Pattern avoiding permutations in genome rearrangement problems: the transposition model

G. Cerbai, L. Ferrari

Dipartimento di Matematica e Informatica "U. Dini", Universitá degli Studi di Firenze,
Viale Morgagni 65, 50134 Firenze, Italy
giuliocerbai14@gmail.com,luca.ferrari@unifi.it

GASCom 2018, Athens, 18-20 June 2018.

The genome rearrangement problem

- Given a set of allowed mutations, understand how a genomic segment can evolve.
- Find a minimal sequence of mutations that transforms one genome into another one (*parsimony principle*).

The model

- Genome $\Longrightarrow$ **Permutation**.
- Mutation $\Rightarrow$ **Combinatorial operation**.

("Combinatorics of genome rearrangement", Fertin, Labarre et al., 2009)

Our case: the block transposition operation

The block transposition $\tau(i, j, l)$ transforms π into:

$$\pi_1 \cdots \pi_{i-1} \boxed{\pi_j \pi_{j+1} \cdots \pi_{l-1}} \boxed{\pi_i \pi_{i+1} \cdots \pi_{j-1}} \pi_l \cdots \pi_n.$$

("Sorting by transpositions", Bafna, Pevzner, 1998)

The genome rearrangement problem

- Given a set of allowed mutations, understand how a genomic segment can evolve.
- Find a minimal sequence of mutations that transforms one genome into another one (*parsimony principle*).

The model

- Genome $\Longrightarrow$ **Permutation**.
- Mutation $\Rightarrow$ **Combinatorial operation**.

("Combinatorics of genome rearrangement", Fertin, Labarre et al., 2009)

Our case: the block transposition operation

The block transposition $\tau(i, j, l)$ transforms π into:

$$\pi_1 \cdots \pi_{i-1} \boxed{\pi_j \pi_{j+1} \cdots \pi_{l-1}} \boxed{\pi_i \pi_{i+1} \cdots \pi_{j-1}} \pi_l \cdots \pi_n.$$

("Sorting by transpositions", Bafna, Pevzner, 1998)

The genome rearrangement problem

- Given a set of allowed mutations, understand how a genomic segment can evolve.
- Find a minimal sequence of mutations that transforms one genome into another one (*parsimony principle*).

The model

- Genome $\Longrightarrow$ **Permutation**.
- Mutation $\Rightarrow$ **Combinatorial operation**.

("Combinatorics of genome rearrangement", Fertin, Labarre et al., 2009)

Our case: the block transposition operation

The block transposition $\tau(i, j, l)$ transforms π into:

$$\pi_1 \cdots \pi_{i-1} \boxed{\pi_j \pi_{j+1} \cdots \pi_{l-1}} \boxed{\pi_i \pi_{i+1} \cdots \pi_{j-1}} \pi_l \cdots \pi_n.$$

("Sorting by transpositions", Bafna, Pevzner, 1998)

Genomic distance

Given $\pi, \rho \in S_n$, define:

$$d(\pi, \rho) = \text{minimum number of operations needed to transform } \pi \text{ into } \rho$$

- Lucky case: d is a distance.
- Luckier case: d is **left-invariant**
  $\Rightarrow$ computing d is equivalent to the problem of **sorting a permutation** using the minimum number of allowed operations.

Notations

Given a left-invariant distance, we define:

- $d(\pi) = d(\pi, id)$;
- $B_k(n) = \{\pi \in S_n : d(\pi) \leq k\}$;
- $B_k = \bigcup_{n \geq 0} B_k(n)$.

Genomic distance

Given $\pi, \rho \in S_n$, define:

$$d(\pi, \rho) = \text{minimum number of operations needed to transform } \pi \text{ into } \rho$$

- Lucky case: d is a distance.
- Luckier case: d is **left-invariant**
  $\Rightarrow$ computing d is equivalent to the problem of **sorting a permutation** using the minimum number of allowed operations.

Notations

Given a left-invariant distance, we define:

- $d(\pi) = d(\pi, id)$;
- $B_k(n) = \{\pi \in S_n : d(\pi) \leq k\}$;
- $B_k = \bigcup_{n \geq 0} B_k(n)$.

Genomic distance

Given $\pi, \rho \in S_n$, define:

$$d(\pi, \rho) = \text{minimum number of operations needed to transform } \pi \text{ into } \rho$$

- Lucky case: d is a distance.
- Luckier case: d is **left-invariant**
  $\Rightarrow$ computing d is equivalent to the problem of **sorting a permutation** using the minimum number of allowed operations.

Notations

Given a left-invariant distance, we define:

- $d(\pi) = d(\pi, id)$;
- $B_k(n) = \{\pi \in S_n : d(\pi) \leq k\}$;
- $B_k = \bigcup_{n \geq 0} B_k(n)$.

Genomic distance

Given $\pi, \rho \in S_n$, define:

$$d(\pi, \rho) = \text{minimum number of operations needed to transform } \pi \text{ into } \rho$$

- Lucky case: d is a distance.
- Luckier case: d is **left-invariant**
  $\Rightarrow$ computing d is equivalent to the problem of **sorting a permutation** using the minimum number of allowed operations.

Notations

Given a left-invariant distance, we define:

- $d(\pi) = d(\pi, id)$;
- $B_k(n) = \{\pi \in S_n : d(\pi) \leq k\}$;
- $B_k = \bigcup_{n \geq 0} B_k(n)$.

General problems

1. Characterize and enumerate the permutations of $B_k(n)$;
2. design sorting algorithms and study the related complexity issues;
3. compute the diameter of $B_k(n)$ and S_n;
4. characterize the permutations of $B_k(n)$ and S_n having maximum distance from the identity.

Idea:

Analyze the genomic model in terms of permutation patterns!

("A variant of the tandem duplication-random loss model of genome rearrangement", Bouvel, Rossin 2009)

General problems

1. Characterize and enumerate the permutations of $B_k(n)$;
2. design sorting algorithms and study the related complexity issues;
3. compute the diameter of $B_k(n)$ and S_n;
4. characterize the permutations of $B_k(n)$ and S_n having maximum distance from the identity.

Idea:

Analyze the genomic model in terms of permutation patterns!

("A variant of the tandem duplication-random loss model of genome rearrangement", Bouvel, Rossin 2009)

Permutation patterns

Definition of **Pattern**

A permutation ρ (pattern) is contained in π (we write $\rho \leq \pi$) if π has a subsequence of elements which is order isomorphic to ρ.

- π **avoids** ρ if it doesn't contain ρ.
- $\leq$ is a partial order on the set of all permutations $\mathcal{S}$.

Example:

- $213 \ \leq \ \overline{4}2\overline{3}1\overline{5}$, because $213 \cong 435$ (**occurrence**);
- $213 \ \not\leq \ 24531$.

Permutation class C:

A subset $C \subset \mathcal{S}$ closed downwards under this partial order (ideal).

Remark

A class C is uniquely determined by the minimal elements in $\mathcal{S} \setminus C$ (**basis**)!

Permutation patterns

Definition of **Pattern**

A permutation ρ (pattern) is contained in π (we write $\rho \leq \pi$) if π has a subsequence of elements which is order isomorphic to ρ.

- π **avoids** ρ if it doesn't contain ρ.
- $\leq$ is a partial order on the set of all permutations $\mathcal{S}$.

Example:

- $213 \leq \overline{4}2\overline{3}1\overline{5}$, because $213 \cong 435$ (**occurrence**);
- $213 \not\leq 24531$.

Permutation class C:

A subset $C \subset \mathcal{S}$ closed downwards under this partial order (ideal).

Remark

A class C is uniquely determined by the minimal elements in $\mathcal{S} \setminus C$ (**basis**)!

Patterns in the transposition model

Transposition distance

- $d(\cdot,\cdot)$ block transposition distance;
- $B_k = \{\pi \in \mathcal{S} : d(\pi) \leq k\}$ ball of radius k.

Remark

For each $k \geq 0$, B_k is a permutation class.

Idea

If $\rho \leq \pi$ and $\tau_1, \ldots, \tau_k$ sort π, perform the same block transpositions to one occurrence of ρ, without considering the other elements of π.

Main goals

1. Investigate the structure of B_k (generating permutations).
2. Characterize B_k in terms of avoided patterns (find the basis!).

Patterns in the transposition model

Transposition distance

- $d(\cdot,\cdot)$ block transposition distance;
- $B_k = \{\pi \in \mathcal{S} : d(\pi) \leq k\}$ ball of radius k.

Remark

For each $k \geq 0$, B_k is a permutation class.

Idea

If $\rho \leq \pi$ and $\tau_1, \ldots, \tau_k$ sort π, perform the same block transpositions to one occurrence of ρ, without considering the other elements of π.

Main goals

1. Investigate the structure of B_k (generating permutations).
2. Characterize B_k in terms of avoided patterns (find the basis!).

Tools and notations

- **Strip** of $\pi = \pi_1 \cdots \pi_n$: a maximal consecutive substring $\pi_i, \pi_{i+1}, \ldots, \pi_j$ such that $\pi_{l+1} = \pi_l + 1$ for each $l = i, \ldots, j-1$.
- **Reduced permutation**: each strip has length 1 (i.e. $\pi_{l+1} \neq \pi_l + 1$ for each l).
- **red**(π): the reduced permutation obtained from π by contracting each non trivial strip and suitably rescaling the elements.

Example:

$$\pi = 123/6/45/7 \xrightarrow{contract} 1/6/4/7 \xrightarrow{rescale} 1\ 3\ 2\ 4 = red(\pi).$$

Remark

- $red(\pi) \leq \pi$;
- $d(red(\pi)) = d(\pi)$.

Tools and notations

Monotone inflation

$\pi = \pi_1 \cdots \pi_n$, $v = (v_1, \ldots, v_n)$ vector of non-negative integers; the **monotone inflation** of π through v is the permutation $\pi[v]$ obtained by:

1. replacing each element π_i with id_{v_i};
2. rescaling the new strips in accordance to the order of π.

Example:

$$\pi = 1324,\ v = (2,1,4,0)\ \Rightarrow \pi[v] = \underbrace{12}_{1}\underbrace{7}_{3}\underbrace{3456}_{2}\underbrace{\quad}_{4} = 1273456$$

- $MI(\pi) = \{\pi[v] : v \text{ vector of non-negative integers}\}$;
- $MI(C) = \bigcup_{\pi \in C} MI(\pi)$.

Monotone inflations and geometric grid classes

Remark

M $\{-1,0,1\}$-matrix, $Geom(M)$ the geometric grid class of M.
$\pi \in \mathcal{S}$, M_π its permutation matrix. Then:

1. $Geom(M_\pi) = Geom(M_{red(\pi)})$;
2. $MI(\pi) = Geom(M_\pi)$;
3. $MI(\pi) = MI(red(\pi))$.

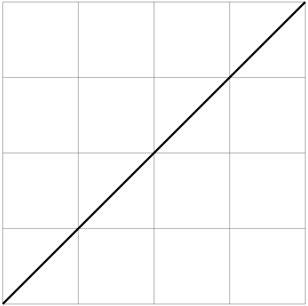
("Geometric grid classes of permutation", Albert, Atkinson et al., 2011)

Corollary

C set of reduced permutations.

- $MI(C)$ is a class of pattern avoiding permutations;
- $MI(C)$ is strongly rational and finitely based.

Starting point: the ball of radius 1



$B_0 = MI(1)$

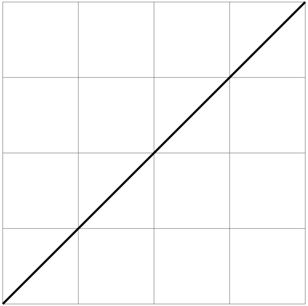
$\Longrightarrow$

$B_1 = MI(1324)$

Geometrical construction:

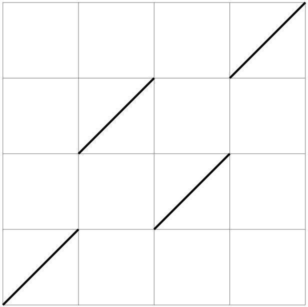
$$1 \xrightarrow{\text{inflation}} 1234 \xrightarrow{\text{transposition}} 1324$$

Starting point: the ball of radius 1



$B_0 = MI(1)$

$\Longrightarrow$

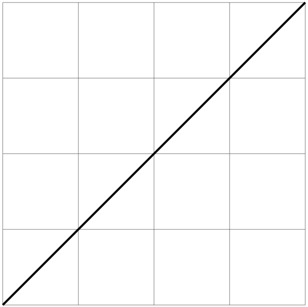


$B_1 = MI(1324)$

Geometrical construction:

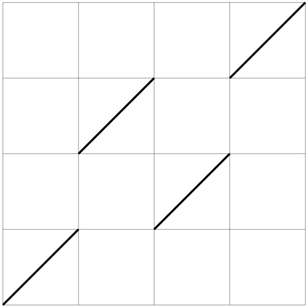
$$1 \xrightarrow{\text{inflation}} 1234 \xrightarrow{\text{transposition}} 1324$$

Starting point: the ball of radius 1



$B_0 = MI(1)$

$\Longrightarrow$



$B_1 = MI(1324)$

Geometrical construction:

$$1 \xrightarrow{inflation} 1234 \xrightarrow{transposition} 1324$$

Starting point: the ball of radius 1

Theorem

1. $B_1 = MI(1324)$ (one reduced **generating permutation**);
2. $\pi \in B_1 \iff \pi$ avoids $321, 2143, 2413, 3142$;
3. the generating function of B_1 is:

$$F(x) = \sum_{n \geq 0} f_n x^n = \frac{1 - 3x + 4x^2 - x^3}{(1-x)^4},$$

where $f_n = \binom{n+3}{3} - 2\binom{n+2}{2} + \binom{n+1}{1} + \binom{n+0}{0}$ [compositions of n].

Starting point: the ball of radius 1

Theorem

1. $B_1 = MI(1324)$ (one reduced **generating permutation**);
2. $\pi \in B_1 \iff \pi$ avoids $321, 2143, 2413, 3142$;
3. the generating function of B_1 is:

$$F(x) = \sum_{n \geq 0} f_n x^n = \frac{1 - 3x + 4x^2 - x^3}{(1-x)^4},$$

where $f_n = \binom{n+3}{3} - 2\binom{n+2}{2} + \binom{n+1}{1} + \binom{n+0}{0}$ [compositions of n].

Starting point: the ball of radius 1

Theorem

1. $B_1 = MI(1324)$ (one reduced **generating permutation**);
2. $\pi \in B_1 \iff \pi$ avoids $321, 2143, 2413, 3142$;
3. the generating function of B_1 is:

$$F(x) = \sum_{n \geq 0} f_n x^n = \frac{1 - 3x + 4x^2 - x^3}{(1-x)^4},$$

where $f_n = \binom{n+3}{3} - 2\binom{n+2}{2} + \binom{n+1}{1} + \binom{n+0}{0}$ [compositions of n].

General case: B_k

Corollary

Let $k \geq 1$.

1. There exist $N = N(k)$ reduced permutations $\alpha^{(1)}, \ldots, \alpha^{(N)}$ of length $3k+1$, each at distance k from the identity, such that:

$$B_k = \bigcup_{j=1}^{N} MI(\alpha^{(j)}).$$

   [We call $\alpha^{(1)}, \ldots, \alpha^{(N)}$ the **generating permutations** of B_k];

2. the generating permutations of B_k are exactly the maximal reduced permutations of B_K;
3. B_k is strongly rational and finitely-based (via geometric grid classes);
4. each permutation of its basis has length at most $3k+1$.
   [easy bound $3k+2$]

General case: B_k

Corollary

Let $k \geq 1$.

1. There exist $N = N(k)$ reduced permutations $\alpha^{(1)}, \dots, \alpha^{(N)}$ of length $3k+1$, each at distance k from the identity, such that:

$$B_k = \bigcup_{j=1}^{N} MI(\alpha^{(j)}).$$

   [We call $\alpha^{(1)}, \dots, \alpha^{(N)}$ the **generating permutations** of B_k];
2. the generating permutations of B_k are exactly the maximal reduced permutations of B_K;
3. B_k is strongly rational and finitely-based (via geometric grid classes);
4. each permutation of its basis has length at most $3k+1$.
   [easy bound $3k+2$]

General case: B_k

Corollary

Let $k \geq 1$.

1. There exist $N = N(k)$ reduced permutations $\alpha^{(1)}, \ldots, \alpha^{(N)}$ of length $3k+1$, each at distance k from the identity, such that:

$$B_k = \bigcup_{j=1}^{N} MI(\alpha^{(j)}).$$

   [We call $\alpha^{(1)}, \ldots, \alpha^{(N)}$ the **generating permutations** of B_k];

2. the generating permutations of B_k are exactly the maximal reduced permutations of B_K;
3. B_k is strongly rational and finitely-based (via geometric grid classes);
4. each permutation of its basis has length at most $3k+1$.
   [easy bound $3k+2$]

General case: B_k

Corollary

Let $k \geq 1$.

1. There exist $N = N(k)$ reduced permutations $\alpha^{(1)}, \ldots, \alpha^{(N)}$ of length $3k+1$, each at distance k from the identity, such that:

$$B_k = \bigcup_{j=1}^{N} MI(\alpha^{(j)}).$$

   [We call $\alpha^{(1)}, \ldots, \alpha^{(N)}$ the **generating permutations** of B_k];
2. the generating permutations of B_k are exactly the maximal reduced permutations of B_K;
3. B_k is strongly rational and finitely-based (via geometric grid classes);
4. each permutation of its basis has length at most $3k+1$.
   [easy bound $3k+2$]

Cheyne Homberger, Vincent Vatter, ***"On the effective and automatic enumeration of polynomial permutation classes"***, 2016.

- More general approach using peg permutation classes;
- algorithm for enumerating any permutation class with polynomial enumeration from a structural description (special case: B_k).

Open problems (still a lot of work...)

1. Enumeration of the generating permutations of B_k.
2. Better understanding of the basis permutations (enumeration, length,...).
3. Use this approach to analyze other genomic distances (reversal, delete-insertion,...).

Cheyne Homberger, Vincent Vatter, ***"On the effective and automatic enumeration of polynomial permutation classes"***, 2016.

- More general approach using peg permutation classes;
- algorithm for enumerating any permutation class with polynomial enumeration from a structural description (special case: B_k).

Open problems (still a lot of work...)

1. Enumeration of the generating permutations of B_k.
2. Better understanding of the basis permutations (enumeration, length,...).
3. Use this approach to analyze other genomic distances (reversal, delete-insertion,...).

A special case: the prefix-transposition distance

Block prefix-transposition:

$$\boxed{\pi_1\pi_2\cdots\pi_{j-1}}\ \boxed{\pi_j\pi_{j+1}\cdots\pi_{l-1}}\pi_l\cdots\pi_n$$

$$\Downarrow$$

$$\boxed{\pi_j\pi_{j+1}\cdots\pi_{l-1}}\ \boxed{\pi_1\pi_2\cdots\pi_{j-1}}\ \pi_l\cdots\pi_n.$$

("Sorting by prefix transpositions", Dias, Meidanis, 2002)

B_1^{pre}

- $B_1^{pre} = MI(213)$;
- $\pi \in B_1^{pre} \iff \pi$ avoids $321, 132$.

A special case: the prefix-transposition distance

Block prefix-transposition:

$$\boxed{\pi_1\pi_2\cdots\pi_{j-1}}\ \boxed{\pi_j\pi_{j+1}\cdots\pi_{l-1}}\pi_l\cdots\pi_n$$

$$\Downarrow$$

$$\boxed{\pi_j\pi_{j+1}\cdots\pi_{l-1}}\ \boxed{\pi_1\pi_2\cdots\pi_{j-1}}\ \pi_l\cdots\pi_n.$$

("Sorting by prefix transpositions", Dias, Meidanis, 2002)

B_1^{pre}

- $B_1^{pre} = MI(213)$;
- $\pi \in B_1^{pre} \iff \pi$ avoids $321, 132$.

Generating permutations

Corollary

For $k \geq 1$, the generating permutations of B_k^{pre} are exactly the reduced permutations of length $2k+1$ and distance k from the identity.

Remark

The inductive construction gives distinct generating permutations.

Let g_k be the number of generating permutations of B_k^{pre}:

$$\begin{cases} B_1^{pre} = MI(213) \implies g_1 = 1; \\ \\ g_k = \binom{2k}{2} \cdot g_{k-1} = \prod_{i=1}^{k} \binom{2i}{i} = \frac{(2k)!}{2^k}, \text{ for } k \geq 2. \end{cases}$$

Basis permutations

Theorem

Every permutation of the basis of B_k^{pre} has length at most $2k+1$.

Easy bound $2k+2$ as in the general case ($2k+1$ tricky).

Example

- $B_2^{pre} = MI(32415, 41325, 31425, 24135, 24315, 42135)$;
- $\pi \in B_2^{pre} \iff \pi$ avoids the patterns:

  1432, 2143, 4321
  13524, 14253, 24351, 25314, 25413,
  35142, 35214, 35241, 41352, 42513,
  42531, 43152, 51324, 52413, 53142.

Pattern avoiding permutations in genome rearrangement problems: the transposition model

G. Cerbai, L. Ferrari

Dipartimento di Matematica e Informatica "U. Dini", Universitá degli Studi di Firenze,
Viale Morgagni 65, 50134 Firenze, Italy
giuliocerbai14@gmail.com,luca.ferrari@unifi.it

GASCom 2018, Athens, 18-20 June 2018.