On the Subgroup Distance Problem

Christoph Buchheim[1], Peter J. Cameron[2], and Taoyang Wu[3]

[1] Department of Computer Science, University of Cologne, Germany
buchheim@informatik.uni-koeln.de
[2] School of Mathematical Sciences, Queen Mary, University of London, U.K.
P.J.Cameron@qmul.ac.uk
[3] Department of Computer Science & School of Mathematical Sciences,
Queen Mary, University of London, U.K.
Taoyang.Wu@dcs.qmul.ac.uk

Abstract. We investigate the computational complexity of finding an element of a permutation group $H \subseteq S_n$ with a minimal distance to a given $\pi \in S_n$, for different metrics on S_n. We assume that H is given by a set of generators, such that the problem cannot be solved in polynomial time by exhaustive enumeration. For the case of the Cayley Distance, this problem has been shown to be NP-hard, even if H is abelian of exponent two [7]. We present a much simpler proof for this result, which also works for the Hamming Distance, the l_p distance, Lee's Distance, Kendall's tau, and Ulam's Distance. Moreover, we give an NP-hardness proof for the l_∞ distance using a different reduction idea. Finally, we settle the complexity of the corresponding fixed-parameter and maximization problems.

1 Introduction

For any metric d on S_n, we define the distance of a permutation $\pi \in S_n$ from a subgroup H of S_n as

$$d(\pi, H) = \min_{\tau \in H} d(\pi, \tau) \ .$$

If H is not given explicitly, but by a set of generators, one can still decide in polynomial time whether $d(\pi, H) = 0$, i.e., whether $\pi \in H$ [8, 4]. However, it is not possible in general to compute the distance $d(\pi, H)$ in polynomial time, unless $\mathrm{P} = \mathrm{NP}$. This was shown by Pinch for the Cayley distance [7]. More precisely, he showed that the following problem is NP-complete in this case:

Problem 1 (Subgroup distance problem). Given $\pi \in S_n$, a set of generators of a subgroup H of S_n, and an integer K, decide whether $d(\pi, H) \leq K$.

In this paper, we present an alternative proof for this result. Furthermore, we show NP-completeness for several other well-known metrics on S_n, namely the Hamming Distance, the l_p distance, the l_∞ distance, Lee's Distance, Kendall's tau, and Ulam's Distance. We list the corresponding definitions in the following, but recommend [3] for further information.

– the *Hamming Distance* between two permutations π and τ is the number of different entries, i.e., $|\{i \mid \pi(i) \neq \tau(i)\}|$
– the *Cayley Distance* is defined as the minimum number of transpositions taking π to τ
– for $p \geq 1$, the l_p *distance* is defined by $\sqrt[p]{\sum_{i=1}^{n}(\pi(i)-\tau(i))^p}$
– the l_∞ *distance* is defined as $\max_{1\leq i\leq n}|\pi(i)-\tau(i)|$
– the *Lee Distance* is $\sum_{i=1}^{n}\min(|\pi(i)-\tau(i)|, n-|\pi(i)-\tau(i)|)$
– *Kendall's tau* is the minimum number of pairwise adjacent transpositions taking π to τ
– *Ulam's Distance* is defined as n minus the length of a longest increasing subsequence in $(\tau\pi^{-1}(1),\ldots,\tau\pi^{-1}(n))$

The subgroup distance problem is related to the weight problem, where one asks for an element $\tau \in H$ with a given distance k to the identity. This problem has been investigated by Cameron and Wu, who showed NP-completeness for all metrics listed above [2]. If the weight problem is restricted to the Hamming distance and to the case $k = n$, the resulting problem is to decide whether the group H contains a fixed-point free permutation, which has been shown to be NP-complete by Buchheim and Jünger [1].

2 NP-completeness for the Hamming Distance

In this section, we prove that the subgroup distance problem is NP-complete for the Hamming distance.

Theorem 1. *The subgroup distance problem for the Hamming distance is NP-complete, even if the permutation group H is abelian of exponent two.*

Proof. For the reduction, we use the decision version of the maximum satisfiability problem with clauses of length two (MAX-2-SAT), which is well-known to be NP-complete [5]. So consider an instance of MAX-2-SAT, consisting of an integer K', of p variables $u_1,\ldots,u_p$ and of q clauses $c_1,\ldots,c_q$ of length two. It is to decide whether there is a truth assignment $\{u_1,\ldots,u_p\} \to \{0,1\}$ satisfying at least K' clauses.

In order to transform this instance to an instance of the subgroup distance problem, first define $K = 6q - 4K'$. Moreover, construct a domain X and a permutation π as follows: for every variable u_i, introduce a set X_i with $6q+2$ points. These points are swapped pairwisely by π. For each clause j, add a set $Y_j = \{a_{j,1},\ldots,a_{j,6}\}$ such that π exchanges $a_{j,1}$ with $a_{j,2}$, $a_{j,3}$ with $a_{j,4}$, and $a_{j,5}$ with $a_{j,6}$. The total size of the domain X, defined as the union of all sets X_i and Y_j, is $p(6q+2)+6q$ and thus polynomial.

Next we define generators for the group H. For each variable u_i, we define two generators $\pi_i(t)$ and $\pi_i(f)$. Both exchange all points in X_i in the same way as π. If u_i appears without negation in the first position of a clause c_j, then $\pi_i(t)$ exchanges $a_{j,1}$ with $a_{j,2}$ and $a_{j,3}$ with $a_{j,4}$; if it appears without negation in the second position, then it exchanges $a_{j,1}$ with $a_{j,2}$ and $a_{j,5}$ with $a_{j,6}$. For a negated appearance, the same is done by $\pi_i(f)$ instead of $\pi_i(t)$. All other points are fixed by $\pi_i(t)$ and $\pi_i(f)$.

Now let $H = \langle \pi_i(t), \pi_i(f) \mid i = 1, \ldots, p \rangle$. It remains to show that $d(\pi, H) \leq K$ if and only if K' clauses from $c_1, \ldots, c_q$ can be simultaneously satisfied. First, let $t \colon \{u_1, \ldots, u_p\} \to \{0, 1\}$ be a truth assignment satisfying at least K' clauses. Consider

$$\tau = \prod_{t(u_i)=1} \pi_i(t) \prod_{t(u_i)=0} \pi_i(f) \in H \ .$$

By construction, τ agrees with π on each X_i. Moreover, it is readily verified that on the clause gadget Y_j the distance between τ and π is 2 if c_j is satisfied by t and 6 otherwise. In summary, we have

$$d(\pi, \tau) = 6q - 4 \cdot |\{j \mid c_j \text{ is satisfied by } t\}| \leq 6q - 4K' = K \ .$$

Now assume that $d(\pi, H) \leq K$. Choose $\tau \in H$ with $d(\pi, \tau) \leq K$. In the composition of τ, exactly one of the generators $\pi_i(t)$ or $\pi_i(f)$ must appear, for each variable u_i. Indeed, as both $\pi_i(t)$ and $\pi_i(f)$ are involutions and H is abelian, we can assume that at most one copy of each appears. If for some variable u_i both $\pi_i(t)$ and $\pi_i(f)$ or neither one appeared, the distance between τ and π on X_i would be $6q + 2 > K$.

We can thus define $t \colon \{u_1, \ldots, u_p\} \to \{0, 1\}$ by setting $t(u_i) = 1$ if and only if $\pi_i(t)$ appears in the composition of τ. Moreover, this implies that the Hamming distance between τ and π on all sets X_i is zero. Arguing as above, one can show that t must satisfy at least K' clauses. □

3 NP-completeness for the l_∞ distance

In this section, we will prove that the subgroup distance problem is NP-complete also for the l_∞ distance. For the weight problem, we know that the l_∞ distance behaves differently from all other metrics considered in this paper, see [2]. The same is true for the subgroup distance problem, as we need a new method to prove its NP-completeness. The difference between l_∞ and other metrics will become even clearer in Sections 5 and 6.

Theorem 2. *The subgroup distance problem for the l_∞ distance is NP-complete, even if the permutation group H is abelian of exponent two.*

Proof. We construct a polynomial-time reduction from the NAE-3-SAT problem. This problem is NP-complete [6]; it is similar to the well known 3-SAT problem but requires a truth assignment such that in no clause all three literals are equal in truth value. More precisely, an instance of the NAE-3-SAT problem is given by p variables $u_1, \ldots, u_p$ and q clauses $c_1, \ldots, c_q$ of length three. The question is whether there is a truth assignment $\{u_1, \ldots, u_p\} \to \{0, 1\}$ such that no clause has all literals true, or all literals false.

In order to transform this instance into an instance of the subgroup distance problem, firstly define $K = 2$. Construct a domain X and a permutation π as follows: for every variable u_i, introduce a set X_i containing 6 points. We may assume that the points in X_i are labelled from 1 to 6. Then π is acting on X_i as $(1,4)(2,5)(3,6)$. For each clause c_j, we use a gadget Y_j of size 4. To simplify the notation, assume Y_j contains nodes $1, 2, 3, 4$. Then π is acting on Y_j as $(1,4)(2)(3)$. Now the domain X is the disjoint union of all X_i and Y_j with suitable adjustments of the labelling.

Next we define generators for the group H. For each variable u_i, we define two generators $\pi_i(t)$ and $\pi_i(f)$. Both exchange all points in X_i in the same way as π. If u_i appears without negation in the first position of a clause c_j, then the action on Y_j is $\pi_i(t) = (1,2)(3,4)$, where we assume again that the points in Y_j are labelled 1 to 4. If it appears without negation in the second position then $\pi_i(t) = (1,3)(2,4)$; if it appears without negation in the third position then we have $\pi_i(t) = (1,4)(2,3)$. For a negated appearance, the same is done by $\pi_i(f)$ instead of $\pi_i(t)$. All other points are fixed by $\pi_i(t)$ and $\pi_i(f)$.

Let $H = \langle \pi_i(t), \pi_i(f) \mid i = 1, \ldots, p \rangle$. It remains to show that $l_\infty(\pi, H) \leq K$ if and only if there exists an assignment such that no clause from $c_1, \ldots, c_q$ has all literals true, or all literals false. First, let $t \colon \{u_1, \ldots, u_p\} \to \{0, 1\}$ be a truth assignment satisfying the requirement of NAE-3-SAT. Consider

$$\tau = \prod_{t(u_i)=1} \pi_i(t) \prod_{t(u_i)=0} \pi_i(f) \in H \ .$$

By construction, τ agrees with π on each X_i. Moreover, one can verify that on the clause gadget Y_j, the l_∞ distance between τ and π is 3 if all literals in c_j have the same truth value with respect to t. Indeed, in this case the induced action of τ on Y_j is trivial. In all other cases, the distance between τ and π is either 1 or 2. In summary, we have $l_\infty(\pi, \tau) \leq 2$ if and only if in all clauses either one or two literals are satisfied.

Now assume that $l_\infty(\pi, H) \leq K = 2$. Choose $\tau \in H$ with $l_\infty(\pi, \tau) \leq 2$. In the composition of τ, exactly one of the generators $\pi_i(t)$ or $\pi_i(f)$ must appear, for each variable u_i. Indeed, as both $\pi_i(t)$ and $\pi_i(f)$ are involutions and H is abelian, we can assume that at most one copy of each appears. If for some variable u_i both $\pi_i(t)$ and $\pi_i(f)$ or neither one appeared, the distance between τ and π on X_i would be $3 > K$.

We can thus define $t \colon \{u_1, \ldots, u_p\} \to \{0, 1\}$ by setting $t(u_i) = 1$ if and only if $\pi_i(t)$ appears in the composition of τ. Moreover, this implies that the l_∞ distance between τ and π on all sets X_i is zero. Arguing as above, one can show that t must satisfy the condition that in all clauses either one or two literals are satisfied. $\square$

4 NP-completeness for other distances

We next show that the subgroup distance problem is NP-complete for all other considered metrics, following the ideas in the proof of Theorem 1. Indeed, we can use exactly the same construction of variable and clause gadgets and the same definition of π. We just have to label the points $a_{j,1}$ to $a_{j,6}$ consecutively in each clause gadget Y_j.

The crucial point in the proof of Theorem 1 is the following: let x_1 (x_2) denote the action on Y_j induced by a satisfied literal in the first (second) position of a clause. Then $d(x_1, \pi) = d(x_2, \pi) = d(x_1x_2, \pi) = 2$ while $d(\pi, e) = 6$. Roughly speaking, if for any metric d we have $d(x_1, \pi) = d(x_2, \pi) = d(x_1x_2, \pi) = a$ while $d(\pi, e) = b$ for some $0 < a < b$, then we can carry over the proof of Theorem 1 with no other change than redefining $K = bq - (b - a)K'$ and making sure that the distance between π and e on each variable gadget is at least bq.

It is readily verified that the parameters a and b can be chosen as follows for the remaining distances defined in the introduction:

distance	a	b
Cayley	1	3
l_p	$\sqrt[p]{2}$	$\sqrt[p]{6}$
Lee	2	6
Kendall's tau	1	3
Ulam	1	3

In summary, we derive

Theorem 3. *The subgroup distance problem is NP-complete for the Cayley Distance, the l_p distance, Lee's Distance, Kendall's tau, and Ulam's Distance, even if the permutation group H is abelian of exponent two.*

For the case of the Cayley Distance, this result has been proved recently by Pinch [7]. However, the proof given here is much simpler.

5 Fixed parameter K

It is worthwhile to have a look at the variant of the subgroup distance problem where the parameter K is considered a constant rather than part of the input. Interestingly, the problem then becomes polynomial for most metrics discussed above. Indeed, as a membership test can be performed in polynomial time [8, 4], it suffices to show that the set

$$X = \{\tau \in S_n \mid d(\tau, e) \leq K\}$$

can be enumerated in polynomial time. For most metrics considered above, this can easily be seen to hold: for the Hamming distance, at most K positions may differ between e and any $\tau \in X$. This implies

$$|X| \leq \binom{n}{K} n^K \in O(n^{2K}) \; .$$

This bound also holds for the Lee distance, as it is always greater or equal to the Hamming distance. Next observe that the Cayley distance is always at least half the Hamming distance, while the l_p distance is at least the p-th root of the Hamming distance. Thus $|X|$ is bounded polynomially for these distances as well. The same result follows for Kendall's tau, which is greater or equal to the Cayley distance by definition. Finally, for Ulam's distance the set X contains all permutations τ such that there is an increasing subsequence in $(\tau^{-1}(1), \tau^{-1}(2), \ldots, \tau^{-1}(n))$ of length at least $n - K$. All these permutations can be constructed by choosing $n-K$ positions, then $n-K$ numbers to be placed increasingly on these positions, and an arbitrary permutation of the remaining K numbers. Thus we derive

$$|X| \leq \binom{n}{K}\binom{n}{K} K! \in O(K!\, n^{2K}).$$

To sum up, we have the following theorem:

Theorem 4. *The subgroup distance problem is in P for the Hamming Distance, the Cayley Distance, the l_p distance, Lee's Distance, Kendall's tau, and Ulam's Distance when K is fixed.*

For the l_∞ distance, the set X actually has exponential size, even for $K = 1$: assume that n is even and consider pairs $\{1, 2\}$, $\{3, 4\}$, ..., $\{n-1, n\}$. Then each pair can be swapped independently without increasing the distance to e beyond one, so that X contains at least $2^{n/2}$ points. This is in contrast to all other metrics, where these swaps would add up (in different ways). In fact, the proof of Theorem 2 shows that for the l_∞ distance the subgroup distance problem remains NP-complete even if K is fixed to 2. Using an appropriate relabelling, the same proof shows NP-completeness for every fixed $K \geq 2$. If we drop the restriction that H is of exponent two, we can even show NP-completeness for $K = 1$.

Theorem 5. *The subgroup distance problem for the l_∞ distance is NP-complete for each fixed $K \geq 1$, even if the permutation group H is abelian.*

Proof. It remains to examine the case $K = 1$. As in the proof of Theorem 2, we use a reduction from NAE-3-SAT. The variable gadgets now contain 4 points each such that π exchanges $(1, 3)$ and $(2, 4)$. The clause gadgets are defined as follows: for every clause c_j, the set Y_j consists of three points $a_{j,1}, a_{j,2}, a_{j,3}$, numbered consecutively. The permutation π exchanges $a_{j,1}$ and $a_{j,3}$.

If u_i appears without negation in any position of a clause c_j, then $\pi_i(t)$ permutes $(a_{j,1}, a_{j,2}, a_{j,3})$ cyclically. For a negated appearance, the same is done by $\pi_i(f)$ instead of $\pi_i(t)$.

As above, let τ be the permutation corresponding to a truth assignment t. Then the action of τ on Y_j is trivial if and only if either none or all of the literals in c_j are satisfied, and it is cyclic otherwise. It is easy to check that the l_∞ distance between τ and π on Y_j is 2 in the first case and 1 in the second. □

6 Maximum subgroup distance

So far we have considered the problem of finding a member of H with minimal distance to the given permutation π. Alternatively, one might ask for an element with a *maximal* distance from π. The corresponding decision problem is

Problem 2 (Maximum subgroup distance problem). Given a permutation $\pi \in S_n$, a set of generators of a subgroup H of S_n, and an integer K, decide whether there is a $\tau \in H$ with $d(\pi, \tau) \geq K$.

This problem is *not* symmetric to the (minimum) subgroup distance problem in general. The complexity status of the maximization version cannot be derived from the complexity status of the minimization version. In particular, we will show that the maximum subgroup problem for the l_∞ distance can be solved in polynomial time, while we saw above that the minimum subgroup distance problem is NP-complete for the same metric.

Theorem 6. *Given $\pi, \tau_1, \ldots, \tau_m \in S_n$, we can find some $\tau \in H = \langle \tau_1, \ldots, \tau_m \rangle$ maximizing $l_\infty(\pi, \tau)$ in polynomial time.*

Proof. We have to find a permutation $\tau \in H$ such that

$$\max_{1 \leq i \leq n} |\pi(i) - \tau(i)|$$

is maximized. Since this value is a priori bounded by n, we can iteratively check for $k = n, n-1, \ldots, 1$ whether H contains an element τ such that $|\pi(i) - \tau(i)| = k$ for some point i. The latter holds if and only if there exists a pair of points (i, j) with $|\pi(i) - j| = k$ and i and j belong to the same orbit of H. The last property can be checked in polynomial time, and the first found τ mapping i to j can be returned. □

We now argue that the maximum subgroup distance problem is NP-complete for all other metrics considered above. The idea of the proof for the minimization variant can be carried over with few changes. The variable and clause gadgets are the same as before. However, the permutation π is now chosen as the identity e, instead of exchanging points pairwisely on the variable and clause gadgets. This makes sure that for every variable exactly one of the two generators $\pi_i(f)$ or $\pi_i(t)$ is chosen.

Basically, it remains to prove that with these definitions on the clause gadgets we have $d(x_1, \pi) = d(x_2, \pi) = d(x_1 x_2, \pi) = a > 0$; note that $d(\pi, e) = 0$ on each clause gadget by definition of a metric. The corresponding numbers are:

distance	a
Hamming	4
Cayley	2
l_p	$\sqrt[p]{4}$
Lee	4
Kendall's tau	2
Ulam	2

Theorem 7. *The maximum subgroup distance problem is NP-complete for the Hamming Distance, the Cayley Distance, the l_p distance, Lee's Distance, Kendall's tau, and Ulam's Distance, even if the permutation group H is abelian of exponent two.*

In contrast to the minimal subgroup distance problem, which is NP-complete for the l_∞ distance when K is fixed, the fixed parameter version of the maximal subgroup distance problem is in P for *all* metrics considered in this paper. Firstly, it is easy to see from the above discussion that this holds for l_∞, so consider the remaining metrics. Following the arguments of Section 5, we know that the following set is of polynomial size for all metrics except l_∞:

$$X = \{\tau \in S_n \mid d(\tau, \pi) \leq K - 1\}$$

Now to check whether there is an element $\tau \in H$ with $d(\tau, \pi) \geq K$, we can equivalently check whether $H \not\subseteq X$, i.e., whether $|H| \neq |H \cap X|$. As X can be enumerated in polynomial time and a membership test for H can be performed in polynomial time, the size of $H \cap X$ can be computed in polynomial time for any $\pi \in S_n$ and any H. On the other hand, the size of H can be computed in polynomial time as well. That is, we have the following theorem:

Theorem 8. *The maximal subgroup distance problem is in P for the Hamming Distance, the Cayley Distance, the l_p distance, the l_∞ distance, Lee's Distance, Kendall's tau, and Ulam's Distance when K is fixed.*

In fact, the above theorem not only gives us the answer for the decision problem, but also provides us an algorithm to find a permutation with distance at least K, if one exists. The key observation is that we can associate an order on the group H via its Cayley graph w.r.t the generators. Therefore we can enumerate the elements in H and compute their distance to π. We stop at the first one whose distance to π is at least K. This algorithm will terminate in polynomial time because the set $H \cap X$ is of polynomial size.

References

1. C. Buchheim and M. Jünger. Linear optimization over permutation groups. *Discrete Optimization*, 2(4):308–319, 2005.
2. P. Cameron and T. Wu. The complexity of the weight problem for permutation groups. *Electronic Notes in Discrete Mathematics*, 2006. to appear.
3. P. Diaconis. *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics, 1988.
4. M. L. Furst, J. Hopcroft, and E. M. Luks. Polynomial time algorithms for permutation groups. In *Proc. 21st IEEE Foundations of Computer Science*, pages 36–41, 1980.
5. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman, 1979.
6. C. H. Papadimitriou. *Computational Complexity.* Addison-Wesley Publishing Company, 1994.

7. R. G. E. Pinch. The distance of a permutation from a subgroup of S_n. *Combinatorics, Probability and Computing*, 2006. To appear.
8. C. C. Sims. Computation with permutation groups. In S. R. Petrick, editor, *Proc. 2nd Symp. on Symbolic and Algebraic Manipulation*, pages 23–28, 1971.