# On closed and open factors of Arnoux-Rauzy words *


Olga Parshina[1,2] and Luca Zamboni[1]

[1]Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 boulevard du 11 novembre 1918, F69622 Villeurbanne Cedex, France

[2]Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Science, 4 Acad. Koptyug avenue, 630090 Novosibirsk, Russia, {parshina,zamboni}@math.univ-lyon1.fr


Given a finite non-empty set $\mathbb{A}$, let $\mathbb{A}^{\mathbb{N}}$ denote the set of (right) infinite words $x = x_1x_2x_3\cdots$ with $x_i \in \mathbb{A}$. For each infinite word $x = x_1x_2x_3\cdots \in \mathbb{A}^{\mathbb{N}}$, the factor complexity $p_x(n)$ counts the number of distinct blocks (or factors) $x_ix_{i+1}\cdots x_{i+n-1}$ of length $n$ occurring in $x$. First introduced by Hedlund and Morse in their seminal 1938 paper [13] under the name of *block growth*, the factor complexity provides a useful measure of the extent of randomness of $x$. Periodic words have bounded factor complexity while digit expansions of normal numbers have maximal complexity. A celebrated theorem of Morse and Hedlund in [13] states that every aperiodic (meaning not ultimately periodic) word contains at least $n+1$ distinct factors of each length $n$. Sturmian words are those aperiodic words of minimal factor complexity: $p_x(n) = n+1$ for each $n \geq 1$.

Other notions of complexity have been successfully used in the study of infinite words and their combinatorial properties [1, 5, 6, 7, 15, 16]. In this note, we introduce and study two new complexity functions based on the notions of open and closed words [8]. We recall that a word $u \in \mathbb{A}^+$ is said to be *closed* if either $u \in \mathbb{A}$ or if $u$ is a complete first return to some proper factor $v \in \mathbb{A}^+$, meaning $u$ has precisely two occurrences of $v$, one as a prefix and one as a suffix. Otherwise, if $u$ is not closed then $u$ is *open*. For example, *abbbab* and *aabaaabaa* are both closed words while *ab* and *abaabbababbaaba* are both open. It is easily seen that all privileged words [15] are closed and hence so are all palindromic factors of rich words [9]. The terminology open and closed was first introduced by the authors in [3] although the notion of a closed word had already been introduced earlier by A. Carpi and A. de Luca in [4]. For a nice overview of open and closed words we refer the reader to the recent survey article by G. Fici [8].

To each infinite word $x \in \mathbb{A}^{\mathbb{N}}$ we consider the functions $f_x^c, f_x^o : \mathbb{N} \to \mathbb{N}$ which count the number of closed and open factors of $x$ of each length $n \in \mathbb{N}$. We study the behaviour of these complexity functions for Arnoux-Rauzy words [2]. Recall an infinite word $x \in \mathbb{A}^{\mathbb{N}}$ is called an *Arnoux-Rauzy word* if it is recurrent and if $x$ contains, for each $n \geq 0$, precisely one right special factor of length $n$ which is a prefix of $|\mathbb{A}|$-many factors of $x$ of length $n+1$ and precisely one left special factor of length $n$ which is a suffix of $|\mathbb{A}|$-many factors of $x$


*This work was performed within the framework of the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR), and has been supported by RFBS grant 18-31-00009

of length $n+1$. In particular one has $p_x(n) = (|\mathbb{A}|-1)n+1$ and each factor $u$ of $x$ has precisely $|\mathbb{A}|$ distinct complete first returns. Arnoux-Rauzy words were first introduced in [2] in the special case of a 3-letter alphabet. Let us note that in case $|\mathbb{A}| = 2$, then $x$ is Sturmian. Since for any word $x \in \mathbb{A}^{\mathbb{N}}$ we have that $f_x^c(n) + f_x^o(n) = p_x(n)$, it suffices to understand the behaviour of $f_x^c(n)$.

Our main result in Theorem 1 below provides an explicit formula for the closed complexity function $f_x^c(n)$ for an Arnoux-Rauzy word $x$ on a $t$-letter alphabet $\mathbb{A}$. The formula is expressed in terms of two related sequences associated to $x$. The first is the sequence $(b_k)_{k\geq 0}$ of the lengths of the bispecial factors $\varepsilon = B_0, B_1, B_2, \ldots$ of $x$, ordered in increasing length. The second is the sequence $(p_a^{(k)})_{a\in\mathbb{A}}^{k\geq 0}$ where for each $k \geq 0$, the $t$ coordinates of $(p_a^{(k)})_{a\in\mathbb{A}}$ are the lengths of the $t$ first returns in $x$ to $B_k$. More precisely, $p_a^{(k)} = |R_a^{(k)}| - b_k$ where $R_a^{(k)}$ is the complete first return to $B_k$ in $x$ beginning in $B_k a$. Both sequences have already been extensively studied in the literature. In particular, following [11] one has that

$$b_k = \frac{\sum_{a\in\mathbb{A}} p_a^{(k)} - t}{t-1}.$$

Furthermore, for each $k \in \mathbb{N}$, the coordinates of $(p_a^{(k)})_{a\in\mathbb{A}}$ are coprime and each is a period of the word $B_k$. Moreover, $B_k$ is an extremal Fine and Wilf word i.e., any word $u$ having periods $(p_a^{(k)})_{a\in\mathbb{A}}$ and of length greater than $b_k$ is a constant word, i.e., $u = a^n$ for some $n$ (see [17]). The sequence $(p_a^{(k)})_{a\in\mathbb{A}}^{k\geq 0}$ is computed recursively as follows : $p_a^{(0)} = 1$ for each $a \in \mathbb{A}$. For $k \geq 1$, let $a \in \mathbb{A}$ be the unique letter such that $aB_{k-1}$ is a right special factor of $x$. Then $p_a^{(k)} = p_a^{(k-1)}$, and $p_b^{(k)} = p_b^{(k-1)} + p_a^{(k-1)}$ for $b \in \mathbb{A} \setminus \{a\}$.

**Theorem 1.** *Let $x \in \mathbb{A}^{\mathbb{N}}$ be an Arnoux-Rauzy word. For each $k \in \mathbb{N}$ and $a \in \mathbb{A}$ set $I_{k,a} = [b_{k-1} - p_k + p_a^{(k)} + 2, b_k + p_a^{(k)}]$ where $p_k = \min_{b\in\mathbb{A}}\{p_b^{(k)}\}$. Let*

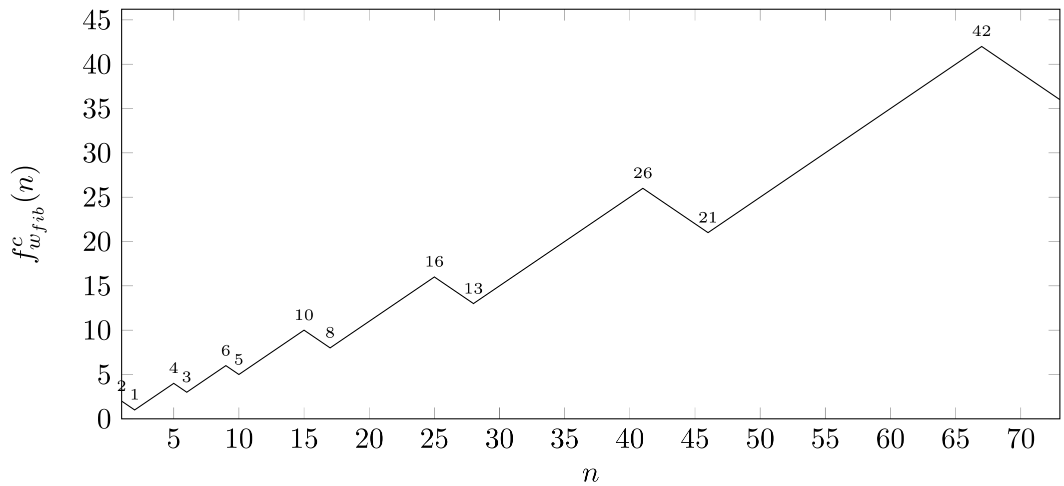
$$F(a,n) = \sum_{\substack{k\in\mathbb{N} \\ n\in I_{k,a}}} (d(n, I_{k,a}) + 1) \tag{1}$$

*where for $n \in I_{k,a}$, the quantity $d(n, I_{k,a})$ denotes the minimal distance from $n$ to the endpoints of the interval $I_{k,a}$. Then the number of closed factors of $x$ for each length $n$ is $f_x^c(n) = \sum_{a\in\mathbb{A}} F(a,n)$.*

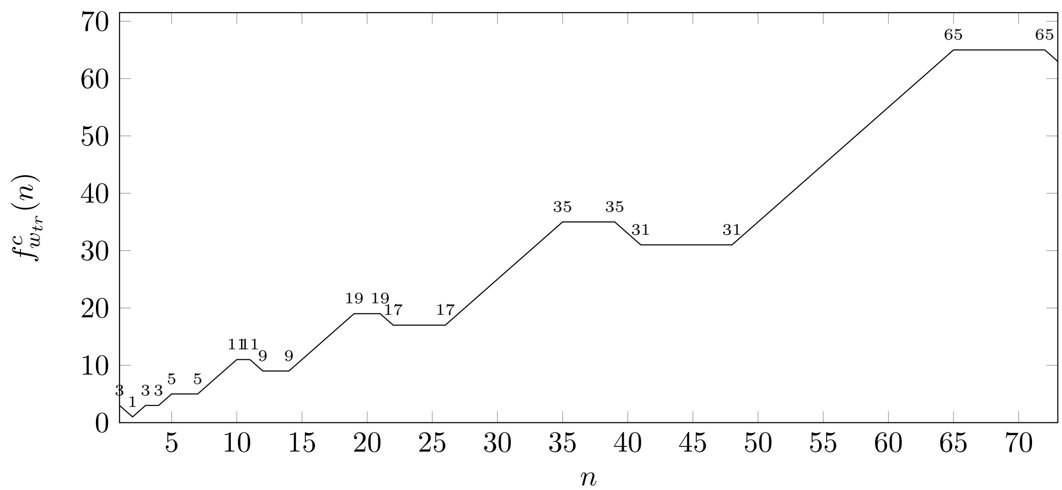
It is easily checked that the length of each interval $I_{k,a}$ is $2p_k - 2$ and that for each fixed $n$ the sum in (1) is actually a finite sum.

The following figures illustrate the behaviour of the closed complexity function $f_x^c$ in the case of the Fibonacci word and the Tribonacci word.

The number of closed factors in Fibonacci word




The number of closed factors in Tribonacci word




It is evident that in general $f_x^c$ is not monotone. However as a consequence to Theorem 1 we are able to show :

**Corollary 2.** *Let $x$ be an Arnoux-Rauzy word. Then* $\liminf f_x^c(n) = +\infty$.

In contrast, it is shown in [14] that for any paperfolding word $x$, $\liminf f_x^c(n) = 0$, in other words, for infinitely many $n$, $x$ has no closed factors of length $n$.

## References

[1] J.-P. Allouche, M. Baake, J. Cassaigne and D. Damanik, Palindrome complexity, Selected papers in honor of Jean Berstel, Theoret. Comput. Sci., 292 (2003), pp. 9–31.

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

[3] M. Bucci, A. De Luca, G. Fici, Enumeration and Structure of Trapezoidal Words, Theoret. Comput. Sci. 468 (2013) pp. 12–22.

[4] A. Carpi, A. de Luca, Periodic-like words, periodicity and boxes, Acta. Inform. 37 (2001), pp. 597–618.

[5] J. Cassaigne, G. Fici, M. Sciortino and L.Q. Zamboni, Cyclic complexity of words, J. Combin. Theory Ser. A, 145 (2017), pp. 36–56.

[6] E. Charlier, S. Puzynina, L.Q. Zamboni, On a group theoretic generalization of the Morse-Hedlund theorem, Proceedings of the AMS, 145 (2017), pp. 3381–3394.

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

[8] G. Fici, Open and closed words, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS No. 123 (2017), p. 138–147.

[9] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, European J. Combin. 30 (2009), no. 2, pp. 510–531.

[10] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoret. Comput. Sci., 136 (1994), pp. 361–385

[11] J. Justin, On a paper by Castelli, Mignosi, Restivo, RAIRO: Theoret.Informatics Appl. 34 (2000),pp. 373–377.

[12] T. Kamae and L.Q. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words, Ergodic Theory Dynam. Systems 22 (2002), pp. 1191–1199.

[13] M. Morse, G. Hedlund, Symbolic dynamics, Amer. J. Math. 60 (1938), pp. 815–866.

[14] O. Parshina, On the number of closed factors in the paperfolding word, preprint in preparation.

[15] J. Peltomäki, Introducing privileged words: privileged complexity of Sturmian words, Theoret. Comput. Sci. 500 (2013), pp. 57–67.

[16] G. Richomme, K. Saari and L.Q. Zamboni, Abelian complexity of minimal subshifts, J. Lond. Math. Soc. (2), 83 (2011), pp. 79–95.

[17] R. Tijdeman, L.Q. Zamboni, Fine and Wilf words for any periods, Indag. Math. (N.S.) 14 (2003), pp. 135–147.