CLUSTERING WORDS AND INTERVAL EXCHANGES

$A = \{1, \ldots, r\}$ or $A = \{a_1 < a_2 < \cdots < a_r\}$.
Primitive word $w = w_1 \cdots w_n$: not a power of another word.

Parikh vector of w : $(n_1, \ldots, n_k)$, n_i : number of occurrences of a_i in w.
Conjugates : $w_{i,1} \cdots w_{i,n} = w_i \cdots w_n w_1 \cdots w_{i-1}$, $1 \leq i \leq n$, ordered by ascending lexicographical order.
Burrows-Wheeler transform $B(w) = w_{1,n} w_{2,n} \cdots w_{n,n}$.

$w = 2314132 \rightarrow$
$132231\underline{4}$
$141322\underline{3}$
$223141\underline{3}$
$231413\underline{2}$
$314132\underline{2}$
$322314\underline{1}$
$413223\underline{1}$
$B(w) = 4332211$.

π--**clustering :** $B(w) = a_{\pi 1}^{n_{\pi 1}} \cdots a_{\pi r}^{n_{\pi r}}$,for $\pi \neq Id$ permutation on $\{1, \ldots, r\}$. **perfectly clustering** : π-clustering for $\pi i = r + 1 - i$, $1 \leq i \leq r$.

WHICH ARE THE CLUSTERING WORDS ?

Perfectly clustering implies strongly (or circularly) rich (Restivo, Rosone) : w^2 has $|w^2| + 1$ distinct palindromic factors.

Clustering on two letters implies Sturmian.

No converse.

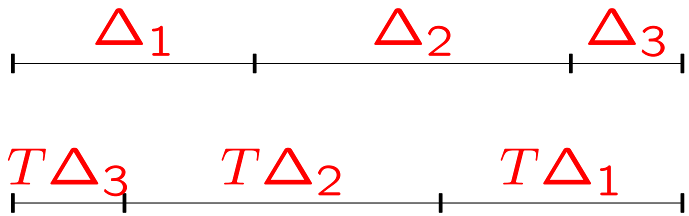
Theorem 1 *The following are equivalent :*

1. w *is* π*-clustering,*
2. ww *occurs in a trajectory of a minimal discrete* r*-interval exchange with permutation* π*,*
3. ww *occurs in a trajectory of a continuous* r*-interval exchange with permutation* π *satisfying the i.d.o.c. condition.*

Continuous r-interval exchange : defined by a probability vector $(\alpha_1, \alpha_2, \ldots, \alpha_r)$, and permutation π by

$$Tx = x + \tau_i \quad x \in \Delta_i,$$

$\Delta_i = \left[\sum_{j<i} \alpha_j, \sum_{j \le i} \alpha_j\right[$, $\tau_i = \sum_{\pi^{-1}(j)<\pi^{-1}(i)} \alpha_j - \sum_{j<i} \alpha_j$.



Discrete r-interval exchange : defined on a set of $n_1 + \cdots + n_r$ points $x_1, \ldots, x_{n_1+\cdots+n_r}$

$$Tx_k = x_{k+s_i} \quad x_k \in \Delta_i,$$

$\Delta_i = \{x_k, \sum_{j<i} n_j < k \le \sum_{j \le i} n_j\}$, $s_i = \sum_{\pi^{-1}(j)<\pi^{-1}(i)} n_j - \sum_{j<i} n_j$.

Example $1122334 \to 4332211$.

Minimal : no invariant subset (nonempty, closed).

Trajectory : $x_n = i$ if $T^n x$ belongs to Δ_i, $1 \leq i \leq r$.

I.d.o.c. condition : technical, stronger than minimality, weaker than total irrationality,.

For $r = 2$ I.d.o.c. interval exchange $=$ irrational rotation. Trajectories $=$ Sturmian infinite words.

ELEMENTS OF PROOF

A clustering word defines a discrete interval exchange.

Lemma 1 *If w is π-clustering, the mapping $w_{1,j} \mapsto w_{n,j}$ defines a discrete r-interval exchange transformation with length vector $(n_1, n_2, \ldots, n_r)$, and permutation π.*

$w = 2314132 \to B(w) = 4332211$.

$1 - - - - - 4$

$1 - - - - - 3$

$2 - - - - - 3$

$2 - - - - - 2$

$3 - - - - - 2$

$3 - - - - - 1$

$4 - - - - - 1$

Continuous and discrete interval exchanges produce the same finite words.

MINIMALITY AND INVERTIBILITY

Lemma 2 *(Crochemore, Désarménien, Perrin or Mantaci, Restivo, Rosone, Sciortino) If w and w' are words such that $B(w) = B(w')$, then w and w' are cyclically conjugate.*

Lemma 3 *If the discrete r-interval exchange T with length vector $(n_1, n_2, \ldots, n_r)$, and permutation π is not minimal, the word $(\pi 1)^{n_{\pi 1}} \ldots (\pi r)^{n_{\pi r}}$ has no primitive pre-image by the Burrows-Wheeler transform.*

$111233444 \rightarrow 444332111$ is not minimal and gives two perfectly clustering words on smaller alphabets, 41 and 323.

WEAKER HYPOTHESES

Non-primitivity. For w not primitive, the Burrows-Wheeler can be defined (the lexicographical order is not strict). Lemma 3 : is not valid : $3333222211 = B(1322313223)$ though the discrete 3-interval exchange is not minimal, A modified Theorem 1 holds.

32221 has no antecedent, primitive or not, by the Burrows-Wheeler transformation.

Two permutations. $223331111 \to 111133322$ is a minimal discrete 3-interval exchange, $w = 123131312$ is such that ww occurs in trajectories of T but $B(w) = 323311112$.

BUILDING CLUSTERING WORDS

By discrete interval exchanges.

Minimality. Pak and Redlich $\rightarrow$ for $n = 3$ and $\pi 1 = 3, \pi 2 = 2, \pi 3 = 1$ and length vector (n_1, n_2, n_3), the interval exchange is minimal iff $(n_1 + n_2)$ and $(n_2 + n_3)$ are coprime.

$11223333 \rightarrow 333322111$ gives the perfectly clustering word 313131223.

Condition of minimality for $n \geq 4$?

By continuous interval exchanges

Use of the **self-dual induction** or its generalization $\rightarrow$ 13131312222 and 1313122213122131

are perfectly clustering.

$2^m(3141)^n32$ are perfectly clustering for any m and n.

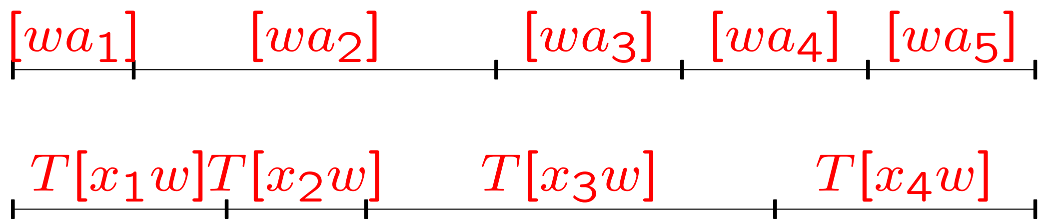
$525243425251615251616152516\!1$ is perfectly clustering.

$412323131241 2$ is π-clustering for $\pi1 = 4, \pi2 = 3, \pi3 = 1, \pi4 = 2$.

CHARACTERIZATION OF TRAJECTORIES

Theorem 2 *A uniformly recurrent infinite word sequence u is a trajectory of an r-interval exchange, defined by permutation π and satisfying the i.d.o.c. condition, if and only if the words of length one occurring in u are $L_1 = \{1, \ldots, r\}$ and it satisfies the following conditions*

- *if w is any word occurring in u, $A(w)$, resp. $D(w)$), the set of all letters x such that xw, resp. wx, occurs in u, is an interval, resp. an interval for the order of π,*
- *if $x \in A(w)$, $y \in A(w)$, $x \leq y$ for the order of π, $z \in D(xw)$, $t \in D(yw)$, then $z \leq t$,*
- *if $x \in A(w)$ and $y \in A(w)$ are consecutive in the order of π, $D(xw) \cap D(yw)$ is a singleton.*



Do these trajectories contain infinitely many ww? Yes for $\pi i = r + 1 - i$, $1 \leq i \leq r$. Open in general.