On the Identity Problem for $SL_2(\mathbb{Z})$

Dr Paul C. Bell

Department of Computer Science
Liverpool John Moores University
p.c.bell.@ljmu.ac.uk

Co-authors for today's topics:
V. Halava, T. Harju, M. Hirvensalo, J. Karhumäki (Turku University, Finland)
I. Potapov (University of Liverpool)
V. Blondel, J.-C. Delvenne, R. Jungers (Université catholique de Louvain)
J. O. Shallit (University of Waterloo, Canada)
S. Chen, L. M. Jackson (Loughborough University)

Outline of the talk

- Introduction and notation
- Mortality problem
- Identity problem over $\mathbb{Z}^{4\times 4}$ - Undecidability
- Identity problem over $SL_2(\mathbb{Z})$ and $GL_2(\mathbb{Z})$ - NP-completeness
- Conclusion

Notations

- We denote an n-dimensional matrix over a semiring $\mathbb{F}$ by $\mathbb{F}^{n\times n}$
- Given a set of matrices $G = \{M_1, M_2, \ldots, M_k\} \subseteq \mathbb{K}^{n\times n}$ (where $\mathbb{K} \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{H}\}$), we denote by $S = \langle G \rangle$ the semigroup generated by G

Decision Problems for Matrix Semigroups

- Given a matrix semigroup S generated by a finite set $G = \{M_1, M_2, \ldots, M_k\} \subseteq \mathbb{K}^{n \times n}$ (where $\mathbb{K} \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{H}\}$):
  - Decide whether the semigroup S
    - contains the zero matrix (MORTALITY PROBLEM)
    - contains the identity matrix (IDENTITY PROBLEM)
    - is free (FREENESS PROBLEM)
    - is bounded, finite, etc.
  - Problem has links to other areas of Computer Science and Mathematics
    - Vector and scalar reachability problems
    - Probabilistic automata, Weighted automata and quantum finite automata
    - Dynamical systems, group theory

Early Reachability Results

- The MORTALITY PROBLEM was one of the earliest undecidability results of reachability for matrix semigroups

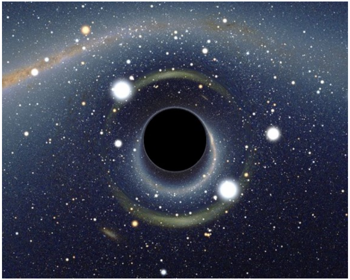
Theorem ([Paterson 70])

The MORTALITY PROBLEM *is undecidable over* $\mathbb{Z}^{3\times 3}$.

Theorem (B., Halava, Harju, Karhumäki, Potapov, 2012 (IJAC))

The MORTALITY PROBLEM *is undecidable for bounded languages:*

$$M_1^{k_1} M_2^{k_2} \cdots M_t^{k_t} = \mathcal{Z}$$



Post's Correspondence Problem (PCP)

Problem (Post's Correspondence Problem (PCP))

Given alphabet Σ, binary alphabet Δ, and morphisms $h, g : \Sigma^* \to \Delta^*$, does there exist $w = x_1 \dots x_k \in \Sigma^+; x_i \in \Sigma$ s.t.

$$h(x_1)h(x_2)\dots h(x_k) = g(x_1)g(x_2)\dots g(x_k)?$$

Theorem (Matiyasevich, Sénizergues, 96)

PCP(7) is undecidable.

Theorem (Neary 15)

PCP(5) is undecidable.

From words to integers

- Let $\sigma(a) = 1, \sigma(b) = 2$ and $\sigma(uv) = 3^{|v|}\sigma(u) + \sigma(v)$ for every $u, v \in \Sigma^*$. Then σ is a monomorphism $\Sigma^* \to \mathbb{N}$.
- We may then define a mapping $\tau : \Sigma^* \times \Sigma^* \mapsto \mathbb{Z}^{3\times 3}$

$$\tau(u, v) = \begin{pmatrix} 1 & \sigma(v) & \sigma(u) - \sigma(v) \\ 0 & 3^{|v|} & 3^{|u|} - 3^{|v|} \\ 0 & 0 & 3^{|u|} \end{pmatrix}$$

- We can prove that $\tau(u_1, v_1) \cdot \tau(u_2, v_2) = \tau(u_1 u_2, v_1 v_2)$ for all $u_1, u_2, v_1, v_2 \in \Sigma^*$, thus τ is a monomorphism.
- Note that $\tau(u, v)_{1,3} = 0$ if and only if $u = v$.
- This technique can be used to show the undecidability of the Mortality Problem via a reduction of PCP.

Semigroup Freeness

Definition (Code)

Let $\mathcal{S}$ be a semigroup and $\mathcal{G}$ a subset of $\mathcal{S}$. We call $\mathcal{G}$ a code if the property

$$u_1 u_2 \cdots u_m = v_1 v_2 \cdots v_n$$

for $u_i, v_i \in \mathcal{G}$, implies that $m = n$ and $u_i = v_i$ for each $1 \leq i \leq n$.

Definition (Semigroup freeness)

A semigroup $\mathcal{S}$ is called free if there exists a code $\mathcal{G} \subseteq \mathcal{S}$ such that $\mathcal{S} = \mathcal{G}^+$.

- For example, consider the semigroup $\{0, 1\}^+$ under concatenation. Then the set $\{00, 01, 10, 11\}$ is a code, but $\{01, 10, 0\}$ is not (since $0 \cdot 10 = 01 \cdot 0$ for example).

Matrix Freeness

Problem (Matrix semigroup freeness)

SEMIGROUP FREENESS PROBLEM - *Given a finite set of matrices $\mathcal{G} \subseteq \mathbb{Z}^{n\times n}$ generating a semigroup $\mathcal{S}$, does every element $M \in \mathcal{S}$ have a single, unique factorisation over $\mathcal{G}$? Alternatively, is $\mathcal{G}$ a code?*

Theorem (Klarner, Birget and Satterfield, 91)

The semigroup freeness problem is undecidable over $\mathbb{N}^{3\times 3}$

- *Undecidability holds even over $\mathbb{N}^{3\times 3}_{uptr}$ [Cassaigne, Harju and Karhumäki, 99]*

Matrix Freeness in Dimension 2

- Let $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 5 \\ 0 & 5 \end{pmatrix}$, is $\{A, B\}$ a code?
- Two groups of authors independently showed that in fact the following equation holds and thus the generated semigroup is not free[Gawrychowskia et al. 2010], [Cassaigne et al. 2012]:

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

and no shorter non-trivial equation exists.

- **Open Problem** - Determine the decidability of the Freeness Problem over $\mathbb{N}^{2\times 2}$ (even for two matrices, or when all matrices are upper triangular).

The Identity Problem

Problem (The Identity Problem)

Given a matrix semigroup S generated by a finite set $G = \{M_1, M_2, \ldots, M_k\} \subseteq \mathbb{Z}^{n \times n}$, determine if $I_n \in \langle G \rangle$, where I_n is the n-dimensional multiplicative identity matrix.

Known results

- For commuting matrices the Membership and Vector Reachability problems are decidable in PTIME for matrices of all dimensions (over algebraic numbers). [Babai, Beals, Cai, Ivanyos, Luks, 1996]
- Identity problem, Mortality problem, Freeness, Vector Reachability in $SL_2(\mathbb{Z})$ are NP-Hard [B., Hirvensalo, Ko, Potapov, 2012-2016]

The Identity Problem

Theorem (Choffrut, Karhumäki 05)

The IDENTITY PROBLEM *is decidable over* $\mathbb{Z}^{2\times 2}$

Theorem (B., Potapov, 2011 (IJFCS))

The IDENTITY PROBLEM *is undecidable over* $\mathbb{Z}^{4\times 4}$.

Theorem (B., Hirvensalo, Potapov, (SODA'17))

The IDENTITY PROBLEM *is NP-complete over* $\mathbb{Z}^{2\times 2}$.

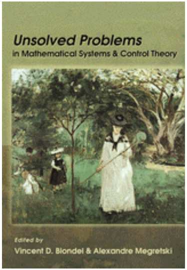


Figure: Unsolved Problems in Mathematical Systems and Control Theory, 309-314. Princeton University Press, Princeton (2004)

Decidability of membership in $\mathrm{SL}_2(\mathbb{Z})$

Special Linear group $\mathrm{SL}_2(\mathbb{Z})$ - 2×2 integer matrices with determinant 1.

Theorem (C. Choffrut and J. Karhumäki, 2005)

Let $M \in \mathrm{SL}_2(\mathbb{Z})$ *and let* F *be a finite collection of matrices from* $\mathrm{SL}_2(\mathbb{Z})$. *Then it is decidable whether* $M \in \langle F \rangle$.

Decidability of membership in $SL_2(\mathbb{Z})$

- $SL_2(\mathbb{Z})$ is generated by $\langle S, T \rangle$, where

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

- Representations of elements of $SL_2(\mathbb{Z})$ using S, T are not unique, for example, $TST = ST^{-1}S^3$
- For a more canonical representation, let

$$R = ST = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}.$$

R has order 6 (thus $R^6 = I$) and S has order 4 (thus $S^4 = I$).

Decidability of membership in $SL_2(\mathbb{Z})$

- Now, $SL_2(\mathbb{Z}) = \langle S, R \rangle$ and the representation is unique
- Each element of $SL_2(\mathbb{Z})$ can be represented as:

$$A = (-1)^{\gamma} R^{n_0} S R^{n_1} S \cdot \ldots \cdot R^{n_{l-1}} S R^{n_l}, \tag{1}$$

where $\gamma \in \{0, 1\}$, $n_i \in \{0, 1, 2\}$ and $n_i \in \{1, 2\}$ for $0 < i < l$.

- Representations of matrices from $SL_2(\mathbb{Z})$ can be exponentially long:

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} = T^m = (-SR)^m = (-1)^m \underbrace{SR \ldots SR}_{m \text{ times}} \tag{2}$$

From Matrices to Words

- The *Projective Special Linear group* is the quotient group

$$\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$$

- Let $s = S\{\pm I\}$ and $r = R\{\pm I\}$ be the projections of S and R in $\mathrm{PSL}_2(\mathbb{Z})$.
- Since $S^2 = R^3 = -I$ in $\mathrm{SL}_2(\mathbb{Z})$ then $s^2 = r^3 = \{\pm I\}$ in $\mathrm{PSL}_2(\mathbb{Z})$.
- Intuitively, $\mathrm{PSL}_2(\mathbb{Z})$ can be taken as $\mathrm{SL}_2(\mathbb{Z})$ by ignoring the sign.

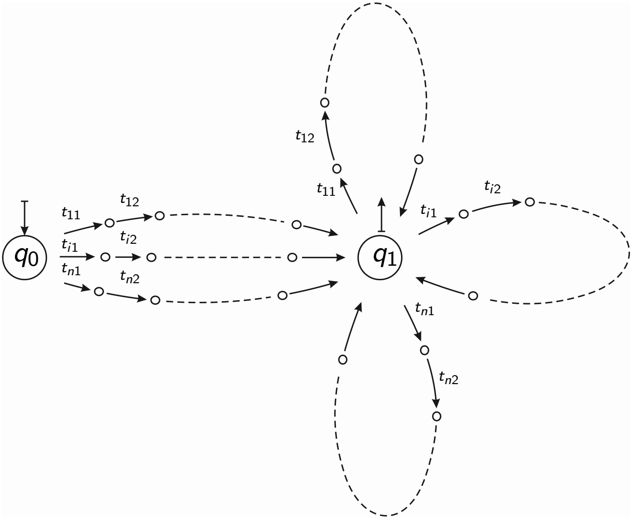
Recognizing the Identity in EXPSPACE

The procedure of Choffrut and Karhumäki:

1. First, a nondeterministic finite automaton over alphabet $\{r, s\}$ recognizing A^+ is constructed;
2. Then ε-transitions are iteratively added to represent the relations $r^3 = s^2 = \varepsilon$ between the nodes (states) as long as possible.

- The procedure ends eventually, since the number of states is finite, although exponential in the description size of A
- The decision whether $\varepsilon \in A^+$ is then made based on the observation whether there is an ε-transition from the initial state to the final state

The 'Petal Automaton'



Difficult cases of the Identity problem

- Problems on words can be encoded into reachability problems over $\mathrm{PSL}_2(\mathbb{Z})$
- Let $\Sigma_t = \{a_1, a_2, \ldots, a_t\}$ be an arbitrary sized group alphabet and $\Sigma_2 = \{a, b\}$, then there exists an injective homomorphism $\alpha : \Sigma_t^* \to \Sigma_2^*$, e.g.,

$$\alpha(a_t) = b^t a b^{-t} \quad \alpha(a_t^{-1}) = b^t a^{-1} b^{-t}$$

Difficult cases of the Identity problem

- Furthermore, there exists an injective homomorphism $f : (\Sigma_2 \cup \overline{\Sigma}_2)^* \to \mathrm{PSL}_2(\mathbb{Z})$ given by:

$$f(a) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, f(b) = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, f(a^{-1}) = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}, f(b^{-1}) = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$$

Exponential Length Solutions

The length of a minimal size identity can be exponential in the description size of the matrix generator [B., Potapov, 2012].

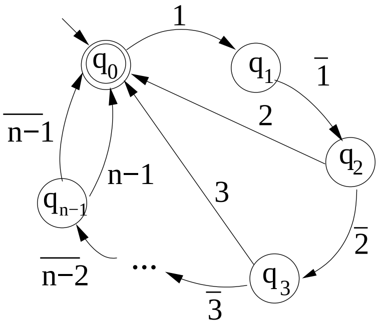


Figure: An automaton from [Ang et al., 2009].

First difficult case

- Let $Q_4 = \{q_i, q_i^{-1} : 1 \leq i \leq 4\}$, $\Sigma_4 = \{i, i^{-1} : 1 \leq i \leq 4\}$ and

$$W = \left\{ \begin{array}{cccc} q_0^{-1}1q_1, & q_2^{-1}2q_0, & q_3^{-1}3q_0, & q_4^{-1}4q_0, \\ q_1^{-1}1^{-1}q_2, & q_2^{-1}2^{-1}q_3, & q_3^{-1}3^{-1}q_4, & q_4^{-1}4^{-1}q_0 \end{array} \right\}$$

- It can be shown that the shortest $\varepsilon \in W^*$ has form:

$$\begin{array}{rclcl} X_1 & = & q_0^{-1}1q_1 \cdot q_1^{-1}1^{-1}q_2 & \equiv & q_0^{-1}q_2 \\ X_2 & = & X_1 \cdot q_2^{-1}2q_0 \cdot X_1 \cdot q_2^{-1}2^{-1}q_3 & \equiv & q_0^{-1}q_3 \\ X_3 & = & X_2 \cdot q_3^{-1}3q_0 \cdot X_2 \cdot q_3^{-1}3^{-1}q_4 & \equiv & q_0^{-1}q_4 \\ X_4 & = & X_3 \cdot q_4^{-1}4q_0 \cdot X_3 \cdot q_4^{-1}4^{-1}q_0 & \equiv & \varepsilon \end{array}$$

- W can be trivially generalised so that it consists of $2k$ elements and the shortest ε uses $2^{k+1} - 2$ elements of W.

Second difficult case

- Consider the subset sum problem: let $S = \{s_1, s_2, \ldots, s_{k-1}\} \subseteq \mathbb{N}$ and $t \in \mathbb{N}$, does there exist some subset $S' \subseteq S$ such that $\sum_{x \in S'} x = t$?
- The problem is well known to be NP-complete

Second difficult case

Using border symbols $\Sigma_k = \{1, 2, \ldots, k, 1^{-1}, 2^{-1}, \ldots, k^{-1}\}$, we may define the following set of words:

$$W' = \left\{ \begin{array}{lllll} 1W_1 2^{-1}, & 2W_2 3^{-1}, & \cdots, & (k-1)W_{k-1}k^{-1}, & kW_t^{-1}1^{-1}, \\ 1 \cdot \varepsilon \cdot 2^{-1}, & 2 \cdot \varepsilon \cdot 3^{-1}, & \cdots, & (k-1) \cdot \varepsilon \cdot k^{-1} & \end{array} \right\}$$

where $W_i = a^{s_i}$ and $W_t^{-1} = a^{-t}$.

Second difficult case

- If $\varepsilon \in W'^{+}$, then it is of the form:

$$\begin{aligned} & 1X_1 2^{-1} \cdot 2X_2 3^{-1} \cdots (k-1)X_{k-1} k^{-1} \cdot kW_t^{-1} 1^{-1}, \\ = \quad & 1X_1 X_2 \cdots X_{k-1} \cdot W_t^{-1} 1^{-1}, \end{aligned}$$

  where $X_i \in \{W_i, \varepsilon\}$
- Equivalent to the subset sum problem
- Monomorphism $f \circ \alpha$ can map this problem to $\mathrm{PSL}_2(\mathbb{Z})$
- Exponentially *many* possible solutions to check

The Structure of an Identity

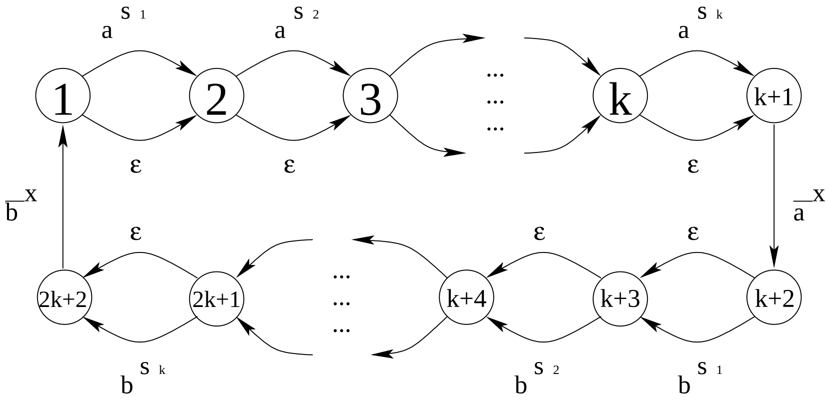


Figure: The structure of a product which forms the identity.

Main results: from EXPSPACE to NP

Theorem

The identity problem over $\mathrm{GL}_2(\mathbb{Z})$ is NP-complete.

Theorem

The problem of determining whether a matrix M is in an arbitrary regular expression $R(a_1, \ldots, a_n) \subseteq \mathrm{GL}_2(\mathbb{Z})$ is in NP.

Theorem

The non-freeness problem for finitely generated semigroups in $\mathrm{GL}_2(\mathbb{Z})$ is NP-complete.

NP solution

Our strategy avoids exponential growth in the graph:

- Following [Gurevich, Schupp], we consider *syllables*, which are a compressed form of word (described next)
- We form a compressed graph and a series of rules to work on those graphs
- The graph size is carefully kept polynomial, and nondeterministically updates edge labels

Words under $PSL_2(\mathbb{Z})$

- Consider the following 'syllables':

$$R_i = \begin{cases} (rs)^{i-1}r & \text{if } i > 0 \\ (r^2s)^{|i|-1}r^2 & \text{if } i < 0 \\ \varepsilon & \text{if } i = 0 \end{cases}$$

  We say that syllable R_i is positive, if $i > 0$, and negative, if $i < 0$.

- An example:

$$\begin{aligned} R_2R_{-5} &= (rs)r(r^2s)^4r^2 = (rs)rr^2s(r^2s)^3r^2 \\ &= r(r^2s)^3r^2 = r(r^2s)(r^2s)^2r^2 = s(r^2s)^2r^2 \end{aligned}$$

Words under $\mathrm{PSL}_2(\mathbb{Z})$

Lemma

Each element $a \in \mathrm{PSL}_2(\mathbb{Z})$ admits a unique representation of the form

$$a = s^{\alpha} R_{n_1} s R_{n_2} s R_{n_3} s \dots s R_{n_l} s^{\beta}, \tag{3}$$

with $\alpha, \beta \in \{0,1\}$ and the representation is alternating. The representation size is polynomial in the representation size of a.

Words under $\mathrm{PSL}_2(\mathbb{Z})$

Lemma

The syllables satisfy the following relations

1. $ss \mapsto \varepsilon$
2. $R_a R_{-a} \mapsto \varepsilon$
3. $R_a R_{-b} \mapsto R_{a-b}s$, *if* $ab > 0$ *and* $abs(b) < abs(a)$
4. $R_a R_{-b} \mapsto sR_{a-b}$, *if* $ab > 0$ *and* $abs(a) < abs(b)$
5. $R_{-1} R_{-1} \mapsto R_1$
6. $R_1 \mapsto R_{-1} R_{-1}$

Pathological cases

The syllables also satisfy pathological relations, for example

$$\begin{aligned} R_1 R_2^t R_1 &\equiv R_{-1} R_{-1} R_2^t R_1 \\ &\equiv R_{-1} s R_1 R_2^{t-1} R_1 \equiv \ldots \\ &\equiv (R_{-1}s)(R_{-1}s)\cdots(R_{-1}s) R_1 R_1 \\ &\equiv (R_{-1}s)^t R_{-1} \equiv R_{-(t+1)} \end{aligned}$$

Syllabic weight

For each syllable in Σ, we now introduce a notion of "weight", which gives a magnitude to each such element.

$$\mathrm{wgt}(z) = \begin{cases} x, \text{ if } z = R_x \text{ and } z \in \Gamma; \\ \pm 2, \text{ if } z \in \{s^\alpha R_{\pm 2} s^\beta \mid \alpha, \beta \in \{0,1\}\}; \\ \pm 1, \text{ if } z \in \{s^\alpha R_{\pm 1} s^\beta \mid \alpha, \beta \in \{0,1\}\}; \\ 0 \text{ if } z \in \{\varepsilon, s\}. \end{cases}$$

Canonical syllabic representation of $\mathrm{PSL}_2(\mathbb{Z})$ elements

Definition

We define the set of syllables $\Omega = \{\varepsilon, s, s^{\alpha}R_{\pm 1}s^{\beta}, s^{\alpha}R_{\pm 2}s^{\beta}\}$, where $\alpha, \beta \in \{0,1\}$. Intuitively, set Ω forms a "neighbourhood" of ε.

Definition (Ω-**Minimal Word**)

A syllabic word $w = w_1 w_2 \cdots w_k \in \Sigma^*$ is called an Ω-*minimal word* if it does not contains syllabic subword that is reducible to any element from Ω.

For example, $R_{10}R_{-5}sR_{-5}$ is Ω-Minimal Word, since $R_{10}R_{-5}sR_{-5} \equiv R_5 ssR_{-5} \equiv R_5 R_{-5} \equiv \varepsilon$, but no shorter syllabic subword of $R_{10}R_{-5}sR_{-5}$ has that property.

NP solution

Our technique avoids exponential growth in the edge set

- Given a matrix set $M = \{M_1, \ldots, M_n\} \subseteq \mathrm{SL}_2(\mathbb{Z})$, the procedure starts with constructing a polynomial size syllabic version of the "daisy graph" $G_M = (Q, E)$
- For nondeterministically chosen vertex pair $q_i, q_j \in Q$, check if there is a path $q_i \to q_j$ with label equivalent to an Ω-minimal word, i.e. one "close" to ε. This may be done via *short*, *medium*, or *long reductions*
- Verify if there is an ε-edge from the initial state q_0 to the final state q_1. The witness for such an edge gives the positive answer to the identity problem.

Short, Medium and Long reductions

We now describe three ways of showing that there is indeed such a path $q_i \to q_j$.

1. Short Reductions. Deal with simple/pathological cases directly.
2. Medium Reductions. Let $|w| > 3$, such that Π *contains no dual edge cycles*, i.e. no pair of edges of the graph is used more than once (excluding ε-edges). Dealt with directly.
3. Long Reductions. Let $|w| > 3$ such that Π contains at least one dual edge cycle, then we call Π a long reduction from q_i to q_j. More complicated to deal with.

Main results: from EXPSPACE to NP

Theorem

The identity problem over $\mathrm{GL}_2(\mathbb{Z})$ is NP-complete.

Theorem

The problem of determining whether a matrix M is in an arbitrary regular expression $R(a_1, \ldots, a_n) \subseteq \mathrm{GL}_2(\mathbb{Z})$ is in NP.

Theorem

The non-freeness problem for finitely generated semigroups in $\mathrm{GL}_2(\mathbb{Z})$ is NP-complete.

Conclusion

- The identity problem in $GL_2(\mathbb{Z})$
- Two new notions of freeness problems for matrix semigroups
- We studied the problems on arbitrary semigroups and bounded languages