On closed and open factors in Arnoux-Rauzy words

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

[1]Institut Camille Jordan UCBL1, Lyon, France
[2]Sobolev Institute of mathematics SB RAS, Novosibirsk, Russia

17th Mons theoretical computer science days
10-14 September, 2018

Notation

- A finite alphabet A is a finite set of symbols
- A finite word over A is a sequence $u = u_0 u_1 \cdots u_{n-1}, u_i \in A, n \geq 0$
- The length of the word u is $|u| = n$
- $u = \mathit{prefix}\,\mathit{factor}\,\mathit{suffix}$, $n = 24$
- The empty word is denoted by ε, $|\varepsilon| = 0$
- An infinite word over A is a sequence $x = x_0 x_1 x_2 \cdots, x_i \in A$

Special factors

Let A be a finite alphabet, and v be a word over A

The factor u of v is

- right special, if ua and ub are factors of v for $a, b \in A$, $a \neq b$
- left special, if au and bu are factors of v for $a, b \in A$, $a \neq b$
- bispecial, if it is both right and left special

Example: $v = independent$
nd is a left special factor of v
de is a right special factor of v
en is a bispecial factor of v

Special factors

Let A be a finite alphabet, and v be a word over A

The factor u of v is

- right special, if ua and ub are factors of v for $a, b \in A$, $a \neq b$
- left special, if au and bu are factors of v for $a, b \in A$, $a \neq b$
- bispecial, if it is both right and left special

Example: $v = independent$
nd is a left special factor of v
de is a right special factor of v
en is a bispecial factor of v

Special factors

Let A be a finite alphabet, and v be a word over A

The factor u of v is

- right special, if ua and ub are factors of v for $a, b \in A$, $a \neq b$
- left special, if au and bu are factors of v for $a, b \in A$, $a \neq b$
- bispecial, if it is both right and left special

Example: $v = independent$

nd is a left special factor of v

de is a right special factor of v

en is a bispecial factor of v

Special factors

Let A be a finite alphabet, and v be a word over A

The factor u of v is

- right special, if ua and ub are factors of v for $a, b \in A$, $a \neq b$
- left special, if au and bu are factors of v for $a, b \in A$, $a \neq b$
- bispecial, if it is both right and left special

Example: $v = independent$
nd is a left special factor of v
de is a right special factor of v
en is a bispecial factor of v

Arnoux-Rauzy words ★

Consider an alphabet $A_t = \{0, 1, \ldots, t-1\}, t \geq 2$.

- A sequence $x \in A_t^\omega$ is an Arnoux-Rauzy word, if for all $n \in \mathbb{N}$ it has $f_x(n) = (t-1)n + 1$ factors of length n, with exactly one right special factor and one left special factor of length n
- In case $t = 2$ the word x is Sturmian

Example

Tribonacci word $x = 0102010010201010201001020 10\cdots$

- $f_x(1) = 3$, bispecial factor of length 1 is **0**
- $f_x(2) = 5$, **10** is right special, **01** is left special

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

Open and closed words★

Let A be a finite alphabet, and v be a finite word over A

- The word v has a border, if it has a proper factor occurring both as a prefix and as a suffix of v:

  $v = ahaha$ is a bordered word with borders a and aha

- A frontier of v is a border having no internal occurrences in v

- A word v is closed if it is of length ≤ 1 or it has a frontier, otherwise v is open

  $v = local$ is closed, l is a frontier of v
  $v = maximum$ is open, m is a border of v
  $v = value$ is unbordered and open

★ Fici, G.: A Classification of Trapezoidal Words. In: WORDS 2011, 8th International Conference on Words. No. 63 in Electronic Proceedings in Theoretical Computer Science (2011), pp. 129–137

Open and closed words★

Let A be a finite alphabet, and v be a finite word over A

- The word v has a border, if it has a proper factor occurring both as a prefix and as a suffix of v:

  $v = ahaha$ is a bordered word with borders a and aha

- A frontier of v is a border having no internal occurrences in v

- A word v is closed if it is of length ≤ 1 or it has a frontier, otherwise v is open

  $v = local$ is closed, l is a frontier of v
  $v = maximum$ is open, m is a border of v
  $v = value$ is unbordered and open

★ Fici, G.: A Classification of Trapezoidal Words. In: WORDS 2011, 8th International Conference on Words. No. 63 in Electronic Proceedings in Theoretical Computer Science (2011), pp. 129–137

Open and closed words★

Let A be a finite alphabet, and v be a finite word over A

- The word v has a border, if it has a proper factor occurring both as a prefix and as a suffix of v:

  $v = ahaha$ is a bordered word with borders a and aha

- A frontier of v is a border having no internal occurrences in v

- A word v is closed if it is of length ≤ 1 or it has a frontier, otherwise v is open

  $v = local$ is closed, l is a frontier of v
  $v = maximum$ is open, m is a border of v
  $v = value$ is unbordered and open

★ Fici, G.: A Classification of Trapezoidal Words. In: WORDS 2011, 8th International Conference on Words. No. 63 in Electronic Proceedings in Theoretical Computer Science (2011), pp. 129–137

Open and closed words*

Let A be a finite alphabet, and v be a finite word over A

- The word v has a border, if it has a proper factor occurring both as a prefix and as a suffix of v:

  $v = ahaha$ is a bordered word with borders a and aha

- A frontier of v is a border having no internal occurrences in v

- A word v is closed if it is of length ≤ 1 or it has a frontier, otherwise v is open

  $v = local$ is closed, l is a frontier of v

  $v = maximum$ is open, m is a border of v

  $v = value$ is unbordered and open

* Fici, G.: A Classification of Trapezoidal Words. In: WORDS 2011, 8th International Conference on Words. No. 63 in Electronic Proceedings in Theoretical Computer Science (2011), pp. 129–137

Open and closed words

Let A be a finite alphabet, and v be a finite word over A

- The word v has a border, if it has a proper factor occurring both as a prefix and as a suffix of v
- A frontier of v is a border having no internal occurrences in v
- A word v is closed if it is a letter or it has a frontier, otherwise v is open

Closed words are also known as periodic-like words★

- A finite word v is called periodic-like if its longest repeated prefix is not a right-special factor of v

★ A.Carpi, A.de Luca: Periodic-like words, periodicity and boxes. Acta. Inform. 37 (2001), pp. 597-618

Open and closed words

Let A be a finite alphabet, and v be a finite word over A

- The word v has a border, if it has a proper factor occurring both as a prefix and as a suffix of v
- A frontier of v is a border having no internal occurrences in v
- A word v is closed if it is a letter or it has a frontier, otherwise v is open

Closed words are also known as complete first returns

- A finite word v is a complete first return to a word u if it has exactly two occurrences of u, one as a prefix and one as a suffix

Complexity functions

Let $t \geq 2$ and $x \in A_t^\omega$ be an Arnoux-Rauzy word, $n \geq 0$

- The function $f_x(n)$ counts the number of factors of length n in x
- $f_x(n) = (t-1)n + 1$
- The function $f_x^c(n)$ counts the number of closed factors of length n in x
- The function $f_x^o(n)$ counts the number of open factors of length n in x
- $f_x^c(n) + f_x^o(n) = f_x(n)$

Aim: To study the function $f_x^c(n)$

Periods of finite Arnoux-Rauzy factors

A period of the word u is a positive integer p,

$$\text{such that } u_i = u_{i+p} \text{ for } i = 0, 1, \cdots |u| - p.$$

Theorem [N.J.Fine, H.S.Wilf (1965)]

If a word u has periods p and q, and has length at least $p + q - gcd(p, q)$, then u has also period $gcd(p, q)$.

[A. de Luca, F.Mignosi (1994)]

If u is a bispecial factor of a Sturmian word, then u has two coprime periods p and q, s.t. $|u| = p + q - 2$.

[J.Justin (2000)]

If u is a bispecial factor of an Arnoux-Rauzy word over $A_t, t \geq 2$, then u has t coprime periods $p_0, p_1, \ldots, p_{t-1}$, such that

$$|u| = \frac{p_0 + p_1 + \cdots + p_{t-1} - t}{t - 1}.$$

Periods of finite Arnoux-Rauzy factors

[J.Justin (2000)]

If u is a bispecial factor of an Arnoux-Rauzy word over $A_t, t \geq 2$,

then u has t coprime periods $p_0, p_1, \ldots, p_{t-1}$, such that $|u| = \dfrac{\sum_{i=0}^{t-1} p_i \; - t}{t-1}$.

Let us call $p_0, p_1, \ldots, p_{t-1}$ critical periods of u.

Lemma 1

Let $B = \{B_k\}_{k=0}^{\infty}$ be the sequence of bispecial factors of an Arnoux-Rauzy word $x \in A_t^{\omega}$, enumerated by increasing of length; the word x' be the unique accumulation point of B, and let $p_0^k, p_1^k, \ldots, p_{t-1}^k$ be crirical periods of B_k. Then:

- For $B_0 = \varepsilon$, $p_i^0 = 1$ for every i;
- If $B_k s$ is a prefix of x' for $k \geq 0$, $s \in A_t$,

  then $p_s^{k+1} = p_s^k$, and $p_r^{k+1} = p_r^k + p_s^k$ for $r \in A \backslash \{s\}$.

The main formula

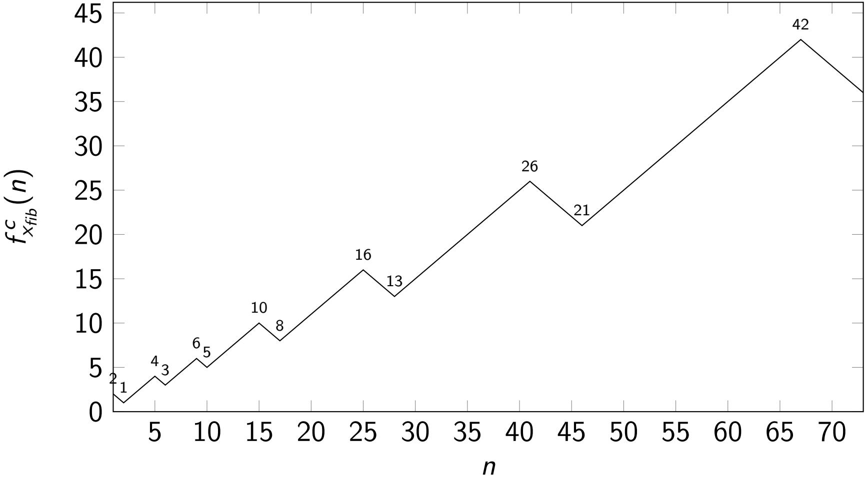
- Arnoux-Rauzy word $\mathbf{x}$ over A_t
- Set of bispecial factors of x is $\{\mathbf{B_i}\}_{\mathbf{i=0}}^{\infty}$, the length of B_i is $\mathbf{b_i}$
- The set of critical periods of B_i is $\{\mathbf{p_0^i}, \mathbf{p_1^i}, \ldots, \mathbf{p_{t-1}^i}\}$

- Take $k \in \omega$ and $a \in A_t$
- The minimal among critical periods of B_k is $\mathbf{p_k}$
- Set the interval $I_{k,a} = [b_k - 2p_k + p_a^k + 2, b_k + p_a^k]$
- The minimal distance from $n \in I_{k,a}$ to the endpoints of $I_{k,a}$ is $\mathbf{d(n, I_{k,a})}$
- The function counting the number of closed factors of length n in x is $\mathbf{f_x^c(n)}$

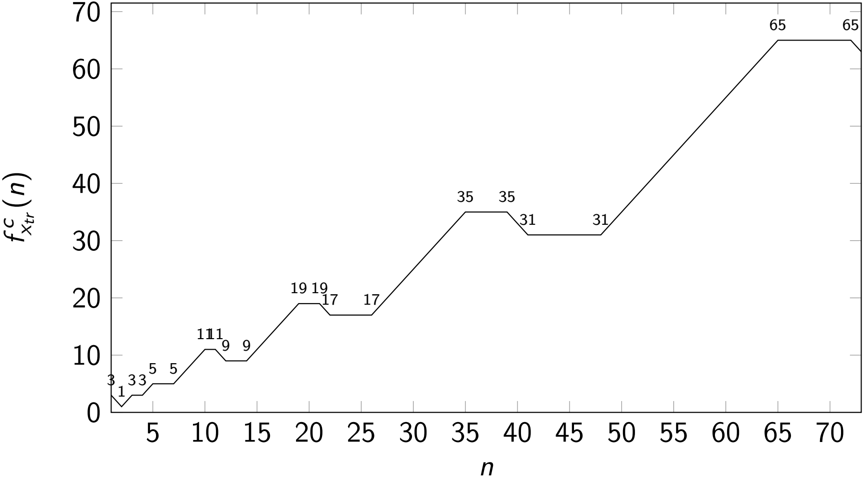
The following holds:

$$f_x^c(n) = \sum_{a \in A_t} \sum_{\substack{k \in \omega \\ n \in I_{k,a}}} (d(n, I_{k,a}) + 1)$$

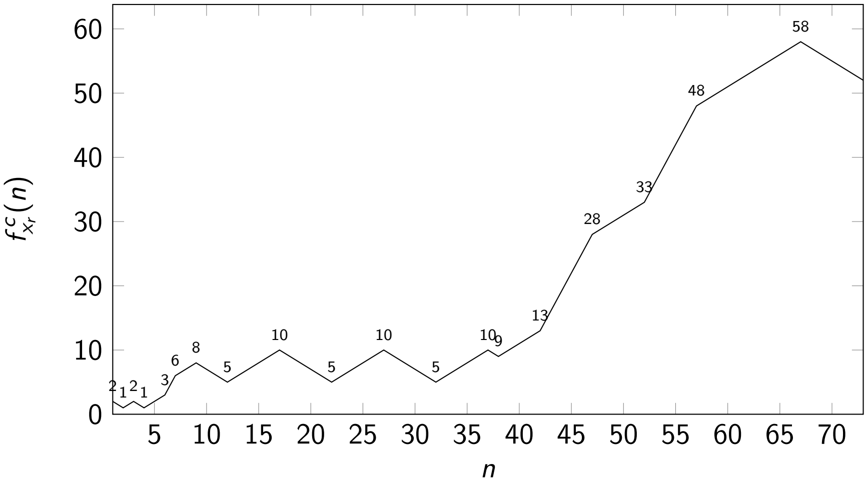
Fibonacci word



Tribonacci word



Sturmian word $x_r = 00010000100001000010000100001000010000\cdots$



Lemma 2

Let x be an Arnoux-Rauzy word. Then

$$\liminf_{n->\infty} f_x^c(n) = +\infty$$

Lemma 2

Let x be an Arnoux-Rauzy word. Then

$$\liminf_{n->\infty} f_x^c(n) = +\infty$$

Lemma 3

Consider the morphism ϕ over the alphabet A_4,
such that $\phi(0) = 02, \phi(1) = 03, \phi(2) = 12, \phi(3) = 14.$
Let y be a fixed point of ϕ begining with letter 0. Then

$$\liminf_{n->\infty} f_y^c(n) = 0$$

$$\mathbf{y} = \mathbf{021203120213031202120313021303120212031202} \cdots$$

Thank you for your attention!