Balanced words with factor frequencies

V. Berthé and J. M. Thuswaldner

Université de Paris, France
and
University of Leoben, Austria

Expanding Dynamics Seminar (via zoom), October 13, 2020

Balance on factors

- Let $\mathcal{A} = \{1, \dots, d\}$ be a finite alphabet.
- Let $u \in \mathcal{A}^{\mathbb{N}}$ be a sequence over $\mathcal{A}$.
- The language $L(u) \subset \mathcal{A}^*$ of u is the set of subwords of u.
- Let $v = v_0 \dots v_{n-1}$ and $w = w_0 \dots w_{k-1}$ elements of $\mathcal{A}^*$ (words). Then

  $$|v|_w := \#\{i : v_i \dots v_{i+k-1} = w_0 \dots w_{k-1}\}$$

  (number of occurrences of w in v).
- For $v_0 \dots v_{n-1} \in \mathcal{A}^*$ we set $|v| = n$ (length of v)

Factor frequencies

Let $u = u_0 u_1 \ldots \in \mathcal{A}^{\mathbb{N}}$ be given.

- We say that the factor $w \in L(u)$ has a frequency in u if

$$f_w := \lim_{n \to \infty} \frac{|u_0 \ldots u_{n-1}|_w}{n}$$

  exists.

- We say that u has factor frequencies if each $w \in L(u)$ has a frequency in u.

Balance and symbolic discrepancy

Let $u \in \mathcal{A}^{\mathbb{N}}$ be given.

- u is finitely balanced on factors if for each $w \in L(u)$ there exists $C_w \in \mathbb{N}$ such that for all $v, v' \in L(u)$ with $|v| = |v'|$ we have
$$\big||v|_w - |v'|_w\big| \le C_w.$$

- u has finite symbolic discrepancy if for each $w \in L(u)$ there exists $C'_w \in \mathbb{N}$ such that each $v \in L(u)$ has a frequency in u and
$$\big||v|_w - f_w|v|\big| \le C'_w$$
holds.

The problem

Problem

Let $f = (f_1, \dots f_d) \in \mathbb{R}_+^d$ with $\|f\|_1 = 1$ be given. Find a sequence $u \in \mathcal{A}^{\mathbb{N}}$ such that

- The letter frequency of a in u is f_a for each $a \in \mathcal{A}$.
- u has factor frequencies.
- u is finitely balanced on factors.

How small can the balance constant be? To what extent can we prescribe the factor frequencies? Necessary:

$$f_w = \sum_{a \in \mathcal{A}} f_{aw} = \sum_{a \in \mathcal{A}} f_{wa} \qquad \text{(for each } w \in \mathcal{A}^*\text{)}.$$

Consequences

- The shift $(\mathcal{O}(u), \Sigma)$ is uniquely ergodic.
- u has finite symbolic discrepancy.

What is known so far

- The Sturmian case ($d = 2$).
- Using symbolic words that are constructed in terms of generalized continued fraction algorithms we can do the cases $d \in \{3, 4\}$ for a.a. frequency vectors $(f_1, \dots, f_d)$.
- The main reason why we cannot go to bigger d is the fact that it is not known if the classical continued fraction algorithms we use are strongly convergent in high dimensions.
- Exception: Sequences related to the Arnoux–Rauzy continued fraction algorithm can be defined for all dimensions, but not for a.a. frequency vectors (the "allowed" vectors are related to the Rauzy gasket).
- Note that billiard words are not balanced on factors.

Factor complexity

Let $u \in \mathcal{A}^{\mathbb{N}}$ be given.

- Let
$$p_u : \mathbb{N} \to \mathbb{N}, \quad n \mapsto \#\{w \in L(u) \ : \ |w| = n\}$$
be the factor complexity function of u.

- We are interested to choose u in a way that it has low (i.e. linear) factor complexity.

- This is true for Sturmian sequences with $p_u(n) = n + 1$.

- It is also true for sequences defined in terms of the Cassaigne–Selmer continued fraction algorithm that was designed in a way that $p_u(n) = 2n + 1$.

- What is the smallest factor complexity we can get?