Undecidability of Infinite Post Correspondence Problem for Instances of Size 9

Vesa Halava* Tero Harju

Department of Mathematics and
TUCS - Turku Centre for Computer Science
University of Turku
FIN-20014 Turku
Finland
E-mail: {vehalava,harju}@utu.fi

January 18, 2012

Abstract

In the infinite Post Correspondence Problem an instance (h, g) consists of two morphisms h and g, and the problem is to determine whether or not there exists an infinite word ω such that $h(\omega) = g(\omega)$. This problem was shown to be undecidable by K. Ruohonen (1985) in general. Recently V. D. Blondel and V. Canterini (*Theory Comput. Syst.* 36, 231–245, 2003) showed that this problem is undecidable for domain alphabets of size 105. Here we give a proof that the infinite Post Correspondence Problem is undecidable for instances where the morphisms have domains of 9 letters. The proof uses a recent result of Matiyasevich and Sénizergues and a modification of a result of Claus.

1 Introduction

An *instance* of the *Post Correspondence Problem* (PCP, for short), consists of two morphisms $h, g\colon A^* \to B^*$, where A and B are (finite) alphabets. In the Post Correspondence Problem it is asked whether or not an instance (h, g) has a *solution* w, i.e., a nonempty word $w \in A^*$ such that $h(w) = g(w)$. The *size* of the instance (h, g) is defined to be the cardinality $|A|$ of its domain alphabet A.

It is well known that the PCP is undecidable in its general form; see Post [10]. The borderline between decidable and undecidable sets of instances has been investigated in several occasions by restricting the instances of the PCP. For example, it is an easy exercise to show that the unary PCP, where the domain alphabet has only one letter, is decidable. An instance (h, g) of size two is said to be *binary*. It was proved in [3] that the PCP is decidable for binary instances; see also [5] for a somewhat simpler proof. On the other hand, the

*Corresponding author.

PCP is undecidable for instances with domain alphabets A satisfying $|A| \geq 7$; Matiyasevich and Sénizergues [9].

In this paper we shall consider infinite solutions of the instances (h, g). Two (finite) words u and v are said to be *comparable*, if one is a prefix of the other. Let $\omega = a_1 a_2 \cdots$ be an infinite word over A where $a_i \in A$ for each index $i = 1, 2, \cdots$. Note that $h(\omega) = g(\omega)$ if the morphisms h and g *agree* on ω, that is, if $h(u)$ and $g(u)$ are comparable for all finite prefixes u of ω. We also say that such an infinite word ω is an *infinite solution* of the instance $I = (h, g)$.

Juhani: "We write" pois

The problem whether or not a given instance of the PCP has an infinite solution is called naturally the *infinite PCP*, or ωPCP, for short. It was shown by Ruohonen [11] that there is no algorithm to determine whether a general instance of the PCP has an infinite solution. It was proved by Blondel and Canterini [1] using undecidability of the halting problem of the Turing machine that the ωPCP is undecidable for instances of size 105.

It was proved in [4] that the ωPCP is decidable for marked instances of the PCP. Later, using the previous result, it was shown in [6] that the ωPCP is decidable for all binary instances.

In this paper we shall prove that the ωPCP is undecidable for instance of size 9. Our proof rests on a result of Matiyasevich and Sénizergues [9], which states that there exists a 3-rule semi–Thue system with undecidable termination problem. From that, by modifying a construction of Claus [2], we obtain the desired result. We also prove that it is undecidable for instances of size 6, whether they have non-ultimately periodic infinite solution.

We shall now fix some notation. Let A be an alphabet. For a set $K \subseteq A^+$ of finite words, let

$$K^\omega = \{w_1 w_2 \cdots \mid w_i \in K \text{ for all } i \geq 1\}$$

be the set of all infinite concatenations of words from K. For a singleton set $K = \{w\}$, we let w^ω denote $\{w\}^\omega$. In particular, A^ω consists of all (one-way) infinite words $a_1 a_2 \cdots$ over the alphabet A. An infinite word $\omega \in A^\omega$ is called *ultimately periodic*, if it can be written in the form $\omega = uv^\omega$ for some finite words u and v.

The *empty word* is denoted by ε. A word $u \in A^*$ is said to be a *prefix* of a word $v \in A^*$, denoted by $u \leq v$, if $v = uw$ for some $w \in A^*$.

2 Semi–Thue systems

A *semi–Thue system* $T = (\Sigma, R)$ consists of an alphabet $\Sigma = \{a_1, a_2, \ldots, a_n\}$ and a relation $R \subseteq \Sigma^* \times \Sigma^*$, the elements of which are called the *rules* of T. For two words $u, v \in \Sigma^*$, we write $u \to_T v$, if there are words u_1 and u_2 such that

$$u = u_1 x u_2 \quad \text{and} \quad v = u_1 y u_2 \quad \text{where } (x, y) \in R.$$

Let $\to_T^*$ be the reflexive and transitive closure of the relation $\to$. Therefore, we have $u \to_T^* v$ if and only if either $u = v$ or there exists a finite sequence of words $u = v_1, v_2, \ldots, v_n = v$ such that $v_i \to_T v_{i+1}$ for each $i = 1, 2, \ldots, n-1$.

The *word problem* for a semi-Thue system $T = (\Sigma, R)$ is stated as follows: given two words $w_1, w_2 \in \Sigma^*$ determine whether or not $w_1 \to_T^* w_2$ holds in T. In the *individual word problem* we are given a fixed word w_0 and we ask, for input words w, whether or not $w \to_T^* w_0$ holds.

Let $w_0 \in \Sigma^*$ be a word, and $T = (\Sigma, R)$ a semi–Thue system. If there does not exist any infinite sequences of words $w_1, w_2, \ldots$ such that $w_i \to_T w_{i+1}$ for all $i \geq 0$, then we say that T *terminates* on w_0. Thus T terminates on w_0 if all derivations starting from w_0 are of finite length. In the *termination problem* we are given a word w_0 and a semi–Thue system T and it is asked whether or not T terminates on w_0.

$T \longleftrightarrow w_0$

The following remarkable results were proved in [9].

Theorem 1. *There exists 3–rule semi–Thue system with undecidable individual word problem and there exists 3–rule semi–Thue system with undecidable termination problem.*

3 Bounds for ωPCP

The next Theorem for the general Post Correspondence Problem is due to Claus [2].

Theorem 2. *If there is a semi–Thue system with n rules having an undecidable word problem, then the PCP is undecidable for instances of size $n + 4$.*

We shall now recall the construction of Claus, because it gives a nice partial result for undecidability of the ωPCP(see also [7], [8]). Let $T = (\Sigma, R)$ be a semi–Thue system. Note first that we may assume that Σ is binary. Indeed, for $\Sigma = \{a_1, a_2, \ldots, a_k\}$, define a coding $\varphi\colon \Sigma^* \to \{a, b\}^*$ with $\varphi(a_i) = ab^i a$ for all i. Then let $R' = \{(\varphi(u), \varphi(v)) \mid (u, v) \in R\}$ be a new set of rules, and define $T' = (\{a, b\}, R')$. It is immediate that $w \to_T w'$ in T if and only if $\varphi(w) \to_{T'} \varphi(w')$ in T'. It follows that if T has undecidable (individual) word problem or termination problem, then so does the semi–Thue system T'.

ref's

Next we define two special morphisms. For any alphabet Y and a nonempty word $s \in Z^*$, let $\ell_s, r_s\colon Y^* \to (Y \cup \{s\})^*$ be the left and right *desynchronizing morphisms* defined by

$$\ell_s(a) = sa \quad \text{and} \quad r_s(a) = as$$

for all letters $a \in Y$. Notice that for all words w, we have $\ell_s(w) \cdot s = s \cdot r_s(w)$. Later, we shall mostly use these morphisms for a single letter $s = d$ or its second power $s = d^2$.

Assume now that $T = (\{a, b\}, R)$ is a semi–Thue system, where $R = \{t_1, t_2, \ldots, t_n\}$ such that $t_i = (u_i, v_i)$. We may suppose without restriction that the rules $t_i \in R$ are encoded with φ, i.e., $u_i, v_i \in (abb^*a)^*$. In the following we shall consider R also an alphabet. Let $f = aa$ be a special word used as marker. Note that aa is not an image of φ.

*

Let $u, v \in \{a, b\}^*$ be two given words. We define morphisms

$$h, g\colon (\{a, b, d, e\} \cup R)^* \to \{a, b, d, e\}^*$$

by

$$\begin{aligned}
&h(x) = \ell_d(x), \quad && g(x) = r_d(x), \quad \text{for } x \in \{a, b\}, \\
&h(t_i) = \ell_d(v_i), \quad && g(t_i) = r_d(u_i), \text{ for } t_i \in R, \\
&h(d) = \ell_d(uf), && g(d) = d, \\
&h(e) = de, && g(e) = r(fv)e,
\end{aligned}$$

It can be proved, see e.g. [2] or [7], that the solutions (if exist) of (h, g) are necessarily of the form ref

$$dw_1 f w_2 f \cdots f w_m e,$$

where each w_i has the form

$$w_i = x_{i_0} t_{i_1} x_{i_1} t_{i_2} \cdots t_{p_i} x_{p_i} \tag{1}$$

for some words x_{i_j} not containing letters from R. Moreover, we have $w_i \to_T^* w_{i+1}$ for $i = 1, 2, \ldots, m-1$. Note that it is possible that $p_i = 0$, in which case w_i contains no letters from R.

For decision problems concerning infinite solutions, the construction of Claus is not directly useful. Indeed, the instances defined above have trivial infinite solutions, for example, $d(uf)^\omega$ is always an infinite solution. Still, we are able to prove

Lemma 1. *If the termination problem is undecidable for n-rule semi–Thue system, then it is undecidable for instances of the PCP size $n + 3$ whether or not there exists an infinite solution that is not ultimately periodic.*

Proof. Let $T = (\{a, b\}, R)$ be a n-rule semi–Thue system with undecidable termination problem provided by Theorem 1, and let the rules in T be $t_i = (u_i, v_i)$ for $i = 1, 2, \ldots, n$. Let u be the input word. Define the instance (h, g) as above except that the letter e is omitted, i.e., $h, g\colon (\{a, b, d\} \cup R)^* \to \{a, b, d\}^*$. We need to prove that there is an infinite derivation in T starting from u if and only if (h, g) has an infinite solution that is not ultimately periodic.

Assume first that there is an infinite derivation

$$u = w_1 \to_T w_2 \to_T \cdots,$$

where $u = w_1 = x_1 u_{i_1} y_1$ and $w_j = x_{j-1} v_{i_{j-1}} y_{j-1} = x_j u_{i_j} y_j$ for all $j \geq 2$. By the construction of the morphisms h and g, we have that $h(x_j t_{i_j} y_j) = \ell_d(w_{j+1})$ and $g(x_j t_{i_j} y_j) = r_d(w_j)$, and this gives us an infinite solution

$$\omega = d x_1 t_{i_1} y_1 f x_2 t_{i_2} y_2 f \cdots \tag{2}$$

of the instance (h, g). If ω is not ultimately periodic, we are done. Therefore, assume that ω is ultimately periodic. Notice that for all $w \in \{a, b, d\}^*$, we have $h(w) = \ell_d(w)$ and $g(w) = r_d(w)$. Therefore we can define a new infinite solution by

$$\omega' = d(w_1 f) x_1 t_{i_1} y_1 f (w_2 f)^2 x_2 t_{i_2} y_2 f (w_3 f)^3 \cdots$$

which is not ultimately periodic for any ω, since the words $w_i \in \{a, b, d\}^*$ do not contain letters from R.

In the other direction, assume that there is a non-ultimately periodic infinite solution ω of the instance (h, g). Then, necessarily $\omega = dw_1 f w_2 f \cdots$, where each w_i is as in (1). Since ω is not ultimately periodic, the set $I = \{i \mid p_i > 1 \text{ for } w_i\}$ is infinite. Indeed, if I is finite and $z = \max I$, then clearly $w_i = w_{z+1}$ for all $i \geq z+1$, which makes ω ultimately periodic; a contradiction. It is obvious that infinite I yields an infinite derivation for u in T.

Since the termination problem is undecidable for T, we conclude that the existence of non-ultimately periodic solutions to instances (h, g) is also undecidable. □

With Theorem 1, the previous lemma yields a nice corollary.

Corollary 1. *It is undecidable for an instance I of size 6 of the PCP whether or not I has an infinite solution that is not ultimately periodic.* muoto

Next we shall now prove our main result. In the previous construction of the instance (h, g), the problem is the existence of trivial infinite solutions. We need to redefine h and g so that in each infinite solution there will be a letter from R between every two occurrences of the special word f. For this, we need a two more letters in the domain alphabet.

Lemma 2. *If the termination problem is undecidable for a semi–Thue system T with n rules, then the ωPCP is undecidable for instances of size $n + 6$.*

Proof. Let again $T = (\{a, b\}, R)$ be a n-rule semi–Thue system with undecidable termination problem provided by Theorem 1, and let the rules in T be $t_i = (u_i, v_i)$ for $i = 1, 2, \ldots, n$. We can assume that the rules t_i are encoded with φ. Let u be the input word.

The domain alphabet of our instance will be $A = \{a_1, a_2, b_1, b_2, d, \#\} \cup R$, where d is for begin and synchronization and $\#$ is special separator of the words in a derivation. Define the morphisms $h, g\colon A^* \to \{a, b, d, \#\}^*$ by

$$\begin{aligned}
&h(x_1) = dxd, && g(x_1) = xdd, && \text{for } x \in \{a, b\},\\
&h(x_2) = ddx, && g(x_2) = xdd, && \text{for } x \in \{a, b\},\\
&h(t_i) = d^{-1}\ell_{d^2}(v_i), && g(t_i) = r_{d^2}(u_i), && \text{for } t_i \in R,\\
&h(d) = \ell_{d^2}(u)dd\#d, && g(d) = dd, &&\\
&h(\#) = dd\#d, && g(\#) = \#dd. &&
\end{aligned}$$

In the special case, where $v_i = \varepsilon$, we define $h(t_i) = d$.

Each infinite solution of (h, g) is of the form

$$dw_1\#w_2\#w_3\#\cdots, \tag{3}$$

where

$$w_j = x_j t_{i_j} y_j \tag{4}$$

for some $t_{i_j} \in R$, $x_j \in \{a_1, b_1\}^*$ and $y_j \in \{a_2, b_2\}^*$ for all j. Indeed, the image $g(w)$ is always of the form $r_{d^2}(v)$, and therefore, by the form of h, between two separators $\#$ there must occur exactly one letter $t \in R$. Also, the separator $\#$ must be followed by words in $\{a_1, b_1\}^*$ before the next occurrence of a letter $t \in R$. By the form of $h(t)$ the following word before next separator must be in $\{a_2, b_2\}^*$. The form (3) follows when we observe that there must be infinitely many separators $\#$ in each infinite solution. Indeed, all solutions begin with a d, and there is one occurrence $\#$ in $h(d)$ and no occurrences of $\#$ in $g(d)$. Later each occurrences of $\#$ is produced from $\#$ by both g and h. Therefore there are infinitely many letters $\#$ in each infinite solution.

Define a mapping $\alpha\colon \{a_1a_2, b_1, b_2, d\}^* \to \{a, b\}^*$ which removes the indices and the letters d, i.e., $\alpha(z_i) = z$ for all $z \in \{a, b\}$, and $\alpha(d) = \varepsilon$. For w_j from (4), we have $\alpha(g(w_j)) = \alpha(x_j)u_{i_j}\alpha(y_j)$ and $\alpha(h(w_j)) = \alpha(x_j)v_{i_j}\alpha(y_j)$. Since the morphism h runs ahead of g, necessarily $\alpha(g(w_j)) = \alpha(h(w_{j-1})$ for $j \geq 2$

and $\alpha(g(dw_1)) = u = \alpha(h(d))\#^{-1}$. Let then $u_i = \alpha(g(w_i))$. Now, it is easy to conclude that (h, g) has an infinite solution (3) if and only if

$$u = u_1 \to u_2 \to u_3 \to \ldots,$$

which proves the claim. □

Theorem 1 and Lemma 2 yield our main result,

Corollary 2. *It is undecidable whether or not an instance of size 9 of the PCP has an infinite solution.*

Note that it is not known whether the ωPCP is undecidable for instance of size $3 \leq n \leq 8$.

Finally, we note that in Theorem 4.1 of [1] it was proved that if the ωPCP is undecidable for instances of size n, then the *isolation threshold problem* for the *probalistic finite automata* with two letters and 4n states and the *isolated threshold existence problem* for probalistic finite automata with two letters and 22n + 44 states, are undecidable (for definitions, see [1]). It follows from Corollary 2 the these problems are undecidable for 36 and 242 states, respectively. The bounds proven in [1] were 420 and 2354 states.

Acknowledgement. We thank Prof. Juhani Karhumäki for useful discussions.

References

[1] V. D. Blondel, and V. Canterini. Undecidable problems for probabilistic automata of fixed dimension. *Theory Comput. Syst.*, 36(3), 231–245, 2003.

[2] V. Claus. Some remarks on PCP(k) and related problems. *Bull. EATCS*, 12:54–61, 1980.

[3] A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg. The (generalized) Post Correspondence Problem with lists consisting of two words is decidable. *Theoret. Comput. Sci.*, 21:119–144, 1982.

[4] V. Halava and T. Harju. Infinite solutions of the marked Post Correspondence Problem. In J. Karhumäki, W. Brauer, H. Ehrig and A. Salomaa, editors, *Formal and Natural Computing*, volume 2300 of *Lecture Notes in Comput. Sci.*, pages 57–68. Springer-Verlag, 2002.

[5] V. Halava, T. Harju, and M. Hirvensalo. Binary (generalized) Post Correspondence Problem. *Theoret. Comput. Sci.*, 276:183–204, 2002.

[6] V. Halava, T. Harju, and J. Karhumäki. Decidability of the binary infinite Post Correspondence Problem. *Discrete Appl. Math.*, 130:521–526, 2003.

[7] T. Harju and J. Karhumäki. Morphisms. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 1. pp. 439–510, Springer-Verlag, 1997.

[8] T. Harju, J. Karhumäki, and D. Krob. Remarks on generalized Post correspondence problem. In C. Puech and R. Reischuk, editors, *STACS'96*, volume 1046 of *Lecture Notes in Comput. Sci.*, pages 39–48. Springer-Verlag, 1996.

[9] Y. Matiyasevich and G. Sénizergues. Decision problems for semi-Thue systems with a few rules. *Theor. Comput. Sci.* 330(1):145-169, 2005.

[10] E. Post. A variant of a recursively unsolvable problem. *Bull. of Amer. Math. Soc.*, 52:264–268, 1946.

[11] K. Ruohonen. Reversible machines and Post's correspondence problem for biprefix morphisms. *J. Inform. Process. Cybernet.* EIK, 21(12):579–595, 1985.