Perfectly clustering words: Induction and morphisms

Mélodie Lapointe

Département mathématiques et statistique
Université de Moncton

8 mai 2023
One World Combinatorics on Words Seminar

Outline

We studied a family of words called

perfectly clustering words.

In this talk, we want to show

- morphisms sending perfectly clustering words to another perfectly clustering words
- an induction on discrete interval exchange transformation
- a relation between perfectly clustering words and band bricks over certain algebras

Perfectly clustering words

Perfectly clustering words

The Burrows-Wheeler transform is a function on the word defined as follows

1. takes all the conjugates of a word
2. sort them in lexicographic order
3. return a word which is the concatenation of the last letter of each conjugates

Perfectly clustering words

The Burrows-Wheeler transform is a function on the word defined as follows

1. takes all the conjugates of a word
2. sort them in lexicographic order
3. return a word which is the concatenation of the last letter of each conjugates

Example :

a n a n a s

Perfectly clustering words

The Burrows-Wheeler transform is a function on the word defined as follows

1. takes all the conjugates of a word
2. sort them in lexicographic order
3. return a word which is the concatenation of the last letter of each conjugates

Example :

```
a n a n a s
  n a n a s a
    a n a s a n
      n a s a n a
        a s a n a n
          s a n a n a
```

Perfectly clustering words

The Burrows-Wheeler transform is a function on the word defined as follows

1. takes all the conjugates of a word
2. sort them in lexicographic order
3. return a word which is the concatenation of the last letter of each conjugates

Example :

a n a n a s	a n a n a s
n a n a s a	a n a s a n
a n a s a n	a s a n a n
n a s a n a	n a n a s a
a s a n a n	n a s a n a
s a n a n a	s a n a n a

Perfectly clustering words

The Burrows-Wheeler transform is a function on the word defined as follows

1. takes all the conjugates of a word
2. sort them in lexicographic order
3. return a word which is the concatenation of the last letter of each conjugates

Example :

a n a n a s	a n a n a s
n a n a s a	a n a s a n
a n a s a n	a s a n a n
n a s a n a	n a n a s a
a s a n a n	n a s a n a
s a n a n a	s a n a n a

$$\mathrm{BWT}(\mathrm{ananas}) = \mathrm{snnaaa}$$

Perfectly clustering word

- $|w|_a$ denote the number of occurrences of the letter a in w.
- A word w is π-clustering if

$$\mathrm{TBW}(w) = a_{\pi(1)}^{|w|_{a_{\pi(1)}}} a_{\pi(2)}^{|w|_{a_{\pi(2)}}} \dots a_{\pi(r)}^{|w|_{a_{\pi(r)}}}$$

  and $\pi \neq id$.
- A word w is perfectly clustering if

$$\mathrm{TBW}(w) = a_r^{|w|_{a_r}} a_{r-1}^{|w|_{a_{r-1}}} \dots a_1^{|w|_{a_1}}$$

Example :

Words	appartement	aluminium	ananas
BWT	tptmeepaanr	mmnauuiil	snnaaa

Also call words with simple Burrows-Wheeler transform.

Why study perfectly clustering words?

- On binary alphabet, they are Christoffel words.

Theorem (Mantaci, Restivo et Sciortino, 2003)

A binary word w is perfectly clustering if and only if w is a conjugate of a Christoffel word.

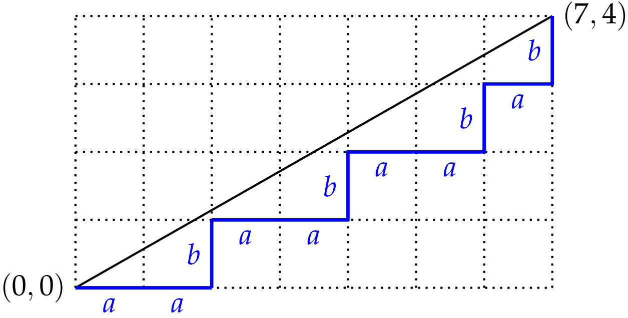
- They are acting as interval exchange transformation.

Theorem (Ferenczi and Zamboni, 2013)

A word w is a perfectly clustering word if and only if the mapping from the last column to the first column is a minimal symmetric discrete interval exchange transformation.

Christoffel words

Examples : The Christoffel words of slope $7/4$



Also known as Standard words, central words or periodic mechanical words.

Generalization of Christoffel words

- finite episturmian words : Factor of an infinite episturmian word.
- A infinite episturmian word $w \in \mathcal{A}^{\omega}$ if $Fact(t)$ closed under reversal and at most one left special factor of each length.
- Episturmian words $\neq$ Perfectly clustering words
- Episturmian words $\cap$ Perfectly clustering words $\neq \emptyset$.

(Restivo, Rosone 2009)

Using morphisms to construct perfectly clustering words

Main goals

Recall : A morphism is a map ϕ between $\mathcal{A}^*$ and $\mathcal{A}^*$ such that for all $u, v \in M$,

$$\phi(uv) = \phi(u)\phi(v).$$

Theorem (Berstel and de Luca, 1997)

A word w is a Christoffel word if and only if there exists a sequence of morphisms

$$\chi = \chi_1 \circ \cdots \circ \chi_n$$

where $\chi_i \in \{G = (a, ab), \widetilde{D} = (ab, b)\}$ such that

$$\chi_1 \circ \cdots \circ \chi_n(ab) = w.$$

Main goals

Recall : A morphism is a map ϕ between $\mathcal{A}^*$ and $\mathcal{A}^*$ such that for all $u, v \in M$,

$$\phi(uv) = \phi(u)\phi(v).$$

Theorem (Berstel and de Luca, 1997)

A word w is a Christoffel word if and only if there exists a sequence of morphisms

$$\chi = \chi_1 \circ \cdots \circ \chi_n$$

where $\chi_i \in \{G = (a, ab), \widetilde{D} = (ab, b)\}$ such that

$$\chi_1 \circ \cdots \circ \chi_n(ab) = w.$$

Can we describe perfectly clustering words using morphisms ?

Main goals

Theorem (Simpson and Puglisi, 2008)

A word $w \in \{a, b, c\}^*$ is perfectly clustering if and only if there exists a sequence of functions

$$\chi = \chi_1 \circ \cdots \circ \chi_n$$

where $\chi_i \in \{\phi, \theta, \psi\}$ such that

$$\chi_1 \circ \cdots \circ \chi_n(m) = w$$

where m is a conjugates to a Christoffel words.

Main goals

Our solution : Free group morphism

Free group

Recall that

- The inverse of an element $l \in \mathcal{F}(\mathcal{A})$ is denoted by l^{-1}.
- Each element of the free group may be represented by a reduced word, which is a product of the letters or their inverses, without the factors xx^{-1} or $x^{-1}x$ for $x \in \mathcal{A}$.
- An element w of the free group is called positive if $w \in \mathcal{A}^*$

Ternary alphabet

	a	b	c
λ_a	a	ab	ac
λ_b	ab^{-1}	b	bc
λ_c	ac^{-1}	bc^{-1}	c

	a	b
f_b	a	b
f_a	b	c

Theorem

If w is a Lyndon perfectly clustering word on $\{a, b, c\}$, there exists a sequence of group morphisms, $g_1, g_2, \ldots, g_k \in \{\lambda_a, \lambda_b, \lambda_c, \lambda_a^{-1}, \lambda_b^{-1}, \lambda_c^{-1}\}$ and $f \in \{f_a, f_b\}$ such that

$$g_1 \circ \cdots \circ g_k \circ f(m_w) = w$$

where m_w is a Christoffel word.

General case

For each ℓ in A_r

$$\lambda_\ell(a) = \begin{cases} a\ell^{-1}, & \text{if } a < \ell; \\ a, & \text{if } a = \ell; \\ \ell a, & \text{if } a > \ell; \end{cases} \quad \text{and} \quad \rho_\ell(a) = \begin{cases} a\ell, & \text{if } a < \ell; \\ a, & \text{if } a = \ell; \\ \ell^{-1}a, & \text{if } a > \ell. \end{cases}$$

Let f_{ℓ,A_r} be a monoid morphism A_r^* to A_{r+1}^* defined by

$$f_{\ell,A_r}(a_i) = \begin{cases} a_i & \text{if } a_i < \ell, \\ a_{i+1} & \text{otherwise,} \end{cases}$$

where $a_i \in A_r$.

General case

Theorem

Let w be a Lyndon complete perfectly clustering word on the totally ordered alphabet A. There exists a sequence of free group morphisms, namely $g = g_1 \circ \cdots \circ g_k$, such that

$$g(a) = w$$

and $g_i \in \{\lambda_{\ell_j}, \rho_{\ell_j}, \lambda_{\ell_j} \circ f_{\ell_j,B}, \rho_{\ell_{i+1}} \circ f_{\ell_{j+1},B} \mid \ell_j \in A \text{ and } B \subset A\}$ for $i \in \{1, \ldots, k\}$.

Example : The word *adbcbdadbd* is perfectly clustering and its sequence g is

$$\lambda_b \circ f_{b,\{a,b,c\}} \circ \rho_c \circ \lambda_a \circ f_{a,\{a,b\}} \circ \rho_b.$$

Idea of proof

- Relation between λ_ℓ and λ_ℓ^{-1} :

$$\widetilde{\tau} \circ \lambda_\ell^{-1} = \lambda_{\tau(\ell)} \circ \widetilde{\tau}$$

  with $\tau(a_k) = a_{r-k+1}$ for all $a_k \in \mathcal{A}$.
- Let w be a perfectly clustering. Then $\lambda_\ell(w)$ is positive and perfectly clustering if

$$\sum_{j>\ell} |w|_j > \sum_{j<\ell} |w|_j.$$

- A word w is perfectly clustering if and only if $\widetilde{\tau}(w)$ is perfectly clustering.
- Let w be a perfectly clustering. Then $\lambda_\ell^{-1}(w)$ is positive and perfectly clustering if

$$\sum_{j>\ell} |w|_j < \sum_{j<\ell} |w|_j.$$

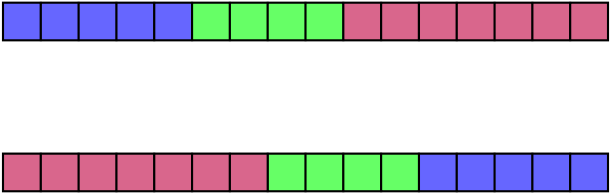
Idea proof Lyndon word

- The maps λ_ℓ^{-1} from $\mathcal{A}^*$ to $(\mathcal{A} \cup \mathcal{A}^{-1})^*$ is increasing.
- Let $w \in \mathcal{A}^*$ be a Lyndon word. If $\lambda_\ell^{-1}(w)$ is positive, then $\lambda_\ell^{-1}(w)$ is a Lyndon word.
- Let w is a Lyndon perfectly clustering word. If $\lambda_\ell(w)$ is positive, then $\lambda_\ell(w)$ is a Lyndon word.

Induction on symmetric discrete interval exchange transformation

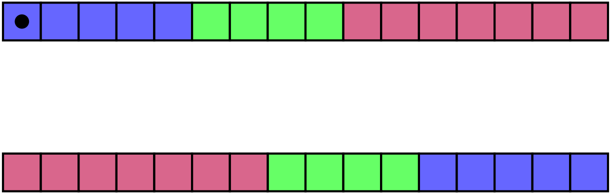
Symmetric discrete interval exchange transformation

A symmetric discrete r-interval exchange transformation with length vector $c = (c_1, c_2, \ldots, c_r)$ defined on a set of $|c|$ points.



Symmetric discrete interval exchange transformation

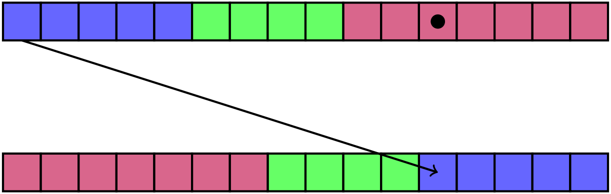
A symmetric discrete r-interval exchange transformation with length vector $c = (c_1, c_2, \ldots, c_r)$ defined on a set of $|c|$ points.



a

Symmetric discrete interval exchange transformation

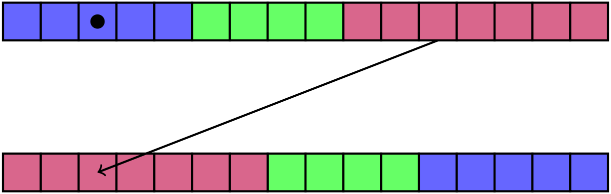
A symmetric discrete r-interval exchange transformation with length vector $c = (c_1, c_2, \ldots, c_r)$ defined on a set of $|c|$ points.



ac

Symmetric discrete interval exchange transformation

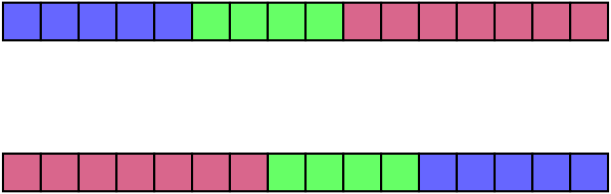
A symmetric discrete r-interval exchange transformation with length vector $c = (c_1, c_2, \ldots, c_r)$ defined on a set of $|c|$ points.



aca

Symmetric discrete interval exchange transformation

A symmetric discrete r-interval exchange transformation with length vector $c = (c_1, c_2, \ldots, c_r)$ defined on a set of $|c|$ points.



acacacbbcacacbbc...

Perfectly clustering words VS SDIET

Theorem (Ferenczi and Zamboni, 2013)

A word w is a perfectly clustering word if and only if the mapping from the last column to the first column is a minimal symmetric discrete interval exchange transformation.

a	n	a	n	a	s
a	n	a	s	a	n
a	s	a	n	a	n
n	a	n	a	s	a
n	a	s	a	n	a
s	a	n	a	n	a

Perfectly clustering words VS SDIET

Theorem (Ferenczi and Zamboni, 2013)

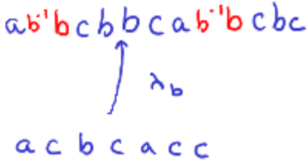
A word w is a perfectly clustering word if and only if the mapping from the last column to the first column is a minimal symmetric discrete interval exchange transformation.

a	n	a	n	a	s
a	n	a	s	a	n
a	s	a	n	a	n
n	a	n	a	s	a
n	a	s	a	n	a
s	a	n	a	n	a

Induction

- Let T be a minimal symmetric discrete k-interval exchange transformation on $U = \{1, \ldots, n\}$.
- Let $(I_a)_{a\in\mathcal{A}}$ a partition of U and $(J_a)_{a\in\mathcal{A}}$ another partition of U such that $T(I_a) = J_a$.
- Define $N_a = I_a \cap J_a$ for all $a \in \mathcal{A}$.
- If one of the $N_a \neq \emptyset$. Denotes by b one of the letter such that $N_a \neq \emptyset$.
- Takes $N = N_b \cup \bigcup_{a\in\mathcal{A}-\{b\}} I_a$
- Then the induced transformation by T on I is a minimal symmetric discrete interval exchange transformation with at most k intervals. (L. 2019)

Examples



Link between perfectly clustering words and band bricks over certain gentle algebra

The gentle algebra Λ_n

The following quiver with relations :

$$Q_n : 1 \underset{\beta_1}{\overset{\alpha_1}{\leftleftarrows}} 2 \underset{\beta_2}{\overset{\alpha_2}{\leftleftarrows}} \cdots \underset{\beta_{n-1}}{\overset{\alpha_{n-1}}{\leftleftarrows}} n \qquad R_n : \beta_i \alpha_{i+1} = 0,\ \alpha_i \beta_{i+1} = 0$$

gives the generators and relations of Λ_n. Its indecomposable representations :

1. string representations given by a certain words in the arrows of Q_n
2. band representations $B_{z,m,\lambda}$ given by a certain non-oriented cycles z in Q_n and two parameters.

Representation of Λ_n

- We defined the cycle $z_i = \alpha_1\alpha_2\ldots\alpha_{i-1}\beta_{i-1}^{-1}\ldots\beta_2^{-1}\beta_1^{-1}$ for each $i \in \{1,2,\ldots,n\}$
- For each primitive word $w = a_1a_2\ldots a_r \in \{1,2,\ldots,n\}^*$, we defined the cycle $\varphi(w) = z_{a_1}z_{a_2}\ldots z_{a_r}$.

$$Q_n : 1 \underset{\beta_1}{\overset{\alpha_1}{\leftleftarrows}} 2 \underset{\beta_2}{\overset{\alpha_2}{\leftleftarrows}} \cdots \underset{\beta_{n-1}}{\overset{\alpha_{n-1}}{\leftleftarrows}} n \qquad R_n : \beta_i\alpha_{i+1} = 0,\ \alpha_i\beta_{i+1} = 0$$

Theorem (Dequêne, L., Palu, Plamondon. Reutenauer, Thomas)

A primitive word w on n letters is perfectly clustering if and only if the band Λ_n-module $B_{\varphi(w),1,\lambda}$ is a brick for some (equivalently any) $\lambda \in k^\times$.

A representation M of Λ_n is called a brick if $\mathrm{End}_{\lambda_n}(M) \cong k$.

Gessel-Reutenauer bijection

The Burrows-Wheeler transformation is a particular case of the Gessel-Reutenauer bijection.

Definition

The Gessel-Reutnauer bijection is a map Φ sending each word $w \in A^*$ to the multiset of primitive necklaces obtain by

1. computing the standard permutation of w, $st(w)$
2. computing all the cycles of the inverse of $st(w)$
3. replacing the number i by the i-th letter of w.

Example : *baacbcab*

$$\begin{aligned} st(baacbcab) &= 41275836 \\ st(baacbcab)^{-1} &= 23715846 \\ \text{cycles } st^{-1} &= (1,2,3,7,4),(5),(6,8) \\ \Phi(baacbcab) &= \{(baaac),(b),(cb)\} \end{aligned}$$

Algebra to Surface to Dyck word

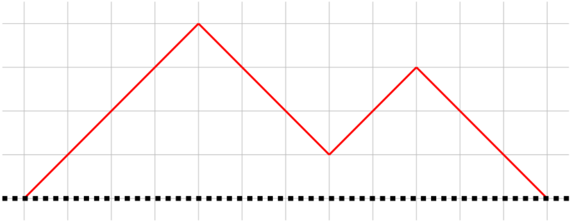
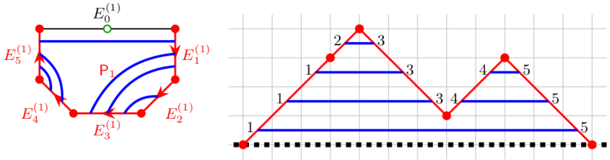


FIGURE 11. Dyck path of Dyck word *uuuuddduuddd*.



The g-vector $(-3, -1, 3, -2, 3)$ and the words constructing along the curves $M_{(-3,-1,3,-2,3)} = \{(54545131), (3231)\}$

Perfectly clustering words and band bricks

Theorem (Dequêne, L., Palu, Plamondon. Reutenauer, Thomas)

Let $(a_1, \ldots, a_n)$ be the g-vector of a simple closed multislalom with a_1 a negative integer and a_i a non-negative integer for $2 \leq i \leq n$. Let $M_{(a_1,\ldots,a_n)}$ be the multiset of circular words defined by $(a_1, \ldots, a_n)$. Then,

$$f(M_{(a_1,\ldots,a_n)}) = \Phi(n^{a_n} \ldots 2^{a_2}), \tag{1}$$

where f is the erasing morphism $f(1) = \varepsilon$ and $f(i) = i$ for $i \in \{2, \ldots, n\}$.

Words in the multiset $\Phi(w)$

Lemma

Let w be a weakly decreasing word. Then, each circular word in $\Phi(w)$ is perfectly clustering.

Sketch of proof :

- u a necklace in $\Phi(w)$.
- $u_1 \neq u_2$ conjugates of u
- $u_1^\omega < u_2^\omega$ iff $u_1 < u_2$
- The last column of the tableau of u is weakly decreasing since $\Phi(w)$ is weakly decreasing.

Number of conjugacy classes

Corollary (Dequêne, L., Palu, Plamondon. Reutenauer, Thomas)

Let $n \geq 1$ *and* $(\alpha_2, \ldots, \alpha_n)$ *be a* $(n-1)$*-tuple of nonnegative integers. The number of distinct conjugacy classes of words appearing in* $\Phi(n^{\alpha_n} \ldots 2^{\alpha_2})$ *is at mots* $\lceil (n-1) \rceil / 2$*.*

Thank you