

Some enumeration results for sorting signed permutations by reversals

ASU Discrete Math Seminar

Dana C. Ernst
Northern Arizona University
March 25, 2022

Joint with F. Awik, F. Burkhart, H. Denoncourt, T. Rosenberg, A. Stewart

Brief Introduction to Genetics

- **DNA:** Double helix of nucleotides, complementary pairs A–T, G–C.
- **Gene:** Sequence of nucleotides, codes a specific protein.
- **Chromosome:** Ordering device for genes.
- **Genome:** Collection of chromosomes.
- **Mutations:** Two types:
 - **Point Mutations:** Mutations at the level of nucleotides.
 - **Genome Rearrangements:** Structural mutations to chromosomes at level of genes. Types: deletions, duplications, **translocation**, **inversion**, fission, fusion, etc.
- **Edit Distance:** The minimum number of genome rearrangements required to transform one genome into another. Approximates evolutionary distance.
 - mouse $\xrightarrow{251}$ human (149 inversions, 93 translocations, 9 fissions)
 - cabbage $\xrightarrow{3}$ turnip (all inversions)

Mathematical Model

- Two closely-related species typically have similar gene orders. Comparing two similar sequences of genes yields two permutations or signed permutations (depending on the mutation you want to model), one for each species.
- Each number in the permutation or signed permutation represents either a single gene or a conserved block of genes (sign of the number indicates the orientation of the gene).
- Translocation = Block Interchange:

$$5 \quad \boxed{2 \ 1} \quad 4 \quad \boxed{3 \ 7 \ 6} \mapsto 5 \quad \boxed{3 \ 7 \ 6} \quad 4 \quad \boxed{2 \ 1}$$

- Inversion = Reversal:

$$5 \quad \boxed{-2 \ -1 \ 4 \ -3} \quad -7 \ 6 \mapsto 5 \quad \boxed{3 \ -4 \ 1 \ 2} \quad -7 \ 6$$

Definition

Let T be generating set for S_n (respectively, S_n^\pm) such that $\rho^{-1} = \rho$ for all $\rho \in T$. For permutations (respectively, signed permutations) π and σ , we define the **distance** $d_T(\pi, \sigma)$ to be the minimum number of generators $\rho_1, \dots, \rho_k \in T$ such that

$$\pi \circ \rho_1 \circ \dots \circ \rho_k = \sigma.$$

Notation and Terminology

- $\text{Rk}_k(S_n, d_T) := \{\pi \in S_n \mid d_T(\pi) = k\}$ = perms in S_n of distance k
- $\text{rk}_k(S_n, d_T) := |\text{Rk}_k(S_n, d_T)|$ = # of perms in S_n of distance k
- $d_T^{\max}(S_n) := \max\{d_T(\pi) \mid \pi \in S_n\}$ = diameter of Cayley diagram
- A **maximal permutation** is a permutation that attains maximal distance.
- $\text{rk}_{\max}(S_n, d_T) := \#$ of maximal perms in S_n

Let T be the collection of transpositions in S_n and let $d_t(\cdot)$ be the corresponding distance ($t =$ transposition).

- $d_t(\pi) = n - \text{cyc}(\pi)$
- $\text{rk}_k(S_n, d_t) = \#$ of perms in S_n with $n - k$ cycles $= S(n, n - k)$
= Stirling numbers of the 1st kind
- $d_t^{\max}(S_n) = n - 1$
- $\text{Rk}_{\max}(S_n, d_t) =$ collection of n -cycles in S_n
- $\text{rk}_{\max}(S_n, d_t) = (n - 1)!$

Sorting By Adjacent Transpositions

Let T be the collection of adjacent transpositions in S_n and let $d_{at}(\cdot)$ be the corresponding distance. (at = adjacent transposition)

- $d_{at}(\pi) = \text{inv}(\pi) = \#$ of inversions in $\pi =$ Coxeter length
- $\text{rk}_K(S_n, d_{at}) = \#$ of perms in S_n with k inversions $= I(n, k)$
= Inversion/Mahonian numbers
- $d_{at}^{\max}(S_n) = \binom{n}{2}$
- $\text{Rk}_{\max}(S_n, d_{at}) = \{[n \cdots 321]\}$
- $d_{at}(S_n, \max) = 1$

Sorting By Block Interchanges

Let T be the collection of block interchanges in S_n and let $d_{bi}(\cdot)$ be the corresponding distance. (bi = block interchange)

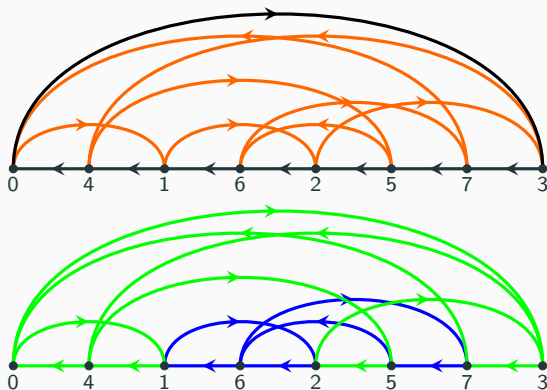
- $d_{bi}(\pi) = \frac{n + 1 - \text{cyc}(\text{DBG}(\pi))}{2}$
- $\text{rk}_k(S_n, d_{bi}) = \#$ of perms in S_n such that DBG has $n + 1 - 2k$ cycles
= $H(n, n + 1 - 2k)$ = Hultman numbers
- $d_{bi}^{\max}(S_n) = \lfloor \frac{n}{2} \rfloor$
- $\text{rk}_{\max}(S_n, d_{bi}) = \begin{cases} H(n, 1), & \text{if } n \text{ even} \\ H(n, 2), & \text{if } n \text{ odd} \end{cases}$

Note that

$$H(n, 1) = \begin{cases} \frac{2n!}{n+2}, & \text{if } n \text{ even} \\ 0, & \text{if } n \text{ odd.} \end{cases}$$

Example of Directed Breakpoint Graph

Directed breakpoint graph for $\pi = [4, 1, 6, 2, 5, 7, 3]$:



$$d_{bi}(\pi) = \frac{n + 1 - \text{cyc}(\text{DBG}(\pi))}{2} = \frac{7 + 1 - 2}{2} = 3$$

Sorting By Adjacent Block Interchanges

Let T be the collection of adjacent block interchanges in S_n and let $d_{abi}(\cdot)$ be the corresponding distance. (abi = adjacent block interchange)

- $d_{abi}(\pi) = ???$ (numerous formulas for lower and upper bounds)
- Special case: $d_{abi}([n \cdots 321]) = \lfloor \frac{n}{2} \rfloor + 1$
- $rk_k(S_n, d_{abi}) = ???$
- $d_{abi}^{\max}(S_n) = ???$ but $d_{abi}^{\max}(S_n) \geq \lfloor \frac{n+1}{2} \rfloor + 1$
- $rk_{\max}(S_n, d_{abi}) = ???$

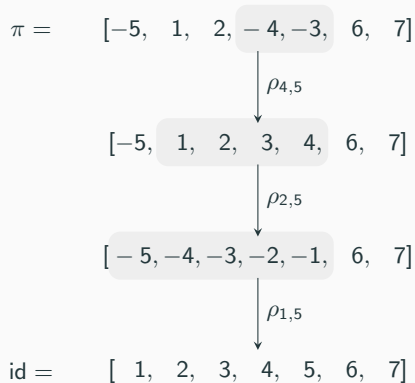
Let S_n^\pm be the set of signed permutations on $\{1, 2, \dots, n\}$. A **reversal** ρ_{ij} acts on a signed permutation π by reversing the order of values in positions i through j and changing all of their signs:

$$\pi \circ \rho_{ij} = [\pi_1, \dots, \pi_{i-1}, -\pi_j, -\pi_{j-1}, \dots, -\pi_{i+1}, -\pi_i, \pi_{j+1}, \dots, \pi_n].$$

Note that $\rho_{i,i}$ is the reversal that changes the sign in the i th position. Let T be the collection of reversals, so that $S_n = \langle T \rangle$ and let $d_r(\cdot)$ be the corresponding distance. ($r = \text{reversal}$)

$$|T| = \binom{n+1}{2}.$$

Example



Definition

Define S_{2n}^0 to be the set of unsigned permutations on $\{0, 1, 2, \dots, 2n + 1\}$ such that 0 and $2n + 1$ are fixed points. We define the **expansion transformation** from a signed permutation $\pi \in S_n^\pm$ to an unsigned permutation $\pi' \in S_{2n}^0$ as follows:

$$\pi'_0 = 0, \pi'_{2n+1} = 2n + 1,$$

and for all other values, if $\pi_i > 0$, then

$$\pi'_{2i-1} = 2\pi_i - 1, \pi'_{2i} = 2\pi_i,$$

while if $\pi_i < 0$, then

$$\pi'_{2i-1} = 2|\pi_i|, \pi'_{2i} = 2|\pi_i| - 1.$$

Note that the expansion transformation is injective, which implies that the process is uniquely reversible for an unsigned permutation in the image.

Breakpoint Diagram

Definition

The **breakpoint diagram** of π , denoted $BG(\pi)$, is a graph with colored edges constructed as follows.

- vertex set: $\{\pi'_0, \pi'_1, \dots, \pi'_{2n+1}\}$;
- black edge set: $\{\{\pi'_{2i}, \pi'_{2i+1}\} \mid 0 \leq i \leq n\}$;
- **orange** edge set: $\{\{2i, 2i + 1\} \mid 0 \leq i \leq n\}$.

Example



Reversal Distance Formula

Theorem (Hannenhalli & Pevzner)

The reversal distance of any signed permutation $\pi \in S_n^\pm$ is given by

$$d_r(\pi) = n + 1 - c(\pi) + h(\pi) + f(\pi)$$

- $c(\pi) := \#$ of cycles in $BG(\pi)$,
- $h(\pi) := \#$ of “hurdles” in $BG(\pi)$,
- $f(\pi)$ is 1 if π is a “fortress” and 0 otherwise.

Example

For $\pi = [-5, 1, 3, 2, 4, 6, -7, 8, 11, 10, 9]$, it turns out that $c(\pi) = 5$, $h(\pi) = 2$, and π is not a fortress, and so $d_r(\pi) = 11 + 1 - 5 + 2 + 0 = 9$.



Cyclic Shift of Breakpoint Diagram

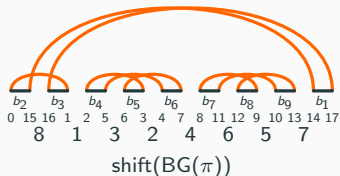
Definition

Let b_1, \dots, b_{n+1} denote the black edges of $BG(\pi)$ (from left to right). The **cyclic shift** of $BG(\pi)$, denoted $\text{shift}(BG(\pi))$, is the diagram obtained by shifting b_i to $b_{i-1} \pmod{n+1}$ while preserving the connections of the gray and black edges between vertices.

Example



→



Theorem

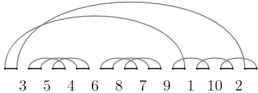
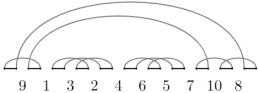
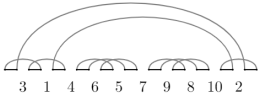
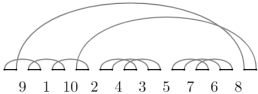
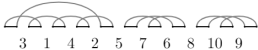
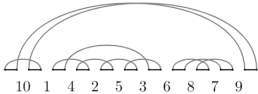
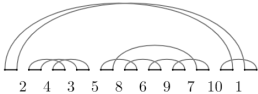
If $\pi \in S_n^\pm$, then $\text{shift}(\text{BG}(\pi))$ is the breakpoint diagram for a signed permutation in S_n^\pm , denoted $\text{shift}(\pi)$. Moreover, $d_r(\pi) = d_r(\text{shift}(\pi))$.

Definition

For $\pi, \gamma \in S_n^\pm$, define $\pi \sim \gamma$ if we can obtain $\text{BG}(\gamma)$ from $\text{BG}(\pi)$ by a sequence of cyclic shifts. If $\pi \sim \gamma$, we say that π and γ are **shift equivalent**. Define the **shift equivalence class** of $\pi \in S_n^\pm$ via

$$[\pi] = \{\gamma \in S_n^\pm \mid \gamma \sim \pi\}.$$

Example



Theorem (Folklore?)

$$d_r^{\max}(S_n^{\pm}) = \begin{cases} n, & n = 1, 3 \\ n + 1, & \text{otherwise.} \end{cases}$$

Theorem

Let $\pi \in S_n^{\pm}$ be a maximal signed permutation. Then

1. π is not a fortress;
2. π only contains positive entries;
3. All cycles of $BG(\pi)$ are hurdles \implies all cycles “sit side by side” or there is one that “covers” and the rest sit “side by side”;
4. Every element of $[\pi]$ is also a maximal signed permutation.

Definition

A **composition** of n is an ordered list of positive integers whose sum is n , denoted

$$\alpha = (\alpha_1, \dots, \alpha_k).$$

We refer to each α_i as a **part** of the composition. Let $C(n)$ denote the set of all compositions on n .

Example

$$C(4) = \{(1, 1, 1, 1), (1, 2, 1), (1, 1, 2), (2, 1, 1), (3, 1), (1, 3), (2, 2), (4)\}.$$

Definition

We define

$$C_{\text{odd}}^{>1}(n) := \{(\alpha_1, \dots, \alpha_k) \in C(n) \mid \text{each } \alpha_i \text{ is odd and greater than } 1\}$$

and let $c_{\text{odd}}^{>1}(n) := |C_{\text{odd}}^{>1}(n)|$.

Theorem

We have $c_{\text{odd}}^{>1}(1) = c_{\text{odd}}^{>1}(2) = 0$, $c_{\text{odd}}^{>1}(3) = 1$ and for $n \geq 4$

$$c_{\text{odd}}^{>1}(n) = c_{\text{odd}}^{>1}(n-2) + c_{\text{odd}}^{>1}(n-3).$$

The first few terms of the sequence are

$$0, 0, 1, 0, 1, 1, 1, 2, 2, 3.$$

It turns out that $c_{\text{odd}}^{>1}(n)$ is the Padovan sequence (OEIS A000931).

Theorem

For $n \neq 1, 3$, we have

$$\text{rk}_{\max}(S_n^{\pm}, d_r) = \sum_{(\alpha_1, \dots, \alpha_k) \in C_{\text{odd}}^{>1}(n+1)} \left(\prod_{i=1}^k \frac{2(\alpha_i - 1)!}{\alpha_i + 1} \right) \cdot \begin{cases} \alpha_1, & \text{if } k \neq 1 \\ 1, & \text{if } k = 1. \end{cases}$$

Remark

- Note that $\frac{2(\alpha_i - 1)!}{\alpha_i + 1} = H(\alpha_i - 1, 1)$ (where α_i is always odd).
- The complexity is subject to finding the compositions in $C_{\text{odd}}^{>1}(n+1)$.
- The first few terms of $\text{rk}_{\max}(S_n^{\pm}, d_r)$ when $n \neq 1, 3$ are 1, 8, 3, 180, 64, 8067.

Conjecture

We conjecture that

$$\lim_{n \rightarrow \infty} \frac{\text{rk}_{\max}(S_n, d_r)}{2(n-1)!} = 1 \quad \text{if } n \text{ is odd,}$$

$$\lim_{n \rightarrow \infty} \frac{\text{rk}_{\max}(S_n, d_r)}{2(n-3)!} = 1 \quad \text{if } n \text{ is even.}$$

If true, then if we choose a signed permutation uniformly at random, the probability of selecting a maximal signed permutation is about $n/2^n$ for n odd and $n(n-1)(n-2)/2^n$ for n even. That is, as n grows, it is exponentially unlikely to choose a maximal signed permutation at random.

We can partition the collection of signed permutations in S_n^\pm of reversal distance k according to the number of “trivial cycles” in their breakpoint diagrams. This yields

$$\text{rk}_k(S_n^\pm, d_r) = \sum_{i=0}^{n+1} a_{i,k} \binom{n+1}{i+1},$$

where $a_{i,k} := \#$ signed perms in S_i^\pm of reversal distance k with no trivial cycles. But some leading terms and trailing terms are 0.

Theorem

$$\text{rk}_k(S_n^\pm, d_r) = a_{k-1,k} \binom{n+1}{k} + a_{k,k} \binom{n+1}{k+1} + \cdots + a_{2k-1,k} \binom{n+1}{2k}.$$

This is a polynomial in n of degree $2k$ with rational coefficients.

Determining closed forms for $\text{rk}_k(S_n^\pm, d_r)$ using the above theorem is dependent on having values for $a_{k-1,k}, \dots, a_{2k-1,k}$. These values are independent of $n!$

Further Enumeration (continued)

Using brute-force computations (Python and Java), we have obtained data for $a_{k-1,k}, \dots, a_{2k-1,k}$ when $1 \leq k \leq 5$. This yields the following:

- $\text{rk}_1(S_n^\pm, d_r) = \frac{n(n+1)}{2} = \binom{n+1}{2}$
- $\text{rk}_2(S_n^\pm, d_r) = \frac{n(n-1)(n+1)^2}{6}$ (OEIS A004320... Aztec diamonds)
- $\text{rk}_3(S_n^\pm, d_r) = \frac{n^2(n-1)(n+1)(n+2)(7n-11)}{144}$
- $\text{rk}_4(S_n^\pm, d_r) = \text{Ugly}$ (not real-rooted)
- $\text{rk}_5(S_n^\pm, d_r) = \text{Ugly}$ (not real-rooted)

Moreover, for $n \neq 1, 3$, we have

$$\text{rk}_{\max}(S_n^\pm, d_r) = a_{n,n+1}.$$

Interesting side story. . .

Definition

We call a signed permutation $\pi \in S_n^\pm$ **terminal** if $d_r(\pi \circ \rho_{ij}) \leq d_r(\pi)$ for all ρ_{ij} .

Note that every maximal signed permutation in S_n^\pm is terminal. However, there exist terminal permutations that are not maximal! Terminal mean maximal in the language of posets as opposed to distance.

Example

Let $\pi = [2, -3, 1, -4] \in S_4^\pm$. It turns out that $d_r(\pi) = 4$ while $d_r(\pi \circ \rho_{ij}) \leq 4$ for all reversals ρ_{ij} , which implies that π is terminal but not maximal. However, the maximal reversal distance in S_4^\pm is 5.

Computing the first several terms of $\sum_{k=0}^{n+1} a_{n,k}$ coincides with OEIS A061714, which counts the number of circular permutations on $0, 1, \dots, 2n - 1$ where every two elements $2i, 2i + 1$ are adjacent and no two elements $2i - 1, 2i$ are adjacent. There is a connection to the Traveling Salesman Problem...