

Transposition Distance

- The term *transposition* comes from biology and refers to *transposons*, which are sequences of DNA that *can be displaced in a genome*.
- Note. Algebraic transposition is a 2-cycle in a permutation, for (i, j) which represents $i \rightarrow j$ and $j \rightarrow i$.

Definition (Interval of permutation)

For permutation $\pi = (\pi_1 \pi_2 \dots, \pi_n)$, an interval is a set $\{\pi_a, \pi_{a+1}, \dots, \pi_b\}$ with $1 \leq a \leq b \leq n$. (If the permutation is a signed permutation, then we take absolute values.) The elements π_a and π_b are the extremities of the interval.

Transposition Distance Continued

Definition (Transposition $\tau(i, j, k)$)

For $\pi = (\pi_1 \pi_2 \dots, \pi_n)$ in S_n , the transposition $\tau(i, j, k)$ with $1 \leq i < j < k \leq n$ applied to π exchanges the intervals determined respectively by i and $j-1$, and by j and $k-1$, transforming π into

$\pi \circ \tau(i, j, k)$ where $\tau(i, j, k) = (1 \ 2 \ \dots \ i-1 \ j \ j+1 \ \dots \ k-1 \ i \ i+1 \ \dots \ j-1 \ k \ \dots \ n)$.

Example $\pi = (4 \ 1 \ 6 \ 2 \ 5 \ 7 \ 3)$, $\tau(3, 5, 6)$ ($i = 3, j = 5$ and $k = 6$)

$$\pi \circ \tau(3, 5, 6) = (4 \ 1 \ 5 \ 6 \ 2 \ 7 \ 3)$$

Definition (Transposition distance)

Let $td(\pi)$ denote the transposition distance of π . Then after $td(\pi)$ transpositions, we end it up with the identity permutation i .

Proposition Let the number of breakpoints of π be $bp(\pi)$. Then $td(\pi) \geq bp(\pi)/3$.

Proof. We can eliminate at most three break points in one transposition. ■

Better Bounds

- Breakpoints: **ascents** $\pi_{j+1} > \pi_j$ but $\pi_j + 1 < \pi_{j+1}$.
descents $\pi_{j+1} < \pi_j$.
- We use **$des(\pi)$** to denote the **number of descents in π** .

Proposition *For every π in S_n , $td(\pi) \geq \max\{des(\pi), des(\pi^{-1})\}$.*

Proof. It suffices to claim that the number of descents decrease at most for each transposition applies to the permutation.



Continued of Proof

Assume that there are three descents which are converted to ascents. Then, the possible locations are $\pi_{i-1} > \pi_i$ but $\pi_{j-1} > \pi_j$ and $\pi_{k-1} > \pi_k$. Let them be renamed as a, b, c, d, e and f respectively. Then, we have

$a > b, c > d$ and $e > f$. Now, following the composition with $\tau(i, j, k)$, we have the total order as: $a-d \text{ --- } e-b \text{ --- } c-f \text{ ---}$. In order to convert them into ascents, we have “ $a < d, e < b$, and $c < f$ ”. But, this is not possible.

First, if $a < d$ and $e < b$, then $c > d > a > b > e > f$, a contradiction to $c < f$.

Second, if $a < d$ and $c < f$, then $b < a < d < c < f < e$, a contradiction to

$b > e$. Finally, if $e < b$ and $c < f$, then $a > b > e > f > c > d$, a contradiction to $a < d$.



Finding $td(\pi)$ by Breakpoint Graph $G(\pi)$

- A **strip** of π is a maximal interval of π **containing no breakpoints**.

Example $\pi = (4 \text{ } 5 \text{ } 6 \text{ } 3 \text{ } 1 \text{ } 7 \text{ } 8 \text{ } 2) \longrightarrow \pi' = (4 \text{ } 3 \text{ } 1 \text{ } 5 \text{ } 2)$

() π and π' have the same transposition distance.*

Proposition For each permutation π in S_n , $td(\pi) \geq (p(\pi) - c(\pi))/2$.

Notice that $c(\pi)$ is the maximum number of alternating cycles in the decomposition of $G(\pi)$.

Proof. It suffices to claim that a transposition may increase the number of alternating cycles by at most two.

Example of Idea in Proof

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

$G(\pi)$ can be decomposed into 4 alternating cycles: $(2, 3)$, $(6, 5)$, $(4, 6, 7, 5)$, $(0, 2, 1, 3, 4, 1)$.

Apply transposition $\tau(1, 3, 4)$, we have $\pi' = (1 \ 2 \ 3 \ 4 \ 6 \ 5)$. Now, $G(\pi')$ can be decomposed into more alternating cycles: $(0, 1)$, $(1, 2)$, $(2, 3)$, $(3, 4)$, $(5, 6)$ and $(4, 6, 7, 5)$ where $\{4, 6\}$ and $\{5, 7\}$ are blue edges, $\{6, 7\}$ and $\{4, 5\}$ are red edges. Finally, apply $\tau(5, 6, 7)$, we will obtain more alternating cycles.

Reference: V. Bafna and P. A. Pevzner, Sorting by transpositions, SIAM J. Discrete Math., Vol. 11, No. 2, 224 – 240.



Algorithm about Transpositions

Definition

The Γ – graph of a permutation $\pi = (\pi_1 \pi_2 \dots, \pi_n)$ is the digraph $\Gamma(\pi)$ with ordered vertex set $\{\pi_1, \pi_2, \dots, \pi_n\}$ and (π_i, π_j) is an arc of $\Gamma(\pi)$ if

$\pi_i = j$.

Example

$$\pi = (4 \ 1 \ 6 \ 2 \ 5 \ 7 \ 3)$$

The arcs of $\Gamma(\pi)$ are $(1, 4), (2, 1), (3, 6), (4, 2), (5, 5), (6, 7), (7, 3)$. They form three odd directed cycles: $(1, 4, 2), (3, 6, 7), (5)$.

(*) These cycles can also be obtained by finding the cycle

decomposition of $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 1 & 6 & 2 & 5 & 7 & 3 \end{bmatrix}$.

Continued

Proposition

For all π in S_n , $td(\pi) \leq n - c_o(\Gamma(\pi))$ where $c_o(\Gamma(\pi))$ denotes the number of odd cycles in $\Gamma(\pi)$.

Exercise 8. Find an algorithm for finding $td(\pi)$.

Proposition

Let $|LIS(\pi)|$ denote the length of a longest increasing subsequence of the sequence $\langle \pi_1, \pi_2, \dots, \pi_n \rangle$. Then, $td(\pi) \leq n - |LIS(\pi)|$.

Proof. Increasing the length of $LIS(\pi)$ one by one. (See it?)

Example (4 1 6 2 5 7 3) Move 6 to the location between 5 and 7 to increase the length of $LIS(\pi)$, from 4 to 5.

(*) $n - |LIS(\pi)|$ is also denoted by $Ulam(\pi)$, called Ulam's distance.

Coding a Permutation

Definition (left code)

Given a permutation $\pi = (\pi_1 \pi_2 \dots \pi_n)$, the left code of the element π_i of π , denoted by $lc(\pi_i) =_{\text{def}} \# \text{ of } \pi_j > \pi_i \text{ where } 0 \leq j \leq i-1, 1 \leq i \leq n$.

Definition (right code)

Given a permutation $\pi = (\pi_1 \pi_2 \dots \pi_n)$, the right code of the element π_i of π , denoted by $rc(\pi_i) =_{\text{def}} \# \text{ of } \pi_j < \pi_i \text{ where } i+1 \leq j \leq n, 1 \leq i \leq n$.

Example $\pi = (6 \ 3 \ 2 \ 1 \ 4 \ 5)$, $lc(\pi) = (0, 1, 2, 3, 1, 1)$ and

$rc(\pi) = (5, 2, 1, 0, 0, 0)$. If π is the identity permutation, then both the left code and right code are *zero vector*.

Plateau of Sequences

Definition (Plateaus)

Let s be a sequence of non-negative integers. A plateau of s is any maximal length subsequence of contiguous elements in s that have the same value. The number of plateaus in s is denoted by $\rho(s)$.

Example Let $s = (0, 1, 2, 3, 1, 1)$. Then, $\rho(s) = 4$.

Let $s = (5, 2, 1, 0, 0, 0)$. Then, $\rho(s) = 3$.

Proposition

Given a permutation $\pi = (\pi_1 \pi_2 \dots \pi_n)$, the left most plateau of $lc(\pi)$ can be removed by a transposition to the left without creating any new plateaus in the code. (The statement is also true for right code.)

Proof of Proposition

Proof. Let the leftmost plateau be from position i to $j-1$. Then, all entries of $lc(\pi_i)$ before position i is “0”. That is, $\pi_1 \pi_2 \dots \pi_{i-1}$ are in increasing order. So, if $lc(\pi_i) = a$, then $\tau(i, j, i-a)$ removes the plateau from the left. ■

Example $\pi = (6 \ 3 \ 2 \ 1 \ 4 \ 5), lc(\pi) = (0, 1, 2, 3, 1, 1)$

$\pi' = (3 \ 6 \ 2 \ 1 \ 4 \ 5), lc(\pi') = (0, 0, 2, 3, 1, 1)$

$\pi'' = (2 \ 3 \ 6 \ 1 \ 4 \ 5), lc(\pi'') = (0, 0, 0, 3, 1, 1)$

Proposition

Let $\pi = (\pi_1 \ \pi_2 \ \dots, \ \pi_n)$ be a permutation. Then,
 $td(\pi) \leq \min \{\rho(lc(\pi)), \rho(rc(\pi))\}.$

Prefix Transposition

Definition

A prefix transposition is a transposition $\tau(i, j, k)$ such that $i = 1$. That is, we transpose an interval from j to $k-1$ to become the new prefix. The transposition is $\tau(1, j, k) = (j \ j+1 \ \dots \ k-1 \ 1 \ 2 \ \dots \ j-1 \ k \ \dots \ n)$. We use $ptd(\pi)$ to denote the sorting distance of π .

Known Results

1. $ptd(\pi) \leq n - \log_8 n$. (Bounding prefix-transposition distance for strings and permutation.)
2. $ptd(\pi) \leq (3n+1)/4$. (Edit distance and factorizations of even permutations, A. Labarre, LNCS, 5193(2008), 635-646.)

Recommended Project: A study of prefix-transposition.

Insertion of Leading Elements

- Among the study of prefix transpositions, the case $\tau(1, 2, k)$ has more impact. It is known as “*insertion of the leading elements*”.

Theorem

The number of insertions required to sort a permutation π in S_n by head insertions is $n - h$, where h is the largest integer such that the last h elements of π form an increasing substring.

Before giving a proof, let's see an example. (5 4 1 6 3 2) $\longrightarrow$

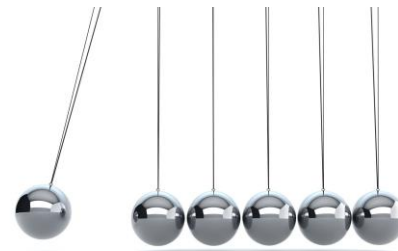
(4 1 6 3 2 5) $\longrightarrow$ (1 6 3 2 4 5) $\longrightarrow$ (6 3 1 2 4 5) $\longrightarrow$

(3 1 2 4 5 6) $\longrightarrow$ (1 2 3 4 5 6) $h = 1$ in this case.

Proof of Theorem

Proof. Since we can insert the leading element to a position to increase the length h of the increasing substring, the *upper bound* can be obtained subsequently. Moreover there are $n - h$ elements to insert, thus we need *at most $n - h$ leading insertions*.

On the other hand, since the increasing substring starts at π_{n-h+1} , the element π_{n-h} has to be moved to somewhere to the right of π_{n-h+1} , so are $\pi_1, \pi_2, \dots, \pi_{n-h-1}$. Hence, *at least $n - h$ insertions* are needed. This concludes the proof. ■



Review of Reversals

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 $\pi^{i,j}$.

Example $(5 \ 4 \ 1 \ 6 \ 3 \ 2) \longrightarrow (1 \ 4 \ 5 \ 6 \ 3 \ 2) \longrightarrow (1 \ 2 \ 3 \ 6 \ 5 \ 4) \longrightarrow (1 \ 2 \ 3 \ 4 \ 5 \ 6)$.

The **prefix reversal** of the segment $\pi_1 \pi_2 \dots \pi_j$, $1 < j$ is the permutation

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

(*) It takes more steps to sort a permutation. For example, it takes 7 prefix reversals to sort $(5 \ 4 \ 1 \ 6 \ 3 \ 2)$. (?)

The Pancake Problem

- The prefix reversal problem is also known as the pancake problem.

Example (1 7 5 3 6 4 2 8)
 (5 7 1 3 6 4 2 8)
 (6 3 1 7 5 4 2 8)
 (1 3 6 7 5 4 2 8)
 (4 5 7 6 3 1 2 8)
 (6 7 5 4 3 1 2 8)
 (7 6 5 4 3 1 2 8)
 (2 1 3 4 5 6 7 8)
 (1 2 3 4 5 6 7 8)



(*) See the pancakes?



Comments on Reversals

- In general, if $rd(\pi)$ denotes the reversal sorting number and $prd(\pi)$ denotes the prefix reversal number, then $rd(\pi) \leq prd(\pi)$.
- *In 1979, William H. Gates and Christos H. Papadimitriou wrote a paper about bounds for sorting by prefix reversals, which was published in Discrete Math. 27, 45 – 57.*
- **Results**

Let π be a permutation in S_n , $prd(n)$ denotes the maximum value of $prd(\pi)$ for all permutations π in S_n . Then,

1. $prd(n) \leq (5n+5)/3$ and
2. $prd(n) \geq 17n/16$ for n a multiple of 16.

Project Recommended: A study of prefix reversals problem.

Permutations with Large Reversal distance

- It is easy to see that if π is not an identity permutation, then the reversal distance (to identity) is at least one.
- Clearly, $(n \ n-1 \ n-2 \ \dots \ 3 \ 2 \ 1)$ has the shortest distance 1. Now, which permutations have the largest distance to identity?

- Gollan's Permutations

$\pi_1 = (1 \ 3 \ 5 \ \dots \ n-1 \ n \ n-2 \ \dots \ 6 \ 4 \ 2)$, n is even; and

$\pi_2 = (1 \ 3 \ 5 \ \dots \ n \ n-1 \ n-2 \ \dots \ 6 \ 4 \ 2)$, n is odd.

(*) $rd(\pi_1) = rd(\pi_2) = n - 2$. (?)

Problem 1. Find all permutations π such that $rd(\pi) = n - 2$.
2. Find $prd(n)$. (Project recommended.)

Edit Distance

Problem

Let $ed(\pi)$ be the edit distance such that we can use transpositions and reversals for editing permutation π . Estimate $ed(\pi)$.

(*) We can also edit the permutation by using block interchanges, that is we interchange the blocks with the same length.

Definition ($\beta(l, j, k, l)$)

$(\pi_1 \pi_2 \dots \pi_{i-1} \pi_i \pi_{i+1} \dots, \pi_{j-1} \pi_j \pi_{j+1} \dots \pi_{k-1} \pi_k \pi_{k+1} \dots \pi_{l-1} \pi_l \pi_{l+1} \dots \pi_n)$



$(\pi_1 \pi_2 \dots \pi_{i-1} \pi_k \pi_{k+1} \dots, \pi_{l-1} \pi_j \pi_{j+1} \dots \pi_{k-1} \pi_i \pi_{i+1} \dots \pi_{j-1} \pi_l \pi_{l+1} \dots \pi_n)$

$\beta(l, j, k, l) = (1 \ 2 \ \dots \ i-1 \ k \ k+1 \ \dots \ l-1 \ j \ \dots \ k-1 \ i \ i+1 \ \dots \ j-1 \ l \ l+1 \ \dots \ n)$

Element Exchange Distance

- This is a special case of block interchange distance, namely, we consider all blocks of length 1.

Proposition

Let $\text{exc}(\pi)$ denote the element exchange distance of π in S_n . Then, $\text{exc}(\pi) \leq n - c(\Gamma(\pi))$ where $c(\Gamma(\pi))$ is the number of cycles in $\Gamma(\pi)$.

Example $\pi = (4 \text{ } 1 \text{ } 6 \text{ } 2 \text{ } 5 \text{ } 7 \text{ } 3)$ Three cycles are; $(1, 4, 2)$, $(3, 6, 7)$, (5) .

$(4 \text{ } 2 \text{ } 6 \text{ } 1 \text{ } 5 \text{ } 7 \text{ } 3)$

$(1 \text{ } 2 \text{ } 6 \text{ } 4 \text{ } 5 \text{ } 7 \text{ } 3)$

$(1 \text{ } 2 \text{ } 7 \text{ } 4 \text{ } 5 \text{ } 6 \text{ } 3)$

$(1 \text{ } 2 \text{ } 3 \text{ } 4 \text{ } 5 \text{ } 6 \text{ } 7)$

Proof of the Proposition

Idea

It takes at most $k - 1$ exchanges to change a cycle of order k into ascending order, see the example in previous page. Now, if there are t cycles of order $k_1, k_2, \dots, k_t$ respectively, then it takes at most $(k_1 - 1) + (k_2 - 1) + \dots + (k_t - 1) = n - t$ steps to sort the permutation. This concludes the proof. ■

