On the Computational Complexity of Matrix Semigroup Problems

Paul C. Bell[1] and Igor Potapov[2]

[1] Department of Computer Science, Loughborough University, Loughborough, LE11 3TU, UK, p.bell@lboro.ac.uk
[2] Department of Computer Science, University of Liverpool, Ashton Building, Ashton St, Liverpool, L69 3BX, UK, potapov@liverpool.ac.uk

Abstract. Most computational problems for matrix semigroups and groups are inherently difficult to solve and even undecidable starting from dimension three. The questions about the decidability and complexity of problems for two-dimensional matrix semigroups remain open and are directly linked with other challenging problems in the field.
In this paper we study the computational complexity of the problem of determining whether the identity matrix belongs to a matrix semigroup (the Identity Problem) generated by a finite set of 2×2 integral unimodular matrices. The Identity Problem for matrix semigroups is a well-known challenging problem, which has remained open in any dimension until recently. It is currently known that the problem is decidable in dimension two and undecidable starting from dimension four.
In particular, we show that the Identity Problem for 2×2 integral unimodular matrices is NP-hard by a reduction of the Subset Sum Problem and several new encoding techniques. An upper bound for the nontrivial decidability result by C. Choffrut and J. Karhumäki is unknown. However, we derive a lower bound on the minimum length solution to the Identity Problem for a constructible set of instances, which is *exponential* in the number of matrices of the generator set and the maximal element of the matrices. This shows that the most obvious candidate for an NP algorithm, which is to guess the shortest sequence of matrices which multiply to give the identity matrix, does not work correctly since the certificate would have a length which is exponential in the size of the instance.
Both results lead to a number of corollaries confirming the same bounds for vector reachability, scalar reachability and zero in the right upper corner problems.

1 Introduction

Most computational problems for matrix semigroups and groups are inherently difficult to solve and in many cases can even be impossible, due to a number of undecidability results which have been shown for matrices of dimension greater than two. Some examples of such problems are the membership problem (including the special cases of the mortality and identity problems), vector reachability, freeness problems and emptiness of semigroup intersections [24, 6, 12].

There are various techniques and methods for embedding universal computation into three and four dimensional matrix semigroups. In particular, in dimension three, problems such as membership, vector reachability and freeness are undecidable for integral matrices [4, 15, 21, 26]. However there are at least two problems which are still open in dimension three which are the membership problem in matrix groups and the Identity Problem in matrix semigroups.

IDENTITY PROBLEM [10]: Decide whether a given finitely generated integral matrix semigroup contains the identity matrix.

The Identity Problem for matrix semigroups is a well-known challenging problem (see "Unsolved Problems in Mathematical Systems and Control Theory" [10]) which is also equivalent to another fundamental problem in Group Theory: given a finitely generated matrix semigroup S, decide whether a subset of the generator of S generates a non-trivial group (Group Problem). The status of the Identity Problem was unknown for all dimensions starting from two until recently. It was first shown in [18] that the Identity Problem for integral matrices of dimension 2 is decidable and later in [7] it was proven that the problem is undecidable starting from dimension 4.

In this paper we investigate the complexity status of the Identity Problem for matrix semigroups of bounded dimension. In contrast to the polynomial time algorithm for the general membership problem in matrix groups generated by unimodular matrices [20], we show that the Identity Problem (membership of identity matrix) for matrix semigroups generated by 2×2 integral unimodular matrices is NP-hard.

It was previously known that the membership problem for commutative matrices of unbounded dimension with a variable sized generator [2, 3] and the bounded membership problem (for commutative 2×2 matrices over $\mathbb{N}$) are both NP-complete [14]. At the same time, the membership problem for finitely generated commutative matrix semigroups over an algebraic number field F is decidable even in polynomial time however [2].

Thus, removing the bound on the length of matrix products makes the problem to be solved much faster in the case of commutative matrices, which is not the case for 2×2 non-commutative matrices, as we show in this paper.

It is known that the bounded membership problem for the zero matrix (the k-mortality problem) is NP-hard for semigroups generated by a pair of matrices (where the dimension is variable) [9]. It is also not difficult to show that the general membership problem for 2×2 matrix semigroups is NP-hard by an encoding of the Subset Sum Problem, as stated in [13]. However, it seems a similar simple approach does not work for the case of the Identity Problem which we are interested in. This is mainly because for the Identity Problem, all elements within a product must cancel and this leads to constraints on possible encodings that can be used. The encodings shown in this paper avoid these problems by using an extended alphabet and the concepts of cycles and border letters.

We also investigate a lower bound on the minimum length solution to the Identity Problem for a constructible set of instances, which is shown to be exponential in the number of matrices of the generator
and the maximal element of the generator matrices. This shows that the most obvious candidate for an NP algorithm, which is to guess the shortest sequence of matrices which multiply to give the identity matrix, does not work correctly since the certificate would have a length which is exponential in the size of the instance.

This paper is organised as follows. The next section contains basic notations and preliminaries. In the third section we prove our first main result, that the Identity Problem for 2×2 integral matrix semigroups is NP-hard, which also implies NP-hardness of the Group Problem. Then in Section 4 we explore a number of corollaries confirming the NP-hardness for vector reachability, scalar reachability and zero in the right upper corner problems for 2×2 matrices. Finally in Section 5 we show an exponential lower bound on the minimum length solution to all of the above mentioned problems and in Section 6 we end the paper with some conclusions and directions for further research.

2 Notations and Preliminaries.

Let $M = \{M_1, M_2, \ldots, M_r\} \subseteq \mathbb{Z}^{n\times n}$ be a set of r integral matrices of dimension n. We use the notation $\langle M \rangle$ to denote the semigroup generated by M.

Given an alphabet $\Gamma = \{1, 2, \ldots, m\}$, a word w is an element $w \in \Gamma^*$. We denote the concatenation of two words u and v by either $u \cdot v$ or uv if there is no confusion. For a letter $a \in \Gamma$, we denote by $\overline{a}$ or a^{-1} the inverse letter of a, such that $a\overline{a} = \varepsilon$ where ε is the empty word. We also denote $\overline{\Gamma} = \Gamma^{-1} = \{\overline{1}, \overline{2}, \ldots, \overline{m}\}$ and for a word $w = w_1 w_2 \cdots w_n$, we denote $\overline{w} = w^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$. For a word $w = w_1 w_2 \cdots w_n$, we define $\text{pref}(w) = w_1$ and $\text{suff}(w) = w_n$, i.e. pref gives the first letter (prefix) of a word and suff gives its last letter (suffix).

Let $\Sigma = \Gamma \cup \overline{\Gamma}$. Using the notation of [1], we shall also introduce a reduction mapping which removes factors of the form $a\overline{a}$ for $a \in \Sigma$. To that end, we define the relation $\vdash \subseteq \Sigma^* \times \Sigma^*$ such that for all $w, w' \in \Sigma^*$, $w \vdash w'$ if and only if there exists $u, v \in \Sigma^*$ and $a \in \Sigma$ where $w = ua\overline{a}v$ and $w' = uv$. We may then define by $\vdash^*$ the reflexive and transitive closure of $\vdash$. The following Lemma was shown in [1].

Lemma 1. *[1] For each $w \in \Sigma^*$ there exists exactly one word $r(w) \in \Sigma^*$ such that $w \vdash^* r(w)$ does not contain any factor of the form $a\overline{a}$, with $a \in \Sigma$.*

The word $r(w)$ is called the reduced representation of word $w \in \Sigma^*$. As an example, we see that if $w = 132\overline{2}1\overline{1}\,\overline{3}\,\overline{1} \in \Sigma^*$, then $r(w) = \varepsilon$.

Using standard notations, a deterministic finite automaton (DFA) is given by quintuple $(Q, \Sigma', \delta, q_0, F)$ where Q is the set of states, Σ' is the *input alphabet*, $\delta : Q \times \Sigma' \to Q$ is the *transition function*, $q_0 \in Q$ is the initial state and $F \subseteq Q$ is the set of final states of the automaton. We may extend δ in the usual way to have domain $Q \times \Sigma'^*$. Given a deterministic finite automaton A, the language

recognised by A is denoted by $L(A) \subseteq \Sigma'^*$, i.e. for all $w \in L(A)$, it holds that $\delta(q_0, w) \in F$.

The most known problem for matrix semigroups is the Membership Problem:

Membership Problem: Given a matrix $X \in \mathbb{Z}^{n\times n}$ and a finite set of matrices $M \subseteq \mathbb{Z}^{n\times n}$, does there exist a product $Y_1Y_2\cdots Y_r$, with each $Y_i \in M$ such that $Y_1Y_2\cdots Y_r = X$? In other words, is $X \in \langle M\rangle$? In the Membership Problem, when matrix X is the identity matrix, we call this problem the Identity Problem.

3 The Identity Problem

In this section we show our first main result, that the Identity Problem is NP-hard for two-dimensional integral matrix semigroups. Our reduction will use the following well-known NP-complete problem.

Subset Sum Problem - Given a positive integer x and a finite set of positive integer values $S = \{s_1, s_2, \ldots, s_k\}$, does there exist a nonempty subset of S which sums to x?

We will require the following encoding between words over an arbitrary group alphabet and a binary group alphabet, which is well known from the literature.

Lemma 2. *Let $\Sigma' = \{z_1, z_2, \ldots, z_l\}$ be a group alphabet and $\Sigma_2 = \{c, d, \overline{c}, \overline{d}\}$ be a binary group alphabet. Define the mapping $\alpha : \Sigma' \to \Sigma_2^*$ by:*

$$\alpha(z_i) = c^i d\overline{c}^i, \alpha(\overline{z_i}) = c^i\overline{d}\overline{c}^i,$$

where $1 \leq i \leq l$. Then α is a monomorphism (see [8] for more details). Note that α can be extended to domain Σ'^* in the usual way.

Lemma 3. *[25] Let $\Sigma_2 = \{c, d, \overline{c}, \overline{d}\}$ be a binary group alphabet and define $f : \Sigma_2^* \to \mathbb{Z}^{2\times 2}$ by:*

$$f(c) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, f(\overline{c}) = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}, f(d) = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, f(\overline{d}) = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}.$$

Then mapping f is a monomorphism.

The above two morphisms give a way to map words from an arbitrary sized alphabet into the set of two-dimensional integral unimodular matrices. We will later require the following lemma concerning mappings f and α to allow us to argue about the size of elements of matrices in $f \circ \alpha$.

Lemma 4. *Let α and f be mappings as defined in Lemma 2 and Lemma 3, then:*

$$f(\alpha(z_j^s)) = f((c^j d\overline{c}^j)^s) = \begin{pmatrix} 1+4sj & -8sj^2 \\ 2s & 1-4sj \end{pmatrix}.$$

Proof. We see that $f((c^j d \overline{c}^j)^s) = f(c^j d^s \overline{c}^j)$. Clearly, $f(c^j) = \begin{pmatrix} 1 & 2j \\ 0 & 1 \end{pmatrix}$, $f(\overline{c}^j) = \begin{pmatrix} 1 & -2j \\ 0 & 1 \end{pmatrix}$ and $f(d^s) = \begin{pmatrix} 1 & 0 \\ 2s & 1 \end{pmatrix}$. Since f is a monomorphism, $f(c^j d^s \overline{c}^j) = f(c^j) f(d^s) f(\overline{c}^j)$. Multiplication of the given matrices proves the result. □

Theorem 1. *The* Identity Problem *is NP-hard for two-dimensional integral matrices.*

Proof. We shall use an encoding of the Subset Sum Problem into a set of two-dimensional integral matrices. Define an alphabet

$$\Sigma = \{1, 2, \ldots, 2k+2, \overline{1}, \overline{2}, \ldots, \overline{(2k+2)}, a, b, \overline{a}, \overline{b}\}.$$

We now define a set of words W which will encode the Subset Sum Problem (SSP) instance. Note that the length of words in the following set is not bounded by a polynomial of the size of the SSP instance, however this is only a transit step and will not cause a problem in the final encoding.

$$W = \begin{array}{ll} \{1 \cdot a^{s_1} \cdot \overline{2}, & 1 \cdot \varepsilon \cdot \overline{2}, \\ 2 \cdot a^{s_2} \cdot \overline{3}, & 2 \cdot \varepsilon \cdot \overline{3}, \\ \vdots & \vdots \\ k \cdot a^{s_k} \cdot \overline{(k+1)}, & k \cdot \varepsilon \cdot \overline{(k+1)}, \\ (k+1) \cdot \overline{a}^x \cdot \overline{(k+2)}, & \\ (k+2) \cdot b^{s_1} \cdot \overline{(k+3)}, & (k+2) \cdot \varepsilon \cdot \overline{(k+3)}, \\ (k+3) \cdot b^{s_2} \cdot \overline{(k+4)}, & (k+3) \cdot \varepsilon \cdot \overline{(k+4)}, \\ \vdots & \vdots \\ (2k+1) \cdot b^{s_k} \cdot \overline{(2k+2)}, & (2k+1) \cdot \varepsilon \cdot \overline{(2k+2)}, \\ (2k+2) \cdot \overline{b}^x \cdot \overline{1}\} \subseteq \Sigma^* & \end{array}$$

The structure of W is that each word contains *'border letters'* on the left and right from the set $\Sigma \setminus \{a, b, \overline{a}, \overline{b}\}$. We shall show that if there exists a word $w = w_1 w_2 \cdots w_j \in W^+$ such that $r(w) = \varepsilon$, then we may assume w has the form that w_1 is from the first row of W, w_2 is from the second row of W and so on, with w_j from the last row of W (and thus $j = 2k+2$). Furthermore we show that if there exists such a w, then the SSP instance has a solution.

Figure 1 shows the way in which the words of W can be combined to give the identity. The above assumption will mean that we start from node 1 of the graph and choose either a^{s_1} or ε to move to node 2. This corresponds to w_1 being equal to either $1 \cdot a^{s_1} \cdot \overline{2}$ or $1 \cdot \varepsilon \cdot \overline{2}$. We follow such non-deterministic choices from node 1 until we have a cycle of the graph. At this point, if we chose $s_{i1}, s_{i2}, \ldots, s_{il}$, such that they sum to x, then the reduced representation of w will equal the empty word as required. If there does not exist a solution to the subset sum problem, then it will not be possible to reach the empty word. We shall now prove these claims formally.

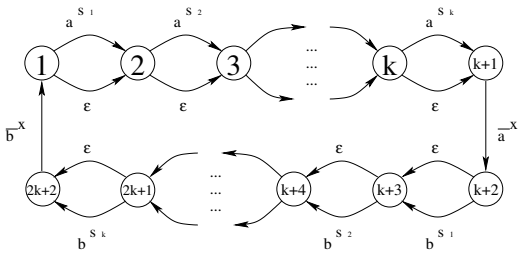

Fig. 1. The structure of a product which forms the identity.

We define *'border letters'* as letters from $\Sigma \setminus \{a, b, \overline{a}, \overline{b}\}$ and the *inner border letters* of a word as all border letters excluding the first and last. We call a word a '*partial cycle*' if all inner border letters in that word are inverse to a consecutive inner border letter. For example $2\,a^{s_2}\,\overline{3}\cdot 3\,\varepsilon\,\overline{4}\cdot 4\,a^{s_4}\,\overline{5}$ is a partial cycle, whereas $2\cdot a^{s_2}\cdot\overline{3}5\cdot a^{s_5}\cdot\overline{6}$ is not (since inner border letters $\overline{3}$ and 5 do not cancel). We call a word $u \in W^+$ '*cyclical*' if it is a partial cycle and $\text{pref}(u) = \text{suff}(u)^{-1}$.

Assume there exists $w = w_1 w_2 \cdots w_j \in W^+$ such that $r(w) = \varepsilon$. We may decompose w into subwords $w = u_1 u_2 \cdots u_m$ such that each $u_i \in W^+$ is a partial cycle of maximal size. This means that all inner border letters in each u_i are inverse to a consecutive inner border letter, but $\text{suff}(u_i)$ is not inverse to $\text{pref}(u_{i+1})$ for each $1 \leq i \leq m-1$. Note that within $r(u_i)$, all inner border letters will be cancelled.

Assume that for all u_i, we have that $r(u_i) \neq \varepsilon$. This implies that $\text{pref}(u_i)$ and $\text{suff}(u_i)$ are not cancelled within $r(u_i)$ because within $r(u_i)$ all inner border letters are cancelled and thus for the first and last border letters to be cancelled, it must be with each other, implying that the entire word must cancel. By the definition of the u_i elements, this means that $r(w) \neq \varepsilon$ since $\text{suff}(u_i)$ is not inverse to $\text{pref}(u_{i+1})$ for each $1 \leq i \leq m-1$. Thus if $r(w) = \varepsilon$, there must also exist a partial cycle u_i such that $r(u_i) = \varepsilon$. Without loss of generality, we may assume then that $w = u_i$ for some $1 \leq i \leq m$.

We may also assume that w_1 equals either $1 \cdot a^{s_1} \cdot \overline{2}$ or $1 \cdot \varepsilon \cdot \overline{2}$ (i.e. it is from the top row of W). This follows since one of these words must be in any product giving the identity by the property that each element contains a border letter i on the left and $\overline{(i+1)}$ on the right (or else $\overline{1}$), thus for complete cancellation, a word from every row must be present (we may think of this as a cycle of the graph in Figure 1 from a node back to itself). Clearly, if $w' w_1 w'' = \varepsilon$, for any $w', w'' \in W^*$, then also $w_1 w'' w' = \varepsilon$, thus we may assume w_1 is from the top row of W.

Since w is a partial cycle and $r(w) = \varepsilon$, we must also have that w is a cyclical word as defined above. This follows since if w is a partial cycle but not cyclical, then $\text{pref}(w) \neq \text{suff}(w)^{-1}$. However for a partial cycle, all of its inner border letters are cancelled within its reduced representation, and thus there is no way for letters $\text{pref}(w)$ and $\text{suff}(w)$ to be cancelled.

The above properties imply that w has the following form after cancelling the inner border letters:

$$w = 1(a^{y_1}\overline{a}^{x}b^{y_2}\overline{b}^{x})^{+}\overline{1} \quad ; y_1, y_2, x \geq 0.$$

Note that elements y_1, y_2 may be any sum of the SSP elements $s_1, s_2, \ldots, s_k \in \mathbb{Z}^+$ according to which word was chosen from each row of W, see also Figure 1. Thus, if there exists some combination equal to x (in other words a solution to the SSP instance), then letting $y_1 = y_2 = x$ means that $r(w) = \varepsilon$. If there exists no solution to the SSP instance, then clearly $y_1 \neq x$ and $y_2 \neq x$ since each y_1, y_2 is a summation of the elements s_i from the SSP instance. This means that each $r(a^{y_1}\overline{a}^{x}b^{y_2}\overline{b}^{x}) \neq \varepsilon$ implying the (nonempty) reduced representations start with a letter from $\{a, \overline{a}\}$ and end with one from $\{b, \overline{b}\}$. Thus the concatenation of such strings, when reduced, do not cancel with each other proving that if there exists no solution to the SSP instance, there does not exist a word $w \in W^+$ such that $r(w) = \varepsilon$.

We now move to the second half of the proof, to show how to encode the set W into matrices and that the resulting instance size is polynomial is the size of the SSP instance. Given an instance of the Subset Sum Problem, use monomorphism $f \circ \alpha$ defined in Lemma 2 on set W and Lemma 3 to give:

$$W' = \{f(\alpha(w_i)) | w_i \in W\} \subseteq \mathbb{Z}^{2\times 2}$$

with the same cardinality as W.

We see that there are $4k + 2$ matrices in set W'. Note that each word in W contains two border letters and possibly a power of a single letter from Σ. By Lemma 4, we see that applying $f \circ \alpha$ to each word in W thus gives a set of matrices M such that their size (i.e. the number of bits required to represent them) is a polynomial in $\sum_{j=1}^{k} s_j$ and x. This follows since Lemma 4 gives a 'size' of the matrix generated from the power of a single letter in Σ and thus proves the result.

Note that we can move directly from the SSP instance to W' without having to construct W. □

4 Related NP-hard problems

In this section we shall explore some corollaries of Theorem 1 for related problems over two dimensional integer matrix semigroups. Since the identity matrix is idempotent (i.e. $I^2 = I$), we see that $\langle I \rangle = \{I\}$. Theorem 1 proves that determining if a semigroup S (which is finitely generated by matrices from $\mathrm{SL}(2, \mathbb{Z})$) contains the identity matrix is NP-hard, therefore taking the intersection of S and $\langle I \rangle$ immediately proves the following corollary.

Corollary 1. *Determining whether the intersection of two finitely generated two-dimensional integral matrix semigroups is empty is NP-hard.*

It is not difficult to show that the problem of determining if the identity matrix belongs to a finitely generated matrix semigroup and the problem of determining if a subset of the generator set of the semigroup forms a non-trivial group are both equivalent. Since, as explained in the proof of Theorem 1, all matrices of the generator set must belong to any product equaling the identity matrix, we see that the following problem regarding groups is also NP-hard.

Corollary 2. *Given a finite set of two-dimensional integer matrices, determining if they generate a group is NP-hard.*

We now state the following problem which has received interest in the literature due to its relation to the Skolem-Pisot problem and the mortality problem:

Zero in the Right Upper Corner (ZRUC) Problem: Given a set of k integer matrices of dimension n: $M = \{M_1, M_2, \ldots, M_k\} \subseteq \mathbb{Z}^{n\times n}$, does there exist any matrix X in the semigroup generated by M such that the upper right entry of X is zero? (i.e. does there exist $X \in \langle M\rangle$ such that $X_{1,n} = 0$?) We denote by ZRUC(k, n) the problem with k matrices of dimension n.

The ZRUC problem ZRUC$(1, n)$ is closely related to the well-known Skolem-Pisot Problem which is to determine whether or not a given linear recurrent sequence (of some depth) ever contains a zero. The problem is known to be decidable for linear recurrent sequences of up to depth 5 [22], but for arbitrary depths, it is a long standing open problem. It was shown in [11] that Skolem's problem is NP-hard.

Generalizing the problem to more than one matrix, it was shown in [16] that ZRUC$(7, 3)$ and ZRUC$(2, 24)$ are both undecidable (the latter result was later improved to show the undecidability of ZRUC$(2, 10)$ in [23]). Decidability is currently (to the authors' knowledge) open for ZRUC$(k, 2)$. Using our previous considerations, we may obtain the following result concerning ZRUC$(k, 2)$.

Corollary 3. *The ZRUC$(k, 2)$ problem is NP-hard.*

Proof. Using the notation of Lemma 2, Lemma 3 and Theorem 1, we shall prove that for any word $w \in \Sigma'$, $f(\alpha(w))_{1,2} = 0$ implies $f(\alpha(w)) = I$, in other words the only possible matrix in the semigroup we generate using a subset sum problem instance with a zero in the right upper corner is the identity matrix (which Theorem 1 proves in NP-hard to decide).

According to Lemma 2, for any word $w \in \Sigma'$, we see that $\alpha(w)$ ends in $\overline{c}$. Under mapping f defined in Lemma 3, we therefore see that $f(\alpha(w))$ always has the final matrix $f(\overline{c}) = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$.

Assume for some $w \in \Sigma'$ that $f(\alpha(w))_{1,2} = 0$. Define $Y = f(\alpha(w))f(\overline{c})^{-1}$ and therefore $Y \cdot f(\overline{c})$ has a zero in the right upper corner. Let $Y = \begin{pmatrix} x & y \\ z & v \end{pmatrix}$ and we observe:

$$f(\alpha(w)) = Y \cdot f(\overline{c}) = \begin{pmatrix} x & y \\ z & v \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} x & y-2x \\ z & v-2z \end{pmatrix}$$

Since $y-2x=0$ and (since the range of f belongs to $\mathrm{SL}(2,\mathbb{Z})$) $\det(Y \cdot f(\overline{c}))=1$, then $x(v-2z)=1$. This implies $x=1$ and $(v-2z)=1$ or $x=-1$ and $(v-2z)=-1$. In the latter case we can square the matrix and still obtain a zero in the upper right corner, thus assume $x=v-2z=1$. Thus we see that $f(\alpha(w))=f(d)^{z/2}$. However, we note that $d \notin \{\alpha(w')|w' \in \Sigma'^*\}$. Actually for any $w' \in \Sigma'^*$, $\alpha(w')$ is either empty or has first letter c and last letter $\overline{c}$ by the proof of α being a monomorphism. Thus $z=0$ which implies that $f(\alpha(w))=I$ in the case that the upper right element is zero. Determining if I is in the semigroup was shown to be NP-hard in Theorem 1 thus completing the proof. □

It is clear that in Theorem 1 the only diagonal matrix possibly present in the generated semigroup is the identity matrix. This immediately gives the following corollary.

Corollary 4. *Determining whether a finitely generated two-dimensional integer matrix semigroup contains any diagonal matrix is NP-hard.*

The same problem over integer matrices of dimension four was shown to be undecidable in [5]. Corollary 3 allows us to also prove the following two problems are NP-hard:

Scalar Reachability Problem: Given a set of k integer matrices of dimension n: $M=\{M_1,M_2,\ldots,M_k\} \subseteq \mathbb{Z}^{n\times n}$, two vectors $\rho,\tau \in \mathbb{Z}^n$ and some integer d. Does there exist any matrix X in the semigroup generated by M such that $\rho^T X \tau = k$?

Vector Reachability Problem: Given a set of k integer matrices of dimension n: $M=\{M_1,M_2,\ldots,M_k\} \subseteq \mathbb{Z}^{n\times n}$ and two vectors $\rho,\tau \in \mathbb{Z}^n$. Does there exist any matrix X in the semigroup generated by M such that $X\rho=\tau$?

Corollary 5. *The* Scalar Reachability Problem *and* Vector Reachability Problem *are NP-hard over two-dimensional integer matrices.*

Proof. We first prove the result concerning the Scalar Reachability Problem. Let $d=0$, $\rho_1=(1,0)^T$ and $\tau=(0,1)^T$. Let the matrix semigroup $\langle M \rangle$ be generated as in Corollary 3. We notice that for any $X \in \mathbb{Z}^{2\times 2}$ then $\rho_1^T X \tau = X_{1,2}$. Determining whether this is zero was proven to be NP-hard in Corollary 3.

For the Vector Reachability Problem, let $\rho_2=\tau=(0,1)^T$. If the identity matrix belongs to the semigroup then $I\rho_2=\beta$. For any $X=\begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{pmatrix} \in \mathbb{Z}^{2\times 2}$, we see that $X\rho_2=(x_{1,2},x_{2,2})^T$. As it was shown in Corollary 3 the fact that $x_{1,2}=0$ implies that $x_{2,2}=1$ for a particular set of matrices from $\mathrm{SL}(2,\mathbb{Z})$ in NP-hardness result. Therefore deciding whether τ is reachable from ρ_2 can solve the ZRUC problem which was proven NP-hard in Corollary 3 and thus vector reachability is NP-hard. □

5 Size of Solutions to the Identity Problem

The complexity of 2×2 integral matrix semigroups is quite surprising, for example we mention here the following recent result. Given the following matrices:

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & 5 \\ 0 & 5 \end{pmatrix},$$

it was mentioned as an open problem in [15] to determine whether or not the semigroup generated by $\{A, B\}$ is free (in other words whether they satisfy any non-trivial equation). Although the problem at first appears to be straightforward (since one matrix is diagonal, the other is upper-triangular), it turns out to be deceptively difficult and was not solved until recently in [17, 19] where the authors independently showed that in fact the following equation holds and thus the generated semigroup is not free:

$$AB^{10}A^2BA^2BA^{10} = B^2A^6B^2A^2BABABA^2B^2A^2BAB^2.$$

In fact, it is also reported in [17] that no shorter non-trivial equation exists.

Given a set of matrices we now consider a lower bound on the length of a solution to the Identity Problem.

We may ask the question: if the identity matrix belongs to the semigroup generated by M, say $M_{i1}M_{i2}\cdots M_{ir} = I$ with each $M_{ij} \in M$, then does there exist a lower bound on the size of r purely in terms of $|M|$? It is not difficult to see that this is not the case, for example let $D = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $C = D^{-s}$ for some large $s > 0$, then the semigroup generated by $\{D, C\}$ contains I since $D^sC = I$, but the length of this product can be set arbitrarily large by increasing s. Thus it makes sense for us to consider a bound using not only the cardinality of the generator, but also the maximal absolute value of an element of any of the matrices in M, since this is a metric taking into account the overall 'size' of the generator.

Proposition 1. *There exists a set of matrices $M = \{M_1, M_2, \ldots, M_n\} \subseteq \mathbb{Z}^{2\times 2}$ where the maximum element of any matrix in M is $O(n^4)$ such that $I \in \langle M \rangle$ and the minimal length product over M equal to I is of length $2^n - 2$, which is exponential in the number of matrices in the generator and the maximal element of any matrix in M.*

Proof. We adapt the proof of a related result over *deterministic finite automata* (DFA) recently shown in [1]. Define alphabets $\Gamma = \{1, 2, \ldots, n-1\}$, $\overline{\Gamma} = \{\overline{1}, \overline{2}, \ldots, \overline{(n-1)}\}$ and $\Sigma = \Gamma \cup \overline{\Gamma}$. It is shown in [1] that for any $n \geq 3$, there exists a DFA A_n, with $n+1$ states over Σ, such that for any word $w \in \Sigma^*$ where $w \in L(A_n)$ and $r(w) = \varepsilon$ then $|w| \geq 2^{n-1}$. Their proof is constructive and we shall now show an adaption of it which when combined with our earlier encoding will prove our claim. Let $Q = \{q_0, \ldots, q_{n-1}\}$ and q_0 be the initial state and $\{q_0\}$ the set of final states. We define the transition function $\delta : Q \times \Sigma^* \to Q$ of the DFA such that:

$$\delta(q_a, c) = \begin{cases} q_1, & \text{if } c = 1 \text{ and } a = 0; \\ q_{a+1}, & \text{if } c = \overline{a} \text{ and } 1 \leq a \leq n-2; \\ q_0, & \text{if either } c = a \text{ and } 2 \leq a \leq n-1, \\ & \text{or } c = \overline{(n-1)} \text{ and } a = n-1. \end{cases}$$

All other transitions are not defined. The structure of this DFA can be seen in Figure 2. Using the encoding idea from earlier in this paper, we may encode this problem into a set of words (the example following this proof shows the form of the set of words for the instance where $n = 4$). Let $W = \{\overline{q_a} \cdot c \cdot \delta(q_a, c)\} \subseteq (\overline{Q}\Sigma Q)^*$ be a set over every rule of the DFA transition function.

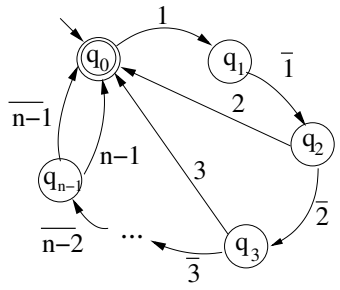

Fig. 2. A deterministic finite automaton such that the minimal non empty word w such that $r(w) = \varepsilon$ and $\delta(q_0, w) \in F$ is of length $2^n - 2$.

Let $w = w_1 w_2 \cdots w_{|w|} \in W^+$ be such that $r(w) = \varepsilon$. We may assume without loss of generality that $w_1 = \overline{q_1} \cdot 1 \cdot \delta(q_1, 1)$. This follows from the construction above, whereby to obtain the identity, each element must be used at least one time. As before, since for any $w', w'' \in W^*$ we have that $w'w_1w'' = \varepsilon$ implies $w_1w''w' = \varepsilon$, we may assume w_1 is of this form.

We may use the proof given in [1] to prove that the length of any nonempty product equal to the identity is $2^n - 2$. We see that the element w_1 corresponds to being in state q_1 and reading in symbol 1, each successive concatenation of elements allows cancellation only if that concatenation corresponds to traversing the automaton according some letter from Σ in the transition function. In this case, all *state symbols*, excluding the first and last will be cancelled and what remains will be a word $c' \in \Sigma^+$. If this word is such that $r(c') = \varepsilon$, then $r(w) = \varepsilon$ implies that the final letter of w is $\overline{q_1}$ and this corresponds to beginning in state q_1 of the DFA of [1], reading in a word whose reduction is empty and returning to the initial state. Any such product has length at least $2^n - 2$ as is shown in [1] (we are using a slight variant of their automaton so our bound is slightly different).

Using Lemma 2, we may encode the set of words W into a set W' over a binary alphabet. Note that the number of letters (excluding inverses) in the

words of W is $|\Gamma| + |Q| = 2n - 1$, thus the maximum size word encoding of morphism α of Lemma 2 is $4n - 1$. The words of W' are thus of the form:

$$c^{i_1} d_1 \overline{c}^{i_1} \cdot c^{i_2} d_2 \overline{c}^{i_2} \cdot c^{i_3} d_3 \overline{c}^{i_3} \qquad | \ 1 \leq i_1, i_2, i_3 \leq 2n - 1,$$

where $d_1, d_2, d_3 \in \{d, \overline{d}\}$. We may apply mapping α from Lemma 3 to each of these words to move from words to 2×2 integral matrices. Lemma 4 can be used to show that the largest absolute entry of any $\alpha(c^{i_j} d_j \overline{c}^{i_j})$ is $O(n^2)$ since $i_j \leq 2n - 2$. In fact we see that $\alpha(c^{i_1} d_1 \overline{c}^{i_1} \cdot c^{i_2} d_2 \overline{c}^{i_2} \cdot c^{i_3} d_3 \overline{c}^{i_3}) = O(n^4)$, since the maximal element of

$$\begin{pmatrix} 1 + 4i_1 & -8i_1^2 \\ 2 & 1 - 4i_1 \end{pmatrix} \begin{pmatrix} 1 + 4i_2 & -8i_2^2 \\ 2 & 1 - 4i_2 \end{pmatrix} \begin{pmatrix} 1 + 4i_3 & -8i_3^2 \\ 2 & 1 - 4i_3 \end{pmatrix}$$

has order $O(n^4)$ using the fact that $i_1, i_2, i_3 \leq 2n-1$ and straightforward matrix multiplication. These matrices of the form $\alpha(c^{i_1} d_1 \overline{c}^{i_1} \cdot c^{i_2} d_2 \overline{c}^{i_2} \cdot c^{i_3} d_3 \overline{c}^{i_3})$ generate M and thus the maximal element size of any matrix in M is $O(n^4)$ as required. Since the length of any (nonempty) product equaling the identity has been shown to be $2^n - 2$, we see that this length is exponential in the number of matrices in the generator and the maximal element as required. □

Example. Let us consider case $n = 4$, thus the automaton has states $Q = \{q_0, q_1, q_2, q_3\}$ and alphabet $\Sigma = \{1, 2, 3, \overline{1}, \overline{2}, \overline{3}\}$. By the encoding of Proposition 1, we obtain the following words (see Figure 2):

$$W = \left\{ \begin{matrix} \overline{q_0} 1 q_1, \ \overline{q_2} 2 q_0, \ \overline{q_3} 3 q_0, \\ \overline{q_1} \overline{1} q_2, \ \overline{q_2} \overline{2} q_3, \ \overline{q_3} \overline{3} \overline{q_0} \end{matrix} \right\} \subseteq (\overline{Q} \Sigma Q)^*.$$

Let $w = w_1 w_2 \dots w_{|w|} \in W^+$ be such that $r(w) = \varepsilon$ and $w_1 = \overline{q_0} 1 q_1$. We shall now examine the form of a product which gives the identity in this case. Let:

$$\begin{aligned} X_1 &= \overline{q_0} 1 q_1 \cdot \overline{q_1} \overline{1} q_2 \\ X_2 &= X_1 \cdot \overline{q_2} 2 q_0 \cdot X_1 \cdot \overline{q_2} \overline{2} q_3 \\ X_3 &= X_2 \cdot \overline{q_3} 3 q_0 X_2 \cdot \overline{q_3} \overline{3} \overline{q_0} \end{aligned}$$

and note that $r(X_1) = \overline{q_0} q_2$, $r(X_2) = \overline{q_0} q_3$ and thus $r(X_3) = \varepsilon$ is a solution. The solution has length $|w| = 2^n - 2 = 14$ as expected.

Finally we note that the matrix encoding used in Proposition 1 also gives us information regarding the length of solutions of the Zero in the Right Upper Corner Problem, Scalar Reachability Problem and Vector Reachability Problem over two-dimensional integer matrices. Using the encoding of Proposition 1 and the proofs of Corollary 3 and Corollary 5 immediately lead to the following result.

Corollary 6. *There exist instances of the* Zero in the Right Upper Corner Problem, Scalar Reachability Problem *and* Vector Reachability Problem *over two dimensional integer matrices where the minimum length of their solution is exponential in the cardinality of the generator set and the maximal size of any element of a matrix in the generator.*

6 Conclusion

In this paper we show that the Identity Problem for 2×2 integral unimodular matrices is NP-hard by a reduction of the Subset Sum Problem and several new encoding techniques. The exact complexity of the problem is not known. However if a detailed analysis of an upper bound for the nontrivial decidability result by C. Choffrut and J. Karhumäki will show that the algorithm is in NP then one can immediately show NP-completeness of Identity Problem. Our result about the Identity Problem reveals a number of corollaries confirming the same NP-hardness result for vector reachability, scalar reachability and zero in the right upper corner problems. Since the decidability status for these problems is not yet known, it is difficult to predict how close our results are to their real upper bounds.

In addition to that, the following problem seems particularly noteworthy and related to the results of this paper, however a direct application of the proposed techniques that are used in Theorem 1 do not appear to work.

Open Problem - Given a finite set, X, of 2-dimensional integer matrices, what is the complexity of determining if the semigroup generated by X is free? For example, is it NP-hard to determine if the matrices in X satisfy any non-trivial equation?

The complexity of these problems may possibly go beyond the NP class and the further study of complexity bounds for low-dimensional problems is an important research direction that may help to get a better understanding of computations in 2×2 matrix semigroups and one-dimensional affine transformations (affine maps).

At the end of the paper we derived an *exponential* lower bound on the minimum length solution to the Identity and other problems, the question as to whether this lower bound can be improved is another open problem that may help to clarify the computational complexity issues in matrix semigroups.

References

1. Ang, T., Pighizzini, G., Rampersad, N., Shallit, J.: Automata and reduced words in the free group, *arXiv:0910.4555*, 2009.
2. Babai, L., Beals, R., Cai, J.-Y., Ivanyos, G., Luks, E. M.: Multiplicative equations over commuting matrices, *Proc. of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA 96, 1996.
3. Beaudry, M.: Membership testing in commutative transformation semigroups, *Information and Computation*, **79**(1), October 1988, 84–93.
4. Bell, P. C., Potapov, I.: On undecidability bounds for matrix decision problems, *Theoretical Computer Science*, **391**(1-2), 2008, 3–13.
5. Bell, P. C., Potapov, I.: Periodic and infinite traces in matrix semigroups, *Current Trends in Theory and Practice of Computer Science (SOFSEM)*, **LNCS 4910**, 2008, 148–161.
6. Bell, P. C., Potapov, I.: Reachability problems in quaternion matrix and rotation semigroups, *Information and Computation*, **206**(11), 2008, 1353–1361.

7. Bell, P. C., Potapov, I.: On the undecidability of the identity correspondence problem and its applications for word and matrix semigroups, *International Journal of Foundations of Computer Science*, **21**(6), 2010, 963–978.
8. Birget, J.-C., Margolis, S.: Two-letter group codes that preserve aperiodicity of inverse finite automata, *Semigroup Forum*, **76**(1), 2008, 159–168.
9. Blondel, V., Tsitsiklis, J.: When is a pair of matrices mortal?, *Information Processing Letters*, **63**, 1997, 283–286.
10. Blondel, V. D., Cassaigne, J., Karhumäki, J.: Problem 10.3, Freeness of multiplicative matrix semigroups, *Unsolved Problems in Mathematical Systems and Control Theory* (V. D. Blondel, A. Megretski, Eds.), Princeton University Press, 2004.
11. Blondel, V. D., Portier, N.: The presence of a zero in an integer linear recurrent sequence is NP-hard to decide, *Linear Algebra and its Applications*, 2002, 91–98.
12. Bournez, O., Branicky, M.: On the mortality problem for matrices of low dimensions, *Theory of Computing Systems*, **35**(4), 2002, 433–448.
13. Cai, J.-Y., Fuchs, W. H., Kozen, D., Liu, Z.: Efficient average-case algorithms for the modular group, *The 35th Annual Symposium on Foundations of Computer Science (FOCS)*, 1994.
14. Cai, J.-Y., Liu, Z.: The bounded membership problem of the monoid $\mathrm{SL}_2(N)$, *Mathematical Systems Theory*, **29**(6), 1996, 573–587.
15. Cassaigne, J., Harju, T., Karhumäki, J.: On the undecidability of freeness of matrix semigroups, *International Journal of Algebra and Computation*, **9**(3-4), 1999, 295–305.
16. Cassaigne, J., Karhumäki, J.: Examples of undecidable problems for 2-generator matrix semigroups, *Theoretical Computer Science*, **204(1-2)**, 1998, 29–34.
17. Cassaigne, J., Nicolas, F.: On the decidability of semigroup freeness, *arXiv:0808.3112v2 [cs.DM]*, September 2008.
18. Choffrut, C., Karhumäki, J.: Some decision problems on integer matrices, *Informatics and Applications*, **39**, 2005, 125–131.
19. Gawrychowski, P., Gutan, M., Kisielewicz, A.: On the problem of freeness of multiplicative matrix semigroups, *Theoretical Computer Science*, **411**(7-9), February 2010, 1115–1120.
20. Gurevich, Y., Schupp, P.: Membership problem for the modular group, *SIAM J. Comput.*, **37**(2), 2007, 425–459.
21. Halava, V., Harju, T., Hirvensalo, M.: Undecidability bounds for integer matrices using Claus instances, *International Journal of Foundations of Computer Science (IJFCS)*, **18,5**, 2007, 931–948.
22. Halava, V., Harju, T., Hirvensalo, M., Karhumäki, J.: Skolem's problem - on the border between decidability and undecidability, *TUCS Technical Report Number 683*, 2005.
23. Halava, V., Hirvensalo, M.: Improved matrix pair undecidability results, *Acta Informatica*, **44(3-4)**, 2007, 191–205.
24. Harju, T.: Post correspondence problem and small dimensional matrices, *Lecture Notes in Computer Science*, **LNCS 5583**, 2009, 39–46.
25. Lyndon, R. C., Schupp, P. E.: *Combinatorial Group Theory*, Springer-Verlag, 1977.
26. Mihailova, A.: The occurrence problem for a direct product of groups, *Doklady Akademii Nauk*, **119**(in Russian), 1958, 1103–1105.