Computing maximal reversal distance of signed permutations

Joint Mathematics Meetings 2018

Tanner Rosenberg
Northern Arizona University
January 10, 2018

Joint work with Dana C. Ernst

Biology Big Picture

- DNA: Double helix of nucleotides, complementary pairs A–T, G–C.
- Gene: Sequence of nucleotides, codes a specific protein.
- Chromosome: Ordered set of genes.
- Genome: Set of ordered sequences of conserved blocks of genes, each gene having orientation given by location on DNA.
- Mutations: Deletions, translocations, duplications, fusions, fissions, ***reversals***.

Biology to Mathematics

- The edit distance between two genomes is the minimum number of mutations required to transform one into another, approximates evolutionary distance.
    - mouse $\xrightarrow{251}$ human (149 reversals)
    - cabbage $\xrightarrow{3}$ turnip (all reversals)
- Comparing two similar sequences of genes appearing along a chromosome in two species yields two signed permutations.
- Reversal distance between two signed permutations is minimum number of reversals needed to transform one into the other. Good estimate of evolutionary distance.
- Reinterpret as sorting problem. Reversal distance for signed permutation π is minimum number of reversals needed to sort π to identity.

Sorting by Reversals

$$\pi = -5 \quad +1 \quad +2 \quad -4 \quad -3 \quad +6 \quad +7$$

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

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

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

$$\text{id} = +1 \quad +2 \quad +3 \quad +4 \quad +5 \quad +6 \quad +7$$

$$d_{rev}(\pi) \leq 3$$

Two Problems of Interest

Definition

The maximal reversal distance for signed permutations of length n is defined to be the greatest reversal distance among all signed permutations of length n.

$$d_{max}(n) := \max\{d_{rev}(\pi) \mid \pi \in S_n^{\pm}\}$$

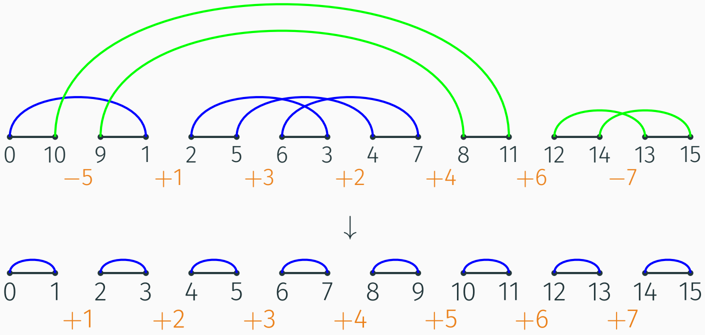
Problem 1

For given n, what is $d_{max}(n)$?

Problem 2

How many signed permutations in $S_n^{\pm}$ have maximal reversal distance?

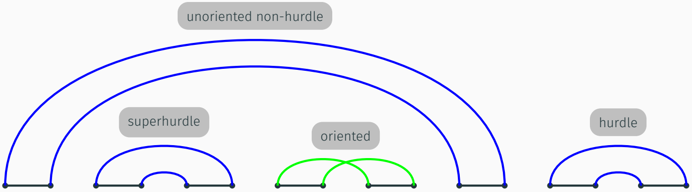
Breakpoint Graphs



Cycles Types

- Oriented = Good
- Unoriented = Bad

Hurdles, Superhurdles, & Fortresses



Cycle Component Types

- **Hurdle:** A “bad” collection of overlapping cycles that either covers all other bad collections or covers no bad collections.
- **Superhurdle:** A hurdle that if removed creates a new hurdle.
- **Fortress:** A permutation with an odd number of hurdles where every hurdle is a superhurdle.

Maximal Reversal Distance

Theorem (Hannenhalli & Pevzner 1999)
The reversal distance for a signed permutation π of length n is

$$d_{rev}(\pi) = n + 1 - c(\pi) + h(\pi) + f(\pi),$$

where $c(\pi)$ is the number of cycles, $h(\pi)$ is the number of hurdles, and $f(\pi) = 1$ if π is a fortress and 0 otherwise.

Claim (Entry A131209 on OEIS)
The maximal reversal distance for signed permutations of length n is given by

$$d_{max}(n) = \begin{cases} n, & \text{if } n \in \{1, 3\} \\ n + 1, & \text{otherwise} \end{cases}$$

Number of Maximal Permutations

Python Code

n	# attaining $d_{max}(n)$
1	1
2	1
3	25
4	8
5	3
6	180
7	64

```
def all_permutations(n,l):
    gens = {}
    for i in range(n):
        for j in range(i,n):
            gens['s'+str(i+1)+str(j+1)]=reversal(i,j)

    spins = {'':range(1,n+1)}
    while len(spins) < factorial(n)*(2**n):
        for y in spins.keys():
            for x in gens:
                a = gens[x](spins[y])
                if a not in spins.values():
                    spins[str(x)+str(y)] = a

    for n in spins:
        if len(n)/3 >= l:
            print n, spins[n]

```

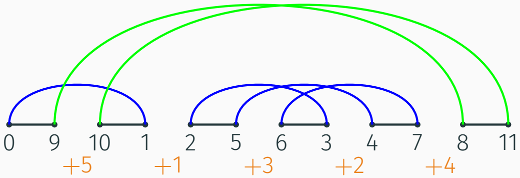
Conjecture

The number of signed permutations that attain the maximal reversal distance depends on the parity of n.

Observations

Examples of Maximal Permutations

$+2 \ +1$
$+2 \ -3 \ -1$
$+2 \ +1 \ +4 \ +3$
$+5 \ +1 \ +3 \ +2 \ +4$



Conjecture

All maximal permutations are positive (for $n \neq 3$)

Next Steps

- Find number of maximal permutations of length n.
- Examine structure of maximal permutations.
- What structures in breakpoint graph are permissible?