A characterization of balanced episturmian sequences

Laurent Vuillon

LAMA, Université de Savoie, (LACIM, UQAM, CRM)

Joint work with Geneviève Paquin (LACIM, UQAM)

Outline

- Sturmian sequences
- Arnoux-Rauzy sequences
- Balance and Imbalance
- Balanced non ultimately periodic sequences
- Balanced periodic sequences and Fraenkel's conjecture
- Balanced episturmian words

Iterated morphisms

Alphabet $\mathcal{A} = \{a, b\}$.

$\sigma(a) = ab$ et $\sigma(b) = a$.

Rules:

If $w = w_1 w_2 \cdots w_n$ with $w_i \in \mathcal{A}$ then

$\sigma(w_1 w_2 \cdots w_n) = \sigma(w_1)\sigma(w_2) \cdots \sigma(w_n)$

Fibonacci sequence

Fixed point of $\sigma(a) = ab$, $\sigma(b) = a$

We search $X \in \mathcal{A}^{\mathbb{N}}$ such that $\sigma(X) = X$

Iterations:

$\sigma(a) = ab$

$\sigma^2(a) = \sigma(ab) = ab.a$

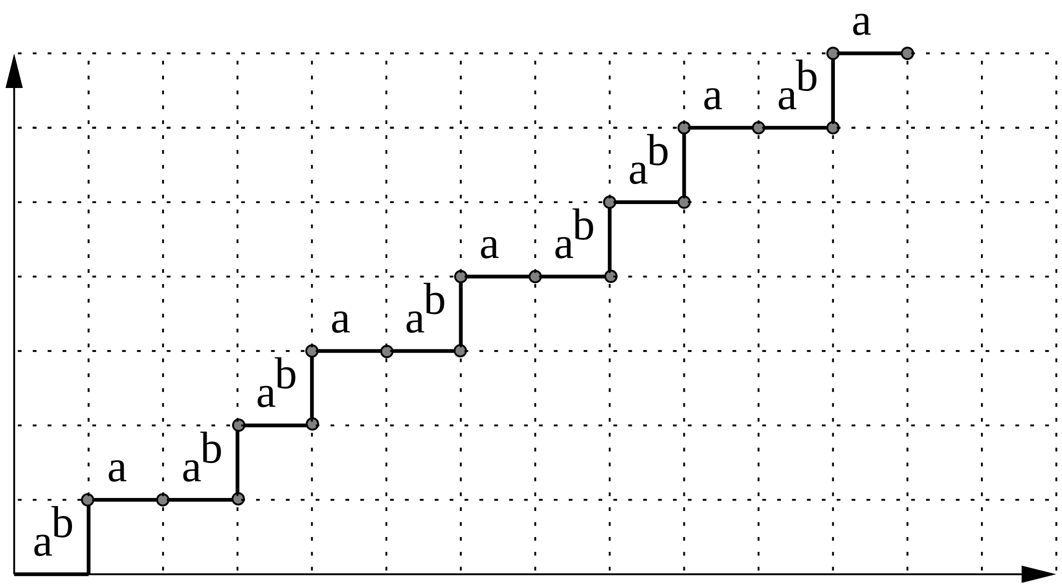
$\sigma^3(a) = abaab = aba.ab$

$\sigma^4(a) = abaab.aba$

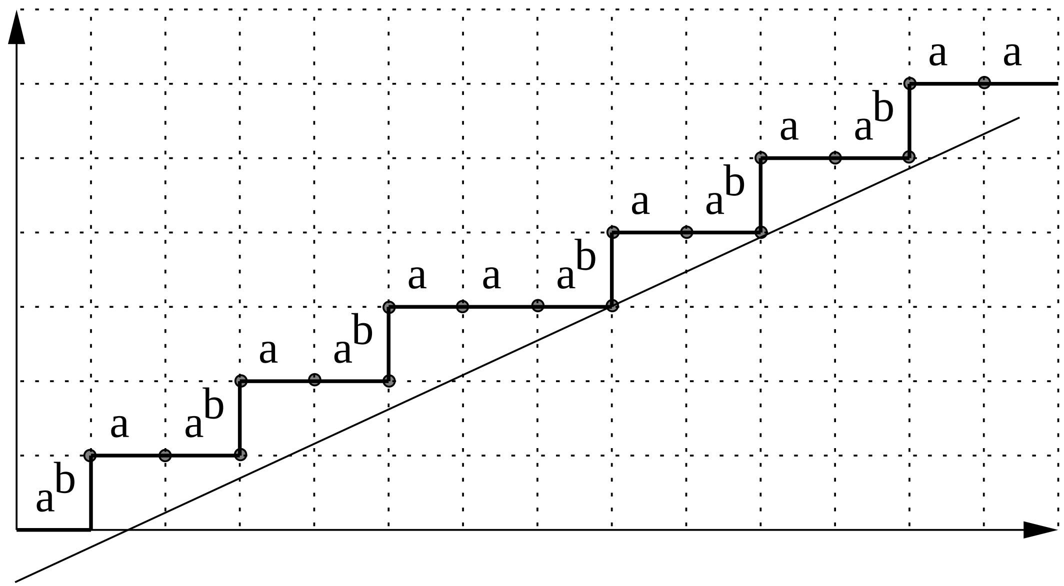
...

$\sigma^{n+2}(a) = \sigma^{n+1}(a).\sigma^n(a)$

Representation of the Fibonacci sequence



Discrete line



Definition

An infinite sequence X is a *sturmian sequence if the complexity function of* X *is given by* $p(n) = n + 1$ *for all* n.

$$X = abaababaabaababaababaabaabab \cdots$$

$$p(1) = \textit{Card } \{a, b\} = 2,$$

$$p(2) = \textit{Card } \{ab, ba, aa\} = 3,$$

$$p(3) = \textit{Card } \{aba, baa, bab, aab\} = 4.$$

For each length there exists a unique word with two right prolongations and a unique word with two left prolongations.

Generalization

Tribonacci sequence $\sigma(1) = 12$, $\sigma(2) = 13$ and $\sigma(3) = 1$.

Fixed point: $\sigma(T) = T = 1213121121312121312112131 21\cdots$

An Arnoux-Rauzy sequence (1991) is an non ultimately periodic and recurrent infinite sequence on a three-letter alphabet such that:

for each length there exists a unique word with three right prolongations and a unique word with three left prolongations.

Example for the length one: $12, 13, 11$ and $21, 31, 11$

Thus the complexity function for Arnoux-Rauzy sequences is

$p(n) = 2n + 1$.

Balance

Theorem .1 (Morse, Hedlund (1940)). *A non ultimately periodic infinite sequence x is a sturmian sequence if and only if*

$$\forall n \in \mathbb{N}, \forall w, w' \in L_n(x)\ ||w|_a - |w'|_a| \leqslant 1.$$

$$x = abaababaabaababaababaabaabab\cdots$$

$$w = aaba$$

$$w' = baab$$

Imbalanced in Arnoux Rauzy sequences

Definition .2. *An infinite sequence x is a c-balanced sequence on each letter if* $\forall a \in \mathcal{A}, \forall n \in \mathbb{N}, \forall w, w' \in L_n(x)\ ||w|_a - |w'|_a| \leqslant c$.

Fixed point: $\sigma(T) = T = 1213121121312121312112131 21 \cdots$

We can check that the Tribonacci sequence is not 1-balanced (not balanced).

$$w = 131, w' = 212$$

The Tribonacci sequence is 2-balanced on each letter.

For each $i \in \{1,2,3\}, \sigma_i(i) = i,\ \sigma_i(j) = ij, i \neq j$

σ_k of an AR-sequence remains an AR-sequence.

$$\Sigma = \sigma_1^2 \sigma_2^2 \sigma_1 \sigma_3$$

Σ increases the Imbalanced of AR-sequences:

$$\Sigma(212) = 2112111211211211121112112, |\Sigma(212)|_2 = 7$$

$$\Sigma(131) = 113112112111211211311, |\Sigma(131)|_2 = 4$$

$\Sigma(T)$ is an AR-sequence which is not 2-balanced.

Theorem .3 (Cassaigne, Ferenczi, Zamboni (2000)). *One can construct an Arnoux-Rauzy sequence which is not c-balanced for any c.*

1-Balance

$\forall a \in \mathcal{A}, \forall n \in \mathbb{N}, \forall w, w' \in L_n(x), ||w|_a - |w'|_a| \leqslant 1.$

Graham (1973) and Hubert (2000) show that infinite non ultimately periodic sequences balanced on each letter are constructed by a modification of Sturmian sequences.

For example, we build a non ultimately periodic sequence on a four-letter alphabet by modification of the Fibonacci sequence:

$$X = abaababaabaababaababaabaabab \cdots$$

We replace periodically a by $(a_1, a_2, \cdots a_{k_1})$ and b by $(b_1, b_2, \cdots b_{k_2})$.

Where a_i's and b_j's are pairwise distincts.

An infinite periodic sequence w has *constant gaps if the number of letters between two occurrences of successive letter a_i of w is constant for each i.*

Words $(a_1, a_2, \cdots, a_{k_1})$ and $(b_1, b_2, \cdots, b_{k_2})$ must be with constant gaps.

Examples

$(abac)^\omega$ *is with constant gaps;*

$(abaac)^\omega$ *is not with constant gaps.*

$$X = abaababaabaababaababaabaabab \cdots ,$$

we replace periodically the occurrences of the letter a by the constant gaps word $(cdce)^\omega$:

$$cbdcbebcdbcebcbdcbebcdbcebcb \cdots .$$

Covering of integers

Graham (1973) presents is result using covering of integers by Beatty sequences of the form $\lceil \alpha n + \beta \rceil$.

Theorem .4 (Skolem (1957)-Fraenkel (1973)). *The Beatty sequences* $\lceil \alpha_1 n + \beta_1 \rceil$ *et* $\lceil \alpha_2 n + \beta_2 \rceil$ *cover the integer if and only if*

$$\frac{1}{\alpha_1} + \frac{1}{\alpha_2} = 1 \text{ et } \frac{\beta_1}{\alpha_1} + \frac{\beta_2}{\alpha_2} \in \mathbb{Z}.$$

$$x = abaababaabaababaababaabaabab \cdots ,$$

the first Beatty sequence gives the following set of indices

$$\{1, 3, 4, 6, 8, 9, 11, 12, 14, 16, 17, 19 \cdots\}$$

and the second

$$\{2, 5, 7, 10, 13, 15, 18 \cdots\}.$$

If we cover the integers by three or more Beatty sequences $\lceil \alpha_i n + \beta_i \rceil$ (with $i = 1, 2, \cdots k, k \geqslant 3$) with all distinct frequencies of letters (i.e. the α_i's two by two distinct), then Graham (1973) shows that the coefficients α_i remain rational.

This implies in particular that the associated infinite sequence is periodic and balanced.

Thus we are searching periodic balanced sequences.

Periodic Balanced sequences

Conjecture .5 (Fraenkel (1973)). *The unique solution (up to a permutation of letters) of balanced sequence on each letter where* $|\mathcal{A}| = k \geqslant 3$ *with all distinct frequencies of letters is*

$$(Fr_k)^\omega = (Fr_{k-1} k Fr_{k-1})^\omega$$

where $Fr_3 = 1213121$.

Conjecture is true for $k = 3, 4, 5, 6$ by Altman, Gaujal, Hordijk, Tijdeman. For $k = 7$ by Barát, Varjú. Several Classes by Fraenkel and Simpson.

Open problem: find a tractable characterization of periodic balanced sequences.

Idea for a subclass : to control sequences using palindromic closure.

Two different generalizations of Sturmian sequences :

- The set of episturmian sequences;

- The set of balanced sequences.

The two notions coincide for Sturmian sequences, which are both non ultimately periodic episturmian and non ultimately periodic balanced sequences over a 2-letter alphabet.

When the alphabet has 3 letters or more, the two notions no longer coincide.

In particular, episturmian sequences are generally unbalanced over a k-letter alphabet, for $k \geqslant 3$.

Question: which sequences are both episturmian and balanced.

Definitions

The *palindromic right closure [de Luca (1997)]* of $w \in \mathcal{A}^*$ is the *shortest palindrome* $u = w^{(+)}$ *with* w *as prefix.*

The set of factors of $s \in \mathcal{A}^\omega$ *is denoted* $F(s)$ *and* $F_n(s) = F(s) \cap \mathcal{A}^n$ *is the set of all factors of* s *of length* $n \in \mathbb{N}$.

The alphabet of s *is* Alph$(s) = F(s) \cap \mathcal{A}$ *and* Ult(s) *is the set of letters occuring infinitely often in* s.

Definition of standard episturmian sequences introduced by Droubay, Justin and Pirillo (2001):

Definition .6. *An infinite sequence* s *is* *standard episturmian* *if it satisfies the following condition:*

There exists an infinite sequence $u_1 = \varepsilon, u_2, u_3, \ldots$ *of palindromes and an infinite sequence* $\Delta(s) = x_1 x_2 \ldots.$, $x_i \in \mathcal{A}$, *such that each of the words* u_{n+1} *defined by* $u_{n+1} = (u_n x_n)^{(+)}$, $n \geqslant 1$, *with* $u_1 = \varepsilon$, *is a prefix of* s. *Then* $\mathrm{Pal}(x_1 x_2 \cdots x_n)$ *denotes the word* u_{n+1}.

The sequence $\Delta(s)$ is called the *directive sequence of the standard episturmian sequence* s *and we write* $s = \mathrm{Pal}(\Delta(s))$.

Property of the operator Pal

Lemma .7. *Let $x \in \mathcal{A}$. If w is x-free, then*

$$\mathsf{Pal}(wx) = \mathsf{Pal}(w)x\mathsf{Pal}(w).$$

If x occurs in w write $w = w'xw''$ with w'' x-free, then

$$\mathsf{Pal}(wx) = \mathsf{Pal}(w)\mathsf{Pal}(w')^{-1}\mathsf{Pal}(w).$$

Let $w = \mathsf{Pal}(123) = 1213121$.

Then, $\mathsf{Pal}(123 \cdot 4) = \mathsf{Pal}(123) \cdot 4 \cdot \mathsf{Pal}(123) = 121312141213121$ and

$\mathsf{Pal}(1223 \cdot 2) = \mathsf{Pal}(1223) \cdot \mathsf{Pal}^{-1}(w') \cdot \mathsf{Pal}(1223)$

$= 12121312121(121)^{-1}12121312121 = 1212131212121312121,$ with $w = 1223$ and $w' = 12$.

The directive sequence allows to construct easily standard episturmian sequences.

Tribonacci

The Tribonacci sequence T is a standard episturmian sequence with directive sequence

$$\Delta(T) = (123)^{\omega}$$

and then, $u_1 = \varepsilon$, $u_2 = \underline{1}$, $u_3 = (1\underline{2})^{(+)} = 1\underline{2}1$, $u_4 = (121\underline{3})^{(+)} = 121\underline{3}121$, ...,

$$T = \underline{12}1\underline{3}121\underline{1}213121\underline{2}131211213121\ldots$$

Arnoux-Rauzy and episturmian sequences

A standard episturmian sequence $s \in \mathcal{A}^\omega$ is said to be $\mathcal{B}$*-strict* if

$$\mathrm{Ult}(\Delta(s)) = \mathrm{Alph}(s) = \mathcal{B} \subseteq \mathcal{A};$$

that is every letter in $\mathcal{B} = \mathrm{Alph}(s)$ occurs infinitely many times in its directive sequence $\Delta(s)$.

In particular, the $\mathcal{A}$-strict episturmian sequences correspond to the Arnoux-Rauzy sequences.

1) Let s be a standard episturmian sequence with the directive sequence $\Delta(s) = 1232\ldots$. Then,

$$s = \mathsf{Pal}(1232\ldots) = 121312121312 1\cdots,$$

which contains the factors 212 and 131. Thus, s is unbalanced over the letter 2.

2) Let t be a standard episturmian sequence with the directive sequence $\Delta(t) = 12131\ldots$. Then

$$t = \mathsf{Pal}(12131\ldots) = 1211213121121121312 1121\ldots,$$

which contains the factors 11211 and 21312. Thus, t is unbalanced over the letter 1.

3) Let u be a standard episturmian sequence with the directive sequence $\Delta(u) = 12341\ldots$. Then

$$u = \mathsf{Pal}(12341\ldots) = 121312141213121121312141213121\ldots,$$

which is a balanced prefix.

Balance condition depends on where the repeated letters occur.

Proposition .8. *Let* $\Delta(s)$ *be the directive sequence of a balanced standard episturmian sequence* s *over a* k*-letter alphabet* $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geqslant 3$. *Let* k *be the first repeated letter of* $\Delta(s)$. *If* $k \neq s_1$, *then the directive sequence can be written as* $\Delta(s) = 12 \ldots (k-1)k^{\omega}$, *up to letter permutation.*

Proposition .9. *Let $\Delta(s)$ be the directive sequence of a balanced standard episturmian sequence s over a k-letter alphabet, $k \geqslant 3$. If $\Delta(s) = 1^{\ell}z$, with $z \in \mathcal{A}^{\omega}$, $z_1 \neq 1$ and $\ell \geqslant 2$, then $\Delta(s) = 1^{\ell}23\ldots(k-1)k^{\omega}$, up to letter permutation.*

Proof. Let $\Delta(s) = 1^{\ell}z$ be the directive sequence of a balanced standard episturmian sequence s, with $z_1 \neq 1$, $\ell \geqslant 2$.

Assume $|z|_1 > 0$. Then, $\Delta(s) = 1^{\ell}z'1z''$, with $z' \neq \varepsilon$ and $|z'|_1 = 0$.

Since s is over at least a 3-letter alphabet, there exists at least one letter α in z' or z'' distinct from z'_1 and 1. At its first occurrence in s, it is preceded and followed by 1^{ℓ}.

Then,

$$s = \underline{1}^{\ell}\underline{z}'_1 1^{\ell} \ldots 1^{\ell} z'_1 1^{\ell} \underline{1} z'_1 \ldots,$$

which contains the factors $z'_1 1^{\ell+1} z'_1$ and $1^{\ell}\alpha 1^2$, hence $|z|_1 = 0$.

Since the alphabet is finite, there is at least one letter distinct from 1 which occurs twice in z. Let us consider the first repeated one in z, namely γ.

Then, $\Delta(s) = 1^\ell u\gamma v\gamma w$, with $|u\gamma v|_i = 1$ or $0, \forall i \in \mathcal{A}$. Assume $v \neq \varepsilon$ and let $p = \mathsf{Pal}(1^\ell u)$.

Then,

$$s = p\underline{\gamma}p\underline{v}_1 p\gamma p \ldots p\gamma p\underline{\gamma} \ldots$$

which contains the factors $\gamma p\gamma$, pv_1p_1, $v_1 \neq \gamma$ and $p_1 = 1$. It follows that $v = \varepsilon$.

Let us now consider $\Delta(s) = 1^\ell u\gamma^2 w$, which we rewrite as $\Delta(s) = 1^\ell u\gamma^m w'$, with $|u|_i = 1$ or 0 $\forall i \in \mathcal{A}$, $m \geqslant 2$ and assume $w'_1 \neq \gamma$.

Then,

$$s = p(\underline{\gamma}p)^m \underline{w'_1} p_1 \ldots$$

which contains the factors $\gamma p\gamma$ and pw'_1p_1. Hence, $w' = \gamma^\omega$ and the conclusion follows. ■

Theorem .10 (G. Paquin, L. V. (2007)). *Any balanced standard episturmian sequence s over an alphabet with 3 or more letters has a directive sequence, up to a letter permutation, in one of the three following families of sequences:*

a) $\Delta(s) = 1^n 23 \ldots (k-1)(k)^\omega$, *with* $n \geqslant 1$,

b) $\Delta(s) = 12 \ldots (k-1)1k \ldots (k+\ell-1)(k+\ell)^\omega$, *with* $\ell \geqslant 1$,

c) $\Delta(s) = 123 \ldots k(1)^\omega$,

where $k \geqslant 3$.

Recall the following result of Droubay, Justin and Pirillo (2001)

Theorem .11. *A standard episturmian sequence s is ultimately periodic if and only if its directive sequence $\Delta(s)$ has the form $w\alpha^\omega$, $w \in \mathcal{A}^*$, $\alpha \in \mathcal{A}$.*

Corollary .12. *Every balanced standard episturmian sequence on 3 or more letters is ultimately periodic.*

Corollary .13. *None of the Arnoux-Rauzy sequences ($\mathcal{A}$-strict episturmian sequences) are balanced.*

A standard episturmian sequence cannot be both periodic and $\mathcal{A}$-*strict*.

Corollary .14. *Any balanced standard episturmian sequence* s*, over an alphabet with more than 2 letters, is in one of the following families, up to letter permutation:*

a) $s = p(k-1)p\,(kp(k-1)p)^{\omega}$, *with* $p = \mathrm{Pal}(1^n 2 \ldots (k-2))$;

b) $s = p(k+\ell-1)p\,[(k+\ell)p(k+\ell-1)p]^{\omega}$,

with $p = \mathrm{Pal}(123 \ldots (k-1)1k \ldots (k+\ell-2))$;

c) $s = [\mathrm{Pal}(123 \ldots k)]^{\omega}$,

where $k \geqslant 3$.

Proposition .15. *Every balanced standard episturmian sequence* s*, over a* k*-letter alphabet,* $k \geqslant 3$*, with different frequencies for every letter can be written as in Corollary .14 c).*

Fraenkel's conjecture for episturmian sequences

As for every episturmian sequence t one could find a standard episturmian sequence s such that $F(s) = F(t)$, the result of Proposition .15 can be extended to any balanced episturmian sequence.

Theorem .16. *Let* s *be a balanced episturmian sequence over a* k*-letter alphabet* $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geqslant 3$, *with all distinct frequencies of letters. Then,* $s = [\mathrm{Pal}(123 \ldots k)]^{\omega}$ *up to letter permutation.*

For $k = 3, 4$, we obtain respectively $s = (1213121)^{\omega}$ and $t = (121312141213121)^{\omega}$.

Concluding remarks

Billiard sequences over a k-letter alphabet are $(k-1)$-balanced.

Furthermore, we notice that Fraenkel's sequences are periodic billiard sequences.

It would be interesting to study the billiard sequences which are balanced.

This class contains at least Fraenkel's sequences, and perhaps some other interesting sequences.

The second direction is directly related to the original form of Fraenkel's conjecture.

To prove this conjecture, it will be useful to have the property that balanced sequences over an alphabet with more than 2 letters and with pairwise distinct frequencies of letters, are given by directive sequences.

Combining our results with this conjecture would give a proof of Fraenkel's conjecture.

Open problem: Prove that balanced sequences over an alphabet with more than 2 letters and with pairwise distinct frequencies of letters are of the form

$$(P)^\omega \text{ with } P \text{ is a palindrome}$$