ELSEVIER

Theoretical Computer Science 276 (2002) 183–204

Theoretical Computer Science

www.elsevier.com/locate/tcs

Binary (generalized) Post Correspondence Problem

Vesa Halava[a,b,*], Tero Harju[a], Mika Hirvensalo[a,b,1]

[a] *Department of Mathematics, University of Turku, FIN-20014 Turku, Finland*
[b] *TUCS-Turku Centre for Computer Science, Lemminkäisenkatu 14 A, FIN-20520, Turku, Finland*

Received August 2000; received in revised form December 2000; accepted February 2001
Communicated by A. Salomaa

Abstract

We give a new proof for the decidability of the binary Post Correspondence Problem (PCP) originally proved in 1982 by Ehrenfeucht, Karhumäki and Rozenberg. Our proof is complete and somewhat shorter than the original proof although we use the same basic idea. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Post Correspondence Problem; Binary Post Correspondence Problem; Generalised Post Correspondence Problem; Marked Morphisms; Decidability

1. Introduction

Let A and B be two finite alphabets and h, g be two morphisms $h, g : A^* \rightarrow B^*$. 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 [8] that this problem is undecidable in general. Such a word w that $h(w) = g(w)$ is called a *solution* of the *instance* (h, g) of the PCP.

In the *binary* PCP we assume that the *size* of the instance (h, g) is two i.e., $|A| = 2$. This problem was proved to be decidable by Ehrenfeucht et al. [2]. Here we shall give a new shorter proof to this binary case, although we use the same basic idea as [2]. Our proofs are combined from [4, 3], and we have added details to the proof to make it easier to read. Also, although we restrict to the binary PCP, we shall achieve more information than really needed for the binary case.

[*] Corresponding author. Department of Mathematics, University of Turku, FIN-20014, Turku, Finland. Tel.: +358-2-333-6675; fax: +358-2-241-6595.

E-mail addresses: vehalava@utu.fi (V. Halava), harju@utu.fi (T. Harju), mikhirve@utu.fi (M. Hirvensalo).

[1] Supported by the Academy of Finland under grant 44087.

PII: S0304-3975(01)00157-8

Note that it is also known that if $|A|\geqslant 7$, then the PCP remains undecidable, see [7]. The decidability status is open for $3\leqslant |A|\leqslant 6$.

Another important problem is the *generalized* PCP, GPCP for short. It consists of two morphisms $h,g : A^* \to B^*$ and words $p_1, p_2, s_1, s_2 \in B^*$. The GPCP is to tell whether or not there exists a nonempty word $w\in A^*$ such that

$$p_1h(w)s_1 = p_2g(w)s_2.$$

Here again w is called a *solution*. We shall denote the instance of the GPCP by $((p_1, p_2), h, g, (s_1, s_2))$. The pair (p_1, p_2) is called the *begin words* and (s_1, s_2) is called the *end words*. Note that also for the GPCP it is known that it is decidable, if $|A|\leqslant 2$, see [2], and undecidable, if $|A|\geqslant 7$, see [6]. As for the PCP, the decidability status of the GPCP is open for the alphabet size between these two bounds.

The basic idea in [2] is that each instance (h, g) of the binary PCP is either

(1) *periodic*, i.e., $h(A^*)\subseteq u^*$, where $u\in B^*$, or
(2) it can be reduced to an equivalent instance of the binary generalized PCP with marked morphisms,

and then it is proved that both of these two cases are decidable. Recall that a morphism h is called *marked* if the images of all letters begin with a different letter, i.e., $h(x)$ and $h(y)$ start with a different letter whenever $x, y\in A$ and $x\neq y$.

For the decidability of the periodic case, see [2, 5]. We shall also present a proof in the next section. The *binary* GPCP was shown to be decidable for marked morphisms in [2]. This proof is by case analysis and it is rather long. We shall give here a new proof, which follows the lines of [3], where it was shown that the GPCP is decidable for marked morphisms with any alphabet size. Since here we shall concentrate only on the binary case, the decidability proof becomes more elementary and shorter than that in [3].

Our proof for the decidability of the marked binary GPCP uses the idea of reducing a problem instance to finitely many new instances such that at least one of these new instances has a solution if and only if the original one has. Then by iterating this reduction we shall finally get to (finitely many) new instances, where the decision is easy to do.

Note that in the PCP and GPCP we may always assume that the image alphabet B is binary, since any B can be injectively encoded to $\{0,1\}^*$. For example, if $B=\{b_1, b_2, \ldots, b_m\}$, then $\varphi : B\to \{0,1\}^*$, where

$$\varphi(b_i) = 01^i \quad \text{for all } 1 \leqslant i \leqslant m,$$

is such an encoding. Therefore, in the binary case we shall assume that $A=B=\{0,1\}$.

We shall first fix 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 $x\leqslant y$. A prefix of length k of y is denoted by $\mathrm{pref}_k(y)$. Also, if $x\neq\varepsilon$ and $z\neq\varepsilon$ in $y=xz$, then x is a *proper* prefix of y, and, as usual, this is denoted by $x<y$. We say that x and y are *comparable* if $x\leqslant y$ or $y\leqslant x$.

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 $x \preccurlyeq y$ and, if $x \neq \varepsilon$ and $z \neq \varepsilon$, then x is called a *proper* suffix of y, denoted by $x \prec y$.

If $x = yz$ then we also denote that $y = xz^{-1}$ and $z = y^{-1}x$.

2. The periodic case

We shall begin with the easier part of the solution and consider first the instances of the (binary) PCP, where one of the morphisms is periodic. To prove this result we shall need lemma, which states a property of the *one counter languages* or context-free languages, see [1].

Lemma 1. *Let $\rho : A^* \to \mathbb{Z}$ be a monoid morphism into the additive group of integers and let $R \subseteq A^*$ be a regular language. It is decidable whether $\rho^{-1}(0) \cap R \neq \emptyset$.*

Proof. Here the language $\rho^{-1}(0)$ is a one counter language and one counter languages are closed under the intersection with regular languages. The emptiness problem is decidable for one counter languages and even for context-free languages, see for example [9]. □

The proof of the next theorem is from [5], see also [2].

Theorem 1. *PCP is decidable for instances (h, g), where h is periodic.*

Proof. Let $h, g : A^* \to B^*$ and assume that h is periodic and $h(A^*) \subseteq u^*$ for a word $u \in B^*$. Define a morphisms ρ by

$$\rho(a) = |h(a)| - |g(a)|$$

for all $a \in A$. Define a regular set $R = g^{-1}(u^*) \setminus \{\varepsilon\}$. Now

$$\rho^{-1}(0) = \{v \mid |h(v)| = |g(v)|\}$$

and $w \in \rho^{-1}(0) \cap R$ if and only if $w \neq \varepsilon$, $g(w) \in u^*$ and $|g(w)| = |h(w)|$. In other words we have $g(w) = h(w)$ for some $w \neq \varepsilon$ if and only if $\rho^{-1}(0) \cap R \neq \emptyset$. By Lemma 1 the latter property is decidable and therefore the claim follows. □

Note that the above proof holds for all alphabet sizes, not only for the binary case.

3. From PCP to GPCP

Let $h : \{0,1\}^* \to \{0,1\}^*$ be a morphism that is not periodic. Define the mapping $h^{(1)}$ by

$$h^{(1)}(x) = \mathrm{pref}_1(h(x))^{-1} h(x)\, \mathrm{pref}_1(h(x)) \quad \text{for } x = 0, 1.$$

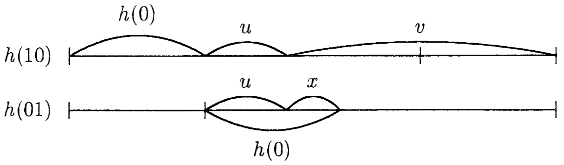

Fig. 1. Case $k=1$, $m<|h(1)|$.

In other words the images of $h^{(1)}$ are the cyclic shifts of the images of h. Now define recursively $h^{(i+1)}=(h^{(i)})^{(1)}$. Clearly

$$h^{(i)}(x) = \mathrm{pref}_j(h(x))^{-1}h(x)\,\mathrm{pref}_j(h(x)),$$

where $j\leqslant |h(x)|$ and $j\equiv i \pmod{|h(x)|}$.

For any two words $u,v\in A^*$ it is well known that $uv=vu$ if and only if u and v are powers of a common word. It follows from this that the maximal common prefix of $h(01)$ and $h(10)$ has length at most $|h(01)|-1$.

Lemma 2. *Let z_h be the maximal common prefix of $h(01)$ and $h(10)$ and $m=|z_h|$. Then $h^{(m)}$ is a marked morphisms and $h^{(m)}(w)=z_h^{-1}h(w)z_h$, for all $w\in\{0,1\}^*$. Moreover, for any w, if $|h(w)|\geqslant m$, then $z_h\leqslant h(w)$.*

Proof. We may assume by symmetry that $|h(1)|\geqslant|h(0)|$. Assume first that $m<|h(0)|$. Then, clearly, $h^{(m)}(0)$ and $h^{(m)}(1)$ begin with different letters by the maximality of the z_h.

If $m\geqslant|h(0)|$, then $h(10)=h(0)^kuv$ for some $k>0$, $u,v\in\{0,1\}^*$ and $m=|h(0)^ku|$. And if $|uv|\geqslant|h(0)|$, then $h(01)=h(0)^kuxw$, where $ux=h(0)$ and $w\in\{0,1\}^*$, otherwise $ux\leqslant h(0)$ and $w=\varepsilon$.

Since $h(1)=h(0)^kuvh(0)^{-1}$, it follows that $h^{(m)}(1)=vh(0)^{-1}h(0)^ku=vh(0)^{k-1}u$. Now $h^{(m)}(0)=xzu$, where $z=\varepsilon$, if $|uv|\geqslant|h(0)|$, and since v and x begin with different letters, $h^{(m)}$ is marked, see also Fig. 1.

Finally we deduce that

$$h^{(m)}(0) = xzu = (h(0)^ku)^{-1}h(0)h(0)^ku = z_h^{-1}h(0)z_h$$

and

$$h^{(m)}(1) = vh(0)^{k-1}u = (h(0)^ku)^{-1}h(0)^kuvh(0)^{-1}h(0)^ku = z_h^{-1}h(1)z_h.$$

Therefore for all $w\in A^*$, $h^{(m)}(w)=z_h^{-1}h(w)z_h$, and the last part of the claim follows directly from this. □

Note that if h is already marked, then $z_h=\varepsilon$.

Let (h,g), where $h,g:\{0,1\}^*\to\{0,1\}^*$, be an instance of the binary PCP. Assume further that h and g are nonperiodic. Let z_h be as above, $m=|z_h|$ and $n=|z_g|$. We may assume by symmetry that $m\geqslant n$. We now have the following lemma.

Lemma 3. *The instance (h,g) of the binary PCP has a solution if and only if $z_g \leqslant z_h$ and the instance $((z_g^{-1}z_h,\varepsilon),h^{(m)},g^{(n)},(\varepsilon,z_g^{-1}z_h))$ of the binary GPCP has a solution.*

Proof. It is obvious that if an instance (h,g) of the PCP has a solution, then $z_g \leqslant z_h$. This can be seen if we assume that w is solution such that $n,m \leqslant |h(w)|$, then $z_h \leqslant h(w)$ and $z_g \leqslant g(w)$.

Assume first that the instance of the GPCP has a solution w, i.e.,

$$z_g^{-1}z_h h^{(m)}(w) = g^{(n)}(w)z_g^{-1}z_h$$

and therefore

$$z_g^{-1}h(w)z_h = z_g^{-1}g(w)z_h.$$

This is true if and only if

$$h(w) = g(w).$$

Assume then that (h,g) has a solution w. Since $h^{(m)}$ and $g^{(n)}$ are morphisms, we get that

$$h(w) = z_h(z_h^{-1}h(w)z_h)z_h^{-1} = z_g(z_g^{-1}g(w)z_g)z_g^{-1} = g(w)$$

and therefore

$$z_h h^{(m)}(w)z_h^{-1} = z_g g^{(n)}(w)z_g^{-1}.$$

This is true if and only if

$$(z_g^{-1}z_h)h^{(m)}(w) = g^{(n)}(w)(z_g^{-1}z_h).$$

This proves the claim. □

4. Marked PCP

In this section we shall consider the solution method to the marked (binary) PCP. The proofs of the lemmata in this section are from [4], and we shall prove the results for all alphabet sizes.

A *block* of an instance $I=(h,g)$, where $h,g: A^* \to B^*$, of the marked PCP is a pair $(u,v) \in A^+ \times A^+$ such that $h(u)=g(v)$ and for all nonempty prefixes $u_1 \leqslant u$, $v_1 \leqslant v$, $h(u_1)=g(v_1)$ 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. A letter $b \in B$ is a *block letter* if there is a block (u,v) such that $b \leqslant h(u)$ and $b \leqslant g(v)$. In other words, b is the first letter of the images of a block. Accordingly, a block is a minimal nontrivial solution of the equation $h(x)=g(y)$.

Lemma 4. *Let (h,g) be an instance of the marked PCP for $h,g:A^* \to B^*$. Then for each letter $a \in A$, there exists at most one block (u,v) such that $a \leqslant u$. In particular, the instance (h,g) has at most $|A|$ blocks. Moreover, the blocks of (h,g) can be effectively found.*

Proof. Consider any pair (u,v) of words such that $h(u)$ and $g(v)$ are comparable and $h(u) \neq g(v)$. Since h and g are marked, there exists a unique $a \in A$ such that $h(ua)$ and $g(v)$ or $h(u)$ and $g(va)$ are comparable if $h(u) < g(v)$ or $g(v) < h(u)$, respectively. Since the morphisms are marked, it is clear that the first letter of u determines uniquely the first letter of v and the claim follows from this inductively.

The latter claim is evident, since $\{u \mid \exists v : h(u) = g(v)\}$ is a regular set. □

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

$$A' = \{b \in B \mid b \text{ is a block letter}\}. \tag{1}$$

Note that $|A'| \leqslant |A|$ although $A' \subseteq B$, since there are at most $|A|$ blocks by Lemma 4.

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

$$h'(a) = u \quad \text{and} \quad g'(a) = v, \tag{2}$$

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 $x \in (A')^*$.*

Proof. (i) This is clear since the different block words for h (and g) begin with different letters.

(ii) Assume that I has a solution w. Then w has two factorizations, $w = u_1 \ldots u_n = v_1 \ldots v_n$ such that $h(u_i) = g(v_i)$, i.e., (u_i,v_i) is a block for some letter $a_i \in A'$, for $i = 1,\ldots,n$. Then $w' = a_1 \ldots a_n$ is a solution for I', since $h'(w') = w = g'(w')$.

Assume that I' has a solution $w' = a_1 \ldots a_n$. Then there are blocks (u_i,v_i) for a_i, $i = 1,\ldots,n$. Now $h(u_1 \ldots u_n) = g(v_1 \ldots v_n)$ and $h'(w') = u_1 \ldots u_n = v_1 \ldots v_n = g'(w')$. Therefore, $u_1 \ldots u_n$ is a solution of I.

(iii) Let $x = x_1 \ldots x_k$, where $x_i \in A'$ for all $i = 1,\ldots,k$. By the definitions, for all x_i there exists a block (u_i,v_i) such that $h'(x_i) = u_i$, $g'(x_i) = v_i$ and $h(u_i) = g(v_i)$. Therefore, the claim follows. □

The definition of a successor gives inductively a sequence of instances I_i, where $I_0 = I$ and $I_{i+1} = I_i'$. Note that the reduction of an instance I to its successor I' was already used in [2], but the reduction was done only once. The difference here is that we shall iterate this reduction. The decidability of the marked PCP in [4] was eventually based on the fact that the successor sequence defined above has only finitely many distinct instances. In [4] two measures were used for an instance I of the marked PCP, namely the size of the alphabet and the *suffix complexity*:

$$\sigma(I) = \left|\bigcup_{a\in A}\{x \mid x \prec g(a)\}\right| + \left|\bigcup_{a\in A}\{x \mid x \prec h(a)\}\right|.$$

It is clear that for alphabet sizes of I' and I we have $|A'| \leqslant |A|$. Note that if we are studying the binary case, then we know that if the alphabet size decreases, then we get to the unary case, where the PCP becomes decidable. That $\sigma(I') \leqslant \sigma(I)$ is not so straightforward.

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

Proof. Let

$$G = \bigcup_{a\in A}\{x \mid x \prec g(a)\}, \qquad G' = \bigcup_{a\in A'}\{x \mid x \prec g'(a)\},$$

$$H = \bigcup_{a\in A}\{x \mid x \prec h(a)\}, \qquad H' = \bigcup_{a\in A'}\{x \mid x \prec h'(a)\}.$$

Let $s \in G'$. Then there exists at least one block (u, v), where $s \preccurlyeq v$. Let $v' = vs^{-1}$ and for some $u' \leqslant u$, $h(u') = g(v')z$ and $z \in H$.

Let $p: G' \to H$ be a function, where $p(s)$ is the z above with the minimal length. By the markedness this z is unique, since $z \leqslant g(s)$, and therefore p is an injective function. Similarly we can define an injective function from H' to G. The claim follows by the injectivity. □

The previous lemma together with $|A'| \leqslant |A|$ yields the following result.

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

Proof. We may assume that in I the alphabets are the same, i.e., $A = B$. Now for all $i \geqslant 0$ the alphabets are subsets of A. Since there are only finitely many morphisms from A^* to A^* with the lengths of images of letters under the bound $\sigma(I)$, we eventually get some instance I_{n_0} twice. The rest of the claim follows from the determinism of the successors. □

The previous lemma means that after n_0 consecutive successors the instances begin to cycle: $I_{n_0}, \ldots, I_{n_0+d} = I_{n_0}, \ldots$.

Lemma 8. *The sequence* $I_i = (h_i, g_i)$ *has the following properties.*

(i) *The size of the alphabet is constant and* $\sigma(I_i) = \sigma(I_{n_0})$ *for all* $i \geqslant n_0$.

(ii) *The instance* I_0 *of the marked PCP has a solution if and only if, for all* $i \geqslant n_0$, I_i *has a one letter solution.*

Proof. Case (i) follows from the definition of n_0.

For (ii), we may assume that $n_0 = 0$. By the proof of Lemma 5, case (ii), for every solution x_i to some I_i, there is a solution x_{i+1} to I_{i+1} such that $x_i = g_{i+1}(x_{i+1}) = h_{i+1}(x_{i+1})$. Suppose x_0 is a solution of a minimum length to I_0. Now by the above relation between the solutions, there is a solution x_d to I_d, where d is as in Lemma 7 such that

$$x_0 = g_1(x_1) = g_1g_2(x_2) = \cdots = g_1g_2\ldots g_d(x_d),$$

$$x_0 = h_1(x_1) = h_1h_2(x_2) = \cdots = h_1h_2\ldots h_d(x_d).$$

Since the g_i and h_i cannot be length-decreasing, we have $|x_0| \geqslant |x_d|$. But x_0 was chosen to be a minimum length solution and x_d is also a solution to $I_d = I_0$, and therefore necessarily $|x_0| = |x_d|$ and the morphisms $g_0(= g_d), \ldots, g_{d-1}, h_0(= h_d), \ldots, h_{d-1}$ map the letters occurring in x_d to letters. But then the first letter of x_d is already a solution to I_0 and, by the proof of Lemma 5, all instances in the loop have a one letter solution. This proves case (ii). □

Theorem 2. *The marked PCP is decidable.*

Proof. By constructing the successor sequence we will meet one of the following cases: (1) the alphabet size is one, (2) the suffix complexity goes to zero or (3) we have a cyclic sequence. The first two are easy to decide, and by Lemma 8 we can decide the third case by checking whether there is a solution of length one and the claim follows. □

Note that we can also decide, whether an instance of the marked PCP has a solution beginning with a fixed letter a, since we may map back the found one letter solutions as in the proof of Lemma 8 and check whether one of these begins with a.

5. Block structure in the marked binary GPCP

The instances

$$I = ((p_1, p_2), h, g, (s_1, s_2)) \tag{3}$$

of the (binary) GPCP 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 $p_1 \leqslant p_2$ or $p_2 \leqslant p_1$, and $s_1 \preccurlyeq s_2$ or $s_2 \preccurlyeq s_1$.

We shall extract the definition of the successor of the marked (binary) PCP to the marked (binary) GPCP. All definitions in this section apply for any alphabet size, not only for binary, therefore, we use arbitrary alphabet A as the domain alphabet, and whenever we consider the binary case, it shall be mentioned.

If the instance of the PCP is neither marked nor periodic, then we transform it to an instance of the marked GPCP as was done in Section 3.

Assume that we have an instance

$$((p_1, p_2), h, g, (s_1, s_2)),$$

where $h, g : A^* \to B^*$, $p_1 = \varepsilon$ or $p_2 = \varepsilon$ and $s_1 = \varepsilon$ or $s_2 = \varepsilon$.

For $b \in B$ we construct the blocks for (h, g) as in the case of the PCP. We shall also construct the so called *begin block* (x, y), where $p_1 h(x) = p_2 g(y)$ and there does not exists $r < x$ and $s < y$ such that $p_1 h(r) = p_2 g(s)$. The begin block is constructed as the blocks: we generate a sequence (x_i, y_i) such that $p_1 h(x_i)$ and $p_2 g(y_i)$ are comparable for all $i \geqslant 1$. The begin block is (x_i, y_i), for the minimal i such that $p_1 h(x_i) = p_2 g(y_i)$. Note that the begin block is unique if it exists and, if $p_1 = p_2 (= \varepsilon)$, then $x = y = \varepsilon$.

For the end words $s_1, s_2 \in B^*$ 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 $u_1 \leqslant u$ and $v_1 \leqslant v$. Let

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

be the set of all end blocks for the letter $a \in B$.

Lemma 9. *Let $I = ((p_1, p_2), h, g, (s_1, s_2))$ be an instance of the marked GPCP, $s_1 = \varepsilon$ or $s_2 = \varepsilon$ and a be 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*

$$\{(xu^k, yv^k w) \mid k \geqslant 0\} \quad \textit{and} \quad \{(xu^k w, yv^k) \mid k \geqslant 0\}$$

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: if $a \leqslant s_1$, then 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) such that $a \leqslant g(v_1)$, $u_i \leqslant u_{i+1}$, $v_i \leqslant v_{i+1}$ and that $h(u_i)$ and $g(v_i)$ are always comparable (as in Lemma 4). Whenever $h(u_i)z_i = g(v_i)$ for some $z_i \leqslant s_1$, 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 we may achieve an end block for several different i's.

If a is a block letter for a block (u,v), then always $u_i \leqslant u$ and $v_i \leqslant v$ 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 $z_i \leqslant s_1$. Claim (i) follows hereby.

By the above considerations, if E_a is infinite, then a is not a block letter, and the sequence (u_i,v_i) is infinite in order to get infinitely many possible z_i. This is possible only if there are words $x,y,u,v \in A^*$ such that $|h(u)| = |g(v)| \geqslant 1$ and a word $s \in B^*$ such that $h(x)s = g(y)$ and $h(xu)s = g(yv)$. Note that x, xy are prefixes of some u_i for large enough i and similarly y, yv are prefixes of v_i, and there are only finitely many different words s, since the morphisms are marked. Now for $i \geqslant |x| + |y|$ we have $(u_i,v_i) = (xu^k u', yv^k v')$, where $u' \leqslant u$ and $v' \leqslant v$. As above, any end block is of form $(u_i, v_i w_i)$. If $(xu', yv'w)$ and $(xuu', yvv'w')$ are end blocks for some $w,w' \in A^*$, then $w = w'$ and $(xu^k u', yv^k v'w)$ is an end block for all $k \geqslant 0$. Therefore, for $i \geqslant |x| + |y|$, 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 $v = v'v''$). 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 9(ii) *extendible end blocks*.

Let $I = ((p_1,p_2),h,g,(s_1,s_2))$ be an instance of the marked GPCP. For a solution $w \in A^*$, $p_1 h(w) s_1 = p_2 g(w) s_2$, of I,

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

is a *block decomposition* for w, if

(i) (u_1,v_1) is the begin block
(ii) (u_i,v_i) is a block for each $i = 2,3,\dots,k$,
(iii) (u_{k+1},v_{k+1}) is an (s_1,s_2)-end block.

Note that we have a special case when $k = 0$ in above. This means that there are no blocks and the begin and end blocks coincide. Then $p_1 h(w) s_1 = p_2 g(w) s_2$, but there does not exist any $u,v \leqslant w$ such that $p_1 h(u) = p_2 g(v)$.

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

Lemma 10. *Every solution $w \in A^*$ of I has a unique block decomposition.*

Note that, since in the block decomposition the k may be 0, it is necessary to construct also the set

$$E_p = \{(u,v) \mid p_1 h(u) s_1 = p_2 g(v) s_2 \text{ and if } (x,y) \text{ is the begin block,} \\ \text{then } u \leqslant x,\ v \leqslant y\}.$$

E_p is the set of end blocks for the pair (p_1,p_2). Clearly, if the begin block exists, then E_p is finite. Moreover, if the begin block does not exist, then E_p can be infinite as in Lemma 9. This case will be studied in Lemma 12.

$h(w)$	$p_1h(u_1)$	$h(u_2)$	$\cdots$	$h(u_k)$	$h(u_{k+1})s_1$
$g(w)$	$p_2g(v_1)$	$g(v_2)$	$\cdots$	$g(v_k)$	$g(v_{k+1})s_2$

Fig. 2. Block decomposition of a solution w.

Let $I=((p_1,p_2),h,g,(s_1,s_2))$ be an instance of the binary marked GPCP. In the binary case we have three choices for a solution:
(0) There are no blocks in the solution.
(1) Exactly one block is used in the solution.
(2) Two blocks are used in the solution.
Here the expression ‘used blocks’ mean the number of different blocks in the block decomposition (Fig. 2).

We shall use the next lemma to prove that the solutions of the type (0) and (1) can be effectively found.

Lemma 11. *Let $x,y,u,v,w,z \in A^*$ be fixed words. It is decidable, whether the pair (xu^kw, yv^kz) is a solution to I for some $k>0$, i.e., whether $p_1h(xu^kw)s_1 = p_2g(yv^kz)s_2$ and $xu^kw = yv^kz$ for some $k>0$.*

Proof. If (xu^kw, yv^kz) is a solution for some k, then

$$p_1h(xu^kw)s_1 = p_2g(yv^kz)s_2.$$

We obtain

$$|p_2|-|p_1|+|s_2|-|s_1|+|g(yz)|-|h(xw)| = k(|h(u)|-|g(v)|), \tag{4}$$

where the left-hand side does not depend on k. Now if this equation holds, then either $|h(u)|=|g(v)|$ or there is a unique k satisfying it. Therefore, we assume that $|h(u)|=|g(v)|$, since in the other case the uniqueness of k guarantees the decidability.

Now if (4) holds for some k, then it holds for all k. And consequently $|p_1h(xu^kw)s_1| = |p_2g(yv^kz)s_2|$ for all k, and the difference $|p_1h(xu^kw)|-|p_2g(yv^kz)|$ is constant. We may assume by symmetry that $|p_1h(x)|>|p_2g(y)|$. Let ℓ be the least number such that $|p_2g(yv^\ell)|>|p_1h(x)|$. Now, if $p_1h(xu^kw)s_1 = p_2g(yv^kz)s_2$ for some $k \geqslant \ell$, then also $p_1h(xu^\ell w)s_1 = p_2g(yv^\ell z)s_2$, since the possible overflow in $p_1h(xu^kw)$ and $p_2g(yv^kz)$ is unique by the length argument.

We have proved that if $|h(u)|=|g(v)|$ then there are at most $\ell+1$ different cases to check for solutions. Clearly these instance can be decided, since we have either one or $\ell+1$ k's to check whether $p_1h(xu^kw)s_1 = p_2g(yv^kz)s_2$ and $xu^kw = yv^kz$. And since this ℓ can be effectively found, we have proved the claim. □

For the case (0) we prove

Lemma 12. *Let I be an instance of the marked binary GPCP as above. It is decidable, whether I has a solution of type* (0). *Moreover, it is decidable, whether E_p contains a solution.*

Proof. If I has a solution w of type (0), then (w,w) is either in E_p or $w=u_1u_2=v_1v_2$, where (u_1,v_1) is the begin block and (u_2,v_2) is an end block. In other words, we need to check whether there exists a solution of these forms.

Consider the set E_p first. If E_p is finite than the decision is easy. Assume next that E_p is infinite, This implies that E_p is union of a finite set and finitely many extendible end block by Lemma 9(ii). Assume that an extendible end block (xu^kw, yv^k) or (xu^k, yv^kw) is in E_p. By Lemma 11, it is decidable whether the extendible end block contains a solution (note that $z=\varepsilon$), and since there may exist only finitely many extendible end blocks, we have completed the first part of the proof.

In the second case, the solutions of the form $w=u_1u_2=v_1v_2$, where (u_1,v_1) the begin block and (u_2,v_2) is an end block, also reduces to Lemma 11. Since the begin block is unique, the end blocks are the ones to consider. If the number of end blocks is finite, then the decision is easy. And if there is an extendible end block, say (xu^kw, yv^k), then we search for the solution in (u_1xu^kw, v_1yv^k), for $k\geqslant 0$, and this can be done by Lemma 11 (replace x by u_1x and y by v_1y). □

We can also prove that the solutions of type (1) can be effectively found. This is a consequence of Lemma 11.

Lemma 13. *It is decidable, whether an instance of the marked binary GPCP has a solution of type* (1).

Proof. Assume that only one block is used in the solution, i.e., the solution w is of the form $w=t_1t^{\ell}t_2=s_1s^{\ell}s_2$, where (t_1,s_1) is the begin block, (t,s) is a block for some letter a and (t_2,s_2) is an end block. Now for a fixed end block (t_2,s_2) the decision, whether there is a solution in the $(t_1t^{\ell}t_2, s_1s^{\ell}s_2)$, for $\ell>0$, can be done by Lemma 11.

In the solutions of type (1), the harder case seems to be the possible extendible end block. Assume therefore that there is an extendible end block (xu^kw, yv^k). By Lemma 9 and the fact that h and g are marked, the block (t,s) and this extendible end block necessarily begin with different letters.

We should now decide whether for some ℓ and k, $(t_1t^{\ell}xu^kw, s_1s^{\ell}yv^k)$ is a solution. But also this case reduces to Lemma 11. We have two cases, assume first that $t_1t^n\neq s_1s^n$ for all n. Then, for all n, there is a non-empty overflow r $(r=(t_1t^n)^{-1}s_1s^n$ or $r=(s_1s^n)^{-1}t_1t^n)$, if the words are comparable. If the words are not comparable for some n, then there is no solution for $\ell\geqslant n$. Now the first letter of r is what makes ℓ unique in this case. Assume that there is such an ℓ for which we have a solution. Then for $n=\ell$ the first letter of r is equal to the first letter of xu or yv, which is different from the first letter of t or s, respectively. Therefore, $t_1t^{\ell+1}$ and $s_1s^{\ell+1}$ are not comparable, and there cannot be solutions for the powers greater then this fixed ℓ.

We can effectively find such an ℓ, if we construct the pairs of words $(u_i,v_i)=(t_1t^n, s_1s^m)$, where $(u_0,v_0)=(t_1,s_1)$ and $(u_{i+1},v_{i+1})=(t_1t^{n+1}, s_1s^m)$, if $t_1t^n<s_1s^m$, and $(u_{i+1},v_{i+1})=(t_1t^n, s_1s^{m+1})$, if $s_1s^m<t_1t^n$. In other words, we construct the solution as the blocks. Now there are only finitely many different overflows in these pairs and if

suitable possible overflow exists we can find it. On the other hand, if no such overflow exists, then we will have a same overflow twice or the pair is no longer comparable. And since ℓ is unique, we may replace x with $t_1 t^\ell x$ and y with $s_1 s^\ell y$ in Lemma 11 and the decidability follows.

The other case is that, for some n, $t_1 t^n = s_1 s^n$ (and $h(t_1 t^n) = g(s_1 s^n)$). If there now is a solution, then necessarily $t^m x u^k w = s^m y v^k$ and $h(t^m x u^k w)s_1 = g(s^m y v^k)s_2$ for some m and k. Moreover, if $|t| \neq |s|$, then m is unique as ℓ in the previous case if it exists. And if $|t| = |s|$, then it is enough to check, whether $x u^k w = y v^k$ and $h(x u^k w)s_1 = g(y v^k)s_2$, which can be done by Lemma 11. □

As a corollary we get

Corollary 14. *The unary GPCP is decidable.*

Proof. Since all the solutions of the unary GPCP are of the type (0) or (1), the claim follows from Lemmas 12 and 13. □

From now on, we shall concentrate on type (2) solutions. Note that, since we are considering the binary GPCP, in this case no extendible end block may occur by Lemma 9.

Next we define the successors of the instances $I = ((p_1, p_2), h, g, (s_1, s_2))$ of the marked GPCP. Assume that the begin block (x, y) exists, and that $x \leqslant y$ or $y \leqslant x$ and set $p_1' = \varepsilon$, $p_2' = x^{-1}y$ or $p_1' = y^{-1}x$, $p_2' = \varepsilon$, respectively. Let (h', g') be the successor of (h, g) and let (u, v) be any end block of I. Then

$$I'(u, v) = ((p_1', p_2'), h', g', (s_1', s_2'))$$

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.

Lemma 14. *An instance $I = ((p_1, p_2), h, g, (s_1, s_2))$ has a solution if and only if the successor $I'(u, v) = ((p_1', p_2'), h', g', (s_1', s_2'))$ has a solution for some end block (u, v). Moreover, each solution w to I can be written as $w = xh'(w')u = yg'(w')v$, where w' is a solution of I', (x, y) is the begin block 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 $2 \leqslant i \leqslant k$, (u_1, v_1) is the begin block and (u_{k+1}, v_{k+1}) is an end block. Clearly $u_1 \leqslant v_1$ or $v_1 \leqslant u_1$ and $u_{k+1} \preccurlyeq v_{k+1}$ or $v_{k+1} \preccurlyeq u_{k+1}$. If the first cases hold, then $p_1' = \varepsilon$, $p_2' = u_1^{-1} v_1$ and $s_1' = \varepsilon$, $s_2' = v_{k+1} u_{k+1}^{-1}$ and $I'(u_{k+1}, v_{k+1}) = ((p_1', p_2'), h', g', (s_1', s_2'))$. Now

$$h'(a_2 \ldots a_k) = u_1^{-1} w u_{k+1}^{-1} = p_2' g'(a_2 \ldots a_k) s_2',$$

i.e., $I'(u_{k+1}, v_{k+1})$ has a solution $w' = a_2 \dots a_k$ and $w = u_1 h'(w') u_{k+1} = u_2 g'(w') v_{k+1}$. The other cases are similar.

Assume then that

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

has a solution w', i.e., $p'_1 h'(w') s'_1 = p'_2 g'(w') s'_2$. Then also $xh'(w')u = yg'(w')v$, where (x, y) is the begin block, and by Lemma 5(iii),

$$\begin{aligned} p_1 h(xh'(w')u)s_1 = p_1 h(x) h(h'(w')) h(u) s_1 &= p_2 g(y) g(g'(w')) g(v) s_2 \\ &= p_2 g(yg'(w')v) s_2 \end{aligned}$$

and so $xh'(w')u = yg'(w')v$ is a solution of I. □

Note that if the begin block does not exist, then there is no successors, and the only possible solutions are in E_p, but this case is decidable by Lemma 12. On the other hand, if the end words disappear, the instance is decidable by the next lemma.

Lemma 15. *Let $I = ((p_1, p_2), h, g, (s_1, s_2))$ be an instance of the binary marked GPCP. For the cases, where $s_1 = s_2 = \varepsilon$, the GPCP is decidable.*

Proof. Let # be a new symbol not in $\{0, 1\}$. Extend the morphisms h and g in a following way,

$$h(\#) = \# p_1 \quad \text{and} \quad g(\#) = \# p_2.$$

Now (h, g) is an instance of the marked PCP, and we can decide whether or not it has a solution beginning with #. □

6. Cycling instances

Let $I = ((p_1, p_2), h, g, (s_1, s_2))$ be an instance of the marked binary GPCP. By Lemma 14 we can reduce the instance I to its successors for all end blocks. The problem in this approach is that by Lemma 9, I potentially has infinitely many successors. However, if there is an extendible end block, then the solutions of the instance are necessarily of type (0) or (1) and these instances are decidable by Lemmata 12 and 13. Therefore we may concentrate on the case where there are no extendible end blocks and, since the unary GPCP is decidable, the alphabet size is 2.

By Lemma 14, I has a solution if and only if one of the successors has. If the suffix complexity goes to zero at some step, then we can always decide whether the successors have a solution (in these cases, $|h(a)| = |g(a)| = 1$ for all letters a). Thus we can solve the original problem. Otherwise, by Lemma 7, there is a number n_0 such that $(h_{i+d}, g_{i+d}) = (h_i, g_i)$ for each $i \geqslant n_0$, i.e., the morphisms start to cycle. Clearly to decide the marked binary GPCP it suffices to show how to solve these cycling instances.

By a *successor sequence* we mean a sequence

$$((p_1^{(0)},p_2^{(0)}),h_0,g_0,(s_1^{(0)},s_2^{(0)})),\dots,((p_1^{(i)},p_2^{(i)}),h_i,g_i,(s_1^{(i)},s_2^{(i)})),\dots \tag{5}$$

of instances of the marked GPCP such that each

$$I_{i+1}=((p_1^{(i+1)},p_2^{(i+1)}),h_{i+1},g_{i+1},(s_1^{(i+1)},s_2^{(i+1)}))$$

is a successor of $I_i=((p_1^{(i)},p_2^{(i)}),h_i,g_i,(s_1^{(i)},s_2^{(i)}))$.

Notice that if $\mathscr{I}_i=((p_1^{(i)},p_2^{(i)}),h_i,g_i,S_i)$, where S_i is the set of all pairs of the end words, is the set of all ith members in the successor sequences, we can assume that
(A) There is a begin block for all i, and
(B) $s_1s_2\neq\varepsilon$ for each $(s_1,s_2)\in S_i$.

For, if condition (A) does not hold, we know that no instance in $\mathscr{I}_{i+1}$ is defined and the only possible solutions are in E_p and if (B) is not satisfied by an instance, then that instance reduces to the marked PCP, which is decidable by Lemma 15.

We shall next show how to treat the instances that begin to cycle, i.e., for which there exists an integer d such that for all successor sequences (5) $(h_i,g_i)=(h_{i+d},g_{i+d})$ for all $i\geqslant 0$. We shall call such an instance I_0 a *loop instance*, and d the *length of the loop*.

Notice that we always choose such d that also $(p_1^{(i)},p_2^{(i)})=(p_1^{(i+d)},p_2^{(i+d)})$. The fact that such d exists, can been seen for example from the construction in the proof of Lemma 15. In fact, let us forget the end words and consider only the begin words and the morphisms in an instance. Let $I_\#=(h,g)$ be the instance of the marked PCP defined in the proof of Lemma 15. Since we assumed that the morphisms are cyclic and there always exists a begin block, then in the successor sequence $I_\#$ we shall eventually get a same instance twice by Lemma 7. Therefore, also the begin words are cyclic.

Notice that, since, by Lemma 8, the alphabet size does not decrease, there is a block for both letters in $\{0,1\}$. In particular, there cannot be extendible end blocks.

Lemma 16. *Assume that the instances cycle as in* (5) *and that a solution exists. Then we have two cases*:
(i) *If* $p_1^{(0)}=\varepsilon=p_2^{(0)}$ *then the minimal solution of* I_0 *is* w, *where the initial letter* a *of* w *satisfies* $h_0(a)\neq g_0(a)$. *Hence* $h_i(a)\neq g_i(a)$ *for all* $i\geqslant 0$.
(ii) *If* $p_1^{(0)}\neq p_2^{(0)}$, *then a minimal solution* w *does not have a prefix* u *such that* $p_1^{(0)}h(u)=p_2^{(0)}g(u)$.

Proof. For case (i), if $h_0(a)=g_0(a)$ for the initial letter a, then $a^{-1}w$ is a shorter solution. Therefore $h_0(a)\neq g_0(a)$, and, if $h_i(a)=g_i(a)$, for some $i\geqslant 0$, then it is true also for all $j>i$ and $|h_j(a)|=1=|g_j(a)|$. But since I_0 is a cycling instance and $I_0=I_k$ for some $k\geqslant i$, then also $h_0(a)=g_0(a)$, a contradiction.

For case (ii), if there exists such a u, then $p_1^{(k)}=\varepsilon=p_2^{(k)}$ for some k, and therefore the same holds for all $j\geqslant k$. But, since the instance is cycling, $(p_1^{(0)},p_2^{(0)})=(p_1^{(t)},p_2^{(t)})$ for some $t\geqslant k$, and we get a contradiction, since $p_1^{(0)}\neq p_2^{(0)}$. □

Hereafter we will assume that $p_1^{(0)} \neq p_2^{(0)}$, since case (i) reduces to the (cycling) instances $((h(a), g(a)), h_0, g_0, (s_1^{(0)}, s_2^{(0)}))$ for each a such that $h_0(a) \neq g_0(a)$.

We would like to have some upper bound for the lengths of the new end blocks in the loop (5). We demonstrate that there is a limit number L such that, if a solution exists, then the minimal solution is found in some sequence (5) shorter than L. Moreover, this limit can be effectively found, and the main result follows from this.

In what follows, we assume that $I = ((p_1, p_2), h, g, (s_1, s_2))$ has a minimal solution w such that if $p_1 h(u) = p_2 g(u)$, then u is not a prefix of w. Then this minimal solution is unique, since we assumed that $p_1 \neq p_2$ and 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

$$I_{i+1} = I_i(u_i, v_i), \tag{6}$$

where (u_i, v_i) is the end block of the minimal solution of I_i. This successor sequence is called *the branch of the 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 $I_i = ((p_1^{(i)}, p_2^{(i)}), h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$ be an instance in the branch of the minimal solutions and w_i be the minimal solution of I_i. Recall that we permanently assume that $s_1^{(i)} s_2^{(i)} \neq \varepsilon$ and $p_1^{(0)} \neq p_2^{(0)}$, which implies that also $p_1^{(i)} \neq p_2^{(i)}$ for each i.

Lemma 17. *Let w_i be the minimal solution of I_i and let $(h_i, g_i) = (h_{i+d}, g_{i+d})$, $(p_1^{(i)}, p_2^{(i)}) = (p_1^{(i+d)}, p_2^{(i+d)})$ for each i. Then $w_{i+d} \leqslant w_i$ but $w_{i+d} \neq w_i$ for each i.*

Proof. The instances

$$I_d = ((p_1^{(i)}, p_2^{(i)}), h_i, g_i, (s_1^{(i)}, s_2^{(i)})) \quad \text{and} \quad I_{i+d} = ((p_1^{(i)}, p_2^{(i)}), h_i, g_i, (s_1^{(i+d)}, s_2^{(i+d)}))$$

share the begin block and the marked morphisms, so clearly $w_i \leqslant w_{i+d}$ or $w_{i+d} \leqslant w_i$, since the minimal solutions cannot have u, such that $p_1^{(\ell)} h_\ell(u) = p_2^{(\ell)} g_\ell(u)$, as a prefix (recall that $p_1^{(i)} \neq p_2^{(i)}$)). If w is a minimal solution to some instance I, then by Lemma 14, there is a solution w' to the successor of I such that $w = xh'(w')u = yg'(w')v$. Since $s_1^{(i)} s_2^{(i)} \neq \varepsilon$, then also $uv \neq \varepsilon$ (and $p_1^{(i)} \neq p_2^{(i)}$, then $xy \neq e$) and consequently $|w| > |w'|$, because the morphisms are nonerasing. Hence $|w_{i+1}| + 1 \leqslant |w_i|$. Inductively, $|w_{i+t}| + t \leqslant |w_i|$ for all t, which proves the claim. □

As a byproduct we obtain

Lemma 18. *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. □

Let $I = ((p_1, p_2), h, g, (s_1, s_2))$ be an instance of the marked binary GPCP. We assume now, by symmetry, that $s_2 \leqslant s_1$. An end block (u, v) of the instance I satisfies the equation

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

where $s = s_1 s_2^{-1}$. 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 19. *Let $I_i = ((p_1^{(i)}, p_2^{(i)}), h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$ 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 $p_1^{(i)} h_i(w_{i+d}) s_1^{(i+d)} = p_2^{(i)} g_i(w_{i+d}) s_2^{(i+d)}$ is a prefix of $p_1^{(i)} h_i(w_i)$ and $p_2^{(i)} g_i(w_i)$.*

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 17, $w_d \leqslant w_0$. Therefore, $p_1^{(0)} h_0(w_d) \leqslant p_1^{(0)} h_0(w_0)$ and $p_2^{(0)} g_0(w_d) \leqslant p_2^{(0)} g_0(w_0)$. We shall next prove that $|h_0(w_d) s_1^{(d)}| \leqslant |h_0(w_0)|$. Assume on the contrary that $|h_0(w_d) s_1^{(d)}| > |h_0(w_0)|$. By the proof of Lemma 14,

$$w_0 = x_1 h_1(x_2 h_2(\ldots h_{d-1}(x_{d-1} h_d(w_d) u_{d-1}) \ldots u_2) u_1$$

and therefore

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

Hence

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

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

$$p_1^{(0)} h_0(w_d) s_1^{(d)} = p_2^{(0)} g_0(w_d) s_2^{(d)} = p_2^{(0)} g_0(w_d) \leqslant p_2^{(0)} g_0(w_0)$$

and since $p_1^{(0)} h_0(w_0)$ and $p_2^{(0)} g_0(w_0)$ are comparable and $|h_0(w_d) s_1^{(d)}| \leqslant |h_0(w_0)|$, necessarily $p_1^{(0)} h_0(w_d) s_1^{(d)} \leqslant p_1^{(0)} h_0(w_0)$, too (Fig. 3). □

The previous lemma will be used in the proof of 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 $((p_1, p_2), h, g, (s_1, s_2))$ is a word $z = g(v_1) g(v_2) \ldots g(v_k)$ such that

(1) $v_1 v_2 \ldots v_k$ 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 $g(v_1) \ldots g(v_{k-1})$,

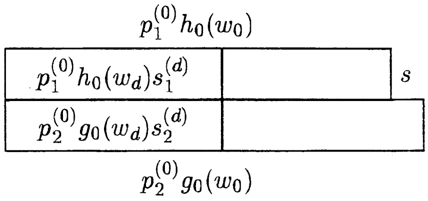

Fig. 3. Prefix property.

(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 if it exists. Hence we can define the integer k to be the *g-covering length* of the occurrence of u (in $g(w)$).

The h-block covering is defined analogously.

In what follows, we shall concentrate on the coverings and covering lengths of the end words s_2 and s_1 as they occur as a factor in $g(w)$ and $h(w)$.

Lemma 20. *Let $I_i = ((p_1^{(i)}, p_2^{(i)}), h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$ be the branch of the minimal solutions of a cycling instance having loop length d, and w_i be the minimal solution of I_i. Then the h_i and g_i coverings of $s_1^{(i)}$ and $s_2^{(i)}$ exist for all $i \geqslant d$.*

Proof. By Lemma 19, for $i \geqslant d$, the words $s_1^{(i)}$ and $s_2^{(i)}$ are factors in $h_i(w_{i-d})$ and $g_i(w_{i-d})$, since $p_1^{(i)}h_i(w_i)s_1^{(i)}$ and $p_2^{(i)}g_i(w_i)s_2^{(i)}$ are prefixes of $p_1^{(i)}h_i(w_{i-d})$ and $p_2^{(i)}g_i(w_{i-d})$. □

Note that in the next lemma the occurrences of $s_1^{(i)}$ and $s_2^{(i)}$ in $h_i(w_{i-d})$ and $g_i(w_{i-d})$ consider are exactly the suffixes in $p_1^{(i)}h_i(w_i)s_1^{(i)}$ and $p_2^{(i)}g_i(w_i)s_2^{(i)}$.

Lemma 21. *Let $I_i = ((p_1^{(i)}, p_2^{(i)}), h_i, g_i, (s_1^{(i)}, s_2^{(i)}))$ be the branch of the minimal solutions of a cycling instance having loop length d, and w_i be the minimal solution of I_i. For all $i \geqslant d$,*

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

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

Proof. By Lemma 20 the coverings exist. We will prove only case (i), the other one is analogous. To simplify the notations, we denote $I = I_i = ((p_1, p_2), h, g, (s, \varepsilon))$ and $I_{i+1} = I'$. Now either $I' = ((p_1', p_2'), h', g', (s', \varepsilon))$ or $I' = ((p_1', p_2'), h', g', (\varepsilon, s'))$. 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 $a \leqslant u$ be the first letter of u. Since $h(u)s = g(v)$ and $u \preccurlyeq v$ or $v \preccurlyeq u$, we have two cases to consider.

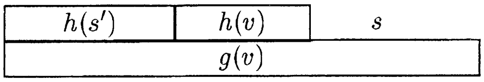

Fig. 4. Picture of case (1).

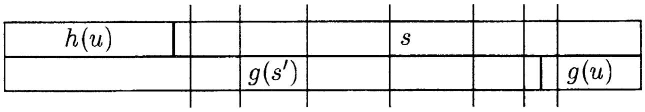

Fig. 5. Block covering of $h(u)s = g(s')g(u)$. The vertical lines illustrate the block covering.

(1) If $u = s'v$, then $|s'| \leqslant |u|$. 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 $s' \leqslant h'(a)$, i.e. the h'-covering length of s' is 1, see Fig. 4.

(2) Assume then that $v = s'u$, and so $I' = ((p_1', p_2'), h', g', (\varepsilon, s'))$. 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 case (1), $h(u)$ is covered by a single g-block longer than $h(u)$, see Fig. 5 for an illustration). We show then that 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 $w = dh'(w') = eg'(w')s'$, where (d, e) is a beginning block for I, satisfies

$$p_1h(w)g(s') = p_1h(d)hh'(w')g(s') = p_2g(e)gg'(w')g(s') = p_2g(w)$$

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 $dh'(w') = eg'(w')s'$ cutting s' can be mapped injectively to block borderlines in $p_1h(w)g(s') = p_2g(w)$ that cut $g(s')$, see Fig. 6. Let $y' \leqslant w'$ be a word that determines a block borderline in $dh'(w') = eg(w')s'$ that cuts s'. That is, $dh'(y') = eg'(z')$ for some word z' and $eg'(w') \leqslant dh'(y')$. Then $z' = w'x'$ for some x' that satisfies $g'(x') \leqslant s'$. Now the word $y = eg'(z')$ is a prefix of w, since

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

and g is marked. But y also determines a block borderline in the word $p_2g(w) = p_1h(w)$ $g(s')$, since $p_2g(y) = p_2g(e)g(g'(z')) = p_1h(d)h(h'(z'))$. This borderline cuts $g(s')$, because

$$p_2g(y) = p_2g(e)g(g'(z')) = p_1h(d)h(h'(z'))$$
$$= p_1h(d)h(h'(w'))h(h'(x')) = p_1h(d)h(w)h(h'(x'))$$

and hence $p_1h(w) \leqslant p_2g(y)$. Notice finally that the word y determines z' uniquely, since g is injective and z' determines y' by $eg'(z') = dh'(y')$. (Recall that h is injective.) □

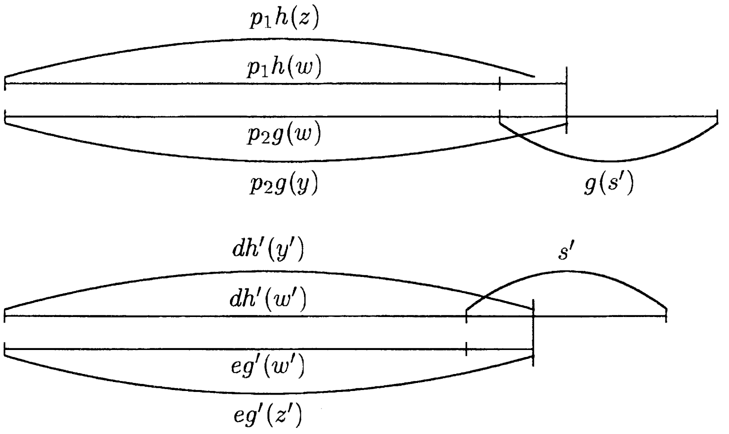

Fig. 6. Relation between $g(s')$ and s'.

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 $\mathscr{I}_d$ found by the first d reductions. If I_0 has a solution then there is a unique $I \in \mathscr{I}_d$ 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 $\mathscr{I}_d$. It now follows by Lemma 21 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 $i \geqslant d$, the end words of I_i are not longer than $M(\sigma(I)+1)$.*

7. Decidability results

Now we are ready to prove our main results.

Theorem 3. *The binary marked GPCP is decidable.*

Proof. We have already proved that the marked binary GPCP is decidable in the unary case and the solutions of type (0) and (1) can be found. It remains to be shown how to find type (2) solutions, i.e., how to solve the binary marked GPCP for the cycling instances I_0.

A cycling instance has the blocks for the both letters and $p_1^{(i)} \neq p_2^{(i)}$ for all successors. In particular, there are no extendible end blocks and only finitely many successors. The successor relation naturally defines a tree $\mathscr{T}$ 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 $\mathscr{T}$ 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 we check whether there are solutions of type (0), (1) or an end block $(u,v) \in E_p$ such that $u = v$. And for all vertices $I = ((p_1, p_2), h, g, (s_1, s_2))$ that have $s_1 s_2 = \varepsilon$, we can always decide if they have a solution by Lemma 15. 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 $\mathscr{T}$ is more specific: Only the successors $I = ((p_1, p_2), h, g, (s_1, s_2))$ that satisfy $|s_1 s_2| \leqslant M(\sigma(I_0) + 1)$ 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 $\mathscr{T}$ will eventually contain an instance twice, thus I_0 has no solution by Lemma 18, unless some vertex $I = ((p_1, p_2), h, g, (s_1, s_2))$ has a solution for some vertex $(u,u) \in E_p$. □

As we saw in Section 3, the binary PCP is decidable if and only if the binary marked GPCP is. Therefore, Theorem 3 has the following corollary.

Theorem 4. *The binary PCP is decidable.*

Proof. First, if one of the morphisms is periodic, it can be decided by Theorem 1, and if the instance is marked, then it can be decided by Theorem 2. Otherwise we construct the equivalent instance of the binary marked GPCP.

The binary marked GPCP is decidable by Theorem 3. The decision procedure achieved reduces an instance of the binary marked GPCP to finitely many simpler equivalent instances. By continuing this reduction to each reduced instance we create a successor tree, where the decision is done in each path separately according to the following seven rules:

(i) If we get unary successors, then we can decide these successors by Corollary 14.

(ii) If we get an extendible end block, then the solutions are of the type (0) or (1) and these cases can be decided by Lemmata 12 and 13, respectively.

(iii) If we get end block (u,u), then $s_1' = s_2' = \varepsilon$ for some successor, and this is decidable by Lemma 15.

(iv) If we get an instance which already occurred in the path, then the instances in this path cannot have a solution by Lemma 8.

(v) If the lengths of the end words break the computable limit $M(\sigma(I_0) + 1)$, then we do not have to continue this branch, since it is not in the branch of minimal solutions by Corollary 2.

(vi) If we get an end block (u,u) in E_p, then we have a solution. Also, if there is no begin block then the possible solutions are in E_p. These cases are decidable by Lemma 12.

(vii) If there are no end blocks, then there are no solutions. □

8. Conclusions and open problems

We have proved that in the binary case the Post Correspondence Problem is decidable. Our solutions are based on the construction of the successors, which is equivalent to the original instance in the decidability sense. Then after doing this reduction sufficiently many times we obtain instances, where the decision is easy to do.

We note that an instance of the binary GPCP can also be reduced to an instance of the binary *marked* GPCP using almost similar arguments as in Lemma 3. Therefore we also gave a new proof to the decidability of the binary GPCP.

As open problems we state the following immediate questions:

- Decidability of the PCP and the GPCP in the ternary case, i.e., $|A| = 3$.
- Decidability of the *strongly 2-marked* PCP. A morphism is strongly 2-marked if each image of a letter has a unique prefix of length 2. See also [4].

There is also a very important open question considering the form of the solutions of the binary PCP.

- Let (h,g) be an instance of the binary PCP, where h is nonperiodic. Is it true that all the solution are from the set $\{u,v\}^+$ for some, possibly equal, words u and v? See also [5].

Acknowledgements

Authors would like to thank the anonymous referees for useful comments.

References

[1] J. Berstel, Transductions and Context-Free Languages, Teubner, Stuttgart, 1979.

[2] A. Ehrenfeucht, J. Karhumäki, G. Rozenberg, The (generalized) Post correspondence problem with lists consisting of two words is decidable, Theoret. Comput. Sci. 21 (2) (1982) 119–144.

[3] V. Halava, T. Harju, M. Hirvensalo, Generalized post correspondence problem for marked morphisms, Internat. J. Algebra Comput. 10(6) (2000) 757–772.

[4] V. Halava, M. Hirvensalo, R. de Wolf, Marked PCP is decidable, Theoret. Comput. Sci. 255 (1–2) (2001), 193–204.

[5] T. Harju, J. Karhumäki, Morphisms, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, vol. 1, Springer, Berlin, 1997, pp. 439–510.

[6] T. Harju, J. Karhumäki, D. Krob, Remarks on Generalized Post Correspondence Problem, Lecture Notes in Computer Science, vol. 1046, Springer, Berlin, 1996, pp. 39–48.

[7] Y. Matiyasevich, G. Sénizergues, Decision problems for semi-Thue systems with a few rules, Proc. 11th IEEE Symp. on Logic in Computer Science, 1996, pp. 523–531.

[8] E.L. Post, A variant of a recursively unsolvable problem, Bull. Amer. Math. Soc. 52 (1946) 264–268.

[9] A. Salomaa, Formal Languages, Academic Press, New York, 1973.