RAIRO-Theor. Inf. Appl. 47 (2013) 195–199
DOI: 10.1051/ita/2013033

Available online at:
www.rairo-ita.org

A NOTE ON CONSTRUCTING INFINITE BINARY WORDS WITH POLYNOMIAL SUBWORD COMPLEXITY*

Francine Blanchet-Sadri[1], Bob Chen[2] and Sinziana Munteanu[3]

Abstract. Most of the constructions of infinite words having polynomial subword complexity are quite complicated, *e.g.*, sequences of Toeplitz, sequences defined by billiards in the cube, etc. In this paper, we describe a simple method for constructing infinite words w over a binary alphabet $\{a, b\}$ with polynomial subword complexity p_w. Assuming w contains an infinite number of a's, our method is based on the gap function which gives the distances between consecutive b's. It is known that if the gap function is injective, we can obtain at most quadratic subword complexity, and if the gap function is blockwise injective, we can obtain at most cubic subword complexity. Here, we construct infinite binary words w such that $p_w(n) = \Theta(n^\beta)$ for any real number $\beta > 1$.

Mathematics Subject Classification. 68R15.

Keywords and phrases. Binary words, subword complexity, gap function.

* *This material is based upon work supported by the National Science Foundation under Grant No. DMS–1060775. We thank the referees of a preliminary version of this paper as well as Aleksandar Chakarov, Lucas Manuelli, Jarett Schwartz, and Slater Stich for their very valuable comments and suggestions.*

[1] Department of Computer Science, University of North Carolina, PO Box 26170, Greensboro, NC 27402–6170, USA. blanchet@uncg.edu

[2] Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, Dept 0112, LaJolla, CA 92093–0112, USA

[3] School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213–3891, USA

Article published by EDP Sciences

© EDP Sciences 2013

https://doi.org/10.1051/ita/2013033 Published online by Cambridge University Press

1. Introduction

Let w be an infinite word. A subword of w is a block of consecutive letters of w. The subword complexity function p_w of w gives for each positive integer n, the number $p_w(n)$ of distinct subwords of length n of w. The subword complexity of infinite words has become an important research topic that finds applications in dynamical systems, ergodic theory, number theory, and theoretical computer science [12]. We refer the reader to [1,2,12] for some surveys on subword complexity.

Specific methods have been given for constructing infinite words with subword complexity functions that are linear [5, 9, 13, 15–18], polynomial [3, 4, 6, 7, 11–14], intermediate [10], and exponential [2]. In the case of polynomial complexity, most of the constructions are quite complicated, *e.g.*, sequences of Toeplitz, sequences defined by billiards in the cube, etc.

Gheorghiciuc [13] studied the subword complexity of a class of infinite words over the binary alphabet $\{a, b\}$. Assuming that such words contain an infinite number of a's, her approach is based on the concept of *gap function* that gives the distances between consecutive b's. She described a method for computing the subword complexity of a word w whose gap function h is injective or blockwise injective (the latter is defined by "$h(i) = h(j)$ for $i < j$ implies $h(i) = h(i+1) = \cdots = h(j)$"). In the case of an injective gap function, $p_w(n) = O(n^2)$, while in the case of a blockwise injective gap function, $p_w(n) = O(n^3)$. She also obtained a necessary and sufficient condition for a function to be the subword complexity of an infinite binary word with an increasing gap function.

In this paper, we provide a simple way of constructing infinite binary words with polynomial subword complexity. Reference [13], Theorem 1.3 shows that if an infinite word w over $\{a, b\}$ has a gap function that is increasing, then $p_w(n) = O(n^2)$. Here for each real number $\beta > 1$, we construct an infinite word w over $\{a, b\}$ such that $p_w(n) = \Theta(n^\beta)$, whose gap function is not strictly increasing. In order to do this, we require that the b's be not spread "too far" apart. Similar constructions were used recently by Blanchet–Sadri *et al.* [8] who showed that there are infinite *partial* words that have subword complexities not achievable by infinite *full* words.

2. A Construction of infinite binary words with polynomial subword complexity

Let w be an infinite word over $\{a, b\}$, *i.e.*, w is a function from $\mathbb{N} = \{0, 1, 2, \ldots\}$ to $\{a, b\}$. Assuming that w contains an infinite number of a's, let $H : \mathbb{N}^+ \to \mathbb{N}^+$ be a strictly increasing function such that for each $i \in \mathbb{N}$, $w(i) = b$ if and only if $i = H(m) - 1$ for some $m \in \mathbb{N}^+$. In other words, the mth b occurs at position $H(m) - 1$.

The *gap function* $h : \mathbb{N}^+ \to \mathbb{N}^+$ of w gives the distances between consecutive b's in w, *i.e.*, for each $m \in \mathbb{N}^+$, $h(m) = H(m+1) - H(m)$.

https://doi.org/10.1051/ita/2013033 Published online by Cambridge University Press

If w has only a finite number N of b's, then we take $H(m) = \infty$ for $m > N$ and $h(m) = \infty$ for $m \geq N$.

We obtain quadratic growth if the gap function is injective.

Theorem 2.1. *If an infinite word w over $\{a, b\}$ has a gap function that is injective, then $p_w(n) = O(n^2)$.*

Proof. There are only $n + 1$ subwords of length n containing fewer than two b's. The function, on the set of subwords of length n containing at least two b's, that maps a subword u to the pair

$$(\text{position in } u \text{ of first occurrence of } b, \text{distance between 1st and 2nd } b \text{ in } u)$$

is injective, and each coordinate in the codomain is bounded by n. So there are at most $O(n^2)$ subwords of length n. □

However, we can obtain arbitrary polynomial growth if the gap function is not strictly increasing, as in the following theorem.

Theorem 2.2. *Let $\beta > 1$ be a real number, and let $\alpha = \frac{\beta}{\beta-1}$. Let $H(m) = \lceil m^\alpha \rceil$, and let w be the infinite word whose letter at position i is defined by*

$$w(i) = \begin{cases} b, & \text{if} \quad i = H(m) - 1 \quad \text{for some } m; \\ a, & \text{otherwise.} \end{cases}$$

Then $p_w(n) = \Theta(n^\beta)$.

Proof. Note that $\alpha > 1$ and that there are exactly $n+1$ subwords of length n that contain at most one b.

Now consider the set of subwords of length n with at least two b's. Pick M large enough so that h is strictly increasing beyond M, *i.e.*, $h(M+1) < h(M+2) < \cdots$. Then for $n > H(M+2)$, the pattern of b's form a unique fingerprint for a subword u of length n (provided u has at least two b's). Now, if m is the smallest integer so that $h(m) \geq n$, then we will get a unique subword starting at every position until $H(m-1)$.

Let $G(x) = x^\alpha$. Then

$$G'(x) = \frac{\mathrm{d}}{\mathrm{d}x} x^\alpha = \alpha x^{\alpha-1}, \quad G'^{-1}(x) = \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} x^{\frac{1}{\alpha-1}},$$

and set $m^* = G'^{-1}(n) = \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} n^{\frac{1}{\alpha-1}}$.

https://doi.org/10.1051/ita/2013033 Published online by Cambridge University Press

If we let $g(x) = G(x+1) - G(x) = (x+1)^\alpha - x^\alpha$, then for all x there is a real number r such that

$$h(x) = \lceil (x+1)^\alpha \rceil - \lceil x^\alpha \rceil = (x+1)^\alpha - x^\alpha + r = g(x) + r \text{ and } |r| < 1.$$

Based on the definition of m and increasingness of h, $h(m-2) < n-1$ and $h(m+1) \geq n+1$, and then $g(m-2) < n$ and $g(m+1) \geq n$.

Hence $g(m-2) < n \leq g(m+1)$, and it follows that $m-2 < m^* < m+2$. Now,

$$H(m^*) = \left\lceil \left(\frac{1}{\alpha}\right)^\beta n^\beta \right\rceil = \Theta(n^\beta).$$

But H increases, so

$$H(m-2) \leq H(m^*) \leq H(m+2),$$

and it follows that $H = \Theta(n^\beta)$. Because $\beta > 1$, we have

$$p_w(n) = H(m-1) + n + 1 = \Theta(n^\beta),$$

as desired. □

References

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

[2] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press (2003).

[3] P. Arnoux, C. Mauduit, I. Shiokawa and J. Tamura, Complexity of sequences defined by billiards in the cube. *Bull. Soc. Math. France* **122** (1994) 1–12.

[4] P. Arnoux, C. Mauduit, I. Shiokawa and J. Tamura, Rauzy's conjecture on billiards in the cube. *Tokyo J. Math.* **17** (1994) 211–218.

[5] P. Arnoux and G. Rauzy, Représentation géométrique de suites de complexité $2n+1$. *Bull. Soc. Math. France* **119** (1991) 199–215.

[6] Y. Baryshnikov, Complexity of trajectories in rectangular billiards. *Communicat. Math. Phys.* **174** (1995) 43–56.

[7] N. Bedaride, Directional billiard complexity in rational polyhedra. *Regular Chaotic Dyn.* **3** (2003) 97–106.

[8] F. Blanchet-Sadri, A. Chakarov, L. Manuelli, J. Schwartz and S. Stich, Constructing partial words with subword complexities not achievable by full words. *Theoret. Comput. Sci.* **432** (2012) 21–27.

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

[10] J. Cassaigne, Constructing infinite words of intermediate complexity. In *DLT 2002, 6th International Conference on Developments in Language Theory*, Kyoto, Japan. Springer-Verlag, Berlin. *Lect. Notes Comput. Sci.* **2450** (2003) 173–184.

[11] J. Cassaigne and J. Karhumäki, Toeplitz words, generalized periodicity and periodically iterated morphisms. *European J. Combin.* **18** (1997) 497–510.

[12] S. Ferenczi, Complexity of sequences and dynamical systems. *Discrete Math.* **206** (1999) 145–154.

https://doi.org/10.1051/ita/2013033 Published online by Cambridge University Press

[13] I. Gheorghiciuc, The subword complexity of a class of infinite binary words. *Adv. Appl. Math.* **39** (2007) 237–259.
[14] M. Koskas, Complexité des suites de Toeplitz. *Discrete Math.* **183** (1998) 161–183.
[15] M. Lothaire, *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge (2002).
[16] M. Morse and G.A. Hedlund, Symbolic dynamics. *Amer. J. Math.* **60** (1938) 815–866.
[17] M. Morse and G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.* **62** (1940) 1–42.
[18] G. Rote, Sequences with subword complexity $2n$. *J. Number Theory* **46** (1993) 196–213.

Communicated by J. Kari.
Received May 18, 2012. Accepted January 8, 2013.

https://doi.org/10.1051/ita/2013033 Published online by Cambridge University Press