Journal of Automata, Languages and Combinatorics **u** (v) w, x–y

Ultimately Constant Abelian Complexity of Infinite Words

Aleksi Saarela [1]

Department of Mathematics and Turku Centre for Computer Science TUCS,
University of Turku, 20014 Turku, Finland
e-mail: `amsaar@utu.fi`

ABSTRACT

It is known that there are recurrent words with constant abelian complexity three, but not with constant complexity four. We prove that there are recurrent words with ultimately constant complexity c for every c.

Keywords: combinatorics on words, abelian complexity, Sturmian words

1. Introduction

The factor complexity (or subword complexity) function ρ_w of a word w maps a positive integer n to the number of different factors of w of length n. There have been a lot of research on the factor complexity of infinite words, see e.g. [1].

The concept of factor complexity can be modified by counting only the number of different commutative images of factors of certain length. This gives the abelian factor complexity function ρ_w^{ab}. It was defined an studied in [7], although it has appeared implicitly before. For example, in [4] it was proved that an aperiodic word w is Sturmian if and only if $\rho_w^{\mathrm{ab}}(n) = 2$ for every n.

Does there exist a word w such that $\rho_w^{\mathrm{ab}}(n) = 3$ for all n? This question was raised by G. Rauzy. In questions like this it is natural to consider only recurrent words, i.e. words where every factor appears infinitely often, because there are trivial nonrecurrent examples. In [7] a recurrent example was given, and it was conjectured that there is no recurrent word w such that $\rho_w^{\mathrm{ab}}(n) = 4$ for all n. This conjecture was proved in [5].

We prove that the situation changes, if the complexity is required to be only ultimately constant. For every $c \geq 2$, we give an example of a recurrent binary word w such that $\rho_w^{\mathrm{ab}}(n) = c$ for all $n \geq c - 1$. These examples are as close to a constant complexity as possible for binary words.

[1] Supported by the Academy of Finland under grant 121419 and by the Turku University Foundation

2. Ultimately Constant Complexity

For basics on combinatorics on words we refer to [3] and [6]. In particular, see [2] for information on Sturmian words.

Let Σ be a finite alphabet and Σ^ω be the set of all right infinite words over Σ. Let ε be the empty word. If $w \in \Sigma^\omega$, let $F_n(w)$ be the set of all factors of w of length n, and let $\#F_n(w)$ be the size of this set. For a finite word u and a letter a, let $|u|$ be the length of u, and let $|u|_a$ be the number of occurrences of a in u.

Words u and v are *abelian equivalent*, if $|u|_a = |v|_a$ for every $a \in \Sigma$. This is denoted by $u \sim_{\mathrm{ab}} v$

The *factor complexity* of a word w is the function $\rho_w : \mathbb{N}_1 \to \mathbb{N}_1, \rho_w(n) = \#F_n(w)$, where $\mathbb{N}_1 = \{1, 2, 3, \dots\}$.

If R is an equivalence relation, we can define a modified factor complexity function by letting $\rho_w^R(n) = \#F_n(w)/R$ be the number of equivalence classes intersecting $F_n(w)$. For $R = \sim_{\mathrm{ab}}$ we get the *abelian factor complexity*, denoted by ρ_w^{ab}.

Example 1 Let $w = 0120^\omega \in \{0,1,2\}^\omega$. Now $F_1(w) = \{0,1,2\}$, $F_2(w) = \{01, 12, 20, 00\}$ and $F_n(w) = \{0120^{n-3}, 120^{n-2}, 20^{n-1}, 0^n\}$ for $n \geq 3$, so $\rho_w(1) = 3$ and $\rho_w(n) = 4$ for $n \geq 2$. Also $\rho_w^{\mathrm{ab}}(1) = 3$ and $\rho_w^{\mathrm{ab}}(2) = 4$, but $\rho_w^{\mathrm{ab}}(n) = 3$ for $n \geq 3$ because $0120^{n-3} \sim_{\mathrm{ab}} 120^{n-2}$.

An infinite word w is *recurrent*, if each of its factors appears infinitely often in it. It is *uniformly recurrent*, if every factor appears infinitely often with bounded gaps, that is for every factor u there exists an integer n such that u is a factor of every word in $F_n(w)$.

It is natural to ask whether for a given k there exists a word $w \in \{0, 1, \dots, k-1\}^\omega$ such that $\rho_w^{\mathrm{ab}}(n) = k$ for all n. There is a trivial nonrecurrent example of such a word:

$$w = 01 \dots (k-2)(k-1)^\omega.$$

If w is required to be recurrent (or uniformly recurrent), the question becomes more interesting. It is well known that for $k = 2$ any Sturmian word will do. For $k = 3$ we can take the image of any aperiodic word in $\{0,1\}^\omega$ under the morphism $0 \mapsto 012, 1 \mapsto 021$ (see [7]). For $k \geq 4$ there are no such words (see [5]).

If we slightly relax the condition by requiring $\rho_w^{\mathrm{ab}}(n) = k$ only for all sufficiently large n, then we will show that there are suitable words for every k. The words we give are morphic images of Sturmian words and thus uniformly recurrent.

We will consider the binary case $\Sigma = \{0, 1\}$. Then u and v are abelian equivalent if and only if $|u| = |v|$ and $|u|_1 = |v|_1$. Now the abelian complexity of a word $w \in \Sigma^\omega$ can be given as follows:

$$\rho_w^{\mathrm{ab}}(n) = \max\{|u|_1 : u \in F_n(w)\} - \min\{|u|_1 : u \in F_n(w)\} + 1.$$

For binary words, $\rho_w^{\mathrm{ab}}(n) \leq n+1$. If we want $\rho_w^{\mathrm{ab}}(n) = k+1$ to hold for as many n as possible, the best we can hope for is that it holds for $n \geq k$. This is indeed possible.

Table 1: Here is a diagram related to the proof of Theorem 1 in the case where $D = h(d)$ and a prefix of $h(e)$ is a suffix of V. It is also possible that a prefix of $h(d)$ is a suffix of V and $E = \varepsilon$.

	U			
	A		B	
$h(a)$		$h(u)$	$h(b)$	

	V				
	C		D	E	
$h(c)$		$h(v)$	$h(d)$	$h(e)$	

Theorem 1 *Let $k \geq 1$. Let $W = h(w)$, where $w \in \{0,1\}^\omega$ is a Sturmian word and h is the morphism $0 \mapsto 0^k, 1 \mapsto 1^k$. Now $\rho_W^{\mathrm{ab}}(n) = k+1$ for all $n \geq k$.*

Proof. First we prove that for every $n \geq k$ there are equally long words U and V such that they are factors of W and $||U|_1 - |V|_1| = k$. This guarantees that $\rho_W^{\mathrm{ab}}(n) \geq k+1$ for all $n \geq k$. Let $n = qk + r$, where $0 \leq r < k$. Because w is Sturmian, it has a right special factor u of length q, and another equally long factor v such that $||u|_1 - |v|_1| = 1$. Let $a \in \{0,1\}$ be such that va is a factor of w. Now ua is also a factor of w, and we can take U and V to be the prefixes of $h(ua)$ and $h(va)$ of length n.

Next we prove that if U and V are equally long factors of W, then $||U|_1 - |V|_1| \leq k$. This guarantees that $\rho_W^{\mathrm{ab}}(n) \leq k+1$ for all n. Let $U = Ah(u)B$ and $V = Ch(v)D'$, where $|A|, |C| < k$, $|u| = |v|$, and $|B| < k$ or $|D'| < k$. Because of the symmetry between U and V, we can assume that $|A| \geq |C|$ and $|B| < k$. There are two possibilities for D': if $|D'| \geq k$, then we let $D' = DE$, where $|D| = k$, and if $|D'| < k$, then we let $D' = DE$, where $D = D'$ and $E = \varepsilon$. Now there are five letters $a, b, c, d, e \in \{0,1\}$ such that aub and $cvde$ are factors of w, A and C are proper suffixes of $h(a)$ and $h(c)$, and B, D and E are prefixes of $h(b)$, $h(d)$ and $h(e)$ (see Table 1). We assume that

$$|V|_1 - |U|_1 > k \tag{1}$$

and derive a contradiction (the case $|U|_1 - |V|_1 > k$ is symmetric because of the symmetry between 0's and 1's). There are three cases.

First, let $|u|_1 - |v|_1 = 1$. Now (1) is the same as $|CDE|_1 - |AB|_1 > 2k$, which is impossible, because $|CDE| = |AB| < 2k$.

Second, let $|u|_1 - |v|_1 = 0$. Now (1) is the same as $|CDE|_1 - |AB|_1 > k$, which can also be written as $|AB|_0 - |CDE|_0 > k$. Thus $a = b = 0$. If $E = \varepsilon$, then $|CE| < k$, and otherwise $|D| = k$ and $|CE| = |AB| - |D| < k$. Because $|CE|, |D| \leq k$, it must be $d = 1$, and $c = 1$ or $e = 1$. But then cvd or vde has two more 1's than aub, which is not possible, because they are factors of a Sturmian word w.

Third, let $|u|_1 - |v|_1 = -1$. Now (1) is the same as $|CDE|_1 - |AB|_1 > 0$, which can also be written as $|AB|_0 - |CDE|_0 > 0$. This means that $a = 0$ or $b = 0$. But then $c = d = 0$, because otherwise cv or vd would have two more 1's than au or ub. So it must be $e = 1$ and $E \neq \varepsilon$. Now $D = 0^k$, so $|AB|_0 > |CDE|_0 \geq |D|_0$ only if $a = b = 0$. But then vde has two more 1's than aub. This completes the proof. □

References

[1] J.-P. Allouche, Sur la complexité des suites infinies. *Bull. Belg. Math. Soc.* **1** (1994) 2, 133–143.

[2] J. Berstel, P. Seebold, Sturmian Words. In: M. Lothaire (ed.), *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[3] C. Choffrut, J. Karhumäki, Combinatorics of Words. In: G. Rozenberg, A. Salomaa (eds.), *Handbook of Formal Languages*. 1, Springer-Verlag, 1997, 329–438.

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

[5] J. Currie, N. Rampersad, Recurrent words with constant Abelian complexity. 2009. Available at http://arxiv.org/abs/0911.5151.

[6] M. Lothaire, *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[7] G. Richomme, K. Saari, L. Q. Zamboni, Abelian complexity in minimal subshifts. 2009. Submitted, available at http://arxiv.org/abs/0911.2914.