Factor-balanced S-adic languages

Léo Poirier and Wolfgang Steiner

arXiv:2211.14076

ENS Lyon IRIF, CNRS, Université Paris Cité

One World Combinatorics on Words Seminar, March 6, 2023

Balancedness

language $\mathcal{L}$ over a finite alphabet A is

- *C-balanced w.r.t. $v \in A^*$*:

$$|w|_v - |w'|_v \leq C \qquad \text{for all } w, w' \in \mathcal{L} \text{ with } |w| = |w'|$$

($|w|_v$ denotes the number of occurrences of v in w, $|w|$ the length of w)

- *balanced w.r.t. $v \in A^*$*: C-balanced w.r.t. $v \in A^*$ for some $C \geq 0$
- *(C-)balanced for length n*: (C-)balanced for all $v \in A^n$
- *letter-(C-)balanced*: (C-)balanced for length 1
- *factor-(C-)balanced*: (C-)balanced for all lengths $n \geq 1$

factor-balanced $\iff$ balanced w.r.t. all $v \in A^*$

$\not\Longleftrightarrow$ factor-C-balanced for some $C \geq 1$

Examples of (factor-)balanced languages

Morse–Hedlund '40:
language of a Sturmian word is letter-1-balanced

Fagnot–Vuillon '02:
language of a Sturmian word is factor-balanced, more precisely $|v|$-balanced w.r.t. all v, factor-C-balanced for some $C \geq 1$ if (and only if) the slope has bounded partial quotients

language of the Thue-Morse word is letter-2-balanced
(if $|w| = 2n + 1$, then $(|w|_0, |w|_1) \in \{(n, n{+}1), (n{+}1, n)\}$;
if $|w| = 2n$, then $(|w|_0, |w|_1) \in \{(n, n), (n{-}1, n{+}1), (n{+}1, n{-}1)\}$)

Berthé–Cecchi Bernales '19:
language of the Thue-Morse word is NOT balanced for length 2

Berthé–Cecchi Bernales–Durand–Leroy–Perrin–Petite '21:
For S-adic languages defined by sequences of (left or right) proper unimodular morphisms, factor-balancedness is equivalent to letter-balancedness.

Frequency vector, factorial languages

Proposition 1 (cf. Berthé–Tijdeman '02, Adamczewski '03)

If there is a (frequency) vector $(f_a)_{a \in A}$ *such that* $||w|_a - f_a|w|| \leq C$ *for all* $a \in A$, $w \in \mathcal{L}$, *then* $\mathcal{L}$ *is letter-*$(2C)$*-balanced.*

If $\mathcal{L}$ *is an infinite letter-*C*-balanced factorial language, then there exists* $(f_a)_{a \in A}$ *such that* $||w|_a - f_a|w|| \leq C$ *for all* $a \in A$, $w \in \mathcal{L}$.

($\mathcal{L}$ is *factorial* if $\mathcal{F}(\mathcal{L}) = \mathcal{L}$, where $\mathcal{F}(\mathcal{L})$ denotes the set of factors of words in $\mathcal{L}$)

n-coding

the *n-coding* of a word $a_1 a_2 \cdots a_N \in A^N$ is the word over the alphabet A^n defined by

$$(a_1 a_2 \cdots a_N)^{(n)} = (a_1 \cdots a_n)(a_2 \cdots a_{n+1}) \cdots (a_{N-n+1} \cdots a_N) \in (A^n)^{N-n+1},$$

$(a_1 a_2 \cdots a_N)^{(n)}$ is the empty word if $n > N$

$|w|_v = |w^{(n)}|_v$ if $v \in A^n$

$\mathcal{L}$ balanced for length n $\Leftrightarrow$ $\{w^{(n)} : w \in \mathcal{L}\}$ letter-balanced (alphabet A^n)

Morphisms preserve letter-balancedness

σ *morphism* (or substitution): $\sigma(vw) = \sigma(v)\sigma(w)$

Proposition 2

If $\mathcal{L} \subset A^*$ is a letter-balanced factorial language and $\sigma : A^* \to B^*$ a morphism, then $\mathcal{F}(\sigma(\mathcal{L}))$ is letter-balanced.

Proof: If $\mathcal{L}$ is letter-C-balanced, then for $w, w' \in \mathcal{F}(\sigma(\mathcal{L}))$, $|w| = |w'|$,

$$w = x\,\sigma(y)\,z, \quad w' = x'\,\sigma(y')\,z',$$

with $y, y' \in \mathcal{L}$, $|y| = |y'|$, $|x\,z|, |x'z'| \le (2+C\#A)\max_{a\in A}|\sigma(a)|$, and

$$|\sigma(y)|_b - |\sigma(y')|_b \le \sum_{a\in A} \big||y|_a - |y'|_a\big|\,|\sigma(a)|_b \le C\#A \max_{a\in A}|\sigma(a)|.$$

Proposition 3

If $\mathcal{L} \subset A^*$ is a factorial language, $\sigma : A^* \to B^*$ a morphism with invertible incidence matrix $\mathcal{M}_\sigma$ and $\mathcal{F}(\sigma(\mathcal{L}))$ is letter-balanced, then $\mathcal{L}$ is letter-balanced.

(*incidence matrix* $\mathcal{M}_\sigma = (|\sigma(b)|_a)_{a\in A, b\in B}$)

Proof: Use Proposition 1.

Non-erasing morphisms preserve balancedness for length n, proper morphisms increase length for balancedness

Proposition 4

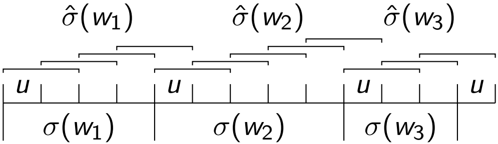
If $\mathcal{L} \subset A^*$ is a factorial language that is balanced for length n, $\sigma : A^* \to B^*$ a morphism, $u \in B^*$ a (possibly empty) word that is a prefix of $\sigma(a)u$ for all $a \in A$ (or a suffix of $u\sigma(a)$ for all $a \in A$), then $\mathcal{F}(\sigma(\mathcal{L}))$ is balanced for length $\min_{w \in A^{n-1} \cap \mathcal{L}} |\sigma(w)| + |u| + 1$. In particular,

- *$\mathcal{F}(\sigma(\mathcal{L}))$ is balanced for length n if σ is non-erasing,*
- *$\mathcal{F}(\sigma(\mathcal{L}))$ is balanced for length $n+1$ if σ is left or right proper.*

(σ is *left* (resp. *right*) *proper* if $\sigma(a)$ starts (resp. ends) with the same letter for all $a \in A$)

Proof: case $n = 1$, $\sigma(a)$ starts with $u \in B$ for all $a \in A$:

$$\text{morphism } \hat{\sigma} : A^* \to (B^2)^*,\ w \mapsto (\sigma(w)u)^{(2)}$$



If $\mathcal{L}$ is letter-balanced, then $\mathcal{F}(\hat{\sigma}(\mathcal{L}))$ is letter-balanced, thus $\mathcal{F}(\sigma(\mathcal{L}))$ is balanced for length 2.

In general, for $m \le \min_{w \in A^{n-1} \cap \mathcal{L}} |\sigma(w)| + |u| + 1$, we can define a morphism $\hat{\sigma} : (A^n \cap \mathcal{L})^* \to (B^m)^*$ mapping n-codings of $\mathcal{L}$ to m-codings of $\mathcal{F}(\sigma(\mathcal{L}))$ by

$$\hat{\sigma}(a_1 a_2 \cdots a_n) = \big(\sigma(a_1)\mathrm{pref}_{m-1}(\sigma(a_2 \cdots a_n)u)\big)^{(m)}.$$

Then $\big(\sigma(w)u\big)^{(m)} = \hat{\sigma}\big(w^{(n)}\big)\, \big(\sigma(\mathrm{suff}_{n-1}(w))u\big)^{(m)}$ for all $w \in \mathcal{L}$. Since $\mathcal{L}$ is balanced for length n, the set $\mathcal{L}^{(n)}$ of n-codings of words in $\mathcal{L}$ is letter-balanced, $\mathcal{F}(\hat{\sigma}(\mathcal{L}^{(n)}))$ is letter-balanced, thus $\mathcal{F}(\sigma(\mathcal{L}))$ is balanced for length m.

Morphisms preserve factor-balancedness

Theorem 1
If $\mathcal{L} \subset A^*$ is a factor-balanced factorial language and $\sigma : A^* \to B^*$ a morphism, then $\mathcal{F}(\sigma(\mathcal{L}))$ is factor-balanced.

For non-erasing morphisms, this is a corollary of Proposition 4.
For morphisms with erasing letters, use Propositions 1 and 4.

S-adic languages

sequence of morphisms $\boldsymbol{\sigma} = (\sigma_k)_{k\geq 0}$, $\sigma_k : A_{k+1}^* \to A_k^*$,

$$\sigma_{[k,n)} = \sigma_k \circ \sigma_{k+1} \circ \cdots \circ \sigma_{n-1}$$

language of $\boldsymbol{\sigma}$:

$$\mathcal{L}_{\boldsymbol{\sigma}} = \left\{ w \in A_0^* \,:\, w \in \mathcal{F}(\sigma_{[0,n)}(A_n)) \text{ for infinitely many } n \right\}$$

language of $\boldsymbol{\sigma}$ at level k:

$$\begin{aligned} \mathcal{L}_{\boldsymbol{\sigma}}^{(k)} &= \left\{ w \in A_k^* \,:\, w \in \mathcal{F}(\sigma_{[k,n)}(A_n)) \text{ for infinitely many } n \right\} \\ &= \mathcal{F}\big(\sigma_{[k,n)}(\mathcal{L}_{\boldsymbol{\sigma}}^{(n)})\big) \quad \text{for all } n \geq k \end{aligned}$$

$(\sigma_k)_{k\geq 0}$ is left (resp. right) *proper* if $\forall k \geq 0\ \exists n > k$ such that $\sigma_{[k,n)}$ is left (resp. right) proper

$(\sigma_k)_{k\geq 0}$ is *everywhere growing* if $\lim_{k\to\infty} \min_{a\in A_k} |\sigma_{[0,k)}(a)| = \infty$

Factor-balanced S-adic and substitutive languages

Theorem 2

- *If $\boldsymbol{\sigma}$ is a left or right proper sequence of morphisms and $\mathcal{L}_{\boldsymbol{\sigma}}^{(k)}$ is letter-balanced for infinitely many k, then $\mathcal{L}_{\boldsymbol{\sigma}}$ is factor-balanced.*
- *If $\boldsymbol{\sigma} = (\sigma_k)_{k\geq 0}$ is left or right proper, $\mathcal{L}_{\boldsymbol{\sigma}}$ is letter-balanced and the incidence matrices $\mathcal{M}_{\sigma_k}$ are invertible, then $\mathcal{L}_{\boldsymbol{\sigma}}$ is factor-balanced.*
- *If $\boldsymbol{\sigma}$ is everywhere growing and $\mathcal{L}_{\boldsymbol{\sigma}}^{(k)}$ is balanced for length 2 for infinitely many k, then $\mathcal{L}_{\boldsymbol{\sigma}}$ is factor-balanced.*
- *If $\sigma : A^* \to A^*$ is a morphism such that σ^k is left or right proper for some $k \geq 1$ and $\mathcal{L}_\sigma$ is letter-balanced, then $\mathcal{L}_\sigma$ is factor-balanced.*
- *If $\sigma : A^* \to A^*$ is an everywhere growing morphism and $\mathcal{L}_\sigma$ is balanced for length 2, then $\mathcal{L}_\sigma$ is factor-balanced. (cf. Queffélec '87, Adamczewski '03, '04)*

(for $\sigma : A^* \to A^*$, $\mathcal{L}_\sigma = \{w \in A^* : w \in \mathcal{F}(\sigma^n(A)) \text{ for infinitely many } n\}$)

Thue–Morse–Sturmian languages

$$L: 0 \mapsto 0, \quad 1 \mapsto 10, \qquad M: 0 \mapsto 01, \quad 1 \mapsto 10, \qquad R: 0 \mapsto 01, \quad 1 \mapsto 1,$$

$\mathcal{L}_M$ is the language of the Thue–Morse word,
for $\boldsymbol{\sigma} \in \{L, R\}^\infty$ not ending with L^∞ or R^∞, $\mathcal{L}_{\boldsymbol{\sigma}}$ is Sturmian

Proposition 5

For all $\boldsymbol{\sigma} \in \{L, M, R\}^\infty$, $\mathcal{L}_{\boldsymbol{\sigma}}$ is letter-2-balanced.

Theorem 3

For $\boldsymbol{\sigma} = (\sigma_k)_{k \geq 0} \in \{L, M, R\}^\infty$, $\mathcal{L}_{\boldsymbol{\sigma}}$ is factor-balanced if and only if $\sigma_k \neq M$ for infinitely many k.

$\mathcal{F}(\sigma \circ L(\mathcal{L}_M))$ is balanced for length $|\sigma(0)|+1$ for any morphism σ
but $\mathcal{F}(\sigma \circ L(\mathcal{L}_M))$ is not balanced w.r.t. $\sigma(01010)$ for $\sigma \in \{L, M, R\}^*$.

Theorem 3

For $\boldsymbol{\sigma} = (\sigma_k)_{k\geq 0} \in \{L, M, R\}^\infty$, $\mathcal{L}_{\boldsymbol{\sigma}}$ *is factor-balanced if and only if* $\sigma_k \neq M$ *for infinitely many* k.

Proof: If $\sigma_k \in \{L, R\}$ for infinitely many k, then $\boldsymbol{\sigma}$ is right proper, hence $\mathcal{L}_{\boldsymbol{\sigma}}$ is factor-balanced by Proposition 5 and Theorem 2.

Otherwise, $\mathcal{L}_{\boldsymbol{\sigma}} = \mathcal{F}(\sigma(\mathcal{L}_M))$ for some $\sigma \in \{L, M, R\}^*$.
By Berthé–Cecchi Bernales '19, $\mathcal{L}_M$ is not balanced w.r.t. 11 (because its frequency is $\frac{1}{6}$ and $6 \nmid 2^n = |M^n(1)|$ for all n).
Also, u_n, u'_n defined by

$$u_1 = 00, \quad M^2(u_{2n-1}) = 0\, u_{2n}\, 0, \quad M^2(u_{2n}) = 1\, u_{2n+1}\, 1,$$
$$u'_1 = 01, \quad M^2(u'_{2n-1}) = u'_{2n}\, 01, \quad M^2(u'_{2n}) = u'_{2n+1}\, 10,$$

are in $\mathcal{L}_M$ for all n, and

$$|u_{2n}|_{00} - |u'_{2n}|_{00} = |u'_{2n}|_{01} - |u_{2n}|_{01} = |u'_{2n}|_{10} - |u_{2n}|_{10} = |u_{2n}|_{11} - |u'_{2n}|_{11} = n.$$

For all $\sigma \in \{L, M, R\}^*$, $w \in \mathcal{L}_M$, we have $|\sigma(w)|_{\sigma(011)} = |w|_{011}$, thus

$$|\sigma(u_{2n})|_{\sigma(011)} - |\sigma(u'_{2n})|_{\sigma(011)} = |u_{2n}|_{011} - |u'_{2n}|_{011} = |u_{2n}|_{11} - 1 - |u'_{2n}|_{11} = n - 1.$$

Therefore, $\boldsymbol{\sigma}$ is not factor-balanced if $\boldsymbol{\sigma}$ ends with M^∞.

Letter-balancedness of (primitive) S-adic languages

Necessary conditions:

$$\bigcap_{n \geq 0} M_{\sigma_{[0,n)}} \mathbb{R}_+^{\# A_n} = \mathbb{R}_+ \, (f_a)_{a \in A_0}$$

$$\left\{ |\sigma_{[0,n)}(b)|_a - f_a \, |\sigma_{[0,n)}(b)| \, : \, a \in A_0, b \in A_n, n \geq 0 \right\} \quad \text{bounded}$$

not always sufficient: see Cassaigne–Ferenczi–Zamboni '00, Cassaigne–Ferenczi–Messaoudi '08 and Berthé–Cassaigne–Steiner '13 for Arnoux–Rauzy words, Delecroix–Hejda–Steiner '13 for Brun words, Andrieu '18 for Cassaigne words