \in \mathbb{N} : w = u_{kl_i+i}u_{kl_i+i+1}\cdots u_{kl_i+i+|w|-1}\ .$$

Note that all modulo-recurrent words are recurrent. The class of modulo-recurrent words includes words of diverse complexity, for instance Sturmian words or words with maximal complexity:

Proposition 3 [7] *Sturmian words are modulo-recurrent.*

Proposition 4 *Let* $u \in A^\omega$ *be an infinite word such that* $P(u, n) = (\#A)^n$ *for all* $n \in \mathbb{N}$*. Then* u *is modulo-recurrent.*

Proof. If $P(u, n) = (\#A)^n$ for all $n \in \mathbb{N}$, then $L(u) = A^*$. Let $w \in A^*$ and $k \geq 1$. Choose $j \in \mathbb{N}$ such that $|w|+j \equiv 1 \pmod{k}$, and $a \in A$. Then the word $(wa^j)^k$ occurs at some position n in u. It follows that w occurs at positions $n + i(|w| + j)$ in u for $i \in \{0, 1, \ldots, k-1\}$, hence at every position modulo k. □

A consequence of Proposition 4 is that almost every infinite word is modulo-recurrent, in the following sense: choose an infinite word u in A^ω at random, each letter being independently chosen in A according to a uniform law; then, with probability 1, the word u is modulo-recurrent. Indeed, it is known that $L(u) = A^*$ for almost every u.

Modulo-recurrent words can be characterized in terms of window complexity:

Theorem 1 *Let* u *be a recurrent infinite word. Then, the following assertions are equivalent:*

1. *The word* u *is modulo-recurrent.*
2. $\forall n \geq 0,\ P_f(u, n) = P(u, n)$.

Proof. Let u be a modulo-recurrent word. Since $P_f(u, n) \leq P(u, n)$ by Proposition 1, we need only to check that $P(u, n) \leq P_f(u, n)$. Let w be a factor of length n in u. Then, w appears in u at any position modulo n, in particular at a position $\equiv 0 \pmod n$. So, there exists $k \in \mathbb{N}$ such that $w = u_{kn}u_{kn+1}\cdots u_{(k+1)n-1}$. Hence, we have the inclusion

$$L_n(u) \subseteq \left\{u_{kn}u_{kn+1}\cdots u_{(k+1)n-1} : k \geq 0\right\}$$

and thus

$$P(u, n) \leq P_f(u, n).$$

Conversely, suppose that

$$\forall n \geq 0,\ P_f(u, n) = P(u, n).$$

Then, for every integer n, any factor of u of length n appears in u at least at one position $\equiv 0 \pmod n$. Let w be a factor of u of length n and k a positive integer. Let us consider an integer i such that $0 \leq i < k$. We have to show that w appears in u at a certain position $\equiv i \pmod k$. As u is a recurrent infinite word, we can find some words x and y such that xwy is a factor of u of length $|xwy| \equiv 0 \pmod k$, with $|x| = i$.

It follows that there exists an integer l such that xwy appears in u at position $l|xwy|$, *i.e.*, $xwy = u_{l|xwy|}u_{l|xwy|+1}\cdots u_{(l+1)|xwy|-1}$. Thus,

$$w = u_{l|xwy|+i}u_{l|xwy|+i+1}\cdots u_{l|xwy|+i+n-1}.$$

Therefore, w appears at a position $\equiv i \pmod k$. □

Note that Theorem 1 does not hold for non-recurrent words. Indeed, the word $u = ab^\omega$ satisfies $P_f(u, n) = P(u, n) = 2$ for all $n \geq 1$ (and of course $P_f(u, 0) = P(u, 0) = 1$), but it is not modulo-recurrent.

3.3 Window complexity and automatic words

One very interesting way to generate infinite words is to proceed by iterating a substitution on a letter.

A substitution is a map $f : A \longrightarrow A^*$. It can be naturally extended to a morphism from A^* to A^*, and to a map from A^∞ to A^∞.

If there exists a constant σ such that $|f(a)| = \sigma$ for all $a \in A$, then we say that f is σ-uniform (or just uniform, if σ is clear from the context). A 1-uniform morphism is called a coding.

Let f be a substitution on A^*. A word w on the alphabet A such that $f(w) = w$ is said to be a fixed point of f. If f is a non-erasing morphism and there exists a letter $a \in A$ such that $f(a) = am$ with $|m| > 0$, then we say that f is prolongable on a. In this case, the sequence a, $f(a)$, $f^2(a)$,... converges to the infinite word

$$u = amf(m)f^2(m)\dots f^k(m)\dots$$

which is a fixed point of f.

An infinite word is said to be σ-automatic if it is the image under a coding of a fixed point of a σ-uniform morphism, for $\sigma \geq 2$. Indeed, such a word is recognizable by a σ-automaton [5].

Proposition 5 *Let u be a σ-automatic infinite word. Then the sequence of integers $(P_f(u, n))_{n\in\mathbb{N}}$ is not strictly increasing.*

Proof. Let $u = g(v)$, where v is a fixed point of the σ-uniform morphism f and g is a coding. Then, for all $n \in \mathbb{N}$, we have $P_f(u, \sigma^n) \leq P_f(v, 1)$ since the window factors of length σ^n of u are the words $g(f^n(a))$ for $a \in L_1(v)$. Since the sequence $(P_f(u, n))_{n\in\mathbb{N}}$ contains a bounded subsequence, it is not strictly increasing. □

3.4 Bounded window complexity

We know that the complexity function of an eventually periodic word is bounded. By Proposition 1, it follows that the window complexity of an eventually periodic word is also bounded. More precisely:

Proposition 6

1. *If u is a τ-periodic word, then $P_f(u, n) \leq \dfrac{\tau}{\gcd(n, \tau)}$.*
2. *If u is eventually τ-periodic, then for n large enough,*

$$P_f(u, n) \leq 1 + \frac{\tau}{\gcd(n, \tau)}.$$

Proof.

1. Let $n \in \mathbb{N}$. A window factor of length n of u can be written as

$$u_{kn}u_{kn+1}\cdots u_{(k+1)n-1},$$

   and it is entirely determined by $kn \bmod \tau$, which takes exactly $\dfrac{\tau}{\gcd(n,\tau)}$ different values.

2. If n is large enough, then $u = wv$ where $|w| = n$ and v is τ-periodic. Then $P_f(u, n) \leq 1 + P_f(v, n)$.

$\square$

Since the window complexity of any eventually periodic word is bounded, a natural question is what happens for non-eventually periodic infinite words. Contrarily to the situation with the usual complexity function, bounded window complexity does not imply eventual periodicity. We present below a non-eventually periodic infinite word whose window complexity is bounded.

Consider the sequence $(n_i)_{i\geq 0}$ such that $n_0 = 0$ and for all $i \geq 0$, $n_{i+1} = n_i! + n_i$; and let $t = t_0t_1t_2\cdots$ be the infinite word defined by $t_n = 1$ if there exists $i \in \mathbb{N}$ such that $n = n_i$ and $t_n = 0$ otherwise.

The first few terms of (n_i) and t are:

i	0	1	2	3	4	5	$\cdots$
n_i	0	1	2	4	28	$28! + 28$	$\cdots$

$$t = 11101000000000000000000000001000\ldots.$$

Let us note that t is neither eventually periodic nor recurrent.

Proposition 7 *The window complexity of the infinite word* t *defined above satisfies* $P_f(t, 0) = 1$*,* $P_f(t, 1) = 2$*, and*

$$\forall n \geq 2,\ P_f(t, n) = 3\ .$$

Proof. Obviously, $P_f(t, 0) = 1$ and $P_f(t, 1) = \#A = 2$.

We see from the first terms of t that 11, 10, 00 all occur in the 2-window decomposition of t. Morever, since all n_i are even except $n_1 = 1$, we have $t_{2l+1} = 0$ for $l \geq 1$, therefore $t_{2l}t_{2l+1}$ cannot be equal to 01. So $P_f(t, 2) = 3$.

More generally, let $n \geq 2$, let i be the smallest integer such that $n \leq n_i$, and let $r = n_i \bmod n$. Then $n_{i-1} < n \leq n_i$ and $i \geq 2$. We first prove by induction on j that $n_j \equiv r \pmod{n}$ for all $j \geq i$. This obviously holds for $j = i$. Assume that $n_j \equiv r \pmod{n}$ for some $j \geq i$. Then $n_{j+1} - r = n_j! + (n_j - r)$, which is a multiple of n since $n_j \geq n$.

There are at least 3 window factors of length n: $t_0 t_1 \cdots t_{n-1}$, with 11 as a prefix, occurring at position 0; 0^n, occurring at position $n_{i+1} + n - r$ (since $n_{i+1} < n_{i+1} + n - r < n_{i+1} + n - r + n - 1 < n_{i+2}$); and $0^r 1 0^{n-r-1}$, occurring at position $n_i - r$.

Assume now that $w = t_{ln} t_{ln+1} \cdots t_{(l+1)n-1}$ is a window factor of length n of t. If it starts with 11, then it must be the prefix of length n, since 11 does not occur in t after position 1. Otherwise, $l \geq 1$. For $0 \leq k \leq n-1$, $ln + k = n_j$ is only possible if $k = r$, since $n_j \equiv r \pmod{n}$ if $j \geq i$, and $n_j < n$ if $j < i$. Hence w is either 0^n or $0^r 1 0^{n-r-1}$. We have shown that there is not other window factor of length n, *i.e.*, $P_f(t, n) = 3$. □

By Proposition 2, and since the word t is non-eventually periodic, we have $n + 1 \leq P(t, n) \leq 9(n-1)$ for $n \geq 2$. Actually, one can prove that $P(t, n) = n + o(\log n)$.

4 Some questions

We conclude with a few open questions.

- By Proposition 6, we know that if u is an eventually periodic infinite word then its window complexity function $P_f(u, \cdot)$ is bounded. Also, we have presented (Proposition 7) an infinite word, non-eventually periodic and non-recurrent, such that its window complexity function is bounded. Does there exist some infinite recurrent and non-eventually periodic word for which the window complexity function is bounded?

- Among infinite words with bounded window complexity, a subclass of particular interest is that of words with eventually constant window complexity, *i.e.*, verifying the following property:

$$\exists n_0,\ c \in \mathbb{N} : \forall n \geq n_0,\ \ P_f(u, n) = c. \tag{1}$$

  Eventually constant words have eventually constant window complexity. The example constructed in Proposition 7 shows that even non-eventually periodic words may have eventually constant window complexity.

  It would be interesting to see if there exist recurrent words, or better automatic words, that possess Property (1).

- We know by Proposition 5 that the window complexity function of an automatic word is not strictly increasing, and even contains a bounded subsequence. On the other hand, by Theorem 1, modulo-recurrent words have strictly increasing window complexity. Do there exist non-modulo-recurrent words with strictly increasing window complexity?

References

[1] J.-P. Allouche, Sur la complexité des suites infinies, *Bull. Belg. Math. Soc.* **1** (1994), 133–143.

[2] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, *Theoret. Comput. Sci.*, **292** (2003), 9–31.

[3] S. V. Avgustinovich, D. G. Fon-Der-Flaass, A. E. Frid, Arithmetical complexity of infinite words, in: *Words, Languages and Combinatorics III*, pp. 51–62, World Scientific, Singapore, 2001.

[4] J. Cassaigne, Complexité et facteurs spéciaux, *Bull. Belg. Math. Soc.*, **4** (1997), 67–88.

[5] A. Cobham, Uniform tag sequences, *Math. Systems Theory*, **6** (1972), 164–192.

[6] A. Ehrenfeucht, K. P. Lee, G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interaction, *Theoret. Comput. Sci.*, **1** (1975), 59–75.

[7] I. Kaboré, T. Tapsoba, Combinatoire de mots récurrents de complexité $n+2$, *RAIRO-Theoret. Inform. Appl.*, **41** (2007), 425–446.

[8] I. Kaboré, T. Tapsoba, Complexité par fenêtre, in: *Proceedings of the 6^{th} International Conference on Words (WORDS 2007)*, pp. 184–188, Institut de Mathématiques de Luminy, Marseille (France), 2007.

[9] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, 2002.

[10] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in: *ICALP'84*, Lect. Notes Comput. Sci. 172, Springer-Verlag, 1984 pp. 380–389.

[11] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lect. Notes Math. 1794, Springer-Verlag, 2002.

[12] T. Tapsoba, Suites infinies, complexité et géométrie, *Rev. CAMES, Série A*, **6** (2008), 94–96.

[13] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Skr. I. Kl. Christiana*, **7** (1906), 1–22.

[14] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I. Math. Nat. Kl. Christiana*, **1** (1912), 1–67.

Received: October 4, 2010