2. The Signed Reversal Distance

Literature:

Hannenhalli, S., & Pevzner, P. A. (1999). *Transforming cabbage into turnip: Polynomial algorithm for sorting signed permutations by reversals.* Journal of the ACM, Vol. 46(1), pp. 1–27.

Bergeron, A., Mixtacki, J., & Stoye, J. (2005). *Chapter 10: The Inversion Distance Problem.* In: Gascuel, O. (ed.) Mathematics of Evolution and Phylogeny. Oxford University Press, pp. 262-290.

Definition 2. *A* reversal $\rho(i,j)$ *in a signed permutation* σ *reverts the order and sign of all elements of the interval* $\sigma[i,j]$.

Problem 1 (Signed Reversal Distance)**.** *Given two signed permutations* π *and* σ*, find* $srd(\pi,\sigma)$*, the minimum number of reversals needed to transform* π *into* σ*.*

In practice, a permutation is always sorted towards the identity $\mathbf{id} = (0 \cdots n+1)$. Therefore, the notation $srd(\pi) := srd(\pi, \mathbf{id})$ is used subsequently.

Example 2. *Let's sort the signed permutation* $\pi^2 = (0\ -3\ -4\ 1\ -5\ -2\ 6)$*:*

$$\pi^2 \circ \rho(5,6) = (0\ -3\ -4\ 1\ 2\ 5\ 6)$$
$$\pi^2 \circ \rho(5,6) \circ \rho(3,5) = (0\ -3\ -2\ -1\ 4\ 5\ 6)$$
$$\pi^2 \circ \rho(5,6) \circ \rho(3,5) \circ \rho(2,4) = (0\ 1\ 2\ 3\ 4\ 5\ 6)$$

The reversal sequence is optimal, i.e., a solution to Problem 1 (a proof will be discussed in the next chapter), thus $srd(\pi^2) = 3$.

Definition 3. *In a signed permutation, a pair of consecutive elements* $i \cdot (i+1)$ *or* $-(i+1) \cdot -i$ *is a* (conserved) adjacency *(ADJ) and otherwise a* breakpoint *(BP).*

Example 2 (continued)**.** *Permutation* π^2 *has 6 breakpoints and 0 adjacencies.*

Definition 4 (Breakpoint Graph)**.** *The* breakpoint graph *of a signed permutation* π *is the graph* $BG(\pi) = (V,E)$*, whose vertex set* V *contains, for each element* $1 \le g \le n$*, two vertices* g^t *and* g^h *called the* tail *and the* head *of element* g*, plus two vertices* 0^h *and* $(n+1)^t$*. The edge set* E *is the union of two perfect matchings* R *and* D *of* V*:*

- *"reality edges"* R *contains edge from* π_i^h *if* π_i *is non-negative, and from* π_i^t *otherwise, to* π_i^t *if* π_{i+1} *is non-negative, and to* π_{i+1}^h *otherwise, for* $0 \le i \le n$*.*

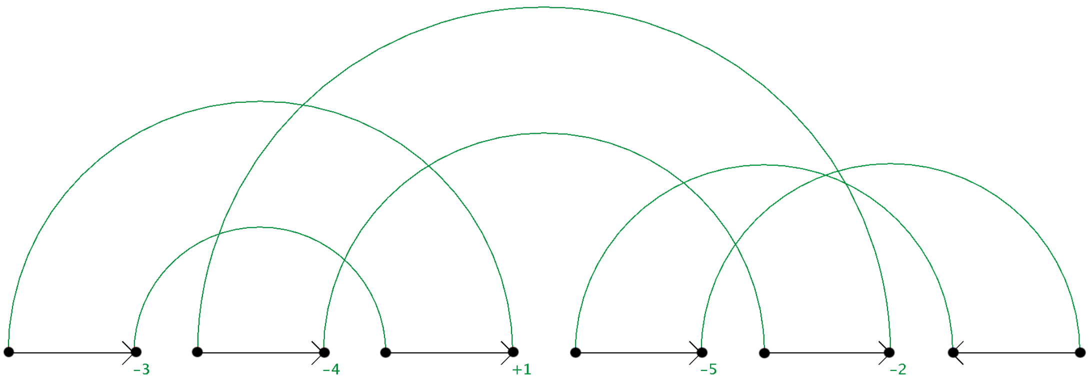


Figure 2.1.: Breakpoint graph of permutation π^2 with reality edges (black arrows) and desire edges (green). Source: Modified figure drawn by program `InversionVisualisation`.

- *"desire edges"* $D := \{\{g_h, (g+1)_t\} \mid 0 \leq g \leq n\}$

Example 2 (continued)**.** *$BG(\pi^2)$ is shown in Figure 2.1.*

Definition 5 (Orientation of Cycles)**.** *Two reality edges in the same cycle are* convergent *if, traversing the cycle, both edges are entered from the right or both from the left side, otherwise they are* divergent*. A desire edge is called* oriented *if its two incident reality edges are divergent, otherwise the edge is* unoriented*. A non-trivial cycle is called* oriented *if it contains at least one oriented (desire) edge.*

Example 3. *The breakpoint graph of $\pi^3 = (0\ -2\ -3\ 1\ 4\ 6\ 5\ 7\ 8)$ (shown in Figure 2.2) has two unoriented and one oriented cycle.*

What happens when we apply a reversal on the same cycle?

- *Type I*: divergent edges $\rightarrow$ breaks the cycle $\Delta c = +1$.
- *Type II*: convergent edges $\rightarrow$ $\Delta c = 0$, but may change cycle orientation.

What about different cycles?

- *Type III*: Merges two cycles $\rightarrow$ $\Delta c = -1$, but may change cycle orientation.

Lemma 1. *A reversal changes the number of cycles of the BP graph at most by* 1*.*

The breakpoint graph $BP(\mathbf{id})$ of an identity permutation **id** corresponding to n genes has $n+1$ cycles. Together with Lemma 1 a lower bound on $srd(\pi)$ can be derived:

$$srd(\pi) \geq n + 1 - c\,,$$

where c is the number of cycles in $BG(\pi)$.

This bound is usually *tight*, that is, most of the times it is exactly the reversal distance. It is not tight whenever it is not possible to increase the number of cycles in $BG(.)$ with a reversal.

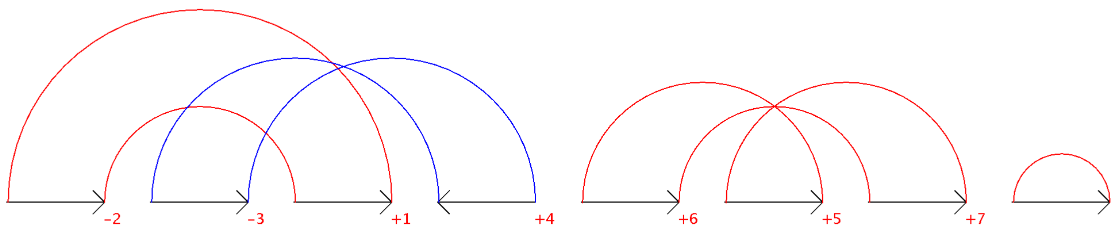


Figure 2.2.: Breakpoint graph of permutation π^3 with oriented (blue) and unoriented (red) desire edges. Source: Figure drawn by Program `InversionVisualisation`.

2.1. Of hurdles, super-hurdles and fortresses

The breakpoint graph of permutation π^3 has four cycles, thus $srd(\pi^3) \geq 7 + 1 - 4$. But π^3 cannot be sorted in less than 5 reversals. Why? Because of *unoriented components*!

Definition 6 (Component)**.** *A* component *is an interval,* $(i^h ... (i + j)^t)$ *for* $i, i + j \geq 0$ *or* $((i + j)^t ... i^h)$ *for* $i + j, i < 0$*, whose set of unsigned elements is* $\{i, ..., i + j\}$*, but not the union of smaller such intervals.*

Example 3 (continued)**.** *Permutation* π^3 *has three components:* $[0 - 2 - 31(4]65[7)8]$

Two cycles are called *interleaving* if they have crossing edges. We make the following observation:

Observation 1. *Components correspond to maximal subsets of interleaving cycles.*

Components have an orientation, according to the orientation of their constituting cycles:

Definition 7. *A component is* trivial *if it is of the form* $(i, i + 1)$ *or* $(-(i + 1), -i)$ *and otherwise* non-trivial*. A component is* unoriented *if it is non-trivial and all its elements have the same sign, otherwise it is* oriented*.*

Example 3 (continued)**.** *Permutation* π^3 *has 1 unoriented and 2 oriented (one of which is trivial) components.*

Observation 2. *In the breakpoint graph, a component is oriented if it is either an unoriented cycle of size 2, or it contains at least one oriented cycle, otherwise it is unoriented.*

A *hurdle* is an unoriented component that does not separate two other unoriented components. Removing a hurdle requires an additional reversal. But some hurdles cause, when removed, the creation of a new hurdle. Such hurdles are called *super-hurdles*. Super-hurdles can be removed in most cases, except for a *fortress*, which is a permutation that has has an odd number of hurdles, and all are super-hurdles.

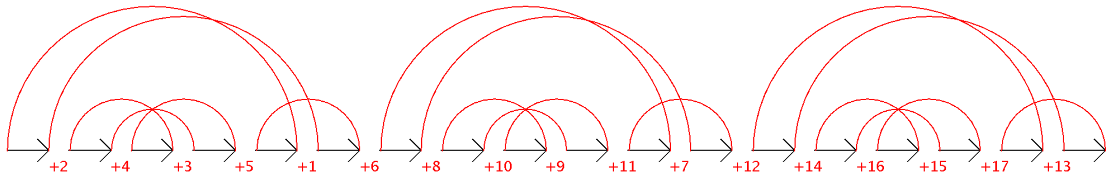


Figure 2.3.: A fortress of a permutation with 17 elements.

Example 4. *Figure 2.3 shows a fortress* (2 4 3 5 1 6 8 10 9 11 7 12 14 16 15 17 13).

Theorem 1 (Hannenhalli-Pevzner Duality Theorem)**.** *The reversal distance for a signed permutation π of n elements is*

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

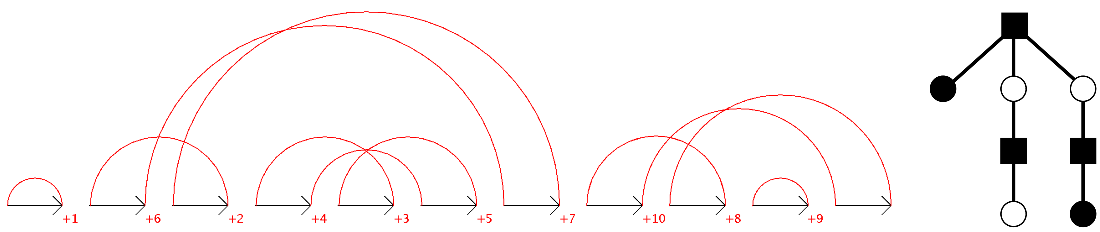
where c is the number of cycles, h the number of hurdles, and $f = 1$ if π has a fortress, and $f = 0$ otherwise.

2.2. Computing the reversal distance without hurdles and fortresses in $O(n)$ time

Observation 3. *Two components are either disjoint, nested, or chained.*

Example 5. *The permutation $\pi^4 = (0\ 1\ 6\ 2\ 4\ 3\ 5\ 7\ 10\ 8\ 9\ 11)$ has three chained components, two of which contain each another nested component. Figure 2.4 (a) shows $BG(\pi^4)$.*

Definition 8 (Component Tree)**.** *Given a permutation π and its components, the* component tree T_π *is constructed as follows:*



(a) (b)

Figure 2.4.: (a) Breakpoint graph of and (b) component tree of permutation π^4 with nodes corresponding to oriented and unoriented components colored black and white, respectively.

1. *Each component is represented by a round node.*
2. *Each maximal chain is represented by a square node, containing its children.*
3. *A square node is the child of the smallest component that contains the chain.*

Example 5 (continued)**.** *Figure 2.4 (b) shows the component tree of* π^4*.*

A reversal has the following effect on components:

- Within the same component C:
  - If C is oriented, at least one reversal exists which does not create new unoriented components and which splits a cycle in two.
  - If C is unoriented, then no reversal can split one of its cycles, nor create a new component. However, the inversion of a single element will orient C without changing the number of cycles.
- An inversion that has its two endpoints in different components A and B destroys, or orients, all components on the path from A to B in T_P, without creating new unoriented components.

The additional inversion required to orient unoriented components can be computed by means of the following tree cover:

Definition 9 (Component Tree Cover)**.** *A* cover C *of* T_π *is a collection of paths joining all the unoriented components of* π*, such that each terminal node of a path belongs to a unique path. A path is* short *if it contains only one component, otherwise it is* long*.*

The cost $t(C)$ *of a cover* C *is the sum of costs of all paths, whereby a short path has cost* 1 *and a long path has cost* 2*.*

Example 6. π^3 *has one cover with cost* 3*.*

This leads to the following theorem:

Theorem 2. *The reversal distance of a signed permutation* π *of* n *elements is*

$$srd(\pi) = n + 1 - c + t\,,$$

where c *is the number of cycles and* t *the cost of an optimal cover of the component tree* T_π*.*

It is not necessary to compute an optimal cover in order to assess its cost:

Theorem 3. *Let* T' *be the smallest unrooted subtree of* T_π *that contains all unoriented components of* π*. Let* l *be the number of leaves of* T'*, then*

- *if* l *is either (i) even or (ii) odd and one of the leaves is on a short branch, then* $t = l$*,*
- *otherwise* $t = l + 1$*.*

The reversal distance $srd(\pi)$ can be computed in $O(n)$ time, because both, the costs of an optimal cover of the component tree and the number of cycles of the breakpoint graph, can by computed in $O(n)$ time.