WORDS 2013 — Turku, Finland — September 16th, 2013

Which Arnoux-Rauzy words are 2-balanced?

Julien Cassaigne

Institut de mathématiques de Luminy - CNRS, Marseille, France

cassaigne@iml.univ-mrs.fr

Which Arnoux-Rauzy words are 2-balanced?

- Sturmian words
- Arnoux-Rauzy words and their directive sequence
- A balanced Arnoux-Rauzy word: Tribonacci
- An unbalanced Arnoux-Rauzy word
- A sufficient condition for 2-balance
- A necessary condition for 2-balance
- 2-balance cannot be expressed with a rational language

Sturmian words

Let $\alpha \in [0,1] \setminus \mathbb{Q}$, $\rho \in \mathbb{R}$, and define $w_n = \lfloor \alpha(n+1) + \rho \rfloor - \lfloor \alpha n + \rho \rfloor$.

$w = (w_n)_{n \in \mathbb{N}}$ is an infinite word on $A = \{0, 1\}$:
a Sturmian word of slope α.

Its language of factors $L(w)$ depends only on α.

Example. Fibonacci word
$\alpha = \rho = \frac{3-\sqrt{5}}{2}$
$w = 0100101001001010010100\ldots$.

Sturmian words and substitutions

Let σ_0 and σ_1 be substitutions defined by:
$\sigma_0(0) = 0$, $\sigma_0(1) = 01$, $\sigma_1(0) = 10$, $\sigma_1(1) = 1$.

Then any Sturmian word w can be written as $w = S^\varepsilon(\sigma_i(w'))$,
where S is the shift, $\varepsilon \in \{0, 1\}$, $i \in A$, and $w' \in A^\mathbb{N}$.
Moreover w' is again Sturmian, so the process can be repeated.

We then obtain a sequence $(i_n) \in A^\mathbb{N}$, the directive sequence of w,
and Sturmian words $w^{(n)}$ obtained from w by de-substitution:
$w = S^k\Big((\sigma_{i_0} \circ \sigma_{i_1} \circ \cdots \circ \sigma_{i_{n-1}})(w^{(n)})\Big)$ for some $k \in \mathbb{N}$.

Directive sequence and continued fraction

Conversely: let $(i_n) \in A^{\mathbb{N}}$, not eventually constant.
Let $u_n = (\sigma_{i_0} \circ \sigma_{i_1} \circ \cdots \circ \sigma_{i_{n-1}})(i_n)$.
Then the u_n are prefixes of increasing length of a Sturmian word w.

The associated slope is given by the continued fraction
$\frac{1}{\alpha} - 1 = [a_0; a_1, a_2, \ldots]$ where $(i_n) = 0^{a_0} 1^{a_1} 0^{a_2} 1^{a_3} \ldots$.

Example. If $\alpha = \frac{5-\sqrt{10}}{5}$, then $1/\alpha - 1 = [1; 1, 2, 1, 1, 2, 1, 1, 2, \ldots]$, and the directive sequence is $010010110100\ldots$, so the associated Sturmian word is the fixed point of $\sigma_0 \circ \sigma_1 \circ \sigma_0^2 \circ \sigma_1 \circ \sigma_0 \circ \sigma_1^2$.

Balance

In a Sturmian word of slope α, we expect a factor of length n to contain approximately αn occurrences of 1.

Indeed, this number is either $\lfloor \alpha n \rfloor$ or $\lceil \alpha n \rceil$. If u and v are two factors of length n, then $\Big| |u|_1 - |v|_1 \Big|$ is at most 1. A word with this property is 1-balanced.

Theorem. [Morse Hedlund 1940]
An infinite word is Sturmian if and only if it is both 1-balanced and aperiodic.

Arnoux-Rauzy words

How to generalize Sturmian words to a larger alphabet? There is no unique answer. Arnoux-Rauzy words are a possible generalization [Arnoux Rauzy 1991].

Let $A = \{1, 2, 3\}$ (generalization to arbitrary alphabets is immediate). Define substitutions σ_i by $\sigma_i(i) = i$ and $\sigma_i(j) = ij$ if $j \neq i$.

Let $(i_n) \in A^{\mathbb{N}}$ be a sequence where each symbol occurs infinitely often.
Let $u_n = (\sigma_{i_0} \circ \sigma_{i_1} \circ \cdots \circ \sigma_{i_{n-1}})(i_n)$.
Then the u_n are prefixes of increasing length of an infinite word w. Any word w' such that $L(w') = L(w)$ is an Arnoux-Rauzy word (AR-word), or ternary strict episturmian word, of directive sequence (i_n).

Tribonacci word

The Tribonacci word is the standard AR-word t of directive sequence $123123123\ldots$. It is a fixed point of $\sigma_1 \circ \sigma_2 \circ \sigma_3$, or more simply of τ: $\tau(1) = 12$, $\tau(2) = 13$, $\tau(3) = 1$.

Let $P_n = \chi(t_0 \ldots t_{n-1})$, where $\chi(u) = (|u|_1, |u|_2, |u|_3)$.
The points P_n form a broken line in $\mathbb{R}^3$. Since τ is a Pisot substitution, their distance to the straight line $\mathbb{R}V$ is bounded, where V is the Pisot eigenvector of τ.

It follows that t is C-balanced for some C: if u and v are two factors of the same length of t, then $\Big||u|_i - |v|_i\Big|$ is bounded by C for every $i \in A$.

Actually, t is 2-balanced [Richomme Saari Zamboni 2009] (proof: later).

Equivalent definition of balance

Let $E_w = \{\chi(u) - \chi(v) \,:\, u, v \in L(w)\}$
(u and v may have different lengths).
Let $C_w = \mathsf{sup}\{\mathsf{max}(|x|, |y|, |z|) \,:\, (x, y, z) \in E_w \text{ and } x + y + z = 0\}$.
Then w is C-balanced if and only if $C \geq C_w$.

Proposition. If $w = \sigma_i(w')$, then E_w can be deduced from $E_{w'}$ by:
$E_w = \{(x+y+z+\delta, y, z) \,:\, (x, y, z) \in E_{w'}, \delta \in \{-2, -1, 0, 1, 2\}\}$ if $i = 1$;
$E_w = \{(x, x+y+z+\delta, z) \,:\, (x, y, z) \in E_{w'}, \delta \in \{-2, -1, 0, 1, 2\}\}$ if $i = 2$;
$E_w = \{(x, y, x+y+z+\delta) \,:\, (x, y, z) \in E_{w'}, \delta \in \{-2, -1, 0, 1, 2\}\}$ if $i = 3$.

Action of a substitution on balance

Proposition. If $w = \sigma_i(w')$, then E_w can be deduced from $E_{w'}$ by:
$E_w = \{(x+y+z+\delta, y, z) : (x,y,z) \in E_{w'}, \delta \in \{-2,-1,0,1,2\}\}$ if $i = 1$;
$E_w = \{(x, x+y+z+\delta, z) : (x,y,z) \in E_{w'}, \delta \in \{-2,-1,0,1,2\}\}$ if $i = 2$;
$E_w = \{(x, y, x+y+z+\delta) : (x,y,z) \in E_{w'}, \delta \in \{-2,-1,0,1,2\}\}$ if $i = 3$.

Example.

2122123	32123	$(1,2,-1)$	
$\downarrow \sigma_1$	$\downarrow \sigma_1$		
121121211213	131211213	$(2,2,-1)$	$\delta = 0$
121121211213	31211213	$(3,2,-1)$	$\delta = 1$
1211212112131	31211213	$(4,2,-1)$	$\delta = 2$
121121211213	1312112131	$(1,2,-1)$	$\delta = -1$
21121211213	1312112131	$(0,2,-1)$	$\delta = -2$

An unbalanced Arnoux-Rauzy word

It was for some time expected that AR-words would all be C-balanced for some C.

Theorem. [Cassaigne Ferenczi Zamboni 2000]
There exists an unbalanced AR-word: an AR-word w with $C_w = \infty$.

Actually, a large class of AR-words generate weak-mixing subshifts, and are therefore unbalanced [Cassaigne Ferenczi Messaoudi 2008].

Construction

Lemma 1. [Cassaigne Nardi 2013]
If $(k+1, -k, k-1) \in E_{w'}$, and $w = (\sigma_1^{2k+5} \circ \sigma_2^{2k+3} \circ \sigma_3^{2k+1})(w')$, then $(k+4, -k-3, k+2) \in E_w$.

Lemma 2. For any K there exists an AR-substitution μ_K such that, for any AR-word w, $C_{\mu_K(w)} \geq K$.

Construction

Lemma 3. $C_{\sigma_i(w)} \geq \frac{1}{2} C_w$.

Now construct inductively (K_n) such that, for any AR-word w, $C_{(\mu_{K_1} \circ \cdots \circ \mu_{K_n})(w)} \geq n$.

A sufficient condition for 2-balance

Theorem. [Berthé Cassaigne Steiner 2012]
Let $X = \{iiji, iijj, ijiji, ijijj \ : \ i, j \in A$ and $i \neq j$. If the directive sequence of an AR-word w contains no factor in X, then w is 2-balanced.

Corollary. Tribonacci word is 2-balanced, as its directive sequence is $123123\ldots$

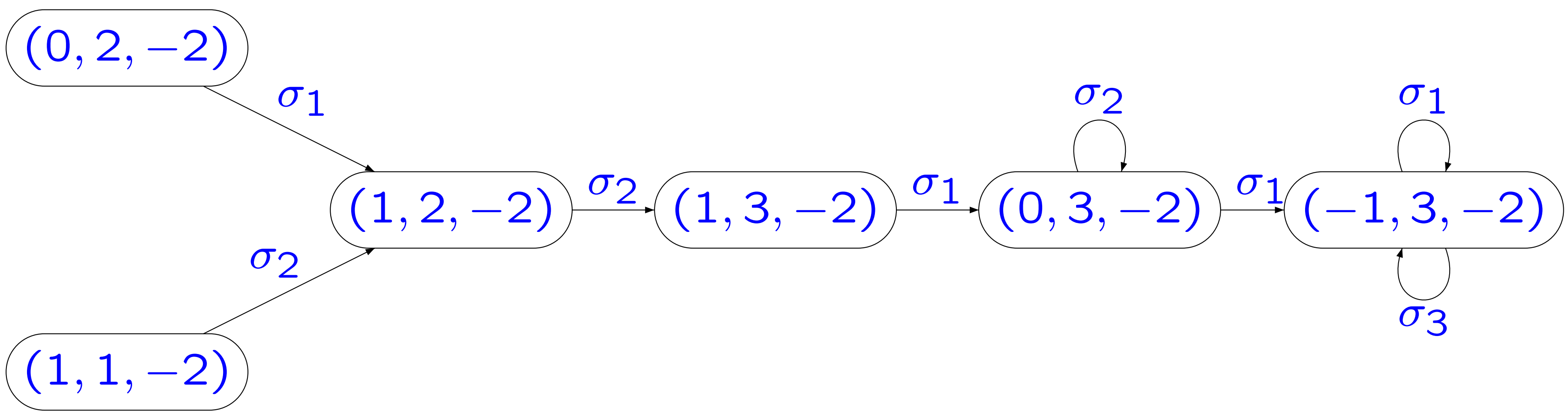
Actually, an AR-word satisfying this condition is strongly 2-balanced: all de-substituted words are also 2-balanced.

A necessary condition for 2-balance

Theorem. [Berthé Cassaigne Steiner 2012]
If the directive sequence of an AR-word w starts with a word in $Y = \{1,3\}^*12^*12\{1,2\}$, then $C_w \geq 3$.

Proof

Both $(0, 2, -2)$ and $(1, 1, -2)$ occur in every E_w.



2-balance and rational languages

Is there any hope to replace Y with a larger rational language, so that the previous theorem becomes an equivalence?

Equivalently: let P be the set of prefixes of directive sequences of 2-balanced AR-words. Is P rational?

Theorem. [Cassaigne Nardi 2013]
P is not a rational language.

2-balance and rational languages

Theorem. [Cassaigne Nardi 2013]
P is not a rational language.

Proof. Assume that P is recognized by a DFA with $n-1$ states. Let w the AR-word directed by $(123)^l(1^n2^n3^n)^\omega$. If l is chosen large enough, then w is 2-balanced, and thus its prefixes are in P.

By a pumping argument, we construct $N > 2^{3l+3}$ such that all prefixes of $(123)^l(1^N2^N3^N)^\omega$ are in P; but the AR-word with this directive sequence is not 2-balanced.

Perspectives

Since rational languages are not sufficient, we should try other classes, e.g. context-free languages.

We have characterized strong 2-balanced ternary AR-words. What about strong C-balanced generalized AR-words on a d-letter alphabet?

Does controling balance help with multidimensional continued fraction algorithms?