Decidability of the Membership Problem for 2×2 integer matrices*

Igor Potapov† Pavel Semukhin‡

Abstract

The main result of this paper is the decidability of the membership problem for 2×2 nonsingular integer matrices. Namely, we will construct the first algorithm that for any nonsingular 2×2 integer matrices $M_1, \ldots, M_n$ and M decides whether M belongs to the semigroup generated by $\{M_1, \ldots, M_n\}$. Our algorithm relies on a translation of numerical problems on matrices into combinatorial problems on words. It also makes use of some algebraic properties of well-known subgroups of $\mathrm{GL}(2, \mathbb{Z})$ and various new techniques and constructions that help to convert matrix equations into the emptiness problem for intersection of regular languages.

1 Introduction

Matrices and matrix products play a crucial role in the representation and analysis of various computational processes, i.e., linear recurrent sequences [19, 28, 29], arithmetic circuits [16], hybrid and dynamical systems [27, 3], probabilistic and quantum automata [8], stochastic games, broadcast protocols [15], optical systems [17], etc. Unfortunately, many simply formulated and elementary problems for matrices are inherently difficult to solve even in dimension two, and most of these problems become undecidable in general starting from dimension three or four. One of such hard questions is the *Membership problem* in matrix semigroups:

Membership problem: Given a finite set of $m \times m$ matrices $F = \{M_1, M_2, \ldots, M_n\}$ and a matrix M. Determine whether there exist an integer $k \geq 1$ and $i_1, i_2, \ldots, i_k \in \{1, \ldots, n\}$ such that

$$M_{i_1} \cdot M_{i_2} \cdots M_{i_k} = M.$$

In other words, determine whether a matrix M belongs to the semigroup generated by F.

In this paper we solve an open problem by showing that the membership is decidable for the semigroups of 2×2 nonsingular matrices over integers. The membership problem was intensively studied since 1947 when A.Markov showed that this problem is undecidable for matrices in $\mathbb{Z}^{6\times 6}$ even for a specific fixed set F [25]. Later, M. Paterson [30] in 1970 showed that a special case of the membership problem when M is equal to a zero matrix (known as *Mortality problem*) is undecidable for matrices in $\mathbb{Z}^{3\times 3}$. The decidability status of another special case of the membership problem — the *Identity problem* (i.e., when $M = I$, the identity matrix) — was unknown for a long time and was only recently shown to be undecidable for integer matrices starting from dimension four [6], see also the solution to Problem 10.3 in [9]. The undecidability of the identity problem means that the *Group problem* (of whether a matrix semigroup over integers forms a group) is undecidable starting from dimension four. A more recent survey of undecidable problems can be found in [10].

The undecidability proofs in matrix semigroups are mainly based on various techniques and methods for embedding universal computations into matrix products. The case of dimension two is the most intriguing one since there is some evidence that if these problems are undecidable, then this cannot be proved using any previously known constructions. In particular, there is no injective semigroup morphism from pairs of words over any finite alphabet (with at least two elements) into complex 2×2 matrices [11], which means that the coding of independent pairs of words in 2×2 complex matrices is impossible and the exact encoding of the Post Correspondence Problem or a computation of the Turing Machine cannot be used directly for proving undecidability in 2×2 matrix semigroups over $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{C}$. The only undecidability in the case of 2×2 matrices has been shown so far is the membership, freeness and vector reachability problems over quaternions [4] or more precisely in the case of diagonal matrices over quaternions, which are simply double quaternions.

The problems for semigroups are rather hard, but there was steady progress on decidable fragments over the last few decades. First, both membership and vector reachability problems were shown to be decidable in polynomial time for a semigroup generated by a single $m \times m$ matrix (known as the *Orbit problem*) by Kannan and Lipton [21] in 1986. Later, in 1996 this decidability result was extended to a more general case of commutative matrices [1]. The generalization of this

*This research was supported by EPSRC grant EP/M00077X/1.

†Department of Computer Science, University of Liverpool. Email: potapov@liverpool.ac.uk

‡Department of Computer Science, University of Liverpool. Email: semukhin@liverpool.ac.uk

result for a special class of non-commutative matrices (a class of row-monomial matrices over a commutative semigroup satisfying some natural effectiveness conditions) was shown in 2004 in [22]. Even now we still have long standing open problems for matrix semigroups generated by a single matrix, see, for example, the *Skolem Problem* about reaching zero in a linear recurrence sequence (LRS), which in matrix form is a question of whether any power of a given integer matrix A has zero in the right upper corner [13, 14]. It was recently shown that the decidability of either Positivity or Ultimate Positivity for integer LRS of order 6 would entail some major breakthroughs in analytic number theory. The decidability of each of these problems, whether for integer, rational, or algebraic linear recurrence sequences, is open, although partial results are known [16, 27, 28, 29].

Due to a severe lack of methods and techniques the status of decision problems for 2×2 matrices (like membership, vector reachability, freeness) remains a long standing open problem. Recently, a new approach of translating numerical problems on 2×2 integer matrices into a variety of combinatorial and computational problems on words over group alphabet and studying their transformations as specific rewriting systems have led to new results on decidability and complexity. In particular, this approach was successfully used to show the decidability of the membership problem for the semigroups of $\mathrm{GL}(2,\mathbb{Z})$ [12] in 2005 and of the mortality problem for 2×2 integer matrices with determinants $0,\pm 1$ [26] in 2008. It also found applications in the design of the polynomial time algorithm for the membership problem for the modular group [18] in 2007. Furthermore, it was used to show NP-hardness for most of the reachability problems in dimension two [5, 7] in 2012 and to prove decidability of the vector/scalar reachability problems in $\mathrm{SL}(2,\mathbb{Z})$ [31] in 2015.

The main ingredient of the translation into combinatorial problems on words is the well-known result that the groups $\mathrm{SL}(2,\mathbb{Z})$ and $\mathrm{GL}(2,\mathbb{Z})$ are finitely generated. For example, $\mathrm{SL}(2,\mathbb{Z})$ can be generated by a pair of matrices

$$S=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix} \quad \text{and} \quad R=\begin{bmatrix}0 & -1\\ 1 & 1\end{bmatrix}$$

with the following relations: $S^4=I$, $R^6=I$ and $S^2=R^3$. So, we can represent a matrix $M\in\mathrm{SL}(2,\mathbb{Z})$ as a word in the alphabet $\{S,R\}$.

In [12] both the *Identity* and the *Group* problems are shown to be decidable in $\mathbb{Z}^{2\times 2}$. Moreover, it was also claimed more generally that it is decidable whether or not a given nonsingular matrix belongs to a given finitely generated semigroup over integers. Unfortunately, it appears that the proof of this more general claim (i.e., when we consider matrices with determinants different from ± 1) has a significant gap, and it only works for a small number of special cases. Namely, in the very end of the proof of Theorem 1 after translating the membership from $\mathrm{GL}(2,\mathbb{Z})$ to $\mathrm{SL}(2,\mathbb{Z})$, the authors describe a very short reduction from the membership problem in $\mathbb{Z}^{2\times 2}$ to the one in $\mathrm{SL}(2,\mathbb{Z})$ using some incorrect assumptions. For instance, it was assumed that if X is an integer matrix with determinant one and Z is a nonsingular integer matrix, then there exists an integer matrix Y satisfying the equation $ZX=YZ$. However, this is not true and here is a simple counterexample. Let $Z=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$ and $X=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$, then from $ZX=YZ$ it follows that

$$Y=ZXZ^{-1}=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\cdot\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\cdot\begin{bmatrix}1 & 0\\ 0 & \frac{1}{2}\end{bmatrix}=\begin{bmatrix}0 & -\frac{1}{2}\\ 2 & 0\end{bmatrix}.$$

So Y has fractional coefficients, and if the matrices X and Z were in the generating set, then the argument from [12] would not work.

The main result of this paper is that the membership problem is decidable for the semigroups of 2×2 nonsingular integer matrices. Our proof provides an algorithm for solving this problem, which is based on the translation of the numerical problem on matrices into combinatorial problems on words and regular languages.

The novelty of our approach in comparison to the decidability of the membership for $\mathrm{GL}(2,\mathbb{Z})$ from [12] is in the use of the Smith normal form of a matrix (Theorem 2.2) and in proving the existence of the automaton $\mathcal{F}_D(\mathcal{A})$ which recognizes conjugates of the matrices accepted by a given automaton $\mathcal{A}$ with a diagonal matrix D (see Section 2.2). In fact, the construction of $\mathcal{F}_D(\mathcal{A})$ is the most crucial ingredient of our algorithm. It is based on a nontrivial combination of algebraic properties of $\mathrm{GL}(2,\mathbb{Z})$ and automata theory. The reader is referred to the proof of Theorem 2.1 in which we give a high-level description of our algorithm and highlight the main ideas behind it.

2 Main result

The semigroup of 2×2 integer matrices is denoted by $\mathbb{Z}^{2\times 2}$. We use $\mathrm{SL}(2,\mathbb{Z})$ to denote the special linear group of 2×2 matrices with integer coefficients, i.e.,

$$\mathrm{SL}(2,\mathbb{Z})=\{M\in\mathbb{Z}^{2\times 2}:\det(M)=1\}$$

and $\mathrm{GL}(2,\mathbb{Z})$ to denote the general linear group, i.e.,

$$\mathrm{GL}(2,\mathbb{Z})=\{M\in\mathbb{Z}^{2\times 2}:\det(M)=\pm 1\}.$$

A matrix is called *nonsingular* if its determinant is not equal to zero.

If F is a finite collection of matrices from $\mathbb{Z}^{2\times 2}$, then $\langle F\rangle$ denotes the semigroup generated by F (including the identity matrix), that is, $M \in \langle F\rangle$ if and only if $M = I$ or there are matrices $M_1, \ldots, M_n \in F$ such that $M = M_1 \cdots M_n$.

The main result of our paper is presented in Theorem 2.1 which states that the membership problem for nonsingular integer matrices in dimension two is decidable.

THEOREM 2.1. *There is an algorithm that decides for a given finite collection F of nonsingular matrices from $\mathbb{Z}^{2\times 2}$ and a nonsingular matrix $M \in \mathbb{Z}^{2\times 2}$ whether $M \in \langle F\rangle$.*

Proof sketch. Let $\{M_1, \ldots, M_n\}$ be all matrices from F whose determinant is different from ± 1, and let $\mathcal{S}^{\pm 1}$ be the semigroup which is generated by all matrices from F with determinant ± 1, that is,

$$\mathcal{S}^{\pm 1} = \langle\, F \cap \mathrm{GL}(2,\mathbb{Z})\,\rangle.$$

Then it is not hard to see that $M \in \langle F\rangle$ if and only if $M \in \mathcal{S}^{\pm 1}$ or there is a sequence of indices $i_1, \ldots, i_t \in \{1, \ldots, n\}$ and matrices $A_1, \ldots, A_{t+1}$ from $\mathcal{S}^{\pm 1}$ such that

$$M = A_1 M_{i_1} A_2 M_{i_2} \cdots A_t M_{i_t} A_{t+1}.$$

The first basic observation is that the value of t is bounded. Indeed, since $|\det(M_{i_s})| \geq 2$, for every $s = 1, \ldots, t$, we have that $t \leq \log_2 |\det(M)|$. So to decide whether or not $M \in \langle F\rangle$ we first need to check whether $M \in \mathcal{S}^{\pm 1}$. If $M \notin \mathcal{S}^{\pm 1}$, then we need to go through all sequences $i_1, \ldots, i_t \in \{1, \ldots, n\}$ of length up to $\log_2 |\det(M)|$ and for every such sequence check whether there are matrices $A_1, \ldots, A_{t+1}$ from $\mathcal{S}^{\pm 1}$ such that $M = A_1 M_{i_1} A_2 M_{i_2} \cdots A_t M_{i_t} A_{t+1}$. The rest of the paper is devoted to the proof that these problems are algorithmically decidable.

In Section 2.1 we describe an algorithm that decides whether $M \in \mathcal{S}^{\pm 1}$. In fact, in Proposition 2.2 we prove a stronger statement that it is decidable whether $M \in \mathcal{S}$, where $\mathcal{S}$ is an arbitrary regular subset of $\mathrm{GL}(2,\mathbb{Z})$, that is, a subset which is defined by a finite automaton. It will be easy to see that any semigroup in $\mathrm{GL}(2,\mathbb{Z})$, and in particular $\mathcal{S}^{\pm 1}$, is a regular subset. To prove this result we will introduce a notion of a canonical word and show that every matrix from $\mathrm{GL}(2,\mathbb{Z})$ can be represented by a unique canonical word (Corollary 2.1). Next, for any automaton $\mathcal{A}$ that defines a regular subset $\mathcal{S}$ of $\mathrm{GL}(2,\mathbb{Z})$, we construct another automaton $\mathrm{Can}(\mathcal{A})$ that accepts only canonical words and which defines the same regular subset $\mathcal{S}$ as the automaton $\mathcal{A}$ (see Proposition 2.2 and the Appendix). Finally, if w is the canonical word that represents a given matrix M, then $M \in \mathcal{S}$ if and only if $\mathrm{Can}(\mathcal{A})$ accepts w.

Proposition 2.2 provides an alternative proof for the decidability of the membership in $\mathrm{GL}(2,\mathbb{Z})$ presented in [12]. In fact, the construction of $\mathrm{Can}(\mathcal{A})$ was in part inspired by a similar construction from [12]. The difference of our approach is that we do not add new symbols to the alphabet during the construction of $\mathrm{Can}(\mathcal{A})$, which makes it easier to use in the next steps of the algorithm. The other key ideas of our proof, which are explained below, are new and original, and they did not appear in [12].

In Section 2.2 we show the decidability of the second problem in the special case when $t = 1$. Again, in Corollary 2.3 we prove a more general statement that for any two nonsingular matrices M_1 and M_2 from $\mathbb{Z}^{2\times 2}$ and regular subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ of $\mathrm{GL}(2,\mathbb{Z})$, it is decidable whether there are matrices $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ such that $A_1 M_1 A_2 = M_2$. One of the ingredients of our solution is the uniqueness of the Smith normal form of a matrix. A Smith normal form of a matrix $M \in \mathbb{Z}^{2\times 2}$ is a diagonal matrix D of special type such that $M = EDF$, where $E, F \in \mathrm{GL}(2,\mathbb{Z})$ (see Theorem 2.2 for the formal definition). It immediately follows from Theorem 2.2 that if M_1 and M_2 have different Smith normal forms, then the equation $A_1 M_1 A_2 = M_2$ cannot hold for any $A_1, A_2 \in \mathrm{GL}(2,\mathbb{Z})$ because the Smith normal form of M_1 is equal to that of $A_1 M_1 A_2$. So, we only need to consider the case when M_1 and M_2 have the same Smith normal form.

To explain in more details the key ideas behind our solution, let us assume that $M_1 = M_2 = D$, where D is a diagonal matrix of the form $\begin{bmatrix} 1 & 0 \\ 0 & n \end{bmatrix}$. As we show in the proof of Corollary 2.3, the general case can be easily reduced to this special case. So, we want to know if there are matrices $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ such that $A_1 D A_2 = D$. Our goal is to express this matrix equation as a relation between two regular languages and ultimately reduce this problem to the emptiness problem for intersection of regular languages. In order to do so, we first rewrite the equation $A_1 D A_2 = D$ as

$$D^{-1} A_1 D = A_2^{-1} \quad \text{or} \quad A_1^D = A_2^{-1},$$

where A_1^D is the conjugate of A_1 with D. Now, let $\mathcal{A}_1$ and $\mathcal{A}_2$ be finite automata that define the regular subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ of $\mathrm{GL}(2,\mathbb{Z})$. We will construct two automata $\mathcal{F}_D(\mathcal{A}_1)$ and $\mathrm{Inv}(\mathcal{A}_2)$ such that $\mathcal{F}_D(\mathcal{A}_1)$ recognizes the conjugates of matrices from $\mathcal{S}_1$ with D and $\mathrm{Inv}(\mathcal{A}_2)$ recognizes the inverses of matrices from $\mathcal{S}_2$. Finally, we will show that the equation $A_1^D = A_2^{-1}$ has a solution $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ if and only if the languages of the automata $\mathrm{Can}(\mathcal{F}_D(\mathcal{A}_1))$ and

$\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_2))$ have empty intersection (see Proposition 2.4).

The construction of $\mathrm{Inv}(\mathcal{A})$ is rather straightforward. On the other hand, the existence of the automaton $\mathcal{F}_D(\mathcal{A})$ is quite nontrivial, and its construction lies at the core of our algorithm. Informally speaking, $\mathcal{F}_D(\mathcal{A})$ computes the conjugates of the words $w \in L(\mathcal{A})$ symbol by symbol. For example, if $w = SRSR$, then $w^D = S^D R^D S^D R^D$. The main obstacle here is that the matrices S^D and R^D have fractional coefficients and hence do not belong to $\mathrm{GL}(2,\mathbb{Z})$, while w^D may still be an integer matrix. For example, if $D = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, then $S^D = \begin{bmatrix} 0 & -2 \\ \frac{1}{2} & 0 \end{bmatrix}$ and $R^D = \begin{bmatrix} 0 & -2 \\ \frac{1}{2} & 1 \end{bmatrix}$ but

$$(RSRS)^D = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}^D = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

To find a way around this obstacle, we will use new ideas that are based on an interplay between automata theory and algebra. Namely, we will show that the set of matrices $A \in \mathrm{GL}(2,\mathbb{Z})$ for which A^D also belongs to $\mathrm{GL}(2,\mathbb{Z})$ forms a subgroup, denoted $H(n)$, that has a finite index in $\mathrm{GL}(2,\mathbb{Z})$ (Theorem 2.3). Using this result we will show that the subgroup $H(n)$ is a regular subset of $\mathrm{GL}(2,\mathbb{Z})$. In fact, the states of the automaton $\mathcal{A}_{H(n)}$ that recognizes $H(n)$ are encoded by the right cosets of $H(n)$ (see Lemma 2.1). The automaton $\mathcal{A}_{H(n)}$ will help us to rewrite any word w that represents a matrix from $H(n)$ as a product of matrices whose conjugates with D have only integer coefficients. This will allow us to compute w^D symbol by symbol without using fractional matrices, that is, without going outside of $\mathrm{GL}(2,\mathbb{Z})$. An informal description of how this can be done is given at the beginning of the proof of Theorem 2.4.

Finally, in Section 2.3 we describe an algorithm for the general case. Namely, in Theorem 2.5 we will prove that for any nonsingular matrices $M_1, \ldots, M_t$ from $\mathbb{Z}^{2\times 2}$ and for any regular subsets $\mathcal{S}_1, \ldots, \mathcal{S}_t$ of $\mathrm{GL}(2,\mathbb{Z})$ that are defined by finite automata $\mathcal{A}_1, \ldots, \mathcal{A}_t$, respectively, it is decidable whether there are matrices $A_1 \in \mathcal{S}_1, \ldots, A_t \in \mathcal{S}_t$ such that

$$A_1 M_1 \cdots A_{t-1} M_{t-1} A_t = M_t. \tag{2.1}$$

In order to do this, we will construct an automaton $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$ such that the equation (2.1) has a solution $A_1 \in \mathcal{S}_1, \ldots, A_t \in \mathcal{S}_t$ if and only if the languages of the automata $\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_t))$ and $\mathrm{Can}(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))$ have empty intersection.

The proof is done by induction on t. For $t = 2$, the construction of $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$ is based on the construction of $\mathcal{F}_D(\mathcal{A}_1)$. For the inductive step, we will rewrite (2.1) as a system of two equations

$$A_1 M_1 \ldots A_{t-2} M_{t-2} U D_{t-1} V = M_t$$
$$A_{t-1} M_{t-1} A_t V^{-1} = U D_{t-1}$$

where D_{t-1} is the Smith normal form of M_{t-1} and U, V are unknown matrices from $\mathrm{GL}(2,\mathbb{Z})$. From these formulas one can conclude that it is quite natural to define $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$ as an automaton that recognizes the following union of regular languages[1]

$$\bigcup_{U \in \mathrm{GL}(2,\mathbb{Z})} L\Big(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-3}, \mathcal{A}_{t-2}, M_1, \ldots, M_{t-3}, M_{t-2} U D_{t-1}; M_t) \cdot \mathcal{F}(\mathcal{A}_{t-1}, M_{t-1}; U D_{t-1})\Big).$$

The obvious problem with this approach is that we need to take an infinite union because there are infinitely many matrices $U \in \mathrm{GL}(2,\mathbb{Z})$. To solve this problem, we have discovered a simple but very useful result that roughly speaking states that when we consider all possible Smith normal forms UDV for a fixed diagonal matrix D, we can assume that matrix U comes from a finite set of matrices (see Lemma 2.2). Using this fact, we can replace an infinite union in the above formula with a finite union, which gives us the desired regular language. □

The complexity of our algorithm is in EXPTIME. This is because a canonical word w that represents a matrix M has length exponential in the binary presentation of M, and so the construction of the automaton for the semigroup $\mathcal{S}^{\pm 1}$ takes exponential time. The next steps of the algorithm can be done in polynomial time.

Moreover, our algorithm can be extended to check the membership not only for semigroups in $\mathbb{Z}^{2\times 2}$ but for arbitrary regular subsets of nonsingular integer matrices.

Recently a simpler case of the membership of the identity matrix in $\mathrm{SL}(2,\mathbb{Z})$ was shown to be NP-complete [2]. However we do not know what is the exact complexity of the membership problem for nonsingular matrices from $\mathbb{Z}^{2\times 2}$.

2.1 Decidability of the membership problem in $\mathrm{GL}(2,\mathbb{Z})$. We will use an encoding of matrices from $\mathrm{GL}(2,\mathbb{Z})$ by words in alphabet $\Sigma = \{X, N, S, R\}$. For

[1] In this formula the central dot denotes concatenation of two finite automata.

this we define a mapping $\varphi : \Sigma \to \mathrm{GL}(2,\mathbb{Z})$ as follows:

$$\varphi(X) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \varphi(N) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$
$$\varphi(S) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \varphi(R) = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$$

We can extend φ to the morphism $\varphi : \Sigma^* \to \mathrm{GL}(2,\mathbb{Z})$ in a natural way. It is a well-known fact that morphism φ is surjective, that is, for every $M \in \mathrm{GL}(2,\mathbb{Z})$ there is a word $w \in \Sigma^*$ such that $\varphi(w) = M$.

DEFINITION 2.1. *We call two words w_1 and w_2 from Σ^** equivalent*, denoted $w_1 \sim w_2$, if $\varphi(w_1) = \varphi(w_2)$. Two languages L_1 and L_2 in the alphabet Σ are* equivalent*, denoted $L_1 \sim L_2$, if*
(i) for each $w_1 \in L_1$, there exists $w_2 \in L_2$ such that $w_1 \sim w_2$, and
(ii) for each $w_2 \in L_2$, there exists $w_1 \in L_1$ such that $w_2 \sim w_1$.
In other words, $L_1 \sim L_2$ if and only if $\varphi(L_1) = \varphi(L_2)$. Two finite automata $\mathcal{A}_1$ and $\mathcal{A}_2$ with alphabet Σ are equivalent*, denoted $\mathcal{A}_1 \sim \mathcal{A}_2$, if $L(\mathcal{A}_1) \sim L(\mathcal{A}_2)$.*

DEFINITION 2.2. *A subset $\mathcal{S} \subseteq \mathrm{GL}(2,\mathbb{Z})$ is called* regular *or* automatic *if there is a regular language L in alphabet Σ such that $\mathcal{S} = \varphi(L)$.*

Throughout the paper we will use the following abbreviation: if n is a positive integer and $V \in \Sigma$, then V^n denotes the word of length n which contains only letter V, and V^0 denotes the empty word.

DEFINITION 2.3. *A word $w \in \Sigma^*$ is called a* canonical word *if it has the form*

$$w = N^{\delta} X^{\gamma} S^{\beta} R^{\alpha_0} S R^{\alpha_1} S R^{\alpha_2} \dots S R^{\alpha_{n-1}} S R^{\alpha_n},$$

where $\beta, \delta, \gamma \in \{0,1\}$, $\alpha_0, \dots, \alpha_{n-1} \in \{1,2\}$, and $\alpha_n \in \{0,1,2\}$. In other words, w is canonical *if it does not contain subwords SS or RRR. Moreover, letter N may appear only once in the first position, and letter X may appear only once either in the first position or after N.*

We will make use of Corollary 2.1 below which states that every matrix from $\mathrm{GL}(2,\mathbb{Z})$ can be represented by a unique canonical word.

PROPOSITION 2.1. ([23, 24, 32]) *For every matrix $M \in \mathrm{SL}(2,\mathbb{Z})$, there is a unique canonical word w such that $M = \varphi(w)$. Note that w does not contain letter N because $\varphi(N) \notin \mathrm{SL}(2,\mathbb{Z})$.*

COROLLARY 2.1. *For every $M \in \mathrm{GL}(2,\mathbb{Z})$, there is a unique canonical word w such that $M = \varphi(w)$.*

Proof. If $\det(A) = 1$, that is, $M \in \mathrm{SL}(2,\mathbb{Z})$, then by Proposition 2.1 there is a unique canonical word w such that $M = \varphi(w)$. If $\det(A) = -1$, then $\varphi(N)^{-1}M \in \mathrm{SL}(2,\mathbb{Z})$ and again by Proposition 2.1 there is a unique canonical word w such that $\varphi(N)^{-1}M = \varphi(w)$ or $M = \varphi(Nw)$. Note that Nw is also a canonical word since w does not contain letter N. □

PROPOSITION 2.2. *There is an algorithm that for any regular subset $\mathcal{S} \subseteq \mathrm{GL}(2,\mathbb{Z})$ and a matrix $M \in \mathrm{GL}(2,\mathbb{Z})$ decides whether $M \in \mathcal{S}$.*

Proof. Let L be a regular language such that $\mathcal{S} = \varphi(L)$, and let $\mathcal{A}$ be a finite automaton that recognizes L, that is, $L = L(\mathcal{A})$. The words in L do not have to be in canonical form. So, we will construct a new automaton $\mathrm{Can}(\mathcal{A})$ whose language contains only canonical words and such that $\mathrm{Can}(\mathcal{A})$ is equivalent to $\mathcal{A}$, that is, $\varphi(L(\mathrm{Can}(\mathcal{A}))) = \varphi(L(\mathcal{A})) = \mathcal{S}$. The construction of $\mathrm{Can}(\mathcal{A})$ consists of a sequence of transformations that insert new paths and ε-transitions into $\mathcal{A}$. The detailed description of this construction is given in the Appendix.

Using the automaton $\mathrm{Can}(\mathcal{A})$ we can decide whether $M \in \mathcal{S}$. Indeed, by Corollary 2.1, there is a unique canonical word w that represents the matrix M, i.e., $M = \varphi(w)$. Now, we have the following equivalence: $M \in \mathcal{S}$ if and only if $w \in L(\mathrm{Can}(\mathcal{A}))$. Therefore, to decide whether $M \in \mathcal{S}$, we need to check whether w is accepted by $\mathrm{Can}(\mathcal{A})$. □

Note that any finitely generated semigroup $\langle M_1, \dots, M_n \rangle$ in $\mathrm{GL}(2,\mathbb{Z})$ is a regular subset. Indeed, let $w_1, \dots, w_n$ be canonical words that represent the matrices $M_1, \dots, M_n$, respectively, and consider a regular language $L = (w_1 + \dots + w_n)^*$. Clearly $\varphi(L) = \langle M_1, \dots, M_n \rangle$, and hence the semigroup $\langle M_1, \dots, M_n \rangle$ is regular. So as a corollary from Proposition 2.2 we obtain the decidability of the membership problem for the semigroups in $\mathrm{GL}(2,\mathbb{Z})$.

COROLLARY 2.2. *The membership problem for $\mathrm{GL}(2,\mathbb{Z})$ is decidable. That is, there is an algorithm that for a given finite collection of matrices $M_1, \dots, M_n$ and M from $\mathrm{GL}(2,\mathbb{Z})$, decides whether $M \in \langle M_1, \dots, M_n \rangle$.*

2.2 Special case: $A_1 M_1 A_2 = M_2$. In this section we show that for any two nonsingular matrices M_1 and M_2 from $\mathbb{Z}^{2\times 2}$ and regular subsets $\mathcal{S}_1$ and $\mathcal{S}_2$, it is decidable whether there exist matrices $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ such that $A_1 M_1 A_2 = M_2$ (Corollary 2.3). First, we prove this statement in the case when $M_1 = M_2 = D$, where D is a diagonal matrix in the Smith normal form (Proposition 2.4).

For the proof of this result we will use a few algebraical facts and results that are explained below. The most important of them is the following theorem about the Smith normal form of a matrix.

THEOREM 2.2. (SMITH NORMAL FORM [20]) *For any matrix $A \in \mathbb{Z}^{2\times 2}$, there are matrices E, F from $\mathrm{GL}(2,\mathbb{Z})$ such that*

$$A = E \begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix} F$$

for some $t_1, t_2 \in \mathbb{Z}$ with $t_1 \mid t_2$. The diagonal matrix $\begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix}$, which is unique up to the signs of t_1 and t_2, is called the Smith normal form *of A. Moreover, E, F, t_1, and t_2 can be computed in polynomial time.*

DEFINITION 2.4. *If H is a subgroup of G, then the sets $gH = \{gh : h \in H\}$ and $Hg = \{hg : h \in H\}$, for $g \in G$, are called the* left *and* right cosets *of H in G, respectively. An element g is called a* representative *of the left coset gH (respectively, of the right coset Hg).*

The collection of left cosets or right cosets of H form a disjoint partition of G. Moreover, the number of left cosets is equal to the number of right cosets, and this number is called the index *of H in G, denoted $|G : H|$.*

For every natural $n \geq 1$, let us define the following subgroups of $\mathrm{GL}(2,\mathbb{Z})$:

$$H(n) = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathrm{GL}(2,\mathbb{Z}) \ : \ n \text{ divides } a_{21} \right\},$$

$$F(n) = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathrm{GL}(2,\mathbb{Z}) \ : \ n \text{ divides } a_{12} \right\}.$$

Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ be any matrix from $\mathrm{GL}(2,\mathbb{Z})$ and let $D = \begin{bmatrix} m & 0 \\ 0 & mn \end{bmatrix}$ be a diagonal matrix in the Smith normal form, where $m, n \neq 0$. Then the conjugate of A with D is equal to $A^D = D^{-1}AD = \begin{bmatrix} a_{11} & na_{12} \\ \frac{1}{n}a_{21} & a_{22} \end{bmatrix}$. From this formula we see that if $A^D \in \mathrm{GL}(2,\mathbb{Z})$, then n divides a_{21}. On the other hand, if a_{21} is divisible by n, then A^D is in $\mathrm{GL}(2,\mathbb{Z})$, and in fact in $F(n)$. Thus we have the following criterion.

PROPOSITION 2.3. *Suppose A is in $\mathrm{GL}(2,\mathbb{Z})$ and D is a diagonal matrix of the above form, then $A^D \in \mathrm{GL}(2,\mathbb{Z})$ if and only if $A \in H(n)$. Moreover, if $A \in H(n)$, then $A^D \in F(n)$.*

THEOREM 2.3. *The subgroups $H(n)$ and $F(n)$ have finite index in $\mathrm{GL}(2,\mathbb{Z})$. Furthermore, there is an algorithm that for a given n computes representatives of the left and right cosets of $H(n)$ and $F(n)$ in $\mathrm{GL}(2,\mathbb{Z})$.*

Proof. We will only show how to compute representatives of the left cosets of $H(n)$ because the other cases are similar. For each pair of indices i, j such that $0 \leq i, j \leq n-1$, let us define a matrix $W_{i,j}$ as follows. Let $W_{i,0}$ be the identity matrix for $i = 0, \ldots, n-1$. If $j > 0$, then consider $d = \gcd(i,j)$ and let i_0 and j_0 be such that $i = i_0 d$ and $j = j_0 d$. Since i_0, j_0 are relatively prime, there exist integers u and v such that $ui_0 + vj_0 = 1$. Hence if we let $W_{i,j} = \begin{bmatrix} u & v \\ -j_0 & i_0 \end{bmatrix}$, then $W_{i,j}$ belongs to $\mathrm{GL}(2,\mathbb{Z})$.

Now, consider an arbitrary matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ from $\mathrm{GL}(2,\mathbb{Z})$. Let $a_{11} = i + nk$ and $a_{21} = j + nl$, where $0 \leq i, j \leq n-1$. We will show that $W_{i,j}A \in H(n)$. If $j = 0$, then $a_{21} = nl$ is divisible by n, and hence $A \in H(n)$. Since we defined $W_{i,0}$ to be the identity matrix, it follows that $W_{i,0}A = A \in H(n)$. If $j > 0$, then let $d = \gcd(i,j)$ and let i_0, j_0 be such that $i = i_0 d$ and $j = j_0 d$. In this case

$$W_{i,j}A = \begin{bmatrix} u & v \\ -j_0 & i_0 \end{bmatrix} \begin{bmatrix} di_0 + nk & a_{12} \\ dj_0 + nl & a_{22} \end{bmatrix},$$

and the lower left corner of $W_{i,j}A$ is equal to

$$-j_0 d i_0 - j_0 nk + i_0 d j_0 + i_0 nl = n(-j_0 k + i_0 l),$$

which is divisible by n. Thus $W_{i,j}A \in H(n)$.

So we showed that for any matrix $A \in \mathrm{GL}(2,\mathbb{Z})$ there is a pair i, j such that $W_{i,j}A \in H(n)$ or, equivalently, $A \in W_{i,j}^{-1}H(n)$. Therefore, the collection

$$\{W_{i,j}^{-1}H(n) \ : \ 0 \leq i, j \leq n-1\}$$

contains all left cosets of $H(n)$ in $\mathrm{GL}(2,\mathbb{Z})$. In particular, the index of $H(n)$ in $\mathrm{GL}(2,\mathbb{Z})$ is bounded by n^2.

Note that some of the cosets in

$$\{W_{i,j}^{-1}H(n) \ : \ 0 \leq i, j \leq n-1\}$$

may be equal to each other. In fact, two cosets $W_{i_1,j_1}^{-1}H(n)$ and $W_{i_2,j_2}^{-1}H(n)$ are equal if and only if $W_{i_1,j_1}W_{i_2,j_2}^{-1} \in H(n)$. Since the domain of the subgroup $H(n)$ is a computable set, the equality of two cosets is a decidable property. Therefore, we can algorithmically choose a collection of pairwise nonequivalent representatives of the left cosets of $H(n)$ in $\mathrm{GL}(2,\mathbb{Z})$. □

LEMMA 2.1. *Let $L_{H(n)}$ and $L_{F(n)}$ be the languages that correspond to the subgroups $H(n)$ and $F(n)$, respectively, that is,*

$$L_{H(n)} = \{w \in \Sigma^* : \varphi(w) \in H(n)\} \quad \text{and}$$
$$L_{F(n)} = \{w \in \Sigma^* : \varphi(w) \in F(n)\}.$$

Then $L_{H(n)}$ and $L_{F(n)}$ are regular languages.

Proof. We will show that $L_{H(n)}$ is regular by constructing an automaton $\mathcal{A}_{H(n)}$ that recognizes it. The proof for $L_{F(n)}$ is similar.

We will use the fact that $H(n)$ has a finite index in $\mathrm{GL}(2,\mathbb{Z})$. Actually, the states of the automaton $\mathcal{A}_{H(n)}$ will be encoded by the right cosets of $H(n)$. Namely, let $U_0, U_1, \ldots, U_k$ be pairwise nonequivalent representatives of the right cosets of $H(n)$ in $\mathrm{GL}(2,\mathbb{Z})$, which can be computed by Theorem 2.3. We will assume that U_0 is the identity matrix. The automaton $\mathcal{A}_{H(n)}$ will have $k+1$ states $u_0, u_1, \ldots, u_k$, where u_0 is the only initial and the only final state of $\mathcal{A}_{H(n)}$. The transitions of $\mathcal{A}_{H(n)}$ are defined as follows:

(2.2) $\mathcal{A}_{H(n)}$ has a transition $u_i \xrightarrow{\sigma} u_j$ if and only if $U_i\varphi(\sigma)U_j^{-1} \in H(n)$.

This equivalence will play an important role in the construction of the automaton $\mathcal{F}_D(\mathcal{A})$ from Theorem 2.4 below, which is in turn a crucial ingredient of our main algorithm.

Note that for every i and σ, there is exactly one j such that

$$U_i\varphi(\sigma) \in H(n)U_j \quad \text{or equivalently}$$
$$U_i\varphi(\sigma)U_j^{-1} \in H(n),$$

which means that the automaton $\mathcal{A}_{H(n)}$ is deterministic.

We now show that the language of $\mathcal{A}_{H(n)}$ is equal to $L_{H(n)}$. Take any word $w = \sigma_1\sigma_2\ldots\sigma_t \in \Sigma^*$ and consider a run $\rho = u_{i_0}u_{i_1}\ldots u_{i_t}$ of $\mathcal{A}_{H(n)}$ on w. Note that $i_0 = 0$, and $u_{i_0} = u_0$ is the initial state. Since $\mathcal{A}_{H(n)}$ has transitions $u_{i_{s-1}} \xrightarrow{\sigma_s} u_{i_s}$, for $s = 1, \ldots, t$, we have that

$$U_{i_{s-1}}\varphi(\sigma_s)U_{i_s}^{-1} \in H(n) \text{ for } s = 1, \ldots, t.$$

Since by assumption $U_{i_0} = U_0 = I$, we can rewrite $\varphi(w) = \varphi(\sigma_1)\varphi(\sigma_2)\ldots\varphi(\sigma_t)$ as $\varphi(w) =$

$$(U_{i_0}\varphi(\sigma_1)U_{i_1}^{-1})(U_{i_1}\varphi(\sigma_2)U_{i_2}^{-1})\cdots(U_{i_{t-1}}\varphi(\sigma_t)U_{i_t}^{-1})U_{i_t}.$$

If $u_{i_t} = u_0$, that is, if w is accepted by $\mathcal{A}_{H(n)}$, then $i_t = 0$ and $U_{i_t} = U_0 = I \in H(n)$. This implies that $\varphi(w) \in H(n)$ because for all $s = 1, \ldots, t$ we have $U_{i_{s-1}}\varphi(\sigma_s)U_{i_s}^{-1} \in H(n)$.

On the other hand, if $\varphi(w) \in H(n)$, then it must be that $U_{i_t} \in H(n)$, which can only happen if $i_t = 0$ and hence $u_{i_t} = u_0$. This means that w is accepted by $\mathcal{A}_{H(n)}$. Therefore, we proved that $L(\mathcal{A}_{H(n)}) = L_{H(n)}$. □

Example. Let us consider the case when $n = 2$. It's not hard to compute that the subgroup $H(2)$ has index 3 in $\mathrm{GL}(2,\mathbb{Z})$ and representatives of the right cosets of $H(2)$ in $\mathrm{GL}(2,\mathbb{Z})$ are

$$U_0 = I, \quad U_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \text{ and } U_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

These representatives correspond to the states u_0, u_1 and u_2 of the automaton $\mathcal{A}_{H(2)}$, respectively, where u_0 is the only initial and the only final state. In order to compute the transitions of $\mathcal{A}_{H(2)}$, we will use the equivalence (2.2) above. For instance, we have

$$U_0\varphi(R)U_1^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \in H(2).$$

Therefore, $\mathcal{A}_{H(2)}$ has a transition $u_0 \xrightarrow{R} u_1$. The full transition diagram of $\mathcal{A}_{H(2)}$ is given in Figure 1.

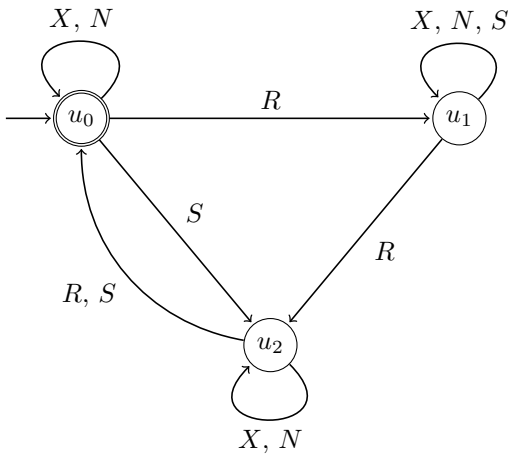

Figure 1: The transition diagram of $\mathcal{A}_{H(2)}$.

Now, for any automaton $\mathcal{A}$ with alphabet Σ we construct two automata $\mathrm{Inv}(\mathcal{A})$ and $\mathcal{F}_D(\mathcal{A})$, where D is a diagonal matrix in the Smith normal form. The automaton $\mathrm{Inv}(\mathcal{A})$ recognizes inverses to the words from $L(\mathcal{A})$, that is:

(1) For every $w \in L(\mathcal{A})$, there exists $w' \in L(\mathrm{Inv}(\mathcal{A}))$ such that $\varphi(w') = \varphi(w)^{-1}$, and

(2) For every $w' \in L(\mathrm{Inv}(\mathcal{A}))$, there exists $w \in L(\mathcal{A})$ such that $\varphi(w) = \varphi(w')^{-1}$.

In other words, for any matrix $A \in \mathrm{GL}(2,\mathbb{Z})$,

$$A \in \varphi(L(\mathcal{A})) \text{ if and only if } A^{-1} \in \varphi(L(\mathrm{Inv}(\mathcal{A}))).$$

Construction of the automaton $\mathrm{Inv}(\mathcal{A})$**.** We will make use of the following equivalences, which are easy to check: $X^{-1} \sim X$, $N^{-1} \sim N$, $S^{-1} \sim S^3$, and

$R^{-1} \sim R^5$. Informally speaking, to construct $\mathrm{Inv}(\mathcal{A})$ we want to reverse the transitions in $\mathcal{A}$ and replace the labels by their inverses. More formally, $\mathrm{Inv}(\mathcal{A})$ will have the same states as $\mathcal{A}$ plus some newly added states as explained below. The initial states of $\mathrm{Inv}(\mathcal{A})$ are the final states of $\mathcal{A}$, and the final states of $\mathrm{Inv}(\mathcal{A})$ are the initial states of $\mathcal{A}$. For every transitions of the form $q \xrightarrow{X} q'$ and $q \xrightarrow{N} q'$ in $\mathcal{A}$ we add the transitions $q' \xrightarrow{X} q$ and $q' \xrightarrow{N} q$ to $\mathrm{Inv}(\mathcal{A})$, respectively. Furthermore, for every transitions of the form $q \xrightarrow{S} q'$ and $q \xrightarrow{R} q'$ in $\mathcal{A}$ we add the paths

$$q' \xrightarrow{S} p_1 \xrightarrow{S} p_2 \xrightarrow{S} q$$

and

$$q' \xrightarrow{R} p_3 \xrightarrow{R} p_4 \xrightarrow{R} p_5 \xrightarrow{R} p_6 \xrightarrow{R} q$$

to $\mathrm{Inv}(\mathcal{A})$, respectively, where $p_1, p_2, \ldots, p_6$ are newly added states. It is not hard to verify that $\mathrm{Inv}(\mathcal{A})$ has the desired properties.

The purpose of the automaton $\mathcal{F}_D(\mathcal{A})$ is to recognize conjugates of the words from $L(\mathcal{A})$ with a matrix D. Its construction is given in the next theorem. We will be only interested in those conjugates that have integer coefficients, that is, belong to $\mathrm{GL}(2,\mathbb{Z})$. Recall that by Proposition 2.3, the conjugate A^D of a matrix A belongs to $\mathrm{GL}(2,\mathbb{Z})$ if and only if $A \in H(n)$. This explains why in Theorem 2.4 we consider only those words w that belong to the intersection $L(\mathcal{A}) \cap L_{H(n)}$.

THEOREM 2.4. *Let $\mathcal{A}$ be an automaton with alphabet Σ and let $D = \begin{bmatrix} m & 0 \\ 0 & mn \end{bmatrix}$ be a diagonal matrix in the Smith normal form, where $m, n \neq 0$. Then there is an automaton $\mathcal{F}_D(\mathcal{A})$ with the following properties:*

1. *For every $w \in L(\mathcal{A}) \cap L_{H(n)}$, there exists $w' \in L(\mathcal{F}_D(\mathcal{A}))$ such that $\varphi(w') = \varphi(w)^D$.*
2. *For every $w' \in L(\mathcal{F}_D(\mathcal{A}))$, there exists $w \in L(\mathcal{A}) \cap L_{H(n)}$ such that $\varphi(w)^D = \varphi(w')$.*

In other words, $\varphi(L(\mathcal{F}_D(\mathcal{A}))) = \{\varphi(w)^D : \text{where } w \in L(\mathcal{A}) \text{ and } \varphi(w) \in H(n)\}$.

Informal description of the automaton $\mathcal{F}_D(\mathcal{A})$. One of the ideas behind the construction of $\mathcal{F}_D(\mathcal{A})$ is to use the following property of conjugation: $(AB)^D = A^D B^D$ for any 2×2 matrices A and B. This property allows us to compute the conjugate of every word from $L(\mathcal{A})$ symbol by symbol. For example, if we take $w = RSRS$, then

$$\varphi(w)^D = \varphi(R)^D \varphi(S)^D \varphi(R)^D \varphi(S)^D.$$

The main problem with this approach is that $\varphi(S)^D$ and $\varphi(R)^D$ may have fractional coefficients, and hence do not belong to $\mathrm{GL}(2,\mathbb{Z})$, while $\varphi(w)^D$ may still be in $\mathrm{GL}(2,\mathbb{Z})$. For example, if $D = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ then

$$\varphi(S)^D = \begin{bmatrix} 0 & -2 \\ \frac{1}{2} & 0 \end{bmatrix} \text{ and } \varphi(R)^D = \begin{bmatrix} 0 & -2 \\ \frac{1}{2} & 1 \end{bmatrix}$$

but

$$\varphi(RSRS)^D = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}^D = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

To overcome this problem, we will use the following trick: Recall that we are only interested in those conjugates that have integer coefficients. By Proposition 2.3, if $\varphi(w)^D$ is in $\mathrm{GL}(2,\mathbb{Z})$, then w is in $L_{H(n)}$, and hence there is an accepting run of the automaton $\mathcal{A}_{H(n)}$ on w.

Consider again the word $w = RSRS$ and the matrix D from the example above. As we have shown, $\varphi(w)^D \in \mathrm{GL}(2,\mathbb{Z})$ and hence w is in $L_{H(2)}$. Let u_0, u_1, u_1, u_2, u_0 be an accepting run[2] of $\mathcal{A}_{H(2)}$ on w. Now, we rewrite $\varphi(RSRS)$ as follows

$$\varphi(RSRS) = (U_0\varphi(R)U_1^{-1})(U_1\varphi(S)U_1^{-1})(U_1\varphi(R)U_2^{-1})(U_2\varphi(S)U_0^{-1}). \tag{2.3}$$

This is a valid equation because in the proof of Lemma 2.1 we assumed that U_0 is the identity matrix. By the equivalence (2.2) from Lemma 2.1, each of the factors $U_0\varphi(R)U_1^{-1}$, $U_1\varphi(S)U_1^{-1}$, $U_1\varphi(R)U_2^{-1}$, and $U_2\varphi(S)U_0^{-1}$ is in $H(2)$ and so the matrices $(U_0\varphi(R)U_1^{-1})^D$, $(U_1\varphi(S)U_1^{-1})^D$, $(U_1\varphi(R)U_2^{-1})^D$, and $(U_2\varphi(S)U_0^{-1})^D$ are in $\mathrm{GL}(2,\mathbb{Z})$, that is, have only integer coefficients. Thus we can compute each of these factors separately and then multiply them to obtain $\varphi(w)^D$. The key point here is that by doing so we can compute the value of $\varphi(w)^D$ term by term without introducing fractional coefficients, that is, without going outside of $\mathrm{GL}(2,\mathbb{Z})$.

The role of the matrices U_0, U_1, U_2 in the equation (2.3) from the above example is to replace the fractional matrices $\varphi(S)^D$ and $\varphi(R)^D$ with integer matrices $(U_0\varphi(R)U_1^{-1})^D$, $(U_1\varphi(S)U_1^{-1})^D$, $(U_1\varphi(R)U_2^{-1})^D$, and $(U_2\varphi(S)U_0^{-1})^D$. An interesting point here is that the matrices U_0, U_1, U_1, U_2, U_0 correspond to the accepting run u_0, u_1, u_1, u_2, u_0 of $\mathcal{A}_{H(2)}$ on $RSRS$. This trick works because $RSRS$ is accepted by $\mathcal{A}_{H(2)}$ and hence in the equation (2.3) the first and the last matrix is $U_0 = U_0^{-1} = I$. If we take a word w not accepted by $\mathcal{A}_{H(2)}$, then we obtain an invalid equation because

[2] We remind the reader that u_0 is the only initial and the only final state of $\mathcal{A}_{H(2)}$.

the last matrix will be different from U_0^{-1}. For example, u_0, u_1, u_1, u_2 is a run of $\mathcal{A}_{H(2)}$ on RSR, but since $U_2 \neq I$, we have the inequality

$$\varphi(RSR) \neq (U_0\varphi(R)U_1^{-1})(U_1\varphi(S)U_1^{-1})(U_1\varphi(R)U_2^{-1}).$$

This gives an idea how to construct the automaton $\mathcal{F}_D(\mathcal{A})$. We need to replace every transition $u_i \xrightarrow{\sigma} u_j$ in $\mathcal{A}_{H(n)}$ with a new path that corresponds to the conjugate of $U_i\varphi(\sigma)U_j^{-1}$ with D. Namely, we compute the matrix $(U_i\varphi(\sigma)U_j^{-1})^D$, which by the equivalence (2.2) from Lemma 2.1 belongs to $\mathrm{GL}(2,\mathbb{Z})$, and find a canonical word v such that $\varphi(v) = (U_i\varphi(\sigma)U_j^{-1})^D$. After that we replace the transition $u_i \xrightarrow{\sigma} u_j$ in $\mathcal{A}_{H(n)}$ with a new path labelled by the word v as explained in the example below.

Example. Let $D = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and consider the transition $u_0 \xrightarrow{R} u_1$ of the automaton $\mathcal{A}_{H(2)}$, which is shown on Figure 1. We have

$$\begin{aligned}(U_0\varphi(R)U_1^{-1})^D &= \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^D = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \\ &= \varphi(RRSRRS).\end{aligned}$$

Hence the transition $u_0 \xrightarrow{R} u_1$ is replaced by a path

$$u_0 \xrightarrow{R} p_1 \xrightarrow{R} p_2 \xrightarrow{S} p_3 \xrightarrow{R} p_4 \xrightarrow{R} p_5 \xrightarrow{S} u_1,$$

where $p_1, \ldots, p_5$ are newly added states. Similarly, for the transition $u_1 \xrightarrow{S} u_1$ of $\mathcal{A}_{H(2)}$ we have

$$\begin{aligned}(U_1\varphi(S)U_1^{-1})^D &= \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix}^D = \begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix} \\ &= \varphi(SRSRRS).\end{aligned}$$

Hence the transition $u_1 \xrightarrow{S} u_1$ is replaced by a path

$$u_1 \xrightarrow{S} p_6 \xrightarrow{R} p_7 \xrightarrow{S} p_8 \xrightarrow{R} p_9 \xrightarrow{R} p_{10} \xrightarrow{S} u_1,$$

where $p_6, \ldots, p_{10}$ are newly added states. On the other hand, the transition $u_1 \xrightarrow{R} u_2$ of $\mathcal{A}_{H(2)}$ is replaced by an ε-transition $u_1 \xrightarrow{\varepsilon} u_2$ because

$$(U_1\varphi(R)U_2^{-1})^D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^D = I = \varphi(\varepsilon).$$

Note that the automaton constructed in this way recognizes conjugates of all words from $L_{H(n)}$ with matrix D. However we are interested only in those words that belong to $L(\mathcal{A})$. So before applying the above procedure we first need to take the direct product of $\mathcal{A}$ and $\mathcal{A}_{H(n)}$. Formally, this construction with all necessary details is explained below.

Formal proof of Theorem 2.4. Let $\mathcal{A}$ be a finite automaton with alphabet Σ and let $D = \begin{bmatrix} m & 0 \\ 0 & mn \end{bmatrix}$ be a diagonal matrix in the Smith normal form, where $m, n \neq 0$.

Suppose that $\mathcal{A}$ has the states $q_0, q_1, \ldots, q_t$. Recall from the proof of Lemma 2.1 that the automaton $\mathcal{A}_{H(n)}$, which recognizes $L_{H(n)}$, has the states $u_0, u_1, \ldots, u_k$, where u_0 is the only initial and the only final state. First, we construct an automaton $\mathcal{A}'$ for the language $L(\mathcal{A}) \cap L_{H(n)}$ by taking the direct product of $\mathcal{A}$ and $\mathcal{A}_{H(n)}$. Namely, $\mathcal{A}'$ has the states (q_i, u_j), for $i = 0, \ldots, t$ and $j = 0, \ldots, k$. The initial states of $\mathcal{A}'$ are of the form (q_i, u_0), where q_i is an initial state of $\mathcal{A}$, and the final states of $\mathcal{A}'$ are of the form (q_i, u_0), where q_i is a final state of $\mathcal{A}$. Furthermore, there is a transition from (q_i, u_j) to $(q_{i'}, u_{j'})$ labelled by σ if and only if there are transitions $q_i \xrightarrow{\sigma} q_{i'}$ and $u_j \xrightarrow{\sigma} u_{j'}$ in $\mathcal{A}$ and $\mathcal{A}_{H(n)}$, respectively.

Next we replace every transition in $\mathcal{A}'$ with a new path as follows. Let $(q_{i_1}, u_{j_1}) \xrightarrow{\sigma} (q_{i_2}, u_{j_2})$ be a transition in $\mathcal{A}'$. Therefore, there must be a transition $u_{j_1} \xrightarrow{\sigma} u_{j_2}$ in $\mathcal{A}_{H(n)}$. By the equivalence (2.2) from Lemma 2.1 we have that $U_{j_1}\varphi(\sigma)U_{j_2}^{-1}$ is in $H(n)$. Hence, by Proposition 2.3, $(U_{j_1}\varphi(\sigma)U_{j_2}^{-1})^D$ belongs to $\mathrm{GL}(2,\mathbb{Z})$. Let $v = \sigma_1 \ldots \sigma_s \in \Sigma^*$ be a canonical word[3] such that $\varphi(v) = (U_{j_1}\varphi(\sigma)U_{j_2}^{-1})^D$. Then we replace the transition $(q_{i_1}, u_{j_1}) \xrightarrow{\sigma} (q_{i_2}, u_{j_2})$ with a path of the form

$$(q_{i_1}, u_{j_1}) \xrightarrow{\sigma_1} p_1 \xrightarrow{\sigma_2} \cdots \xrightarrow{\sigma_{s-1}} p_{s-1} \xrightarrow{\sigma_s} (q_{i_2}, u_{j_2}),$$

where $p_1, \ldots, p_{s-1}$ are new states added to $\mathcal{A}'$. Let $\mathcal{F}_D(\mathcal{A})$ be an automaton that we obtain after applying the above procedure to $\mathcal{A}'$.

To prove the first property, take any $w = \sigma_1 \ldots \sigma_s \in L(\mathcal{A}) \cap L_{H(n)}$. Then there must be an accepting run

$$\rho = (q_{i_0}, u_{j_0})(q_{i_1}, u_{j_1}) \ldots (q_{i_s}, u_{j_s})$$

of $\mathcal{A}'$ on w. By the construction, for every transition $(q_{i_{r-1}}, u_{j_{r-1}}) \xrightarrow{\sigma_r} (q_{i_r}, u_{j_r})$ in the run ρ, there is a path in $\mathcal{F}_D(\mathcal{A})$ from $(q_{i_{r-1}}, u_{j_{r-1}})$ to (q_{i_r}, u_{j_r}) labelled by a word w_r such that $\varphi(w_r) = (U_{j_{r-1}}\varphi(\sigma_r)U_{j_r}^{-1})^D$, where $U_{j_{r-1}}\varphi(\sigma_r)U_{j_r}^{-1} \in H(n)$. If we let $w' = w_1 \ldots w_s$, then w' is accepted by $\mathcal{F}_D(\mathcal{A})$. To prove that $\varphi(w') = \varphi(w)^D$, we first note that since $w \in L_{H(n)}$, the run $u_{j_0}u_{j_1} \ldots u_{j_s}$

[3] Actually, we can take v to be any word that represents $(U_{j_1}\varphi(\sigma)U_{j_2}^{-1})^D$. The fact that it is canonical is not important for our construction.

is an accepting run of $\mathcal{A}_{H(n)}$ on w, and in particular $j_0 = j_s = 0$. Since $U_{j_0} = U_{j_s} = U_0 = I$, we can rewrite $\varphi(w)$ as

$$\varphi(w) = U_{j_0}^{-1}(U_{j_0}\varphi(\sigma_1)U_{j_1}^{-1})(U_{j_1}\varphi(\sigma_2)U_{j_2}^{-1})\cdots \cdots(U_{j_{s-1}}\varphi(\sigma_s)U_{j_s}^{-1})U_{j_s} = = (U_{j_0}\varphi(\sigma_1)U_{j_1}^{-1})(U_{j_1}\varphi(\sigma_2)U_{j_2}^{-1})\cdots(U_{j_{s-1}}\varphi(\sigma_s)U_{j_s}^{-1}).$$

In the last equality we used the fact that $U_{j_0} = U_{j_s} = I$. Recall that for each $r = 1, \ldots, s$, we have $\varphi(w_r) = (U_{j_{r-1}}\varphi(\sigma_r)U_{j_r}^{-1})^D$. Therefore, $\varphi(w)^D =$

$$(U_{j_0}\varphi(\sigma_1)U_{j_1}^{-1})^D(U_{j_1}\varphi(\sigma_2)U_{j_2}^{-1})^D\cdots(U_{j_{s-1}}\varphi(\sigma_s)U_{j_s}^{-1})^D = \varphi(w_1)\varphi(w_2)\cdots\varphi(w_s) = \varphi(w').$$

This proves the first property of $\mathcal{F}_D(\mathcal{A})$.

To prove the second property of $\mathcal{F}_D(\mathcal{A})$, take any $w' \in L(\mathcal{F}_D(\mathcal{A}))$ and consider an accepting run of $\mathcal{F}_D(\mathcal{A})$ on w'. This run passes through some states of the form (q_i, u_j), that are present in both $\mathcal{F}_D(\mathcal{A})$ and $\mathcal{A}'$, and some new states that exist only in $\mathcal{F}_D(\mathcal{A})$. Let $(q_{i_0}, u_{j_0}), (q_{i_1}, u_{j_1}), \ldots, (q_{i_s}, u_{j_s})$ be the subsequence of the states of the first type which appear in the accepting run of $\mathcal{F}_D(\mathcal{A})$. They naturally divide w' into subwords $w' = w_1 w_2 \ldots w_s$, where w_r is a label of the path from $(q_{i_{r-1}}, u_{j_{r-1}})$ to (q_{i_r}, u_{j_r}) for $r = 1, \ldots, s$. By construction of $\mathcal{F}_D(\mathcal{A})$, for each $r = 1, \ldots, s$, there exists a symbol $\sigma_r \in \Sigma$ for which there is a transition $(q_{i_{r-1}}, u_{j_{r-1}}) \xrightarrow{\sigma_r} (q_{i_r}, u_{j_r})$ in $\mathcal{A}'$ and, moreover, $U_{j_{r-1}}\varphi(\sigma_r)U_{j_r}^{-1} \in H(n)$ and $\varphi(w_r) = (U_{j_{r-1}}\varphi(\sigma_r)U_{j_r}^{-1})^D$.

Let $w = \sigma_1\sigma_2\ldots\sigma_s$, then $q_{i_0}q_{i_1}\ldots q_{i_s}$ will be an accepting run of $\mathcal{A}$ on w and $u_{j_0}u_{j_1}\ldots u_{j_s}$ will be an accepting run of $\mathcal{A}_{H(n)}$ on w. Thus $w \in L(\mathcal{A}) \cap L_{H(n)}$. Furthermore, we have $u_{j_0} = u_{j_s} = u_0$ and hence $U_{j_0} = U_{j_s} = I$. So we can rewrite $\varphi(w)$

$$\varphi(w) = U_{j_0}^{-1}(U_{j_0}\varphi(\sigma_1)U_{j_1}^{-1})(U_{j_1}\varphi(\sigma_2)U_{j_2}^{-1})\cdots \cdots(U_{j_{s-1}}\varphi(\sigma_s)U_{j_s}^{-1})U_{j_s} = = (U_{j_0}\varphi(\sigma_1)U_{j_1}^{-1})(U_{j_1}\varphi(\sigma_2)U_{j_2}^{-1})\cdots(U_{j_{s-1}}\varphi(\sigma_s)U_{j_s}^{-1}).$$

From this we obtain the following equalities

$$\varphi(w)^D = (U_{j_0}\varphi(\sigma_1)U_{j_1}^{-1})^D(U_{j_1}\varphi(\sigma_2)U_{j_2}^{-1})^D\cdots \cdots(U_{j_{s-1}}\varphi(\sigma_s)U_{j_s}^{-1})^D = \varphi(w_1)\varphi(w_2)\cdots\varphi(w_s) = \varphi(w').$$

This proves the second property of $\mathcal{F}_D(\mathcal{A})$. □

Proposition 2.4. *Let D be a diagonal matrix in the Smith normal form and let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two regular subsets of $\mathrm{GL}(2,\mathbb{Z})$. Then it is decidable whether there exist matrices $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ which satisfy the equation $A_1 D A_2 = D$.*

Proof. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be finite automata such that $\mathcal{S}_1 = L(\mathcal{A}_1)$ and $\mathcal{S}_2 = L(\mathcal{A}_2)$, respectively. We will show that the equation $A_1 D A_2 = D$ has a solution for some $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ if and only if[4]

$$L(\mathrm{Can}(\mathcal{F}_D(\mathcal{A}_1))) \cap L(\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_2))) \neq \emptyset.$$

The proof of the proposition then follows from the fact that the intersection emptiness problem for regular languages is decidable.

First, suppose there exist matrices $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ such that $A_1 D A_2 = D$. Let $w_1 \in L(\mathcal{A}_1)$ and $w_2 \in L(\mathcal{A}_2)$ be such that $\varphi(w_1) = A_1$ and $\varphi(w_2) = A_2$, respectively. Also let $D = \begin{bmatrix} m & 0 \\ 0 & mn \end{bmatrix}$ for some $m, n \neq 0$. We can rewrite the equation $A_1 D A_2 = D$ as $A_1^D = A_2^{-1}$. From this we can see that the matrix A_1^D must have integer coefficients. Hence, by Proposition 2.3, $A_1 \in H(n)$ and $w_1 \in L_{H(n)}$. Since $w_1 \in L(\mathcal{A}_1) \cap L_{H(n)}$, there is $w_1' \in L(\mathcal{F}_D(\mathcal{A}_1))$ such that $\varphi(w_1') = \varphi(w_1)^D = A_1^D$. Furthermore, there is $w_2' \in L(\mathrm{Inv}(\mathcal{A}_2))$ such that $\varphi(w_2') = \varphi(w_2)^{-1} = A_2^{-1}$. Since $A_1^D = A_2^{-1}$, we have $\varphi(w_1') = \varphi(w_2')$. In other words, w_1' and w_2' are equivalent. Let w be a canonical word such that $w \sim w_1' \sim w_2'$, then $w \in L(\mathrm{Can}(\mathcal{F}_D(\mathcal{A}_1))) \cap L(\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_2)))$.

Now, suppose there is a word w that belongs to $L(\mathrm{Can}(\mathcal{F}_D(\mathcal{A}_1))) \cap L(\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_2)))$. Hence there are words w_1' and w_2' such that $w \sim w_1' \sim w_2'$ and $w_1' \in L(\mathcal{F}_D(\mathcal{A}_1))$ and $w_2' \in L(\mathrm{Inv}(\mathcal{A}_2))$. Therefore, there exists $w_1 \in L(\mathcal{A}_1) \cap L_{H(n)}$ such that $\varphi(w_1)^D = \varphi(w_1')$. Also there exists $w_2 \in L(\mathcal{A}_2)$ such that $\varphi(w_2)^{-1} = \varphi(w_2')$. Let $A_1 = \varphi(w_1)$ and $A_2 = \varphi(w_2)$. Then we have $A_1^D = \varphi(w_1)^D = \varphi(w_1') = \varphi(w_2') = \varphi(w_2)^{-1} = A_2^{-1}$, which is equivalent to $A_1 D A_2 = D$. Moreover, since $w_1 \in L(\mathcal{A}_1)$ and $w_2 \in L(\mathcal{A}_2)$, we have that $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$. □

Corollary 2.3. *Let M_1 and M_2 be nonsingular matrices from $\mathbb{Z}^{2\times 2}$ and let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular subsets of $\mathrm{GL}(2,\mathbb{Z})$. Then it is decidable whether there exist matrices $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ such that $A_1 M_1 A_2 = M_2$.*

Proof. The idea of the proof is to compute the Smith normal forms of M_1 and M_2 and then reduce the equation $A_1 M_1 A_2 = M_2$ for regular subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ to an equation of the form $A_1 D A_2 = D$ for different regular subsets $\mathcal{S}_1'$ and $\mathcal{S}_2'$, where D is a diagonal matrix in the Smith normal form.

Let D_1 and D_2 be the Smith normal forms of M_1 and M_2, respectively, that is,

$$M_1 = E_1 D_1 F_1 \quad \text{and} \quad M_2 = E_2 D_2 F_2$$

[4] We remind that the construction of the automaton $\mathrm{Can}(\mathcal{A})$ is described in the Appendix.

for some $E_1, F_1, E_2, F_2 \in \mathrm{GL}(2,\mathbb{Z})$. Without loss of generality, we can assume that D_1 and D_2 have strictly positive diagonal coefficients. Note that if the equation $A_1 M_1 A_2 = M_2$ has a solution for some $A_1, A_2 \in \mathrm{GL}(2,\mathbb{Z})$, then, by Theorem 2.2, M_1 and M_2 must have the same Smith normal form. Therefore, if $D_1 \neq D_2$, then the equation does not have a solution.

So suppose that $D = D_1 = D_2$ is the Smith normal form of M_1 and M_2. Then $A_1 M_1 A_2 = M_2$ is equivalent to $A_1(E_1 D F_1)A_2 = E_2 D F_2$, which we can rewrite as

$$(E_2^{-1} A_1 E_1) D (F_1 A_2 F_2^{-1}) = D.$$

Let

$$\mathcal{S}_1' = \{E_2^{-1} A E_1 : A \in \mathcal{S}_1\} \text{ and}$$
$$\mathcal{S}_2' = \{F_1 A F_2^{-1} : A \in \mathcal{S}_2\}.$$

Then $\mathcal{S}_1'$ and $\mathcal{S}_2'$ are regular subsets of $\mathrm{GL}(2,\mathbb{Z})$ because E_1, F_1, E_2, and F_2 are some fixed matrices. Now, it is not hard to see that the equation $A_1 M_1 A_2 = M_2$ has a solution A_1, A_2 such that $A_1 \in \mathcal{S}_1$ and $A_1 \in \mathcal{S}_2$ if and only if the equation $A_1' D A_2' = D$ has a solution A_1', A_2' such that $A_1' \in \mathcal{S}_1'$ and $A_2' \in \mathcal{S}_2'$. By Proposition 2.4, this problem is decidable. □

2.3 General case: $A_1 M_1 \dots A_{t-1} M_{t-1} A_t = M_t$. In order to prove an analogue of Corollary 2.3 in the general case, we will extend the construction of the automaton $\mathcal{F}_D(\mathcal{A})$ to build an automaton $\mathcal{F}(\mathcal{A}_1, \dots, \mathcal{A}_{t-1}, M_1, \dots, M_{t-1}; M_t)$ (where $\mathcal{A}_1, \dots, \mathcal{A}_{t-1}$ are finite automata in alphabet Σ and $M_1, \dots, M_{t-1}, M_t$ are nonsingular matrices from $\mathbb{Z}^{2\times 2}$) which will have the following properties:

(1) If $w_1 \in L(\mathcal{A}_1), \dots, w_{t-1} \in L(\mathcal{A}_{t-1})$ and there is a matrix $A \in \mathrm{GL}(2,\mathbb{Z})$ which satisfies the equation

$$\varphi(w_1) M_1 \dots \varphi(w_{t-1}) M_{t-1} A = M_t,$$

then there is

$$w \in L(\mathcal{F}(\mathcal{A}_1, \dots, \mathcal{A}_{t-1}, M_1, \dots, M_{t-1}; M_t))$$

such that

$$\varphi(w_1) M_1 \dots \varphi(w_{t-1}) M_{t-1} \varphi(w)^{-1} = M_t$$

(and hence $A = \varphi(w)^{-1}$).

(2) If $w \in L(\mathcal{F}(\mathcal{A}_1, \dots, \mathcal{A}_{t-1}, M_1, \dots, M_{t-1}; M_t))$, then there are $w_1 \in L(\mathcal{A}_1)$, …, $w_{t-1} \in L(\mathcal{A}_{t-1})$ such that

$$\varphi(w_1) M_1 \dots \varphi(w_{t-1}) M_{t-1} \varphi(w)^{-1} = M_t.$$

Construction of the finite automaton $\mathcal{F}(\mathcal{A}_1, \dots, \mathcal{A}_{t-1}, M_1, \dots, M_{t-1}; M_t)$. The construction will be done by induction on t. We will use the following notations: If $\mathcal{A}_1$ and $\mathcal{A}_2$ are finite automata in alphabet Σ, then $\mathcal{A}_1 \cdot \mathcal{A}_2$ denotes the concatenation of $\mathcal{A}_1$ and $\mathcal{A}_2$. If $\mathcal{A}$ is an automaton and $w \in \Sigma^*$, then $\mathcal{A} \cdot w$ denotes an automaton that recognizes the language $L(\mathcal{A}) \cdot \{w\} = \{uw : u \in L(\mathcal{A})\}$. Similarly, $w \cdot \mathcal{A}$ is an automaton that recognizes $\{w\} \cdot L(\mathcal{A}) = \{wu : u \in L(\mathcal{A})\}$.

First, we construct an automaton $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$, which will serve as a base for induction. Let D_1 and D_2 be diagonal matrices with nonnegative coefficients which are equal to the Smith normal forms of M_1 and M_2, respectively. If $D_1 \neq D_2$, then define $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$ to be an automaton that accepts the empty language. Otherwise, let $D = D_1 = D_2$ be the common Smith normal form of M_1 and M_2, and suppose $M_1 = E_1 D F_1$ and $M_2 = E_2 D F_2$ for some matrices $E_1, F_1, E_2, F_2 \in \mathrm{GL}(2,\mathbb{Z})$. Let $w(E_1)$, $w(F_1)$, $w(E_2^{-1})$ and $w(F_2^{-1})$ be canonical words that represent the matrices E_1, F_1, E_2^{-1} and F_2^{-1}, respectively, and define $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$ to be the following automaton

$$\mathcal{F}(\mathcal{A}_1, M_1; M_2) = w(F_2^{-1}) \cdot \mathcal{F}_D\big(w(E_2^{-1}) \cdot \mathcal{A}_1 \cdot w(E_1)\big) \cdot w(F_1).$$

The proof that the automaton $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$ indeed satisfies the desired properties is given in the following proposition.

PROPOSITION 2.5. *Let $\mathcal{A}_1$ be a finite automaton in alphabet Σ, and let M_1 and M_2 be nonsingular matrices from $\mathbb{Z}^{2\times 2}$. Then the automaton $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$ has the following properties:*

(1) If $w_1 \in L(\mathcal{A}_1)$ and there is a matrix $A \in \mathrm{GL}(2,\mathbb{Z})$ which satisfies the equation $\varphi(w_1) M_1 A = M_2$, then there is $w \in L(\mathcal{F}(\mathcal{A}_1, M_1; M_2))$ such that $\varphi(w_1) M_1 \varphi(w)^{-1} = M_2$ (and hence $A = \varphi(w)^{-1}$).

(2) If $w \in L(\mathcal{F}(\mathcal{A}_1, M_1; M_2))$, then there is $w_1 \in L(\mathcal{A}_1)$ such that $\varphi(w_1) M_1 \varphi(w)^{-1} = M_2$.

Proof. Note that if M_1 and M_2 have different Smith normal forms, then by the uniqueness part of Theorem 2.2 the equation $A_1 M_1 A_2 = M_2$ cannot have a solution $A_1, A_2 \in \mathrm{GL}(2,\mathbb{Z})$. Therefore, in this case both properties of $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$ are trivially satisfied. Now, suppose that $D = \begin{bmatrix} m & 0 \\ 0 & mn \end{bmatrix}$ is the common Smith normal form of M_1 and M_2 and let E_1, F_1, E_2, F_2 be matrices form $\mathrm{GL}(2,\mathbb{Z})$ such that $M_1 = E_1 D F_1$ and $M_2 = E_2 D F_2$.

To see that the first property of $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$ holds, let's take any $w_1 \in L(\mathcal{A}_1)$ for which there is a matrix $A \in \mathrm{GL}(2,\mathbb{Z})$ that satisfies the equation $\varphi(w_1)M_1A = M_2$. Hence we have that

$$\varphi(w_1)E_1DF_1A = E_2DF_2,$$

which is equivalent to

$$F_2^{-1}(E_2^{-1}\varphi(w_1)E_1)^DF_1 = A^{-1}.$$

Because F_2^{-1}, F_1, and A^{-1} are matrices from $\mathrm{GL}(2,\mathbb{Z})$, we conclude that $(E_2^{-1}\varphi(w_1)E_1)^D$ is in $\mathrm{GL}(2,\mathbb{Z})$. Then, by Proposition 2.3, we have $E_2^{-1}\varphi(w_1)E_1 \in H(n)$ or, equivalently,

$$w(E_2^{-1}) \cdot w_1 \cdot w(E_1) \in L_{H(n)}.$$

By the first property of Theorem 2.4, there exists $w' \in L\big(\mathcal{F}_D\big(w(E_2^{-1}) \cdot \mathcal{A}_1 \cdot w(E_1)\big)\big)$ such that

$$\varphi(w') = \varphi\big(w(E_2^{-1}) \cdot w_1 \cdot w(E_1)\big)^D = (E_2^{-1}\varphi(w_1)E_1)^D.$$

Let $w = w(F_2^{-1}) \cdot w' \cdot w(F_1)$. Then obviously w is in $L(\mathcal{F}(\mathcal{A}_1, M_1; M_2))$. Moreover,

$$\varphi(w) = F_2^{-1}\varphi(w')F_1 = F_2^{-1}(E_2^{-1}\varphi(w_1)E_1)^DF_1.$$

The last equation is equivalent to

$$\varphi(w_1)E_1DF_1\varphi(w)^{-1} = E_2DF_2,$$

which is the same as $\varphi(w_1)M_1\varphi(w)^{-1} = M_2$. Hence the first property holds.

We now prove the second property of $\mathcal{F}(\mathcal{A}_1, M_1; M_2)$. Let's take any $w \in L(\mathcal{F}(\mathcal{A}_1, M_1; M_2))$. Then there exists $w' \in L\big(\mathcal{F}_D\big(w(E_2^{-1}) \cdot \mathcal{A}_1 \cdot w(E_1)\big)\big)$ such that $w = w(F_2^{-1}) \cdot w' \cdot w(F_1)$. By the second property of the construction $\mathcal{F}_D$, there exists $w_1 \in L(\mathcal{A}_1)$ such that

$$w(E_2^{-1}) \cdot w_1 \cdot w(E_1) \in L_{H(n)}$$

and

$$\varphi(w') = \varphi\big(w(E_2^{-1}) \cdot w_1 \cdot w(E_1)\big)^D.$$

The last two conditions are equivalent to the facts that $E_2^{-1}\varphi(w_1)E_1 \in H(n)$ and $\varphi(w') = (E_2^{-1}\varphi(w_1)E_1)^D$. From the equation $w = w(F_2^{-1}) \cdot w' \cdot w(F_1)$ we have that $\varphi(w) = F_2^{-1}\varphi(w')F_1$. Therefore,

$$\varphi(w) = F_2^{-1}(E_2^{-1}\varphi(w_1)E_1)^DF_1.$$

The last equation is equivalent to $\varphi(w_1)E_1DF_1\varphi(w)^{-1} = E_2DF_2$, which is the same as $\varphi(w_1)M_1\varphi(w)^{-1} = M_2$. This proves the second property. □

We now explain how to construct an automaton $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$. For convenience, the description of this construction is enclosed in the Proposition 2.6 below.

The next lemma will play an important role in the proof of the inductive step in Proposition 2.6. Informally speaking, it states that when we consider all possible Smith normal forms UDV for a fixed D, we can assume that U comes from a finite set of matrices.

LEMMA 2.2. *Let $D = \begin{bmatrix} m & 0 \\ 0 & mn \end{bmatrix}$ be a diagonal matrix in the Smith normal form and let $U_0, \ldots, U_k$ be representatives of the right cosets of $H(n)$ in $\mathrm{GL}(2,\mathbb{Z})$. Then*

$$\big\{ UDV \ : \ U, V \in \mathrm{GL}(2,\mathbb{Z})\big\} = \bigcup_{i=0}^{k} \big\{U_iDV \ : \ V \in \mathrm{GL}(2,\mathbb{Z})\big\}.$$

Proof. Consider a matrix $M = UDV$ for some $U, V \in \mathrm{GL}(2,\mathbb{Z})$ and choose i such that $U \in U_iH(n)$. In this case we have that $U_i^{-1}U \in H(n)$, and thus $(U_i^{-1}U)^D$ belongs to $\mathrm{GL}(2,\mathbb{Z})$ by Proposition 2.3. Let

$$V' = (U_i^{-1}U)^DV \in \mathrm{GL}(2,\mathbb{Z}).$$

Then we have an equality

$$M = UDV = U_iDD^{-1}U_i^{-1}UDV = U_iDV',$$

and hence

$$M \in \big\{U_iDV \ : \ V \in \mathrm{GL}(2,\mathbb{Z})\big\}.$$

The inclusion in the other direction is obvious. □

PROPOSITION 2.6. *Let $\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}$ be finite automata in alphabet Σ, and $M_1, \ldots, M_{t-1}, M_t$ be nonsingular matrices from $\mathbb{Z}^{2\times 2}$. Then there is an automaton $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$ which has the following properties:*

(1) If $w_1 \in L(\mathcal{A}_1), \ldots, w_{t-1} \in L(\mathcal{A}_{t-1})$ and there is a matrix $A \in \mathrm{GL}(2,\mathbb{Z})$ which satisfies the equation

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}A = M_t,$$

then there is

$$w \in L(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))$$

such that

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}\varphi(w)^{-1} = M_t$$

(and hence $A = \varphi(w)^{-1}$).

(2) If $w \in L(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))$*, then there are* $w_1 \in L(\mathcal{A}_1)$, $\ldots$, $w_{t-1} \in L(\mathcal{A}_{t-1})$ *such that*

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}\varphi(w)^{-1} = M_t.$$

Proof. The case when $t = 2$ was proved in Proposition 2.5. Suppose that the proposition holds for $t-1$, and thus we have a construction of the automata of the form $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-2}, M_1, \ldots, M_{t-2}; M_{t-1})$ which satisfy the properties (1) and (2) above. Using induction on t, we will show how to construct an automaton $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$.

Let $D_{t-1} = \begin{bmatrix} m & 0 \\ 0 & mn \end{bmatrix}$ be equal to the Smith normal form of the matrix M_{t-1} and let $U_0, \ldots, U_k$ be representatives of the right cosets of $H(n)$, which can be computed by Theorem 2.3. Then we define $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$ to be an automaton that recognizes the following union of regular languages

$$\bigcup_{i=0}^{k} L\Big(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-3}, \mathcal{A}_{t-2}, M_1, \ldots, M_{t-3}, M_{t-2}U_iD_{t-1}; M_t) \cdot \mathcal{F}(\mathcal{A}_{t-1}, M_{t-1}; U_iD_{t-1})\Big).$$

To see that the first property holds for $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$, let's take

$$w_1 \in L(\mathcal{A}_1), \ldots, w_{t-1} \in L(\mathcal{A}_{t-1}),$$

and suppose there is a matrix $A \in \mathrm{GL}(2, \mathbb{Z})$ which satisfies the equation

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}A = M_t.$$

By Lemma 2.2, there is $i \in \{0, \ldots, k\}$ and $V \in \mathrm{GL}(2, \mathbb{Z})$ such that $\varphi(w_{t-1})M_{t-1}A = U_iD_{t-1}V$. So the above equation is equivalent to the following system of equations

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-2})M_{t-2}U_iD_{t-1}V = M_t$$
$$\varphi(w_{t-1})M_{t-1}AV^{-1} = U_iD_{t-1}.$$

Since $V \in \mathrm{GL}(2, \mathbb{Z})$, by the inductive hypothesis there is a word u such that

$$u \in L\Big(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-3}, \mathcal{A}_{t-2}, M_1, \ldots, M_{t-3}, M_{t-2}U_iD_{t-1}; M_t)\Big)$$

and

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-2})M_{t-2}U_iD_{t-1}\varphi(u)^{-1} = M_t.$$

Moreover, since $AV^{-1} \in \mathrm{GL}(2, \mathbb{Z})$, by Proposition 2.5, there is a word $v \in L(\mathcal{F}(\mathcal{A}_{t-1}, M_{t-1}; U_iD_{t-1}))$ such that $\varphi(w_{t-1})M_{t-1}\varphi(v)^{-1} = U_iD_{t-1}$. Combining the last two equations together we obtain that

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}\varphi(v)^{-1}\varphi(u)^{-1} = M_t$$

or, equivalently,

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}\varphi(uv)^{-1} = M_t.$$

Note that

$$uv \in L\Big(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-3}, \mathcal{A}_{t-2}, M_1, \ldots, M_{t-3}, M_{t-2}U_iD_{t-1}; M_t) \cdot \mathcal{F}(\mathcal{A}_{t-1}, M_{t-1}; U_iD_{t-1})\Big)$$

and hence $uv \in L(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))$. Therefore, property (1) holds.

To show the second property, let's take $w \in L(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))$. Then there is $i \in \{0, \ldots, k\}$ such that

$$w \in L\Big(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-3}, \mathcal{A}_{t-2}, M_1, \ldots, M_{t-3}, M_{t-2}U_iD_{t-1}; M_t) \cdot \mathcal{F}(\mathcal{A}_{t-1}, M_{t-1}; U_iD_{t-1})\Big).$$

Therefore, there are words u and v such that

$$u \in L\Big(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-3}, \mathcal{A}_{t-2}, M_1, \ldots, M_{t-3}, M_{t-2}U_iD_{t-1}; M_t)\Big)$$

and $v \in L(\mathcal{F}(\mathcal{A}_{t-1}, M_{t-1}; U_iD_{t-1}))$. By Proposition 2.5, there is $w_{t-1} \in L(\mathcal{A}_{t-1})$ such that

$$\varphi(w_{t-1})M_{t-1}\varphi(v)^{-1} = U_iD_{t-1}.$$

Furthermore, by the inductive hypothesis, there are $w_1 \in L(\mathcal{A}_1), \ldots, w_{t-2} \in L(\mathcal{A}_{t-2})$ such that

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-2})M_{t-2}U_iD_{t-1}\varphi(u)^{-1} = M_t.$$

Combining the last two equation together we obtain

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}\varphi(v)^{-1}\varphi(u)^{-1} = M_t.$$

Note that $\varphi(w)^{-1} = \varphi(v)^{-1}\varphi(u)^{-1}$, and hence we have

$$\varphi(w_1)M_1 \ldots \varphi(w_{t-1})M_{t-1}\varphi(w)^{-1} = M_t.$$

Therefore, property (2) holds. □

THEOREM 2.5. *Let* $M_1, \ldots, M_t$ *be nonsingular matrices from* $\mathbb{Z}^{2\times 2}$ *and let* $\mathcal{S}_1, \ldots, \mathcal{S}_t$ *be regular subsets of* $\mathrm{GL}(2, \mathbb{Z})$*. Then there is an algorithm that decides whether there exist matrices* $A_1 \in \mathcal{S}_1, \ldots, A_t \in \mathcal{S}_t$ *such that* $A_1M_1 \ldots A_{t-1}M_{t-1}A_t = M_t$.

Proof. Let $\mathcal{A}_1, \ldots, \mathcal{A}_t$ be finite automata such that $\mathcal{S}_i = \varphi(L(\mathcal{A}_i))$, for each $i = 1, \ldots, t$. Now, consider an automaton $\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t)$ which was constructed in the proof of Proposition 2.6. We will show the following equivalence: there exist matrices $A_1 \in \mathcal{S}_1, \ldots, A_t \in \mathcal{S}_t$ that satisfy the equation $A_1 M_1 \ldots A_{t-1} M_{t-1} A_t = M_t$ if and only if

$$L\big(\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_t))\big) \cap L\big(\mathrm{Can}(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))\big) \neq \emptyset.$$

The statement of the theorem then follows from the decidability of the emptiness problem for regular languages.

First, suppose that there are matrices $A_1 \in \mathcal{S}_1, \ldots, A_t \in \mathcal{S}_t$ that satisfy the equation $A_1 M_1 \ldots A_{t-1} M_{t-1} A_t = M_t$. Then there are words $w_1 \in L(\mathcal{A}_1), \ldots, w_t \in L(\mathcal{A}_t)$ such that

$$\varphi(w_1) M_1 \ldots \varphi(w_{t-1}) M_{t-1} \varphi(w_t) = M_t.$$

By property (1) of Proposition 2.6, there is a word $u \in L(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))$ such that

$$\varphi(w_1) M_1 \ldots \varphi(w_{t-1}) M_{t-1} \varphi(u)^{-1} = M_t.$$

In particular, we have $\varphi(w_t) = \varphi(u)^{-1}$. Furthermore, by the construction of $\mathrm{Inv}(\mathcal{A}_t)$, there is a word $v \in L(\mathrm{Inv}(\mathcal{A}_t))$ such that $\varphi(v) = \varphi(w_t)^{-1}$. So we have $\varphi(u) = \varphi(w_t)^{-1} = \varphi(v)$, that is, $u \sim v$. Let w be the canonical word that is equivalent to u and v. Then

$$w \in L\big(\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_t))\big) \cap L\big(\mathrm{Can}(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))\big).$$

On the other hand, suppose there is a word w such that

$$w \in L\big(\mathrm{Can}(\mathrm{Inv}(\mathcal{A}_t))\big) \cap L\big(\mathrm{Can}(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))\big).$$

Then there are words u and v such that $u \sim v \sim w$ and $u \in L(\mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}, M_1, \ldots, M_{t-1}; M_t))$ and $v \in L(\mathrm{Inv}(\mathcal{A}_t))$. Hence there is $w_t \in L(\mathcal{A}_t)$ such that $\varphi(w_t) = \varphi(v)^{-1}$. Also by property (2) of Proposition 2.6, there are words $w_1 \in L(\mathcal{A}_1), \ldots, w_{t-1} \in L(\mathcal{A}_{t-1})$ such that

$$\varphi(w_1) M_1 \ldots \varphi(w_{t-1}) M_{t-1} \varphi(u)^{-1} = M_t.$$

Since $v \sim u$, we have that $\varphi(u)^{-1} = \varphi(v)^{-1} = \varphi(w_t)$. Therefore, the above equation is equivalent to

$$\varphi(w_1) M_1 \ldots \varphi(w_{t-1}) M_{t-1} \varphi(w_t) = M_t.$$

Now, if we let $A_1 = \varphi(w_1), \ldots, A_t = \varphi(w_t)$, then for each $i = 1, \ldots, t$ we have $A_i \in \mathcal{S}_i$, and these matrices satisfy the equation $A_1 M_1 \ldots A_{t-1} M_{t-1} A_t = M_t$. □

3 Appendix

Construction of the automaton $\mathrm{Can}(\mathcal{A})$**.** Below is a detailed description of the automaton $\mathrm{Can}(\mathcal{A})$ which is used in the proofs of Propositions 2.2 and 2.4 and Theorem 2.5.

Let $\mathcal{A}$ be a finite automaton with alphabet Σ. We will construct a new automaton $\mathrm{Can}(\mathcal{A})$ such that the language of $\mathrm{Can}(\mathcal{A})$ contains only canonical words and $\mathrm{Can}(\mathcal{A}) \sim \mathcal{A}$, that is, $\varphi(L(\mathrm{Can}(\mathcal{A}))) = \varphi(L(\mathcal{A}))$. In order to do this, we will define a sequence of transformations called Red, F_N and F_X which will have the following properties:

- $\mathrm{Can}(\mathcal{A}) = F_X \circ \mathrm{Red} \circ F_N(\mathcal{A})$,
- $L(F_N(\mathcal{A})) \subseteq \{X, S, R\}^* \cup N\{X, S, R\}^*$, that is, $F_N(\mathcal{A})$ accepts only those words that have at most one occurrence of N which may appear only in the first position,
- $L(\mathrm{Red} \circ F_N(\mathcal{A})) \subseteq \{X, S, R\}^* \cup N\{X, S, R\}^*$ and, moreover, $\mathrm{Red} \circ F_N(\mathcal{A})$ accepts only those words that do not contain subwords of the form XX, $SX^{\alpha}S$ and $RX^{\alpha_1}RX^{\alpha_2}R$ for any $\alpha, \alpha_1, \alpha_2 \in \{0, 1\}$,
- $F_X \circ \mathrm{Red} \circ F_N(\mathcal{A})$ accepts only canonical words,
- finally, we will have the equivalences

$$\mathcal{A} \sim F_N(\mathcal{A}) \sim \mathrm{Red} \circ F_N(\mathcal{A}) \sim \sim F_X \circ \mathrm{Red} \circ F_N(\mathcal{A}) = \mathrm{Can}(\mathcal{A}).$$

We now describe each of these transformations in detail.

Transformation F_N**.** We will make use of the following equivalences which can be easily verified: $X \sim NXN$, $S \sim NXSN$, and $R \sim NSR^2SN$.

First, for every transition $q \xrightarrow{X} q'$ which appears in $\mathcal{A}$, we add new states p_1, p_2 and a new path of the form

$$q \xrightarrow{N} p_1 \xrightarrow{X} p_2 \xrightarrow{N} q'.$$

Note that since $X \sim NXN$, the addition of such paths produces an equivalent automaton. Similarly, for any transition $q \xrightarrow{S} q'$ in $\mathcal{A}$, we add new states p_1, p_2, p_3 and a path

$$q \xrightarrow{N} p_1 \xrightarrow{X} p_2 \xrightarrow{S} p_3 \xrightarrow{N} q'.$$

Finally, for any transition $q \xrightarrow{R} q'$ in $\mathcal{A}$, we add new states p_1, p_2, p_3, p_4, p_5 and a path

$$q \xrightarrow{N} p_1 \xrightarrow{S} p_2 \xrightarrow{R} p_3 \xrightarrow{R} p_4 \xrightarrow{S} p_5 \xrightarrow{N} q'.$$

Again, the addition of such paths produces an equivalent automaton. Let us call this automaton $\mathcal{A}_1$.

Now, for every pair of states q, q' in $\mathcal{A}_1$, which are connected by a path labelled with NN, we add an ε-transition $q \xrightarrow{\varepsilon} q'$. We repeat this procedure iteratively until no new ε-transitions of this type can be added. Let $\mathcal{A}_2$ be the resulting automaton. Note that since NN is equivalent to the empty word, which represents the identity matrix I, the automaton $\mathcal{A}_2$ is equivalent to $\mathcal{A}_1$ and hence to $\mathcal{A}$.

Let $F_N(\mathcal{A})$ be an automaton that recognizes the intersection

$$L(\mathcal{A}_2) \cap (\{X,S,R\}^* \cup N\{X,S,R\}^*).$$

Obviously, the language of $F_N(\mathcal{A})$ is a subset of $\{X,S,R\}^* \cup N\{X,S,R\}^*$, so we only need to show that $F_N(\mathcal{A}) \sim \mathcal{A}$. Take any $w_1 \in L(F_N(\mathcal{A}))$, then $w_1 \in L(\mathcal{A}_2)$ and since $\mathcal{A}_2 \sim \mathcal{A}$, there is $w_2 \in L(\mathcal{A})$ such that $w_1 \sim w_2$. Next, we need to prove that for any $w_2 \in L(\mathcal{A})$, there is $w_1 \in L(F_N(\mathcal{A}))$ such that $w_2 \sim w_1$.

Let us take any $w_2 \in L(\mathcal{A})$. To construct the required word w_1, we first need to find all occurrences of letter N in w_2. For example, suppose that

$$w_2 = u_1 N u_2 N \dots u_{n-1} N u_n,$$

where each $u_i \in \{X,S,R\}^*$. If the number of N's is odd, then in each subword u_i with odd i we replace every occurrence of X, S, and R with NXN, $NXSN$, and NSR^2SN, respectively, and leave u_i's with even i unchanged. On the other hand, if the number of N's is even, then we apply such substitution to each u_i with even i and leave u_i's with odd i unchanged. Let w' be the resulting word. Then by construction $w' \sim w_2$ and $w' \in L(\mathcal{A}_1)$. Next, we repeatedly remove all occurrences of the subword NN from w'. This will give us a word $w_1 \sim w' \sim w_2$ such that $w_1 \in L(\mathcal{A}_2)$ and w_1 contains at most one letter N, which may appear only in the first position. Hence $w_1 \in L(F_N(\mathcal{A}))$. This idea is illustrated by the following example. Let

$$w_2 = SXNRNRSNS \in L(\mathcal{A}),$$

so w_2 contains an odd number of N's and hence

$$\begin{aligned} w' &= (NXSN)(NXN)NRN(NSR^2SN)(NXSN)NS \\ &= NXS(NN)X(NN)R(NN)SR^2S(NN)RS(NN)XS(NN)S. \end{aligned}$$

In the above formula the parentheses are inserted only to visually separated subwords in w'. After removing subwords NN from w' we obtain

$$w_1 = NXSXRSR^2SXSS \in L(F_N(\mathcal{A}))$$

such that $w_1 \sim w_2$. The next example illustrates the same idea for an even number of N's. Let

$$w_2 = SXNRNRSNSN \in L(\mathcal{A}),$$

then

$$\begin{aligned} w' &= SXN(NSR^2SN)NRSN(NXSN)N \\ &= SX(NN)SR^2S(NN)RS(NN)XS(NN). \end{aligned}$$

After removing NN from w' we obtain

$$w_1 = SXSR^2SRSXS \in L(F_N(\mathcal{A}))$$

such that $w_1 \sim w_2$. This completes the proof that $F_N(\mathcal{A}) \sim \mathcal{A}$.

Transformation Red**.** To construct $\mathrm{Red} \circ F_N(\mathcal{A})$ from $F_N(\mathcal{A})$ we will make use of the following equivalences $SS \sim X$ and $RRR \sim X$. We will also use the fact that X commutes with S, R, and N, and that XX is equivalent to the empty word.

First, we apply the following procedure to $F_N(\mathcal{A})$:

(1) For any pair of states q, q' in $F_N(\mathcal{A})$ that are connected by a path labelled with XX, we add an ε-transition $q \xrightarrow{\varepsilon} q'$.

(2) For any pair of states q, q' in $F_N(\mathcal{A})$ that are connected by a path labelled with $SX^\alpha S$, where $\alpha \in \{0,1\}$ (recall that X^0 denotes the empty word), we add a new transition $q \xrightarrow{X^\beta} q'$, where $\beta = 1-\alpha$.

(3) For any pair of states q, q' in $F_N(\mathcal{A})$ that are connected by a path labelled with $RX^{\alpha_1}RX^{\alpha_2}R$, where $\alpha_1, \alpha_2 \in \{0,1\}$, we add a new transition $q \xrightarrow{X^\gamma} q'$, where $\gamma \in \{0,1\}$ is such that

$$\gamma \equiv \alpha_1 + \alpha_2 + 1 \mod 2.$$

We repeat the above steps iteratively until no new transitions can be added.

Let $\mathcal{A}'$ be the resulting automaton. By construction, we have $\mathcal{A}' \sim F_N(\mathcal{A})$. Let $\mathcal{L}_{\mathrm{Red}}$ be the regular language which consists of all words in alphabet Σ that do not contain subwords of the form XX, $SX^\alpha S$ and $RX^{\alpha_1}RX^{\alpha_2}R$ for any $\alpha, \alpha_1, \alpha_2 \in \{0,1\}$. Define $\mathrm{Red} \circ F_N(\mathcal{A})$ as an automaton that accepts the language $L(\mathcal{A}') \cap \mathcal{L}_{\mathrm{Red}}$. It is not hard to see that the language of $\mathrm{Red} \circ F_N(\mathcal{A})$ is contained in

$$\mathcal{L}_{\mathrm{Red}} \cap (\{X,S,R\}^* \cup N\{X,S,R\}^*).$$

What is left to show is that $\mathrm{Red} \circ F_N(\mathcal{A}) \sim F_N(\mathcal{A})$. If $w_1 \in L(\mathrm{Red} \circ F_N(\mathcal{A}))$, then $w_1 \in L(\mathcal{A}')$, and hence $w_1 \sim w_2$ for some $w_2 \in L(F_N(\mathcal{A}))$ because $\mathcal{A}' \sim F_N(\mathcal{A})$. On the other hand, if $w_2 \in L(F_N(\mathcal{A}))$, then we can repeatedly remove subwords XX from w_2 and replace subwords of the form $SX^\alpha S$ and $RX^{\alpha_1}RX^{\alpha_2}R$, for $\alpha, \alpha_1, \alpha_2 \in \{0,1\}$, with X^β and X^γ, respectively, where $\beta = 1-\alpha$ and $\gamma \in \{0,1\}$ is such that

$$\gamma \equiv \alpha_1 + \alpha_2 + 1 \mod 2.$$

Let w_1 be a resulting word that does not contain subwords XX, $SX^{\alpha}S$ and $RX^{\alpha_1}RX^{\alpha_2}R$ for any $\alpha, \alpha_1, \alpha_2 \in \{0,1\}$. Then $w_1 \sim w_2$ and

$$w_1 \in L(\mathcal{A}') \cap \mathcal{L}_{\mathrm{Red}} = L(\mathrm{Red} \circ F_N(\mathcal{A})).$$

Transformation F_X. The words accepted by $\mathrm{Red} \circ F_N(\mathcal{A})$ are almost in canonical form with the exception that the letter X may appear in the middle of a word. To get rid of such X's we use a similar idea as in the construction of $F_N(\mathcal{A})$. Namely, we will use the following equivalences: $S \sim XSX$ and $R \sim XRX$. Note that we will not need the equivalence $N \sim XNX$ because the letter N can appear only at the beginning of a word.

To construct $\mathrm{Can}(\mathcal{A}) = F_X \circ \mathrm{Red} \circ F_N(\mathcal{A})$ from $\mathrm{Red} \circ F_N(\mathcal{A})$, we do the following. First, for every transition $q \xrightarrow{S} q'$ which appears in $\mathrm{Red} \circ F_N(\mathcal{A})$, we add new states p_1, p_2 and a new path of the form

$$q \xrightarrow{X} p_1 \xrightarrow{S} p_2 \xrightarrow{X} q'.$$

Similarly, for every transition $q \xrightarrow{R} q'$ which appears in $\mathrm{Red} \circ F_N(\mathcal{A})$, we add new states p_1, p_2 and a new path of the form

$$q \xrightarrow{X} p_1 \xrightarrow{R} p_2 \xrightarrow{X} q'.$$

After that we iteratively add ε transitions $q \xrightarrow{\varepsilon} q'$ for every pair of states q, q' that are connected by a path with label XX. We do this until no new ε-transitions can be added.

Let $\mathcal{A}'$ be the resulting automaton, which is by construction equivalent to $\mathrm{Red} \circ F_N(\mathcal{A})$. Let $\mathcal{L}_{\mathrm{Can}}$ be the regular language which consists of all canonical words in alphabet Σ. Define $\mathrm{Can}(\mathcal{A}) = F_X \circ \mathrm{Red} \circ F_N(\mathcal{A})$ as an automaton that accepts the language $L(\mathcal{A}') \cap \mathcal{L}_{\mathrm{Can}}$. Therefore, $\mathrm{Can}(\mathcal{A})$ accepts only canonical words.

The proof that $\mathrm{Can}(\mathcal{A}) \sim \mathrm{Red} \circ F_N(\mathcal{A})$ is similar to the proof that $F_N(\mathcal{A}) \sim \mathcal{A}$ given above. If $w_1 \in L(\mathrm{Can}(\mathcal{A}))$, then $w_1 \in L(\mathcal{A}')$ and hence $w_1 \sim w_2$ for some $w_2 \in L(\mathrm{Red} \circ F_N(\mathcal{A}))$ because $\mathcal{A}' \sim \mathrm{Red} \circ F_N(\mathcal{A})$. On the other hand, if $w_2 \in L(\mathrm{Red} \circ F_N(\mathcal{A}))$, then to construct $w_1 \in L(\mathrm{Can}(\mathcal{A}))$ such that $w_1 \sim w_2$ we first find all occurrences of the letter X in w_2. For example, let w_2 has the form

$$w_2 = Nu_1Xu_2X \dots u_{n-1}Xu_n$$

or the form

$$w_2 = u_1Xu_2X \dots u_{n-1}Xu_n,$$

where each $u_i \in \{S, R\}^*$. If the number of X's is odd, then in each u_i with odd i we replace every occurrence of R and S with XRX and XSX, respectively, and leave u_i's with even i unchanged. If the number of X's is even, then we do the same substitution in all u_i's with even i and leave u_i's with odd i unchanged. After that we remove all occurrences of XX. If w_1 is a resulting word, then $w_1 \sim w_2$ and $w_1 \in L(\mathcal{A}')$. Moreover, since w_1 is in canonical form, we also have $w_1 \in L(\mathrm{Can}(\mathcal{A}))$. This idea is illustrated by the following example. Suppose

$$w_2 = NSRXSXRRX,$$

then after replacing suitable occurrences of R and S with XRX and XSX, respectively, we obtain the word

$$N(XSX)(XRX)XSX(XRX)(XRX)X = NXS(XX)R(XX)S(XX)R(XX)R(XX).$$

After removing all occurrences of XX we obtain the word

$$w_1 = NXSRSRR \sim w_2$$

which is in canonical form, and hence $w_1 \in L(\mathrm{Can}(\mathcal{A}))$. This completes the construction of $\mathrm{Can}(\mathcal{A})$.

References

[1] L. Babai, R. Beals, J.-y. Cai, G. Ivanyos, and E. M. Luks. Multiplicative equations over commuting matrices. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '96, pages 498–507, Philadelphia, PA, USA, 1996. Society for Industrial and Applied Mathematics.

[2] P. Bell, M. Hirvensalo, and I. Potapov. The identity problem for matrix semigroups in SL(2, Z) is NP-complete. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, 2017.

[3] P. Bell and I. Potapov. On undecidability bounds for matrix decision problems. *Theoretical Computer Science*, 391(1-2):3–13, 2008.

[4] P. Bell and I. Potapov. Reachability problems in quaternion matrix and rotation semigroups. *Information and Computation*, 206(11):1353–1361, 2008.

[5] P. C. Bell, M. Hirvensalo, and I. Potapov. Mortality for 2x2 matrices is NP-hard. In B. Rovan, V. Sassone, and P. Widmayer, editors, *Mathematical Foundations of Computer Science 2012*, volume 7464 of *Lecture Notes in Computer Science*, pages 148–159. Springer Berlin Heidelberg, 2012.

[6] P. C. Bell and I. Potapov. On the undecidability of the identity correspondence problem and its applications for word and matrix semigroups. *Int. J. Found. Comput. Sci.*, 21(6):963–978, 2010.

[7] P. C. Bell and I. Potapov. On the computational complexity of matrix semigroup problems. *Fundam. Inf.*, 116(1-4):1–13, 2012.

[8] V. D. Blondel, E. Jeandel, P. Koiran, and N. Portier. Decidable and undecidable problems about quantum automata. *SIAM J. Comput.*, 34(6):1464–1473, June 2005.

[9] V. D. Blondel and A. Megretski, editors. *Unsolved problems in mathematical systems and control theory*. Princeton, NJ: Princeton University Press, 2004. http://press.princeton.edu/math/blondel/solutions.html.

[10] J. Cassaigne, V. Halava, T. Harju, and F. Nicolas. Tighter undecidability bounds for matrix mortality, zero-in-the-corner problems, and more. *CoRR*, abs/1404.0644, 2014.

[11] J. Cassaigne, T. Harju, and J. Karhumaki. On the undecidability of freeness of matrix semigroups. *International Journal of Algebra and Computation*, 09(03n04):295–305, 1999.

[12] C. Choffrut and J. Karhumaki. Some decision problems on integer matrices. *RAIRO-Theor. Inf. Appl.*, 39(1):125–131, 2005.

[13] V. Chonev, J. Ouaknine, and J. Worrell. The orbit problem in higher dimensions. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 941–950, 2013.

[14] V. Chonev, J. Ouaknine, and J. Worrell. On the complexity of the orbit problem. *J. ACM*, 63(3):23, 2016.

[15] J. Esparza, A. Finkel, and R. Mayr. On the verification of broadcast protocols. In *Logic in Computer Science, 1999. Proceedings. 14th Symposium on*, pages 352–359, 1999.

[16] E. Galby, J. Ouaknine, and J. Worrell. On Matrix Powering in Low Dimensions. In E. W. Mayr and N. Ollinger, editors, *32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015)*, volume 30 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 329–340, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[17] A. Gerrard and J. M. Burch. *Introduction to matrix methods in optics*. Dover Publications, Inc., New York, 1994. Corrected reprint of the 1975 original.

[18] Y. Gurevich and P. Schupp. Membership problem for the modular group. *SIAM J. Comput.*, 37(2):425–459, May 2007.

[19] V. Halava, T. Harju, M. Hirvensalo, and J. Karhumaki. Skolem's problem — on the border between decidability and undecidability. Technical Report 683, Turku Centre for Computer Science, 2005.

[20] R. Kannan and A. Bachem. Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix. *SIAM J. Comput.*, 8(4):499–507, 1979.

[21] R. Kannan and R. J. Lipton. Polynomial-time algorithm for the orbit problem. *J. ACM*, 33(4):808–821, Aug. 1986.

[22] A. Lisitsa and I. Potapov. Membership and reachability problems for row-monomial transformations. In *Mathematical Foundations of Computer Science 2004, 29th International Symposium, MFCS 2004, Prague, Czech Republic, August 22-27, 2004, Proceedings*, pages 623–634, 2004.

[23] R. C. Lyndon and P. E. Schupp. *Combinatorial group theory*. Springer-Verlag, Berlin-New York, 1977. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89.

[24] W. Magnus, A. Karrass, and D. Solitar. *Combinatorial group theory*. Dover Publications, Inc., New York, revised edition, 1976. Presentations of groups in terms of generators and relations.

[25] A. Markov. On certain insoluble problems concerning matrices. *Doklady Akad. Nauk SSSR*, 57(6):539–542, June 1947.

[26] C. Nuccio and E. Rodaro. Mortality problem for 2×2 integer matrices. In *SOFSEM 2008: Theory and Practice of Computer Science, 34th Conference on Current Trends in Theory and Practice of Computer Science, Nový Smokovec, Slovakia, January 19-25, 2008, Proceedings*, pages 400–405, 2008.

[27] J. Ouaknine, J. a. S. Pinto, and J. Worrell. On termination of integer linear loops. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 957–969. SIAM, 2015.

[28] J. Ouaknine and J. Worrell. On the positivity problem for simple linear recurrence sequences,. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II*, pages 318–329, 2014.

[29] J. Ouaknine and J. Worrell. Ultimate positivity is decidable for simple linear recurrence sequences. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II*, pages 330–341, 2014.

[30] M. S. Paterson. Unsolvability in 3×3 matrices. *Studies in Appl. Math.*, 49:105–107, 1970.

[31] I. Potapov and P. Semukhin. Vector reachability problem in SL(2, Z). In *41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, August 22-26 - Kraków, Poland*, pages 84:1–14, 2016.

[32] R. A. Rankin. *Modular forms and functions*. Cambridge University Press, Cambridge-New York-Melbourne, 1977.