Complexity of Cayley Distance and other General Metrics on Permutation Groups

[1]Thaynara Arielly de Lima and [1,2]Mauricio Ayala-Rincón
Grupo de Teoria da Computação, Departamentos de [1]Matemática e [2]Ciência da Computação
Universidade de Brasília, Brasília D.F., Brazil
Email: {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 modeling natural phenomena by permutations of practical interest. 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 proof 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. Although, recently a shorter and more general proof than Pinch's one was given by Buchheim, Cameron and Wu, we believe the correction of Pinch's proof presented in this paper is of great interest because it correctly relates the Cayley distance problem with a maximal routing problem giving an additional perspective in relation to Buchheim et al. recent proof from which only the usual logical satisfiability perspective of distance problems is observable.**

I. Introduction

Among the variety of distance problems on permutation groups, the subgroup distance problem is of great interest. In this context, several metrics can be considered.

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

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

In the biological context several metrics on S_n can be found giving rise to different instances of the genome rearrangement problem. Take a class of operations that changes the order of genes of an organism, without modifying or destroying them. The genome rearrangement problem consists in finding the minimum number of these operations necessary to transform a genome into another one. We can consider the genes order in an organism represented by a permutation $\pi \in S_n$ [BP96]. As a class of operation one can consider, for instance, *reversals* that are permutations $\rho \in S_n$ presented by permutation cycles of the form

$$(i \quad j)(i+1 \quad j-1)\dots(i+\lfloor\frac{j-i}{2}\rfloor+1 \quad i+\lceil\frac{j-i}{2}\rceil+1),$$

for $1 \leq i < j \leq n$.

The effect of applying a reversal is to invert a piece of the genome of an organism. The sorting by reversals problem (*MIN-SBR*) consists in finding the minimum number of reversals to transform a permutation π in the permutation identity, denoted as id. The reversal distance is the minimum number of reversals for an instance of *MIN-SBR*. It is a metric on S_n.

Other metrics on S_n are well-known, for instance the Hamming distance, l_p distance, l_∞ distance, Lee distance, Kendall's tau distance, Ulam's distance and Cayley distance [AJ08], [BCW09].

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 unsigned reversals, *MIN-SBR* is $\mathcal{NP}$-hard [Cap97], whereas for signed reversals, sorting by reversals is polynomial [BMY01]. For the other metrics mentioned above, *SDP* is $\mathcal{NP}$-complete [AJ08], [BCW09].

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.

In this work a proof is given of the fact that *SDP* with respect to the Cayley distance is $\mathcal{NP}$-complete. The proof follows Pinch's approach [Pin07] that is based on two polynomial reductions: from *3SAT* into the problem of finding a routing which respects a polarisation, of maximum cardinality, in a switching circuit and then, from the latter problem into *SDP* for the Cayley distance. Although Pinch's proof was published in 2007, previous drafts were available since 1992. The contribution of this work is to correct the first reduction presented in Pinch's proof in two ways: firstly, by stating correct polarised switching circuits in the reduction from *3SAT* to these circuits, specifically for clauses with two and three variables and secondly, by simplifying the width of polarisation in the switching circuits used in the first reduction.

As it will be showed, a simpler and more general proof of the $\mathcal{NP}$-completeness of the SDP which applies also to the Cayley distance was presented in [BCW09]. This proof directly reduces satisfiability problems into SDP problems without the intermediate step which relates routing problems in switching circuits with SDP problems.

In the second Section, is is proved that *SDP* for the Cayley distance is $\mathcal{NP}$-complete. In the third Section it is made explicit the flaw in Pinch's proof and presented a sketch of Bucheim's et al. proof about $\mathcal{NP}$-completeness of SDP for different measures.

II. *SDP* for the Cayley distance is $\mathcal{NP}$-complete

We will present a correct and detailed proof of this fact pointing out the problems in Pinch's work [Pin07].

Let $S \subset S_n$ be a set of permutations of the form $\gamma_j = (x_j^1 \quad y_j^1) \ldots (x_j^{r_j} \quad y_j^{r_j})$, where all x_j^i and y_j^i are different. We call S *Involutions with Disjoint Support (IDS)*. The *width* of an IDS is defined as the maximum number of 2-cycles (r_j) in its permutations (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 $G(V, E)$ 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. Each in and out-edge of v has a different label in $\{1, \ldots, \partial(v)\}$. 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. 1. Note that, there is a correspondence between routings of a switching circuit G and decompositions of the edge set into directed cycles of G.

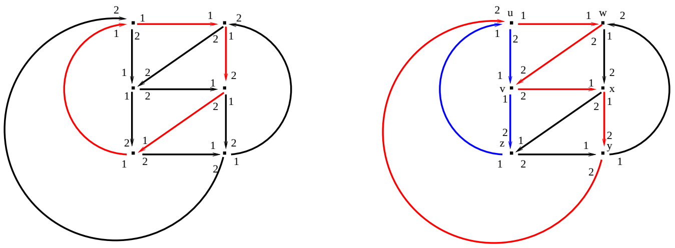

Fig. 1. The first circuit shows a decomposition in two directed cycles with routing $\rho = id$ for all vertices; the second circuit shows a decomposition in three cycles given by the routing $\rho(u) = \rho(w) = \rho(z) = (1 \quad 2)$ and $\rho(v) = \rho(x) = \rho(y) = id$

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. 1 respect polarisation with a unique equivalence class (Fig. 1 to the left) and two equivalence classes (Fig. 1 to the right). Note that, for distinct labels and routings, the decomposition into cycles changes.

The *Polarised-Switching-Circuit-Routing* (*PSCR*) is defined 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 instead 4 as presented here:

1) Prove that *Width$_4$-Valency$_2$-Routing* is $\mathcal{NP}$-complete;
2) Show the existence of an equivalence between *Width$_w$-Valency$_2$-Routing* and *IDS$_w$-Subgroup-Distance*.

Applying both these results one obtains that IDS$_4$-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.

A. Width$_4$-Valency$_2$-Routing *is $\mathcal{NP}$-complete*

The first step in Pinch's paper is in fact a attempt 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 detail in III-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, $\rho(v) = (1 \quad 2) \in S_2$, 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 a 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. 2 and 3.

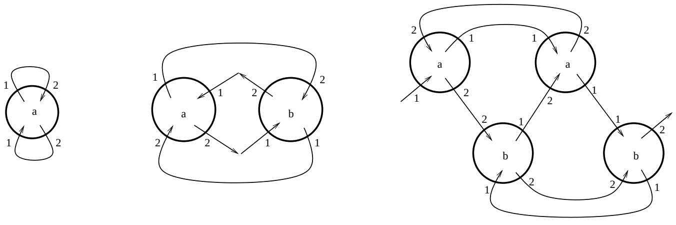

Fig. 2. Switching circuits $I(a)$, $E(a, b)$ and $F(a, b)$ for unary clauses, equality between variables and binary clauses

Proposition 2.1: Properties of the switching circuits I, E, F and A

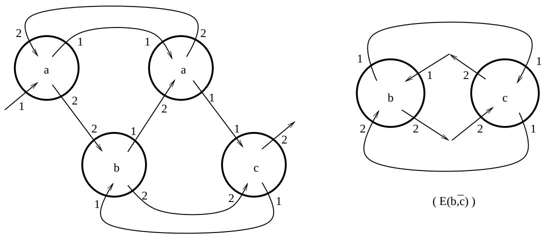

Fig. 3. Switching circuit $A(a, b, c)$ for tertiary clauses

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. 4. 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. 5. 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. 6. 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. 7. Thus $A(a, b, c)$ will have four circuits. □

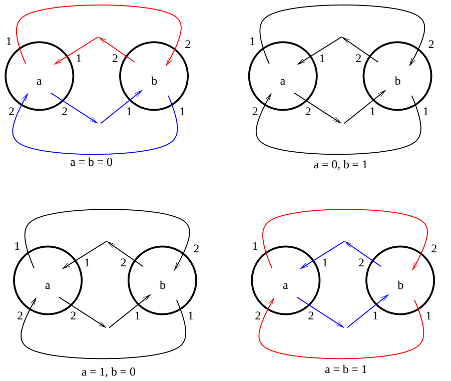

Fig. 4. Cycles of circuit $E(a, b)$ for $a = b$ and $a \neq b$

Theorem 2.2: There is a polynomial reduction from *3-SAT* to *Width*$_4$*-Valency*$_2$*-Routing*.

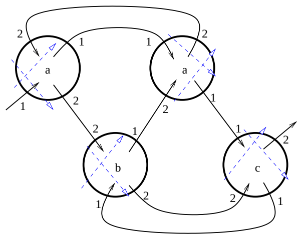

Fig. 5. Cycle of left circuit of $A(a, b, c)$ for $a = b = c = 0$

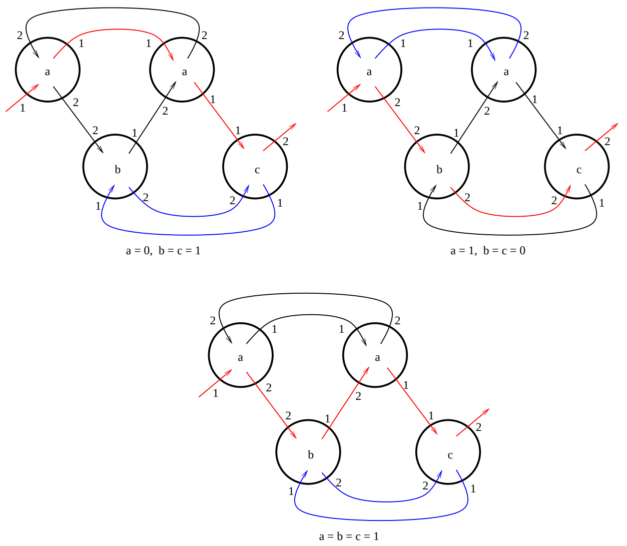

Fig. 6. Cycles of left circuit of $A(a, b, c)$ for $a \neq b = c$ and $a = b = c = 1$

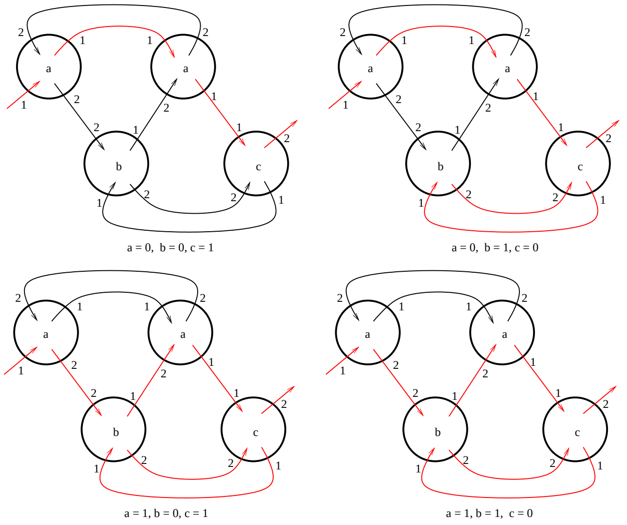

Fig. 7. Cycles of left circuit of $A(a, b, c)$ for $b \neq c$

Proof: Let φ an instance of *3-SAT* that is a Boolean formula over variables $x_1, \ldots, x_n$, 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 $(y_j^1 \equiv y_j^2) \wedge \ldots \wedge (y_j^{(r_i-1)} \equiv y_j^{r_i})$ 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 number 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 labeled 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 of the form 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 2.1, 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. □

B. Width$_w$-Valency$_2$-Routing *problem polynomially reduces to* IDS$_w$-Subgroup-Distance

In Pinch's work it is proved in fact a polynomial equivalence between both problems.

Theorem 2.3 ([Pin07]): There is a polynomial equivalence between the *Width$_w$-Valency$_2$-Routing* and *IDS$_w$-Subgroup-Distance* problems.

To understand this equivalence, consider $(G(V, E), T)$, a polarised switching circuit, where each vertex has valency two and each equivalence class has width at most w. Let $|E| = n$ and associate a different number in $\{1, \ldots, n\}$ to each edge. Construct a permutation π as follows: for each edge e, let v be the vertex such that e is an input edge in v and define $\pi(e)$ as the edge f out of v such that the labels of e and f as an input and an output edge of v are equal. Construct an instance of the *IDS* problem from G as follows: for each equivalence class in T, $C_j = \{v_j^i \mid i = 1, \ldots, r_j\}$, let γ_j be a generator given as the product of transpositions $(f_j^i \quad g_j^i)$, 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 $\eta \in H = \langle\{\gamma_j\}\rangle$. The correspondence between the problems is understood, based on the observation that a transposition can split a cycle permutation at most into two cycles, from which 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.

This construction also helps to understand how to build a corresponding polarized switching circuit from an instance of the *IDS$_w$-Subgroup-Distance* problem.

To conclude the $\mathcal{NP}$-completeness, it is necessary to prove that *SDP* restricted to the Cayley distance is in $\mathcal{NP}$. For this, some remarks will be done.

Given two permutations $\pi, \sigma \in S_n$, if the Cayley distance between π and σ is k, denoted as $d(\pi, \sigma) = k$, then there is a sequence of k transpositions $\rho_1, \ldots, \rho_k$, such that $\pi\rho_1 \ldots \rho_k = \sigma$; or equivalently, $\rho_1 \ldots \rho_k = \pi^{-1}\sigma$. Thus, calculating the Cayley distance between two permutations π and σ is equivalent to decompose the permutation $\pi^{-1}\sigma$ as a minimum product of transpositions.

Theorem 2.4 ([Mac95]): A permutation in S_n cannot be written as the product of fewer than $n - r$ transpositions, where r is the number of disjoint cycles in the permutation.

For example, consider the permutation $\pi = (12)(345) \in S_5$; it consists of two disjoint cycles. Thus, by Theorem 2.4, at least three transpositions are necessary to represent this permutation. Namely, $\pi = (12)(34)(35)$.

Proposition 2.5: Given a cycle $(\pi_1 \ldots \pi_t)$, one always can write it as the product of $t - 1$ transpositions.

In fact, observe that $(\pi_1 \ldots \pi_t) = (\pi_1\pi_2)(\pi_1\pi_3) \ldots (\pi_1\pi_t)$.

Consider a permutation $\pi \in S_n$ consisting of k disjoint cycles; this is a permutation $\pi = \pi_1 \ldots \pi_k$, where each $\pi_i, 1 \leq i \leq k$, corresponds to a cycle in π, and whenever $l \in \{1, \ldots, n\}$ is in cycle π_i, this element is not in cycle π_j, for $j \neq i$. Denote as $n_i, 1 \leq i \leq k$, the length of cycle π_i. Note that $n_1 + \ldots + n_k = n$. By Proposition 2.5, each cycle π_i of π can be decomposed as the product of $n_i - 1$ distinct transpositions. Thus, the permutation π can be decomposed in $n_1 - 1 + \ldots + n_k - 1 = n - k$ transpositions. By Theorem 2.4, this is the minimum number of transpositions in which π can be written.

Now, consider a permutation $\pi \in S_n$, a set of generators of a subgroup H of S_n and an integer k. Non deterministically, choose a permutation $\sigma \in H$. Decompose the permutation $\pi^{-1}\sigma$ as it is done in the Proposition 2.5. This polynomial procedure checks whether the Cayley distance between π and the choosed permutation is smaller or equal than k. Repeatedly application of this non-deterministic polynomial verification procedure is applied for computing the Cayley distance. This concludes the proof that *SDP* restricted to Cayley distance is in $\mathcal{NP}$.

III. Related work

Although Pinch's proof was available since 1992, it was published only in 2007 and subsequently referenced by Buchheim et al. [BCW09] without any mention to the flaws reported in this paper and detailed in Subsection III-A. Even, more recently, after the publication of the elegant proof developed by Buchheim et al., that will be detailed in Subsection III-B, other authors have referenced Pinch's proof without mentioning these flaws. Among the papers that referenced Pinch's proof attempt, one can mention [CW07] and [CW10], whose main subject is the $\mathcal{NP}$-completeness of the *Weight Problem* restricted to several distances over S_n, where the *Weight Problem*, with respect to the distance d, consists in, given generators for a group G and an integer k, find an element $g \in G$ such that $d(g, e) = k$, where e is the identity permutation. In these two papers, as in [BCW09], the Hamming, Cayley, l_p, l_∞, Lee, Kendall's tau and Ulam distances are considered.

Also, in Bogaerts' Thesis [Bog09], Pinch's work is referenced. One of the main objectives of this work is, given a fixed number n and a distance d, to study the maximum length of a permutation code. Bogaerts asserts that several problems are related with decoding a permutation code, among them, the *SDP* problem.

A. Flaws in Pinch's proof attempt

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. 8 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.

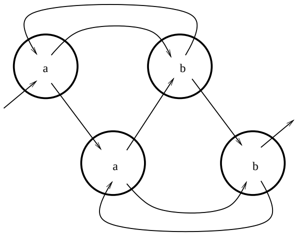

Fig. 8. Switching circuit $F'(a, b)$

Proposition 3.1: There is no possible labeling 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. 9 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.

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.

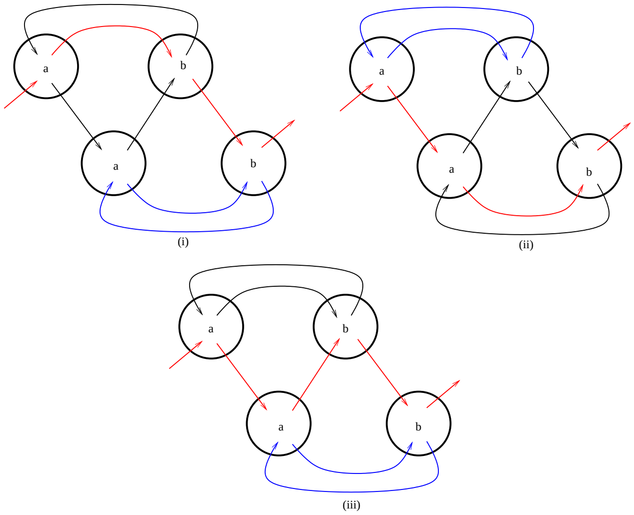

Fig. 9. Possible decompositions in three cycles of $F'(a, b)$ for the routing $a = b = 1$

Changing the routing from $a = b = 1$ to $a = b = 0$ in each of these cases gives the decomposition in cycles depicted in Fig. 10, 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.

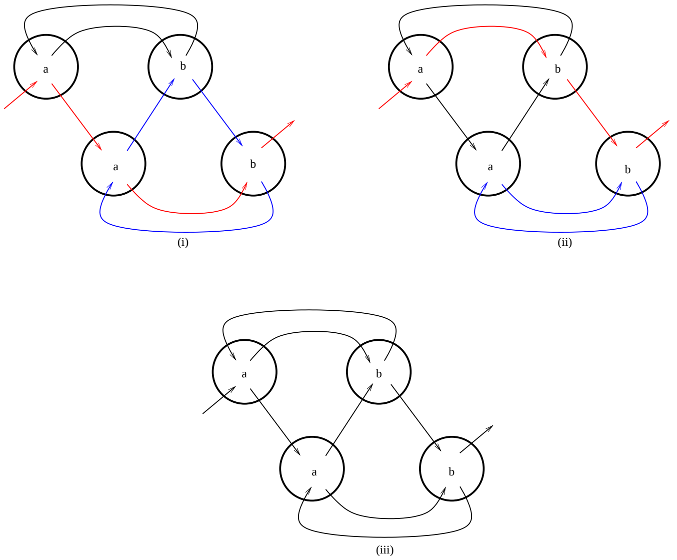

Fig. 10. Cycle decompositions of $F'(a, b)$ for the routing $a = b = 0$

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. □

B. A general proof of $\mathcal{NP}$-completeness for SDP

A simpler and more general proof of the $\mathcal{NP}$-completeness of the SDP which applies also to the Cayley distance was given in [BCW09]. Here, it is important to stress that although this proof is simpler it relates SDP only with satisfiability and because of this it is relevant the result presented in this work establishing a correct relation of SDP with routing problems.

To prove that the SDP restricted to the Cayley distance is $\mathcal{NP}$-complete one reduces the problem of maximum SAT with clauses of length two, *MAX-2-SAT* to it. *MAX-2-SAT* is well-known to be $\mathcal{NP}$-complete. Below the reduction from this problem to SDP for the Cayley distance is sketched.

Consider an instance of *MAX-2-SAT*: given an integer K and a formula φ, in conjunctive normal form, consisting of p variables, $u_1, \ldots, u_p$ and q clauses, $c_1, \ldots, c_q$, each clause of length two. The problem is to decide whether there exists a truth assignment for the variables such that, at least, K clauses are satisfied.

Firstly, one constructs a permutation π on a domain $X = \cup_{i=1..p} X_i \cup \cup_{j=1..q} Y_j$, where: for each variable u_i, we associate a set of size $6q+2$, $X_i = \{x_{i,1}, \ldots, x_{i,6q+2}\}$. The elements of X_i are swapped pairwise by π. Thus, the factor $\Pi_{i,j,k}(x_{i,j}\, x_{i,k})$, where $1 \leq i \leq p$; $j,k \in \{1, \ldots 6q+2\}$ and each $x_{i,j}$ appears exactly once in the multiplicand, is in the cycle structure of π. Additionally, for each clause $j = 1..q$, $Y_j = \{a_{j,1}, \ldots, a_{j,6}\}$, such that π acts on Y_j as the permutation $(a_{j,1}\, a_{j,2})(a_{j,3}\, a_{j,4})(a_{j,5}\, a_{j,6})$. Observe that the size of X, $p(6q+2)+6q$, it is polynomial on the length of φ.

Secondly, one defines the generators of group H as follows: for each variable u_i, consider two permutations $\pi_i(t)$ and $\pi_i(f)$. Both, $\pi_i(t)$ and $\pi_i(f)$, agree with permutation π on X_i. If u_i appears positively in the first position of a clause c_j, then $\pi_i(t)$ exchange $a_{j,1}$ with $a_{j,2}$ and $a_{j,3}$ with $a_{j,4}$; if u_i appears positively in the second position, then $pi_i(t)$ exchange $a_{j,1}$ with $a_{j,2}$ and $a_{j,5}$ with $a_{j,6}$. If a variable u_i appears negated, the same is done by $\pi_i(f)$ instead of $\pi_i(t)$. The remaining elements of X are fixed by $\pi_i(t)$ and $\pi_i(f)$. Let $H = \langle \pi_i(t), \pi_i(f) | i = 1..p \rangle$ be the group generated by $\pi_i(t)$ and $\pi_i(f)$ and let $K' = 3q - 2K$.

Finally, to conclude, one proves that $d(\pi, H) \leq K$‘ if, and only if, at least K clauses are satisfiable. Consider $t : \{u_1, \ldots, u_p\} \rightarrow \{0,1\}$, a truth assignment satisfying, at least, K clauses. Let $\tau = \Pi_{t(u_i)=1}\pi_i(t)\Pi_{t(u_i)=0}\pi_i(f) \in H, 1 \leq i \leq p$. Note that, the permutation τ acts on each X_i as the permutation π. Furthermore, one can verify, by analysis of the behavior of τ on each set Y_j, that the distance between π and τ is 1 if the clause c_j is satisfied and 3, otherwise. Since, at least, K clauses are satisfied, thus $d(\pi, \tau) \leq 3(q-K) + 1K = 3q - 2K = K'$. Conversely, suppose $d(\pi, H) \leq K'$. Thus, there is $\tau \in H$ such that $d(\pi, \tau) \leq K'$. Note that, if, for any variable u_i, both generators, $\pi_i(t)$ and $\pi_i(f)$ simultaneously either appear or not in the cycle structure of τ, then the distance between τ and π on X_i would be $6q+2 > K'$. Therefore, exactly one of $\pi_i(t)$ or $\pi_i(f)$ appear in τ, for each variable u_i. Define $t : \{u_1, \ldots, u_p\} \rightarrow \{0,1\}$, where $t(u_i) = 1$ if, and only if, $\pi_i(t)$ is in the composition of τ. In this way, τ agree with π on each set X_i. Thus, $d(\pi, \tau) \leq K' = 3q - 2K$ in case that at least K clauses were satisfied.

This elegant proof was introduced in [BCW09] for the Hamming distance. For that distance one define $K' = 6q - 4K$. Let, $K' = bq - (b-a)K$ and one can obtain, only by changing the parameters a and b, proofs of $\mathcal{NP}$-completeness of SDP for l_p, Lee, Kendall's tau and Ulam distances. The proof presented in this section for the Cayley distance, is obtained setting these parameters as $a = 1$ and $b = 3$.

IV. 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 or not $\mathcal{NP}$-complete.

References

[AJ08] Vikraman Arvind and Pushkar S. Joglekar. Algorithmic problems for metrics on permutation groups. In *Proc. 34th Conference on Current Trends in Theory and Practice of Computer Science SOFSEM*, volume 4910 of *Lecture Notes in Computer Science*, pages 136–147, 2008.

[BCW09] Christoph Buchheim, Peter J. Cameron, and Taoyang Wu. On the subgroup distance problem. *Discrete Mathematics*, 309(4):962–968, 2009.

[BMY01] David A. Bader, Bernard M. E. Moret, and Mi Yan. A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. In Frank K. H. A. Dehne, Jörg-Rüdiger Sack, and Roberto Tamassia, editors, *WADS*, volume 2125 of *Lecture Notes in Computer Science*, pages 365–376. Springer, 2001.

[Bog09] Mathieu Bogaerts. *Codes et tableaux de permutations: construction, énumération et automorphismes*. PhD thesis, Université Libre de Bruxelles, Faculté des sciences, Département de Mathématiques, Bruxelles, 2009.

[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.

[CW07] Peter J. Cameron and Taoyang Wu. The complexity of the weight problem for permutation groups. *Electronic Notes in Discrete Mathematics*, 28:109–116, 2007.

[CW10] Peter J. Cameron and Taoyang Wu. The complexity of the weight problem for permutation and matrix groups. *Discrete Mathematics*, 310(3):408–416, 2010.

[Mac95] George Mackiw. Permutations as products of transpositions. *The American Mathematical Monthly*, 102(5):pp. 438–440, 1995.

[Pin07] Richard G. E. Pinch. The Distance of a Permutation from a Subgroup of S_n. In G. Brightwell, I. Leader, A. Scott, and A. Thomason, editors, *Combinatorics and Probability*, pages 473–480. Cambridge University Press, 2007.