

4. The Algebraic Theory of Genome Rearrangements

Literature:

Meidanis J., & Dias Z. (2000). *An Alternative Algebraic Formalism for Genome Rearrangements*. In: Sankoff D., Nadeau J.H. (eds) *Comparative Genomics*. Computational Biology, Springer, Vol. 1, pp. 213-223.

Dias, Z., & Meidanis, J. (2001). *Genome Rearrangements Distance by Fusion, Fission, and Transposition is Easy*. Proceedings of SPIRE 2001. pp. 250-253.

Note that although this chapter is confined to the study of unsigned permutations, the discussed theory can be extended to signed permutations. For further details, see Meidanis and Diaz (2000).

4.1. Composition of permutations

Have you ever wondered what the “ \circ ” operator does? What does $\pi \circ \rho$ mean?

The symbol “ \circ ” denotes the composition (multiplication) operator of permutations which is, just as the composition of functions, applied from right to left, i.e. in composition $\pi \circ \sigma$, we first apply sigma (to the identity permutation $(1 \cdots n)$), and then we apply π , resulting in the permutation $(\pi_{\sigma_1} \cdots \pi_{\sigma_n})$.

Example 8. Two permutations $\pi = (2\ 1\ 4\ 3)$, $\sigma = (1\ 4\ 2\ 3)$;

$$\pi \circ \sigma = (\pi_1\ \pi_4\ \pi_2\ \pi_3) = (2\ 3\ 1\ 4)$$

Then a reversal $\rho(2, 3)$ corresponds to permutation $(1\ 3\ 2\ 4)$ and $\pi \circ \rho(2, 3) = (2\ 4\ 1\ 3)$. The set of all permutations of n elements, denoted Π_n , under composition operator \circ forms a group (Π_n, \circ) . Its neutral element is $(1 \cdots n)$ and the inverse of a permutation $(\pi_1 \pi_2 \cdots \pi_n)$ is obtained by exchanging positions and elements in π , i.e. $\pi_{\pi_i}^{-1} = i$.

Note that in general $\pi \circ \sigma \neq \sigma \circ \pi$.

4.2. Cycles of a permutation

A permutation can also be represented by a *composition of one or more cycles*:

Definition 15. A cycle of a permutation π , denoted by $C = (i_1, \dots, i_n)$ is a set of elements such that for $1 \leq j \leq n - 1$ holds that $\pi_{i_j} = i_{j+1}$ and $\pi_{i_n} = i_1$.

Important: we distinguish between a permutation and its cycle decomposition by delimiting elements in the latter by “,”.

Theorem 6. Every permutation has a unique disjoint cycle decomposition¹.

Example 9. The disjoint cycle composition of σ is:

$$\begin{array}{cccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \pi^1 = & (2 & 1 & 4 & 3 & 5 & 8 & 6 & 7) = (1, 2)(3, 4)(5)(6, 7, 8) \end{array}$$

Note:

- $(1\ 2\ 3) \neq (1, 2, 3)$ but $(1, 2, 3) = (2, 3, 1) = (3, 1, 2)$
- the inverse of a cycle is the cycle in which elements have reversed order, the neutral element is $(1) \cdots (n)$
- cycles with a single element are usually omitted from the cycle representation e.g. $\pi^1 = (1, 2)(3, 4)(6, 7, 8)$

Permutations can also be decomposed into multiple (non-disjoint) cycles. Alternative representations of the cycle $\pi^2 = (1, 2, 3, 4, 5)$ include

$$\begin{aligned} & (1, 2, 3)(3, 4, 5) \\ & (1, 2)(2, 3)(3, 4)(4, 5) \\ & (1, 5)(1, 4)(1, 3)(1, 2). \end{aligned}$$

The order of non-disjoint cycles is no longer arbitrary e.g. $(1, 2)(2, 3) \neq (2, 3)(1, 2)$.

Algorithm 1 describes the procedure of obtaining the disjoint cycle decomposition of sequence of cycles.

Example 10. $\pi^3 = (1, 5, 3, 2)$, $\pi^4 = (1, 4)(3, 5)$:

$$\pi^3 \pi^4 = (1, 5, 3, 2)(1, 4)(3, 5) = (1, 4, 5, 2)$$

4.3. Algebraic theory meets genome rearrangements (or: going beyond reversals)

Definition 16. A genome is an (unsigned) permutation π in disjoint cycle notation where each cycle corresponds to a circular chromosome.

¹up to the order of cycles and the rotation of elements within cycles

Algorithm 1 Construction of a disjoint cycle decomposition from a sequence of cycles

Input: Collection of cycles $\mathcal{C} = C^1, C^2, \dots$ **Output:** Collection of disjoint cycles $\mathcal{C}' = C'^1, C'^2, \dots$

```
1:  $i_1 = 1$ 
2: let  $\mathcal{C}'$  be an empty collection of cycles
3: add new cycle  $C' = (i_1, )$  to  $\mathcal{C}'$ .
4: for  $j = 2..n$  do
5:    $i_j = i_{j-1}$ 
6:   traverse from right to left through each cycle  $C$  of  $\mathcal{C}$ , update  $i_j \leftarrow C_{k+1}$  if and
   only if  $C_k = i_j$ 
7:   if  $i_j = C'_1$  then
8:     let  $i_j$  be the next higher number not contained in any cycle of  $\mathcal{C}'$ 
9:     associate variable  $C'$  with a new cycle  $(i_j, )$ 
10:    add  $C'$  to  $\mathcal{C}'$ 
11:   else
12:     append  $i_j$  to cycle  $C'$ 
13:   end if
14: end for
15: return  $\mathcal{C}'$ 
```

The following discussion is restricted to circular chromosomes, but the algebraic theory can be extended to linear chromosomes. For further details, see [Feijão, P., & Meidanis, J. \(2013\). *Extending the Algebraic Formalism for Genome Rearrangements to Include Linear Chromosomes*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 10\(4\), pp. 819–831.](#)

Example 11. $\pi^5 = (1, 4, 5)(2, 3)$ is a genome with two chromosomes.

A rearrangement in a genome π can be modeled by a product with a permutation ρ .

Example 12. $\rho = (2, 4, 5)$

$$\rho\pi^2 = (2, 4, 5)(1, 2, 3, 4, 5) = (1, 4, 2, 3, 5)$$

Permutation ρ is a transposition of the blocks $[2, 3]$ and $[4]$.

Applying a 2-cycle $\rho = (a, b)$ to a genome has the following effect:

- Fission: if a and b are in the same cycle, this cycle is split in two, separating a and b
- Fusion: if a and b are in different cycles, the cycles are joined into one

Example 13. *Fission:* $(4, 5) \pi^5 = (1, 5)(2, 3)(4)$; *fusion:* $(2, 5) \pi^5 = (1, 4, 2, 3, 5)$

4.4. The rearrangement power of permutations

Definition 17. The norm of a permutation π , denoted by $\|\pi\|$, is the minimum number of 2-cycles needed to decompose π .

The norm of a permutation can be seen as a measure of its rearrangement power.

Observation 6. The norm of a cycle with k elements is $k - 1$.

Observation 7. The norm of a permutation π with n elements and c cycles is $\|\pi\| = n - c$.

Example 14. $\|\pi^5\| = 5 - 2 = \|(1, 4)(4, 5), (2, 3)\| = 3$

Problem 3 (Sorting by Algebraic Rearrangements). Given genomes π and σ , find permutations $\rho_1, \rho_2, \dots, \rho_k$ that transform π into σ such that the algebraic distance $ad(\pi, \sigma) := \sum_{i=1}^k \|\rho_i\|$ is minimum.

How to compute $ad(\pi, \sigma)$?

$$\begin{aligned}
 \rho_k \cdots \rho_2 \rho_1 \pi &= \sigma && \Leftrightarrow \\
 \rho_k \cdots \rho_2 \rho_1 &= \sigma \pi^{-1} && \Rightarrow \\
 \|\rho_k \cdots \rho_2 \rho_1\| &= \|\sigma \pi^{-1}\| \\
 \sum_{i=1}^l \|\rho_i\| &\geq \|\sigma \pi^{-1}\| && \text{(norm property)} \\
 ad(\pi, \sigma) &\geq \|\sigma \pi^{-1}\|
 \end{aligned}$$

Observation 8. We can obtain the rearrangement operations by decomposing $\sigma \pi^{-1}$ and $\|\sigma \pi^{-1}\|$ serves as lower bound of $ad(\pi, \sigma)$.

Thus, a solution to Problem 3 is to find a minimal 2-cycle decomposition $\rho_k \cdots \rho_2 \rho_1$ of $\sigma \pi^{-1}$. Then, by definition, $ad(\pi, \sigma) = \|\sigma \pi^{-1}\| = k - 1$. Each 2-cycle corresponds to a fusion, fission.

A fission followed by a fusion can model a transposition:

Example 15. $\pi^6 = (1, 3, 4, 2, 5, 6)$:

$$(2, 4)(4, 6) \pi^6 = (2, 4, 6) \pi^6 \quad // \text{ transposition of interval } [2..5]$$

Observation 9. If elements a, b and c are in the same cycle in π and appear in this very order, then $\rho = (abc)$ is a transposition in π .

Therefore, the algebraic distance $ad(\pi, \sigma) = \|\sigma \pi^{-1}\|$ is also known as *fusion, fission and transposition* (FFT) distance. But, in terms of costs, a transposition will still cost 1 fusion and 1 fission.