On the undecidability of the identity correspondence problem and its applications for word and matrix semigroups

Citation: BELL, P. and POTAPOV, I., 2010. 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), pp.963-978.

On the Undecidability of the Identity Correspondence Problem and its Applications for Word and Matrix Semigroups

Paul C. Bell, Igor Potapov

Department of Computer Science, The University of Liverpool,
Email: p.bell@lboro.ac.uk (P. Bell), potapov@liverpool.ac.uk (I. Potapov)

Abstract. In this paper we study several closely related fundamental problems for words and matrices. First, we introduce the Identity Correspondence Problem (ICP): whether a finite set of pairs of words (over a group alphabet) can generate an identity pair by a sequence of concatenations. We prove that ICP is undecidable by a reduction of Post's Correspondence Problem via several new encoding techniques. In the second part of the paper we use ICP to answer a long standing open problem concerning matrix semigroups: "Is it decidable for a finitely generated semigroup S of integral square matrices whether or not the identity matrix belongs to S?". We show that the problem is undecidable starting from dimension four even when the number of matrices in the generator is 48. From this fact, we can immediately derive that the fundamental problem of whether a finite set of matrices generates a group is also undecidable. We also answer several questions for matrices over different number fields. Apart from the application to matrix problems, we believe that the Identity Correspondence Problem will also be useful in identifying new areas of undecidable problems in abstract algebra, computational questions in logic and combinatorics on words.

Keywords: Combinatorics on Words, Group problem, Post's Correspondence Problem, Matrix Semigroups, Undecidability.

1 Introduction

Combinatorics on words has strong connections to several areas of mathematics and computing. It is well known that words are very suitable objects to formulate fundamental properties of computations. One such property that may be formulated in terms of operations on words is the exceptional concept of undecidability. A problem is called undecidable if there exists no algorithm that can solve it. A famous example is Post's Correspondence Problem (PCP) originally proved undecidable by Emil Post in 1946 [21]. It plays a central role in computer science due to its applicability for showing the undecidability of many computational problems in a very natural and simple way.

There are surprisingly many easily defined problems whose decidability status is still open. In some cases we believe that an algorithm solving the problem may

exist, but finding it would require the solution to fundamental open problems in mathematics. For other problems, the current tools for showing undecidability are not directly applicable and new techniques need to be invented to explore the border between decidable and undecidable problems.

In this paper, we introduce the Identity Correspondence Problem (ICP) in the spirit of Post's Correspondence Problem : whether a finite set of pairs of words (over a group alphabet) can generate an identity pair by a sequence of concatenations. We prove that ICP is undecidable by a reduction of Post's Correspondence Problem via several new encoding techniques that are used to guarantee the existence of an identity pair only in the case of a correct solution existing for the PCP instance. It is our belief that the Identity Correspondence Problem may be useful in identifying new areas of undecidable problems related to computational questions in abstract algebra, logic and combinatorics on words.

In the second part of the paper, we use the Identity Correspondence Problem to answer several long standing open problems concerning matrix semigroups [6]. Taking products of matrices is one of the fundamental operations in mathematics. However, many computational problems related to the analysis of matrix products are algorithmically hard and even undecidable. Among the oldest results is a remarkable paper by M. Paterson, where he shows that it is undecidable whether the multiplicative semigroup generated by a finite set of 3×3 integer matrices contains the zero matrix (also known as the mortality problem), see [20]. Since then, many results were obtained about checking the freeness, boundedness and finiteness of matrix semigroups and the decidability of different reachability questions such as the membership problem, vector reachability, scalar reachability etc. See [2–5, 8–10, 13] for several related decidability results.

The membership problem asks whether a particular matrix is contained within a given semigroup. The membership problem is undecidable for 3×3 integral matrix semigroups due to Paterson's results and also for the special linear group SL(4, $\mathbb{Z}$) of 4×4 integer matrices of determinant 1, shown by Mikhailova [18].

Another important problem in matrix semigroups is the Identity Problem: Decide whether a finitely generated integral matrix semigroup contains the identity matrix. The Identity Problem is equivalent to the following Group Problem: given a finitely generated semigroup S, decide whether a subset of the generator of S generates a non-trivial group. In general, it is undecidable whether or not the monoid described by a given finite representation is a group. However, this decision problem is reducible to a very restricted form of the uniform word problem and it does not immediately imply that the Group Problem in finitely generated semigroups (without a set of relations) is undecidable [19].

The question about the membership of the identity matrix for matrix semigroups is a well known open problem and was recently stated in "Unsolved Problems in Mathematical Systems and Control Theory", [6] and also as Problem 5 in [14]. The embedding methods used to show undecidability in other results do not appear to work here [6]. As far as we know, only two decidability

results are known for the Identity Problem. Very recently the first general decidability result for this problem was proved in the case of 2×2 integral matrix semigroups, see [10]. It is also known that in the special case of commutative matrix semigroups, the problem is decidable in any dimension [1].

In this paper we apply ICP to answer the long standing open problem: "Is it decidable for a finitely generated semigroup S of square integral matrices whether or not the identity matrix belongs to S?". We show that the Identity Problem is undecidable starting from dimension four even when the number of matrices in the generator is fixed. In other words, we can define a class of finite sets $\{M_1, M_2, \ldots, M_k\}$ of four dimensional matrices such that there is no algorithm to determine whether or not the identity matrix can be represented as a product of these matrices. From this fact, we can immediately derive that the fundamental problem of whether a finite set of 4×4 matrices generates a group is also undecidable. In our proofs we use the fact that free groups can be embedded into the multiplicative group of 2×2 integral matrices. This allows us to transfer the undecidability of ICP into undecidability results on matrices.

We also provide a number of other corollaries. In particular, the Identity and Group problems are undecidable for double quaternions and a set of rotations on the 3-sphere. Therefore, there is no algorithm to check whether a set of linear transformations or a set of rotations in dimension 4 is reversible. Also, the question of whether any diagonal matrix can be generated by a 4×4 integral matrix semigroup is undecidable.

2 Identity Correspondence Problem

Notation: Given an alphabet $\Sigma = \{a, b\}$, we denote the concatenation of two letters $x, y \in \Sigma$ by xy or $x \cdot y$. A *word* over Σ is a concatenation of letters from alphabet Σ, i.e., $w = w_1 w_2 \cdots w_k \in \Sigma^*$. We denote throughout the paper the *empty word* (or identify element) by ε. We shall denote a *pair of words* by either (w_1, w_2) or $\frac{w_1}{w_2}$.

The free group over a generating set H is denoted by $\mathrm{FG}(H)$, i.e., the free group over two elements a and b is denoted as $\mathrm{FG}(\{a, b\})$. For example, the elements of $\mathrm{FG}(\{a, b\})$ are all the words over the alphabet $\{a, b, a^{-1}, b^{-1}\}$ that are reduced, i.e., that contain no subword of the form $x \cdot x^{-1}$ or $x^{-1} \cdot x$ (for $x \in \{a, b\}$). Note that $x \cdot x^{-1} = x^{-1} \cdot x = \varepsilon$.

Problem 1 *Identity Correspondence Problem (ICP) - Let $\Sigma = \{a, b\}$ be a binary alphabet and*

$$\Pi = \{(s_1, t_1), (s_2, t_2), \ldots, (s_m, t_m)\} \subseteq \mathrm{FG}(\Sigma) \times \mathrm{FG}(\Sigma).$$

Determine if there exists a nonempty finite sequence of indices $l_1, l_2, \ldots, l_k$ where $1 \leq l_i \leq m$ such that

$$s_{l_1} s_{l_2} \cdots s_{l_k} = t_{l_1} t_{l_2} \cdots t_{l_k} = \varepsilon,$$

where ε is the empty word (identity).

A first step towards the proof of undecidability of Problem 1 was shown in [2] where the following theorem was presented (although in a different form).

Theorem 1. *[2] - Index Coding PCP - Let* $\Sigma = \{a, b\}$ *be a binary alphabet and*

$$X = \{(s_1, t_1), (s_2, t_2), \ldots, (s_f, t_f)\} \subseteq \mathrm{FG}(\Sigma) \times \mathrm{FG}(\Sigma).$$

It is undecidable to determine if there exists a finite sequence $l_1, l_2, \ldots, l_k$ *where* $1 \leq l_i \leq f$ *and* exactly one $l_i = f$ *such that*

$$s_{l_1} s_{l_2} \cdots s_{l_k} = t_{l_1} t_{l_2} \cdots t_{l_k} = \varepsilon.$$

Unfortunately, Theorem 1 cannot be directly used to prove the Identity Problem or the Group Problem are undecidable. We may, however, immediately use Problem 1 for this purpose (and do so in Section 3) once we have proved that it is undecidable.

The reason Theorem 1 does not prove Problem 1 is undecidable is the restriction that the final pair of words (s_f, t_f) is used exactly one time. Despite many attempts, it is not clear how one may remove this restriction in the construction of the proof, since it is essential that this pair be used once to avoid the pathological case of several incorrect solutions cancelling with each other and producing an identity element.

The main idea of this paper is to show a new non-trivial encoding which contains the encoding used in Theorem 1 but avoids the requirement that a specific element be used one time. The idea is that by encoding the set X *four times* using four different alphabets and adding 'borders' to each pair of words such that for cancellation to occur, each of these alphabets must be used in a specific (cyclic) order, any incorrect solutions using a single alphabet will not be able to be cancelled later on. More details of this encoding with four alphabets will be given later, in Lemmas 4, 5 and 6 and the example that follows them, which provides some intuition as to why three alphabets is not sufficient in the encoding.

We shall reduce a restricted form of Post's Correspondence Problem (PCP) [13] to the Identity Correspondence Problem in a constructive way. We shall require the following theorem:

Theorem 2. *[13, 17] Restricted PCP - Let* $\Sigma = \{a, b\}$ *be a binary alphabet and*

$$P = \{(u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)\} \subseteq \Sigma^* \times \Sigma^*$$

be a set of pairs of words where $n \geq 3$*. It is undecidable to determine if there exists a finite sequence of indices* $l_1, l_2, \ldots, l_k$ *with each* $2 \leq l_i \leq n-1$ *such that:*

$$u_1 u_{l_1} u_{l_2} \cdots u_{l_k} u_n = v_1 v_{l_1} v_{l_2} \cdots v_{l_k} v_n.$$

This result holds even for $n = 7$*.*

We now show the reduction of an instance of the Restricted Post's Correspondence Problem of Theorem 2 to an instance of the Identity Correspondence Problem. Let here and throughout $\Sigma = \{a, b\}$ and define new alphabets $\Gamma_i = \{a_i, b_i\}$ for $1 \leq i \leq 4$ and $\Gamma_B = \{x_j | 1 \leq j \leq 8\}$ such that the alphabets are distinct (specifically, the intersection of the free groups generated by any two different alphabets equals $\{\varepsilon\}$). Let us define mappings $\delta_i : \mathrm{FG}(\Sigma) \to \mathrm{FG}(\Gamma_i)$ by $\delta_i(a) = a_i$, $\delta_i(b) = b_i$, $\delta_i(a^{-1}) = a_i^{-1}$ and $\delta_i(b^{-1}) = b_i^{-1}$ for $1 \leq i \leq 4$. Note that each δ_i is a homomorphism that may be applied to words over $\mathrm{FG}(\Sigma)$ in a natural way.

Let $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4 \cup \Gamma_B$. Define $\phi_i : \mathbb{Z}^+ \to \{a_i, b_i\}^*$ by $\phi_i(j) = a_i^j b_i$. Similarly, let $\psi_i : \mathbb{Z}^+ \to \{a_i^{-1}, b_i^{-1}\}^*$ be defined by $\psi_i(j) = (a_i^{-1})^j b_i^{-1}$. These morphisms will be used to ensure a product is in a specific order. As an example of these morphisms we see that $\phi_2(3) = a_2 a_2 a_2 b_2$ and $\psi_3(2) = a_3^{-1} a_3^{-1} b_3^{-1}$.

Let $P = \{(u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)\} \subseteq \Sigma^* \times \Sigma^*$ be a given Restricted PCP instance. We shall define an instance of ICP consisting of a set of $8(n-1)$ pairs of words:

$$W = W_0 \cup W_1 \cup \ldots \cup W_{15} \subseteq \mathrm{FG}(\Gamma) \times \mathrm{FG}(\Gamma)$$

$$W_0 = \left\{ \frac{x_8}{x_8} \cdot \frac{v_{11}^{-1} u_{11}}{b_1} \cdot \frac{x_1^{-1}}{x_1^{-1}} \right\}, \quad W_1 = \left\{ \frac{x_1}{x_1} \cdot \frac{u_{1j}}{\phi_1(j)} \cdot \frac{x_1^{-1}}{x_1^{-1}} | 2 \leq j \leq n-1 \right\},$$
$$W_2 = \left\{ \frac{x_1}{x_1} \cdot \frac{u_{1n} v_{1n}^{-1}}{b_1^{-1}} \cdot \frac{x_2^{-1}}{x_2^{-1}} \right\}, \quad W_3 = \left\{ \frac{x_2}{x_2} \cdot \frac{v_{1j}^{-1}}{\psi_1(j)} \cdot \frac{x_2^{-1}}{x_2^{-1}} | 2 \leq j \leq n-1 \right\},$$
$$W_4 = \left\{ \frac{x_2}{x_2} \cdot \frac{v_{21}^{-1} u_{21}}{b_2} \cdot \frac{x_3^{-1}}{x_3^{-1}} \right\}, \quad W_5 = \left\{ \frac{x_3}{x_3} \cdot \frac{u_{2j}}{\phi_2(j)} \cdot \frac{x_3^{-1}}{x_3^{-1}} | 2 \leq j \leq n-1 \right\},$$
$$W_6 = \left\{ \frac{x_3}{x_3} \cdot \frac{u_{2n} v_{2n}^{-1}}{b_2^{-1}} \cdot \frac{x_4^{-1}}{x_4^{-1}} \right\}, \quad W_7 = \left\{ \frac{x_4}{x_4} \cdot \frac{v_{2j}^{-1}}{\psi_2(j)} \cdot \frac{x_4^{-1}}{x_4^{-1}} | 2 \leq j \leq n-1 \right\},$$
$$W_8 = \left\{ \frac{x_4}{x_4} \cdot \frac{v_{31}^{-1} u_{31}}{b_3} \cdot \frac{x_5^{-1}}{x_5^{-1}} \right\}, \quad W_9 = \left\{ \frac{x_5}{x_5} \cdot \frac{u_{3j}}{\phi_3(j)} \cdot \frac{x_5^{-1}}{x_5^{-1}} | 2 \leq j \leq n-1 \right\},$$
$$W_{10} = \left\{ \frac{x_5}{x_5} \cdot \frac{u_{3n} v_{3n}^{-1}}{b_3^{-1}} \cdot \frac{x_6^{-1}}{x_6^{-1}} \right\}, \quad W_{11} = \left\{ \frac{x_6}{x_6} \cdot \frac{v_{3j}^{-1}}{\psi_3(j)} \cdot \frac{x_6^{-1}}{x_6^{-1}} | 2 \leq j \leq n-1 \right\},$$
$$W_{12} = \left\{ \frac{x_6}{x_6} \cdot \frac{v_{41}^{-1} u_{41}}{b_4} \cdot \frac{x_7^{-1}}{x_7^{-1}} \right\}, \quad W_{13} = \left\{ \frac{x_7}{x_7} \cdot \frac{u_{4j}}{\phi_4(j)} \cdot \frac{x_7^{-1}}{x_7^{-1}} | 2 \leq j \leq n-1 \right\},$$
$$W_{14} = \left\{ \frac{x_7}{x_7} \cdot \frac{u_{4n} v_{4n}^{-1}}{b_4^{-1}} \cdot \frac{x_8^{-1}}{x_8^{-1}} \right\}, \quad W_{15} = \left\{ \frac{x_8}{x_8} \cdot \frac{v_{4j}^{-1}}{\psi_4(j)} \cdot \frac{x_8^{-1}}{x_8^{-1}} | 2 \leq j \leq n-1 \right\},$$

where $u_{ik} = \delta_i(u_k)$, $v_{ik} = \delta_i(v_k)$ for $1 \leq k \leq n$ and $1 \leq i \leq 4$, thus $u_{ik} \in \{a_i, b_i\}^*$ and $v_{ik}^{-1} \in \{a_i^{-1}, b_i^{-1}\}^*$. Given any two words $w_1, w_2 \in \mathrm{FG}(\Gamma)$, recall that we denote by $\frac{w_1}{w_2}$ the pair of words $(w_1, w_2) \in \mathrm{FG}(\Gamma) \times \mathrm{FG}(\Gamma)$ in the above table.

Note that each word in each pair from W_i has a so called 'border letter' on the left and right from $\mathrm{FG}(\Gamma_B)$. These are used to restrict the type of sequence[1] that can lead to an identity pair. The central element of each word (i.e. excluding the 'border letters') corresponds to particular words from P and we encode instance P four times separately, first in W_0, W_1, W_2, W_3, secondly in W_4, W_5, W_6, W_7 etc. using different alphabets for each encoding [2]. This may be seen in Figure 1, where A, B, C and D each separately encode instance P.

[1] The only sequences that may lead to an identity pair should be of the form of a cycle or a nested insertion of cycles as we shall show in Lemma 1.

[2] In the case of an incorrect solution for the Restricted PCP instance (i.e. an index sequence $i_1, \ldots, i_k$ such that $u_{i_1} \cdots u_{i_k} \neq v_{i_1} \cdots v_{i_k}$), the use of different alphabets

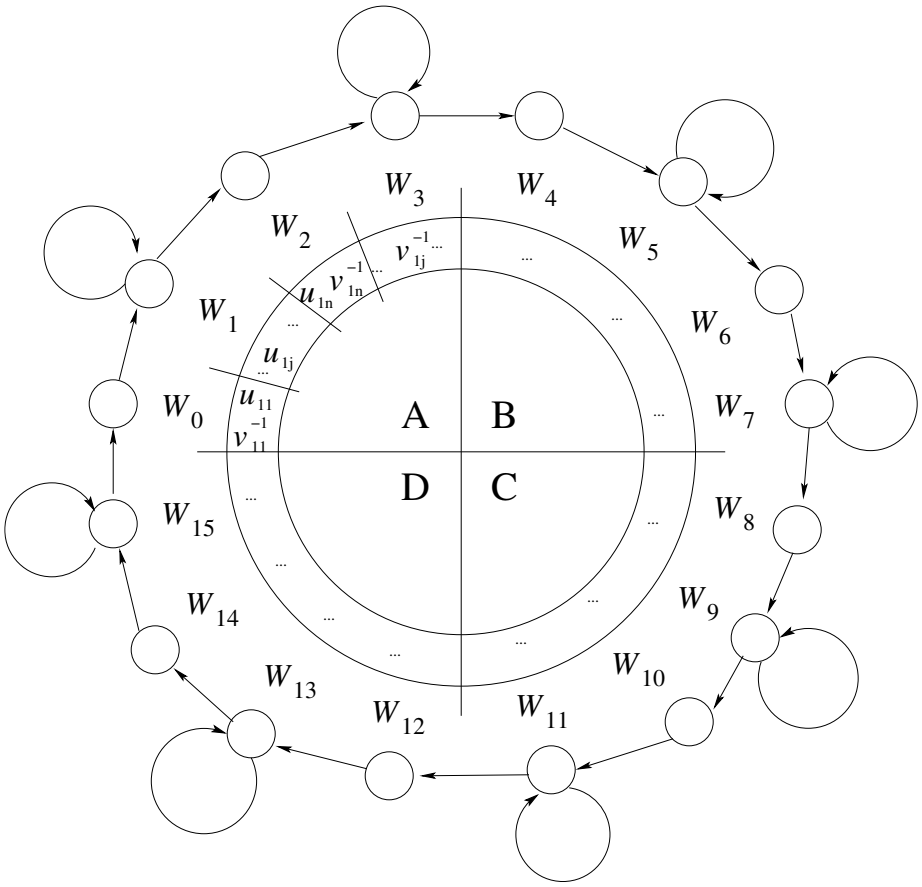

Fig. 1. The structure of a product which forms the identity.

This forms the set $W = \{(s_1, t_1), (s_2, t_2), \ldots, (s_m, t_m)\} \subset \mathrm{FG}(\Gamma) \times \mathrm{FG}(\Gamma)$. Let us define the s_i-word to mean the first word from pair of words (s_i, t_i) and the t_i-word for the second word of this pair. The t_i-words from each pair in W use an encoding which ensures that the set of s_i-words is concatenated in a particular order within each A, B, C and D part. We show in Lemma 2 that this encoding enforces a correct encoding of the Restricted PCP instance P within each part if that part gets reduced to two letters in the second word (the first and last 'border letters'). We adapt here our recently introduced index encoding technique from [2].

One of the important encoding concepts is a *cycle* of set W. We see that the first and last letters from each word of any pair of words from set $W_i \subset W$

for the four parts creates a sequence of non-empty parts that cannot be trivially cancelled from the left or right side.

only cancel with a pair of words from set $W_{i+1 \bmod 16}$ for $0 \leq i \leq 15$ and with elements from W_i itself if $i \bmod 2 \equiv 1$. We shall now define a 'cycle' of set W.

Definition 1. *An element $w \in W^*$ is called a cycle of W if it is of the form:*

$$w = w_i \cdot w_{(i+1) \bmod 16} \cdot \ldots \cdot w_{(i+15) \bmod 16} \in W^* \tag{1}$$

for some i: $0 \leq i \leq 15$, where $w_y \in W_y$ if $y \bmod 2 \equiv 0$ and $w_y \in W_y^*$ if $y \bmod 2 \equiv 1$.

For example a cycle could use element W_4 followed by a product of elements from W_5, then element W_6, followed by a product of elements from W_7 etc. As previously mentioned, the idea of the encoding is that a correct solution to the Restricted PCP instance P will be encoded *four times* in a correct solution to W, in elements from $\{W_0, \ldots, W_3\}, \{W_4, \ldots, W_7\}, \{W_8, \ldots, W_{11}\}$ and $\{W_{12}, \ldots, W_{15}\}$ separately.

We now define a *pattern generated by cycle insertions*. By this, we mean a product where cycles can be inserted within other cycles or appended to the end of them. For example, given an element $q_1 q_2 q_3 q_4 q_3' q_1' q_5 \in W^*$ where $q_1 q_1'$, q_2, $q_3 q_3'$, q_4 and q_5 are all cycles, then this would form a pattern generated by cycle insertions since it can be decomposed into cycles being nested or concatenated in the required way.

Lemma 1. *If instance W of the Identity Correspondence Problem has a solution, it must be constructed by an element $w \in W^*$ which forms either a single cycle or a pattern generated by cycle insertions (including concatenation).*

Proof. It is not difficult to see that the border symbols from Γ_B give us constraints on the type of patterns which can be considered as possible solutions to the ICP instance, i.e., which may have some form of word cancellation. These constraints can be considered as the state system represented in Figure 1 and require that a cycle is completed in a clockwise direction. Note that for any $w \in W_y^*$ where $y \bmod 2 \equiv 1$, it holds that w is not equal to $(\varepsilon, \varepsilon)$ since the left and right borders of each word are separated by a nonempty word from an alphabet not containing inverse elements. Let us assume that we have a pair of words from W_i for some i. The only possible way to cancel its border symbols is to complete a chain of cancellations by inverse border elements which will correspond to a clockwise traversal of states represented in Figure 1. Since at any time we can start to build a new cycle and all cycles must be completed we have that the only sequence of word pairs that can be equal to $(\varepsilon, \varepsilon)$ must be represented as a single cycle or a pattern generated by cycle insertions.

Definition 2. *For any product $Y \in W^*$ we shall denote by a* decomposition by parts *of Y, the decomposition $Y = Y_1 Y_2 \cdots Y_k$ where for each $1 \leq i \leq k$, if $Y_i \subset \mathrm{FG}(\Gamma_i \cup \Gamma_B) \times \mathrm{FG}(\Gamma_i \cup \Gamma_B)$ then $Y_{i+1} \subset \mathrm{FG}(\Gamma_j \cup \Gamma_B) \times \mathrm{FG}(\Gamma_j \cup \Gamma_B)$ where $1 \leq i, j \leq 4$ and $i \neq j$.*

For a cycle Q, the decomposition by parts of Q clearly gives either 4 or 5 parts in the decomposition. For example, we may have $Q = X_1 X_2 X_3 X_4 X_5$ where

$X_1 \in \mathrm{FG}(\Gamma_i \cup \Gamma_B) \times \mathrm{FG}(\Gamma_i \cup \Gamma_B)$ and thus $X_2 \in \mathrm{FG}(\Gamma_{(i+1 \bmod 4)} \cup \Gamma_B) \times \mathrm{FG}(\Gamma_{(i+1 \bmod 4)} \cup \Gamma_B)$ etc. X_5 is either empty or uses the same alphabet as X_1.

The s_i-words from each A, B, C, D part of Figure 1 will store all words from the instance of Restricted PCP, P separately using distinct alphabets. If we concatenate the s_i-words of one of these parts in the correct order and have the empty word (excluding initial and final 'border letters'), then this corresponds to a solution of P. By a correct order, we mean that if we have $u_{i1}u_{i2}\cdots u_{ik}$ for example, then they should be concatenated with $(v_{i1}v_{i2}\cdots v_{ik})^{-1} = v_{ik}^{-1}\cdots v_{i2}^{-1}v_{i1}^{-1}$. If the concatenation of these words equals ε, then we have a correct solution to P.

Let us here illustrate this fact with a simple example. Take a (standard) PCP instance $P = \{(aab, a), (a, baa)\}$. Clearly we have a solution to this instance since $(aab, a)(aab, a)(a, baa)(a, baa) = (aabaabaa, aabaabaa)$. Using the above encoding, we can alternatively write this as:

$$(aab)(aab)(a)(a) \cdot (baa)^{-1}(baa)^{-1}(a)^{-1}(a)^{-1} = \varepsilon,$$

which can also be seen as a solution where the words on the left are from the first words of each pair in P and the words on the right are the inverse of the second words from P. The idea is that on the right, using the inverse elements of the alphabet, we should have a palindrome of the word on the left and they should occur in the correct order. Here we used the sequence $1, 1, 2, 2$ on the left thus used the reverse sequence on the right, namely $2, 2, 1, 1$.

The encoding in the second words using ϕ_i, ψ_i and $\{b_i, b_i^{-1} | 1 \leq i \leq 4\}$ is used to ensure that any solution to W *must* use such a correct ordering in each A, B, C, D part. The next lemma formalizes this concept and is a modification of the technique presented in [2]. It also can be seen as a variant of Index Coding PCP, see [4], which is simpler to prove.

Lemma 2. *Given any part $X \in \mathrm{FG}(\Gamma_j \cup \Gamma_B) \times \mathrm{FG}(\Gamma_j \cup \Gamma_B)$, if the second word of X consists of only the initial and final 'border letters' $x_p x_q^{-1} \in \Gamma_B^*$, then the second word of X must be of the form*

$$x_p \cdot b_j \phi_j(z_1)\phi_j(z_2)\cdots\phi_j(z_k) \cdot b_j^{-1} \cdot \psi_j(z_k)\cdots\psi_j(z_2)\psi_j(z_1) \cdot x_q^{-1},$$

where $2 \leq z_1, z_2, \ldots, z_k \leq n-1$. (This corresponds to a 'correct' palindromic encoding of the Restricted PCP instance P within this part. We see that all elements except x_p and x_q^{-1} will be cancelled.)

Proof. Since X is a single part, we see that $(p, q) \in \{(8, 2), (2, 4), (4, 6), (6, 8)\}$ depending on the type of part X. Let us consider the case that X is a product over elements from $W_0 \cup W_1 \cup W_2 \cup W_3$ (part A in Figure 1). The proof for the other 'parts', B, C and D is analogous. Consider the morphisms used in the second words of these elements. If we have for example a word starting with the element from W_0, by the choice of 'border letters', it must be followed by an element from W_1^* or W_2 for cancellation to occur. In the former case (excluding 'border letters') it will thus be of the form $b_1 a_1^{z_1} b_1 \cdot a_1^{z_2} b_1 \cdots a_1^{z_k} b_1$ where each

$2 \leq z_i \leq n-1$. The only way to cancel this final b_1 is to eventually use W_2 (even if we use no element from W_1^*) since this is the only element whose second word starts with b_1^{-1} and this is the only other element within W cancelling the 'border letter' of W_1.

After this we must use an element from W_3^* to cancel the a_1 values since each $\psi(i)$ starts with a_i^{-1}. It is not difficult to see that we in fact must use these elements in the order $\psi(z_k) \cdot \psi(z_{k-1}) \cdots \psi(z_1)$ otherwise the 'b_1^{-1}' at the end of some $\psi(z_j)$ will not be cancelled on the left. The only way to cancel this 'b_1^{-1}' would be to use W_1 but this cannot follow W_3 by the choice of 'border letters'. With a correct sequence of W_3 elements, all the letters of the second words will cancel leaving the empty word ε (again excluding the 'border letters'). If we do not use this sequence, by the choice of the morphisms ϕ and ψ, the letters cannot be reduced to ε. See [2] for further details.

Finally note that if we do not start with the element from W_0 then, since the left 'border letter' of the pair of words in this element is x_8, we cannot use it to cancel the product later on, since this border essentially splits the pair of words in two. It is not difficult to see that without this element we cannot reduce a product to ε however since without the b_i element in the second word to cancel the last letter of W_2 or W_3 elements, they cannot be reduced. Thus we must have the given structure given in the lemma.

Lemma 3. *If there exists a solution to the Restricted PCP instance P, then there exists a solution to the Identity Correspondence Problem instance W.*

Proof. Assume we have a solution to P with indices $2 \leq i_1, i_2, \ldots, i_k \leq n-1$, i.e., $u_1 u_{i_1} \cdots u_{i_k} u_n = v_1 v_{i_1} \cdots v_{i_k} v_n$. We can explicitly define a product which will give a correct solution to the ICP instance W. Define a word $w = w_0 w_1 \cdots w_{15} \in W^*$ such that each $w_i \in W_i^*$. If $i \bmod 1 \equiv 0$, then $|w_i| = 1$. If $i \bmod 4 \equiv 1$, then w_i will be chosen from W_i^* using the indices $2 \leq i_1, i_2, \ldots, i_k \leq n-1$ for j. Finally, if $i \bmod 4 \equiv 3$, then w_i will be chosen from W_i^* using the indices $2 \leq i_k, i_{k-1}, \ldots, i_1 \leq n-1$ for j. A simple computation shows that since this sequence gave a correct solution to P, then w will be equal to $(\varepsilon, \varepsilon)$ and thus a solution to the ICP instance W.

Let us introduce several notations which will be useful for the analysis of cancellations that may occur in the construction. We shall define four 'types' of parts, A, B, C, D where type A parts use alphabet $\mathrm{FG}(\Gamma_1 \cup \Gamma_B) \times \mathrm{FG}(\Gamma_1 \cup \Gamma_B)$, type B parts use $\mathrm{FG}(\Gamma_2 \cup \Gamma_B) \times \mathrm{FG}(\Gamma_2 \cup \Gamma_B)$, type C parts use $\mathrm{FG}(\Gamma_3 \cup \Gamma_B) \times \mathrm{FG}(\Gamma_3 \cup \Gamma_B)$ and type D parts use $\mathrm{FG}(\Gamma_4 \cup \Gamma_B) \times \mathrm{FG}(\Gamma_4 \cup \Gamma_B)$ as in Figure 1. A cycle thus has a decomposition which is a permutation of $ABCD$.

We shall now define a function $\zeta : W^* \to \mathbb{N}$. Given any product $Y \in W^*$ with the decomposition by parts $Y = Y_1 Y_2 \cdots Y_k$, we first define the set of pairs of words $\{Z_1, Z_2, \ldots, Z_k\}$ where Z_i is a pair of words constructed from Y_i where we exclude the initial and final letters (from Γ_B) in each pair of words in Y_i. We let $\zeta(Y)$ denote the sum of non-identity words from $\{Z_1, Z_2, \ldots, Z_k\}$. Note that $Z_i \in \mathrm{FG}(\Gamma_j \cup \Gamma_B) \times \mathrm{FG}(\Gamma_j \cup \Gamma_B)$ for some $1 \leq j \leq 4$.

Thus for a single cycle Q, $0 \leq \zeta(Q) \leq 10$ since it can be decomposed to a maximum of 5 parts. If the first and second words in each decomposed part have a non-reducible word in between the borders, we have that $\zeta(Q)$ is equal to 10. If $\zeta(Q)$ equals 0, it means that all words in between the border elements are reducible to identity.

Lemma 4. *If there exists no solution to the encoded Restricted PCP instance P then for any cycle $Q \in W^+$ having decomposition by parts $Q = X_1X_2X_3X_4X_5$, the following holds:*

- $X_r \neq (\varepsilon, \varepsilon)$ *for all* r *where* $1 \leq r \leq 5$*;*
- $4 \leq \zeta(Q)$*;*
- $Q \neq (\varepsilon, \varepsilon)$*, i.e., a single cycle cannot be a solution to the Identity Correspondence Problem.*

Proof. Let Q be a single cycle of the form (1). Since it is a cycle, the 'border letters' of each pair will all cancel with each other and thus we may ignore letters from $\mathrm{FG}(\Gamma_B)$ (except for the first and last such border letters). Let $Q = X_1X_2X_3X_4X_5$ be its decomposition by parts (thus X_5 can be empty and four of the 'parts' use different alphabets).

Let us consider some X_r where $1 \leq r \leq 5$. We will show that X_r cannot be equal to $(\varepsilon, \varepsilon)$.

Since Q is a cycle, which has a specific structure, the first word of X_r, when concatenated, equals $v_{p1}^{-1}u_{p1}u_{pj_1}\cdots u_{pj_h}u_{pn}v_{pn}^{-1}v_{pk_l}^{-1}\cdots v_{pk_1}^{-1}$ for some $1 \leq p \leq 4$ and $h, l \geq 0$. If $j_i = k_i$ for all $1 \leq i \leq h$ with $h = l$ then this is a correct encoding of the Restricted PCP instance P which we have assumed has no solution, thus this word does not equal ε in this case. Therefore the elements must not be in a correct sequence if the first word equals ε. In this case however, the second word will now not equal ε by the choice of the morphisms ϕ_i and ψ_i as shown in Lemma 2. If we have such an incorrect ordering then when we multiply the second set of words (since also each morphism uses a different alphabet) they never equal ε which is not difficult to see.

So assuming that there is no solution to the Restricted PCP instance P, for any part X_r, $X_r \neq (\varepsilon, \varepsilon)$, i.e., at least one word in the pairs of words of each part does not equal ε (even ignoring initial and final border letters). Thus, crucially, if there exists no solution to the encoded Restricted PCP instance P, then $4 \leq \zeta(ABCD) \leq 8$ for a cycle $ABCD \in W^*$.

It follows from Lemma 4 that the statements of Lemma 1 can be restricted further. Lemma 1 asserts that the solution of ICP can be either a single cycle or a pattern that is formed by a nested insertion of cycles (including concatenation). It follows from Lemma 4 that if the Restricted PCP instance P does not have a solution, then a single cycle cannot be equal to $(\varepsilon, \varepsilon)$. We prove now that any solution to the corresponding ICP instance W cannot be in the form of cycle insertion unless the solution is in the form of a concatenation of several cycles each of which starts with the same element.

Lemma 5. *If there exists no solution to the Restricted PCP instance P, any solution to the corresponding ICP instance W cannot be in the form of cycle insertion unless the solution is in the form of a concatenation of several cycles each of which starts with the same element.*

Proof. Let us assume that a sequence of indices gives us a solution to ICP in the form LQR, where $L, Q, R \in W^+$ and Q is a cycle. We show that if $LQR = (\varepsilon, \varepsilon)$ then Q is not inserted inside of any other cycles and LQR is a concatenation of cycles each of which starts with the same element.

If LQR is equal to $(\varepsilon, \varepsilon)$ then $QRL = (\varepsilon, \varepsilon)$. By l, r, q let us define pairs of words constructed from L, Q, R where we exclude the initial and final border letters.

Let us assume that the single cycle Q is in the form where it starts and finishes with border letters $\frac{x_i}{x_i}$ and $\frac{x_i^{-1}}{x_i^{-1}}$, i.e., $Q = \frac{x_i}{x_i} \cdot q \cdot \frac{x_i^{-1}}{x_i^{-1}}$ where element q, when reduced (i.e. removing consecutive inverse elements), is in $\mathrm{FG}(\Gamma') \times \mathrm{FG}(\Gamma')$ where $\Gamma' = \Gamma \setminus \Gamma_B$, $Q \neq (\varepsilon, \varepsilon)$ and $LR = \frac{x_k}{x_k} \cdot l \cdot \frac{x_j^{-1}}{x_j^{-1}} \cdot \frac{x_j}{x_j} \cdot r \cdot \frac{x_k^{-1}}{x_k^{-1}}$ for some border letters $x_j, x_k \in \Gamma_B$.

Since q cannot be equal to $(\varepsilon, \varepsilon)$ by Lemma 4 and $QRL = (\varepsilon, \varepsilon)$ we have that the cycle Q can only be cancelled by a concatenation with RL. Thus the reduced form of rl is in $\mathrm{FG}(\Gamma') \times \mathrm{FG}(\Gamma')$ and RL must therefore be in the form of concatenations of cycles starting with a border symbol x_i: $RL = \frac{x_i}{x_i} \cdot r \cdot \frac{x_k^{-1}}{x_k^{-1}} \cdot \frac{x_k}{x_k} \cdot l \cdot \frac{x_i^{-1}}{x_i^{-1}}$. We see that QRL is therefore a concatenation of cycles.

Since the cycle Q can be factorized into two parts Q_1, Q_2 separated by border letters x_k, x_k^{-1}, i.e. $Q = Q_1 Q_2 = \frac{x_i}{x_i} \cdot q_1 \cdot \frac{x_k^{-1}}{x_k^{-1}} \cdot \frac{x_k}{x_k} \cdot q_2 \cdot \frac{x_i^{-1}}{x_i^{-1}}$, we have that

$$\begin{aligned} LQR &= \frac{x_k}{x_k} \cdot l \cdot \frac{x_i^{-1}}{x_i^{-1}} \cdot \frac{x_i}{x_i} \cdot q \cdot \frac{x_i^{-1}}{x_i^{-1}} \cdot \frac{x_i}{x_i} \cdot r \cdot \frac{x_k^{-1}}{x_k^{-1}} \\ &= \frac{x_k}{x_k} \cdot l \cdot \frac{x_i^{-1}}{x_i^{-1}} \cdot \frac{x_i}{x_i} \cdot q_1 \cdot \frac{x_k^{-1}}{x_k^{-1}} \cdot \frac{x_k}{x_k} \cdot q_2 \cdot \frac{x_i^{-1}}{x_i^{-1}} \cdot x_i \cdot r \cdot \frac{x_k^{-1}}{x_k^{-1}} \\ &= \frac{x_k}{x_k} \cdot l \cdot q_1 \cdot \frac{x_k^{-1}}{x_k^{-1}} \cdot \frac{x_k}{x_k} \cdot q_2 \cdot r \cdot \frac{x_k^{-1}}{x_k^{-1}}. \end{aligned}$$

Thus LQR is in the form of concatenation of cycles starting from a border letter x_k as required.

In the next lemma, we show that if the encoded Restricted PCP instance P has no solution, then a concatenation of cycles also cannot form a solution.

Lemma 6. *Given an instance of the Identity Correspondence Problem W encoding an instance P of Restricted Post's Correspondence Problem, if there exists no solution to P then for any product $X \in W^+$, it holds that $X \neq (\varepsilon, \varepsilon)$, i.e., if there is no solution to P, there is no solution to W.*

Proof. Let $X = X_1 X_2 \cdots X_k$ be the decomposition by parts of X. Assume $X = (\varepsilon, \varepsilon)$ is a solution to W, then since P has no solution by our assumption, Lemma 5 proves that X is a concatenation of cycles, each of which begins

with the same element. Note further that if any concatenation of cycles $c_{h_1} \cdots c_{h_l}$ (where each cycle starts with the same element) equals $(\varepsilon, \varepsilon)$, then this implies that we may cyclically permute the product so that it begins with element w_0 (at least one w_0 element must be present in any product of W giving an identity pair since we require at least one cycle).

Due to the 'border constraints', Lemma 5 gives us a restricted form of sequences that may lead to an identity pair, i.e., a type A pair of words must be followed by a type B pair of words which must be followed by a type C pair of words etc. This implies that at least one (cyclic) permutation of X must be of the form $ABCD \cdot ABCD \cdots ABCD$ if it equals $(\varepsilon, \varepsilon)$ since a single cycle is not a solution to W by Lemma 4.

Assuming that there is no solution to the Restricted PCP instance P, for any part, Y_i, we proved in Lemma 4 that $Y_i \neq (\varepsilon, \varepsilon)$, i.e., at least one word in the pairs of words of each part does not equal ε (even excluding initial and final border letters). Thus, crucially, if there exists no solution to P, then $4 \leq \zeta(ABCD) \leq 8$ for any cycle $ABCD \in W^*$.

We have that $\zeta(Q_1) \geq 4$ for any cycle $Q_1 \in W^*$. We shall now prove that $\zeta(Q_1 Q_2) \geq 4$ where $Q_2 \in W^*$ is also a cycle, i.e., by adding another cycle to the existing one, the number of 'empty parts' does not decrease. This means that we cannot reduce such a product to $(\varepsilon, \varepsilon)$ and thus if there exists no solution to instance P, there exists no solution to the Identity Correspondence Problem instance W as required. To see this, consider how many parts can be cancelled by adding a cycle. For example if the first word of Q_1 has an A part which cancels with the A part of Q_2, then the first word for the B, C, D parts of Q_1 must be ε. But since no part can be equal to $(\varepsilon, \varepsilon)$ we know that in Q_1, the second word of the B, C, D parts must not equal ε. The only element that can cancel the second word of Q_1 is thus the D part of Q_2. However this implies that the second word of the A, B, C parts of Q_2 all equal ε, thus the first word of the B, C parts of Q_2 cannot be ε and we have at least four non-ε parts (the first and second words of the B, C parts).

The same argument holds to cancel any part, thus we cannot reduce more than 4 parts by the concatenation of any two cycles. The first word can cancel at most two parts and the second words can cancel at most two parts but since we start with eight nonempty parts we remove only four parts at most leaving four remaining parts. Thus $\zeta(Q_1 Q_2) \geq \zeta(Q_1) + \zeta(Q_2) - 4 \geq 4$ as required. In fact, it is not difficult to see that this argument can be applied iteratively and thus $\zeta(Q_1 Q_2 \cdots Q_m) \geq 4$ always holds for any $m \geq 1$. If there is no solution to the Restricted PCP instance P then a concatenation of cycles cannot form a solution.

As an example of this lemma, take the following decomposition by parts (ignoring 'border letters') where $*_i$ is any nonempty word from $\mathrm{FG}(\Gamma_i)$ (where each $*_i$ is understood to be distinct):

$$ABCD \cdot ABCD = \left(\frac{\varepsilon}{*_1}\frac{*_2}{\varepsilon}\frac{*_3}{\varepsilon}\frac{*_4}{\varepsilon}\right)\left(\frac{\varepsilon}{*_1}\frac{\varepsilon}{*_2}\frac{\varepsilon}{*_3}\frac{*_4}{\varepsilon}\right) = \left(\frac{\varepsilon}{\varepsilon}\frac{*_2}{*_2}\frac{*_3}{*_3}\frac{\varepsilon}{\varepsilon}\right).$$

Here we cancel four parts in total and we are left with another four parts. The next $ABCD$ cycle that we concatenate cannot have $(\varepsilon, \varepsilon)$ for its first two parts however which will not thus cancel with the last non ε part above and thus the next concatenation of $ABCD$ cannot reduce the number of empty parts by less than four as we showed above, this is the iterative argument that we apply.

Theorem 3. *The Identity Correspondence Problem is undecidable for $m = 8(n-1)$ where n is the minimal number of pairs for which Restricted PCP is known to be undecidable (currently $n = 7$).*

Proof. Given an instance of the Identity Correspondence Problem, $W \subseteq \mathrm{FG}(\Gamma) \times \mathrm{FG}(\Gamma)$ which encodes an instance of Restricted Post's Correspondence Problem P. If there exists a solution to P, Lemma 3 shows that there also exists a solution to W. Lemma 6 then shows that if there does not exist a solution to the Restricted PCP instance P, there does not exist a solution to the Identity Correspondence Problem instance either, thus proving its undecidability. Since the restricted version of Post's Correspondence Problem is known to be undecidable for instances of size 7 by Theorem 2, this implies that ICP is undecidable for $m = 48$ by the construction of W.

It remains to prove that we may define the problem over a binary group alphabet $\{a, b, a^{-1}, b^{-1}\}$. This is not difficult however by a standard technique which we now outline. Given a group alphabet $\Sigma_1 = \{y_1, \ldots, y_k, y_1^{-1}, \ldots, y_k^{-1}\}$ and a binary group alphabet $\Sigma_2 = \{a, b, a^{-1}, b^{-1}\}$. Define $\sigma : \Sigma_1 \to \Sigma_2^*$ by $\sigma(y_i) = a^i b$ and $\sigma(y_i^{-1}) = (a^{-1})^i b^{-1}$. It is not difficult to see that this is an injective morphism and applying iteratively to each letter in each word of W proves the undecidability of the Identity Correspondence Problem over a binary group alphabet.

3 Applications of ICP

In this section we will provide a number of new results in semigroups using the undecidability of ICP. We first consider the "Group Problem" defined on a semigroup of pairs of words.

Problem 2 *Group Problem - Given an alphabet $\Sigma = \{a, b\}$, is the semigroup generated by a finite set of pairs of words $P = \{(u_1, v_1), (u_2, v_2), \ldots, (u_m, v_m)\} \subset \mathrm{FG}(\Sigma) \times \mathrm{FG}(\Sigma)$ a group?*

Theorem 4. *The Group Problem is undecidable for $m = 8(n-1)$ pairs of words where n is the minimal number of pairs for which Restricted PCP is known to be undecidable (currently $n = 7$).*

Proof. Let us assume by contradiction that the Group Problem is decidable for a semigroup S defined by pairs of words over a group alphabet and the operation of pairwise concatenation. If the identity element can be generated by the concatenation of word pairs

$$(u_{i_1}, v_{i_1})(u_{i_2}, v_{i_2}) \cdot \ldots \cdot (u_{i_k}, v_{i_k}) = (u_{i_1} u_{i_2} \cdot \ldots \cdot u_{i_k}, v_{i_1} v_{i_2} \cdot \ldots \cdot v_{i_k}) = (\varepsilon, \varepsilon)$$

then any cyclic permutation of words in this concatenation is also equal to $(\varepsilon, \varepsilon)$. Thus every element in the set of all pairs used in the generation of identity has an inverse element and this set generates a subgroup. Therefore the Identity Problem can be solved by checking if some nonempty subset of the original pairs generates a group. If there is a subset of S which generates a group then the identity element is in S. Otherwise the identity element is not generated by S.

It was not previously known whether the Identity Problem for matrix semigroups was decidable for any dimension greater than two. The Identity Problem in the two dimensional case for integral matrices was recently proved to be decidable in [10].

Theorem 5. *Given a semigroup S generated by a fixed number n of square four dimensional integral matrices, determining whether the identity matrix belongs to S is undecidable. This holds even for $n = 48$.*

Proof. We shall use a standard encoding to embed an instance of the Identity Correspondence Problem into a set of integral matrices. Given an instance of ICP say $W \subseteq \Sigma^* \times \Sigma^*$ where $\Sigma = \{a, b, a^{-1}, b^{-1}\}$ generates a free group. Define the morphism $\rho : \Sigma^* \to \mathbb{Z}^{2\times 2}$:

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

It is known from the literature that ρ is an injective homomorphism, i.e., the group generated by $\{\rho(a), \rho(b)\}$ is free, see for example [15]. For each pair of words $(w_1, w_2) \in W$, define the matrix $A_{w_1,w_2} = \rho(w_1) \oplus \rho(w_2)$ where $\oplus$ denotes the direct sum of two matrices. Let S be a semigroup generated by $\{A_{w_1,w_2} | (w_1, w_2) \in W\}$. If there exists a solution to ICP, i.e., $(\varepsilon, \varepsilon) \in W^+$, then we see that $\rho(\varepsilon) \oplus \rho(\varepsilon) = I_4 \in S$ where I_4 is the 4×4 identity matrix. Otherwise, since ρ is an injective homomorphism, $I_4 \notin S$.

It follows from the above construction that another open problem concerning the reachability of any diagonal matrix in a finitely generated integral matrix semigroup stated in [6] and as Open Problem 6 in [14], is also undecidable.

Corollary 1. *Given a finitely generated semigroup of integer matrices S, determining whether there exists any diagonal matrix in S is algorithmically undecidable.*

Proof. This result follows from the proof of Theorem 5. Note that in that theorem, the morphism ρ is injective and thus the only diagonal matrix in the range of ρ is the 2×2 identity matrix I_2 (corresponding to $\rho(\varepsilon)$), since diagonal matrices commute. Clearly then, the only diagonal matrix in the semigroup S of Theorem 5 is given by $\rho(\varepsilon) \oplus \rho(\varepsilon) = I_4$ where I_4 is the 4×4 identity matrix. Since determining if this matrix is in S was shown to be undecidable, it is also undecidable to determine if there exists any diagonal matrix in S.

Theorem 6. *Given a finite set of rotations on the* 3*-sphere. Determining whether this set of rotations generates a group is undecidable.*

Proof. We shall use the notation $\mathbb{H}$ to denote the set of quaternions. More details of quaternions used in this theorem can be found in [5]. The set of all unit quaternions forms the unit 3-sphere and any pair of unit quaternions a and b can represent a rotation in 4 dimensional space. A point $x = (x_1, x_2, x_3, x_4)$ on the 3-sphere may be represented by a quaternion $q_x = x_1 + x_2 i + x_3 j + x_4 k$ and rotated using the operation: $aq_x b^{-1}$. This gives a quaternion $q'_x = x'_1 + x'_2 i + x'_3 j + x'_4 k$ representing the rotated point $x' = (x'_1, x'_2, x'_3, x'_4)$.

We can define a morphism ξ from a group alphabet to unitary quaternions:

$$\xi(a) = \frac{3}{5} + \frac{4}{5} \cdot i; \xi(b) = \frac{3}{5} + \frac{4}{5} \cdot j.$$

It was proven in [5] that ξ is an injective homomorphism. We may thus convert pairs of words from an instance of the Identity Correspondence Problem into pairs of quaternions $\{(a_1, b_1), \ldots, (a_n, b_n)\} \subseteq \mathbb{H} \times \mathbb{H}$. Therefore we reduce the Group Problem for pairs of words over a group alphabet to the question of whether a finite set of rotations, $\{(a_1, b_1), \ldots, (a_n, b_n)\}$, represented by pairs of quaternions, generates a group.

4 Conclusion

In this paper we introduced the Identity Correspondence Problem, proved that it is undecidable and applied it to answer long standing open problems in matrix semigroups. In particular, we proved that the membership problem for the identity matrix in 4×4 integral matrix semigroups is undecidable. The identity matrix membership problem for 2×2 matrix semigroups was shown to be decidable in [10], but the problem in dimension 3 remains open. We believe that the Identity Correspondence Problem will be useful in identifying new areas of undecidable problems not only related to matrix problems but also to computational questions in abstract algebra, logic and combinatorics on words.

Acknowledgements - We would like to thank Prof. Tero Harju for useful discussions concerning this problem and the anonymous referees for their careful checking of this manuscript.

References

1. L. Babai, R. Beals, J. Cai, G. Ivanyos and E. M. Luks, Multiplicative Equations over Commuting Matrices, Proc. 7th ACM-SIAM Symp. on Discrete Algorithms, 498-507, (1996).
2. P. Bell, I. Potapov, On the Membership of Invertible Diagonal and Scalar Matrices, Theoretical Computer Science, 372(1), 37-45, (2007).
3. P. Bell, I. Potapov, On Undecidability Bounds for Matrix Decision Problems, Theoretical Computer Science 391(1-2), 3-13, (2008).

4. P. Bell, I. Potapov, Periodic and Infinite Traces in Matrix Semigroups, SOFSEM 2008: Theory and Practice of Computer Science, Lecture Notes in Computer Science 4910, 148-161 (2008).
5. P. Bell, I. Potapov, Reachability Problems in Quaternion Matrix and Rotation Semigroups, Information and Computation, Volume 206 , Issue 11, 1353-1361, (2008).
6. V. D. Blondel, J. Cassaigne and J. Karhumäki, Problem 10.3, Freeness of Multiplicative Matrix Semigroups. In: V. D. Blondel, A. Megretski (Eds.), Unsolved Problems in Mathematical Systems and Control Theory, Princeton University Press, 309-314, (2004).
7. V. D. Blondel, E. Jeandel, P. Koiran, N. Portier, Decidable and Undecidable Problems about Quantum Automata, SIAM Journal on Computing, 34:6, 1464-1473, (2005).
8. V. D. Blondel, J. Tsitsiklis, The Boundedness of All Products of a Pair of Matrices is Undecidable, Systems and Control Letters, 41:2:135-140, (2000).
9. J. Cassaigne, T. Harju, J. Karhumäki, On the Undecidability of Freeness of Matrix Semigroups, Intern. J. Alg. & Comp. 9, 295-305, (1999).
10. C. Choffrut, J. Karhumäki, Some Decision Problems on Integer Matrices, Theoretical Informatics and Applications, v39, 125-131, (2005).
11. F. D'Allesandro, Free Groups of Quaternions, Intern. J. of Alg and Comp. (IJAC), Volume 14, Number 1, (2004).
12. V. Halava, T. Harju, On Markov's Undecidability Theorem for Integer Matrices, TUCS Technical Report, Number 758, (2006).
13. V. Halava, T. Harju, M. Hirvensalo, Undecidability Bounds for Integer Matrices using Claus Instances, Intern. Journal of Foundations of Computer Science, Vol. 18, No. 5, 931-948, (2007).
14. T. Harju, Post Correspondence Problem and Small Dimensional Matrices, Lecture Notes in Computer Science, Springer Berlin, Volume 5583, 39-46, (2009).
15. R. C. Lyndon, P. E. Schupp, Combinatorial Group Theory, Springer-Verlag, (1977).
16. A. Markov, On Certain Insoluble Problems Concerning Matrices, Doklady Akad. Nauk SSSR, 539-542, (1947).
17. Y. Matiyasevich, G. Senizergues, Decision Problems for Semi-Thue Systems with a Few Rules, Theoretical Computer Science, 330(1), 145-169, (2005).
18. K. A. Mihailova, The Occurrence Problem for a Direct Product of Groups, Dokl. Akad. Nauk 119, 1103-1105, (1958). [in Russian]
19. F. Otto, On Deciding whether a Monoid is a Free Monoid or a Group, Acta Informatica, 23, 99-110, (1986).
20. M. Paterson, Unsolvability in 3×3 Matrices, Studies in Applied Mathematics, 49, (1970).
21. E. Post, A Variant of a Recursively Unsolvable Problem, Bulletin of the American Mathematical Society, 52 : 264-268, (1946).
22. S. Swierczkowski, A Class of Free Rotation Groups, Indag. Math. 5, no.2, 221-226, (1994).