

Algorithms for Genome Rearrangements

Pedro Feijão

Lecture 10 - Algebraic Theory for Genome Rearrangements

Summer 2014

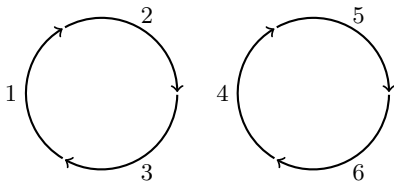
`pfeijao@cebitec.uni-bielefeld.de`

Introduction

- The **Algebraic theory** for genome rearrangements was introduced in 2000, by Meidanis and Dias.
- Its motivation is to use permutation group theory for solving rearrangement problems.
- It has been used to solve several rearrangement problems.

Definitions

- A **permutation** is a bijection in a set E . When $E = \{1, \dots, n\}$, we use the notation S_n .
- Permutations are composed by **k-cycles**, where k is the number of elements in the cycle.
- For instance, $\pi = (1\ 2\ 3)(4\ 5\ 6)$ has two **3-cycles**:



Definitions

- The same cycle can be represented in different ways, by rotating its elements:

$$(1\ 2\ 3\ 4) = (2\ 3\ 4\ 1) = (3\ 4\ 1\ 2) = (4\ 1\ 2\ 3)$$

- A **1-cycle** represents a fixed element, and can be omitted in the notation of a permutation.

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

Definitions

- The **product** of α , β is denoted by $\alpha\beta$, defined as $\alpha\beta(x) = \alpha(\beta(x))$ for $x \in E$.

$$(1\ 2\ 3\ 4)(4\ 1\ 3) = (1\ 4\ 2\ 3)$$

- The **identity permutation** i is the permutation where every element is fixed.
- Every permutation π has an **inverse** π^{-1} such that $\pi\pi^{-1} = \pi^{-1}\pi = i$.
- The inverse of a cycle is obtained by reversing its elements.

$$\pi = (1\ 2\ 3\ 4) \Rightarrow \pi^{-1} = (4\ 3\ 2\ 1)$$

Definitions

- A ***k*-cycle decomposition** of a permutation α is a representation of α as a product of k -cycles, not necessarily disjoint.

$$\begin{aligned}\alpha &= (1\ 2\ 3\ 4\ 5) = (1\ 2)(2\ 3)(3\ 4)(4\ 5) = (1\ 5)(1\ 4)(1\ 3)(1\ 2) \\ &= (1\ 2\ 3)(3\ 4\ 5)\end{aligned}$$

- All permutations have a 2-cycle decomposition.

Definitions

- The **norm** of a permutation α , denoted by $\|\alpha\|$, is the minimum number of 2-cycles needed to decompose α . It can be seen as a measure of its *rearrangement power*.

Example: $\pi = (1\ 2\ 3\ 4) = (1\ 2)(2\ 3)(3\ 4) \Rightarrow \|\pi\| = 3$

The norm of a k -cycle is $k - 1$.

- Alternative equation: $\|\alpha\| = n - c$, where n is the number of elements and c the number of cycles of α .

Examples:

$$\pi = (1\ 2\ 3\ 4) \Rightarrow \|\pi\| = n - c = 4 - 1 = 3$$

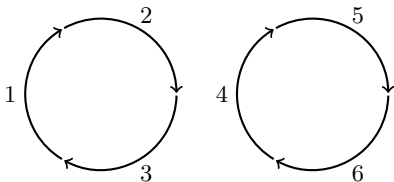
$$\pi' = (1\ 2\ 3\ 4)(5)(6\ 7) \Rightarrow \|\pi'\| = n - c = 7 - 3 = 4$$

Modeling Genomes

An (unsigned) permutation π models a genome where each cycle corresponds to a circular chromosome.

Example:

$\pi = (1\ 2\ 3)(4\ 5\ 6)$ models a genome with 2 circular chromosomes:



Modeling Rearrangements

- A **rearrangement** in a genome π can be modeled by a product with a permutation ρ . For instance:

Consider the permutation $\pi = (1\ 2\ 3\ 4\ 5)$. Applying $\rho = (2\ 4\ 5)$, we have

$$\rho\pi = (2\ 4\ 5)(1\ 2\ 3\ 4\ 5) = (1\ \underline{4}\ \underline{2}\ 3\ 5)$$

In this case, ρ is a *transposition* of the blocks $[2, 3]$ and $[4]$.

- The **weight** of a rearrangement operation ρ is $\|\rho\|$.

Modeling Rearrangements

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

- If a and b are in the same cycle, this cycle is split in two, separating a and b . **(Fission)**
- If a and b are in different cycles, the cycles are joined in one. **(Fusion)**

Examples:

$$\pi = (1\ 2\ 3\ 4\ 5) \text{ and } \rho = (2\ 4) \Rightarrow \rho\pi = (1\ 4\ 5)(2\ 3)$$

$$\pi = (1\ 2\ 3)(4\ 5) \text{ and } \rho = (2\ 4) \Rightarrow \rho\pi = (1\ 4\ 5\ 2\ 3)$$

Genome Rearrangement Problem

The **Algebraic Rearrangement Problem** can be defined as:
Given genomes π and σ , find permutations $\rho_1, \rho_2, \dots, \rho_k$ that **minimally** transform π into σ .

Formally:

- $\rho_k \dots \rho_2 \rho_1 \pi = \sigma$
- The **algebraic distance**, defined as $d(\pi, \sigma) = \sum_{i=1}^k \|\rho_i\|$, is minimum.

Solving the Rearrangement Problem

Since

$$\rho_k \dots \rho_2 \rho_1 \pi = \sigma$$

we arrive at

$$\rho_k \dots \rho_2 \rho_1 = \sigma \pi^{-1} \tag{1}$$

$$\|\rho_k \dots \rho_2 \rho_1\| = \|\sigma \pi^{-1}\|$$

$$\sum_{i=1}^n \|\rho_i\| \geq \|\sigma \pi^{-1}\| \quad (\text{norm property})$$

$$d(\pi, \sigma) \geq \|\sigma \pi^{-1}\|$$

- Eq. (1) shows that rearrangement operations are obtained by decomposing $\sigma \pi^{-1}$.
- Eq. (2) shows that $\|\sigma \pi^{-1}\|$ is a lower bound to the distance.

Sorting by Fusions and Fissions

- The simplest way to decompose $\sigma\pi^{-1}$ is choosing $\rho_k \dots \rho_2\rho_1$ as the 2-cycles in a minimal decomposition of $\sigma\pi^{-1}$.
- Then, each 2-cycle is a fusion or fission, and by definition we have $d(\pi, \sigma) = \|\sigma\pi^{-1}\|$.

Example:

$$\pi = (1\ 3\ 2\ 5\ 4\ 6) \text{ and } \sigma = (1\ 2\ 3\ 4\ 5\ 6)$$

$$\sigma\pi^{-1} = (2\ 4\ 6\ 5\ 3) = (2\ 4)(4\ 6)(6\ 5)(5\ 3) \text{ and } d(\pi, \sigma) = \|\sigma\pi^{-1}\| = 4$$

π can be transformed into σ with 4 fusion/fission operations.

Example

$$\pi = (1\ 3\ 2\ 5\ 4\ 6) \text{ and } \sigma = (1\ 2\ 3\ 4\ 5\ 6)$$
$$\sigma\pi^{-1} = (2\ 4\ 6\ 5\ 3) = (2\ 4)(4\ 6)(6\ 5)(5\ 3)$$

Applying each 2-cycle gives the complete scenario:

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

Adding Transpositions

Repeating the same example,

$$\pi = (1\ 3\ 2\ 5\ 4\ 6) \text{ and } \sigma = (1\ 2\ 3\ 4\ 5\ 6)$$

we can find a different decomposition

$$\sigma\pi^{-1} = (2\ 4\ 6\ 5\ 3) = (5\ 3\ 2\ 4\ 6) = (5\ 3)(3\ 2)(2\ 4)(4\ 6)$$

Applying the first two 2-cycles $(2\ 4)$ and $(4\ 6)$ separately results in a fission followed by a fusion, but the combined result is of a **transposition**.

The same happens with $(5\ 3)$ and $(3\ 2)$.

Example

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

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

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

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

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

or using the transpositions: $(2\ 4)(4\ 6) = (1\ 4\ 6)$ and $(5\ 3)(3\ 2) = (5\ 3\ 2)$:

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

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

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

The distance is still $d(\pi, \sigma) = 4$, because the weight of a transposition is 2.

General formula for Transpositions

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

Example:

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

$$\rho\pi = (1\ 3\ \underline{4}\ \underline{6}\ \underline{2}\ 5)$$

$\rho' = (1\ 4\ 2)$, on the other hand, is **not** a transposition on π .

Good Transpositions

- If $\rho = (a b c)$ is a transposition on π , and elements a , b and c are in the same cycle and in this order in $\sigma\pi^{-1}$, then there exists an optimal decomposition of $\sigma\pi^{-1}$ that contains $(a b c)$.
- Such a $\rho = (a b c)$ is called a **good transposition**, bringing π closer to σ , that is,

$$d(\rho\pi, \sigma) = d(\pi, \sigma) - \|\rho\| = d(\pi, \sigma) - 2$$

Finding decompositions of Transpositions

- If ρ is a good transposition, to find a decomposition of $\sigma\pi^{-1}$ that contains ρ , we multiply $\rho^{-1}\rho$ to the right of $\sigma\pi^{-1}$.

Example:

$$\pi = (1\ 3\ 2\ 5\ 4\ 6) \text{ and } \sigma = (1\ 2\ 3\ 4\ 5\ 6) \Rightarrow \sigma\pi^{-1} = (2\ 4\ 6\ 5\ 3)$$

$\rho = (2\ 6\ 3)$ is a good transposition. Decomposing $\sigma\pi^{-1}$ we get:

$$\sigma\pi^{-1} = \sigma\pi^{-1}(\rho^{-1}\rho) = (2\ 4\ 6\ 5\ 3)(3\ 6\ 2)(2\ 6\ 3) = (3\ 5)(4\ 6)(2\ 6\ 3)$$

ρ is a good transposition because:

$$\|\sigma\pi^{-1}\| = \|(2\ 4\ 6\ 5\ 3)\| = \|(3\ 5)(4\ 6)(2\ 6\ 3)\| = 4$$

$\rho' = (2\ 5\ 6)$, for instance, would not be a good transposition. **(why?)**

Sorting by Fusions, Fissions and Transpositions (FFT)

- Given permutations π and σ , the FFT distance between π and σ is given by

$$d(\pi, \sigma) = \|\sigma\pi^{-1}\|$$

where fissions and fusions have weight 1, and transpositions have weight 2.

- The rearrangement operations can be found by decomposing $\sigma\pi^{-1}$ in 2-cycles (fissions and fusions) and 3-cycles (good transpositions).

Summary

- The algebraic theory uses permutation group properties to solve rearrangement problems.
- The distance is based on the norm of $\sigma\pi^{-1}$, and rearrangement events are found by decomposing $\sigma\pi^{-1}$.
- It can also be used in multichromosomal signed genomes, and on linear chromosomes, not only circular.
- The resulting model is similar to the Double-Cut-and-Join, but with different weights.