Generalized PCP Is Decidable for Marked Morphisms

Vesa Halava

Turku Centre for Computer Science TUCS
Lemminkäisenkatu 14 A, FIN-20520 Turku, Finland.
E-mail: vehalava@cs.utu.fi

Tero Harju

Department of Mathematics, University of Turku
FIN-20014 Turku, Finland.
E-mail: harju@utu.fi

Mika Hirvensalo

Department of Mathematics, University of Turku
FIN-20014, Turku, Finland
and
Turku Centre for Computer Science.
E-mail: mikhirve@cs.utu.fi

Supported by the Academy of Finland under grant 44087

Turku Centre for Computer Science
TUCS Technical Report No 283
May 1999

ISBN 952-12-0470-2
ISSN 1239-1891

Abstract

We prove that the generalized Post Correspondence Problem (GPCP) is decidable for marked morphisms. This result gives as a corollary a shorter proof for the decidability of the binary PCP, proved in 1982 by Ehrenfeucht, Karhumäki and Rozenberg.

Keywords: decidability, marked morphism, generalized Post Correspondence Problem, GPCP

TUCS Research Group
Theory Group: Mathematical Structures in Computer Science

1 Introduction

Let A and B be two finite alphabets and h, g two morphisms $h, g\colon A^* \to B^*$. The *Post Correspondence Problem*, PCP for short, is to determine if there exists a nonempty word $w \in A^*$ such that $h(w) = g(w)$. It was proved by Post [5] that this problem is undecidable. The PCP is one of the most useful problems for deriving other undecidability results.

Restricting the *instances* (h, g) of the PCP may make the problem decidable. For example, if we assume that the *size* of the instance, i.e. $|A|$, is at most two, then the PCP is decidable, see Ehrenfeucht, Karhumäki and Rozenberg [1]. On the other hand it is known that if $|A| \geq 7$, then the problem remains undecidable, see Matiyasevich and Sénizerques [4].

It was proved by Halava, Hirvensalo and de Wolf that the PCP is decidable if the morphisms are *marked* [2]. A morphism h is called marked, if $h(x)$ and $h(y)$ start with a different letter whenever $x, y \in \Sigma$ and $x \neq y$. In this paper we consider a modification of the PCP, called the *generalized PCP*, GPCP for short. An instance of the GPCP consists of two morphisms $h, g\colon A^* \to B^*$ and words $p_1, p_2, s_1, s_2 \in B^*$. The GPCP is to tell whether or not there exists a word $w \in A^*$ such that

$$p_1 h(w) s_1 = p_2 g(w) s_2.$$

We shall denote the instance of the GPCP by $((p_1, p_2), h, g, (s_1, s_2))$. Actually Ehrenfeucht, Karhumäki and Rozenberg proved in [1] that the GPCP for marked morphisms is decidable for binary alphabets A, and since every binary instance of the PCP is either periodic (see [3]) or can be reduced to an instance of the binary GPCP with marked morphisms, they had that the binary PCP is decidable. In this paper we shall prove the following theorem.

Theorem 1. *The GPCP with marked morphisms is decidable for any alphabet size.*

The decidability of the marked GPCP was mentioned as an open problem in [2]. As a general reference for the PCP and the GPCP we give [3].

Our proof of Theorem 1 uses the idea of reducing a problem instance to finitely many new instances such that one of these new instances has a solution if and only if the original one has. We will show that the iterative use of this reduction process will eventually give us the desired decision method.

We fix first some notations. The *empty word* is denoted by ε. A word $x \in A^*$ is said to be a *prefix* of $y \in A^*$, if there is $z \in A^*$ such that $y = xz$. This will be denoted by $x \leq y$. Similarly, a word $x \in A^*$ is said to be a *suffix* of $y \in A^*$, if there is $z \in A^*$ such that $y = zx$. This will be denoted by $x \preccurlyeq y$. We say that x and y are *comparable* if $x \leq y$ or $y \leq x$. Moreover, a word $x \in A^*$ is said to be a *factor* of $y \in A^*$, if there are words $z_1, z_2 \in A^*$ such that $y = z_1 x z_2$.

Finally we introduce a convention that will smoothen some concepts to be used later: If $h : A^* \to B^*$ is a marked morphism, then clearly $|A| \leq |B|$. Renaming the source alphabet A we can always assume even that $A \subseteq B$ and also that for each $a \in A$, $h(a)$ begins with a. However, notice carefully that for a morphism pair $h, g : A^* \to B^*$ given as an instance of the GPCP, we *cannot assume* that *both* $h(a)$ and $g(a)$ begin with a. That $h(a)$ begins with a will be a permanent premise hereafter.

2 Modified instances

Let $I = ((p_1, p_2), h, g, (s_1, s_2))$, where $h, g \colon A^* \to B^*$, be an instance of the marked GPCP. A word $w \in A^*$ is called a *solution* of I, if

$$p_1 h(w) s_1 = p_2 g(w) s_2.$$

A solution $w = w_0 v w_1$ is called *minimal* if for all $v \neq \varepsilon$, $w_0 w_1$ is not a solution.

The instances

$$I = ((p_1, p_2), h, g, (s_1, s_2)) \tag{1}$$

can be reduced to instances, where $p_1 = \varepsilon$ or $p_2 = \varepsilon$ and $s_1 = \varepsilon$ or $s_2 = \varepsilon$, since to have a solution we must have $p_1 \leq p_2$ or $p_2 \leq p_1$, and $s_1 \preccurlyeq s_2$ or $s_2 \preccurlyeq s_1$.

We first modify the marked GPCP by requiring that the solutions begin with a new fixed letter $\#$. For this, let $\#$ be a new letter. If in (1) $p_1 \neq \varepsilon$ or $p_2 \neq \varepsilon$ we extend the morphisms by defining $h(\#) = \# p_1$ and $g(\#) = \# p_2$. On the other hand, if $p_1 = \varepsilon = p_2$, we fix a letter $a_0 \in A$ and define $h(\#) = \# h(a_0)$ and $g(\#) = \# g(a_0)$.

In both cases, the extended morphisms $h, g \colon (A \cup \{\#\})^* \to (B \cup \{\#\})^*$ remain marked, and in the latter case, (1) has a solution that begins with a_0 if and only if the modified instance $((\varepsilon, \varepsilon), h, g, (s_1, s_2))$ has a solution in $\# A^*$. Clearly the marked GPCP is decidable if we can decide whether a solution exists for each $a_0 \in A$.

Therefore we can restrict to the instances of the form

$$(\#, h, g, (s_1, s_2)) \qquad (s_1 = \varepsilon \text{ or } s_2 = \varepsilon), \tag{2}$$

where the solutions w are required to satisfy $w \in \#(A \setminus \{\#\})^*$ and $h(w)s_1 = g(w)s_2$.

If an instance (2) has $s_1 = s_2 = \varepsilon$, then we have to check whether there is a solution of the marked PCP beginning with the letter $\#$, but this is decidable by the result of [2]. Therefore we shall assume that $s_1 s_2 \neq \varepsilon$. We have

Lemma 1. *The GPCP for marked morphisms is decidable if and only if it decidable for the instances $(\#, h, g, (s_1, s_2))$, where $s_1 = \varepsilon$ or $s_2 = \varepsilon$.*

3 Blocks and successors of instances

A *block* of an instance $I = (h, g)$ of the marked PCP is a pair $(u, v) \in A^+ \times A^+$ such that $h(u) = g(v)$ and for all nonempty prefixes $u_1 \leq u$, $v_1 \leq v$, $h(u_1) = g(v_1)$ implies $u_1 = u$ and $v_1 = v$. If there is no danger of confusion, we will also say that $h(u) = g(v)$ is a block. The words u and v are called *block words* (of h and g respectively). Letter $a \in A$ is a *block letter* if there is a block (u, v) such that $a \leq h(u), g(v)$ (now also $a \leq u$, since we assumed that $h(a)$ begins with a and h is marked). Accordingly, a block is a minimal nontrivial solution of the equation $h(x) = g(y)$.

Lemma 2. *Let (h, g) be an instance of the marked PCP for $h, g\colon A^* \to B^*$. Then for each letter $a \in A$, there exists at most one block (u, v) such that $a \leq u$. In particular, the instance (h, g) has at most $|A|$ blocks. Moreover, the blocks of (h, g) can be effectively found.*

Proof. The claim follows since the blocks can be determined inductively: If (u, v) is a block and $a \leq u$, then the first letter of $v = bv'$ is uniquely determined because g is marked. We define $u_1 = a$, $v_1 = b$. Assume that we have found a pair (u_i, v_i) for which $h(u_i)$ and $g(v_i)$ are comparable. If $h(u_i) \neq g(v_i)$ and $h(u_i) \leq g(v_i)$, there is, due to markedness, at most one letter c such that $h(u_i c)$ and $g(v_i)$ are comparable. In this case we set $u_{i+1} = u_i c$, $v_{i+1} = v_i$. The symmetric case $g(v_i) \leq h(u_i)$ is treated similarly. The uniqueness eventually follows from the minimality of the blocks. The latter claim is evident since $\{u \mid h(u) = g(v) \text{ for some } v\}$ is a regular set.
□ □

For two words $s_1, s_2 \in A^*$ with $s_1 = \varepsilon$ or $s_2 = \varepsilon$, a pair (u, v) is called an *end block* (or an (s_1, s_2)*-end block*, to be precise) if $h(u)s_1 = g(v)s_2$ and (u_1, v_1) is not a block for any $u_1 \leq u$ and $v_1 \leq v$. Let

$$E_a = \{(u, v) \mid (u, v) \text{ is an end block and } a \leq h(u) \text{ or } a \leq g(v)\}$$

be the set of all end blocks for the letter a.

Lemma 3. *Let $I = (\#, h, g, (s_1, s_2))$ be an instance of the marked GPCP, $s_1 = \varepsilon$ or $s_2 = \varepsilon$ and a a fixed letter. The set of end blocks E_a is a rational relation and can be effectively found. Moreover,*

(i) If a is a block letter, E_a is finite.

(ii) If E_a is infinite, then it is a union of a finite set and finite number of sets

$$\{(xu^k, yv^k w) \mid k \geq 0\} \text{ and } \{(xu^k w, yv^k) \mid k \geq 0\}$$

for some words u, v, x, y, w.

Proof. Without loss of generality we may assume that $s_2 = \varepsilon$. The end blocks can be found similarly as we found the blocks for a letter a in the proof of Lemma 2: We check first if $s_1 = g(v)$ for some word v. If so, (ε, v) is an end block. Then we construct the sequence (u_i, v_i) as in Lemma 2. Recall that $a \leq u_i \leq u_{i+1}$, $v_i \leq v_{i+1}$ and that $h(u_i)$ and $g(v_i)$ are always comparable. Whenever $h(u_i)z_i = g(v_i)$ for some $z_i \leq s_1$, we can check if there is a word w_i such that $h(u_i)s_1 = g(v_i w_i)$. If such a w_i exists, it is unique because g is marked. Consequently $(u_i, v_i w_i)$ is an end block. Notice that i is not necessarily unique.

If a is a block letter for a block (u, v), then always $u_i \leq u$ and $v_i \leq v$ and the sequence (u_i, v_i) terminates. But then there are only finitely many possible z_i such that $h(u_i)z_i = g(v_i)$ and $z_i \leq s_1$. The claim (i) follows hereby.

If a is not a block letter, we have two possibilities:
(1) An attempt to find a block for the letter a terminates because for some i we have $h(u_i)z_i = g(v_i)$ (resp. $h(u_i) = g(v_i)z_i$) but $h(u_i c)$ and $g(v_i)$ (resp. $h(u_i)$ and $g(v_i c)$) are noncomparable for each c. In this case the sequence (u_i, v_i) is finite and there can be only finitely many end blocks.
(2) The sequence (u_i, v_i) is infinite. This is possible only if there are words $x, y, u, v \in A^*$ such that $|h(u)| = |g(v)| \geq 1$ and $h(x)s = g(y)$ and $h(xu)s = g(yv)$. But then for $i \geq |x| + |y|$ we have $(u_i, v_i) = (xu^k u', yv^k v')$, where $u' \leq u$ and $v' \leq v$. As above, any end block is of form $(u_i, v_i w_i)$. But for $i \geq |x| + |y|$, an end block can always be written as $(xu^k u', yv^k v' w_i)$ or equivalently as $(xu'(u''u')^k, yv'(v''v')^k w_i)$ to get the desired form (here $u = u'u''$ and $v = v'v''$). The claim (ii) follows, since there are only finitely many prefixes u' and v' and there are at most $|x| + |y|$ other potential end blocks. The rationality of E_a follows from the proofs for (i) and (ii). □ □

We shall call $(xu^k, yv^k w)$ and $(xu^k w, yv^k)$ in Lemma 3 (ii) *extendible end blocks.*

Let $I = (\#, h, g, (s_1, s_2))$ be an instance of the marked GPCP. For a solution $w \in A^*$, $h(w)s_1 = g(w)s_2$, of I,

$$w = u_1 u_2 \ldots u_{k+1} = v_1 v_2 \ldots v_{k+1}$$

is a *block decomposition* for w, if

1. (u_i, v_i) is a block for each $i = 1, 2, \ldots, k$,
2. (u_{k+1}, v_{k+1}) is an (s_1, s_2)-end block.

Because the blocks are minimal solutions to $h(u_i) = g(v_i)$, it is easy to see that the following lemma holds.

Lemma 4. *Every solution $w \in A^*$ of I has a unique block decomposition.*

Note that, since the block decomposition of a solution may consist only of an end block, it is necessary to construct also the set $E_\#$.

$h(w)$	$h(u_1)$	$h(u_2)$	$\cdots$	$h(u_k)$	$h(v_{k+1})s_1$
$g(w)$	$g(v_1)$	$g(v_2)$	$\cdots$	$g(v_k)$	$g(v_{k+1})$

Figure 1: Block decomposition of a solution w (for $s_2 = \varepsilon$)

Let $I = (h, g)$ be an instance of the marked PCP with $h, g \colon A^* \to B^*$ and

$$A' = \{a \in A \mid a \text{ is a block letter}\}. \tag{3}$$

We define the *successor* of I to be $I' = (h', g')$, where the morphisms h' and g' are from $(A')^*$ into A^* such that

$$h'(a) = u \quad \text{ and } g'(a) = v, \tag{4}$$

where (u, v) is a block for the letter $a \in A'$.

Lemma 5. *Let $I = (h, g)$ be an instance of the marked PCP and $I' = (h', g')$ be its successor.*
(i) I' is an instance of the marked PCP.
(ii) I has a solution if and only if I' has.
(iii) $hh'(x) = gg'(x)$ for all $x \in (A')^*$.
(iv) If $a \leq h(a)$ for each a, then also $a \leq h'(a)$ for each a.

Proof. This is clear by the construction and by the definition of the blocks.
□ □

The definition of a successor gives inductively a sequence of instances I_i, where $I_0 = I$ and $I_{i+1} = I_i'$. The decidability of the marked PCP in [2] was eventually based on the fact that the successor sequence defined above has only finitely many distinct instances. The authors of [2] used two measures for an instance I of the marked PCP, namely the size of the alphabet and the *suffix complexity*:

$$\begin{aligned}\sigma(I) &= |\cup_{a\in\Sigma}\{x \mid x \text{ is a proper suffix of } g(a)\}| \\ &+ |\cup_{a\in\Sigma}\{x \mid x \text{ is a proper suffix of } h(a)\}|.\end{aligned}$$

It is clear that for alphabet sizes of I' and I we have $|A'| \leq |A|$. That $\sigma(I') \leq \sigma(I)$ is not so straightforward. The next lemma can be found in [2].

Lemma 6. *If I is an instance of the marked PCP and I' is its successor then $\sigma(I') \leq \sigma(I)$.*

The previous lemma together with $|A'| \leq |A|$ yields the following result, see [2].

Lemma 7. *Let I be an instance of the marked PCP. Then there exist n_0 and d such that $I_{i+d} = I_i$ for all $i \geq n_0$. The numbers n_0 and d can be effectively found.*

The previous lemma means that after n_0 consecutive successors the instances begin to cycle: $I_{n_0}, \ldots, I_{n_0+d} = I_{n_0}, \ldots$. In this loop the alphabet size and the suffix complexity are constant.

Next we define the successors of the instances $I = (\#, h, g, (s_1, s_2))$ of the marked GPCP. Let (h', g') be the successor of (h, g) and let (u, v) be any end block of I. Then

$$I'(u, v) = (\#, h', g', (s'_1, s'_2))$$

is the *successor* of I w.r.t. (u, v), where (s'_1, s'_2) is defined as follows: If $v \preccurlyeq u$, then $s'_1 = uv^{-1}$ and $s'_2 = \varepsilon$ and if $u \preccurlyeq v$, then $s'_1 = \varepsilon$ and $s'_2 = vu^{-1}$. Otherwise $I'(u, v)$ is not defined. Note that for $(u, v) \in E_\#$ the successor $I'(u, v)$ is defined if and only if $u = v$, since $\#$ is a special symbol not in A. Moreover, $u = v$ is a solution of I.

Lemma 8. *An instance $I = (\#, h, g, (s_1, s_2))$ has a solution if and only if the successor $I'(u, v) = (\#, h', g', (s'_1, s'_2))$ has a solution for some end block (u, v). Moreover, each solution w to I can be written as $w = h'(w')u = g'(w')v$, where w' is a solution of I' and (u, v) an end block of I.*

Proof. Assume first that I has a solution w with the block decomposition

$$w = u_1 u_2 \ldots u_{k+1} = v_1 v_2 \ldots v_{k+1},$$

where (u_i, v_i) is a block for the letter a_i, for $1 \leq i \leq k$, and (u_{k+1}, v_{k+1}) is an end block. Clearly $u_{k+1} \preccurlyeq v_{k+1}$ or $v_{k+1} \preccurlyeq u_{k+1}$. If the first case holds, then $s'_1 = \varepsilon$, $s'_2 = v_{k+1} u_{k+1}^{-1}$ and $I'(u, v) = (\#, h', g', (s'_1, s'_2))$. Now

$$h'(a_1 \ldots a_k) = w u_{k+1}^{-1} = g'(a_1 \ldots a_k) s'_2,$$

i.e. $I'(u, v)$ has a solution $w' = a_1 \ldots a_k$ and $w = h'(w') u_{k+1} = g'(w') v_{k+1}$. Case $v_{k+1} \preccurlyeq u_{k+1}$ is symmetric.

Assume then that

$$I'(u, v) = (\#, h', g', (s'_1, s'_2))$$

has a solution w' i.e., $h'(w') s'_1 = g'(w') s'_2$. Then also $h'(w')u = g'(w')v$ and by the definition of the end blocks and Lemma 5,

$$h(h'(w')u)s_1 = h(h'(w'))h(u)s_1 = g(g'(w'))g(v)s_2 = g(g'(w')v)s_2,$$

so $h'(w')u = g'(w')v$ is a solution of I. □ □

For $(u, v) \in E_\#$ in the previous theorem $w' = \varepsilon$. Clearly it is of no use to construct the successors of I for $(u, v) \in E_\#$ with $u = v$, since we already found a solution for I. We shall return to this in Lemma 11.

4 Reducing the extendible end blocks

By Lemma 8 we can reduce an instance I to its successors for all end blocks. The problem in this approach is that by Lemma 3, I potentially has infinitely many successors. We shall next show that also the extendible end blocks reduce to a finite number of successors.

Let $I = (\#, h, g, (s_1, s_2))$ and that assume that we have the successors

$$I'(xu^k w, yv^k) = I'_k \text{ for } k \geq 0 \tag{5}$$

with morphisms $h', g' \colon (A')^* \to A^*$ as defined for the successor of (h, g). Successors $I'(xu^k, yv^k w)$ are treated analogously.

Lemma 9. *Let I'_k, for $k \geq 0$, be as in the above and $|u| = |v|$. Then there are only finitely many distinct successors I'_k and they can be effectively found.*

Proof. A successor with respect to $(xu^k w, yv^k)$ is defined only if either $yv^k \preccurlyeq xu^k w$ or $xu^k w \preccurlyeq yv^k$. In the first case, $s'_1 = xu^k w (yv^k)^{-1}$ and $s'_2 = \varepsilon$. Let ℓ be the least number such that $|yv^\ell| > |w|$ and $yv^\ell \preccurlyeq xu^\ell w$. If there is no such ℓ, then there is only finitely many possible k such that $yv^k \preccurlyeq xu^k w$. If $yv^k \preccurlyeq xu^k w$ for some $k > \ell$, then however $s'_1 = xu^k w (yv^k)^{-1} = xu^\ell w (yv^\ell)^{-1}$ and $s'_2 = \varepsilon$. The second case is similar. □ □

Lemma 10. *Assume that the successors I'_k are as in* (5) *and that $|u| \neq |v|$. Then there are only finitely many k such that I'_k can have a solution and these numbers k can be effectively found.*

Proof. Notice first that if $b \leq xu$ and $c \leq yv$, then $g(c)$ begins with b, since $h(xu)$ and $g(yv)$ are comparable. It follows that no $h'(w_1)$ begins with the letter b and no $g'(w_2)$ begins with the letter c. Namely if $h'(b) = bz_1$ and $g'(d) = cz_2$ for some words z_1 and z_2, then by the definition of h' and g', bz_1 and cz_2 are some block words. Both $h(bz_1)$ and $g(cz_2)$ begin with b, so (bz_1, cz_2) would be the block for b, but a block for b cannot exists, because there is an extendible end block for b.

To prove the claim it suffices to show that for all $f \in A'$, there are only finitely many k such that I'_k has an end block for f. For that purpose, assume that I'_k has an end block beginning with $f \in A'$. Recall that I'_k is defined only if $yv^k \preccurlyeq xu^k w$ or $xu^k w \preccurlyeq yv^k$. In the first case, s'_1 is a prefix of $xu^k w$ and $s'_2 = \varepsilon$. In the second case $s'_1 = \varepsilon$ and s'_2 is a prefix of yv^k. Consider now the second case, the first one is symmetric. If $|u| > |v|$, then $xu^k w \preccurlyeq yv^k$ can be true only for finitely many k, so we can assume $|u| < |v|$ as well. As in the proof of Lemma 3, the end block for f is constructed either by findind a word u' satisfying $h'(u')xu^k w = yv^k$ or constructing the sequence $(u_i, v_i) \in (A')^* \times (A')^*$, where $h'(u_i)$ and $g'(v_i)$ are comparable. Consider now the latter case, the final argument concerning u' will also apply to the former case. Because this construction was assumed to yield an end block for f, we must have

$$h'(u_i) = g'(v_i) z_i \tag{6}$$

for some i and $z_i \leq yv^k$. Now z_i and yv^k both begin with c. But no $g'(d)$ begins with c, so (u_{i+1}, v_{i+1}) cannot be defined any more. Hence i and u_i are unique and independent of k. By assumption, this construction leads to the ending block for f, so there is a word u' such that $h'(u_i u') x u^k w = g'(v_i) y v^k$.

We will finally show that the word u' can exist only for finitely many k from which the claim follows. Also the word u' is constructed letter by letter defining $u'_0 = \varepsilon$ and $u'_{j+1} = u'_j a$, where a is the letter for which the words $h'(u_i) h'(u'_j a)$ and $g'(v_i) y^k v$ are comparable. Because h' is marked, such a letter a is always unique, if exists. But the sequence u'_j eventually stops giving u', since $xu^k w$ is a suffix of yv^k that begins with b and no $h'(w)$ begins with b. This means that if u' exists for some k, it cannot exist for larger values of k. Notice also that we can effectively check wheter or not there is a k such that u' exists, i.e., if the sequence u'_j stops. □ □

Let

$$S' = \{(s'_1, s'_2) \mid (\#, h', g', (s'_1, s'_2)) \text{ is a successor of } I\}$$

be the set of the end words of the successor of I. We introduce a collective notation

$$\mathcal{I}' = (\#, h', g', S') \tag{7}$$

to stand for all the successors of I. By Lemma 8, I has a solution if and only if one of the successors (7) has, but by Lemmata 9 and 10, we can always assume that S' is a finite set. Thus we obtain a chain of sets of successors reducing the original instance $I_0 = (\#, h_0, g_0, (s_1, s_2))$ to its successors $\mathcal{I}_1 = (\#, h_1, g_1, S_1)$, then all these to get $\mathcal{I}_2 = (\#, h_2, g_2, S_2)$, etc. If eventually some successors $\mathcal{I}_i = (\#, h_i, g_i, S_i)$ have very simple morphisms (h_i, g_i) (i.e. $|A_i| = 1$ or the suffix complexity is zero), we can decide for each $(\#, h_i, g_i, S_i)$ if there is a solution or not. Thus we could also solve the original problem. Otherwise, we know by Lemma 7 that there is a number n_0 such that $(h_{i+d}, g_{i+d}) = (h_i, g_i)$ for each $i \geq n_0$, i.e. the morphisms start so cycle. Clearly to decide the marked GPCP it suffices to show how to solve these cycling instances. This is done in the next section.

In our solution we do not construct the successors for the end blocks in $E_\#$, but check at each step whether there is $(u, v) \in E_\#$ such that $u = v$.

Lemma 11. *We can effectively check whether there is an end block $(u, v) \in E_\#$ such that $u = v$.*

Proof. First of all, we can build the end blocks starting with # as for any other letter. We have two special cases, (1) we get an extendible end block and (2) we are in an instance with an extendible end block. Otherwise we have finitely many end blocks of finite length. But it is clear that also these special cases can be decided, using the methods in the proofs of Lemmata 9 and 10. □ □

By a *successor sequence* we mean a sequence

$$(\#, h_0, g_0, (s_1^{(0)}, s_2^{(0)})), \dots, (\#, h_i, g_i, (s_1^{(i)}, s_2^{(i)})), \dots$$

of instances of the marked GPCP such that each $(\#, h_{i+1}, g_{i+1}, (s_1^{(i+1)}, s_2^{(i+1)}))$ is a successor of $(\#, h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$.

To end this section, notice that if $\mathcal{I}_i = (\#, h_i, g_i, S_i)$ is the set of all ith members in the successor sequences, we can always assume that

(A) morphism pair (h_i, g_i) has a block for the letter # and
(B) $s_1 s_2 \neq \varepsilon$ for each $(s_1, s_2) \in S_i$.

For if the condition (A) does not hold, we know that no instance in $\mathcal{I}_i$ can have a solution beginning with # and if (B) is not satisfied by an instance, then that instance reduces to the marked PCP, which is decidable by [2].

5 Cycling instances

In this section we show how to treat the instances that begin to cycle, i.e. there exists a d such that for all successor sequences

$$(\#, h_0, g_0, (s_1^{(0)}, s_2^{(0)})), \dots, (\#, h_i, g_i, (s_1^{(i)}, s_2^{(i)})), \dots \tag{8}$$

$(h_i, g_i) = (h_{i+d}, g_{i+d})$ for all $i \geq 0$. We shall call such an instance I_0 a *loop instance*, and d is the *length of the loop.*

Lemma 12. *The sequence $I_i = (h_i, g_i)$ has the following properties.*
(i) The size of the alphabet is constant and $\sigma(I_i) = \sigma(I_0)$ for all $i \geq 0$.
(ii) An instance (h_0, g_0) of the marked PCP has a solution beginning with a if and only if $h_i(a) = a = g_i(a)$ for all $i \geq 0$.

Proof. Both of these cases follow from [2]. □ □

Notice that because the alphabet size does not decrease, there is a block for each letter $a \in A$. In particular, there cannot be extendible end blocks.

Corollary 1. *Assume that the instances cycle as in* (8) *and that a solution exists. Then we have two cases:*
(i) If $h_0(\#) = \# = g_0(\#)$ then the minimal solution of I_0 is $\#w$, where w the initial letter a of w satisfies $h_0(a) \neq g_0(a)$. Hence also $h_i(a) \neq g_i(a)$ for all $i \geq 0$.

(ii) If $h_0(\#) \neq g_0(\#)$, then the minimal solution beginning with $\#$ does not have a solution of the PCP as a prefix.

Hereafter we will assume that $h_0(\#) \neq g_0(\#)$, since the case (i) reduces to the following problem for each a such that $h_0(a) = g_0(a)$: Does the cycling instance I_0 have a solution beginning with a? But this is essentially the original problem, $\#$ replaced with a.

We would like to have some upper bound for the lengths of the new end blocks in the loop (8). We demonstrate that if a solution exists, there is a limit number L such that, if a solution exists, then the minimal solution is found in some sequence (8) shorter than L. Moreover, this limit can be effectively found, hence the main result follows.

In what follows, we assume that $I = (\#, h, g, (s_1, s_2))$ has a minimal solution beginning with $\#$ and which does not have a solution of the PCP as prefix. Then this minimal solution is unique, since the morphisms are marked. Consequently each I has a unique end block (u, v) in the block decomposition of the minimal solution. It follows that there exists a unique successor sequence $I_0, I_1, \ldots$ of instances such that

$$I_{i+1} = I_i(u_i, v_i). \tag{9}$$

and (u_i, v_i) is the end block of the minimal solution (beginning with $\#$) of I_i. This successor sequence is called *the branch of minimal solutions.* Note that we cannot determine, which is the end block of the minimal solution, but the desired limit will be obtained anyway.

Let $I_i = (\#, h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$ be an instace of the branch of the minimal solutions and w_i the minimal solution of I_i. Recall that we permanently assume that $s_1^{(i)} s_2^{(i)} \neq \varepsilon$ and also that $g_0(\#) \neq h_0(\#)$, which implies that also $g_i(\#) \neq h_i(\#)$ for each i.

Lemma 13. *Let w_i be the minimal solution of I_i beginning with $\#$ and $(h_i, g_i) = (h_{i+d}, g_{i+d})$ for each i. Then $w_{i+d} \leq w_i$ but $w_{i+d} \neq w_i$ for each i.*

Proof. The instances

$$I_d = (\#, h_i, g_i, (s_1^{(i)}, s_2^{(i)})) \text{ and } I_{i+d} = (\#, h_i, g_i, (s_1^{(i+d)}, s_2^{(i+d)}))$$

share the marked morphisms, so clearly $w_i \leq w_{i+d}$ or $w_{i+d} \leq w_i$, since the minimal solutions cannot have a solution of PCP as a prefix (recall that $h_i(\#) \neq g_i(\#)$). If w is a minimal solution to some instance I, then by Lemma 8, there is a solution w' to the successor of I such that $w = h'(w')u = g'(w')v$. Notice that w and w' begin with the same letter. Since $s_1^{(i)} s_2^{(i)} \neq \varepsilon$, then also $uv \neq \varepsilon$ and consequently $|w| > |w'|$, because the morphisms are nonerasing. Hence $|w_{i+1}| + 1 \leq |w_i|$ Inductively $|w_{i+t}| + t \leq |w_i|$ for all t. This proves the claim. □ □

As a byproduct we obtain

Lemma 14. *If an instance occurs twice in a successor sequence, it has no solutions.*

Proof. By the proof of the previous lemma, the length of the minimal solution decreases strictly. □ □

An end block (u, v) of an instance $I = (\#, h, g, (s, \varepsilon))$ satisfies the equation

$$h(u)s = g(v).$$

If this is an end block of a solution, then necessarily $u = s'v$ or $v = s'u$ for some word s', and I', the successor of I has the end words (s', ε) or (ε, s'), respectively.

Lemma 15. *Let $I_i = (\#, h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$ be the branch of the minimal solutions of a cycling instance with loop length d. Let also w_i be the minimal solution of I_i. Then $h_i(w_{i+d})s_1^{(i+d)} = g_i(w_{i+d})s_2^{(i+d)}$ is a prefix of $h_i(w_i)$ and $g_i(w_i)$.*

Proof. It suffices to take $i = 0$, the proof is analogous for all other values. Recall also that $s_1^{(t)}s_2^{(t)} \neq \varepsilon$ for each t. By Lemma 13, $w_d \leq w_0$. Therefore $h_0(w_d) \leq h_0(w_0)$ and $g_0(w_d) \leq g_0(w_0)$. We shall next prove that $|h_0(w_d)s_1^{(d)}| \leq |h_0(w_0)|$. Assume on the contrary that $|h_0(w_d)s_1^{(d)}| > |h_0(w_0)|$. By Lemma 8,

$$w_0 = h_1(h_2(\dots h_{d-1}(h_d(w_d)u_{d-1})\dots u_2)u_1$$

and therefore

$$|h_0(w_d)s_1^{(d)}| > |h_0(w_0)| \geq |h_0(w_d)| + \sum_{i=1}^{d-1} |u_i|.$$

Hence

$$|s_1^{(d)}| > \sum_{i=1}^{d-1} |u_i|.$$

This is a contradiction, since $|s_1^{(d)}| \leq |u_{d-1}|$. It follows that $|h_0(w_d)s_1^{(d)}| \leq |h_0(w_0)|$ and similarly we can prove that $|g_0(w_d)s_2^{(d)}| \leq |g_0(w_0)|$. Without loss of generality, we can assume that $s_2^{(d)} = \varepsilon$. Then

$$h_0(w_d)s_1^{(d)} = g_0(w_d)s_2^{(d)} = g_0(w_d) \leq g_0(w_0),$$

and since $h_0(w_0)$ and $g_0(w_0)$ are comparable and $|h_0(w_d)s_1^{(d)}| \leq |h_0(w_0)|$, necessarily $h_0(w_d)s_1^{(d)} \leq h_0(w_0)$, too. □ □

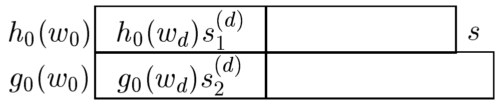

Figure 2: Prefix property

The previous lemma is used in the proof our last lemma, which gives an upper bound for the size of the end blocks in the branch of the minimal solutions.

For an occurrence of a word u in $g(w)$, its *g-block covering* in a solution w of an instance $(\#, h, g, (s_1, s_2))$ is a word $z = g(v_1)g(v_2)\ldots g(v_k)$ such that

1. $v_1v_2\ldots v_k$ is a factor of w,
2. u is a factor of z,
3. u is not a factor of $g(v_2)\ldots g(v_k)$ or $g(v_1)\ldots g(v_{k-1})$,
4. for each i, $g(v_i) = h(u_i)$ is a block for morphism pair (h, g).

Note that a g-block covering for an occurrence of a factor u (in $g(w)$) is unique. Hence we can define the integer k to be the *g-covering length* of the occurrence of u (in w).

Lemma 16. *Let $I_i = (\#, h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$ be the branch of minimal solutions of a cycling instance having loop length d. For all $i \geq d$*

(i) If $s_1^{(i)} \neq \varepsilon$, then the h_{i+1}-covering lengths of $s_1^{(i+1)}$ and $s_2^{(i+1)}$ are at most the g_i-covering length of $s_1^{(i)}$.

(ii) If $s_2^{(i)} \neq \varepsilon$ then the g_{i+1}-covering lengths of $s_1^{(i+1)}$ and $s_2^{(i+1)}$ are at most the h_i-covering length of $s_2^{(i)}$

Proof. By Lemma 15, for $i \geq d$, the words $s_1^{(i)}$ and $s_2^{(i)}$ are covered by some $h_i(w_1)$ and by some $g_i(w_2)$.

We will prove only the case (i), the other one is analogous. To simplify the notations, we denote $I = I_i = (\#, h, g, (s, \varepsilon))$ and $I_{i+1} = I'$. Now either $I' = (\#, h', g', (s', \varepsilon))$ or $I' = (\#, h', g', (s', \varepsilon))$. We have to show that in both cases, the h'-covering length of s' is at most the g-covering length of s.

Assume that (u, v) is the end block of the minimal solution w of I. Let also $a \leq u$ be the first letter of u. Then $h(u)s = g(v)$ and $u \preccurlyeq v$ or $u \preccurlyeq v$, which gives us two cases to consider.

(1) If $u = s'v$, then $|s'| \leq |u|$. But words u and v in the equation $h(u)s = g(v)$ are obtained during the block construction for a letter a. Because there also is a block for letter a, necessarily $s' \leq h'(a)$, i.e. the h'-covering length of s' is 1 (see Figure 3).

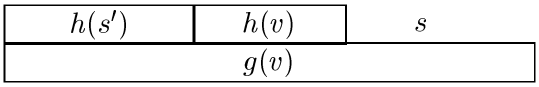

Figure 3: Picture of the case (1)

(2) Assume then that $v = s'u$, so $I' = (\#, h', g', (\varepsilon, s'))$. We observe first that the g-covering length of the word $g(s')$ is at most that of the word s. This is clear, because $g(s')$ shares with s every one of its block factors $g(v_i)$ (including the first one, since as in the case (1), $h(u)$ is covered by a single g-block longer than $h(u)$ (see Figure 4 for illustration)). We show then that

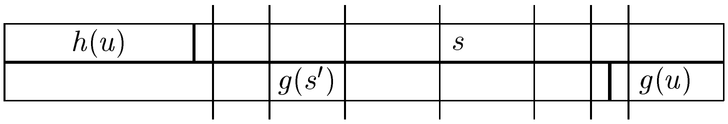

Figure 4: Block covering of $h(u)s = g(s')g(u)$. The vertical lines illustrate the block covering.

the h'-covering of s' is not longer than the g-covering of $g(s')$, from which the claim follows. Let w' be the minimal solution of I'. Then the word $w = h'(w') = g'(w')s'$ satisfies

$$h(w)g(s') = hh'(w')g(s') = gg'(w')g(s') = g(w),$$

and consequently w is a prefix of the minimal solution of I. To show that the h'-covering of s' is not longer than g-covering of $g(s')$, it is sufficient to show that the block borderlines in $h'(w') = g'(w')s'$ cutting s' can be mapped injectively to block borderlines in $h(w)g(s') = g(w)$ that cut $g(s')$ (See Figure 5). Let now $y' \leq w'$ be a word that determines a block borderline in $h'(w') = g(w')s'$ that cuts s'. That is, $h'(y') = g'(z')$ for some word z' and $g'(w') \leq h'(y')$. Then $z' = w'x'$ for some x' that satisfies $g'(x') \leq s'$. Now the word $y = g'(z')$ is a prefix of w, since

$$g(y) = g(g'(z')) = g(g'(w')g'(x')) \leq g(g'(w')s') = g(w)$$

and g is marked. But y also determines a block borderline in word $g(w) = h(w)g(s')$, since $g(y) = g(g'(z)) = h(h'(z))$. This borderline cuts s, because

$$g(y) = g(g'(z')) = h(h'(z')) = h(h'(w'))h(h'(x')) = h(w)h(h'(x')),$$

and hence $h(w) \leq g(y)$. Notice finally that the word y determines z' uniquely since g is injective and z' determines y' by $g'(z') = h'(y')$ (h is also injective). □

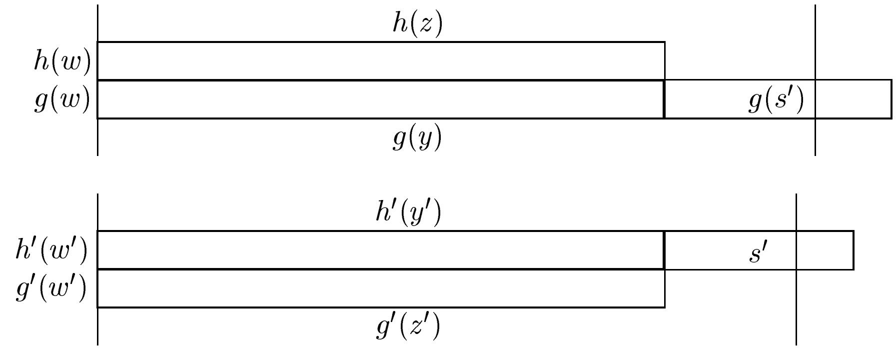

Figure 5: Relation between $g(s')$ and s'

□

The previous lemma gives us a tool for recognizing instances which *are not* in the branch of minimal solutions. Let I_0 be a cycling instance with loop length d and consider *all* the instances $\mathcal{I}_d$ found by the first d reductions. If I_0 has a solution then there is a unique $I \in \mathcal{I}_d$ in the branch of the minimal solutions.

Let M be the maximal g- or h-covering length of all the end words s_1 and s_2 in $\mathcal{I}_d$. It now follows by Lemma 16 that in the branch of the minimal solutions the g_i or h_i-covering length is always less than or equal to M. For a sequence of cycling instances, the suffix complexity is constant $\sigma(I_0)$ and since the blocks of an instance I_i are the images of the successor I_{i+1}, the block length can never be more than $\sigma(I_0) + 1$. By the previous lemma we have

Corollary 2. *Let $I_0, \ldots, I_i, \ldots$ be the branch of the minimal solutions of a cycling instance with loop length d. For each $i \geq d$, the end words of I_i are not longer $M(\sigma(I) + 1)$.*

Now we are ready to prove the Theorem 1.

Proof of Theorem 1. It remains to be shown how to solve the marked GPCP for the cycling instances I_0. A cycling instance has the blocks for all the letters. In particular, there are no extendible end blocks and only finitely many successors. The successor relation naturally defines a tree $\mathcal{T}$ having I_0 as the root, all the successors of I_0 as the vertices and the pairs (I, I') as the edges.

The decision procedure is based on constructing $\mathcal{T}$ partially by first inserting the vertices having depth (the distance from the root) at most d and then computing the number M, the maximal covering length of the end words of instances at the depth d. For all vertices $I = (\#, h, g, (s_1, s_2))$ that have $s_1 s_2 = \varepsilon$, we can always decide if they have a solution or not, and at each step we check whether there is an end block $(u, v) \in E_\#$ such that $u = v$ as in Lemma 11. If some such vertex I has no solution, then I and all the successors of I can be removed. On the other hand, if some such I has a solution, then I_0 also has a solution and the procedure may stop.

For the vertices having depth greater than d, the (partial) construction of $\mathcal{T}$ is more specific: Only the successors $I = (\#, h, g, (s_1, s_2))$ that satisfy $|s_1 s_2| \leq M(\sigma(I_0) + 1)$ are inserted. By Corollary 2, the branch of minimal solutions is included in the partial construction.

But now there are only finitely many instances to be inserted, so each path (successor sequence) in the partially constructed $\mathcal{T}$ will eventually contain an instance twice, thus I_0 has no solution by Lemma 14, unless some vertex $I = (\#, h, g, (\varepsilon, \varepsilon))$ has. $\square$

In [1] Ehrenfeucht, Karhumäki and Rozenberg proved that the binary PCP, i.e. for the instances (h, g), where $h, g \colon A^* \to B^*$ and $|A| = 2$, is decidable if and only if the binary GPCP is decidable for marked morphisms. Therefore Theorem 1 has the following corollary.

Corollary 3. *The binary PCP is decidable.*

References

[1] A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg. *The (generalized) Post correspondence problem with lists consisting of two words is decidable.* Theoretical Computer Science, 21(2):119–144, 1982.

[2] V. Halava, M. Hirvensalo and R. de Wolf. *Decidability and Undecidability of Marked PCP.* STACS'99 (C. Meinel and S. Tison, eds.), Lecture Notes in Comput. Sci, vol 1563, Springer-Verlag, 1999, pp. 207-216.

[3] T. Harju and J. Karhumäki. *Morphisms.* In Handbook of Formal Languages, volume 1, 439–510. edited by G. Rozenberg and A. Salomaa, eds. Springer-Verlag, Berlin, 1997.

[4] Y. Matiyasevich and G. Sénizergues. *Decision problems for semi-Thue systems with a few rules.* In Proceedings of the 11th IEEE Symposium on Logic in Computer Science, pages 523–531, 1996.

[5] E. L. Post. *A variant of a recursively unsolvable problem.* Bulletin of the American Mathematical Society, 52:264–268, 1946.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**