



HAL
open science

Open and closed factors in Arnoux-Rauzy words

Olga Parshina, Luca Q. Zamboni

► **To cite this version:**

Olga Parshina, Luca Q. Zamboni. Open and closed factors in Arnoux-Rauzy words. *Advances in Applied Mathematics*, 2019, 107, pp.22–31. 10.1016/j.aam.2019.02.007 . hal-02072987

HAL Id: hal-02072987

<https://hal.science/hal-02072987>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Open and closed factors in Arnoux-Rauzy words *

Olga Parshina^{1,2} and Luca Zamboni¹

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

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

In Memory of the Late Professor Aldo de Luca

Abstract: Given a finite non-empty set \mathbb{A} , let \mathbb{A}^+ denote the free semigroup generated by \mathbb{A} consisting of all finite words $u_1u_2\cdots u_n$ with $u_i \in \mathbb{A}$. 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 factor $v \in \mathbb{A}^+$, meaning u contains precisely two occurrences of v , one as a prefix and one as a suffix. We study the function $f_x^c : \mathbb{N} \rightarrow \mathbb{N}$ which counts the number of closed factors of each length in an infinite word x . We derive an explicit formula for f_x^c in case x is an Arnoux-Rauzy word. As a consequence we prove that $\liminf_{n \rightarrow \infty} f_x^c(n) = +\infty$.

Keywords: Arnoux-Rauzy word, first return, complexity, return word, Sturmian word, closed word

1 Introduction

Throughout this paper we let $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\omega = \{0, 1, 2, 3, \dots\}$ be the smallest transfinite ordinal. Given a finite non-empty set \mathbb{A} , we let \mathbb{A}^+ denote the

*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.

free semigroup generated by \mathbb{A} consisting of all words $u_1u_2 \cdots u_n$ with $u_i \in \mathbb{A}$, and $\mathbb{A}^{\mathbb{N}}$ 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) $x_i x_{i+1} \cdots x_{i+n-1}$ of length n occurring in x . First introduced by Hedlund and Morse in their seminal 1938 paper [19] 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 [19] 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 $n \geq 1$.

Several notions of complexity have been successfully used in the study of infinite words and their combinatorial properties. They include Abelian complexity [1, 7, 8, 9, 20, 21], palindrome complexity [1], cyclic complexity [7], privileged complexity [20], group complexity [8] and maximal pattern complexity [17] to name just a few. In this paper we introduce and study two new complexity functions based on the notions of open and closed words [13]. 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. If u is not closed then u is said to be *open*. Thus a word $u \in \mathbb{A}^+ \setminus \mathbb{A}$ is closed if and only if it is bordered and its longest border only occurs in u as a prefix and as a suffix. The longest border of a closed word is called *frontier*. For example, *aabaaabaa* is closed and its frontier is equal to *aabaa*. In contrast, *ab* is open as it is unbordered while *abaabbababbaaba* is open since its frontier *aba* occurs internally in u . It is easily seen that all privileged words [20] are closed and hence so are all palindromic factors of rich words [14]. The terminology open and closed was first introduced by the authors in [5] although the notion of a closed word had already been introduced earlier by A. Carpi and A. de Luca in [6]. For a nice overview of open and closed words we refer the reader to the recent survey article by G. Fici [13].

For each infinite word $x \in \mathbb{A}^{\mathbb{N}}$ we consider the functions $f_x^c, f_x^o : \mathbb{N} \rightarrow \mathbb{N}$ which count the number of closed and open factors of x of each length $n \in \mathbb{N}$. In this note we investigate the function f_x^c where x is an Arnoux-Rauzy word. Arnoux-Rauzy words were first introduced in [2] in the special case of a 3-letter alphabet. They are a natural generalization of Sturmian words to alphabets of cardinality greater than two. If $x \in \mathbb{A}^{\mathbb{N}}$ is an Arnoux-Rauzy word, then $p_x(n) = (|\mathbb{A}| - 1)n + 1$ for each $n \in \mathbb{N}$. Moreover each factor u of x has precisely $|\mathbb{A}|$ distinct complete first returns in x .

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} . Since for any word $x \in \mathbb{A}^{\mathbb{N}}$ we have that $f_x^c(n) + f_x^o(n) = p_x(n)$, a formula for $f_x^c(n)$ also yields a formula for $f_x^o(n)$. Our 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 $\varepsilon = B_0, B_1, B_2, \dots$ of x , ordered according to increasing length. The second is the sequence $(p_a^{(k)})_{a \in \mathbb{A}}^{k \in \omega}$ where for each $k \in \omega$, the t coordinates of $(p_a^{(k)})_{a \in \mathbb{A}}$ are the lengths of the t first returns to B_k in x . Both sequences have been extensively studied in the literature. For each $k \in \omega$, the coordinates of $(p_a^{(k)})_{a \in \mathbb{A}}$ are coprime and each is a period of the word B_k . Moreover, each 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 [24]).

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

$$F(a, n) = \sum_{\substack{k \in \omega \\ n \in I_{k,a}}} (d(n, I_{k,a}) + 1), \quad (1)$$

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 $f_x^c(n) = \sum_{a \in \mathbb{A}} F(a, n)$.

For each fixed $n \in \mathbb{N}$ and $a \in \mathbb{A}$, the sum in (1) is finite since it only involves those k for which $n \in I_{k,a}$.

As a corollary of Theorem 1, we show that if x is an Arnoux-Rauzy word, then $\liminf f_x^c(n) = +\infty$. In contrast, it follows from [23] that if x is the regular paperfolding word, then $\liminf f_x^c(n) = 0$, in other words, for infinitely many n , all factors of x of length n are open.

We end this section by recalling a few basic notions in combinatorics on words relevant to the paper. Throughout this text \mathbb{A} will denote a finite non-empty set (the alphabet). For $n \in \mathbb{N}$, let \mathbb{A}^n denote the set of all words $a_1 a_2 \cdots a_n$ with $a_i \in \mathbb{A}$. For $u = a_1 a_2 \cdots a_n \in \mathbb{A}^n$, we let $\bar{u} \in \mathbb{A}^n$ denote the reversal of u , i.e., $\bar{u} = a_n a_{n-1} \cdots a_1$. Let $\mathbb{A}^+ = \bigcup_{n \in \mathbb{N}} \mathbb{A}^n$ denote the free semigroup generated by \mathbb{A} . For $u = a_1 a_2 \cdots a_n \in \mathbb{A}^+$ the quantity n is called the length of u and denoted $|u|$. We set $\mathbb{A}^* = \mathbb{A}^+ \cup \{\varepsilon\}$ where ε is the empty word (of length equal to 0). We let $\mathbb{A}^{\mathbb{N}}$ denote the set of all infinite words $a_1 a_2 a_3 \cdots$ with $a_i \in \mathbb{A}$. For $x \in \mathbb{A}^+ \cup \mathbb{A}^{\mathbb{N}}$ and $v \in \mathbb{A}^*$ we say that v is a factor of x if $x = uv y$ for some $u \in \mathbb{A}^*$ and $y \in \mathbb{A}^* \cup \mathbb{A}^{\mathbb{N}}$. We let $\text{Fac}(x)$ denote the set of all factors of x . A factor v of x is called right (resp. left) special if va and vb (resp. av and bv) are each factors of x for some choice of distinct $a, b \in \mathbb{A}$. A factor which is both right and left special is said to be bispecial. Given factors u and v of x , we say that u is a first return to v in x if uv is a factor of x having precisely two first occurrences of v , one as a prefix and one as a suffix. In which case the word uv is called a complete first return to v .

2 Counting closed factors in Arnoux-Rauzy words

Throughout this section we let \mathbb{A} denote a finite set of cardinality $t \geq 2$. A recurrent word $x \in \mathbb{A}^{\mathbb{N}}$ is called an Arnoux-Rauzy word if x contains, for each $n \geq 0$, precisely one right special factor R_n of length n and one left special factor L_n of length n . Furthermore, R_n is a prefix of t -many distinct factors of x of length $n + 1$ while L_n is a suffix of t -many distinct factors of x of length $n + 1$. In particular one has $p_x(n) = (t - 1)n + 1$ and each factor u of x has precisely t distinct complete first returns. In the special case of a binary alphabet, we see that x is a Sturmian word. Arnoux-Rauzy words constitute a special class of episturmian words (see [3, 12, 16]) and hence each factor u of an Arnoux-Rauzy word is (palindromically) rich, i.e., u contains exactly $|u| + 1$ many distinct palindromic factors (including the empty word ε). We will make use of the following alternative characterisation of rich words given in [12].

Lemma 2.1. *[Proposition 3 in [12]] A word $u \in \mathbb{A}^+$ is rich if and only if for every prefix v of u , the longest palindromic suffix of v is uni-occurrent in v .*

Let us now fix an Arnoux-Rauzy word $x \in \mathbb{A}^{\mathbb{N}}$. Recall that for each length $n \in \omega$ an Arnoux-Rauzy word contains either zero or one bispecial factor of length n . Let $\varepsilon = B_0, B_1, B_2, \dots$ be the sequence of bispecial factors of x ordered according to increasing length. Put $b_k = |B_k|$ so that $0 = b_0 < b_1 < b_2 < \dots$. We recall the following characterization of the bispecial factors B_k of x in terms of palindromic closures (see [10, 12]). For each $k \in \mathbb{N}$ there exists a unique $a_k \in \mathbb{A}$ such that $B_{k-1}a_k$ is a left special factor of x . The sequence $(a_k)_{k \in \mathbb{N}}$ is called the directive sequence of x . It follows that $B_{k-1}a_k$ is a prefix of B_k but in fact B_k is the palindromic closure of $B_{k-1}a_k$, i.e., the shortest palindrome beginning in $B_{k-1}a_k$. More precisely, if we let S_k denote the longest palindromic suffix of $B_{k-1}a_k$ and write $B_{k-1}a_k = x_k S_k$ with $x_k \in \mathbb{A}^*$, then $B_k = x_k S_k \overline{x_k}$ (see for instance Lemma 5 in [10] in case the alphabet \mathbb{A} is binary).

Lemma 2.2. *For each $k \in \mathbb{N}$ we have that S_k is a uni-occurrent factor of B_k . In particular S_k is not a factor of B_{k-1} .*

Proof. Clearly S_k is a factor of B_k . To see that it is uni-occurrent, suppose that S_k occurs more than once in B_k . Since S_k and B_k are each palindromes and $B_k = x_k S_k \overline{x_k}$, it follows that S_k occurs at least twice in $x_k S_k = B_{k-1}a_k$. But this contradicts Lemma 2.1 since S_k was defined as the longest palindromic suffix of $B_{k-1}a_k$. \square

Define $\varphi : \text{Fac}(x) \rightarrow \omega$ by $\varphi(v)$ is the least $k \in \omega$ such that v is a factor of B_k . In particular $\varphi(v) = 0 \Leftrightarrow v = \varepsilon$.

Lemma 2.3. *Let $k \in \mathbb{N}$ and $v \in \mathbb{A}^+$. Then $v \in \varphi^{-1}(k)$ if and only if v is a factor of B_k containing S_k as a factor. In particular each $v \in \varphi^{-1}(k)$ is uni-occurrent in B_k .*

Proof. Suppose v is a factor of B_k containing S_k as a factor. Then by Lemma 2.2, v is not a factor of B_{k-1} and hence not a factor of any B_j with $j < k$. Hence $\varphi(v) = k$. Conversely suppose that $\varphi(v) = k$. Then v is a factor of B_k but not of B_{k-1} . Since

$$B_k = x_k S_k \bar{x}_k = B_{k-1} a_k \bar{x}_k = x_k a_k B_{k-1},$$

it follows that v must contain S_k as a factor. Having established that each $v \in \varphi^{-1}(k)$ contains S_k , it follows by Lemma 2.2 that v is uni-occurrent in B_k . \square

For each $k \in \omega$ and $a \in \mathbb{A}$, let $R_a^{(k)}$ denote the complete first return to B_k in x beginning in $B_k a$ and put $p_a^{(k)} = |R_a^{(k)}| - b_k$. In other words $p_a^{(k)}$ is the length of the first return to B_k determined by $R_a^{(k)}$. We note that $R_a^{(0)} = a$ for each $a \in \mathbb{A}$. The sequence $(p_a^{(k)})_{a \in \mathbb{A}}^{k \in \omega}$ is computed recursively as follows : $p_a^{(0)} = 1$ for each $a \in \mathbb{A}$. For $k \geq 1$, we have $p_{a_k}^{(k)} = p_{a_k}^{(k-1)}$, and $p_b^{(k)} = p_b^{(k-1)} + p_{a_k}^{(k-1)}$ for $b \in \mathbb{A} \setminus \{a_k\}$. It is easily verified by induction that

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

For each $k \in \mathbb{N}$, we set $p_k = p_{a_k}^{(k)}$. Since B_k is a complete first return to B_{k-1} beginning in $B_{k-1} a_k$ i.e., $R_{a_k}^{(k-1)} = B_k$, it follows that

$$p_k = p_{a_k}^{(k)} = p_{a_k}^{(k-1)} = |R_{a_k}^{(k-1)}| - b_{k-1} = b_k - b_{k-1}. \quad (2)$$

It follows immediately from our recursive definition of the $p_a^{(k)}$ that

$$p_k = \min\{p_a^{(k)} \mid a \in \mathbb{A}\}.$$

Lemma 2.4. *Let $k \in \mathbb{N}$ and let J_k denote the interval $[b_k - 2p_k + 2, b_k]$. If $v \in \varphi^{-1}(k)$ then $|v| \in J_k$ and, for each $m \in J_k$, the set $\varphi^{-1}(k)$ contains precisely $d(m, J_k) + 1$ distinct words of length m , where $d(m, J_k)$ is the minimal distance between m and the two boundary points of the interval J_k . In particular $|\varphi^{-1}(k)| = p_k^2$.*

Proof. In view of Lemma 2.3 we have that $v \in \varphi^{-1}(k)$ if and only if v is a factor of B_k which contains S_k as a subfactor. It follows that $|S_k| \leq |v| \leq |B_k|$. Also, since $B_k = B_{k-1} a_k \bar{x}_k$, by (2) we deduce that $p_k = b_k - b_{k-1} = |x_k| + 1$. Furthermore, as $B_k = x_k S_k \bar{x}_k$ we have $|S_k| = |B_k| - 2|x_k| = b_k - 2(p_k - 1) = b_k - 2p_k + 2$. Hence $b_k - 2p_k + 2 \leq |v| \leq b_k$. Now suppose $m \in J_k$. To see that $\varphi^{-1}(k)$ contains $d(m, J_k) + 1$ distinct words of length m we simply use the fact that each $v \in \varphi^{-1}(k)$ contains S_k and is uni-occurrent in B_k (see Lemma 2.3). Finally,

$$|\varphi^{-1}(k)| = 1 + 2 + \cdots + (p_k - 1) + p_k + (p_k - 1) + \cdots + 2 + 1 = 2 \left(\frac{p_k(p_k - 1)}{2} \right) + p_k = p_k^2.$$

\square

Let $k \in \mathbb{N}$ and $v \in \varphi^{-1}(k)$. As a consequence of Lemma 2.3, there exists a unique decomposition $B_k = u_1 v u_2$ with $u_1, u_2 \in \mathbb{A}^*$. In particular, $v u_2$ is right special in x and $u_1 v$ is left special in x . Now suppose u is a closed factor of x with frontier v . In particular u begins and ends in v . Since x is recurrent and aperiodic, it follows that $v u_2$ is a prefix of u and $u_1 v$ is a suffix of u , whence $u_1 u u_2$ is a complete first return to B_k . In fact, $u_1 u u_2$ begins and ends in B_k and does not admit other occurrences of B_k for otherwise v would occur in u internally (meaning not as a prefix or as a suffix). Thus $u_1 u u_2 = R_a^{(k)}$ for some $a \in \mathbb{A}$.

Definition 2.5. Let u be a closed factor of x and $a \in \mathbb{A}$. We say u is of *type* a if and only if either $u = a$ or, if u is closed with frontier $v \in \mathbb{A}^+$, then $u_1 u u_2 = R_a^{(k)}$ where $k = \varphi(v)$ and $B_k = u_1 v u_2$.

If u is a closed factor of x of type $a \in \mathbb{A}$ and frontier $v \in \mathbb{A}^+$, then

$$|u| - |v| = |R_a^{(k)}| - (|u_1| + |u_2| + |v|) = |R_a^{(k)}| - |B_k| = p_a^{(k)}, \quad (3)$$

where $k = \varphi(v)$. We observe that the equality $|u| - |v| = p_a^{(k)}$ in (3) also holds in case $u \in \mathbb{A}$ taking $v = \varepsilon$ and $k = 0$.

Let $C(x)$ denote the set of all closed factors of x and for each $u \in C(x)$ let $\text{fr}(u) \in \mathbb{A}^*$ denote its frontier. By convention we define $\text{fr}(a) = \varepsilon$ for each $a \in \mathbb{A}$. For each $k \in \omega$ and $a \in \mathbb{A}$ we let $C_{k,a}(x)$ denote the set of all closed factors u of x of type a whose frontier $\text{fr}(u)$ belongs to $\varphi^{-1}(k)$.

Lemma 2.6. *The sets $\{C_{k,a}(x) : k \in \omega, a \in \mathbb{A}\}$ define a partition of $C(x)$ and $\text{fr} : C_{k,a}(x) \rightarrow \varphi^{-1}(k)$ is a bijection.*

Proof. Each closed factor $u \in C(x)$ has a unique type and its frontier $\text{fr}(u)$ belongs to $\varphi^{-1}(k)$ for a unique value of $k \in \omega$. Whence each closed factor u of x belongs to a unique $C_{k,a}(x)$. By definition, if $u \in C_{k,a}(x)$ then $\text{fr}(u) \in \varphi^{-1}(k)$. Moreover u is uniquely determined by its frontier $\text{fr}(u)$ and its type. In fact, if $u \in C_{k,a}(x)$ then $u_1 u u_2 = R_a^{(k)}$ where u_1, u_2 are determined by the (unique) factorization $B_k = u_1 \text{fr}(u) u_2$. This proves fr is injective. To see that fr is also surjective, let $v \in \varphi^{-1}(k)$. Then we can write $B_k = u_1 v u_2$ for some $u_1, u_2 \in \mathbb{A}^*$. Hence $R_a^{(k)}$ begins in u_1 and ends in u_2 . It follows that $u = u_1^{-1} R_a^{(k)} u_2^{-1}$ is a closed factor of x of type a and $\text{fr}(u) = v$. \square

Proof of Theorem 1. Fix $n \in \mathbb{N}$. By Lemma 2.6 we have

$$f_x^c(n) = |C(x) \cap \mathbb{A}^n| = \sum_{\substack{k \in \omega \\ a \in \mathbb{A}}} |C_{k,a}(x) \cap \mathbb{A}^n|.$$

Now assume $u \in C_{k,a} \cap \mathbb{A}^n$ and put $v = \text{fr}(u) \in \varphi^{-1}(k)$. Then by (3) we have that $n = |u| = |v| + p_a^{(k)}$. By Lemma 2.4, $|v| = n - p_a^{(k)} \in J_k = [b_k - 2p_k + 2, b_k]$. By Lemma 2.6 the number of words $u \in C_{k,a}(x) \cap \mathbb{A}^n$ is equal to the number of words $v \in \varphi^{-1}(k)$ of length $n - p_a^{(k)}$ which by Lemma 2.4 is equal to $d(n - p_a^{(k)}, J_k) + 1 = d(n, I_{k,a}) + 1$ where $I_{k,a} = [b_k - 2p_k + p_a^{(k)} + 2, b_k + p_a^{(k)}]$. This completes the proof of Theorem 1. \square

In case $|\mathbb{A}| = 2$, i.e., x is Sturmian, each bispecial factor B_k has precisely two first returns, the shortest one is of length p_k , and we let q_k denote the length of the other first return. So for fixed $a \in \mathbb{A}$ and $k \in \mathbb{N}$ we have

$$p_a^{(k)} = \begin{cases} p_k, & \text{if } a = a_k; \\ q_k, & \text{otherwise.} \end{cases}$$

If $a = a_k$ then $I_{k,a} = [q_k, q_k + 2p_k - 2]$ and if $a \neq a_k$ then $I_{k,a} = [2q_k - p_k, 2q_k + p_k - 2]$. Putting $P_k = [q_k, q_k + 2p_k - 2]$ and $Q_k = [2q_k - p_k, 2q_k + p_k - 2]$, we obtain that for a Sturmian word x the number of closed factors of x of each length n is given by

$$f_x^c(n) = \sum_{\substack{k \in \omega \\ n \in P_k}} (d(n, P_k) + 1) + \sum_{\substack{k \in \omega \\ n \in Q_k}} (d(n, Q_k) + 1). \quad (4)$$

Example 2.7. Consider the Fibonacci word

$$x_{fib} = abaababaabaababaa \dots$$

fixed by the morphism $a \mapsto ab$, $b \mapsto a$. Then $p_k = F_k$ and $q_k = F_{k+1}$ where the sequence $(F_k)_{k \in \omega}$ is the Fibonacci sequence given by $F_0 = F_1 = 1$ and $F_{k+1} = F_k + F_{k-1}$ for $k \geq 1$.

Table 1 shows the number of closed factors of length $n \leq 15$ in the Fibonacci word computed using (4).

Table 1: The number of closed factors in the Fibonacci word.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$f_{x_{fib}}^c(n)$	2	1	2	3	4	3	4	5	6	5	6	7	8	9	10

For example, for $n = 11$ we must determine those k for which either $11 \in P_k$ or $11 \in Q_k$. It is easily checked that 11 only belongs to $P_4 = [8, 16]$, $Q_3 = [7, 11]$ and $Q_4 = [11, 19]$. So

$$f_{x_{fib}}^c(11) = d(11, P_4) + 1 + d(11, Q_3) + 1 + d(11, Q_4) + 1 = 4 + 1 + 1 = 6.$$

The graph of the function $f_{x_{fib}}^c$ is shown in Figure 1. The function is clearly not monotone.

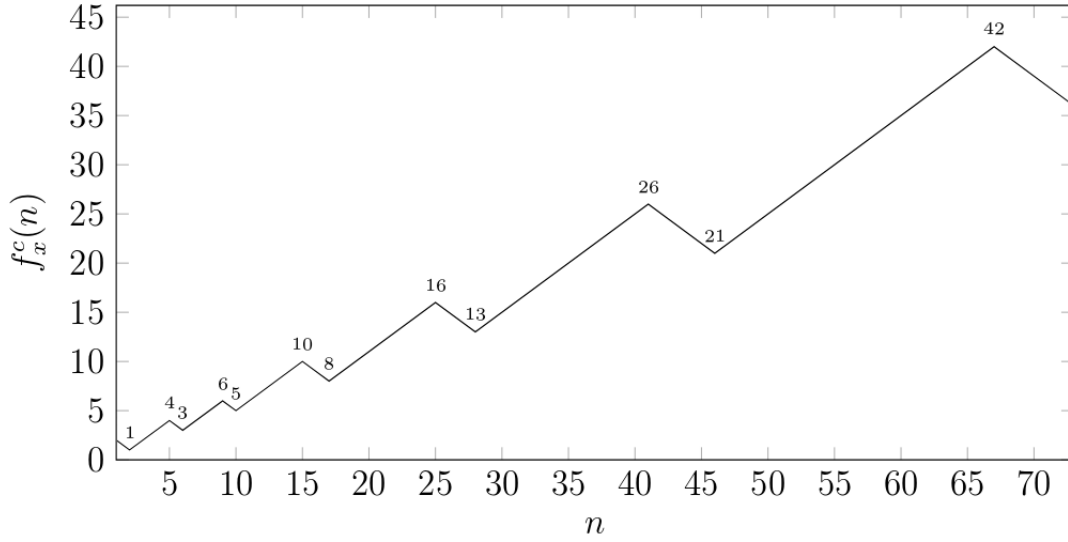


Figure 1: The number of closed factors in the Fibonacci word.

Figure 2 illustrates the behavior of the number of closed factors of the Tribonacci word $x_{trib} \in \{a, b, c\}^{\mathbb{N}}$ defined as the fixed point of the morphism $a \mapsto ab$, $b \mapsto ac$, $c \mapsto a$.

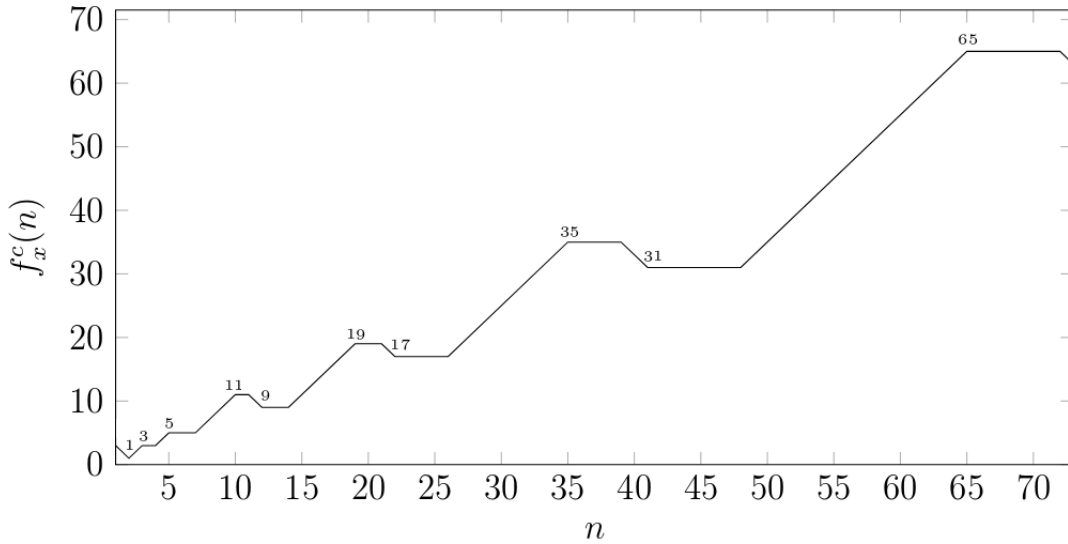


Figure 2: The number of closed factors in the Tribonacci word.

Our last example (Figure 3) illustrates the behavior of the number of closed factors of the Sturmian word $x_r \in \{0, 1\}^{\mathbb{N}}$ whose directive sequence begins with $0010000011110111110101101110011000011 \dots$.

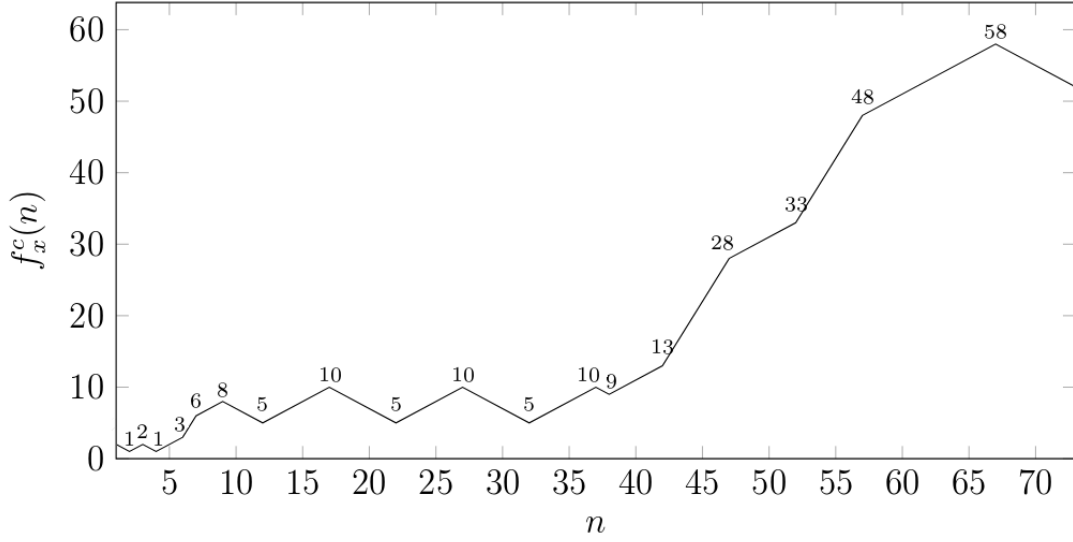


Figure 3: The number of closed factors in the word x_r .

The above examples suggest that the function $f_x^c(n)$ tends to infinity, although it need not be monotone and may contain plateaus and inflection points. Our next result establishes this fact:

Corollary 2.8. *If $x \in \mathbb{A}^{\mathbb{N}}$ is an Arnoux-Rauzy word, then*

$$\liminf_{n \rightarrow \infty} f_x^c(n) = +\infty.$$

Proof. For each $k \in \omega$ set $I_k = [b_k - p_k + 2, b_k + p_k]$, i.e., $I_0 = [1, 1]$ and $I_k = I_{a_k, k}$ for $k \geq 1$. Since $p_k \geq 1$ for each $k \in \mathbb{N}$ it follows that $\mathbb{N} = \bigcup_{k \in \omega} I_k$. Given $m \in \mathbb{N}$, pick j such that $p_j - 1 > 2m$, and put $N = b_j + 2$. We will show that $f_x^c(n) \geq m$ for every $n \geq N$. Notice that since $b_k = b_{k-1} + p_k$, the left hand endpoint of the interval I_{j+1} is $b_j + 2$. Thus for each $n \geq N$ there exists a positive integer $k \geq j$, such that $n \in I_{k+1}$. We have $|I_{k+1}| = 2p_{k+1} - 2 \geq 2p_j - 2 > 4m$. If $d(n, I_{k+1}) \geq m$, then it follows from Theorem 1 that $f_w^c(n) \geq m$. Otherwise we must have either i) $b_{k+1} + p_{k+1} - n < m$ or ii) $n - (b_k + 2) < m$. In case i) we have

$$n - b_{k+1} > p_{k+1} - m \geq p_j - m > m + 1. \quad (5)$$

Since $m + 1 \geq 2$, we have that n also belongs to I_{k+2} . We will show that $d(n, I_{k+2}) \geq m$. By (5) we have $n - (b_{k+1} + 2) \geq m$. Also

$$b_{k+2} + p_{k+2} - n \geq b_{k+2} + p_{k+2} - (b_{k+1} + p_{k+1}) = 2p_{k+2} - p_{k+1} \geq p_{k+2} \geq p_j > 2m + 1.$$

Thus $d(n, I_{k+2}) \geq m$ and hence by Theorem 1 $f_x^c(n) \geq m$.

In case ii) $n < b_k + 2 + m \leq b_k + 2m + 1 < b_k + p_k$, and hence $n \in I_k$. We will show that $d(n, I_k) \geq m$. In fact, $b_k + p_k - n > p_k - 2 - m \geq p_j - 2 - m \geq m$. Moreover, since $n \in I_{k+1}$, we have that $n - (b_{k-1} + 2) \geq b_k + 2 - (b_{k-1} + 2) = p_k \geq p_j > 2m + 1$, and thus $d(n, I_k) \geq m$. \square

While the previous result applies to Arnoux-Rauzy words, for a general aperiodic word x the limit inferior of the function $f_x^c(n)$ need not be infinite. For example, in the case of the regular paperfolding word one has that $\liminf_{n \rightarrow \infty} f_x^c(n) = 0$. In fact, in [23] the authors exhibit an 11-state automaton which accepts the base 2 representation of those n for which there is a closed factor of the regular paperfolding word of length n (see Figure 1 in [23]). As another perhaps simpler example, let x be the fixed point beginning in a of the 2-uniform morphism φ on the alphabet $\{a, b, c, d\}$ given by $\varphi : a \mapsto ac, b \mapsto ad, c \mapsto bc, d \mapsto bd$. Then it is easily shown that all factors of x of length 2^n ($n \in \mathbb{N}$) are open. We remark that this last example is closely related to the regular paperfolding word. In fact, the regular paperfolding word is the image of the fixed point of φ under the mapping which sends a, c to 0 and b, d to 1.

References

- [1] J.-P. Allouche, M. Baake, J. Cassaigne, 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] J. Berstel, Sturmian and episturmian words (A survey of some recent results), in: Proceedings of CAI 2007, Lecture Notes in Computer Science, vol. 4728, 2007, pp. 23–47.
- [4] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, Internat. J. Found. Comput. Sci., 15 (2004), pp. 293–306.
- [5] M. Bucci, A. De Luca, G. Fici, Enumeration and structure of trapezoidal words, Theoret. Comput. Sci. 468 (2013), pp. 12–22.
- [6] A. Carpi, A. de Luca, Periodic-like words, periodicity and boxes, Acta. Inform. 37 (2001), pp. 597–618.
- [7] J. Cassaigne, G. Fici, M. Sciortino, L.Q. Zamboni, Cyclic complexity of words, J. Combin. Theory Ser. A, 145 (2017), pp. 36–56.

- [8] 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.
- [9] E. M. Coven, G. A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory* 7 (1973), pp. 138–153.
- [10] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, *Theoret. Comput. Sci.* 183 (1997), pp. 45–82.
- [11] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words. *Theoret. Comput. Sci.*, 136 (1994), pp. 361–385
- [12] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* 255 (2001), pp. 539–553.
- [13] G. Fici, Open and closed words, *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* No. 123 (2017), p. 138–147.
- [14] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, *European J. Combin.* 30 (2009), no. 2, pp. 510–531.
- [15] J. Justin, On a paper by Castelli, Mignosi, Restivo, *RAIRO: Theoret. Informatics Appl.* 34 (2000), pp. 373–377.
- [16] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* 276 (2002), pp. 281–313.
- [17] T. Kamae, L.Q. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words, *Ergodic Theory Dynam. Systems* 22 (2002), pp. 1191–1199.
- [18] M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading, MA, 1983)
- [19] M. Morse, G. Hedlund, Symbolic dynamics, *Amer. J. Math.* 60 (1938), pp. 815–866.
- [20] J. Peltomäki, Introducing privileged words: privileged complexity of Sturmian words, *Theoret. Comput. Sci.* 500 (2013), pp. 57–67.
- [21] G. Richomme, K. Saari, L.Q. Zamboni, Abelian complexity of minimal subshifts, *J. Lond. Math. Soc. (2)*, 83 (2011), pp. 79–95.
- [22] R. Risley, L. Zamboni, A generalization of Sturmian sequences: Combinatorial structure and transcendence, *Acta Arithmetica* 95.2 (2000), pp. 167–184.
- [23] L. Schaeffer, J. Shallit, Closed, palindromic, rich, privileged, trapezoidal, and balanced words in automatic sequences, *Electron. J. Comb.*, 23(1) (2016), pp. 1–25.

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