

Rearrangement Problems

- The problems are on permutations.

Problem

Given any two permutations π and σ in S_n and a set S of generators of S_n , find a minimum number of permutations in S , $\rho_1, \rho_2, \dots, \rho_k$ such that

$\pi \circ \rho_1 \circ \rho_2 \circ \dots \circ \rho_k = \sigma$. Equivalently, $\sigma^{-1} \circ \pi \circ \rho_1 \circ \rho_2 \circ \dots \circ \rho_k = i$ (identity of S_n) or $\sigma^{-1} \circ \pi = (\rho_1 \circ \rho_2 \circ \dots \circ \rho_k)^{-1} = \rho_k^{-1} \circ \rho_{k-1}^{-1} \circ \dots \circ \rho_1^{-1}$. $\rho_k^{-1} \circ \rho_{k-1}^{-1} \circ \dots \circ \rho_1^{-1}$ is a factorization of minimum length.

- *Finding a minimum length factorization of permutations is an NP-hard problem. (Even and Goldreich, J. Algorithm 2 (1981), 311 – 313.)*

Distance of Permutations

- Given a set of generators S of a permutation group. A distance between two permutations π and σ (with respect to S) can be defined as the minimum number of generators $\rho_1, \rho_2, \dots, \rho_k$ in S such that $\pi \circ \rho_1 \circ \rho_2 \circ \dots \circ \rho_k = \sigma$, denoted by $d_S(\pi, \sigma) = k$.
- Noticed that d_S is a metric. (?)

Definition (Left-invariant)

A distance d on S_n is **left-invariant** if for all π, σ and τ , $d(\pi, \sigma) = d(\tau \circ \pi, \tau \circ \sigma)$.



Sorting Sequences

- Given a permutations π , a sequence of rearrangements $\langle \rho_1, \rho_2, \dots, \rho_k \rangle$ is called a sorting sequence for π if $\pi \circ \rho_1 \circ \rho_2 \circ \dots \circ \rho_k = i$.

How to find the sorting sequences?

Definition (Cayley Graph)

Given a set S of generators of a permutation group G , the Cayley graph associated with (S, G) is the graph whose vertices are the elements of G and two vertices are adjacent if and only if the corresponding permutations can be transformed into another using an element of S .

For example, let $V(G) = S_4$ and $S = \{ (1, 2), (2, 3), (3, 4) \}$. This is also known as **Bubble sort**. The Cayley graph defined accordingly will be a 3-regular graph of order 24.

- Now, the distance of two permutations using S in above example is equal to the shortest distance in $G =_{\text{def}} \text{Cayley}(S, S_4)$.
- Draw the graph $G =_{\text{def}} \text{Cayley}(S, S_4)$ is not difficult. (?)



Reversals Distance

Problem

Use only reversals to find the minimum factorization.

Review

Definition (Reversal of a segment)

$$\text{Let } \pi = (\pi_1 \pi_2 \dots \pi_n) = \begin{bmatrix} 1 & 2 & 3 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{bmatrix}$$

The reversal of the segment $\pi_i \pi_{i+1} \dots \pi_j$, $i < j$ is the permutation

$(\pi_1 \pi_2 \dots \pi_{i-1} \pi_j \pi_{j-1} \dots, \pi_i \pi_{j+1} \pi_{j+2} \dots \pi_n)$, denoted by $\pi^{i,j}$.

(*) $\pi^{i,j}$ can be written as $\pi \circ \rho$ where

$\rho = (1, 2, \dots, i-1, j, j-1, j-2, \dots, i, j+1, \dots, n)$. (The idea can be extended to signed permutations.)

Continued on Reversals

Observation

At least one reversal is needed if (1) $\pi_1 > 1$, (2) $\pi_n < n$ and $\pi_{j+1} > \pi_j + 1$ or $\pi_j > \pi_{j+1}$.

Note

If a group of genes appears consecutively in several species, then they must have been presented in the same order on the ancestral species and were not separated during evolution.

Definition (Linear extension)

The linear extension of (signed or unsigned) permutation π of $[n]$ is the permutation of $[0, n+1]$ defined by $\pi^l = (0 \ \pi_1 \ \pi_2 \ \dots, \ \pi_n \ n+1)$.

Example

(0 4 8 9 7 6 5 1 3 2 10) is a linear extension of (4 8 9 7 6 5 1 3 2).

Breakpoints

Definition (Breakpoints)

Let $\pi = (\pi_1 \pi_2 \dots, \pi_n)$ and $\pi^l = (0 \pi_1 \pi_2 \dots, \pi_n n+1)$ be the linear extension of π . A point of π is an ordered pair (π_i^l, π_{i+1}^l) denoted by $\pi_i^l \cdot \pi_{i+1}^l$ in short, $0 \leq i \leq n$. ($\pi_0 = 0, \pi_{n+1} = n+1$)

Definition

1. If $\pi_i^l + 1 = \pi_{i+1}^l$, then (π_i^l, π_{i+1}^l) is called an **adjacency** point.
2. If $\pi_i^l - 1 = \pi_{i+1}^l$, then (π_i^l, π_{i+1}^l) is called a **reverse adjacency** point.
3. If it is **not an adjacency** point, then it is called a **break point**.
4. If it is **neither an adjacency or reverse adjacency**, then it is called a **strong break point**.

Breakpoints Continued

Example

In $(0\ 4\ 8\ 9\ 7\ 6\ 5\ 1\ 3\ 2\ 10)$, the linear extension of $(4\ 8\ 9\ 7\ 6\ 5\ 1\ 3\ 2)$, $(8, 9)$ is an adjacency, $(7, 6)$, $(6, 5)$ and $(3, 2)$ are reverse adjacency. All the other pairs are strong break points.

Notations

$p(\pi)$: The number of points

$bp(\pi)$: The number of break points

$sb(\pi)$: The number of strong break points

(*) We shall use break points to find the distance between two permutations.

Bounds on Reversal Distance

- Identity permutation has no breakpoints.
- If we aiming at sorting a permutation π into the identity, it suffices to eliminate all the breakpoints in the permutation π .

Proposition

Let $d_R(\pi)$ denote the sorting distance by way of reversals. Then,
 $d_R(\pi) \geq bp(\pi)/2$.

Proof. Since every reversal can eliminate at most two breakpoints, the proof follows. 

(*) This estimation is not quite accurate in general.

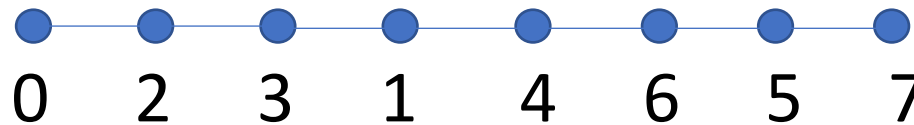
Breakpoint graph of a Permutation

Definition (Breakpoint graph)

The breakpoint graph of a permutation $\pi = (\pi_1 \pi_2 \dots, \pi_n)$ is an edge-colored graph $G(\pi)$ with vertex set $\{\pi_0, \pi_1, \pi_2, \dots, \pi_n, \pi_{n+1}\} = [0, n+1]$. In $G(\pi)$, π_i is joined to π_{i+1} by a blue edge for $0 \leq i \leq n$, π_i is joined to π_j by a red edge if $|\pi_i - \pi_j| = 1$, $0 \leq i \neq j \leq n$.

Example: $\pi = (2 \ 3 \ 1 \ 4 \ 6 \ 5)$

$G(\pi)$



$\{0, 1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{5, 6\}$ and $\{6, 7\}$ are colored red.

Breakpoint Graph Continued

- The graph $G(\pi)$ is an even graph such that all vertices are equal to 4 except 0 and $n+1$ whose degree are 2.
- Since the graph is connected, it is an Eulerian graph and also it can be decomposed into cycles.
- *A cycle in an edge-colored graph is called **alternating** if the colors of every two consecutive edges are distinct.*
- *A **2-edge-colored graph is balanced** if for each vertex colors occurred in its incident edges are balanced, that is the two colors occur the same number of times.*
- (*) It is not difficult to prove that $G(\pi)$ is a balanced graph and thus $G(\pi)$ contains an alternating Eulerian circuit. Moreover, $G(\pi)$ can be decomposed into alternating cycles.*

Example of Alternating Cycles

Example: $\pi = (2 \ 3 \ 1 \ 4 \ 6 \ 5)$

Alternating cycles are:

$(2, 3), (6, 5), (4, 6, 7, 5), (0, 2, 1, 3, 4, 1)$ (Generalized cycle)

(*) We are interested in the decomposition of the breakpoint graph into the maximum number of alternating cycles! The number of such cycles is denoted by $c(\pi)$.

For the above example, $c(\pi) = 4$.



Important Result

Theorem (Bafna and Pevzner, 1996, *SIAM J. on Computing*, 25:272-289)

$d_R(\pi) \geq n + 1 - c(\pi)$ where $c(\pi)$ is the maximum number of alternating cycles in the decomposition of $G(\pi)$.

Proof. If ρ is a reversal, then $c(\pi \circ \rho) - c(\pi) \leq 1$. $d_R(\pi) \geq c(i) - c(\pi) = n + 1 - c(\pi)$.

Example: $\pi = (2 \ 3 \ 1 \ 4 \ 6 \ 5)$

Step 1: Reverse **2 3**, we have (**3 2** 1 4 6 5)

Step 2: Reverse **6 5**, we have (**3 2** 1 4 **5 6**)

Step 3: Reverse **3 2 1**, we have (**1 2 3** 4 **5 6**)

(*) $d_R(\pi) \geq 6 + 1 - 4 = 3$. In fact, the answer is 3.

(*) In general, this bound is tighter than the breakpoints bound $d_R(\pi) \geq bp(\pi)/2$. Though, in this example, they are the same.

Final Comments

- *The goal of finding maximum number of alternating cycles in the decomposition of breakpoint graph is very difficult in general. It is not easy to obtain an explicit solution. But, we are able to estimate the upper bounds such as the theorem mention in previous page.*

