On the Complexity of Diameter and Related Problems in Permutation Groups

Markus Lohrey
Universität Siegen, Germany

Andreas Rosowski
Universität Siegen, Germany

Abstract

We prove that it is Π_2^{P}-complete to verify whether the diameter of a given permutation group $G = \langle A \rangle$ is bounded by a unary encoded number k. This solves an open problem from a paper of Even and Goldreich, where the problem was shown to be NP-hard. Verifying whether the diameter is exactly k is complete for the class consisting of all intersections of a Π_2^{P}-language and a Σ_2^{P}-language. A similar result is shown for the length of a given permutation π, which is the minimal k such that π can be written as a product of at most k generators from A. Even and Goldreich proved that it is NP-complete to verify, whether the length of a given π is at most k (with k given in unary encoding). We show that it is DP-complete to verify whether the length is exactly k. Finally, we deduce from our result on the diameter that it is Π_2^{P}-complete to check whether a given finite automaton with transitions labelled by permutations from S_n produces all permutations from S_n.

2012 ACM Subject Classification Theory of computation → Problems, reductions and completeness

Keywords and phrases algorithms for finite groups, diameter of permutation groups, rational subsets in groups

Digital Object Identifier 10.4230/LIPIcs.ICALP.2023.134

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Funding Both authors were supported by the DFG research project Lo748/12-2.

1 Introduction

Algorithmic problems for finite groups, in particular permutation groups, are an active research at the borderline between mathematics and theoretical computer science. Among the many applications of permutation group algorithms in computer science, let us just mention the work on the graph isomorphism problem that culminated with Babai's quasi-polynomial time algorithm [2]. For a comprehensive introduction into the area permutation group algorithms, see Serres' textbook [23]. In this paper we are concerned with algorithmic problems related to the diameter of finite permutation groups. We start with a few basic definitions.

Let G be a finite group. For a subset $A \subseteq G$ we denote with $\langle A \rangle$ the subgroup of G generated by the elements from A (i.e., the closure of A under the group multiplication). If $\langle A \rangle = G$ then A is called a generating set of G. For $k \geq 0$ we write $A^{\leq k}$ for the set of all products $a_1 a_2 \cdots a_l \in G$ with $l \leq k$ and $a_1, \ldots, a_l \in A$. For an element $g \in \langle A \rangle$ we denote with $|g|_A$ (the A-length of g) the smallest integer k such that $g \in A^{\leq k}$. The diameter $d(G, A)$ of G with respect to the generating A is the smallest number d such that $\langle A \rangle = A^{\leq d}$. Note that such a d exists since G is finite. There is a quite extensive literature on upper and lower bounds on the diameter in various finite groups; see e.g. [3, 4, 5, 6, 7, 8, 9, 13, 17, 19]. Let us mention in this context a famous (and still open) conjecture of Babai and Seress [8] stating that for every finite non-abelian simple group G and every generating set A, $d(G, A)$

50th International Colloquium on Automata, Languages, and Programming (ICALP 2023).
Editors: Kousha Etessami, Uriel Feige, and Gabriele Puppis; Article No. 134; pp. 134:1–134:18
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

is bounded by $\mathcal{O}((\log |G|)^c)$ for some universal constant c. This would imply in particular that the diameters of A_n and S_n (with respect to any generating sets) are bounded by a polynomial in n. The currently best known upper bound is $\exp(\mathcal{O}(\log^4 n \log\log n))$ [13].

Many problems about mechanical puzzles reduce to questions about the diameter of finite groups. As an example let us mention Rubik's cube. For a long time it was open how many moves in Rubik's cube are needed to transform an arbitrary initial configuration into the target configuration. This number is simply the diameter of the so-called Rubik's cube group. The precise value of this diameter was open for a long time. In 2013 Rokicki et al. proved that is 20 [22].

1.1 Computing diameter and length

In the first part of the paper (Sections 3 and 4) we investigate the complexity of certain decision variants of the following computational problems:

(i) computing the length of a given element from a permutation group and

(ii) computing the diameter of a permutation group.

Let us define the problems that we will investigate more precisely. With S_n we denote the group of all permutations on $[1, n] = \{1, \dots, n\}$. In the following problems, a permutation $\pi \in S_n$ is given by the list $\pi(1), \pi(2), \dots, \pi(n)$. The size n of the domain (also called the degree of the permutations) is part of the input. For $A \subseteq S_n$ we write $d(A)$ for $d(\langle A\rangle, A)$. We then define the following computational problems:

▶ **Problem** (UNARY LENGTH).
Input: a set of permutations $A \subseteq S_n$, an element $\pi \in \langle A\rangle$, and a unary encoded number k.[1]
Question: Is $|\pi|_A \leq k$?

▶ **Problem** (UNARY DIAMETER).
Input: a set of permutations $A \subseteq S_n$ and a unary encoded number k.
Question: Is $d(A) \leq k$?

The problems BINARY LENGTH and BINARY DIAMETER are defined in the same way, except that the input number k is given in binary encoding.

Goldreich and Even [11] were the first who obtained results on the complexity of these problems. They proved that UNARY LENGTH is NP-complete and UNARY DIAMETER is NP-hard[2] but the exact complexity of UNARY DIAMETER remained open. A parameterized variant of UNARY LENGTH (with k as the parameter) is studied under the name PERMUTATION GROUP FACTORIZATION in [10] and shown to be W[1]-hard and in W[P]. The binary setting was first studied by Jerrum [15]. He proved that BINARY LENGTH is PSPACE-complete.

We also study exact versions of the above problems:

▶ **Problem** (UNARY EXACT LENGTH).
Input: a set of permutations $A \subseteq S_n$, a permutation $\pi \in \langle A\rangle$, and a unary encoded number k.
Question: Is $|\pi|_A = k$?

[1] It is well-known that there is a polynomial time algorithm that checks whether $\pi \in \langle A\rangle$ holds [12].

[2] The problem UNARY LENGTH is called MGS for "minimum generator sequence" in [11], whereas UNARY DIAMETER is called MBGS for "minimum upper bound on generator sequences". We believe that UNARY LENGTH and UNARY DIAMETER are more suggestive. Another point is that Even and Goldreich do not specify the encoding of integers in their paper, but from the NP-completeness result for UNARY LENGTH, it can deduced that they have the unary encoding of integers in mind.

▶ **Problem** (UNARY EXACT DIAMETER).
Input: a set of permutations $A \subseteq S_n$ and a unary encoded number k.
Question: Is $d(A) = k$?

Again, there are corresponding problems BINARY EXACT LENGTH and BINARY EXACT DIAMETER, where the input number k is given in binary notation.

The first main result of this paper solves the open problem left in [11]: UNARY DIAMETER is Π_2^{P}-complete, where Π_2^{P} is the second universal level of the polynomial time hierarchy. This result also holds for the restriction, where all permutations in the set $A \subseteq S_n$ pairwise commute and have order two (and hence $\langle A \rangle$ is an abelian group of exponent two). Moreover, we also show that BINARY DIAMETER with a set A of pairwise commuting permutations is Π_2^{P}-complete.

The complexity of BINARY DIAMETER for general permutation groups remains open. The problem is easily seen to be in PSPACE. The above mentioned result of Jerrum (PSPACE-completeness of BINARY LENGTH for a binary encoded number k) seems to have no implications for the complexity of BINARY DIAMETER. Nevertheless, we conjecture that BINARY DIAMETER is PSPACE-complete.

We then proceed to show that UNARY EXACT DIAMETER is complete for the complexity class DP_2, which is the class of all intersections of a Π_2^{P}-language and a Σ_2^{P}-language. Hardness for DP_2 already holds for the restriction of UNARY EXACT DIAMETER to abelian permutation groups of exponent two. To get DP_2-hardness, we use the fact that our Π_2^{P}-hardness proof of UNARY DIAMETER already holds for inputs $A \subseteq S_n$ and $k \in \mathbb{N}$ with the promise that the diameter of $\langle A \rangle$ is either k or $k+1$. Using similar techniques we can also show that UNARY EXACT LENGTH is DP-complete, where DP is the class of all intersections of an NP-language and a coNP-language.

1.2 Equality and universality for finite automata over permutation groups

In the second part of the paper (Section 5), we consider problems related to finite automata over permutation groups. The setting is as follows: Consider a nondeterministic finite automaton (NFA) $\mathcal{A}$ over a finite alphabet Σ of input letters and a mapping $h : \Sigma \to S_n$ to a symmetric group. The mapping h extends to a morphism $h : \Sigma^* \to S_n$ from the free monoid Σ^* to the group S_n (we use the same letter h for this extension). We may then ask whether a given permutation π belongs to $h(L(\mathcal{A}))$. This is the *rational subset membership problem for permutation groups*, where the input consists of the NFA $\mathcal{A}$, the mapping $h : \Sigma \to S_n$ (n is also part of the input) and the permutation π. It is shown in [16, 18] that the rational subset membership problem for permutation groups is NP-complete.[3]

To simplify notation, we omit the mapping $h : \Sigma^* \to S_n$ in the following, and replace in the NFA $\mathcal{A}$ every transition label $a \in \Sigma$ by the corresponding permutation $h(a) \in S_n$. Thus, we consider NFAs, where the transitions are labelled with elements of a symmetric group S_n. The set $L(\mathcal{A})$ accepted by $\mathcal{A}$ is then directly interpreted as a subset of S_n. Clearly, every subset of S_n is of the form $L(\mathcal{A})$ for an NFA $\mathcal{A}$ over S_n, but in general the number of

[3] In [18] stronger results are shown: (i) NP-hardness already holds for membership in sets $\pi^*\sigma^*\tau^*$, where π, σ, τ are input permutations, and (ii) membership in NP holds for black-box groups and a restricted class of context-free languages (where terminal symbols are again replaced by permutations). The general membership problem for context-free sets of permutations is PSPACE-complete [18].

transitions of $\mathcal{A}$ must be exponential in n (this follows from a simple counting argument). Note that for a finite set $A \subseteq S_n$ it is straightforward to come up with an automaton $\mathcal{A}$ over S_n with a single state and $|A|$ many transitions such that $L(\mathcal{A}) = \langle A \rangle$.

In Section 5, we consider the *rational equality problem for permutation groups*, RATIONAL EQUALITY for short:

▶ **Problem** (RATIONAL EQUALITY).
Input: two NFAs $\mathcal{A}$ and $\mathcal{B}$ over S_n (n is as usual part of the input).
Question: Does $L(\mathcal{A}) = L(\mathcal{B})$ hold?

As before, we also consider the abelian variant of this problem, where all permutations labelling the transitions of $\mathcal{A}$ and $\mathcal{B}$ pairwise commute. Moreover, we consider the following restriction of RATIONAL EQUALITY.

▶ **Problem** (RATIONAL UNIVERSALITY).
Input: an NFA $\mathcal{A}$ over S_n.
Question: Does $L(\mathcal{A}) = S_n$ hold?

Note that for RATIONAL UNIVERSALITY, the restriction where the permutations appearing in $\mathcal{A}$ pairwise commute is not interesting, since S_n is not abelian for $n \geq 3$.

We show that RATIONAL EQUALITY and RATIONAL UNIVERSALITY are both Π_2^{P}-complete and that Π_2^{P}-hardness for RATIONAL EQUALITY already holds for the abelian case. For the lower bounds we use reductions from UNARY DIAMETER.

Let us finally remark that our upper bound proofs do not use any specific properties of permutation groups. In particular, all upper bounds shown in this paper also hold for the black-box-group setting, where elements of a black-box group G are encoded by bit strings and there are oracles for (i) multiplying two elements of G, (ii) inverting an element of G, and (iii) checking whether two bit strings represent the same element of G (see [23] for more details on black-box groups).

2 Preliminaries

2.1 Background from complexity theory

We assume that the reader has some basic background from complexity theory, see e.g. [1] for more information. The levels Σ_k^{P} and Π_k^{P} of the *polynomial time hierarchy* [24] are defined as follows:

- $\Sigma_0^{\mathsf{P}} = \Pi_0^{\mathsf{P}} = \mathsf{P}$
- $\Sigma_{k+1}^{\mathsf{P}}$ is the set of all languages L such that there exists a language $K \in \Pi_k^{\mathsf{P}}$ and a polynomial p with $L = \{x \mid \exists y \in \{0,1\}^{p(|x|)} : x\#y \in K\}$ (here $\#$ is a separator symbol).
- Π_{k+1}^{P} is the set of all languages L such that there exists a language $K \in \Sigma_k^{\mathsf{P}}$ and a polynomial p with $L = \{x \mid \forall y \in \{0,1\}^{p(|x|)} : x\#y \in K\}$.

In particular, we have $\Sigma_1^{\mathsf{P}} = \mathsf{NP}$ and $\Pi_1^{\mathsf{P}} = \mathsf{coNP}$. We will make use of the computational problem $\forall\exists$SAT, where the input is a $\forall\exists$-formula

$$\Psi = \forall x_1 \cdots \forall x_n \exists y_1 \cdots \exists y_m F(x_1, \ldots, x_n, y_1, \ldots, y_m), \tag{1}$$

where F is a boolean formula in conjunctive normal form built from the boolean variables $x_1, \ldots, x_n, y_1, \ldots, y_m$. The question is whether Ψ holds. This problem is Π_2^{P}-complete [24].

The complexity class DP_k is defined as

$$\mathsf{DP}_k = \{L_1 \cap L_2 \mid L_1 \in \Sigma_k^{\mathsf{P}} \text{ and } L_2 \in \Pi_k^{\mathsf{P}}\}.$$

The class $\mathsf{DP}_1 = \{L_1 \cap L_2 \mid L_1 \in \mathsf{NP} \text{ and } L_2 \in \mathsf{coNP}\}$ is usually denoted by DP. It was defined in [21]. The only mentioning of the classes DP_k for $k \geq 2$ we are aware of is the stack exchange post [20].

2.2 Some notations for permutation groups

Recall that a permutation group is a subgroup of the *symmetric group* S_n for some n, where S_n is the group of all permutations on $[1, n] = \{1, \ldots, n\}$. We will use standard notations for permutation groups; see e.g., [23]. Permutations will be often written by their decomposition into disjoint cycles, called the *disjoint cycle decomposition.* A cycle of length two is a *transposition.* A product of permutations will be evaluated from left to right. For $a \in [1, n]$ and $\pi \in S_n$ we will also write a^π for $\pi(a)$. This fits nicely to the left-to-right evaluation order: $a^{\pi\tau} = (a^\pi)^\tau$. For a permutation π we denote by $\mathrm{ord}(\pi)$ the order of π, i.e., the smallest $k \geq 1$ such that π^k is the identity permutation.

Most of the hardness results in Sections 3 and 4 will be shown for the group $\mathbb{Z}_2^n$ (the n-fold direct product of the group $\mathbb{Z}_2$) for an $n \geq 0$. Clearly, this is an abelian group of exponent two (i.e., every element has order two). The group $\mathbb{Z}_2^n$ is isomorphic to the subgroup of S_{2n} generated by all transpositions $(2i-1, 2i)$ for $i \in [1, n]$. For a finite set V of size n, we will identify $\mathbb{Z}_2^n$ with the group $\mathbb{Z}_2^V$ of all mappings $f : V \to \mathbb{Z}_2$ with the group operation to be pointwise addition modulo 2. We write this abelian group additively. For a function $f : V \to \mathbb{Z}_2 = \{0, 1\}$ we define its support as $\mathrm{supp}(f) = \{v \in V \mid f(v) = 1\}$. For a subset $U \subseteq V$ we denote by $[U] \in \mathbb{Z}_2^V$ the unique group element with $\mathrm{supp}([U]) = U$.

3 Complexity of diameter for permutation groups

We come to the first main result of the paper: UNARY DIAMETER is Π_2^{P}-hard. In the following theorem, the additional statement that the diameter $d(A)$ is either k or $k+1$ will be needed later when we consider UNARY EXACT DIAMETER.

▶ **Theorem 3.1.** *There is a logspace reduction ϕ from $\forall\exists$SAT to UNARY DIAMETER such that for every $\forall\exists$-formula Ψ with $\phi(\Psi) = (A, k)$ we have: $A \subseteq \mathbb{Z}_2^n$ for some n, $\langle A \rangle = \mathbb{Z}_2^n$ and $d(A) \in \{k, k+1\}$.*

Proof. Let us fix a $\forall\exists$-formula Ψ as in (1). We can write F as $F = \bigwedge_{c \in C} c$, where C is a set of clauses (disjunctions of variables and negated variables). We start with several transformations that ensure some additional properties for Ψ.

Step 1. In order to bound the diameter of the group from above by $k+1$ we replace Ψ by the formula

$$\forall x_1 \cdots \forall x_n \exists y_1 \cdots \exists y_m \exists y^* \, G(x_1, \ldots, x_n, y_1, \ldots, y_m, y^*),$$

where y^* is a new variable and

$$G = \neg y^* \wedge \bigwedge_{c \in C} (y^* \vee c).$$

By this it is ensured that for every truth assignment of the universally quantified variables $x_1, \ldots, x_n$, there is a truth assignment of the existentially quantified variables $y_1, \ldots, y_m, y^*$ such that exactly one clause in G is unsatisfied (simply set y^* to the true value 1).

Step 2. Next, it is necessary to ensure that every variable appears in at most d clauses for a fixed constant d. This can be ensured similarly to [25] for 3SAT: for every variable $z \in \{x_1, \dots, x_n, y_1, \dots, y_m, y^*\}$ that appears in $l \geq 4$ clauses in G we introduce new variables $z_1, \dots, z_l$ and replace the i-th occurrence of z by z_i. Then we add the clauses

$$(\neg z_1 \vee z_2), (\neg z_2 \vee z_3), \dots, (\neg z_{l-1} \vee z_l), (\neg z_l \vee z_1) \tag{2}$$

which enforce that $z_1, \dots, z_l$ must get the same truth value. If $z \in \{x_1, \dots, x_n\}$ then we universally quantify z_1 and existentially quantify $z_2, \dots, z_l$. If $z \in \{y_1, \dots, y_m, y^*\}$ then all new variables z_i get existentially quantified. In the resulting formula, every variable occurs in at most 3 clauses.

Step 3. Finally, for our later arguments, it is necessary to add for every universally quantified variable x the trivial clause

$$c_x = (x \vee \neg x).$$

Of course, this trivial clause does not change the truth value of the formula. Now every variable occurs in at most 4 clauses. We still have the property that every truth assignment for the universally quantified variables can be extended by a truth assignment for the existentially quantified variables such that exactly one clause becomes unsatisfied. To see this, consider an arbitrary truth assignment for the universally quantified variables. The clauses c_x that we added in Step 3 are always satisfied. We now assign the truth value 1 to all variables y_i^* that replaced in Step 2 the variable y^* from Step 1. This ensures that all clauses that were derived from clauses $y^* \vee c$ with $c \in C$ in Step 2 are satisfied. All remaining clauses of the form (2) (with $z \neq y^*$) can be easily satisfied. If z_1 is universally quantified (so its truth value is already fixed) then all $z_2, \dots, z_l$ are existentially quantified and we assign to these variables the truth value of z_1. Otherwise $z_1, \dots, z_l$ are all existentially quantified and we can assign the truth value 1 to all of them (the truth value 0 would also work). At this point, only the single clause derived from $\neg y^*$ in Step 2 is not satisfied. Finally, note that each of the three steps preserves the truth value of the $\forall\exists$-formula.

This concludes the preprocessing of the $\forall\exists$-formula Ψ. To simplify notation, we denote the resulting formula again with

$$\Psi = \forall x_1 \cdots \forall x_n \exists y_1 \cdots \exists y_m F(x_1, \dots, x_n, y_1, \dots, y_m). \tag{3}$$

Let $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$. For a variable $z \in X \cup Y$ we denote with $\tilde{z}$ one of the literals z or $\neg z$. Moreover we denote by C the set of clauses of F. A clause is viewed as a set of literals. For a literal $\tilde{z}$ let $C(\tilde{z})$ be the set of all clauses containing $\tilde{z}$. Note that every set $C(\tilde{z})$ has size at most 4. Let

$$V = X \cup Y \cup C$$

(we assume that X, Y, C are pairwise disjoint). We will work in the group $\mathbb{Z}_2^V$ introduced in Section 2.2 and use the notation introduced there.

The logspace reduction ϕ from $\forall\exists$SAT to UNARY DIAMETER is defined by $\phi(\Psi) = (A, k)$ with $k = n + m$ and

$$A = \bigcup_{x \in X} A_x \cup \bigcup_{y \in Y} A_y,$$

where for all universally quantified variables $x \in X$,

$$A_x = \{[\{x\} \cup U] \mid U \subseteq C(x)\} \cup \{[U] \mid U \subseteq C(\neg x), U \neq \emptyset\}$$

and for all existentially quantified variables $y \in Y$,

$$A_y = \{[U] \mid U \subseteq \{y\} \cup C(y), U \neq \emptyset\} \cup \{[U] \mid U \subseteq \{y\} \cup C(\neg y), U \neq \emptyset\}.$$

Since every set $C(\tilde{z})$ has size at most 4, we can construct the instance (A, k) in logspace.

▷ Claim 3.2. $\langle A \rangle = \mathbb{Z}_2^V$.

Proof of Claim 3.2. It suffices to show that every $[\{v\}]$ for $v \in V$ belongs to $\langle A \rangle$. From the definition of A we immediately get $[\{x\}] \in A_x$ for all $x \in X$ and $[\{y\}] \in A_y$ for all $y \in Y$. Consider now a clause $c \in C$ and fix a literal $\tilde{z} \in c$. We have $c \in C(\tilde{z})$. If $z \in Y$ then $[\{c\}] \in A_z$. If $z \in X$ and $\tilde{z} = \neg z$, then again $[\{c\}] \in A_z$. Finally, if $z \in X$ and $\tilde{z} = z$ then $[\{z\}], [\{z, c\}] \in A_z$. Hence, $[\{c\}] = [\{z\}] + [\{z, c\}] \in \langle A \rangle$. ◁

▷ Claim 3.3. Let $f_X : X \to \{0,1\}$ and $f : V \to \{0,1\}$ be elements of $\mathbb{Z}_2^V$ with

$$f(x) = f_X(x)$$

for all $x \in X$ and

$$f(v) = 1$$

for all $v \in Y \cup C$. If $f \in A^{\leq n+m}$, then there is a function $f_Y : Y \to \{0,1\}$ such that $f_X + f_Y$ is a satisfying truth assignment for F.

Proof of Claim 3.3. Suppose that $f \in A^{\leq n+m}$ and hence

$$f = f_1 + \cdots + f_s \tag{4}$$

for an $s \leq n+m$ with $f_i \in A$. Then the right-hand side of (4) must contain for every $x \in X$ a generator from A_x since $f(c_x) = 1$ (recall that we added the clause $c_x = \{x, \neg x\}$ since x is universally quantified) and only generators from A_x set the c_x-value to 1. Moreover, the right-hand side of (4) must contain for every $y \in Y$ a generator from A_y since $f(y) = 1$. Thus, the right-hand side of (4) must contain at least $|X| + |Y| = n + m$ generators. We get $s = n + m$ and obtain

$$f = \sum_{x \in X} g_x + \sum_{y \in Y} g_y \tag{5}$$

with $g_x \in A_x$ and $g_y \in A_y$. For $z \in X \cup Y$ let $C_z = \mathrm{supp}(g_z) \cap C$ be the set of clauses that appear in g_z in the sum (5). For all $y \in Y$ we must have $g_y = [\{y\} \cup C_y]$ since $f(y) = 1$.

We define the function $f_Y : Y \to \{0,1\}$ by

$$f_Y(y) = \begin{cases} 1 & \text{if } C_y \subseteq C(y), \\ 0 & \text{otherwise.} \end{cases}$$

Note that if $f_Y(y) = 0$ then we must have $C_y \subseteq C(\neg y)$.

We claim that $f_X + f_Y$ satisfies F. Consider a clause $c \in C$. Since $f(c) = 1$ there must exist $z \in X \cup Y$ such that $c \in C_z$. If $z = y \in Y$, then one of the following two cases holds:

- $f_Y(y) = 1$, $g_y = [\{y\} \cup C_y]$, and $c \in C_y \subseteq C(y)$, i.e., $y \in c$,
- $f_Y(y) = 0$, $g_y = [\{y\} \cup C_y]$, and $c \in C_y \subseteq C(\neg y)$, i.e., $\neg y \in c$.

In both cases f_Y set a literal from c (either y or $\neg y$) to 1. Now, assume that $z = x \in X$. Then one of the following two cases holds:

- $f(x) = f_X(x) = 1$, $g_x = [\{x\} \cup C_x]$ and $c \in C_x \subseteq C(x)$, i.e., $x \in c$,
- $f(x) = f_X(x) = 0$, $g_x = [C_x]$ and $c \in C_x \subseteq C(\neg x)$, i.e., $\neg x \in c$.

Again, in both cases f_X sets a literal from c (either x or $\neg x$) to 1. $\lhd$

Our proof of Claim 3.3 also shows that $d(A) \leq k$ implies $d(A) = k$: if $d(A) \leq k$ then for any of the functions f from Claim 3.3 we have $|f|_A = k$.

From Claims 3.2 and 3.3 it follows that if $\langle A \rangle = A^{\leq n+m}$ then for every $f_X : X \to \{0,1\}$ there must exist $f_Y : Y \to \{0,1\}$ such that $f_X + f_Y$ satisfies F. Hence, the formula Ψ from (3) holds.

Now suppose that Ψ holds. Let $f \in \mathbb{Z}_2^V$. We want to show $f \in A^{\leq n+m}$. First observe that there are unique functions $f_X : X \to \{0,1\}$ and $g : Y \cup C \to \{0,1\}$ such that $f = f_X + g$. Since Ψ holds, there is a partial truth assignment $f_Y : Y \to \{0,1\}$ such that $f_X + f_Y$ satisfies F. We define for every variable $z \in X \cup Y$ the set $U(z) \subseteq V$ as follows:

- if $z = x \in X$ and $f_X(x) = 1$ then $U(x) = \{x\} \cup C(x)$,
- if $z = x \in X$ and $f_X(x) = 0$ then $U(x) = C(\neg x)$,
- if $z = y \in Y$ and $f_Y(y) = 1$ then $U(y) = \{y\} \cup C(y)$,
- if $z = y \in Y$ and $f_Y(y) = 0$ then $U(y) = \{y\} \cup C(\neg y)$.

Note that $[U(z)] \in A_z$ for every variable $z \in X \cup Y$, except for the case that $z = x \in X$, $f_X(x) = 0$ and $C(\neg x) = \emptyset$ (then, $U(z) = \emptyset$). Define

$$U = \bigcup_{z \in X \cup Y} U(z).$$

Since all clauses evaluate to 1 under $f_X + f_Y$, we have $C \cup Y \subseteq U$. Moreover, $x \in U$ if and only if $x \in \operatorname{supp}(f)$ for all $x \in X$. We therefore have

$$\operatorname{supp}(f) \subseteq U.$$

We can choose pairwise disjoint (possibly empty) subsets $U'(z) \subseteq U(z)$ such that

$$U = \bigcup_{z \in X \cup Y} U'(z)$$

is a partition of U. It follows that

$$\operatorname{supp}(f) = \bigcup_{z \in X \cup Y} (U'(z) \cap \operatorname{supp}(f))$$

is a partition of $\operatorname{supp}(f)$. Let $Z \subseteq X \cup Y$ be the set of all $z \in X \cup Y$ such that $U'(z) \cap \operatorname{supp}(f) \neq \emptyset$. Then we have

$$f = \sum_{z \in Z} [U'(z) \cap \operatorname{supp}(f)]$$

and $|Z| \leq n + m$. It remains to show that $[U'(z) \cap \operatorname{supp}(f)]$ is a generator from A_z. This is clear if $z = y \in Y$ or ($z = x \in X$ and $U(x) = C(\neg x)$). In those case, for every non-empty subset $U' \subseteq U(z)$, $[U']$ belongs to A_z. Finally, if $z = x \in X$ and $U(x) = \{x\} \cup C(x)$ then also $x \in U'(x)$ must hold because $x \in U(x) \subseteq U = \bigcup_{z \in X \cup Y} U'(z)$ and $x \notin U'(z)$ for $z \neq x$. Moreover, $U(x) = \{x\} \cup C(x)$ implies $f(x) = f_X(x) = 1$, i.e., $x \in \operatorname{supp}(f)$. Therefore, $[U'(x) \cap \operatorname{supp}(f)]$ is of the form $[\{x\} \cup C']$ for some $C' \subseteq C(x)$, which belongs to A_x. We obtain $f \in A^{\leq n+m}$, which shows that ϕ is indeed a logspace reduction from $\forall\exists$SAT to UNARY DIAMETER.

▷ Claim 3.4. $d(A) \leq k+1$.

Proof of Claim 3.4. Our preprocessing ensured that every partial truth assignment of the universally quantified variables can be extended by a truth assignment for the existentially quantified variables such that exactly one clause $c \in C$ is unsatisfied. Let $f \in \mathbb{Z}_2^V$. Then there are functions $f_X : X \to \{0,1\}$ and $g : Y \cup C \to \{0,1\}$ such that $f = f_X + g$. Moreover, there is a partial truth assignment $f_Y : Y \to \{0,1\}$ such that $f_X + f_Y$ satisfies all clauses from $C \setminus \{c\}$. As above we define for every variable $z \in X \cup Y$ the set $U(z) \subseteq V$ as follows:

- if $z = x \in X$ and $f_X(x) = 1$ then $U(x) = \{x\} \cup C(x)$,
- if $z = x \in X$ and $f_X(x) = 0$ then $U(x) = C(\neg x)$,
- if $z = y \in Y$ and $f_Y(y) = 1$ then $U(y) = \{y\} \cup C(y)$,
- if $z = y \in Y$ and $f_Y(y) = 0$ then $U(y) = \{y\} \cup C(\neg y)$.

Note that $c = \{\neg y\}$ for an existentially quantified variable $y \in Y$ (see Step 1 in our preprocessing). Hence, we have $c \in C(\neg y)$ and $[\{c\}] \in A_y$ is a generator. We define $U(c) = \{c\}$ and

$$U = \bigcup_{z \in X \cup Y \cup \{c\}} U(z).$$

The rest of the argument is the same as above: Since all clauses except for c evaluate to 1 under $f = f_X + f_Y$, we have $C \cup Y \subseteq U$. Moreover, $x \in U$ if and only if $x \in \mathrm{supp}(f)$ for all $x \in X$. We therefore have

$$\mathrm{supp}(f) \subseteq U.$$

Then there are pairwise disjoint subsets $U'(z) \subseteq U(z)$ such that

$$U = \bigcup_{z \in X \cup Y \cup \{c\}} U'(z) \quad \text{and} \quad \mathrm{supp}(f) = \bigcup_{z \in X \cup Y \cup \{c\}} (U'(z) \cap \mathrm{supp}(f))$$

are partitions of U and $\mathrm{supp}(f)$, respectively. Let $Z \subseteq X \cup Y \cup \{c\}$ be the set of all $z \in X \cup Y \cup \{c\}$ such that $U'(z) \cap \mathrm{supp}(f) \neq \emptyset$. Then we have

$$f = \sum_{z \in Z} [U'(z) \cap \mathrm{supp}(f)]$$

and $|Z| \leq n + m + 1$. As above it can easily be shown that $[U'(z) \cap \mathrm{supp}(f)]$ is a generator. Hence $f \in A^{\leq n+m+1}$. ◁

It now follows that $d(A)$ is either k or $k+1$: if $d(A) \leq k$ then $d(A) = k$ (see the remark after the proof of Claim 3.3), and if $d(A) > k$ then $d(A) = k+1$ by Claim 3.4. ◀

▶ **Corollary 3.5.** *The following problems are all Π_2^{P}-complete:*

(i) *UNARY DIAMETER (without a restriction on the permutation group $\langle A \rangle$),*

(ii) *UNARY DIAMETER restricted to abelian permutation groups $\langle A \rangle$ of exponent two,*

(iii) *BINARY DIAMETER restricted to abelian permutation groups $\langle A \rangle$.*

Proof. In all cases the lower bound follows from Theorem 3.1. It remains to show the upper bound in cases (i) and (iii). For (i), this is straightforward: Let $G = \langle A \rangle$ where $A \subseteq S_n$ is a set of permutations and take a unary encoded $k > 0$. First of all we universally guess an element $\pi \in G$. More precisely, we guess an arbitrary permutation $\pi \in S_n$ and then check in polynomial time (using [12]) whether $\pi \in \langle A \rangle$. If this does not hold, we immediately accept, otherwise we proceed with existentially guessing a sequence $a_1 a_2 \cdots a_l$ with $a_i \in A$ and $l \leq k$. We accept if and only if $a_1 a_2 \cdots a_l = \pi$.

The upper bound in case (iii) can be shown in a similar way. We follow the procedure for UNARY DIAMETER up to the point where we guess a sequence $a_1 a_2 \cdots a_l$ with $a_i \in A$ and $l \leq k$. Since k is given in binary encoding this is not feasible. Instead, we guess for each $a \in A$ a binary encoded number $k_a \geq 0$ whose bit length is bounded by the bit length of k. We accept if and only if the following two conditions hold:

- $\sum_{a \in A} k_a \leq k$,
- $\prod_{a \in A} a^{k_a} = \pi$.

Both conditions can be checked in polynomial time. For the second point note that a^{k_a} can be computed in time $\mathcal{O}(n \log k_a)$ by iterated squaring. ◀

▶ **Theorem 3.6.** *UNARY EXACT DIAMETER is DP_2-complete for general permutation groups as well as abelian permutation groups of exponent two.*

Proof. Let $A \subseteq S_n$ be a set of permutations and let k be a unary encoded number. Then we have $d(A) = k$ if and only if $d(A) \leq k$ and $d(A) > k - 1$. This is the intersection of a Π_2^{P}-property and a Σ_2^{P}-property. Hence, UNARY EXACT DIAMETER belongs to DP_2.

Now let $L = L_1 \cap L_2$ be a language from DP_2 with $L_1 \in \Sigma_2^{\mathsf{P}}$ and $L_2 \in \Pi_2^{\mathsf{P}}$. By Theorem 3.1 we can compute from x two pairs (A_1, k_1) and (A_2, k_2) (with $A_i \subseteq S_{n_i}$ and k_i a unary encoded natural number) such that

- $x \in L_1$ if and only if $d(A_1) = k_1 + 1$ if and only if $d(A_1) \neq k_1$, and
- $x \in L_2$ if and only if $d(A_2) = k_2$ if and only if $d(A_2) \neq k_2 + 1$.

Hence, $x \in L$ if and only if $d(A_1) = k_1 + 1$ and $d(A_2) = k_2$.

Consider the subgroup of $S_{n_2} \times S_{n_2} \leq S_{2n_2}$ generated by

$$B := (A_2 \times \{1\}) \cup (\{1\} \times A_2).$$

Here, 1 denotes the identity permutation. Since we have either $d(A_2) = k_2$ or $d(A_2) = k_2 + 1$ we obtain either $d(B) = 2k_2$ or $d(B) = 2k_2 + 2$.

Finally, consider the subgroup of $S_{n_1} \times S_{n_2} \times S_{n_2} \leq S_{n_1 + 2n_2}$ generated by

$$A := (A_1 \times \{(1,1)\}) \cup (\{1\} \times B).$$

There are four cases for the diameter of the group generated by A:

$$d(A) = \begin{cases} k_1 + 2k_2 & \text{if } d(A_1) = k_1 \quad\;\;\;\; \text{and } d(A_2) = k_2 \\ k_1 + 2k_2 + 1 & \text{if } d(A_1) = k_1 + 1 \text{ and } d(A_2) = k_2 \\ k_1 + 2k_2 + 2 & \text{if } d(A_1) = k_1 \quad\;\;\;\; \text{and } d(A_2) = k_2 + 1 \\ k_1 + 2k_2 + 3 & \text{if } d(A_1) = k_1 + 1 \text{ and } d(A_2) = k_2 + 1. \end{cases}$$

Thus, we have $x \in L$ if and only if $d(A) = k_1 + 2k_2 + 1$, which shows the DP_2-hardness of UNARY EXACT DIAMETER. ◀

4 Complexity of computing the length in permutation groups

Recall that Even and Goldreich [11] proved that UNARY LENGTH is NP-complete. We present below an alternative proof for the NP-hardness, where the reduction has additional properties (similar to Theorem 3.1) that will be needed in order to settle the complexity of UNARY EXACT LENGTH. Our techniques are similar to those from Section 3.

▶ **Theorem 4.1.** *There is a logspace reduction ϕ from SAT to UNARY LENGTH such that for every CNF formula F with $\phi(\Psi) = (A, \pi, k)$ we have: $A \subseteq \mathbb{Z}_2^n$ for some n, $\pi \in \langle A \rangle = \mathbb{Z}_2^n$ and $|\pi|_A \in \{k, k+1\}$.*

Proof. Let $F = \bigwedge_{c \in C} c$ be a conjunction of clauses $c \in C$ with boolean variables from the set X. We preprocess F as in the proof of Theorem 3.1. First, we replace F by $F' = \neg y^* \wedge \bigwedge_{c \in C}(y^* \vee c)$, where $y^* \notin X$ is a new variable ensuring that there is a truth assignment such that exactly one clause in F' is unsatisfied. Moreover, F is satisfiable if and only if F' is satisfiable.

Then we apply the construction of [25] that we also used in the proof of Theorem 3.1 in order to ensure that every variable occurs in at most three clauses. We replace the occurrences of every variable $z \in X \cup \{y^*\}$ that occurs in $l \geq 4$ clauses (negated or unnegated) by new variables $z_1, \dots, z_l$ and add the clauses $z_l \vee \neg z_1$ and $z_i \vee \neg z_{i+1}$ for $i \in [1, l-1]$. Let F'' be the resulting CNF formula. It still has the property that there is a truth assignment such that exactly one clause in F'' is unsatisfied. One can take the truth assignment that sets all variables of F'' to 1. Moreover, F is satisfiable if and only if F'' is satisfiable.

From this consideration, it follows that we can assume that our input CNF formula F has the following two properties:

- Every variable occurs in at most three clauses.
- There is a truth assignment for F such that exactly one clause of F is not satisfied.

Let X be the variables that occur in F and let C be the set of clauses in F. Moreover, let $V = X \cup C$. With $L = X \cup \{\neg x \mid x \in X\}$ we denote the set of all literals.

We reuse several notations that we have introduced in the proof of Theorem 3.1. For a literal $\tilde{x} \in L$ we denote with $C(\tilde{x}) \subseteq C$ the set of all clauses containing $\tilde{x}$. Note that we have $|C(x)| + |C(\neg x)| \leq 3$. For the reduction we work with the group $\mathbb{Z}_2^V$ and use the notations from Section 2.2. Now we define the set A of generators by

$$A = \bigcup_{c \in C} A_c \cup \bigcup_{\tilde{x} \in L} A_{\tilde{x}},$$

where for $x \in X$ and $c \in C$ we take

$$\begin{aligned} A_x &= \{[\{x\} \cup U] \mid U \subseteq C(x)\}, \\ A_{\neg x} &= \{[\{x\} \cup U] \mid U \subseteq C(\neg x)\}, \\ A_c &= \{[\{c\}]\}. \end{aligned}$$

Note that $\langle A \rangle = \mathbb{Z}_2^V$.

We define $\pi = [V]$ and $k = |X|$. This defines our logspace reduction $\phi : F \mapsto (A, \pi, k)$. To compute ϕ in logspace, it is important that all sets $C(\tilde{x})$ have constant size.

Now we show that $|\pi|_A \leq k$ if and only if F is satisfiable. Suppose $|\pi|_A \leq k$. Since $X \subseteq \mathrm{supp}(\pi)$, we need a generator from every $A_x \cup A_{\neg x}$ $(x \in X)$ to produce π. This implies $|\pi|_A = k$ and we can write

$$\pi = \sum_{x \in X} \pi_x$$

with $\pi_x \in A_x \cup A_{\neg x}$. Let $\pi_x = [\{x\} \cup U_x]$ with $U_x \subseteq C$. We define a truth assignment by

$$\sigma(x) = \begin{cases} 1 & \text{if } \pi_x \in A_x, \\ 0 & \text{if } \pi_x \in A_{\neg x} \setminus A_x. \end{cases}$$

for all $x \in X$. From $C \subseteq \mathrm{supp}(\pi)$ it follows that for every clause $c \in C$ there must exist a variable $x \in X$ such that $c \in U_x$. If $\pi_x \in A_x$ (i.e., $\sigma(x) = 1$) then $c \in U_x \subseteq C(x)$, i.e., x appears in the clause c. Hence, π satisfies c. Similarly, if $\pi_x \in A_{\neg x}$ (i.e., $\sigma(x) = 0$) then $\neg x$ appears in the clause c. Therefore, σ satisfies F.

Now suppose that F is satisfiable and let σ be a satisfying truth assignment. Hence, every clause is satisfied. Let $X_0 = \{x \in X \mid \sigma(x) = 0\}$ and $X_1 = \{x \in X \mid \sigma(x) = 1\}$. Then we have

$$\operatorname{supp}(\pi) = V = X \cup C = \bigcup_{x \in X_0} \{x\} \cup C(\neg x) \cup \bigcup_{x \in X_1} \{x\} \cup C(x).$$

Then we can choose for every $x \in X_0$ a subset $U_x \subseteq C(\neg x)$ and for every $x \in X_1$ a subset $U_x \subseteq C(x)$ such that

$$\operatorname{supp}(\pi) = \bigcup_{x \in X} \{x\} \cup U_x$$

is a partition of $\operatorname{supp}(\pi)$. Hence, we have

$$\pi = \sum_{x \in X} [\{x\} \cup U_x].$$

Since $[\{x\} \cup U_x] \in A_x \cup A_{\neg x}$, we finally obtain $|\pi|_A = k$. This show that ϕ is indeed a logspace reduction from SAT to UNARY LENGTH.

We have already noted that $|\pi|_A \leq k$ implies $|\pi|_A = k$. The converse implication is trivially true. Therefore, we have $|\pi|_A \leq k$ if and only if $|\pi|_A = k$. It remains to show that $|\pi|_A \in \{k, k+1\}$. For this it suffices to show $|\pi|_A \leq k+1$.

We know that there is a truth assignment σ such that exactly one clause $c \in C$ is unsatisfied. As above we can choose generators $\pi_x \in A_x \cup A_{\neg x}$ for all $x \in X$ such that

$$\operatorname{supp}(\pi) \setminus \{c\} = \bigcup_{x \in X} \operatorname{supp}(\pi_x)$$

is a partition of $\operatorname{supp}(\pi) \setminus \{x\}$. From this we obtain

$$\pi = [\{c\}] + \sum_{x \in X} \pi_x$$

and hence $|\pi|_A \leq k+1$, which concludes the proof. ◀

▶ **Theorem 4.2.** *UNARY EXACT LENGTH is* DP*-complete for general permutation groups as well as abelian permutation groups of exponent two.*

Proof. Let A be a set of generators of a permutation group, k a unary encoded integer, and $\pi \in \langle A \rangle$ a permutation. Then we have $|\pi|_A = k$ if and only if $|\pi|_A \leq k$ and $|\pi|_A > k-1$. This is the conjunction of an NP-property and a coNP-property. Thus UNARY EXACT LENGTH is the intersection of a language in NP with a language in coNP and therefore belongs to DP.

We show DP-hardness of UNARY EXACT LENGTH by a reduction from SAT-UNSAT. The input for the latter problem is a pair (F, G) of two CNF formulas and the question is whether F is satisfiable and G is unsatisfiable. This problem is known to DP-complete, see [21].

Let (F, G) be an input for SAT-UNSAT. By Theorem 4.1 we can compute from (F, G) in logspace two triples (A_1, π_1, k_1) and (A_2, π_2, k_2) (with $A_i \subseteq S_{n_i}$, $\pi_i \in \langle A_i \rangle$ and k_i a unary encoded natural number) such that

- F is satisfiable if and only if $|\pi_1|_{A_1} = k_1$ if and only if $|\pi_1|_{A_1} \neq k_1 + 1$ and
- G is unsatisfiable if and only if $|\pi_2|_{A_2} \neq k_2$ if and only if $|\pi_2|_{A_2} = k_2 + 1$.

Hence, (F, G) is a positive instance of SAT-UNSAT if and only if $|\pi_1|_{A_1} = k_1$ and $|\pi_2|_{A_2} = k_2+1$.

Consider the subgroup of $S_{n_2} \times S_{n_2} \leq S_{2n_2}$ with the generating set

$$B := (A_2 \times \{1\}) \cup (\{1\} \times A_2).$$

Since we have $|\pi_2|_{A_2} \in \{k_2, k_2+1\}$ we obtain $|(\pi_2,\pi_2)|_B \in \{2k_2, 2k_2+2\}$.

Finally, consider the group $\langle A_1\rangle \times \langle A_2\rangle \times \langle A_2\rangle \leq S_{n_1+2n_2}$ with the generating set

$$A := (A_1 \times \{(1,1)\}) \cup (\{1\} \times B).$$

For the length $|(\pi_1,\pi_2,\pi_2)|_A$ we obtain

$$|(\pi_1,\pi_2,\pi_2)|_A = \begin{cases} k_1 + 2k_2 & \text{if } |\pi_1|_{A_1} = k_1 \quad\;\;\; \text{and } |\pi_2|_{A_2} = k_2 \\ k_1 + 2k_2 + 1 & \text{if } |\pi_1|_{A_1} = k_1 + 1 \text{ and } |\pi_2|_{A_2} = k_2 \\ k_1 + 2k_2 + 2 & \text{if } |\pi_1|_{A_1} = k_1 \quad\;\;\; \text{and } |\pi_2|_{A_2} = k_2 + 1 \\ k_1 + 2k_2 + 3 & \text{if } |\pi_1|_{A_1} = k_1 + 1 \text{ and } |\pi_2|_{A_2} = k_2 + 1. \end{cases}$$

Hence, (F,G) is a positive instance of SAT-UNSAT if and only if $|(\pi_1,\pi_2,\pi_2)|_A = k_1 + 2k_2 + 2$, which concludes the proof. ◀

Since BINARY LENGTH is PSPACE-complete [15], one might expect that also BINARY EXACT LENGTH is PSPACE-complete. The following result confirms this.

▶ **Theorem 4.3.** *BINARY EXACT LENGTH is* PSPACE*-complete.*

Proof of Theorem 4.3. Since PSPACE is closed under complement, and $|\pi|_A = k$ if and only if $\pi \in A^{\leq k}$ and $\pi \notin A^{\leq k-1}$, it follows that also BINARY EXACT LENGTH belongs to PSPACE.

For the lower bound let $A \subseteq S_n$ be a set of permutations on $[1,n]$, $\pi \in \langle A\rangle$ and k be a binary encoded number. We construct from A, π, k in logspace a new instance B, τ, k such that $\pi \in A^{\leq k}$ if and only if $|\tau|_B = k$ holds. This proves that BINARY EXACT LENGTH is PSPACE-complete.

Clearly, $S_n \leq S_m$ for $n \leq m$. In the following, we will identify a permutation $\pi \in S_n$ with a permutation from S_m by defining $a^\pi = a$ for $a \in [n+1, m]$.

Let d be the number of bits of k. Then $\log_2(k) < d \leq \log_2(k)+1$. Let $p_1 = 2, p_2 = 3, \dots, p_d$ be the first d primes. Note that since d is polynomially bounded in the input length, the primes p_i are so too and therefore can be stored in logarithmic space. Let $m = \sum_{i=1}^d p_i$ and let $\alpha_1, \dots, \alpha_d$ be permutations with pairwise disjoint support on $[n+1, n+m]$ such that α_i is a cycle of length p_i. Moreover let $r_1, \dots, r_d \in [0, p_i - 1]$ such that

$$k \equiv r_i \bmod p_i.$$

These numbers can be computed in logspace; see e.g. [14]. Moreover let $\alpha = \alpha_1 \cdots \alpha_d$, $\beta = \alpha_1^{r_1} \cdots \alpha_d^{r_d}$ and $\tau = \pi\beta$. We have

$$\mathrm{ord}(\alpha) = \prod_{i=1}^{d} p_i \geq 2^d > 2^{\log_2(k)} = k.$$

Finally, we define the set of permutations

$$B = \{\gamma\alpha \mid \gamma \in A\} \cup \{\alpha\} \subseteq S_{n+m}.$$

Since $\beta = \alpha^k$, i.e., $\tau = \pi\alpha^k$ and $\mathrm{ord}(\alpha) > k$, we obtain $\pi \in A^{\leq k}$ if and only if $|\tau|_B = k$, which concludes the reduction. ◀

5 Complexity of equality and universality for NFAs over permutation groups

In this section we determine the complexity of RATIONAL EQUALITY and RATIONAL UNIVERSALITY (defined in Section 1.2).

▶ **Theorem 5.1.** *The following problems are Π_2^{P}-complete for permutation groups:*

(i) *RATIONAL EQUALITY*

(ii) *RATIONAL EQUALITY restricted to the case where all permutations in the two input NFAs $\mathcal{A}$ and $\mathcal{B}$ pairwise commute and have order two.*

(iii) *RATIONAL UNIVERSALITY*

Proof. For the upper bounds, we only have to consider RATIONAL EQUALITY. Membership of RATIONAL EQUALITY in Π_2^{P} follows from the fact that the rational subset membership problem for permutation groups (see Section 1.2) is in NP. More precisely, the following formula expresses the equality $L(\mathcal{A}_0) = L(\mathcal{A}_1)$ for two NFAs $\mathcal{A}_0$ and $\mathcal{A}_1$ over S_n:

$$\forall i \in \{0,1\} \forall \pi \in S_n : \pi \notin L(\mathcal{A}_i) \vee \pi \in L(\mathcal{A}_{1-i}).$$

Since the rational subset membership problem for permutation groups is in NP, the above formula is equivalent to a statement of the form

$$\forall i \in \{0,1\} \forall \pi \in S_n \forall u \exists v : u \text{ is not a witness for } \pi \in L(\mathcal{A}_i) \vee v \text{ is a witness for } \pi \in L(\mathcal{A}_{1-i}).$$

Here u and v are bit strings of size polynomial in the input length.

The lower bound in (ii) is a direct consequence of Corollary 3.5, since for a finite set $A \subseteq S_n$ and a unary encoded number k both $\langle A \rangle$ and $A^{\leq k}$ can be defined by logspace computable NFAs.

It remains to show Π_2^{P}-hardness of RATIONAL UNIVERSALITY. For this we give a reduction from UNARY DIAMETER to RATIONAL UNIVERSALITY. Before we come to the actual reduction, let us explain an auxiliary construction. Fix an $n \geq 1$ and consider the symmetric group S_{2n} on the domain $\Omega = [1, 2n]$. We define the following sets of transpositions:

$$T_i = \{(a,b) \mid a,b \in \Omega \setminus \{i\}, a \neq b\} \subseteq S_{2n} \text{ for all } i \in \Omega, \tag{6}$$

$$Z = \{(2i-1, 2i) \mid 1 \leq i \leq n\} \subseteq S_{2n}. \tag{7}$$

Note that $\langle Z \rangle \cong \mathbb{Z}_2^n$ and $\langle T_i \rangle$ is the set of permutations that fix i.

For every $1 \leq i < j \leq 2n$ with $(i,j) \notin Z$ we can construct in space $\mathcal{O}(\log n)$ three automata $\mathcal{A}_{i,j}, \mathcal{B}_{i,j}, \mathcal{C}_{i,j}$ over S_{2n} such that the following hold:

$$L(\mathcal{A}_{i,j}) = \bigcup_{\ell \in \Omega \setminus \{i,j\}} (i,j)(j,\ell)\langle T_i \cap T_j \rangle \tag{8}$$

$$L(\mathcal{B}_{i,j}) = \bigcup_{\ell \in \Omega \setminus \{i,j\}} (j,i)(i,\ell)\langle T_i \cap T_j \rangle \tag{9}$$

$$L(\mathcal{C}_{i,j}) = (i,j)\langle T_i \cap T_j \rangle \tag{10}$$

▷ Claim 5.2. We have

$$\langle Z \rangle \cap \bigcup_{\substack{1 \leq i < j \leq 2n \\ (i,j) \notin Z}} (L(\mathcal{A}_{i,j}) \cup L(\mathcal{B}_{i,j}) \cup L(\mathcal{C}_{i,j})) = \emptyset.$$

Proof of Claim 5.2. Suppose there is a $\tau \in \langle Z \rangle$ such that $\tau \in L(\mathcal{A}_{i,j}) \cup L(\mathcal{B}_{i,j}) \cup L(\mathcal{C}_{i,j})$ for some $1 \leq i < j \leq 2n$ with $(i,j) \notin Z$. For every $a \in [1,n]$ we have either $(2a-1)^\tau = 2a-1$ and $(2a)^\tau = 2a$ or $(2a-1)^\tau = 2a$ and $(2a)^\tau = 2a-1$.

Case 1. $\tau \in L(\mathcal{A}_{i,j})$. Then we can write $\tau = (i,j)(j,\ell)\pi$ with $\ell \in \Omega \setminus \{i,j\}$ and $\pi \in \langle T_i \cap T_j \rangle$. Then we obtain

$$j^\tau = j^{(i,j)(j,\ell)\pi} = i^{(j,\ell)\pi} = i^\pi = i.$$

We can exclude the case $j = j^\tau = i$, since $i < j$. Hence, we have $j^\tau \in \{j+1, j-1\}$. If j is odd we obtain $j+1 = j^\tau = i$, which is a contradiction since $i < j$. If j is even we obtain $j-1 = j^\tau = i$, and hence $(i,j) \in Z$, which is also a contradiction.

Case 2. $\tau \in L(\mathcal{B}_{i,j})$. Then we can write $\tau = (j,i)(i,\ell)\pi$ with $\ell \in \Omega \setminus \{i,j\}$ and $\pi \in \langle T_i \cap T_j \rangle$. In this case we obtain

$$i^\tau = i^{(j,i)(i,\ell)\pi} = j^{(i,\ell)\pi} = j^\pi = j.$$

We can exclude the case $i = i^\tau = j$, since $i < j$. Hence, we have $i^\tau \in \{i+1, i-1\}$. If i is odd we obtain $i+1 = i^\tau = j$ and hence $(i,j) \in Z$, which is a contradiction. If i is even we obtain $i-1 = i^\tau = j$, which contradicts $i < j$.

Case 3. $\tau \in L(\mathcal{C}_{i,j})$. Then we can write $\tau = (i,j)\pi$ with $\pi \in \langle T_i \cap T_j \rangle$ and get

$$i^\tau = i^{(i,j)\pi} = j^\pi = j.$$

We obtain a contradiction in the same way as in Case 2. ◁

▷ Claim 5.3. We have

$$\bigcup_{\substack{1 \le i < j \le 2n \\ (i,j) \notin Z}} (L(\mathcal{A}_{i,j}) \cup L(\mathcal{B}_{i,j}) \cup L(\mathcal{C}_{i,j})) = S_{2n} \setminus \langle Z \rangle. \tag{11}$$

Proof of Claim 5.3. By Claim 5.2 it suffices to show

$$S_{2n} \setminus \langle Z \rangle \subseteq \bigcup_{\substack{1 \le i < j \le 2n \\ (i,j) \notin Z}} (L(\mathcal{A}_{i,j}) \cup L(\mathcal{B}_{i,j}) \cup L(\mathcal{C}_{i,j})). \tag{12}$$

Let $\tau \in S_{2n} \setminus \langle Z \rangle$. We have to show that τ belongs to the union on the right-hand side of (12). Let $\tau = \gamma_1 \cdots \gamma_m$ be the disjoint cycle decomposition of τ. Since $\tau \notin \langle Z \rangle$ we can assume that w.l.o.g. $\gamma_1 \notin \langle Z \rangle$. Let $\alpha = \gamma_1$ and let $\beta = \gamma_2 \cdots \gamma_m$. Then we can write $\tau = \alpha\beta = \beta\alpha$ in which $\alpha = (i_d, \ldots, i_2, i_1)$ is a cycle of length $d \ge 2$. Note that $i_q \ne i_p$ for all $q \ne p$.

Case 1. $d = 2$. W.l.o.g. we can assume $i_1 < i_2$. Then $(i_1, i_2) \notin Z$ and by this the NFA $\mathcal{C}_{i_1,i_2}$ is defined. We have $(i_1,i_2)\pi \in L(\mathcal{C}_{i_1,i_2})$ for all $\pi \in \langle T_{i_1} \cap T_{i_2} \rangle$. Since β fixes i_1 and i_2, we have $\beta \in \langle T_{i_1} \cap T_{i_2} \rangle$ and hence $\tau = (i_1,i_2)\beta \in L(\mathcal{C}_{i_1,i_2})$.

Case 2. $d \ge 3$ and $(i_1, i_2) \notin Z$. Then, $\mathcal{A}_{i_1,i_2}$ is defined if $i_1 < i_2$ and $\mathcal{B}_{i_2,i_1}$ is defined if $i_2 < i_1$. We have

$$\alpha = (i_d, \ldots, i_1) = (i_1, i_2)(i_2, i_3)(i_3, i_4) \cdots (i_{d-1}, i_d).$$

Let $\gamma = (i_3, i_4) \cdots (i_{d-1}, i_d)\beta$. We get $\tau = \alpha\beta = (i_1, i_2)(i_2, i_3)\gamma$. Moreover, $\gamma \in \langle T_{i_1} \cap T_{i_2} \rangle$, since β fixes i_1 and i_2 and $i_q \ne i_p$ for $q \ne p$. If $i_1 < i_2$ we have $(i_1,i_2)(i_2,\ell)\pi \in L(\mathcal{A}_{i_1,i_2})$ for all $\ell \in \Omega \setminus \{i_1, i_2\}$ and $\pi \in \langle T_{i_1} \cap T_{i_2} \rangle$. Hence we obtain $\tau = (i_1,i_2)(i_2,i_3)\gamma \in L(\mathcal{A}_{i_1,i_2})$. If $i_2 < i_1$ we have $(i_1,i_2)(i_2,\ell)\pi \in L(\mathcal{B}_{i_2,i_1})$ for all $\ell \in \Omega \setminus \{i_1, i_2\}$ and $\pi \in \langle T_{i_1} \cap T_{i_2} \rangle$. Thus we analogously obtain $\tau = (i_1,i_2)(i_2,i_3)\gamma \in L(\mathcal{B}_{i_2,i_1})$.

Case 3. $d \geq 3$ and $(i_1, i_2) \in Z$. We then have $(i_2, i_3) \notin Z$ (otherwise, we would get $i_3 = i_1$) and $\mathcal{A}_{i_2,i_3}$ is defined if $i_2 < i_3$ and $\mathcal{B}_{i_3,i_2}$ is defined if $i_3 < i_2$. We have

$$\alpha = (i_d, \dots, i_1) = (i_1, i_d, i_{d-1} \dots, i_2) = (i_2, i_3)(i_3, i_4)(i_4, i_5) \cdots (i_{d-1}, i_d)(i_d, i_1).$$

Let $\gamma = (i_4, i_5) \cdots (i_{d-1}, i_d)(i_d, i_1)\beta$. Then we obtain $\tau = \alpha\beta = (i_2, i_3)(i_3, i_4)\gamma$ (if $d = 3$ we have $\gamma = \beta$ and $i_4 = i_1$). Analogously to Case 2, we obtain $(i_2, i_3)(i_3, i_4)\gamma \in L(\mathcal{A}_{i_2,i_3})$ if $i_2 < i_3$ and $(i_2, i_3)(i_3, i_4)\gamma \in L(\mathcal{B}_{i_3,i_2})$ if $i_3 < i_2$. ◁

Now we come to the reduction from UNARY DIAMETER to RATIONAL UNIVERSALITY. The proof of Theorem 3.1 shows that we can start with an input instance (A, k) of UNARY DIAMETER, where $A \subseteq \mathbb{Z}_2^n$ for some $n \in \mathbb{N}$ and $k \in \mathbb{N}$ is given in unary encoding. We can therefore assume that $\langle A \rangle = \langle Z \rangle$ for the above Z from (7). From A and k we can easily construct in logspace an NFA $\mathcal{A}$ such that

$$L(\mathcal{A}) = A^{\leq k} \cup \bigcup_{\substack{1 \leq i < j \leq 2n \\ (i,j) \notin Z}} (L(\mathcal{A}_{i,j}) \cup L(\mathcal{B}_{i,j}) \cup L(\mathcal{C}_{i,j})) = A^{\leq k} \cup (S_{2n} \setminus \langle A \rangle),$$

where the second equality follows from Claim 5.3. It is also important that k is given in unary encoding, which allows to construct in logspace an NFA for $A^{\leq k}$. We have $L(\mathcal{A}) = S_{2n}$ if and only if $d(A) \leq k$ which concludes the reduction. ◀

Note that in the above proof we write the complement $S_{2n} \setminus \langle A \rangle = S_{2n} \setminus \langle Z \rangle$ as a union of a polynomial number of cosets (see (8)–(10) and (11)). One might ask why we do not write $S_{2n} \setminus \langle A \rangle$ simply as union of cosets of $\langle A \rangle$. The problem is that the latter would require $|S_{2n}|/|\langle A \rangle| = (2n!)/2^n - 1$ cosets, which is not polynomial in n.

6 Open problems

The main open problem that remains is the complexity of BINARY DIAMETER. We conjecture that this problem is PSPACE-complete. Recall that we proved BINARY DIAMETER to be Π_2^{P}-complete for abelian permutation groups. We conjecture that this result can be extended to nilpotent permutation groups (and maybe even solvable permutation groups).

We conjecture that UNARY DIAMETER is Π_2^{P}-complete for input instances (A, k), where A generates the full symmetric group S_n. The Π_2^{P}-completeness of RATIONAL UNIVERSALITY would directly follow from this. Moreover, we mentioned in the introduction the conjecture according to which the diameter of S_n (with respect to any generating set) is bounded by a polynomial in n. This conjecture would imply that BINARY DIAMETER belongs to Π_2^{P} for input instances (A, k), where A generates the full symmetric group S_n.

References

1 Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009. doi:10.1017/CBO9780511804090.

2 László Babai. Graph isomorphism in quasipolynomial time [extended abstract]. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016*, pages 684–697. ACM, 2016. doi:10.1145/2897518.2897542.

3 László Babai, Robert Beals, and Ákos Seress. On the diameter of the symmetric group: polynomial bounds. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004*, pages 1108–1112. SIAM, 2004. URL: https://dl.acm.org/doi/10.5555/982792.982956.

4 László Babai and Thomas P. Hayes. Near-independence of permutations and an almost sure polynomial bound on the diameter of the symmetric group. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005*, pages 1057–1066. SIAM, 2005. URL: http://dl.acm.org/citation.cfm?id=1070432.1070584.

5 László Babai and Gábor Hetyei. On the diameter of random Cayley graphs of the symmetric group. *Combinatorics, Probability & Computing*, 1:201–208, 1992. doi:10.1017/S0963548300000237.

6 László Babai, Gábor Hetyei, William M. Kantor, Alexander Lubotzky, and Ákos Seress. On the diameter of finite groups. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science, FOCS 1990, Volume II*, pages 857–865. IEEE Computer Society, 1990. doi:10.1109/FSCS.1990.89608.

7 László Babai, William M. Kantor, and A. Lubotsky. Small-diameter Cayley graphs for finite simple groups. *European Journal of Combinatorics*, 10(6):507–522, 1989. doi:10.1016/S0195-6698(89)80067-8.

8 László Babai and Ákos Seress. On the diameter of Cayley graphs of the symmetric group. *Journal of Combinatorial Theory, Series A*, 49(1):175–179, 1988. doi:10.1016/0097-3165(88)90033-7.

9 László Babai and Ákos Seress. On the diameter of permutation groups. *European Journal of Combinatorics*, 13(4):231–243, 1992. doi:10.1016/S0195-6698(05)80029-0.

10 Liming Cai, Jianer Chen, Rodney G. Downey, and Michael R. Fellows. On the parameterized complexity of short computation and factorization. *Archive for Mathematical Logic*, 36(4-5):321–337, 1997. doi:10.1007/s001530050069.

11 Shimon Even and Oded Goldreich. The minimum-length generator sequence problem is NP-hard. *Journal of Algorithms*, 2(3):311–313, 1981. doi:10.1016/0196-6774(81)90029-8.

12 Merrick L. Furst, John E. Hopcroft, and Eugene M. Luks. Polynomial-time algorithms for permutation groups. In *Proceedings of the 21st Annual Symposium on Foundations of Computer Science, FOCS 1980*, pages 36–41. IEEE Computer Society, 1980. doi:10.1109/SFCS.1980.34.

13 Harald A. Helfgott and Ákos Seress. On the diameter of permutation groups. *Annals of Mathematics*, 179(2):611–658, 2014. doi:10.4007/annals.2014.179.2.4.

14 William Hesse, Eric Allender, and David A. Mix Barrington. Uniform constant-depth threshold circuits for division and iterated multiplication. *Journal of Computer and System Sciences*, 65(4):695–716, 2002. doi:10.1016/S0022-0000(02)00025-9.

15 Mark Jerrum. The complexity of finding minimum-length generator sequences. *Theoretical Computer Science*, 36:265–289, 1985. doi:10.1016/0304-3975(85)90047-7.

16 Arthur A. Khashaev. On the membership problem for finite automata over symmetric groups. *Discrete Mathematics and Applications*, 32(6):389–395, 2022. doi:10.1515/dma-2022-0033.

17 D. Kornhauser, G. Miller, and P. Spirakis. Coordinating pebble motion on graphs, the diameter of permutation groups, and applications. In *Proceedings of the 25th IEEE Symposium on Foundations of Computer Science, FOCS 1984*, pages 241–250. IEEE Computer Society Press, 1984. doi:10.1109/SFCS.1984.715921.

18 Markus Lohrey, Andreas Rosowski, and Georg Zetzsche. Membership problems in finite groups. In *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science, MFCS 2022*, volume 241 of *LIPIcs*, pages 71:1–71:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.MFCS.2022.71.

19 Pierre McKenzie. Permutations of bounded degree generate groups of polynomial diameter. *Information Processing Letters*, 19(5):253–254, 1984. doi:10.1016/0020-0190(84)90062-0.

20 not A or B (https://cstheory.stackexchange.com/users/38066/not-a-or b). An analog of DP for the second level of the polynomial hierarchy. Theoretical Computer Science Stack Exchange. URL: https://cstheory.stackexchange.com/q/38776.

21 Christos H. Papadimitriou and Mihalis Yannakakis. The complexity of facets (and some facets of complexity). *Journal of Computer and System Sciences*, 28(2):244–259, 1984. doi:10.1016/0022-0000(84)90068-0.

22 Tomas Rokicki, Herbert Kociemba, Morley Davidson, and John Dethridge. The diameter of the Rubik's cube group is twenty. *SIAM Journal of Discrete Mathematics*, 27(2):1082–1105, 2013. doi:10.1137/120867366.

23 Ákos Seress. *Permutation Group Algorithms*. Cambridge Tracts in Mathematics. Cambridge University Press, 2003. doi:10.1017/CBO9780511546549.

24 Larry J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976. doi:10.1016/0304-3975(76)90061-X.

25 Craig A. Tovey. A simplified NP-complete satisfiability problem. *Discrete Applied Mathematics*, 8(1):85–89, 1984. doi:10.1016/0166-218X(84)90081-7.