

Complexity of Cayley Distance and other General Metrics on Permutation Groups

¹Thaynara Arielly de Lima and ^{1,2}Mauricio Ayala-Rincón

Grupo de Teoria da Computação, Departamentos de ¹Matemática e ²Ciência da Computação
Universidade de Brasília, Brasília D.F., Brazil
{thay@mat, ayala@}unb.br

Abstract. Permutation groups arise as important structures in group theory because many algebraic properties about them are well-known, which makes modelling natural phenomena by permutations of practical interest. This paper reviews the complexity of some problems involving permutation groups. Usability of the involved algebraic notions is illustrated by problems such as genome rearrangement by reversals for which it is well-known that for the case of unsigned and signed sorting by reversals the time complexity is, respectively, \mathcal{NP} -hard and \mathcal{P} . Reversal distance is a particular metric and in this work more general metrics on permutation groups are considered emphasizing on the Cayley distance. In particular, we point out an error in one of the polynomial reductions applied in Pinch's approach attempting to prove that the subgroup distance problem for Cayley distance is \mathcal{NP} -complete and following his approach we present a simplified and correct proof of this fact.

1 Introduction

Biological evolution is given by the modifications that occur in genes of living organisms. If these modifications are passed to future generations by genetic material, they are evolutive changes. In this context, investigations arise, among others, about the history of evolution of a specie or evolution from an organism to another. The genome rearrangement problem can be conceived as the obtention of adequate tools for comparing sequences of genes in molecular biology.

Genome rearrangements are mutations that only change the genes order, without modifying or destroying them. Several kinds of transformations can modify the genes order. Among them, one could mention reversals, transpositions and block interchange. Two or more kinds of transformations could occur simultaneously. The number of genome rearrangements necessary to obtain a genome from another one can be used as a distance measure between the two given genomes.

Given a set of admissible operations, the general genome rearrangement problem consists in finding the minimum number of operations to transform a genome into another one.

The problem is modeled as follows. The order of the genes of an organism is characterized as a numeric sequence, where each element of the sequence represents a particular gene. If one knows the orientation of the genes, it can be modelled according to this orientation: if the gene occurs from right to left it is represented by a negative element in the numeric sequence. A positive element is associated otherwise.

Initially, this work reviews the time complexities of the problems of sorting by unsigned and signed reversals following an approach that is based on the representation of permutations by permutation cycles.

The reversal distance is a very particular metric on the symmetric groups and other relevant metrics such as the Cayley and Hamming distances are of interest. Then, in this more general context is presented the main contribution of this work that is a correct proof of the fact that the Subgroup Distance Problem is \mathcal{NP} -complete. The proof follows Pinch's approach that is based on a sequence of polynomial reductions, from which the first one, that is a reduction from $3SAT$ to the problem of finding maximal cycle decomposition routings in polarised switching circuits, is corrected and simplified.

2 Complexity of Sorting by Reversals

Let X be a nonempty finite set and $S_X := \{\alpha \mid \text{bijection from } X \rightarrow X\}$. It is well-known that S_X provided of the composition operation of functions is a group. This is the permutation group of the set X . If $X = \{1, \dots, n\}$ the group is called symmetric group and denoted by S_n .

A permutation $\alpha \in S_n$ is an r -cycle if there exist different elements $a_1, a_2, \dots, a_r \in \{1, 2, \dots, n\}$ such that $\alpha(a_i) = a_{i+1}, 1 \leq i \leq r-1, \alpha(a_r) = a_1$ and $\alpha(j) = j, \forall j \in \{1, 2, \dots, n\} \setminus \{a_1, \dots, a_r\}$; this r -cycle is denoted by $(a_1 \ a_2 \ \dots \ a_r)$. The number r is called the cycle length.

Consider the genes order in an organism represented by the string $\pi = \pi_1 \ \pi_2 \ \dots \ \pi_n$, $\pi_i \in \{1, 2, \dots, n\}, 1 \leq i, j \leq n$ where $\pi_i \neq \pi_j, \forall i \neq j$.

Let $\pi = \pi_1 \ \pi_2 \ \dots \ \pi_n$ be a string. By $\pi(j)$ it is denoted π_j . The permutation in S_n associated to the string π maps π_j to j , for all $1 \leq j \leq n$. For example, we associate the permutation $(2 \ 1 \ 5 \ 4 \ 3) \in S_6$ to the string $2 \ 3 \ 4 \ 5 \ 1 \ 6$.

Now, let σ be a permutation in S_n . To this permutation it is associated a string of n symbols, denoted as σ_{str} , such that $\sigma_{str}(i) := j$, whenever $\sigma(j) = i$ and the symbols of the string are all different and belong to the set $\{1, 2, \dots, n\}$.

In this way, given a permutation π , we can associate a string π_{str} and vice-versa. Thus, it is possible to build an isomorphism between the set S of all strings of length n of different symbols of the set $\{1, 2, \dots, n\}$ and the symmetric group S_n .

A *reversal* is a permutation $\rho \in S_n$ presented by permutation cycles of the form

$$(i \ j)(i+1 \ j-1) \dots (i + \lfloor \frac{j-i}{2} \rfloor + 1 \ i + \lceil \frac{j-i}{2} \rceil + 1), \text{ for } 1 \leq i < j \leq n$$

The genome rearrangement problem of sorting by reversals (*SBR*) can be formulated as follows: given two permutations π and σ in S_n , find the minimum number of reversals that transform π into σ .

The problem above is equivalent to *MIN-SBR*, defined as follows: given a permutation π , find the minimum number of reversals $\rho_1, \rho_2, \dots, \rho_t$ such that $\pi \rho_1 \rho_2 \dots \rho_t = id$, where id denotes the identity permutation. This number is called reversal distance, denoted as $d(\pi)$.

An important tool used in the study of *MIN-SBR* is the *breakpoint graph* of a permutation π . Let $i \sim j$ whenever $|i - j| = 1$ and $i \not\sim j$ otherwise. Let $\pi \in S_n$ be a permutation. Extend π_{str} by adding $\pi_0 = 0$ and $\pi_{n+1} = n + 1$. A pair of consecutive elements π_i and $\pi_{i+1}, 0 \leq i \leq n$ of π_{str} is called an *adjacency* if $\pi_i \sim \pi_{i+1}$ and a *breakpoint* if $\pi_i \not\sim \pi_{i+1}$.

Define a bi-colored graph $G(\pi)$ with $n + 2$ vertices such that each vertex i has the label $\pi_{str}(i)$ if $1 \leq i \leq n + 1$. The edges set is divided in breakpoint edges (black edges) and desired edges (gray edges). We join vertices i and $i + 1$ by a black edge if, and only if, the i and $i + 1$ labels are a breakpoint. Furthermore, there is a gray edge between the vertices i and j if, and only if, $i \not\sim j$, and the i and j labels are adjacency. This graph is called *breakpoint graph* of the permutation π (See Fig. 1 (i)).

Let us denote $b(\pi)$ the number of breakpoints of a permutation π .

In a breakpoint graph $G(\pi)$ of a permutation π , a cycle is called *alternating cycle* if the colors of the edges alternate. $c(\pi)$ denotes the maximum number of edge-disjoint alternating cycles.

In a breakpoint graph of a permutation π , $G(\pi)$, assign each black edge $(i, i + 1)$, with labels π_i and π_{i+1} respectively, an orientation from i to $i + 1$. An alternating cycle of $G(\pi)$ is called *directed* with respect to π if it is possible to walk along the whole cycle traversing each black edge according of its orientation and *undirected* otherwise. In the Figure 1 (ii), the cycle with labels 3, 1, 2, 4 is directed with respect to π , while the cycle with labels 0, 1, 5, 6, 4, 3 is undirected with respect to π .

A characterization of reversal distance by breakpoints and undirected alternating cycles can be gives [BP96,HP95b,HP95a]. Consider a permutation π , its breakpoint graph $G(\pi)$ and a maximum decomposition of $G(\pi)$ into $c(\pi)$ alternating cycles. If every cycle in this decomposition is undirected with respect to π , then $d(\pi) = b(\pi) - c(\pi)$ or $d(\pi) = b(\pi) - c(\pi) + 1$.

This characterization point out the interest on the maximum alternating cycle decomposition problem (*MAX-ACD*) that consists in finding a maximum number of alternating cycles in a decomposition of the breakpoint graph of a permutation π .

The rearrangement problem by reversal was proved \mathcal{NP} -hard by Caprara [Cap97]. Caprara's proof is based on three polynomial reductions:

1. *3SAT* reduces polynomially to *MAX-ECD*;
where, *3SAT* is the language of all satisfiable conjunctive normal form formula in wich all clauses contain at most 3 literals and *MAX-ECD* problem consists in finding a maximum-cardinality cycle decomposition of an Eulerian graph ([Hol81]).
2. *MAX-EDC* reduces polynomially to *MAX-ACD*;
3. *MAX-ACD* reduces polynomial to *MIN-SBR*.

Given a part of genoma, if we know the orientation of the genes, we can associate a signed sequence to the genoma. This is a model of the genome rearrangement problem for signed permutations.

But, given a signed permutation $\bar{\pi}$ of n elements, we can transform $\bar{\pi}$ into $\pi' \in S_{2n}$ such that π' is a unsigned permutation. This transformation is done by replacing the elements $+i$ by $2i - 1, 2i$ and the elements $-i$ by $2i, 2i - 1$. For example, the signed identity permutation $id_+ = +1 + 2 \dots + n$ is mapped into $id = 1 2 \dots 2n$.

If we consider the transformation of a signed permutation $\bar{\pi}$ into a unsigned permutation π' , we have the effect that a reversal on $\bar{\pi}$ can be mimicked by a reversal on π' . Furthermore, if we apply reversals that eliminate a maximum number of breakpoints in a signed permutation, we can mimicked this reversals to eliminate the maximum number of breakpoints in the associated unsigned permutation. Thus the problem of sorting by reversals for signed permutations can be solved by analysis of unsigned permutations.

The transformation above allows us to consider the concepts of adjacency, breakpoints and breakpoint graph for signed permutations. For example, given $\bar{\pi}_{str} = +1 + 4 + 3 + 2$, we can transform into $\pi'_{str} = 1 2 7 8 5 6 3 4$. The breakpoint graph of this permutation as presented in the Figure 1 (iii).

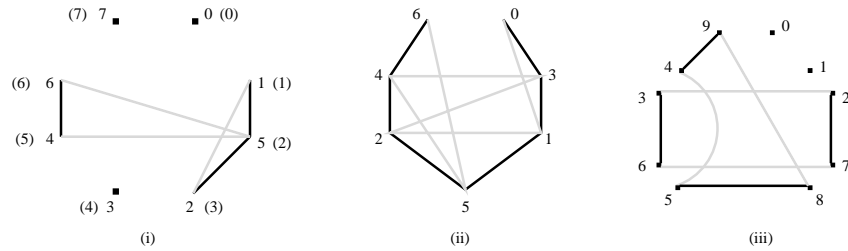


Fig. 1. (i) Breakpoint graph of permutation $\pi = (2345)$, (ii) Directed and undirected cycles in a breakpoint graph of $\pi = (31245)$, (iii) Breakpoint graph of a permutation $\pi' = (37)(48)$

Each node of a breakpoint graph has degree 0 or 2. Thus, the cycle decomposition in a breakpoint graph of a signed permutation is unique. Then, the *MAX-ACD* problem for signed permutation is polynomial. The procedure presented by Caprara [Cap97] to obtain breakpoint graphs such that every alternating cycle is undirected with respect to some permutation can be adapted for signed permutations. Just apply the transformation from signed permutations to unsigned permutations and apply the procedure. Hannenhali and Pevzner [HP95b] proved that $d(\bar{\pi}) = b(\bar{\pi}) - c(\bar{\pi})$ if every cycle of $G(\bar{\pi})$ is undirected with respect to $\bar{\pi}$. Then, it is possible to compute the reversal distance of a signed permutation in polynomial time.

3 Distance of a permutation and Subgroup distance problem

Among the variety of distance problems on permutation groups, the subgroup distance problem is of great interest. The reversal distance is a particular metric on S_n , but other metrics can be used.

A *metric* on the symmetric group S_n is a function $d : S_n \times S_n \rightarrow \mathbb{R}_+$ that, for every π, σ and $\varphi \in S_n$, satisfies:

$$i) d(\pi, \sigma) \geq 0; \quad ii) d(\pi, \sigma) = 0 \text{ if, and only if, } \pi = \sigma; \quad iii) d(\pi, \varphi) \leq d(\pi, \sigma) + d(\sigma, \varphi).$$

A right-invariant metric d on S_n is a metric satisfying $d(\pi, \sigma) = d(\pi\varphi, \sigma\varphi)$, for every $\pi, \sigma, \varphi \in S_n$.

Some metrics on S_n are well-known, for instance the Hamming distance, l_p distance, l_∞ distance and Cayley distance [AJ08], among others.

The *Subgroup Distance Problem (SDP)* with respect to a metric d on S_n is defined as: given a subgroup $H \leq S_n$, a permutation $\pi \in S_n$ and a number $k \in \mathbb{N}^*$, determine whether $d(\pi, H) := \min_{\sigma \in H} d(\pi, \sigma) \leq k$.

Note that, *MIN-SBR* is an instance of *SDP*, just take $H = \langle id \rangle$ and d as the reversal distance.

For the metrics mentioned above, *SDP* is \mathcal{NP} -complete [AJ08]. In the remainder of this section we will follow Pinch's approach used to state this fact for the Cayley distance. We will present a corret and detailed proof of this fact pointed out the problems in Pinch's work [Pin07].

Given two permutations π and $\sigma \in S_n$, the Cayley distance is the minimum number of transpositions (cycles of length two) transforming π into σ .

Let $S \subset S_n$ be a set of elements of the form $\gamma_j = (x_j^1 \ y_j^1) \dots (x_j^{r_j} \ y_j^{r_j})$, where all x_j^i and y_j^i are different. We call S *Involutions with Disjoint Support (IDS)*. For example, if we consider the symmetric group S_6 , $S_1 = \{(1 \ 2), (3 \ 4), (5 \ 6)\}$ and $S_2 = \{(1 \ 3), (2 \ 5), (4 \ 6)\}$ are IDS's of S_6 . The *width* of an IDS is defined as the maximum number of 2-cycles r_j in the elements γ_j . The *SDP* with the subgroup $H := \langle \gamma_j \rangle$ generated by the elements γ_j of an IDS of width w is called the *IDS_w-Subgroup-Distance*.

Additional definitions are necessary. A *switching circuit* is a directed graph G such that for all $v \in V$ the cardinality of input and output edges coincide; for each $v \in V$, its *valency*, denoted as $\partial(v)$, is the number of in edges (which equals the number of out edges) and, where each in and out edge has a different label in $\{1, \dots, \partial(v)\}$. The valency of G is the maximum among the valencies of its vertices. For any edge $(u, v) \in E$, its output label, as an in edge, and its input label, as an out edge are not related. A *routing* ρ for a switching circuit is a choice of a permutation $\rho(v) \in S_{\partial(v)}$, for each vertex $v \in V$. For an example see Fig. 2. Note that, there is a correspondence between routings of a switching circuit G and decompositions of the edge set into directed cycles of G .

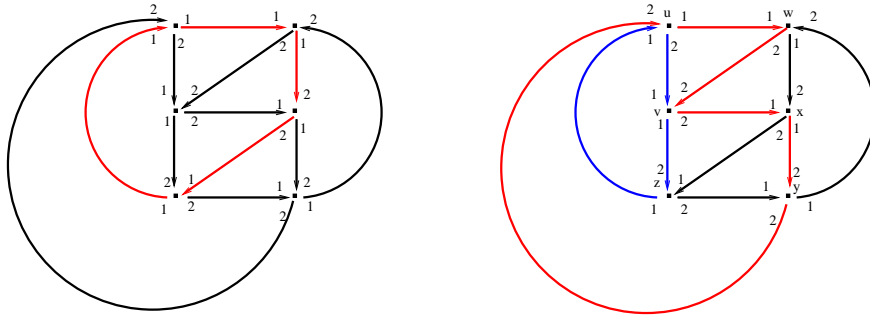


Fig. 2. Figure to the left shows a decomposition in two directed cycles with routing $\rho = id$ for all vertices; Figure to the right shows a decomposition in three cycles given by the routing $\rho(u) = \rho(w) = \rho(z) = (1 \ 2)$ and $\rho(v) = \rho(x) = \rho(y) = id$

A *polarisation* T for a switching circuit $G(V, E)$ is an equivalence relation over the set of vertices V , such that vertices belonging to the same class have the same valency. The pair (G, T) is called a *polarised switching circuit*. Note that vertices having the same valency are not necessarily in the same class. A routing ρ is said to *respect the polarisation* T if $\rho(x) = \rho(y)$, whenever vertices x and y belong to the same class. Routings in the Fig. 2 respect polarisation with a unique equivalence class (Fig. to the left) and two equivalence classes (Fig. to the right).

Note that, for distinct labels and routings, the decomposition into cycles changes.

We define the *Polarised-Switching-Circuit-Routing (PSCR)* as the problem stated as: given a polarised switching circuit (G, T) and a positive integer k , determine the existence of a routing which respects T and has at least k cycles in the associated decomposition in cycles.

The *width of a polarisation* T is defined as the maximum number of vertices in a class of T . We call *Width_w-Valency_n-Routing* the *PSCR* with the width of T restricted to be at most w and $\partial(v)$ of each vertex v restricted to be at most n .

The proof that *SDP* is \mathcal{NP} -complete is made in two steps following [Pin07] approach, but correcting and improving the first step, for which the original proposed width was 6:

1. Prove that *Width₄-Valency₂-Routing* is \mathcal{NP} -complete;
2. Show the existence of an equivalence between *Width_w-Valency₂-Routing* and *IDS_w-Subgroup-Distance*.

Applying both these results one obtains that *IDS₄-Subgroup-Distance* is \mathcal{NP} -complete from which one immediately concludes that *SDP* is \mathcal{NP} -complete as well. In the following subsections proofs of these results are presented.

3.1 $Width_4$ - $Valency_2$ -Routing is \mathcal{NP} -complete

The first step in Pinch's paper is in fact a tentative to prove that $3SAT$ polynomially reduces to $Width_6$ - $Valency_2$ -Routing, but one of the circuits presented is incorrect because it does not satisfy the necessary properties as presented in Appendix A. The current proof is in fact an improvement because in the first step we reduce the width of the routing problem.

A polarised switching circuit (G, T) is *Boolean* if every vertex has valency at most two. To each class C of the polarisation T a Boolean variable $a(C)$ is associated, where $a(C) = 0$ if, and only if, the permutation $\rho(v) = id \in S_2$ and $a(C) = 1$, if, and only if, $\rho(v) = (1\ 2) \in S_2$, for all $v \in C$. There is a straightforward correspondence between routing and designation of boolean values to the vertices of (G, T) . For a negated variable \bar{a} we exchange the input labels 1 and 2 in all the associated vertices.

The reduction in the first step of the proof is based on a representation of unary, binary and tertiary clauses in a formula, instance of $3SAT$, by corresponding switching circuits that have an specific number of cycles exactly when the clauses hold. For Boolean variables a, b and c , we consider the switching circuits $I(a), E(a, b), F(a, b)$ and $A(a, b, c)$ corresponding to unary clauses, equality between variables, binary and tertiary clauses, respectively. See Figs. 3 and 4.

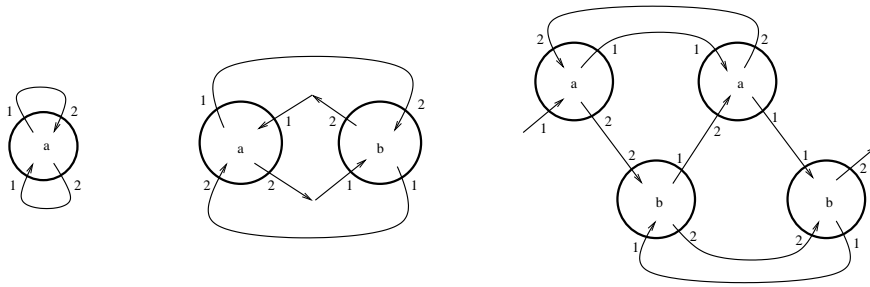


Fig. 3. Switching circuits $I(a), E(a, b)$ and $F(a, b)$ for unary clauses, equality between variables and binary clauses

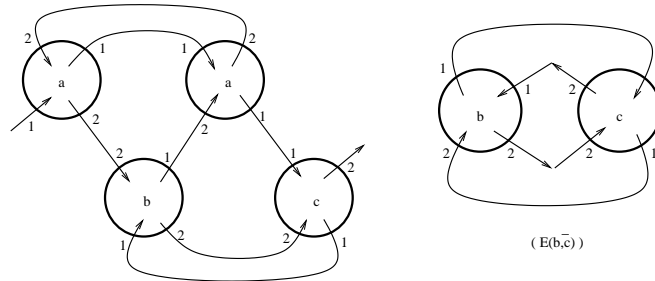


Fig. 4. Switching circuit $A(a, b, c)$ for tertiary clauses

Proposition 1 (Properties of I, E, F and A).

1. the number of cycles for $I(a)$ is 2 if $a = 1$ and 1 otherwise;
2. the number of cycles for $E(a, b)$ is 2 if $a = b$ and 1 otherwise;
3. the number of cycles for $F(a, b)$ is 1 if $a = b = 0$ and 3 otherwise;
4. the number of cycles for $A(a, b, c)$ is 2 if $a = b = c = 0$, 4 otherwise.

Proof: We will demonstrate the item 4, that uses item 2. All other items are proved similarly by case analysis.

Notice that, according to the item 2, the circuit $E(b, c)$ has two cycles whenever $b = c$ and only one otherwise, as depicted in Fig. 5. The right part of the circuit $A(a, b, c)$ is exactly $E(b, \bar{c})$ and consequently this sub circuit will have two cycles if $b \neq c$ and one if $b = c$. In order to conclude the proof of item 4, we will proof the following:

- the left part of $A(a, b, c)$ has one circuit if $a = b = c = 0$. Observe this circuit in Fig. 6. Thus, $A(a, b, c)$ has two circuits in this case.
- the left part of $A(a, b, c)$ has three circuits if $a \neq b = c$ or $a = b = c = 1$. Observe this case in Fig. 7. Thus $A(a, b, c)$ has four circuits in this case.
- the left part of $A(a, b, c)$ has two circuits if $b \neq c$. Observe this case in Fig. 8. Thus $A(a, b, c)$ will have four circuits. \square

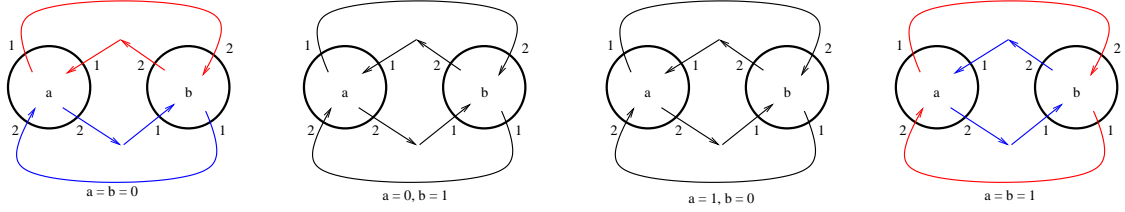


Fig. 5. Cycles of circuit $E(a, b)$ for $a = b$ and $a \neq b$

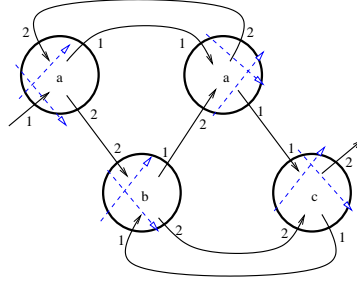


Fig. 6. Cycle of left circuit of $A(a, b, c)$ for $a = b = c = 0$

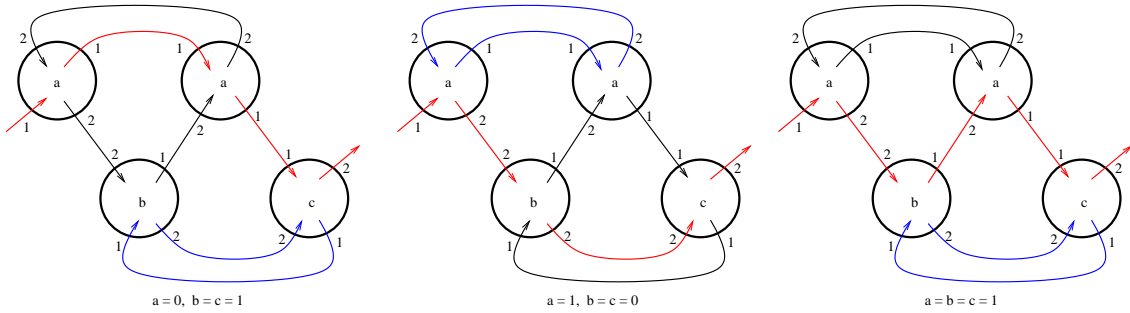


Fig. 7. Cycles of left circuit of $A(a, b, c)$ for $a \neq b = c$ and $a = b = c = 1$

Theorem 1. *There is a polynomial reduction from 3-SAT to $\text{Width}_4\text{-Valency}_2\text{-Routing}$.*

Proof: Let φ an instance of 3-SAT that is a Boolean formula over variables x_1, \dots, x_n , that is a conjunction of k clauses each of which is a disjunction of at most three variables or their negations.

Firstly, one transforms φ into an equivalent formula φ' in this way: for all x_i , replace its j^{th} occurrence in φ by a new variable y_i^j . For all x_i include the conjunction of clauses $(y_j^1 \equiv y_j^2) \wedge \dots \wedge (y_j^{(r_i-1)} \equiv y_j^{r_i})$

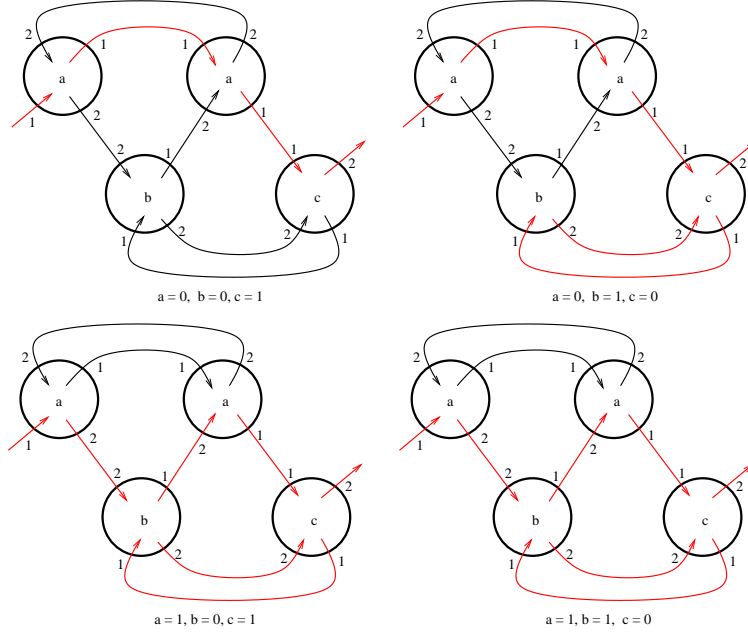


Fig. 8. Cycles of left circuit of $A(a, b, c)$ for $b \neq c$

where the variable x_j occurs r_i times in φ . For example, if $\varphi = (x_1 \vee \bar{x}_2 \vee x_3) \wedge (x_2 \vee \bar{x}_3 \vee x_4) \wedge (\bar{x}_1 \vee x_3)$ then $\varphi' = (y_1^1 \vee \bar{y}_2^1 \vee y_3^1) \wedge (y_2^2 \vee \bar{y}_3^2 \vee y_4^1) \wedge (\bar{y}_1^2 \vee y_3^3) \wedge (y_1^1 \equiv y_1^2) \wedge (y_2^1 \equiv y_2^2) \wedge (y_3^1 \equiv y_3^2) \wedge (y_3^2 \equiv y_3^3)$.

Note that, in fact, φ' is equivalent to φ . Thus, we have the same numbers of satisfying designations. Note also that each variable occurs at most three times in φ' , and exactly once in a disjunction. Therefore, the length of φ' is linear in the length of φ .

Secondly, one will construct a polarised switching circuit Ψ for the formula φ' as the forest consisting of the following circuits:

- for each tertiary clause of the form $(x \vee y \vee z)$ take a circuit $A(x, y, z)$;
- for each binary clause of the form $(x \vee y)$ take a circuit $F(x, y)$;
- for each unary clause of the form (x) take a circuit $I(x)$ and;
- for each clause of the form $(x \equiv y)$ take a circuit $E(x, y)$.

The classes in the polarised switching circuit Ψ are given as the sets of vertices labelled by the same variable of φ' . Then, this polarisation will have exactly n classes, where n is the number of variables in φ' . Observe that each class in this polarisation is involved in at most a circuit of the form A , F or I and, in addition, in at most two circuits E . Therefore each class in the polarised switching circuit Ψ has at most 4 vertices. Thus the size of Ψ is at most $4n$, that is, the size of Ψ is linear in the length of the formula φ' .

Thirdly, denote as a, f, i and e the number of circuits of types A, F, I and E in Ψ , respectively. Consider the number $M = 4a + 3f + 2i + 2e$. And finally, conclude observing that according to Proposition 1, there exists a routing for the polarised circuit Ψ which gives a decomposition into M cycles if, and only if, there exists an assignment of Boolean values for the variables in φ' that satisfies φ' . Namely, notice that a satisfying assignment for φ' corresponds to a routing in Ψ which decomposes into M cycles and vice-versa. \square

3.2 $Width_w$ -Valency₂-Routing problem polynomially reduces to IDS_w -Subgroup-Distance

This proof follows Pinch's approach.

Theorem 2 ([Pin07]). *There is a polynomial equivalence between the $Width_w$ -Valency₂-Routing and IDS_w -Subgroup-Distance problems.*

Proof: On the one side, consider an instance of *Width_w-Valency₂-Routing*, that is, a polarised switching circuit $(G(V, E), T)$, where each equivalence class has width at most w . Let n be the number of edges in G and associate a different number in $\{1, \dots, n\}$ to each edge. Construct a permutation π in this way: for each vertex $v \in V$ and each edge e into v define $\pi(e)$ as the edge f out of v , such that the labels of e as an input edge of v and of f as an output edge of v are equal. Construct an instance of the *IDS* problem from G in this way: for each equivalence class in T , $C_i = \{v_j^i \mid j = 1, \dots, r_j\}$, let γ_j be a generator given as the product of transpositions $(f_j^i \ g_j^i)$, where f_j^i and g_j^i are the edges out the vertex v_j^i . Notice that the number of transpositions in γ_j is at most w , since each equivalence class in T has at most w vertices. Observe that there is a correspondence between routings in the polarised circuit and a cycle decomposition of $G(V, E)$ and the cycles in a permutation $\pi\eta$, where $\eta \in H = \langle \{\gamma_j\} \rangle$.

On the other side, consider an instance of *IDS_w-Subgroup-Distance*, that is an element $\pi \in S_n$ and an *IDS* on a set of t generators $\{\gamma_j\}$, with $\gamma_j = (x_j^1 \ y_j^1) \dots (x_j^{r_j} \ y_j^{r_j})$, all x_j^i, y_j^i distinct and all their $r_j \leq w$. A polarised switching circuit $(G(V, E), T)$ is built from this graph as described below. $V := \{P_1, \dots, P_n\} \cup \{Q_{1,1}, \dots, Q_{t,r_t}\}$. For vertices of the former type $\partial(P_k) = 1$ and of the latter type $\partial(Q_{j,i}) = 2$. For each $1 \leq j \leq t$ and $1 \leq i \leq r_j$, edges from $P_{x_j^i}$ and from $P_{y_j^i}$ to $Q_{j,i}$ are built whose input labels are 1 and 2, respectively; edges from $Q_{j,i}$ to $P_{\pi(x_j^i)}$ and to $P_{\pi(y_j^i)}$ with respective output labels 1 and 2 are built. The polarisation T of G is given by the classes $C_j := \{Q_{j,i} \mid 1 \leq i \leq r_j\}$.

Also in this case there is a correspondence between routings ρ of the polarised switching circuit and permutations of the form $\pi\eta$, where $\eta \in H := \langle \{\gamma_j\}_{1 \leq j \leq t} \rangle$. In this correspondence the number of cycles in the routing ρ is equal to the number of cycles in the permutation $\pi\eta$.

Based on the observation that a transposition can split a cycle permutation at most into two cycles, as explained below, one can conclude that π is within distance d of the group H if and only if there is a routing ρ with at least $n - d$ cycles.

A cycle of length $k > 1$ in a permutation can be factored as minimum product of $k - 1$ transpositions. The proof is by induction in the length k of the cycle. If $k = 2$ then $\pi = (\pi_1 \pi_2)$ that is a transposition. Suppose that the result holds for cycles of length $k > 1$ and consider a cycle $\pi = (\pi_1 \dots \pi_k \pi_{k+1})$. Note that $(\pi_1 \dots \pi_k \pi_{k+1}) = (\pi_1 \dots \pi_k)(\pi_1 \pi_{k+1})$. By induction hypothesis, $(\pi_1 \dots \pi_k)$ can be factored as a minimum product of $k - 1$ transpositions, thus π can be factored as a minimum product of k transpositions.

Since a transposition $\sigma = \sigma^{-1}$, the minimum number of transpositions necessary to transform a permutation π into *id* can be obtained factoring each cycle of π as transpositions. We can interpret graphically in this way. A permutation $\rho \in S_n$ can be observed as a product of disjoint cycles [Rot02]. We construct a directed graph $G(V, E)$, where $V = \{1, \dots, n\}$ and $E = \{(i, j) \mid \rho(i) = j\}$. Let $\pi = (\pi_1 \dots \pi_k)$, for $k > 1$, be one of the cycles in ρ . Thus, each connected component of $G(V, E)$ represents a cycle in ρ . The connected component in G associated with π is illustrated in Fig. 9, where the action of the transposition $(\pi_j \ \pi_{j+1})$ is also given. Notice that the action of any transposition of the form $(\pi_l \ \pi_m)$, where π_l and π_m belong to the cycle π , split the cycle into two cycles.

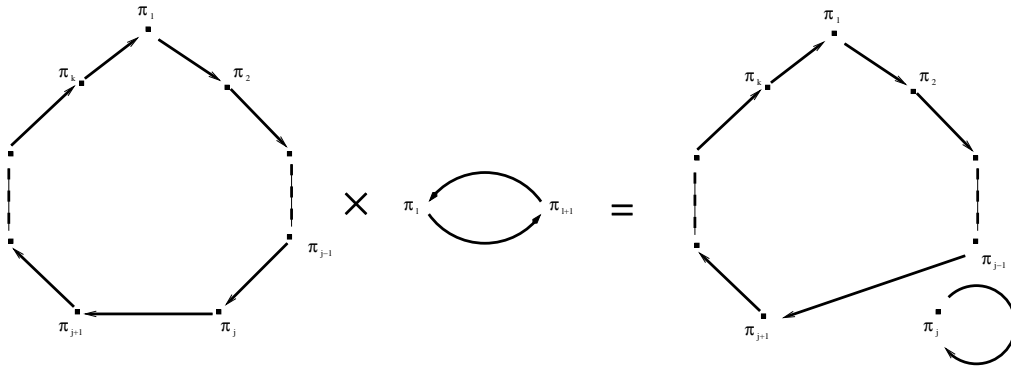


Fig. 9. Cycle in G of the permutation $\pi = (\pi_1 \dots \pi_j \dots \pi_k)$

4 Conclusion

A proof is presented of the fact that the problem of computing the general distance of a given permutation from a subgroup H of the symmetry group S_n is \mathcal{NP} -complete. This proof is based on two time-polynomial reductions: firstly, from $3SAT$ to $Width_4$ - $Valency_2$ - $Routing$ and then, from the latter problem to IDS_4 - $Subgroup$ - $Distance$. The proof follows the approach originally proposed by Pinch in [Pin07], but after detecting an error in the first reduction, that was originally proposed for the problem $Width_6$ - $Valency_2$ - $Routing$, in this paper it is presented a reduction from $3SAT$ to the simpler case of $Width_4$ - $Valency_2$ - $Routing$ problems.

The general subgroup distance problem is closely related with distances in other metrics as the one associated with the case of distance by reversion or other transformations of biological interest. We believe that the formal study of these properties from the algebraic point of view will provide a very strong insight in order to deal with open questions such as whether the reversion distance for unsigned permutations, that is known to be \mathcal{NP} -hard, is \mathcal{NP} -complete.

References

- [AJ08] Vikraman Arvind and Pushkar S. Joglekar. Algorithmic problems for metrics on permutation groups. In Viliam Geffert, Juhani Karhumäki, Alberto Bertoni, Bart Preneel, Pavol Návrat, and Mária Bieliková, editors, *SOFSEM*, volume 4910 of *Lecture Notes in Computer Science*, pages 136–147. Springer, 2008.
- [BP96] Vineet Bafna and Pavel A. Pevzner. Genome Rearrangements and Sorting by Reversals. *SIAM Journal on Computing*, 25(2):272–289, 1996.
- [Cap97] Alberto Caprara. Sorting by reversals is difficult. In *RECOMB*, pages 75–83, 1997.
- [DST01] Frank K. H. A. Dehne, Jörg-Rüdiger Sack, and Roberto Tamassia, editors. *Algorithms and Data Structures, 7th International Workshop, WADS 2001, Providence, RI, USA, August 8-10, 2001, Proceedings*, volume 2125 of *Lecture Notes in Computer Science*. Springer, 2001.
- [Hol81] Ian Holyer. The np-completeness of some edge-partition problems. *SIAM J. Comput.*, 10(4):713–717, 1981.
- [HP95a] S. Hannenhalli and P. A. Pevzner. Reversals do not cut long strips. In Dehne et al. [DST01].
- [HP95b] Sridhar Hannenhalli and Pavel A. Pevzner. Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals. In *STOC*, pages 178–189. ACM, 1995.
- [Pin07] Richard G. E. Pinch. *Combinatorics, Probability and Computing*, chapter The Distance of a Permutation from a Subgroup of S_n , pages 473–479. Cambridge University Press, 2007.
- [Rot02] Joseph J. Rotman. *Advanced Modern Algebra*. Prentice Hall, first edition, 2002.

A Details about Pinch’s proof tentative

This appendix can be dropped in a final version of this paper. Here is presented only for the benefit of the reviewing process. In Pinch’s proof that $3SAT$ polynomially reduces to $Width_6$ - $Valency_2$ - $Routing$, it is incorrectly stated that the switching circuit $F'(a, b)$ in Fig. 10 has 2 cycles whenever $a \neq b$, 3 if $a = b = 1$ and 1 if $a = b = 0$. This switching circuit is given without edge labels and the following proposition establishes that this is in fact impossible.

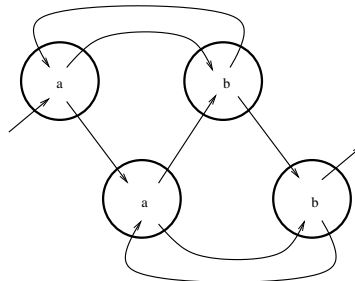


Fig. 10. Switching circuit $F'(a, b)$

Proposition 2. *There is no possible labelling for the edges of the switching circuit $F'(a, b)$ satisfying: the number of cycles in a routing for $F'(a, b)$ is 2 if $a \neq b$, 3 if $a = b = 1$ and 1 if $a = b = 0$.*

Proof: In first place notice that for the routing $a = b = 1$ if $F'(a, b)$ admits in fact a decompositions into three cycles, then, necessarily, one, and only one, of the red sub cycles illustrated in Fig. 11 should be in the decomposition. In second place, observe that for each case the other two cycles in the decomposition in three cycles is univocally determined.

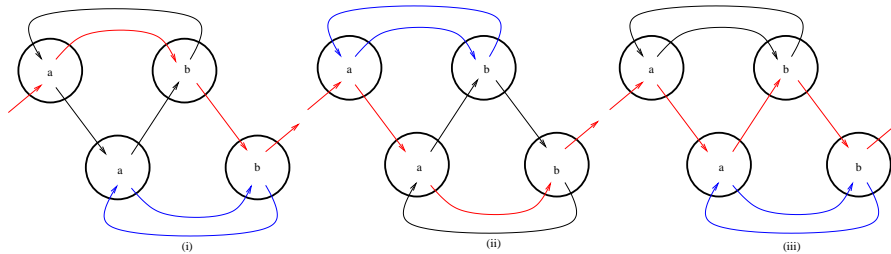


Fig. 11. Possible decompositions in three cycles of $F'(a, b)$ for the routing $a = b = 1$

In third place, for each of these three possibilities, by case analysis, one can prove that for the routings $a \neq b$ and $a = b = 0$ the decomposition into 2 and 1 cycles, resp., is impossible.

Changing the routing from $a = b = 1$ to $a = b = 0$ in each of these cases gives the decomposition in cycles depicted in Fig. 12, from which the cases (i) and (ii), for which this routing gives three cycles, are proved impossible. The sole case that remains to be analysis is the third one.

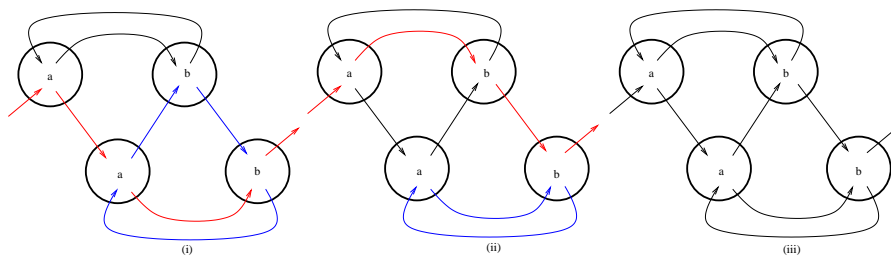


Fig. 12. Cycle decompositions of $F'(a, b)$ for the routing $a = b = 0$

Finally, one observes that the decomposition in cycles for the routings $a \neq b$ for the third case gives in both cases a unique cycle.

This concludes the proof. □