

# PERMUTATIONS AS GENOME REARRANGEMENTS

SIMON GRÜNING

(BLOCK INTERCHANGES)

---

(Simon Grüning) UNIVERSITY OF ZURICH, RÄMISTRASSE 71, 8006 ZURICH  
*E-mail address:* [simon.gruening@uzh.ch](mailto:simon.gruening@uzh.ch).

## 1. THE CYCLE GRAPH AND BLOCK INTERCHANGE DISTANCE

REMARK 1.1. Recall an inversion in  $p = p_1 p_2 \cdots p_n$  is such that  $(i, j \in [n] : p_i > p_j \wedge i < j)$ .

DEFINITION 1.1. A **Block Interchange** is an operation that interchanges two blocks of consecutive entries without rearranging said entries.

EXAMPLE 1.1.  $|34|17|562| \rightarrow |562|17|34|$

DEFINITION 1.2. Let the **Block Interchange Distance** between two  $n$ -permutation  $p, q$ , denoted as  $\text{bid}(p, q)$ , be the smallest number of block interchanges required to transform  $p$  into  $q$ . Let  $\text{bid}(p) := \text{bid}(p, \text{id})$  be the number of interchanges required to sort  $p$ .

REMARK 1.2.  $\text{bid}(p, q)$  is a metric and is right-invariant, just as we have seen for  $\text{btd}(p, q)$ .

DEFINITION 1.3. Let  $p = p_1 p_2 \cdots p_n$  be an  $n$ -permutation. The Cycle Graph  $G(p)$  is a graph of coloured directed edges on the vertex set  $\{0, 1, \dots, n, n+1\}$  constructed as follows:

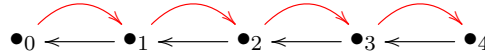
- (1) For every  $i$  with  $1 \leq i \leq n+1$ : Add a black edge from  $p_i$  to  $p_{i-1}$
- (2) For every  $i$  with  $0 \leq i \leq n$ : Add a red edge from  $i$  to  $i+1$

with  $p_0 = 0$  and  $p_{n+1} = n+1$ .

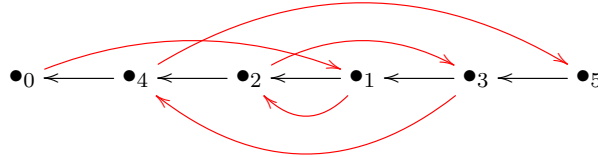
REMARK 1.3. The Cycle Graph has  $2n+2$  edges for any  $n$ -permutation.

EXAMPLE 1.2.

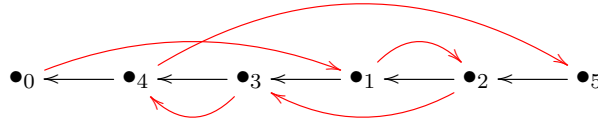
$$p_1 = 123 \implies c(G(p_1)) = 4$$



$$p_2 = 4213 \implies c(G(p_2)) = 1$$



$$p_3 = 4312 \implies c(G(p_3)) = 3$$



For the final permutation  $p_3$  notice that  $\text{bid}(p_3) = 1 = \frac{4+1-3}{2}$ , something which will become relevant in a later theorem.

DEFINITION 1.4. Two cycles are edge-disjoint when they do not share an edge. A cycle on a directed graph is directed if all edges are of the same orientation.

LEMMA 1.1. The cycle graph  $G(p)$  has a unique decomposition into edge-disjoint directed cycles in which the colors of the edges alternate.

*Proof.* It is easy to see that a path of alternating colours cannot branch since by construction there is always only one edge per color directed away from every node. We begin at the 0-node and complete a cycle (this must occur since  $G(p)$  is finite). Once we have done so, we remove the edges of this cycle from  $G(p)$  and iterate. Notice that in removing full cycles the structure is maintained as we only remove disjointly coloured pairs of edges from every node. The cycles are then edge-disjoint by construction.  $\square$

**DEFINITION 1.5.** Let  $c(G(p))$  be the **number of alternating directed cycles** in the decomposition of  $G(p)$ . To avoid confusion, let  $c(\Gamma(p))$  denote the number of cycles in the traditional sense of a permutation. From now on the cycles of  $G(p)$  will refer to the alternating cycles.

**REMARK 1.4.**

- (1) The cycles are only edge-disjoint not necessarily vertex-disjoint.
- (2) If there is an alternating path from one node to another, both nodes are in the same cycle.
- (3) The identity permutation has  $n + 1$  cycles. In fact, it is the only permutation with this many cycles.

Before we can find a formula for  $\text{bid}(p)$  we shall examine how single block interchanges can influence the block interchange distance of a permutation in the following two lemmas.

**DEFINITION 1.6.** Let  $p = p_1 p_2 \cdots p_n$  with  $p \neq \text{id}$ , or in other words,  $p$  contains at least one inversion. Select the inversion in  $p$  ( $i, j \in [n] : p_i > p_j \wedge i < j$ ) such that  $p_j$  is minimal and if necessary, as a second criterion that  $p_i$  is maximal.

Let  $x := p_j, y := p_i$  to ease our eyes. By our selection criterion we find that  $x - 1$  occurs before  $y$  in  $p$ , whilst  $y + 1$  occurs after  $x$ . (In the case of  $x = 1$  simply let  $x - 1 = 0$ )

$$p = \cdots (x - 1) \mid \cdots y \mid \cdots \mid x \cdots \mid (y + 1) \cdots$$

The **Canonical Block Interchange** is defined as the interchange transforming  $p$  into

$$p' = \cdots (x - 1) \mid x \cdots \mid \cdots \mid \cdots y \mid (y + 1) \cdots$$

**LEMMA 1.2. (9.10)** Let  $p \in S_n$  with  $p \neq \text{id}$ . Then there exists a block interchange which increases  $c(G(p))$  by 2.

*Proof.* We shall prove that the canonical block interchange suffices, or symbolically:

$$c(G(p')) = c(G(p)) + 2.$$

Notice first that any block interchange does not affect the red edges in  $G(p)$ . The canonical interchange changes 3 black edges when  $x$  and  $y$  are adjacent, and 4 edges otherwise. We shall perform a proof by cases.

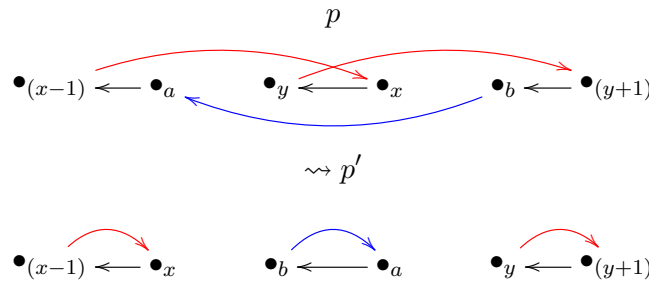
- (1)  $x, y$  adjacent: The cycle decomposition of  $G(p)$  will contain an alternating path through the following nodes in respective order:

$$a \rightarrow (x - 1) \rightarrow x \rightarrow y \rightarrow (y + 1) \rightarrow b$$

where  $a$  is the entry in  $p$  directly to the right of  $(x-1)$  and  $b$  the one to the left of  $(y+1)$ .

$$p = \cdots (x-1) | a \cdots y | \cdots | x \cdots b | (y+1) \cdots$$

Since a path exists, we find that the entries are part of the same cycle  $C$ . After performing the canonical interchange,  $C$  is split into 3 cycles. In the black edges we now have  $x \rightarrow (x-1)$ ,  $(y+1) \rightarrow y$ , and  $a \rightarrow b$  while the red have been untouched as seen in the following illustration:



Notice the arrow in blue is not necessarily a singular edge, however it is a well defined alternating path.

- (2)  $x, y$  not adjacent: We must introduce another notation in this case. Let  $a, b, c, d$  be such that they are adjacent to  $(x-1), y, x, (y+1)$  respectively pair-wise. Since they are neighbours we will find the alternating paths

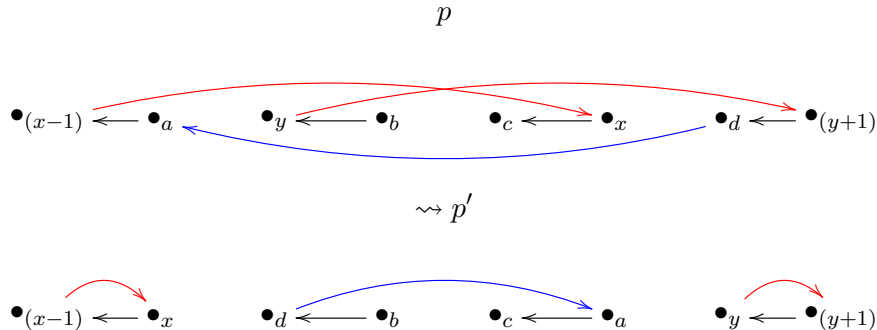
$$P_a = a \rightarrow (x-1) \rightarrow x \rightarrow c$$

and

$$P_b = b \rightarrow y \rightarrow y+1 \rightarrow d$$

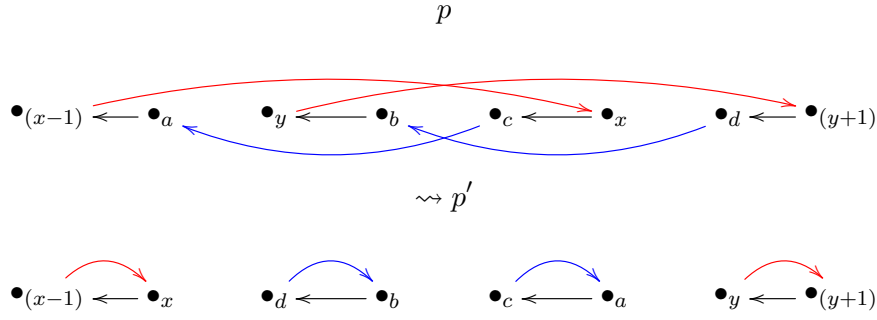
We must make a further division of cases at this point.

- (a)  $P_a, P_b$  are of one cycle: Denote the cycle as  $C_{ab}$ . In this case there exists an alternating path  $A_{da}$  from  $d$  to  $a$  with the final edge being grey. See the following illustration:



where the blue edge is our path  $A_{da}$ , the red edges are kept static, and once more we have split one cycle into three, all else remaining equal.

- (b)  $P_a, P_b$  are of distinct cycles: Denote these cycles  $C_a$  and  $C_b$  respectively. We find that each cycle shall split into two, we shall cement this only with a short illustration as the argument is analogue.



Where the blue edges are our paths  $C_a$  and  $C_b$  exempted  $(x-1) \rightarrow x$  and  $y \rightarrow (y+1)$  correspondingly.

Having exhausted all cases, we conclude the proof. □

LEMMA 1.3. (9.11) *Let  $p \in S_n$ . Any block interchange performed on  $p$  cannot increase  $c(G(p))$  by more than 2.*

*Proof.* We shall only provide a quick sketch of the proof which makes use of the reversibility of block interchanges to create a contradiction.

- (1) Any block interchange changes at most 4 black edges of  $G(p)$ .
  - (2) For a block interchange to create more than 2 cycles, it must break one cycle into 4 smaller cycles.
  - (3) If this were the case, the inverse block interchange would have to turn 4 cycles into one cycle, which leads to contradiction upon closer examination.  $\nexists$
- 

THEOREM 1.1. (9.9) *Let  $p \in S_n$ . Then*

$$bid(p) = \frac{n+1 - c(G(p))}{2}.$$

*Proof.* By the previous two lemmata we have for any  $p \in S_n$  that

$$\frac{n+1 - c(G(p))}{2} \leq bid(p) \leq \frac{n+1 - c(G(p))}{2}$$

since we require a change in  $c(G(p))$  of magnitude  $(n+1) - c(G(p))$  to achieve

$$c(G(\tilde{p})) = n+1 \iff \tilde{p} = id$$

and this can be achieved by half the same number of block interchanges at most and at least, thus culminating our desired equality. □

REMARK 1.5. *We have discovered that we need at most  $\lfloor \frac{n}{2} \rfloor$  block interchanges to sort an  $n$ -permutation. Furthermore  $bid(p) \in \mathbb{N}$  implies that  $(n+1)$  and  $c(G(p))$  must either both be odd or even.*

## 2. AVERAGE NUMBER OF BLOCK INTERCHANGES REQUIRED

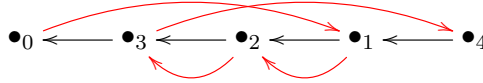
We would now like to find the average number of block interchanges required for any arbitrary  $n$ -permutation. We wish to count the cardinality of the set of  $n$ -permutations restricted to a specific number of cycles in order to calculate this average.

DEFINITION 2.1. *Define the Hultman Number as*

$$S_H(n, k) := |\{\pi \in S_n : c(G(p)) = k\}|$$

EXAMPLE 2.1.  $S_H(3, 4) = 1$  and  $S_H(3, 2) = 5$ . Thus all 5 non-trivial 3-permutations have 2 cycles.

Have a glance at the graph of  $p = 321$ :



THEOREM 2.1. (9.14)

$$S_H(n, k) = |\{(q, r) \in S_n^2 : c(\Gamma(r)) = k \wedge qr = (12 \cdots n(n+1))\}|$$

The previous theorem illustrates the connection between  $c(G(p))$  and  $c(\Gamma(p))$ . It allows us to translate our question into previously explored territory. After some complex and difficult papers alleged to require things such as symmetric functions, character theory, and non-elementary integration, we find our answer:

THEOREM 2.2. (9.23) *The average number of block interchanges needed to sort an  $n$ -permutation is*

$$\frac{1}{2} \left( n - \frac{1}{\left\lfloor \frac{n+2}{2} \right\rfloor} - \sum_{i=2}^n \frac{1}{i} \right)$$